

Research Article

Ethical Pluralism in AI Governance: A Comparative Analysis of Maqasid al-Shariah and International Frameworks in Islamic Finance

Fellah Sofiane^{1*} and Bedarnia Mebarka²

¹University of Biskra, Algeria

²University of Laghouat, Algeria

*Corresponding Author

Fellah Sofiane
(fellahsofiane09@gmail.com)

Article History

Received: 24.07.2026

Accepted: 21.08.2026

Published: 26.09.2026

DOI: 10.61336/cjmr.1603.104

Abstract: Artificial Intelligence (AI) is being adopted at speed in Muslim-Majority Societies (MMS), above all in digital and Islamic financial services. The international instruments that now shape AI governance (the UNESCO Recommendation on the Ethics of AI, the NIST AI Risk Management Framework and the OECD AI Principles) have, however, very little to say about Islamic ethical and legal norms. In this paper we ask whether and on what conditions, selected objectives of Maqasid al-Shariah can be combined with these frameworks to govern AI in digital entrepreneurship and Islamic finance. The design is qualitative and comparative. We coded a purposive corpus (classical and contemporary Islamic legal sources, international AI governance instruments, Shari'ah and regulatory standards issued by AAOIFI and the IFSB and peer-reviewed studies) against a predefined grid of six ethical themes and five dimensions of comparison and then set the resulting propositions against an illustrative regulatory analysis of Malaysia and Algeria. In practice the two normative sources agree on harm prevention, accountability, transparency and dignity. Where they part ways is on the origin of ethical limits and on whether those limits may be balanced against expected benefits. Building on this, we propose the Maqasid AI Filter Framework (MAFF), a Maqasid-AI Compliance Index (MACI) with certification tiers and a three-lines institutional model for Shari'ah AI auditing and we discuss enforcement and the stakeholders involved. The contribution is conceptual; it has yet to be tested empirically.

Keywords: Artificial Intelligence Governance, Maqasid Al-Shariah, Islamic Finance, Explainable AI (XAI), Algorithmic Fairness, Shari'ah Governance, Ethical Pluralism, Digital Entrepreneurship

Copyright @ 2026: This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

INTRODUCTION

AI is entering the global economy quickly. Machine learning, until recently a laboratory matter, now runs a large share of the world's digital infrastructure [1,2]. In high- and low-income countries alike, people reach banks, employers, health services and public offices through digital platforms [3]. As AI takes over forecasting, credit and business scoring and customer management, decisions that used to belong to managers and loan officers are passing, little by little, to software [4].

The consequences for people are real. Much of the harm can be traced to developers and designers who never built proper checks into their systems [5,6]. The best documented examples are well known: hiring software whose logic nobody can see [7], recidivism scores that turned out to be biased [8] and credit-scoring systems that treat groups differently [9]. Tools of this kind tend to entrench the social and economic inequalities that already exist. In Islamic finance the risk takes a particular form. An algorithm that no one can explain may conceal pricing rules or contract terms that amount to *riba* or to excessive *gharar* and lend them a legitimacy they have not earned [10,11].

In response, a number of soft-law and hard-law instruments have been adopted. The UNESCO Recommendation on the Ethics of Artificial Intelligence [4], adopted by 193 Member States, rests on four core values and ten principles, human dignity, transparency and explainability and responsibility and accountability among them. The NIST AI Risk Management Framework [12] proposes an iterative process organised around four functions: Govern, Map, Measure and Manage.

The OECD AI Principles [13] ask for human-centred values and trustworthy AI, while the European Union's AI Act [14] makes several of these commitments legally binding; creditworthiness assessment, for instance, is listed there as a high-risk use. These are serious, carefully built instruments. Their vocabulary, though, is largely secular and liberal and they do not engage directly with other ethical traditions, the Islamic tradition included [15,16].

In Islamic thought, technology is not morally neutral: The way it is used must agree with the principles of the Shariah. The Maqasid al-Shariah, the higher objectives of Islamic law, provide a standard for judging a technology by its effects on the preservation of religion, life, intellect, progeny and wealth. Later scholars widened the list with further values and brought it to bear on governance and on financial systems. Islamic banking and finance is probably where this is most visible, since there the Maqasid serve to interpret regulatory, ethical and legal limits and requirements [17].

Research on AI and Islam has been growing [18-20,21,22,23,24,25]. Two gaps nonetheless remain. This literature rarely touches the legal and technical design of AI systems. And little has been written on using Maqasid al-Shariah to govern AI in Islamic finance, which is surprising given that the Islamic financial services industry held some USD 3.88 trillion in total assets in 2024 and that Islamic banks are already using AI for identity verification, chatbots and the analysis of digital footprints [26].

Our response to these gaps has four parts. First, we build a comparative framework, grounded in an explicit coding grid, that places the five Maqasid objectives next to the principles of international AI governance. Second, we link specific technical tools (explainable AI, fairness metrics, differential privacy and model documentation) to Maqasid-based obligations in Islamic finance, with reference to AAOIFI and IFSB standards. Third, we try out the resulting propositions on two quite different jurisdictions, Malaysia and Algeria and on a stylised Islamic FinTech vignette. Last, we set out a practical governance package made up of the Maqasid AI Filter Framework (MAFF), a Maqasid-AI Compliance Index (MACI) with certification tiers, an institutional model for Shari'ah AI auditing and a stakeholder-based analysis of enforcement.

The research question is this: to what extent can selected aspects of Maqasid al-Shariah be combined with international approaches to AI governance so as to regulate AI in the digital economies of Muslim-majority countries and in Islamic finance in particular? We look at whether a selective and contextual integration of the two normative sources (the term is defined operationally in Section 1.5) can underpin a responsible AI governance framework able to handle the technical, ethical and cultural problems these countries face. Our working proposition is modest. International AI ethics and Maqasid al-Shariah stand on different normative ground and cannot be brought into full alignment; even so, both can inform a single, coherent framework for responsible AI governance (Table 1).

Theoretical Background

Maqasid al-Shariah: Classical Foundations and Contemporary Extensions

Few doctrines in Islamic jurisprudence are as developed as Maqasid al-Shariah. Al-Ghazali [27] identified five essential objectives of the law, namely the preservation of religion (din), life (nafs), intellect ('aql), progeny (nasl) and wealth (mal). Maslahah, in his account, meant protecting these objectives rather than simply pursuing whatever seems useful and he accepted recourse to unrestricted public interest (maslahah mursalah) only when that interest was necessary (daruriyyah), certain (qat'iyyah) and universal (kulliyah). The restriction deserves attention today, as societies consider leaving serious decisions to algorithms. It was Al-Shatibi [28] who gave the theory its systematic shape. He arranged the objectives into daruriyyat (necessities), hajiyyat (needs) and tahsiniyyat (embellishments), attached complementary rules (mukammilat) to each level and asked jurists to weigh the consequences of acts (i'tibar al-ma'alat). That concern with consequences offers a legal and normative yardstick for technologies whose effects only become visible over time.

Ibn Ashur [29] revived the tradition in the twentieth century. He drew a line between the general objectives of the law (maqasid 'ammah) and the objectives proper to particular fields and for financial transactions he insisted on the circulation (rawaj), clarity (wuduh), preservation (hifz), stability (thabat) and justice ('adl) of wealth. He also laid down criteria for recognising a valid objective, which must be established (thubut), evident (zuhur), well defined (indibat) and consistent (ittirad) and he gave real weight to values such as freedom (hurriyyah) and equality (musawat). His criteria are what we use to operationalise the Maqasid side of the comparison (Section 2.3). It is worth noting that Al-Shatibi and Ibn Ashur belonged to the Maliki school, which prevails in the Maghreb, whereas Al-Ghazali was Shafi'i; this connects the framework with South-East Asian jurisdictions such as Malaysia.

Table 1: Comparative Overview: Maqasid al-Shariah Objectives and International AI Governance Principles

Maqasid Objective	AI Governance Principle	Convergence	Key Divergence
Hifz al-Din	Non-discrimination; Cultural rights	Protection of belief and identity	Divine Vs secular authority on moral limits
Hifz al-Nafs	Safety; Risk mitigation; Human oversight	Harm prevention; Human dignity	Absolute prohibition Vs contextual thresholds
Hifz al-'Aql	Transparency; Explainability (XAI)	Cognitive autonomy; Anti-manipulation	Moral duty Vs procedural obligation
Hifz al-Nasl	Fairness; Non-discrimination in hiring	Social equity; Protection of future generations	Intergenerational ethics Vs current-risk framing
Hifz al-Mal	Accountability; Data protection; Fairness	Distributive justice; Prohibition of riba/gharar	Absolute prohibition Vs risk-based thresholds

Source: Authors' own elaboration based on Al-Ghazali [27], Al-Shatibi [28], Auda [30], Kamali [32], UNESCO [4], NIST [12], OECD [13]

Contemporary authors have carried this legacy further. Auda [30] reads the Maqasid through a systems approach centred on human development, openness and purpose and this gives a basis for judging new technologies by the extent to which they serve the objectives of the law. Kamali [31,32] likewise applies the Maqasid to the assessment of law and governance. Both accept that AI and similar technologies may help to realise the objectives of the law; both also hold that moral responsibility for what these technologies do lies with the people who design, deploy and use them [30,32].

