

Artificial Intelligence Ethics through the Lens of Maqāṣid al-Sharī'ah: Constructing a Normative Framework for Civilized Technology Governance

Maman Supardi¹, Hilmiy Hanif², Mursyid Rahman³, Ahmad Fadhil Arham⁴, Andi Muhammad Ali Amiruddin⁵

¹⁻⁵ Universitas Islam Negeri Alauddin Makassar

Article Info

Article history:

Received Jul, 2026

Revised Jul, 2026

Accepted Jul, 2026

Keywords:

Artificial Intelligence Ethics

Maqāṣid Al-Sharī'ah

Islamic Ethics

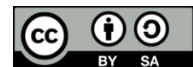
Technology Governance

Normative Framework

ABSTRACT

The accelerating deployment of artificial intelligence (AI) across public and private life has outpaced the moral vocabularies available to evaluate it. Prevailing AI-ethics frameworks, articulated largely within Western traditions, converge on principles such as fairness, transparency, and accountability, yet remain thin on the spiritual and communal dimensions of moral judgement in Muslim societies. This study asks how maqāṣid al-sharī'ah, the higher objectives of Islamic law, may be operationalised as an evaluative instrument for AI governance rather than a declarative ideal. Employing qualitative library research, the study analyses primary Islamic sources alongside peer-reviewed scholarship from 2021 to 2026 through content analysis and philosophical-normative reasoning. The analysis maps four recurring problems (the erosion of privacy, algorithmic bias, opacity and diffuse accountability, and the attrition of human moral agency) and evaluates each against the five essential objectives (al-darūriyyāt al-khams). The resulting framework translates every objective into governance criteria and correlates them with globally recognised AI principles, yielding an eleven-criterion instrument together with an ordering rationale for cases of value conflict. The study contributes an operational, value-based model that complements rather than displaces existing regulatory approaches, offering developers, regulators, and Shariah boards a design vocabulary for anticipating harm before deployment.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Name:

Institution Address:

e-mail:

1. INTRODUCTION

Artificial intelligence has moved, within little more than a decade, from a specialised research programme to a general-purpose technology embedded in healthcare, finance, education, and public administration [1], [2]. Muslim-majority societies participate

fully in this diffusion: algorithmic systems now mediate Shariah screening in Islamic financial institutions, structure everyday access to religious information, and increasingly shape the informational environment in which ordinary believers

form judgements [3],[4]. The stakes are considerable in populous Muslim-majority states such as Indonesia, where rapid digital adoption, a large young online population, and an active religious public sphere converge, and where national data-protection law has only recently begun to address the resulting exposures [42]. The benefits are neither trivial nor speculative. Yet the very capacities that generate them, namely large-scale data ingestion, statistical inference over human behaviour, and progressively autonomous decision-making, also produce harms that are diffuse, delayed, and difficult to attribute to any single actor.

Systematic reviews conducted in markedly different domains converge on a common inventory of concerns. Studies of nursing practice, dentistry, and accounting each independently identify client autonomy, data protection, algorithmic bias, and unclear accountability as the dominant ethical fault lines [5], [6], [7]. That such disparate fields arrive at the same catalogue suggests these are structural properties of the technology rather than sector-specific anomalies. Generative systems have since sharpened the picture, adding the industrial-scale production of synthetic and misleading content to an already crowded risk landscape [8], [9], [10], and raising distributive questions about who bears the costs of automation and who captures its gains [11].

The dominant institutional response has been a proliferation of principles. Since 2016, governments, standards bodies, and intergovernmental organisations, among them the OECD, the IEEE, and the European Union through its Artificial Intelligence Act, have issued overlapping declarations committing to fairness, transparency, accountability, privacy, and human oversight [12], [13]. Comparative analyses of these documents reveal a striking convergence at the level of stated values alongside persistent divergence in enforceability, institutional backing, and cross-border coordination [12], [14]. Two limitations bear directly on the present inquiry. First, the discourse is grounded predominantly in utilitarian,

deontological, and virtue-based traditions, and tends to bracket the spiritual and communal dimensions of moral reasoning that remain decisive for much of the world's population [15], [50]. Second, principle documents are frequently abstract: they enumerate values without specifying how those values are to be ranked when they conflict, as in practice they routinely do [1].

To address these two deficits, this article turns to a tradition that has for centuries supplied a structured method for weighing competing goods in novel circumstances. It is against this background that *maqāṣid al-sharīʿah* offers a distinctive evaluative resource. Systematised most influentially in al-Shāṭibī's *al-Muwāfaqāt*, the tradition organises the aims of the law around five essential interests (*al-ḍarūriyyāt al-khams*): the preservation of religion (*ḥifẓ al-dīn*), life (*ḥifẓ al-nafs*), intellect (*ḥifẓ al-ʿaql*), lineage (*ḥifẓ al-nasl*), and property (*ḥifẓ al-māl*), graded through the categories of necessities, needs, and enhancements [16], [17]. Far from being a purely historical construct, the framework has been revived and extended to contemporary problems including environmental sustainability [18], humanitarian crisis response [19], and the protection of digital rights and privacy [20]. Its anticipatory, objective-driven character suits emerging technologies whose harms cannot be adjudicated by case-specific precedent alone [21].

A growing body of scholarship has begun to bring this tradition into contact with AI. Foundational contributions clarified the general relevance of the *maqāṣid* to AI ethics, particularly in relation to privacy and manipulation [22], while subsequent studies proposed value-mappings in which the five essentials serve as assessment criteria [23], [24], [25]. Others have advanced integrative or procedural proposals: eight Islamic guidelines combining the *maqāṣid* with the principle of moderation (*wasatīyyah*) [26], frameworks oriented toward sustainable digital-transformation governance [27], and multi-step institutional models incorporating impact assessment and oversight councils

[28]. Sector-specific applications have followed in business [29] and Islamic finance [4], alongside a parallel literature examining AI's implications for religious authority, fatwa issuance, and the integrity of juristic reasoning [3], [30], [31].

Despite this momentum, a gap persists. Recent surveys observe that Islamic AI-ethics scholarship remains predominantly normative and declarative, with only limited operationalisation of the *maqāṣid* as a concrete evaluative instrument and a notable scarcity of integrative models connecting Islamic ethics with mainstream AI-ethics frameworks [15]. Empirical evidence reinforces the diagnosis: in a systematic comparison of AI-generated and human-authored fatwas, consideration of *maqāṣid al-sharīʿah* was precisely the dimension on which automated outputs performed worst [32]. The tradition, in short, is frequently invoked in principle but rarely rendered into criteria that a developer, regulator, or Shariah board could actually apply. This study addresses that gap by constructing a normative framework that operationalises the five objectives as evaluative criteria for AI governance and articulates their correspondence with globally recognised principles. It pursues three questions: (1) What are the principal ethical challenges posed by contemporary AI? (2) How do the five objectives of *maqāṣid al-sharīʿah* evaluate these challenges? (3) What normative framework for civilised technology governance follows from that evaluation? The article proceeds through a review of three relevant literatures, an account of method, a four-part analysis, and a concluding synthesis.

2. LITERATURE REVIEW

Three bodies of scholarship converge on the problem addressed here: the contemporary literature on AI ethics and governance, the theory of *maqāṣid al-sharīʿah* and its modern revival, and the emerging intersection between Islamic normative thought and digital technology. Each is

reviewed in turn, with attention to what it settles and what it leaves open.

2.1 Contemporary AI Ethics: Converging Principles, Diverging Enforcement

The field of AI ethics is organised around a compact and remarkably stable set of principles. Analyses of prominent principle documents identify recurring themes, namely privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and the promotion of human values, with several of these appearing in almost every document surveyed [12]. Socio-technical surveys structure the field along similar lines, treating privacy and data protection, transparency and explainability, fairness and equity, responsibility and accountability, and environmental impact as the constitutive concerns of trustworthy AI [1].

Translation into binding instruments has proceeded unevenly. The European Union's Artificial Intelligence Act represents the most comprehensive attempt to date, distinguished by its legally binding character, extraterritorial reach, and institutional innovations, yet criticised for weak enforcement, absent remedy mechanisms, and institutional ambiguity [12]. Elsewhere, governance has taken the form of testing toolkits, voluntary blueprints, and standard-setting by professional bodies, each articulating human-centred, explainable, and fair

operation as design desiderata [2]. Policy analyses of generative systems specifically note that labour-market disruption, copyright uncertainty, and the amplification of societal bias exceed the reach of instruments designed for earlier technological generations [13].

Two critical observations recur. The first concerns trade-offs: strengthening privacy can degrade fairness, and constructing fair models may itself require demographic data that privacy constraints withhold, so that no checklist can guarantee trustworthiness across all dimensions simultaneously [1]. The second concerns legitimacy: survey evidence from rival AI powers indicates that public trust in governance actors tracks geopolitical alignment rather than institutional design alone, complicating the prospect of a single global regime [14]. Sectoral reviews add a third caution, finding existing frameworks insufficiently tailored to the contexts in which AI is actually deployed, and institutional readiness (infrastructure, literacy, oversight) as consequential as principle articulation [5], [33].

2.2 *Maqāṣid al-Sharī'ah: Structure and Contemporary Revival*

Maqāṣid al-sharī'ah denotes the higher objectives that the Lawgiver intends the law to realise. Al-Shāṭibī's *al-Muwāfaqāt* distinguishes the objectives of the Lawgiver from those of the human subject and grades the latter along three scales: the essential (*darūriyyāt*), the necessary (*ḥājīyyāt*), and the complementary (*taḥsīniyyāt*)

[16]. Andalusian scholarship, including the contributions of al-Ghazālī and Ibn Rushd, supplied much of the conceptual groundwork and emphasised the adaptability of the Shariah across contexts, balancing reason, ethics, and textual fidelity [17].

A methodological point deserves emphasis because it constrains the present analysis. The five essentials were not derived by pairing each objective with a single proof-text; al-Shāṭibī established them through *istiqrā'*, an inductive survey of the legal corpus as a whole, supplemented by attention to the 'illah underlying commands and prohibitions [16]. Any contemporary application must therefore reason from the purposes that texts disclose rather than from isolated citation, a discipline the present study observes when invoking Qur'anic material below.

The tradition has been actively extended to modern problems. Scholarship on *ijtihād* argues that *maqāṣid*-oriented reasoning is indispensable for novel cases (*nawāzil*), enhancing the flexibility of the law without loosening its anchorage [21]. Environmental scholarship has re-theorised sustainability within the framework, going so far as to propose the preservation of ecological order as an additional objective [18]. Humanitarian analyses treat the five objectives as a guiding structure for questions of inequality and social unrest [19]. Most pertinently for this study, legal scholarship has begun mapping *ḥifẓ al-'aql* and *ḥifẓ al-māl* onto data-protection and

privacy regulation [20], while a parallel line of argument advances *ḥifẓ al-ʿird*, the protection of dignity and reputation, as an under-explored objective of particular relevance to digital surveillance [34], and applies *maqāṣid* reasoning to questions of digital inclusion [35].

2.3 *The Emerging Intersection of Islamic Ethics and Artificial Intelligence*

Explicitly *maqāṣid*-based engagement with AI is recent but expanding rapidly. Early contributions established the general relevance of the framework to questions of privacy and manipulation and called for alternative ethical paradigms grounded in Islamic thought [22]. Two contributions have been especially influential in framing the field: a virtue-based Islamic approach that situates AI evaluation within the *maqāṣid* while contesting the market logics driving development [49], and a pluralist argument for ethical benchmarking that advances *maṣlaḥa* as the normative guide and shows that Islamic jurisprudence accommodates both welfarist and duty-based readings of it [50]. Parallel work has extended the tradition to the moral status of AI-mediated labour [51]. A second wave produced value-mappings in which the five essentials operate as evaluative criteria: *ḥifẓ al-ʿaql* underwriting resistance to misinformation and *ḥifẓ al-māl* guiding equitable financial technology [23], with parallel proposals emphasising the anticipatory character of *maqāṣid* reasoning relative to

case-specific juristic response [24], [25]. More governance-oriented work has followed, including guidelines pairing the *maqāṣid* with *wasatiyyah* [26], frameworks for sustainable digital transformation [27], and models specifying societal impact assessment, compliance standards, and oversight councils [28]. Sectoral applications now extend to business [29] and Islamic finance, where AI is analysed as a means (*wasīlah*) toward Shariah objectives [4].