Maqasid Al-Shariah in Islamic Financial Governance

The connection between Maqasid al-Shariah and Islamic finance is arguably where current work in Islamic finance and law is most innovative. El-Gamal [10] maintains that restrictions such as those on *riba* and *gharar* are best read as limits on conduct that harms society and therefore ought to be assessed in terms of *Hifz al-Mal*, '*adl* and *dar*' *al-mafسادah*.

Iqbal and Mirakhor [17] describe the welfare and redistributive function of the Islamic financial system, performed through contracts such as *mudarabah* and *musharakah* and through *Zakat* and *Waqf* and they underline its concern with social justice. Governing such a system is becoming harder. In an economy reshaped by AI and financial technology (FinTech), the Shariah supervisory framework may run into side effects nobody intended: interest-based risk pricing hidden in credit scoring, say, or *gharar* in automated financial advice.

Before any algorithm can be examined, two technical distinctions have to be made. Classical jurisprudence separates *riba al-nasi'ah*, an increase charged in exchange for deferring a debt, from *riba al-fadl*, an unequal exchange of *ribawi* items of the same kind, such as gold, silver or currencies [10,32]. The former is relevant to AI-driven credit pricing, rescheduling and late-payment tools that might add a time-based increase to an existing debt. The latter matters for algorithmic trading in currencies, gold and digital assets, where delayed or netted settlement may violate the requirement of hand-to-hand settlement (*yadan bi-yad*). *Gharar* calls for similar care. AAOIFI Shari'ah Standard No. 31 holds that *gharar* invalidates a contract only when it occurs in a commutative contract, is excessive, affects the main subject matter and is not justified by a recognised need; minor *gharar* is tolerated [33]. Such a graded approach gives a more precise benchmark for algorithmic opacity than an outright ban would.

Institutionally, Shari'ah governance in Islamic finance is structured by standard-setting bodies. The AAOIFI governance standards regulate how Shari'ah supervisory boards are appointed and composed and what they have to report. The IFSB guiding principles on Shari'ah governance systems [34], on corporate governance [35] and on the supervision of Shari'ah governance [36] set expectations for competence, independence, internal Shari'ah review and audit. None of these texts was drafted with algorithmic decision-making in view and none, so far, explains how AI models should be reviewed for Shari'ah compliance. That institutional gap is one of the things this paper examines.

International AI Governance Frameworks: Principles and Technical Dimensions

Attempts to regulate AI have multiplied and are dispersed across many institutions. The UNESCO Recommendation [4] is built on four values (human rights and human dignity; living in peaceful, just and interconnected societies; diversity and inclusiveness; environmental flourishing) and on ten principles that include proportionality and do no harm, privacy, transparency and explainability, human oversight and fairness and non-discrimination. The NIST AI RMF [12] splits risk management into four functions (Govern, Map, Measure and Manage) and names seven characteristics of trustworthy AI, among which are validity and reliability, explainability and interpretability, privacy enhancement and fairness with harmful bias managed; its Generative AI Profile extends the same logic to large language models [37]. As for the OECD AI Principles [13], they comprise five value-based principles, including human-centred values and fairness, transparency and explainability and accountability.

Principles, however, achieve little until they are connected to the technical features of AI systems and four technical dimensions are relevant here. The first concerns the kind of model. Interpretable 'glass-box' models (scorecards, logistic regression, shallow decision trees) can be inspected directly. Ensemble models such as random forests and gradient-boosted trees, now dominant in credit scoring, are more accurate but harder to see into and deep neural networks are for the most part opaque. Generative foundation models, Large Language Models (LLMs) among them, add risks of their own, namely invented outputs ('hallucinations') and unpredictable behaviour [37-39].

Explainability is the second dimension. XAI methods may be intrinsic, when the model is interpretable by design, or post hoc; global, when they account for the model as a whole, or local, when they account for a single decision; and model-specific or model-agnostic [38]. LIME approximates a model around one case by means of a simple surrogate [6]. SHAP attributes a prediction to its input features with Shapley values [40]. Counterfactual explanations tell an applicant what would have to change for the outcome to be different [41]. Rudin [39] contends that, where the stakes are high, interpretable models are preferable to post hoc explanations of black boxes and her position sits comfortably with the demands of Shari'ah review.

The third dimension has to do with measuring bias. Among the usual metrics are demographic parity and the disparate impact ratio (frequently judged against the 'four-fifths' threshold), equal opportunity and equalized odds, which compare error rates across groups [42], counterfactual fairness [5] and calibration. If base rates differ from one group to another, these criteria generally cannot be met simultaneously [43,44]; picking a metric is thus already a normative choice [45,46]. The fourth dimension, data and model governance, takes in data lineage and provenance, documentation by means of datasheets [47] and model cards [48], privacy-preserving methods such as differential privacy [49] and, for LLMs, retrieval-augmented generation drawing on verified sources [50] together with red-teaming and output monitoring [37]. Table 2 relates these dimensions to Maqasid-related risks in Islamic finance.

Prior Literature and Research Gaps

The literature on AI and Islamic ethics runs along three main lines. Doctrinal studies usually conclude that AI is acceptable as long as it serves a permissible end [22,25]. Governance-oriented work includes Kannike and Fahm [18], who approach AI governance from an Islamic ethical standpoint and Muchtasor [19], who suggests a way of integrating Islamic ethics into global AI governance. Finally, Masyhadi and Thaib [24] represent the legal-comparative line, setting Maqasid-based and conventional approaches to AI regulation side by side.

For the most part, this work has developed at a distance from the core AI ethics literature. Floridi *et al.* [51] and Floridi and Cowls [52] gather the principles of beneficence, non-maleficence, autonomy, justice and explicability into one framework. Mittelstadt *et al.* [53] chart the epistemic and normative problems that algorithms raise and Mittelstadt [16] cautions that principles cannot make AI ethical on their own when no enforcement mechanisms or professional duties stand behind them. Bostrom and Yudkowsky [54] single out transparency, predictability and the moral status of AI systems as central ethical questions. All of this bears on our study. Mittelstadt's objection to principles without enforcement applies with equal force to Maqasid-based proposals, which is precisely why we move from principles to measurable indicators and institutional responsibilities (Section 5).

Reviewing the literature (Section 2.2) brought three gaps to light. In our corpus, no study maps the five Maqasid objectives onto the values and principles of the main international AI governance instruments by means of an explicit analytical grid. Nor did we find a Maqasid-based governance framework for AI in Islamic finance that explains how algorithms might carry *riba* or *gharar* and how such risks might be detected technically. And no study proposes measurable indicators, or an institutional audit model, that would allow Maqasid-based requirements to operate alongside international frameworks. The present paper takes up all three (Table 3).

Table 2: AI Model Classes, Explainability and Maqasid-Relevant Risks in Islamic Finance

Model class	Typical use in Islamic finance	Explainability	Maqasid-relevant risk	Governance instrument
Scorecards; logistic regression; shallow trees	Financing eligibility; murabaha pricing	Intrinsic, global	Hidden <i>riba</i> -proxy variables (Hifz al-Mal)	Variable review by Shari'ah board; model card
Tree ensembles (random forest, gradient boosting)	Credit scoring; fraud detection	Post hoc (SHAP, LIME)	Opaque exclusion of groups (Hifz al-Nasl); <i>gharar</i> in the basis of decisions	SHAP feature audit; fairness metrics; counterfactual explanations
Deep neural networks	Identity and document verification; transaction monitoring; algorithmic trading	Low; post hoc only	Biometric misidentification (Hifz al-Nafs); <i>riba al-fadl</i> through delayed settlement	Human-in-the-loop; accuracy by group; settlement controls
Generative AI/LLMs	Chatbots; robo-advice; Shari'ah Q&A	Very low; output-level only	Fabricated rulings (Hifz al-Din); manipulation (Hifz al-'Aql)	Retrieval from approved rulings (RAG); red-teaming; ban on autonomous fatwa

Source: Authors' own elaboration based on Arrieta *et al.* [38], Rudin [39], NIST [37] and IFSB [26]

Table 3: Summary of Selected Prior Studies on AI and Islamic Ethics

Author(s)	Focus	Contribution	Limitation Addressed by This Study
Kannike and Fahm [18]	AI governance from Islamic ethics perspective	Ethical principles mapping	No technical AI depth; no governance model
Muchtasor [19]	Islamic integration into global AI governance	Ethical pluralism framework	Limited policy feasibility; no Islamic finance focus
Nikmah and Nasikhin [20]	God-conscious AI; Maqasid in algorithmic design	Normative orientation	No operational model; no XAI/fairness engagement
Najib <i>et al.</i> [11]	AI in Islamic finance; Maqasid perspective	Domain-specific analysis	No structured comparative framework; no technical AI depth
This study	Systematic Maqasid-AI mapping; Islamic finance; governance model	Structured thematic mapping + MAFF, MACI and audit model	Integrates XAI, fairness, differential privacy + Islamic finance governance