A second strand examines AI's implications for religious authority. Analyses of automated fatwa systems conclude that machine-generated rulings cannot substitute for qualified scholars, lacking *ijtihād*, intention (*niyyah*), and contextual sensitivity [3], and theological work has probed deeper questions concerning knowledge production and the Qur'anic claim of inimitability [30]. Empirical evaluation reinforces these reservations: across matched pairs of AI-generated and human-authored fatwas, automated outputs performed adequately on retrieval tasks but poorly on depth of reasoning, contextual sensitivity, and *maqāṣid* consideration, prompting the proposal of multi-pillar governance models built on scholarly supervision and verification [32]. Related studies converge on human-in-the-loop control, verification of the operative cause (*taḥqīq al-manāt*), and provenance protocols as the conditions under which algorithmic assistance remains legitimate [31], [36], [37],

and caution that digital mediation of tafsīr may erode traditional interpretive authority even as it widens access [38].

A third strand addresses digital privacy directly from Islamic sources. Studies ground personal-data protection in *ḥifẓ al-ʿird* and the prohibition of *tajassus*, arguing that Islamic law treats privacy as a dignity interest rather than a merely proprietary one [39], [40]. Jurisprudential analyses derive requirements of informed consent, data integrity, and accountability from Islamic ethical principles and find them broadly convergent with contemporary data-protection standards [41], while comparative legal work examines the integration of these principles into national data-protection regimes [42], [43]. Taken together, the three strands establish that the *maqāṣid* can complement mainstream AI ethics rather than merely oppose it. What remains absent is a framework that renders the objectives into criteria specific enough to guide assessment, the task the present study undertakes.

3. METHODS

This study is qualitative and conceptual in design, employing library research (*baḥṭh maktabī*) as its principal mode of inquiry. The choice follows from the research questions: the object of analysis is a normative structure and its application to a class of technologies, not the behaviour of a measurable population. The study therefore proceeds by reasoned interpretation of texts rather than by fieldwork or statistical inference, a design consistent with prevailing practice in *maqāṣid*-based scholarship on emerging technology [23], [25], [27].

Sources were drawn from two tiers. Primary sources comprise the Qurʾān, the Sunnah, and classical works of *maqāṣid* theory, accessed through established scholarly editions and, for Qurʾanic material, a single reference translation to preserve terminological consistency [44]. Secondary sources comprise peer-reviewed articles and policy documents published between 2021 and 2026, retrieved from indexed academic databases using combinations of the terms artificial intelligence ethics, AI governance, *maqāṣid al-sharīʿah*, Islamic ethics, and digital privacy, combined through Boolean strings of the form (“artificial intelligence” OR “generative AI”) AND (“*maqāṣid al-sharīʿah*” OR “Islamic ethics”) AND (“governance” OR “privacy”). Retrieval was conducted through academic search platforms indexing peer-reviewed literature across publishers, supplemented by citation tracing from the most frequently cited studies in the resulting set. Inclusion required substantive engagement with either the ethical evaluation of AI or the contemporary application of *maqāṣid* reasoning; publications that mentioned either topic only incidentally were excluded. Priority was given to recent work in order to capture the generative-AI phase of the debate, which differs materially from earlier discussions of automation.

Analysis proceeded in three stages. First, content analysis of the secondary corpus was used to identify recurring categories of ethical harm, which were consolidated into four problem clusters. Second, each cluster was examined through philosophical-normative reasoning against the five *ḍarūriyyāt*, with Qurʾanic and juristic material adduced to establish the objective at stake rather than to settle particular rulings; consistent with al-Shāṭibī's inductive method, textual citation is treated as evidence of purpose, not as isolated proof [16]. Third, the resulting judgements were synthesised into governance criteria and correlated with the principle vocabulary of mainstream AI ethics. Interpretive validity was pursued through triangulation: every judgement concerning an objective was checked against at least two

independent scholarly treatments of that objective, and divergences were retained and discussed rather than resolved by preference. The study's conclusions are advanced as a conceptual framework requiring subsequent empirical testing rather than as findings of fact.

4. RESULTS AND DISCUSSION

4.1 Mapping the Ethical Terrain of Contemporary Artificial Intelligence

Content analysis of the reviewed corpus yields four problem clusters that recur across sectors and jurisdictions. They are analytically distinct but causally entangled, and it is their entanglement that defeats piecemeal remedies.

The first cluster concerns the erosion of privacy. Contemporary models depend on the accumulation and processing of personal information at a scale that renders meaningful consent largely fictional; data collected for one purpose is routinely repurposed for another, and individuals lose practical control over who holds information about them [2]. Professional studies confirm that the handling of sensitive data ranks among the foremost concerns of practitioners themselves [7]. The harm is not confined to material loss: surveillance infrastructures expose persons to reputational damage and unwarranted suspicion even when no data breach occurs [34].

The second cluster concerns algorithmic bias. Systems trained on historical data absorb the disparities encoded within it and reproduce them at scale, with documented effects on hiring, credit, and clinical decision-making [1], [2]. Bias is not a defect of implementation alone but a structural consequence of learning from an unequal world, which is why technical mitigation has proved insufficient without institutional oversight [33]. Generative systems extend the problem into representation itself, distributing stereotypes through the content they produce [10].

The third cluster concerns opacity and diffuse accountability. Deep models

function as black boxes whose internal reasoning resists translation into human-intelligible terms, so that even developers cannot fully explain particular outputs [2]. Where explanation fails, responsibility disperses: when harm occurs, it becomes unclear whether fault lies with the designer, the deployer, the data supplier, or the user. Empirical work in professional settings finds that accountability depends less on formal rules than on organisational mechanisms and individual familiarity with the systems in question [7].

The fourth cluster concerns the attrition of human moral agency. As consequential judgements are delegated to automated systems, the human capacity for deliberation is progressively displaced, whether through dependence, manipulation, or the sheer volume of synthetic content competing for attention [8], [9]. Generative technologies intensify the risk by producing fabricated material indistinguishable from authentic testimony [10], though the magnitude of this effect is contested: some communication research argues that fears about AI-driven misinformation are overstated, since exposure remains bounded by demand rather than supply [45]. Distributive analyses add that the burdens and benefits of these transformations fall unevenly across populations [11], [46].