Source: Authors' own elaboration

Table 4: Operational Definitions of Key Constructs

Construct	Operational definition in this study	How it is observed or applied
Selective and contextual integration	Adoption of an international AI governance requirement into a Maqasid-based framework only where (a) it serves at least one of the five objectives, (b) it does not conflict with a definitive (qat'i) Shari'ah rule and (c) it is adjusted to the jurisdiction's legal and institutional context	Each requirement is coded 'integrable', 'integrable with adaptation' or 'non-integrable' in the analytical grid (Table 5)
Functional convergence	Two frameworks prescribe the same or an equivalent practical safeguard (e.g., human oversight) for the same risk, irrespective of their justification	The same safeguard is coded under the same theme in both types of source
Normative divergence	The frameworks differ in the source of the norm (revelation vs. human consensus), its defeasibility (absolute vs. risk-based) or the locus of accountability	Difference recorded on dimensions D1-D3 of the grid
Ethical filter	A design-stage checkpoint derived from one Maqasid objective, consisting of verifiable requirements, technical tests and a red line whose breach prevents deployment	Filters F1-F5 of MAFF (Table 9), scored through MACI (Table 10)
Red line (hard limit)	A use of AI treated as impermissible regardless of expected benefits because it violates a necessity (daruri) objective or a definitive prohibition	Binary gate: any breach sets the MACI score to zero
Pluralistic governance	An arrangement in which a common international baseline coexists with context-specific normative layers (here, Maqasid-based) that are publicly justified and institutionally enforceable	Presence of both layers and of an identifiable enforcing body (Section 4.6)

Source: Authors' own elaboration

Key Constructs and Operational Definitions

A handful of constructs recur throughout the paper. So that they are not misread, Table 4 gives an operational definition of each, stated in terms of what is observed or checked when the construct is applied.

MATERIALS AND METHODS

Research Design and Epistemological Positioning

We adopt a qualitative, comparative-normative design, pairing doctrinal legal analysis, which interprets authoritative texts, with comparative governance analysis, which examines how different normative orders handle the same risks. The three thematic pillars (AI ethics, Maqasid al-Shariah and Islamic finance within digital entrepreneurship) were chosen for three reasons. AI is changing the way people participate in the economies of Muslim-majority countries and finance is where adoption is most advanced and where its effects on individuals are most immediate [26]. Islamic finance is, moreover, the only field in which Shari'ah norms are already embedded in institutions (regulators, Shari'ah boards and standard-setting bodies), which makes it the most realistic place to begin with Maqasid-based AI governance. And the Maqasid are the most elaborate teleological framework that Islamic jurisprudence offers for assessing new situations the texts do not address directly.

On the international side, the comparison is restricted to the UNESCO Recommendation [4], the NIST AI RMF [12] and the OECD AI Principles [13]. The UNESCO text is the only global normative instrument on AI ethics adopted by every UNESCO Member State; the NIST framework is the operational risk-management framework most widely used; and the OECD principles are intergovernmental, endorsed beyond the OECD membership and taken up by the G20. Between them they represent three kinds of instrument: ethical, operational-technical and policy-oriented. The EU AI Act [14] is used only as a secondary reference, since it is binding regional legislation rather than a global framework. Corporate ethics guidelines and sector-specific technical standards were excluded, because neither enjoys comparable public legitimacy or scope.

Source Selection Strategy and Corpus

Three kinds of source make up the corpus. The first is classical and contemporary Islamic legal literature: Al-Ghazali's Al-Mustasfa, Al-Shatibi's Al-Muwafaqat and Ibn Ashur's Maqasid al-Shari'ah al-Islamiyyah, together with Auda [30], Kamali [31,32], El-Gamal [10] and Iqbal and Mirakhor [17]. The second consists of international AI governance instruments and Islamic finance standards and regulations, that is, the UNESCO, NIST and OECD documents, the EU AI Act, the AAQIFI Shari'ah and governance standards, the IFSB guiding principles and the relevant texts of Bank Negara Malaysia and the Bank of Algeria. The third is peer-reviewed research on AI ethics, XAI and algorithmic fairness and on AI and Islamic ethics.

Peer-reviewed literature was searched in Scopus, the Web of Science Core Collection and Google Scholar and the websites of the standard-setting bodies and regulators named above were also consulted. Three groups of terms were combined: ('artificial intelligence' OR 'machine learning' OR algorithm*) AND (maqasid OR shari'ah OR 'Islamic ethics' OR 'Islamic finance') AND (governance OR ethics OR regulation). Further searches dealt with XAI, fairness metrics and AI in credit scoring.

Table 5: Analytical Grid (Coding Matrix) Used for the Comparison

Dimension	Question asked of each source	Coding categories	Example: Hifz al-Mal vs. NIST AI RMF
D1 Source of the norm	Where does normative authority come from?	Revelation and ijtihad; intergovernmental consensus; technical standardization	Prohibition of riba (revelation) Vs fairness and accountability (consensus)
D2 Defeasibility	Can the norm be traded off against expected benefits?	Absolute (red line); presumptive; risk-based	Absolute prohibition Vs organisational risk tolerance
D3 Locus of accountability	Who is answerable for outcomes?	Individual steward; institution; distributed along the AI value chain	Shari’ah board and management Vs AI actors across the lifecycle
D4 Enforcement mechanism	How is compliance secured?	Legal sanction; certification or audit; voluntary adoption	Shari’ah conformity certification Vs voluntary framework
D5 Technical instrument	Which technical control can verify compliance?	XAI; fairness metrics; privacy techniques; documentation; human oversight	SHAP feature audit for riba-proxies Vs Measure function
Outcome codes		Convergent/partially convergent/divergent; integrable/integrable with adaptation/non-integrable	

Source: Authors' own elaboration

For AI-related work the window was 2015--2025, the years in which machine learning spread through financial services; classical and foundational works were admitted with no date limit. To be included, a source had to be (a) a peer-reviewed article, a scholarly book or an official instrument, (b) written in English or Arabic and (c) explicitly concerned with AI or algorithmic decision-making in relation to ethics, law or finance. Opinion pieces and non-scholarly web content were excluded, as were studies of Islam and technology that do not mention AI and duplicate or superseded versions of instruments. Titles and abstracts were screened first and the remaining texts were then read in full. The search returned [N1] records, of which [N2] survived screening and [N3] the full-text reading. Since many studies on AI and Islamic ethics were published in 2024--2025 in recently launched journals, we relied on them chiefly to describe the current debate and grounded our theoretical claims in the classical texts and in established, widely cited AI ethics research.

Analytical Procedure and Coding Grid

The analysis went through four stages. We began by describing each source and extracting its main normative commitments. The material was then coded with a predefined codebook of six ethical themes (justice, harm prevention, accountability, dignity, transparency and privacy), derived from the overlap between the Maqasid literature and the principles of the three international instruments; codes that surfaced along the way, such as intergenerational impact and epistemic authority, were added as coding proceeded. In the third stage, for each Maqasid objective, the coded material was compared on five dimensions (Table 5): source of the norm, whether it can be traded off, locus of accountability, mode of enforcement and available technical tool. Every pairing was classed as convergent, partially convergent or divergent and every international requirement as integrable, integrable with adaptation or non-integrable (Table 4). Finally, the results were brought together in the governance model presented in Section 5 and tried out on the illustrative cases described in Section 2.4.

Several steps were taken to make the analysis trustworthy. We triangulated across the three kinds of source; we applied Ibn Ashur's criteria (thubut, zuhur, indibat and ittirad) to make sure that each Maqasid-based requirement rests on an established objective; and we kept an audit trail tying every cell of Tables 6 and 9 to its sources.

Illustrative Case Analysis

A purely conceptual comparison lacks empirical grounding and for this reason the propositions derived from the grid were tested on two contrasting jurisdictions. Malaysia was selected as a mature and systemically important Islamic finance market with a dual banking system, a statutory Shari’ah governance regime and a public central-bank consultation on AI in the financial sector. Algeria, the jurisdiction closest to the authors' own context, was selected as an emerging case: Islamic banking has been regulated there since 2020 and Shari’ah conformity is certified by a national authority. Only official documents were used. To show how MAFF would operate inside a single firm, we also added a stylised vignette of an Islamic FinTech platform offering credit scoring and robo-advice, built from typical use cases reported by the IFSB [26]. The vignette is purely illustrative and does not describe any existing company.

Scope and Limitations

The scope of the study is limited in four respects. (1) Jurisdiction. The study concerns Muslim-majority countries with an established Islamic finance sector and the illustrative analysis is confined to Malaysia and Algeria; its findings do not automatically apply to countries where Islamic finance is marginal or where Muslims are a minority. (2) Legal schools.

We rely on Sunni jurisprudence and take a comparative, multi-madhab approach that draws mainly on the Shafi'i (Al-Ghazali) and Maliki (Al-Shatibi, Ibn Ashur) traditions. Hanafi and Hanbali positions are considered only where they diverge on a point relevant to AI and Shi'i approaches to the Maqasid are left aside. Where the schools disagree (as on the scope of *riba al-fadl*), AAOIFI standards serve as the practical reference. (3) Feasibility. The governance model is a proposal whose institutional costs, legal compatibility and market acceptance remain untested. (4) Time. AI technologies and regulations evolve quickly and our analysis reflects the instruments available up to 2025.