4.2 Evaluating Artificial Intelligence through the Five Essential Objectives (10pt)

Each cluster is now assessed against the *ḍarūriyyāt*. The Qur'anic material adduced below is invoked to establish the objective at issue and the purpose it protects, in keeping with the inductive method described above; it is not offered as adjudication of particular cases.

Ḥifẓ al-Dīn (preservation of religion). The Qur'ān forecloses compulsion in matters of faith (Q 2:256), which presupposes that religious assent be formed freely and through understanding rather than engineered by external control. Algorithmic recommendation and generative systems now

function as epistemic structures shaping how religious knowledge is produced, ranked, and consumed, and thereby bear directly on this objective [36]. The risk is twofold: the displacement of qualified scholarly authority by automated outputs presented with unwarranted confidence, and the fragmentation of juristic reasoning into instant answers stripped of context [3], [31]. Empirical evaluation showing that AI systems perform worst precisely on maqāsid-aware reasoning gives the concern concrete force [32]. Ḥifẓ al-dīn therefore requires provenance verification for religious citations, explicit communication of uncertainty, and qualified human oversight of any fatwa-like output [36], [37].

Ḥifẓ al-Nafs (preservation of life). The Qur'anic equation of a single unjust killing with the destruction of humanity entire (Q 5:32), read alongside the declaration of human dignity in Q 17:70, establishes that persons are inviolable and not reducible to instruments. Two implications follow for AI. In safety-critical deployments such as clinical decision support and autonomous systems, error tolerance must be calibrated to the gravity of the interest at stake, and reviews of health-sector adoption consistently identify patient safety and institutional readiness as preconditions rather than afterthoughts [5], [33]. More broadly, the dignity clause resists any governance arrangement in which persons figure merely as data subjects to be optimised, which is the precise register in which the dehumanisation objection is best stated [39].

Ḥifẓ al-'Aql (preservation of intellect). This objective bears most directly on the informational harms of generative AI. The Qur'ān prohibits asserting what one does not know (Q 17:36) and requires verification before acting on reported information (Q 49:6). This epistemic discipline, *tabayyun*, maps closely onto contemporary concerns about synthetic media and model confabulation [8], [10]. Existing maqāsid scholarship already treats ḥifẓ al-'aql as the anchor for resistance to misinformation [23] and, in the religious domain, for the

protection of intellectual security against unverified digital content [36]. The objective thus grounds requirements of provenance labelling, verifiability, and cultivation of public digital literacy, while the contested magnitude of misinformation effects [45] counsels proportionate rather than maximal intervention.

Ḥifẓ al-Nasl (preservation of lineage). Understood expansively as the protection of family integrity and intergenerational continuity, this objective addresses harms that surface in relational rather than individual form. Technology-facilitated abuse within marriage, including digital surveillance, doxing, coercive monitoring, and threats to disseminate intimate material, constitutes non-material harm (*ḍarar ma'nawī*) that Islamic family ethics has begun to conceptualise through maqāsid-based frameworks incorporating *amānah*, *satr*, and the prohibition of *tajassus* [47]. The obligation in Q 66:6 to safeguard oneself and one's household supplies the normative anchor. For governance, the objective grounds age-appropriate design, protection against technology-facilitated coercion, and attention to the effects of algorithmic mediation on the social fabric.

Ḥifẓ al-Māl (preservation of property). The prohibition on consuming others' property wrongfully (Q 2:188) and the requirement that transactions rest on mutual consent (Q 4:29) acquire new application once it is recognised that personal data now functions as *māl* in the operative sense: it is appropriable, tradable, and revenue-generating, and its extraction transfers value from the person described to the party that holds it. Classical jurisprudence attaches protection to whatever possesses recognised value and is customarily exchanged, a description behavioural data satisfies more fully than many tangible goods. Where data is extracted without genuine consent and monetised, the transaction fails the consent condition however elaborate the terms of service; Islamic scholarship on digital ethics has drawn precisely this conclusion regarding consumer data [43],[48]. The objective further

bears on distributive questions, since the gains of automation accrue unevenly while its displacement costs fall on identifiable groups [11], [29]. *Ḥifẓ al-māl* therefore demands not only protection against data exploitation but attention to digital inclusion as a condition of equitable participation [35].

4.3 Constructing the Normative Framework

Synthesising the foregoing yields a framework in which each objective generates governance criteria and each criterion corresponds to an established principle in mainstream AI ethics. Table 1 presents the correspondence. The claim is not that the two vocabularies are equivalent. It is that the *maqāṣid* supply a ranking rationale that principle lists conspicuously lack: where values conflict, priority follows the gradation of interests, with *ḍarūriyyāt* taking precedence over *ḥājiyyāt* and *ḥājiyyāt* over *taḥsīniyyāt*, and with the juristic maxim that averting harm precedes securing benefit (*dar' al-mafāṣid muqaddam 'alā jalb al-maṣāliḥ*) governing cases of genuine indeterminacy.

A worked example clarifies how the ordering operates in practice. Consider a clinical decision-support system whose diagnostic accuracy improves materially when trained on patient records reused without individually renewed consent. Here *ḥifẓ al-nafs*, engaged by accuracy in a life-bearing context, stands in tension with *ḥifẓ al-māl* and the dignity interest in personal

information. Both interests are *ḍarūrī*, so precedence cannot be settled by rank alone, and the analysis proceeds to the maxim that averting harm precedes securing benefit. The outcome is neither unrestricted reuse nor prohibition but a conditional permission: the least intrusive data configuration adequate to the clinical purpose, coupled with de-identification, purpose limitation, and independent review. Where the anticipated benefit is marginal rather than life-bearing, the balance inverts and the configuration is refused. The framework therefore yields graded outcomes rather than binary verdicts, precisely the determination that principle lists leave unspecified [1].