RESULTS

Conceptual Findings: Thematic Mapping of Maqasid al-Shariah and International AI Governance

Preservation of Religion (Hifz al-Din): AI, Religious Identity and Epistemic Boundaries: Hifz al-Din obliges states to ensure that their governance systems protect people's religious identity, beliefs and practices. For AI, this translates into legal and policy frameworks that guard against the risks AI may present to them and two risks deserve particular mention. One is the use of AI to give religious advice or to generate fatwas, Qur'anic commentary and other religious texts. AI is not capable of properly evaluating religious legal norms [30,32], so allowing it to issue fatwas would be unethical on the part of the state and legal and policy barriers are required. The other is AI that puts people at a disadvantage because of their religious practices and beliefs, in matters such as dress, food or prayer; here too, legal and policy barriers are needed.

Within AI governance, many standards and frameworks produced by international organisations, the UN and the OECD among them, require AI systems to respect, protect and fulfil freedom of religion and belief. What differs is the priority this right receives. Most international human rights frameworks make a broad commitment to protecting and promoting religious freedom as a human right, whereas Hifz al-Din goes a step further and makes it a binding legal obligation for the individual and for the state alike. When states balance competing interests, therefore, the protection of religion must be given clear priority.

Preservation of Life (Hifz al-Nafs): Safety, Dignity and the Limits of Automated Decision-Making

Hifz al-Nafs protects both the right to life and the right to dignity and it helps to trace red lines for AI governance in several areas. In health care, AI ought not to make diagnoses by itself. In the workplace, algorithmic management and the scoring of employees violate *karamah* (human dignity) and so do social scoring and social credit systems. It follows that a governance framework based on the Maqasid must contain a red line against social scoring and social credit.

Hifz al-Nafs, like the other Maqasid, coincides with international frameworks whenever AI-based technologies violate human dignity. Where the two diverge is in the weight attached to dignity. International principles and frameworks regard dignity as one value among others, to be balanced against them; under Hifz al-Nafs it is absolute and unconditional, since it derives from the Creator.

Preservation of Intellect (Hifz al-'Aql): Cognitive Autonomy, Explainability and Anti-Manipulation

Hifz al-'Aql protects rational capacity, including the ability to reason and to form moral judgements and this has practical implications for AI governance. It justifies methods that make opaque models more intelligible (e.g., LIME and SHAP) wherever reliance on such models produces harmful outcomes. It implies that educational AI should not simply supply students with correct answers, lest they come to depend on AI for moral judgement. It also means that recommender systems which steer behaviour by bypassing people's moral judgement and substituting appeals to emotional and psychological needs for it, are in breach of Hifz al-'Aql [30,32]. International bodies such as NIST, the OECD and UNESCO endorse the same safeguards. The chief difference lies elsewhere: international frameworks regard explainability as a procedure, while Hifz al-'Aql regards it as a moral duty.

Preservation of Progeny (Hifz al-Nasl): Intergenerational Justice and Algorithmic Discrimination

Hifz al-Nasl protects more than biological lineage; it also protects the conditions in which each generation can flourish, fair access to livelihoods and to social roles included [28,32]. In hiring and school admissions, AI trained on biased historical data can hand exclusion down from one generation to the next, as Raghavan *et al.* [7] show for algorithmic hiring. In education, digital learning tools may concentrate good learning in groups that are already advantaged and so deepen inequality. International instruments, including the work of the Committee on the Rights of the Child, focus mostly on present violations of children's rights, whereas Hifz al-Nasl explicitly extends the assessment to future generations. Both are concerned with current violations; they differ in how far ahead they look.

Preservation of Wealth (Hifz al-Mal): AI in Islamic Finance and Distributive Justice

Hifz al-Mal has a direct bearing on AI-driven financial systems, as three examples make clear. Take first the advanced risk-pricing models that Islamic banks may use to assess credit risk. Many present-day Shari'ah boards accept a conventional benchmark rate as a mere pricing reference, though some scholars regard it as undesirable [10].

Table 6: Full Thematic Mapping: Maqasid Objectives, AI Governance Principles, Convergences, Divergences and Governance Implications

Maqasid Objective	AI Governance Principle	Functional Convergence	Normative Divergence	Governance Implication
Hifz al-Din	Non-discrimination; Cultural diversity (UNESCO)	Protection of religious identity from AI bias	Divine obligation Vs procedural rights-based requirement	AI fatwa red line; Shari'ah-compliant content filters
Hifz al-Nafs	Safety; Human oversight; Risk mitigation (NIST, UNESCO)	Harm prevention; Human dignity - strongest convergence	Karamah absolute Vs dignity as one value among many	Mandatory human-in-the-loop; ban on mass biometric surveillance
Hifz al-Aql	Explainability (XAI); Anti-manipulation (UNESCO, NIST)	Cognitive autonomy; transparency as shared imperative	Sacred faculty Vs procedural governance requirement	Mandatory XAI (LIME/SHAP) in credit and education AI; ban on micro-targeted manipulation
Hifz al-Nasl	Fairness; Equity; Non-discrimination (OECD, UNESCO)	Social equity; access to opportunity	Intergenerational scope Vs present-generation focus	Fairness audits in hiring/education AI; intergenerational impact assessments
Hifz al-Mal	Accountability; Data protection; Auditability (NIST)	Distributive justice; fraud prevention; transparency	Riba/gharar absolute prohibition Vs risk-based thresholds	Shari'ah audit of credit scoring; differential privacy in robo-advisory; ban on riba-proxy variables

Source: Authors' own elaboration based on Al-Ghazali [27], Al-Shatibi [28], Auda [30], Kamali [32], El-Gamal [10], Iqbal and Mirakhor [17], UNESCO [4], NIST (2023), OECD [13]

The Shari'ah problem arises when a model's outputs serve to add a time-based increase to an existing debt, through automatic re-pricing, compounding or late-payment charges retained as income, for example; that would constitute riba al-nasi'ah. From a Maqasid perspective, then, such models have to be audited for their outputs and for the way these outputs enter contracts and not only for their inputs ([10,11]. The second example is a robo-advisor that allocates a client's portfolio on the basis of parametric uncertainty the client cannot understand. This may amount to excessive gharar if the uncertainty concerns the main subject of the contract and is not disclosed [33]. The third is high-frequency and algorithmic trading that exploits advantages of information and speed over ordinary investors [17], which raises concerns of gharar and unjust enrichment; algorithmic trading in currencies or gold with delayed or netted settlement may, in addition, breach the spot-settlement rule and thereby involve riba. In all three cases there is a conflict with Hifz al-Mal and with the circulation, clarity and justice of wealth on which Ibn Ashur [29] insisted (Table 6).

Illustrative Case Analysis: Malaysia, Algeria and an Islamic FinTech Vignette

In Malaysia, a mature Shari'ah governance regime coexists with active supervision of AI. The Islamic Financial Services Act 2013 and the central bank's Shariah Governance Policy Document [55] make Shari'ah committees and internal Shari'ah review mandatory. In August 2025 Bank Negara Malaysia issued a discussion paper setting out its proposed approach to AI in the financial sector [56] and its 2024 industry survey found that 71% of banking institutions and 77% of insurers and takaful operators had deployed at least one AI solution. The two regimes, however, run in parallel. The AI discussion paper is technology-neutral and the Shari'ah governance framework was never designed to tell Shari'ah committees how to review algorithmic models. Malaysia therefore possesses both components of MAFF but lacks the link between them.

Algeria is further behind. Regulation No. 20-02 of 15 March 2020 defines the banking operations that belong to Islamic finance and the conditions under which they may be carried out and Instruction No. 20-03 lists the products concerned [57]. An institution wishing to market these products must first obtain a certificate of Shari'ah conformity from the National Shari'ah Authority for Fatwa in the Islamic Finance Industry and only then apply to the Bank of Algeria for authorisation. The framework was consolidated by the Monetary and Banking Law No. 23-09 of 2023 [58] and personal data are protected under Law No. 18-07 of 2018 [59]. At the time of writing, we could not identify any Algerian instrument that addresses AI in financial services or requires Shari'ah certification to extend to algorithmic components. Algeria thus has a centralised certification mechanism, which is an asset for MAFF, since a single national authority could extend certification to AI-based products; what is missing are AI-specific rules and the technical capacity to use the mechanism for this purpose.

The vignette concerns a hypothetical Islamic FinTech start-up offering murabaha-based consumer finance through a mobile application. The firm uses (a) a gradient-boosted credit-scoring model fed with alternative data such as mobile-payment histories, (b) an LLM-based chatbot that answers customers' Shari'ah questions and (c) a robo-advisor that divides savings between sukuk and Shari'ah-screened equity funds. Applying MAFF to this case reveals where the risks lie and which tool addresses each of them. The scoring model passes F5 only if a SHAP-based audit of features and outputs shows that no variable or pricing rule adds a time-based increase to debt and that late-payment charges are given to charity. It passes F4 only if the disparate impact ratio and the equalized-odds gaps between, for instance, urban and rural applicants remain within the limits fixed by the firm's AI ethics committee.

Table 7: Illustrative Comparison of AI and Shari'ah Governance Arrangements in Malaysia and Algeria

Dimension	Malaysia	Algeria	Implication for MAFF
Shari'ah governance basis	Islamic Financial Services Act 2013; BNM Shariah Governance Policy Document (2019); institution-level Shari'ah committees	Regulation 20-02 and Instruction 20-03 (2020); Law 23-09 (2023); national Shari'ah conformity certification	Both offer an institutional anchor for F1-F5
AI-specific financial regulation	BNM discussion paper on AI in the financial sector (2025)	No AI-specific financial-sector instrument identified	Algeria needs an AI layer; Malaysia needs a Shari'ah-AI interface
Data protection	Personal Data Protection Act 2010	Law 18-07 (2018)	Legal basis for the privacy controls of F5
Locus of Shari'ah decision	Institution-level committees under the central Shariah Advisory Council	Centralised national certification	Centralisation eases uniform certification; decentralisation eases model-level review
Main gap	Parallel regimes; no guidance on Shari'ah review of algorithms	No AI rules; limited technical capacity	Joint Shari'ah-AI review function (Section 5.4)

Source: Authors' own elaboration based on Bank Negara Malaysia [55,56], Bank of Algeria [57], People's Democratic Republic of Algeria [58,59] and IFSB [26]

The chatbot breaches F1 if it generates rulings on its own; it becomes acceptable once it is restricted to retrieving rulings already approved by the Shari'ah board and passes difficult questions on to a human scholar. The robo-advisor satisfies F3 and F5 when it provides counterfactual, plain-language explanations of risk and discloses the uncertainty attached to its allocations. What the vignette suggests is that most Maqasid-based requirements can be met with existing techniques, provided that Shari'ah review and technical review are conducted jointly. Table 7 summarises the comparison.

DISCUSSION

From Utilitarian Efficiency to Maslahah-Oriented Ethics

The analysis indicates that the most influential international AI governance frameworks, those of the OECD, UNESCO and NIST, are essentially procedural. They explain how risk is to be identified, measured and managed, but impose few substantive limits on what may be optimised. Justice and fairness are present in these texts, yet usually as values to be weighed against efficiency and innovation. This agrees with Whittlestone *et al.* [60], who show that AI ethics principles are put into practice mainly through procedures and that the tensions among them are rarely resolved and with Jobin *et al.* [15], who find consensus on a small set of principles alongside wide variation in how they are interpreted and applied.

To these critiques we add a proposal: *maslahah* as an additional evaluative standard, one that gives greater weight to substantive justice and to the reduction of harm. The worry about current frameworks is that an outcome judged useful can, in principle, be justified regardless of the harm it causes. *Maslahah*, on the other hand, draws from the five *Maqasid* a set of harm-prevention floors that cannot be bargained away. An AI system that improved overall market efficiency while excluding a particular social group from credit would fail the test of *maslahah*. Our study therefore supports the use of the *Maqasid* as a limit on AI governance, even though our focus is not the same as that of Muchtasor [19], who examines this use of the *Maqasid* at length.

Human Moral Agency, Accountability and the Limits of Algorithmic Ijtihad

There is broad agreement that, for the foreseeable future, AI cannot bear moral responsibility for its actions [25,51]. In religious and legal terms, AI may be a moral patient, but not a moral agent. As moral and legal judgements are increasingly automated, in adjudication and regulatory compliance for instance, there is a danger that *ijtihad* will gradually give way to automated reasoning. We therefore treat algorithmic *ijtihad* as a governance problem that must be ruled out by clear restrictions. According to Auda [30], genuine *ijtihad* demands moral judgement and will exercised in context, something beyond the reach of current or future AI. Whatever automated assistance is employed, genuine *ijtihad* cannot itself be delegated to AI, however sophisticated the system may be.

The principle of stewardship (*khilafah*) places further demands on the design and application of AI governance. If AI systems are tools made and managed by human beings, they must be built in such a way that their stewards remain accountable and have the information required to supervise them. By this standard, the widespread industry practice of dispersing responsibility among developers, users and deploying organisations, so that ultimately no one answers for the outcome, cannot be accepted in Islamic governance. Financial and other institutions offering Shari'ah-compliant services must therefore exercise active oversight of the AI systems they use and this means reviewing how the systems are designed, not merely checking the finished products.

Technical AI Governance and Maqasid Ethics: A Productive Interface

A further contribution of this study lies in its engagement with the technical side of AI. Most research on AI and Islamic ethics [18-22,24,25] remains at the normative level. We maintain that particular technical tools can act as verifiable

means of fulfilling Maqasid-based duties: XAI methods in the case of Hifz al-'Aql, fairness metrics in the case of Hifz al-Nasl, privacy-preserving techniques in the case of Islamic data stewardship. LIME produces locally faithful explanations of individual decisions [6]. SHAP values attribute predictions to features and can thus reveal whether a prohibited proxy is driving a decision [40]. Counterfactual explanations offer the person affected an account on which they can act [41]. None of these tools is neutral, however. Post hoc explanations may be unstable or even misleading [39] and fairness metrics embody competing conceptions of justice that cannot all be met at once [44]. Choosing among them is not a mere technical setting but a normative decision, one in which Shari'ah and ethics committees ought to be involved.

Requiring credit-scoring models to be screened for proxies equivalent to *riba* also broadens algorithmic fairness auditing so that it accounts for the rules of Islamic finance, something neither the technical fairness literature nor the research on AI and Islamic finance has addressed until now [10,11]. Differential privacy may provide comparable protection against informational *gharar*. The examples discussed here concern governance and they could be transferred to other financial systems and other AI settings as well.

Ethical Pluralism and the Global AI Governance Debate

What the analysis shows is that elements of Maqasid al-Shariah and of international AI governance frameworks complement one another. Each addresses, in its own way, the core concerns of governance: preventing harm, ensuring accountability and transparency and respecting human dignity. Their foundations are nevertheless different. International frameworks generally rely on probabilistic, risk-based reasoning, whereas Maqasid al-Shariah rests on absolute, non-relativist grounds. Human dignity illustrates the point well. International law protects it through the assessment and management of risk, while the Maqasid treat it as a fixed limit [27,30,32].

As regards the debate between universalism and contextualism, we share the view of Whittlestone *et al.* [60] that deep normative differences persist beneath the apparent international consensus on AI ethics principles. Even where international frameworks and the Maqasid demand the same safeguards, they justify them on different grounds and weigh them differently. To claim that the two frameworks are fully aligned would be mistaken and so would the claim that their differences are unbridgeable. With Muchtasor [19], we hold that ethical pluralism allows Islamic ethical frameworks to enrich international governance rather than to supplant it.

Critical Appraisal: Juristic Disagreement, Political Risk and Legal Pluralism

The Maqasid-based approach has limits and they need to be acknowledged. It is not, to begin with, a uniform body of doctrine. Jurists disagree on the ranking of the objectives, on whether the list of five is closed or open [29,30], on whether *riba al-fadl* reaches beyond the six items named in the hadith and on specific financial practices such as the use of conventional benchmarks [10]. Different Shari'ah boards may come to different conclusions about one and the same algorithm and firms may then go 'fatwa shopping' for the most permissive opinion. Standardisation through AAOIFI and national Shari'ah authorities lessens this risk without eliminating it.

Incorporating religious norms into state regulation of AI carries political risks too. Religious authority may be invoked to justify surveillance or censorship in the name of protecting religion or public morals and a prohibition on 'AI fatwas' could be stretched to curb legitimate religious or academic expression. The Maqasid contain a corrective of their own here, since the preservation of intellect and the values of freedom and justice emphasised by Ibn Ashur [29] weigh against such uses. Institutional safeguards are needed as well, among them judicial review, transparent Shari'ah decisions and a clearly bounded scope for each filter. Legal pluralism is a third limit. In most Muslim-majority countries the legal system blends civil, commercial and Islamic law and many of these countries have non-Muslim citizens and foreign firms. Maqasid-based requirements are easiest to defend when they apply to Shari'ah-compliant products that consumers choose of their own accord, or when they coincide with general protections (dignity, non-discrimination, privacy) that can be justified on religious and secular grounds alike.

Enforcement: Who Governs Maqasid-Based AI and at What Level?

As a policy instrument, MAFF is only useful if it is clear who is to enforce it and we propose an architecture operating at several levels. Within each institution, the Shari'ah supervisory board, working together with an AI risk or ethics committee, approves AI systems before they go into operation and reviews them periodically, supported by internal Shari'ah review and audit (Section 5.4). Nationally, the central bank or financial regulator, jointly with a national Shari'ah authority (as in Algeria) or a central Shari'ah advisory council (as in Malaysia), issues binding guidance, extends Shari'ah certification to algorithmic components and sanctions breaches. Internationally, AAOIFI and the IFSB can translate MAFF into standards and guiding principles and thereby supply the harmonisation that national regimes lack. Multinational AI platforms that provide models or services to Islamic financial institutions are reached indirectly: The licensed institution remains responsible for outsourced models and regulators can make market access and procurement conditional on documentation (model cards, datasheets) and on a right to audit. The logic is that of the EU AI Act, which imposes obligations on providers and deployers of high-risk systems placed on the EU market [14].