Operationalisation requires three institutional commitments. The first is anticipatory assessment: because *maqāṣid* reasoning is purposive rather than reactive, evaluation must occur at design stage rather than after deployment, a requirement compatible with risk-tiered regulatory models already in force [12], [24]. The second is qualified oversight, whether through Shariah advisory boards for religiously salient applications or human-in-the-loop review for consequential decisions [28], [32]. The third is contestability: those affected by an automated decision must be able to challenge it, which presupposes explanation sufficient for challenge to be meaningful [2], [36]. Together these convert the framework from a taxonomy into an assessment procedure.

Table 1. *Maqāṣid*-based governance criteria and their correspondence with mainstream AI-ethics principles

Objective	Principal AI risk	Governance criterion	Corresponding principle
<i>Ḥifẓ al-Dīn</i>	Displacement of religious authority; unverified religious content	Provenance verification; scholarly oversight of fatwa-like output	Human oversight; accountability
<i>Ḥifẓ al-Nafs</i>	Safety failure; instrumentalisation of persons	Risk calibration to gravity of interest; dignity-preserving design	Safety and robustness; human dignity
<i>Ḥifẓ al-'Aql</i>	Misinformation; confabulation; cognitive dependence	Provenance labelling; verifiability (<i>tabayyun</i>); digital literacy	Transparency and explainability
<i>Ḥifẓ al-Nasl</i>	Technology-facilitated coercion; harm to family integrity	Age-appropriate design; protection from digital surveillance	Harm prevention; protection of minors

Hifẓ al-Māl	Data exploitation; inequitable distribution of gains	Genuine consent; equitable access and digital inclusion	Privacy; fairness and non-discrimination
-------------	--	---	--

Source: Author's Synthesis Based on [1], [2], [12], [15], [23], [36].

4.4 Critical Discussion: Complementarity, Limits, and Objections

The framework proposed here is complementary rather than oppositional. At the level of stated values, the correspondence with mainstream instruments is substantial: privacy, fairness, accountability, transparency, and human oversight all find maqāṣid counterparts, a convergence that comparative studies of Islamic and secular digital ethics have repeatedly documented [41], [43]. Three differences nonetheless matter. First, the maqāṣid supply an ordering principle, whereas principle lists typically leave conflicts unresolved [1]. Second, they extend the moral field to religious and spiritual interests that secular frameworks do not thematise, a difference of scope, not merely of idiom [15], [22]. Third, they locate obligation in trusteeship (amānah) rather than in compliance, which changes what counts as adequate discharge of responsibility.

The framework also speaks to governance already under way in Muslim-majority states. Comparative analysis of national AI strategies across the Gulf Cooperation Council identifies a prevailing posture of soft regulation that articulates ethical principles without binding enforcement, raising concerns about ethics-washing and uneven alignment with international instruments [53]. Studies of Saudi and Emirati frameworks show deliberate efforts to embed Islamic conceptions of privacy and dignity within national AI ethics, indicating that value-based governance is already a live policy project rather than a scholarly abstraction [52]. What such initiatives have generally lacked is an explicit derivational method connecting professed Islamic values to determinate governance requirements, which is the deficit this framework is designed to address. Comparable considerations arise in Islamic finance, where bibliometric mapping documents rapid AI adoption for Shariah-

compliance and risk functions across Indonesia, Malaysia, and the Gulf alongside comparatively little specification of the ethical criteria such systems should satisfy [54]. Realising that specification will require sustained collaboration between jurists and technologists, since neither constituency commands the whole of the competence involved [37].

Three objections deserve acknowledgement. The first is the charge of indeterminacy: critics may object that the five objectives are too abstract to adjudicate concrete design choices. The response is that abstraction is a property of all principle-level frameworks, including secular ones, and that the remedy lies in the assessment procedure specified above rather than in abandoning purposive reasoning [21]. The second is the risk of proof-texting, that is, selecting verses to license conclusions reached on other grounds. This study has sought to mitigate the risk by treating textual material as evidence of purpose within an inductive method rather than as isolated warrant [16], but the danger is real and warrants methodological vigilance in subsequent work. The third concerns pluralism: in religiously diverse polities, a confessional framework cannot serve as the sole basis of public regulation. The appropriate claim is narrower: the framework can guide Muslim developers and institutions, inform Shariah governance bodies, and contribute a distinct voice to global deliberation [15], [29].

Several limitations follow from the study's design. As conceptual research, it produces a framework rather than evidence of its efficacy; empirical validation through case application, expert consultation, or systematic audit of model outputs remains outstanding, and recent scholarship has begun to sketch such an agenda [36]. The analysis also treats the five ḍarūriyyāt as given, whereas contemporary maqāṣid scholarship has proposed additional objectives, the preservation of dignity [34] and of ecological

order [18], whose incorporation would enlarge the framework's reach. Finally, the correspondence with mainstream principles is drawn at the level of stated commitments; whether the two vocabularies would converge in hard cases, where trade-offs bite, is a question this study identifies but does not settle.

5. CONCLUSION

This study set out to determine whether *maqāṣid al-sharīʿah* can function as an operational instrument for evaluating artificial intelligence rather than as a declarative ideal, and the analysis supports an affirmative answer with qualifications. Four ethical problem clusters recur across the recent literature: the erosion of privacy, algorithmic bias, opacity with diffuse accountability, and the attrition of human moral agency. Each engages one or more of the five essential objectives in determinate ways: religious authority and epistemic security under *ḥifẓ al-dīn*, safety and dignity under *ḥifẓ al-nafs*, misinformation and cognitive dependence under *ḥifẓ al-ʿaql*, technology-facilitated harm to family integrity under *ḥifẓ al-nasl*, and data exploitation with distributive inequity under *ḥifẓ al-māl*. Rendering these judgements into an eleven-criterion framework, correlated with the principle vocabulary of mainstream

AI ethics and coupled with an ordering rationale drawn from the gradation of interests, yields a normative structure that complements existing regulatory approaches by supplying what they most conspicuously lack: a principled basis for resolving conflicts among values that all parties already endorse.

Three recommendations follow. For scholars, the priority is empirical: the framework requires validation through case application and systematic auditing of model behaviour in religiously and socially salient domains. Priority candidates include Shariah-screening systems in Islamic financial institutions and digital fatwa platforms, both of which already operate at scale and generate auditable outputs, and the framework would benefit from incorporating recently proposed objectives such as the preservation of dignity and ecological order. For policymakers and Shariah governance bodies, the analysis supports embedding anticipatory *maqāṣid* assessment into existing risk-tiered review, together with qualified oversight and effective contestability for affected persons. For Muslim developers and institutions, the framework offers a design vocabulary rather than a compliance checklist, locating responsibility in trusteeship over the interests the law exists to protect. Its value will ultimately be measured not by its coherence on paper but by whether it changes decisions made before systems are deployed.

REFERENCES

- [1] D. Mbiazi, M. Bhange, M. Babaei, I. Sheth, P. Kenfack, and S. E. Kahou, "Survey on AI Ethics : A Socio-Technical Perspective," 2025, doi: 10.1111/coin.70149.
- [2] M. A. Camilleri, "Artificial intelligence governance : Ethical considerations and implications for social responsibility," no. April 2023, pp. 1–15, 2024, doi: 10.1111/exsy.13406.
- [3] M. Usmonov, "From Human Scholars to AI Fatwas : Media , Ethics , and the Limits of AI in Islamic Religious Communication," vol. 5, no. 1, 2025.
- [4] M. Izzul *et al.*, "Artificial intelligence in islamic finance: an outlook based on maqasid al-shariah," vol. 31, no. 2, 2026.
- [5] S. A. A. Q. Mohammed, Y. M. M. Osman, A. M. Ibrahim, and M. Shaban, "Ethical and regulatory considerations in the use of AI and machine learning in nursing: A systematic review," *Int. Nurs. Rev.*, vol. 72, no. 1, p. e70010, Mar. 2025, doi: <https://doi.org/10.1111/inr.70010>.
- [6] R. B. K., A. M. Luke, V. M. A., K. P. Shetty, and N. A. Valente, "Artificial intelligence-driven dentistry: A systematic review of ethical and legal challenges," *Int. J. Dent.*, vol. 2026, no. 1, 2026, doi: 10.1155/ijod/1870800.
- [7] I. Akin, L. Wilczkowska, M. Akin, and M. Alrawad, "Ethical and data privacy risks in AI-driven accounting: A mixed-methods study," *Hum. Behav. Emerg. Technol.*, vol. 2026, no. 1, 2026, doi:

- 10.1155/hbe2/6048044.
- [8] E. Ferrara, "GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models," *J. Comput. Soc. Sci.*, 2023, [Online]. Available: <https://consensus.app/papers/details/53180ab31a1a5c1aa0490d19dd159b9f/>
 - [9] K. Wach, "The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT," *Entrep. Bus. Econ. Rev.*, 2023, [Online]. Available: <https://consensus.app/papers/details/c8abfcc13baf5aa19fa6357d897ea5ac/>
 - [10] M. Al-kfairy, D. Mustafa, N. Kshetri, M. Insiew, and O. Alfandi, "Ethical challenges and solutions of generative AI: An interdisciplinary perspective," *Informatics*, 2024, [Online]. Available: <https://consensus.app/papers/details/3ada17d220b154c58fa6c446d017b1c9/>
 - [11] V. Capraro, "The impact of generative artificial intelligence on socioeconomic inequalities and policy making," *PNAS Nexus*, 2023, [Online]. Available: <https://consensus.app/papers/details/f5656c44cd1e526e8f513851e457b9e1/>
 - [12] M. Wörsdörfer, "Mitigating the adverse effects of AI with the European Union's Artificial Intelligence Act: Hype or hope?," *Glob. Bus. Organ. Excell.*, vol. 43, no. 3, pp. 106–126, 2023, doi: 10.1002/joe.22238.
 - [13] P. Lorenz, K. Perset, and J. Berryhill, "Initial policy considerations for generative artificial intelligence," *OECD Artif. Intell. Pap.*, 2023, [Online]. Available: <https://consensus.app/papers/details/4e461cae3f3f5c4f894da86412cfedf1/>
 - [14] X. Li, "Public trust in global AI governance across geopolitical rivals," *Glob. Policy*, vol. 17, no. 1, pp. 36–46, 2025, doi: 10.1111/1758-5899.70105.
 - [15] Y. B. Prabowo, "Maqasid Al-Shariah and Artificial Intelligence: Unaddressed Issues in Contemporary AI Ethics Studies," *AL-IMAM J. Islam. Stud. Civiliz. Learn. Soc.*, 2026, [Online]. Available: <https://consensus.app/papers/details/32fcb4a10498572e9c542c5b7694d071/>
 - [16] A. Ikhlās, "The theory of higher objectives and intents of Islamic law (maqasid al-shariah) as an instrument of ijtihad according to Imam al-Shatibi in al-Muwafaqat fi Usul al-Shari'ah," *Media Syari'ah Wahana Kaji. Huk. Islam dan Pranata Sos.*, 2021, [Online]. Available: <https://consensus.app/papers/details/443a8efcfdea5ae3be756fc269ccea4d/>
 - [17] A. A. I. Omar, "The Andalusian legacy in maqasid al-shariah: A comprehensive literature review of classical texts and their modern relevance," *Int. J. Educ. Psychol. Couns.*, 2025, [Online]. Available: <https://consensus.app/papers/details/d8fe195d4a2e576099f9d11c7adf37c9/>
 - [18] S. Al-Tamimi, "Maqasid al-Shari'ah (the higher objectives of Islamic law) as an alternative perspective for contemporary environmental sustainability," *Res. J. Adv. Humanit.*, 2026, [Online]. Available: <https://consensus.app/papers/details/1a5555acd8785ac9a2642b04068410b7/>
 - [19] A. M. Yunus, "The relevance of maqasid al-shariah in addressing humanitarian crises: Insights from Surah al-Fil," *Int. J. Acad. Res. Bus. Soc. Sci.*, 2024, [Online]. Available: <https://consensus.app/papers/details/63a4b37d813e53a1959ed5c3be4a7203/>
 - [20] Arminsyah, "Maqasid al-shariah in contemporary legal systems: An analysis of digital rights and privacy protection," *Al-Qadha J. Huk. Islam dan Perundang-Undangan*, 2025, [Online]. Available: <https://consensus.app/papers/details/2fcf86756d795ab8aedf7e5f4e32cf89/>
 - [21] A. A. M. Hassanein, "The complementarity of ijtihad and the maqasid al-shariah in Islamic law: An analytical study," *Int. J. Acad. Res. Bus. Soc. Sci.*, 2025, [Online]. Available: <https://consensus.app/papers/details/2cf65b1b404355798db9bc12b5207780/>
 - [22] M. Mohadi, "Maqasid Al-Shari'ah and the Ethics of Artificial Intelligence," *J. Contemp. Maqasid Stud.*, 2023, [Online]. Available: <https://consensus.app/papers/details/afd98f2253cf551cb8279c8b5f1f8fee/>
 - [23] Z. Habib, "Ethics of Artificial Intelligence in Maqāṣid Al-Sharī'a's Perspective," *KARSA J. Soc. Islam. Cult.*, 2025, [Online]. Available: <https://consensus.app/papers/details/d7f1e820c02f56f98b1dc4cb41702b9f/>
 - [24] W. Hidayati, "Evaluating the role of Maqasid al-Shari'ah in formulating ethical frameworks for artificial intelligence in Islamic jurisprudence," *Harmon. Philos. Int. J. Islam. Relig. Stud. Sharia*, 2024, [Online]. Available: <https://consensus.app/papers/details/45ff4a2313ac561d99fefaf11649dc39/>
 - [25] M. F. Alamsyah, "Ethics of the Use of Artificial Intelligence (AI) in the Paradigm of Islamic Law," *Solo Int. Collab. Publ. Soc. Sci. Humanit.*, 2025, [Online]. Available: <https://consensus.app/papers/details/67f145e43eb9522cb99374c60cc8d235/>
 - [26] Zulkifli, "Artificial intelligence and Islamic ethical guidelines: A systematic review," 2025. [Online]. Available: <https://consensus.app/papers/details/e74073afd1de55b1a38b3ffab24b2253/>
 - [27] Bahri, "Constructing an artificial intelligence ethics framework based on maqasid syariah for sustainable digital transformation governance," *Khazanah J. Islam. Educ. Sci.*, 2025, [Online]. Available:

- <https://consensus.app/papers/details/be17488bfaaa5e93936a4276840faacb/>
- [28] Z. Baimakhanova, "Integrative Artificial Intelligence Ethics Model Through the Lens of Maqasid Al-Shari'a," *Islam. Zertteuleri (Islamic Stud. Res.*, 2025, [Online]. Available: <https://consensus.app/papers/details/9e1039766d9a5edf84538201ca27e454/>
- [29] N. Samsuddin, "Artificial Intelligence in Business: A Literature Review from maqasid al-Shariah Perspective's," *Contrar. Ser. Law Soc. Justice*, 2026, [Online]. Available: <https://consensus.app/papers/details/ea8568f6aef258fd942d9b842566a90b/>
- [30] S. Malik, "Artificial intelligence and Islamic thought: Two distinctive challenges," *J. Islam. Muslim Stud.*, 2023, [Online]. Available: <https://consensus.app/papers/details/e83f0d44adca5ce3b1195be383a215c9/>
- [31] H. Sya'diyah, "Fiqh algorithm: Finding a common ground between fatwa and artificial intelligence," 2025. [Online]. Available: <https://consensus.app/papers/details/ede51b9f8bd05acdb3d8987feea4332d/>
- [32] S. Isma'il, "The role of artificial intelligence in developing contemporary fatwas and analyzing jurisprudential texts: Empirical evidence and the S.A.F.E.-Fiqh framework," *Glob. J. Artif. Intell. Technol. Dev.*, 2026, [Online]. Available: <https://consensus.app/papers/details/4acf3a7d864753968feaeef7a7be2d6c/>
- [33] W. T. Almagharbeh, M. Alharrasi, M. K. Rony, S. Kabir, S. K. Ahmed, and D. M. Alrazeeni, "Ethical and institutional readiness for artificial intelligence in nursing: An umbrella review," *Int. Nurs. Rev.*, vol. 72, no. 4, 2025, doi: 10.1111/inr.70111.
- [34] I. O. Assayouti, "Digital surveillance and human dignity (Hifz al-'Ird): An Islamic ethical appraisal with reflections from Nigeria," *Int. J. Res. Innov. Soc. Sci.*, 2026, [Online]. Available: <https://consensus.app/papers/details/0f9ac664ba815939b7aa1c11f01b6802/>
- [35] F. N. H. Afif, "Inclusive digital technology and Islamic education: Promoting social equity for persons with disabilities in the digital age," *Multicult. Islam. Educ. Rev.*, 2025, [Online]. Available: <https://consensus.app/papers/details/f309be58e7e45528b515c5288ee3ca2f/>
- [36] A. Abdullah, "Artificial intelligence, religious authority, and digital governance: A maqasid framework for governing religious knowledge and protecting intellectual security," *Int. J. Spec. Islam. Stud.*, 2026, [Online]. Available: <https://consensus.app/papers/details/0dc1df686b275e4d9d3ea38c9aad262/>
- [37] S. F. A. Rahim, "Artificial intelligence for fatwa issuance: Guidelines and ethical considerations," *J. Fatwa Manag. Res.*, 2025, [Online]. Available: <https://consensus.app/papers/details/4ddba4d09efe55d6937637694f5d0e68/>
- [38] A. Sati, "The digital transformation of tafsir and its implications for Islamic legal derivation in the contemporary era," *MILRev Metro Islam. Law Rev.*, 2025, [Online]. Available: <https://consensus.app/papers/details/6298f0cd41f95a6898744d66726b66fb/>
- [39] M. Nidhom, "Does Islam protect digital privacy? An Islamic studies approach to personal data protection," *Al-Afham J. Islam. Stud.*, 2025, [Online]. Available: <https://consensus.app/papers/details/4c614dba06ba5aa48b3ec8a7120a5a1d/>
- [40] Nurhikma, "Batasan privasi dalam hukum Islam: Analisis fenomena oversharing di media sosial," *J. Ilmu Huk.*, 2025, [Online]. Available: <https://consensus.app/papers/details/c318803acc2b5793906a5a89126a73dd/>
- [41] R. Al Qaruty, "Islamic ethical frameworks for privacy, security, and trust in digital social networks: A jurisprudential perspective," 2025. [Online]. Available: <https://consensus.app/papers/details/5d5cbcd7ad15865b16907d702676fc8/>
- [42] T. Putra, "Integrating Islamic laws into Indonesian data protection laws: An analysis of regulatory landscape and ethical considerations," *Contemp. Issues Interfaith Law Soc.*, 2024, [Online]. Available: <https://consensus.app/papers/details/dd3a51274ff05049a0ad66f30e648af5/>
- [43] Bakhrudin, "Islamic perspectives on cybersecurity and data privacy: Legal and ethical implications," *West Sci. Law Hum. Rights*, 2023, [Online]. Available: <https://consensus.app/papers/details/aafeb2907efc5a618daf91a04b317c9b/>
- [44] M. A. S. A. Haleem and Trans., *The Qur'an: A New Translation*. Oxford, U.K.: Oxford Univ. Press, 2016. [Online]. Available: <https://global.oup.com/academic/product/the-quran-9780199535958>
- [45] F. M. Simon, S. Altay, and H. Mercier, "Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown," *Harvard Kennedy Sch. Misinformation Rev.*, 2023, [Online]. Available: <https://consensus.app/papers/details/f8b6b7129a3851d5b81b92b24fe938dc/>
- [46] I. Solaiman, "Evaluating the social impact of generative AI systems in systems and society," *arXiv Prepr.*, 2023, [Online]. Available: <https://consensus.app/papers/details/73e8a2dce0935a2dafa57c8a41f1a15d/>
- [47] M. S. Fajar, "Toxic online relationships in Muslim marriage: Reconstructing prophetic digital ethics in