Table 8: Stakeholder Analysis of Maqasid-Based AI Governance

Stakeholder	Main interest	Role under MAFF	Key tension or risk
Digital entrepreneurs and Islamic FinTech start-ups	Rapid product launch; low compliance cost; market trust	Apply F1-F5 by design; document models	Compliance costs may deter entry; requirements must be proportionate and tiered
Regulators and central banks	Financial stability; consumer protection; innovation	Issue guidance; supervise; link certification to licensing; run regulatory sandboxes	Limited AI expertise; risk of over-regulation
Shari'ah supervisory boards and national Shari'ah authorities	Shari'ah compliance; credibility	Approve and audit AI systems against F1-F5; define red lines	Limited technical competence; divergent opinions; independence
AI developers and vendors (including multinationals)	Scalability; protection of proprietary models	Provide explainability, documentation and audit access	Trade secrecy vs. transparency; limited knowledge of Shari'ah
Consumers and civil society	Fair access; understandable decisions; privacy; religious confidence	Exercise rights to explanation and redress; take part in consultations	Low financial and digital literacy; information asymmetry (gharar)

Source: Authors' own elaboration

Stakeholder Analysis

In practice, governance depends on the interests and capacities of the actors involved. Table 8 sets out the main stakeholders, what each of them wants, the role each would play under MAFF and the principal tensions that implementation will have to manage.

Governance Model and Policy Recommendations

Foundational Principles of the Maqasid-Based AI Governance Model: Four elements form the basis of the model. The first is complementarity, meaning that Maqasid-based governance supplements global AI regulation rather than replacing it. As Auda [30] and Kamali [32] suggest, additional substantive limits can help bring AI regulation into line with the requirements of the Shariah, while AI regulation itself needs moral and ethical content, supported by tools such as explanations, in order to deal with bias and equity. The second element is hard limits. In finance the Shariah sets firm boundaries around transactions involving excessive uncertainty (gharar) or unfair gain (riba) and firm boundaries are likewise needed elsewhere, against the use of AI for mass surveillance or for issuing fatwas, for example. The third element is accountability through human stewardship and the fourth a selective, context-sensitive integration of the Maqasid.

Operational Governance Mechanisms: The Maqasid AI Filter Framework (MAFF)

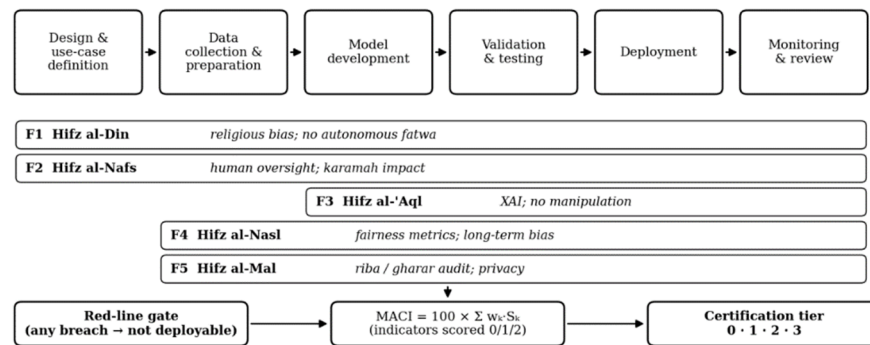
The Maqasid AI Filter Framework (MAFF) is made up of five filters, each named after one of the Maqasid and applied at the design stage of AI systems. They are intended to operate alongside whatever AI governance framework an organisation has already adopted. Several such frameworks exist, the NIST AI RMF and the UNESCO Recommendation among them and organisations that follow them already face a long list of requirements; MAFF adds Maqasid-based provisions to that list for organisations active in Islamic financial services.

MAFF is flexible and can be combined with other AI governance controls. Under Filter F1 (Hifz al-Din), the system must show no religious bias and, whenever it provides religious information, must draw only on sources approved by a qualified Shari'ah authority and must never issue rulings on its own. Filter F2 (Hifz al-Nafs) calls for human validation of decisions affecting life, health, livelihood or liberty and for an assessment of their effects on human dignity (karamah). Filter F3 (Hifz al-'Aql) demands that high-risk systems be explainable to the person affected and that they contain no design features intended to exploit cognitive biases. Filter F4 (Hifz al-Nasl) requires that algorithms be evaluated with explicit fairness metrics, such as the disparate impact ratio, equalized odds or counterfactual fairness, including their long-term effects across generations. Filter F5 (Hifz al-Mal) requires that financial AI systems be checked for Shari'ah compliance and above all for the absence of riba and excessive gharar, before deployment. The filters apply throughout the AI lifecycle rather than in a set sequence, but each carries a red line and a system that crosses any of them cannot be deployed, however well it performs on other criteria (Table 9). Figure 1 locates the filters within the lifecycle and within the institutional audit model.

Measuring Compliance: the Maqasid-AI Compliance Index (MACI) and Certification Tiers

To make the checklist measurable, we propose the Maqasid-AI Compliance Index (MACI). Each filter F_k ($k = 1, \dots, 5$) is evaluated through three indicators (Table 10), each scored 0 (absent), 1 (partly met) or 2 (fully met) on the basis of documentation and technical tests. The filter score S_k equals the sum of its indicator scores divided by the maximum attainable (6), so that $0 \leq S_k \leq 1$. The index is then $MACI = 100 \times (w_1 \cdot S_1 + w_2 \cdot S_2 + w_3 \cdot S_3 + w_4 \cdot S_4 + w_5 \cdot S_5)$, where the weights w_k sum to 1.

A. MAFF embedded in the AI lifecycle



B. Three-lines model for Shari'ah AI auditing

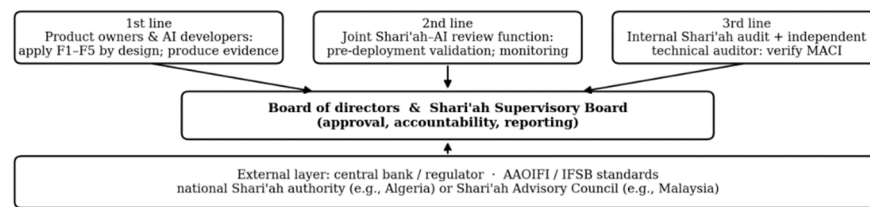


Figure 1: The Maqasid AI Filter Framework (MAFF) Embedded in the AI Lifecycle and the Three-Lines Model for Shari'ah AI Auditing

Table 9: The Maqasid AI Filter Framework (MAFF): Governance Requirements, Technical Instruments and Institutional Responsibilities

Filter	Governance requirement	Technical instrument	Institutional responsibility	Red line/absolute prohibition
F1: Hifz al-Din	Religious identity bias audit; Shari'ah-compliant content policy	Demographic parity audit; content filtering; bias detection on religious proxies	Shari'ah Supervisory Board; AI ethics committee	Autonomous fatwa issuance by AI; AI-driven religious adjudication
F2: Hifz al-Nafs	Mandatory human-in-the-loop for life-affecting decisions; karamah impact assessment	Human oversight protocols; AI impact assessment tools; audit logs	Regulatory authority; institutional AI governance office	Mass biometric surveillance; social scoring systems
F3: Hifz al-'Aql	Mandatory XAI in high-stakes contexts; cognitive autonomy protection	LIME [6]; SHAP [40]; model cards; explainability documentation	AI developers; deploying institutions; independent auditors	Micro-targeted psychological manipulation; black-box systems in credit/education
F4: Hifz al-Nasl	Intergenerational impact assessment; fairness audits in hiring and education AI	Equalized odds [5]; counterfactual fairness; longitudinal bias monitoring	Labour ministry; education regulators; independent fairness auditors	Automated systems with unchecked structural discrimination against protected groups
F5: Hifz al-Mal	Shari'ah audit of AI feature sets; riba/gharar compliance; differential privacy in data processing	Differential privacy [49]; feature importance audit (SHAP); Shari'ah-compliant model cards	Shari'ah Supervisory Board; central bank; Islamic finance regulatory authority	AI systems facilitating riba or gharar in any form; opaque scoring models without Shari'ah audit

Source: Authors' own elaboration based on Al-Ghazali [27], Al-Shatibi [28], Auda [30], Kamali [32], El-Gamal [10], Iqbal and Mirakhor[17], Ribeiro *et al.* [6], Lundberg and Lee [40], Dwork and Roth [49], Kusner *et al.* [5]

Table 10: Indicators of the Maqasid-AI Compliance Index (MACI)