- Indonesian Islamic family law," *Sakina J. Fam. Stud.*, 2026, [Online]. Available: <https://consensus.app/papers/details/c9bf12f182225ffeb81642d801eb3403/>
- [48] A. A. Saputra, "Islamic-based digital ethics: The phenomenon of online consumer data security," *Share J. Ekon. dan Keuang. Islam*, 2022, [Online]. Available: <https://consensus.app/papers/details/41c321e15245598392233cf811772659/>
- [49] A. Raquib, B. Channa, T. Zubair, and J. Qadir, "Islamic virtue-based ethics for artificial intelligence," *Discover Artificial Intelligence*, vol. 2, no. 1, 2022.
- [50] E. Elmahjub, "Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI," *Philosophy & Technology*, vol. 36, no. 4, 2023.
- [51] M. Ghaly, "What makes work 'good' in the age of artificial intelligence (AI)? Islamic perspectives on AI-mediated work ethics," *The Journal of Ethics*, 2023.
- [52] E. Gorian and D. Gribov, "Digital ethics of artificial intelligence (AI) in Saudi Arabia and United Arab Emirates," *Malaysian Journal of Syariah and Law*, 2024.
- [53] M. Albous, et al., "AI governance in the GCC states: A comparative analysis of national AI strategies," *arXiv preprint*, 2025.
- [54] M. Iqbal, et al., "AI in Islamic finance: Global trends, ethical implications, and bibliometric insights," *Review of Islamic Social Finance and Entrepreneurship*, 2025.

BIOGRAPHIES OF AUTHORS

	Maman Supardi, S.Ag  is a graduate student in the Department of Qur'anic Studies and Exegesis (Ilmu al-Qur'an dan Tafsir) at the Postgraduate Program, Universitas Islam Negeri (UIN) Alauddin Makassar, Indonesia. He completed his undergraduate degree in the same field at UIN Alauddin Makassar and is currently pursuing his master's degree there. His research interests include Qur'anic exegesis (tafsir), thematic interpretation of the Qur'an (tafsir mawḍūʿī), maqāṣid al-sharīʿah, and the intersection of Islamic thought with contemporary issues such as technology and ethics. bangsupardi186@gmail.com
	Hilmiy Hanif, S.Ag. is a graduate with a bachelor's degree in Quranic Studies and Exegesis from Universitas Islam Alauddin Makassar. He is currently pursuing a master's degree in Quranic Studies and Exegesis at the Postgraduate School of Universitas Islam Alauddin Makassar. His research interests include contemporary Quranic exegesis, thematic interpretation of the Qur'an (Tafsir Maudhui), contemporary issues and technology, and Islamic thought related to ethics. Email: hilmiyhanif35@gmail.com
	Mursyid Rahman, S.Ag. has received his bachelor's degree in Qur'anic Studies and Exegesis (Ilmu al-Qur'an dan Tafsir) from Universitas Islam Negeri Alauddin Makassar, Indonesia. He has applied for the Master's Program in the same discipline at the Postgraduate Program of UIN Alauddin Makassar. His research interests include Classical Qur'anic Interpretation, Nusantara Tafsir, Linguistic Tafsir (or lughawi Tafsir), and the thought of prominent Qur'anic exegetes (mufasssirrūn). mursyidr2020@gmail.com

	Ahmad Fadhil Arham, S.Ag. received his bachelor's degree in Qur'anic Studies and Exegesis (<i>Ilmu al-Qur'an dan Tafsir</i>) from Universitas Islam Negeri (UIN) Alauddin Makassar, Indonesia. He has applied for the Master's Program in the same discipline at the Postgraduate Program of UIN Alauddin Makassar. His research interests include local Qur'anic exegesis, thematic Qur'anic interpretation (<i>tafsir mawḍū'ī</i>), and contemporary Qur'anic exegesis. ahmadfadhil.ic@gmail.com
	Andi Ali Amiruddin, M.A. is the Head of the Department of Qur'anic Studies and Exegesis (<i>Ilmu al-Qur'an dan Tafsir</i>), Faculty of Ushuluddin and Philosophy, Universitas Islam Negeri (UIN) Alauddin Makassar, Indonesia. His areas of expertise include Qur'anic exegesis (<i>tafsir</i>), thematic interpretation of the Qur'an (<i>tafsir mawḍū'ī</i>), and Qur'anic ecotheology. He is actively engaged in teaching, research, and academic supervision in the field of Qur'anic studies. andiamiruddin@uin-alauddin.ac.id