Filter	Indicator 1	Indicator 2	Indicator 3	Evidence
F1 Hifz al-Din	Religious-bias test on outputs and proxy variables	Religious content restricted to board-approved sources (e.g., RAG)	Escalation of rulings to a human scholar	Test reports; system configuration; escalation logs
F2 Hifz al-Nafs	Human-in-the-loop for consequential decisions	Karamah/AI impact assessment completed	Complaint and redress mechanism	Oversight protocol; impact assessment; redress records
F3 Hifz al-'Aql	Explanation available to affected persons (local XAI or counterfactual)	Model card and explainability documentation	No manipulative design (dark patterns, micro-targeting)	Sample explanations; model card; user-interface review
F4 Hifz al-Nasl	Fairness metrics reported across protected and vulnerable groups	Thresholds met (e.g., disparate impact ratio ≥ 0.8)	Longitudinal bias monitoring	Fairness audit; monitoring dashboard
F5 Hifz al-Mal	Shari'ah audit of features and outputs (no riba-proxies; Late charges to charity)	Disclosure of model uncertainty to clients (gharar)	Privacy protection (data minimisation; differential privacy where feasible)	Shari'ah audit report; disclosures; data-protection assessment

Note: Each indicator is scored 0/1/2; Sk = sum of indicator scores/6; $MACI = 100 \times \Sigma wk \cdot Sk$; any red-line breach sets MACI to 0. Source: Authors' own elaboration

Two design choices reflect the logic of the Maqasid. The first is that red lines take precedence: should any red line in Table 9 be crossed, MACI is set to zero and the system cannot be certified, because the protection of necessities (*daruriyyat*) cannot be exchanged for gains elsewhere [28]. The second is that weights vary by field. Our default is equal weighting ($w_k = 0.2$), but for Islamic financial services a heavier weight on F5 (e.g., 0.30, with 0.175 for each remaining filter) reflects the central objective of the field. Final weights should be fixed through a Delphi consultation involving Shari'ah scholars, AI experts and regulators. We suggest four certification tiers: Tier 0, not deployable (a red line has been crossed); Tier 1, conditional (MACI below 60, with a remediation plan required); Tier 2, compliant (60--84); and Tier 3, exemplary (85 or above). These thresholds are indicative and will need empirical calibration. Consider the credit-scoring model from the vignette in Section 3.6. Assuming it crosses no red line and fully satisfies the indicators of F2, F3 and F5, but lacks religious-bias testing ($S1 = 4/6$) and longitudinal bias monitoring ($S4 = 4/6$), it would obtain roughly 87 with equal weights and roughly 88 with the finance-specific weights and would therefore fall in Tier 3.

An Institutional Model for Shari'ah AI Auditing

MAFF and MACI require an institutional setting. Building on the Three Lines Model of the Institute of Internal Auditors [62] and on the IFSB principles on Shari'ah governance [34,36], we propose the model depicted in Figure 1. In the first line, product owners and AI developers build F1--F5 into the design and generate the evidence listed in Table 10. In the second, a joint Shari'ah--AI review function, which brings the Shari'ah compliance unit together with specialists in model risk and data protection, validates models before deployment and monitors them thereafter. In the third, internal Shari'ah audit, assisted by an independent technical auditor, verifies the MACI score periodically and reports to the board of directors and to the Shari'ah supervisory board. Outside the institution, the regulator and the national Shari'ah authority determine red lines and thresholds, recognise certification and conduct thematic reviews. The model further requires that at least one member of the Shari'ah board or of the review function possess recognised expertise in data science, or else that the board have formal access to such expertise.

Concrete Policy Recommendations

From MAFF, MACI and the evidence presented in Sections 3 and 4 we derive five policy recommendations. The first concerns regulators and central banks in countries with an Islamic financial system, who should require Islamic financial institutions to submit the AI systems they use in customer-facing or Shari'ah-relevant decisions to joint review by the Shari'ah supervisory board and AI governance experts, in order to establish, among other things, whether those systems contain *riba*-equivalents or excessive *gharar*. In Algeria, one way of doing this would be to extend the conformity certification issued by the national Shari'ah authority to algorithmic components.

Second, the national AI strategies of Muslim-majority countries ought to lay down Maqasid-based red lines, ruling out, for example, autonomous religious rulings and indiscriminate mass biometric surveillance, while remaining consistent with international standards such as those of NIST, the OECD and UNESCO.

Third, AI companies and FinTech entrepreneurs that develop, sell or deploy AI systems in Islamic finance should adopt Shari'ah-by-design practices, document their models by means of model cards and datasheets and grant Shari'ah and technical auditors the access required to verify that neither *riba* nor excessive *gharar* is present.

Fourth, international standard-setting bodies such as UNESCO, ISO and the OECD should involve Islamic jurists and specialists in Islamic finance governance in their standard-setting work. AAOIFI and the IFSB, for their part, could build on MAFF to produce a dedicated standard, or a set of guiding principles, on AI in Islamic financial services. In that way the prohibitions of *riba* and *gharar* would be recognised as governance limits specific to the field, alongside limits grounded in human rights.

Fifth, the research agenda should take in case studies of Islamic financial institutions using AI in Malaysia, Saudi Arabia, the United Arab Emirates, Indonesia and Algeria, regulatory impact assessments of MAFF and MACI and a Delphi consultation of Shari'ah scholars, AI experts and other stakeholders aimed at calibrating indicators, weights, thresholds and red lines.

Policy Feasibility Considerations

Whether MAFF can work well depends on a number of conditions. One is the capacity of Shari'ah supervisory boards to evaluate AI systems. Most boards have a sound knowledge of Islamic finance and law but little familiarity with AI, so scholars will need training in model explainability, bias and privacy and Shari'ah review functions will have to recruit data-science expertise. Until such capacity is in place, MAFF can be applied only partially.

A second condition is the development of regulatory frameworks for AI in Islamic finance. Only a small number of central banks and financial regulators have begun to address AI in the financial sector and fewer still have linked AI oversight to Shari'ah governance, as the Malaysian and Algerian cases illustrate [17,11,56].

The third condition is cooperation from technology firms. Many of them are based outside Muslim-majority countries and have few incentives to design their systems around MAFF. Securing their cooperation will require market-access conditions, procurement rules and the involvement of the international organisations mentioned above; at the same time, proportionate and tiered requirements should prevent compliance costs from excluding small entrepreneurs.

CONCLUSION

This study set out to examine whether selected aspects of Maqasid al-Shariah can be combined with international approaches to AI governance in order to regulate AI in digital and Islamic finance. On the strength of a qualitative comparative analysis conducted with an explicit coding grid and tested against an illustrative analysis of Malaysia and Algeria, our answer to the research question is a conditional yes. Integration is possible, provided it is selective, attentive to context and anchored in institutions.

Set against international AI governance principles, the five Maqasid objectives agree with them in practice on preventing harm, protecting human dignity, accountability and transparency. The disagreement concerns the source of the norms and whether they can be traded off. International frameworks are for the most part procedural and risk-based, while the Maqasid impose substantive limits, some of which are absolute. Nowhere is the contrast sharper than in finance, where the prohibitions of *riba* and excessive *gharar* operate as fixed constraints on algorithm design [10,27,30,32].

Such differences do not preclude integration. MAFF converts each of the five objectives into a design-stage filter with verifiable requirements, technical tools and red lines; MACI translates those requirements into a measurable index with certification tiers; and the three-lines model allocates institutional responsibility for Shari'ah AI auditing. The package provides, among other safeguards, Shari'ah audit of credit-scoring systems, explanations through LIME, SHAP or counterfactuals, fairness testing, protection of financial data and outright prohibitions on autonomous fatwas and mass biometric surveillance. It thereby addresses the three gaps identified in the literature, namely the absence of a structured comparative framework, of Maqasid-based limits for AI in Islamic finance and of a practical governance model [10,17,11].

Among our central findings is that AI systems cannot be regarded as moral agents, since they possess neither intention (*niyyah*) nor moral understanding. The highest level of moral and legal judgement has to remain with human beings and Maqasid-based governance does not permit *ijtihad* to be delegated to automated systems, which must operate within safeguards that protect human moral and legal judgement.

The study also suggests that Islamic moral philosophy is able to draw at once on classical Islamic legal thought and on current AI research and that the combination can yield ethical frameworks sufficiently refined and sufficiently broad in scope, to contribute to the global ethical debate. Incorporating Maqasid principles would make international AI standards and frameworks more responsive to the cultures of Muslim-majority countries and communities. The argument for ethical pluralism in global AI regulation is gaining strength and a pluralist framework that takes normative and ethical differences seriously is more legitimate, more durable and better able to earn public trust [19,52,60].

The study is not without limitations. It is conceptual in nature: the illustrative case analysis relies on official documents and a stylised vignette rather than on field data. Future research should therefore include case studies of Islamic financial institutions that employ AI, particularly in Malaysia, Saudi Arabia, the UAE and Algeria, as well as regulatory impact assessments of MAFF and MACI and Delphi or multi-stakeholder consultations to calibrate indicators, weights and red lines. The study also depends on Sunni and mainly Shafi'i and Maliki, readings of the Maqasid; work that brings in Hanafi, Hanbali and Shi'i perspectives would test the robustness of the framework. Lastly, because AI technologies and regulations change rapidly, the framework will require periodic revision, especially where generative AI is concerned.

REFERENCES

1. LeCun, Y. *et al.* "Deep learning." *Nature*, vol. 521, no. 7553, 2015, pp. 436-444.
2. Russell, S. and P. Norvig. *Artificial intelligence: A modern approach*. 4th Edn., Pearson, 2020.
3. World Bank. *The Global Findex Database 2017: measuring financial inclusion and the fintech revolution*. World Bank, 2018.
4. UNESCO. *Recommendation on the ethics of artificial intelligence*. UNESCO, 2021.
5. Kusner, M.J. *et al.* "Counterfactual fairness." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
6. Ribeiro, M.T. *et al.* "Why should I trust you? Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, pp. 1135-1144.
7. Raghuvaran, M. *et al.* "Mitigating bias in algorithmic hiring: Evaluating claims and practices." *Proceedings of the Conference on Fairness, Accountability and Transparency*, ACM, 2020, pp. 469-481.
8. Angwin, J. *et al.* "Machine bias." *ProPublica*, 23 May 2016.

9. Fuster, A. *et al.* "Predictably unequal? The effects of machine learning on credit markets." *The Journal of Finance*, vol. 77, no. 1, 2022, pp. 5-47.
10. El-Gamal, M.A. *Islamic finance: Law, economics and practice*. Cambridge University Press, 2006.
11. Najib, N.W.M. *et al.* "Artificial Intelligence (AI) in Islamic finance: A Maqasid al-Shariah perspective." *International Journal of Law, Government and Communication*, vol. 10, no. 40, 2025, pp. 41-50. doi:10.35631/IJLGC.1040003.
12. NIST. Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology, 2023.
13. OECD. Recommendation of the Council on artificial intelligence (OECD/LEGAL/0449). OECD, 2019.
14. European Union. "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)." *Official Journal of the European Union*, L, 2024/1689, 2024.
15. Jobin, A. *et al.* "The global landscape of AI ethics guidelines." *Nature Machine Intelligence*, vol. 1, no. 9, 2019, pp. 389-399.
16. Mittelstadt, B. "Principles alone cannot guarantee ethical AI." *Nature Machine Intelligence*, vol. 1, no. 11, 2019, pp. 501-507.
17. Iqbal, Z. and A. Mirakhor. *An Introduction to Islamic Finance: Theory and Practice*. 2nd Edn., Wiley, 2011.
18. Kannike, U.M.M. and A.O. Fahm. "Exploring the ethical governance of artificial intelligence from an Islamic ethical perspective." *Jurnal Fiqh*, vol. 22, no. 1, 2025, pp. 134-161. doi:10.22452/fiqh.vol22no1.5.
19. Muchtasor, N.S. "Ethical pluralism in AI policy: A framework for Islamic integration into global AI governance." *Sinergi International Journal of Islamic Studies*, vol. 3, no. 3, 2025, pp. 191-203.
20. Nikmah, M. and Nasikhin. "God-conscious AI: Maqasid al-Shari'ah in algorithmic design." *Salam Institute Islamic Studies*, vol. 2, no. 1, 2025, pp. 13-20.
21. Noordin, M.F. and S.N.S. Zailan. "AI from an Islamic perspective." *Journal of Information Systems and Digital Technologies*, vol. 7, no. 2, 2025, pp. 102-114.
22. Mustapha, R. *et al.* "Islamic perspectives on AI and cybersecurity: Developing ethical frameworks for autonomous security systems." *International Journal of Law, Government and Communication*, vol. 10, no. 40, 2025. doi:10.35631/IJLGC.1040041.
23. Raquib, A. *et al.* "Islamic virtue-based ethics for artificial intelligence." *Discover Artificial Intelligence*, vol. 2, 2022, article 11.
24. Thaib, M. "The interface of Maqashid al-Shariah with human cognitive capacity and AI innovation." *Jurnal Integrasi Ilmu Syariah (Jisrah)*, vol. 6, no. 3, 2025, pp. 134-139.
25. Waqar, K.M. and M. Ibrahim. "Ethical implications of artificial intelligence: An Islamic perspective." *Journal of Religion & Society*, vol. 3, no. 1, 2025, pp. 347-358.
26. IFSB. *Islamic financial services industry stability report 2025*. Islamic Financial Services Board, 2025b.
27. Al-Ghazali, A.H. *Al-Mustasfa min 'ilm al-usul*. Dar al-Kutub al-'Ilmiyyah, 1993. Originally published ca. 1109.
28. Al-Shatibi, I. *Al-Muwafaqat fi usul al-shari'ah*. Dar Ibn 'Affan, 1997. Originally published ca. 1388.
29. Ibn Ashur, M. al-T. *Treatise on Maqasid al-Shari'ah*. Translated by M. El-Tahir El-Mesawi, International Institute of Islamic Thought, 2006.
30. Auda, J. *Maqasid al-Shariah as philosophy of Islamic law: A systems approach*. International Institute of Islamic Thought, 2008.
31. Kamali, M.H. *Principles of Islamic jurisprudence*. 3rd Edn., Islamic Texts Society, 2003.
32. Kamali, M.H. *Shari'ah law: An introduction*. Oneworld, 2008.
33. AAOIFI. *Shari'ah Standards*. Accounting and Auditing Organization for Islamic Financial Institutions, 2017.
34. IFSB. *Guiding principles on Shari'ah governance systems for institutions offering Islamic financial services (IFSB-10)*. Islamic Financial Services Board, 2009.
35. IFSB. *Revised guiding principles on corporate governance for institutions offering Islamic financial services (IFSB-30)*. Islamic Financial Services Board, 2023.
36. IFSB. *Guiding principles for effective supervision of Shari'ah governance (IFSB-31)*. Islamic Financial Services Board, 2025a.
37. NIST. Artificial intelligence risk management framework: generative artificial intelligence profile (NIST AI 600-1). National Institute of Standards and Technology, 2024.
38. Arrieta, A.B. *et al.* "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information Fusion*, vol. 58, 2020, pp. 82-115.
39. Rudin, C. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." *Nature Machine Intelligence*, vol. 1, no. 5, 2019, pp. 206-215.

40. Lundberg, S.M. and S.I. Lee. "A unified approach to interpreting model predictions." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
41. Wachter, S. *et al.* "Counterfactual explanations without opening the black box: Automated decisions and the GDPR." *Harvard Journal of Law and Technology*, vol. 31, no. 2, 2018, pp. 841-887.
42. Hardt, M. *et al.* "Equality of opportunity in supervised learning." *Advances in Neural Information Processing Systems*, vol. 29, 2016.
43. Chouldechova, A. "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments." *Big Data*, vol. 5, no. 2, 2017, pp. 153-163.
44. Kleinberg, J. *et al.* "Inherent trade-offs in the fair determination of risk scores." *Proceedings of the 8th Innovations in Theoretical Computer Science Conference, LIPIcs*, vol. 67, 2017, pp. 43:1-43:23.
45. Barocas, S. *et al.* *Fairness and machine learning: limitations and opportunities*. MIT Press, 2023.
46. Mehrabi, N. *et al.* "A survey on bias and fairness in machine learning." *ACM Computing Surveys*, vol. 54, no. 6, 2021, pp. 1-35.
47. Gebru, T. *et al.* "Datasheets for datasets." *Communications of the ACM*, vol. 64, no. 12, 2021, pp. 86-92.
48. Mitchell, M. *et al.* "Model cards for model reporting." *Proceedings of the Conference on Fairness, Accountability and Transparency*, ACM, 2019, pp. 220-229.
49. Dwork, C. and A. Roth. "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science*, vol. 9, nos. 3-4, 2014, pp. 211-407.
50. Lewis, P. *et al.* "Retrieval-augmented generation for knowledge-intensive NLP tasks." *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459-9474.
51. Floridi, L. *et al.* "AI4People-an ethical framework for a good AI society: Opportunities, risks, principles and recommendations." *Minds and Machines*, vol. 28, no. 4, 2018, pp. 689-707.
52. Floridi, L. and J. Cowls. "A unified framework of five principles for AI in society." *Harvard Data Science Review*, vol. 1, no. 1, 2019.
53. Mittelstadt, B.D. *et al.* "The ethics of algorithms: Mapping the debate." *Big Data and Society*, vol. 3, no. 2, 2016.
54. Bostrom, N. and E. Yudkowsky. "The ethics of artificial intelligence." In: *The Cambridge Handbook of Artificial Intelligence*, K. Frankish and W.M. Ramsey (Eds.), Cambridge University Press, 2014, pp. 316-334.
55. Bank Negara Malaysia. *Shariah governance*. Bank Negara Malaysia, 2019.
56. Bank Negara Malaysia. Discussion paper on artificial intelligence in the Malaysian financial sector. Bank Negara Malaysia, 5 Aug. 2025.
57. Bank of Algeria. Règlement n° 20-02 du 15 mars 2020 définissant les opérations de banque relevant de la finance islamique et les conditions de leur exercice par les banques et établissements financiers. Banque d'Algérie, 2020.
58. People's Democratic Republic of Algeria. Law No. 23-09 of 21 June 2023 on the monetary and banking law. *Journal Officiel de la République Algérienne*, 2023.
59. People's Democratic Republic of Algeria. Law No. 18-07 of 10 June 2018 on the protection of natural persons in the processing of personal data. *Journal Officiel de la République Algérienne*, 2018.
60. Whittlestone, J. *et al.* "The role and limits of principles in AI ethics: Towards a focus on tensions." *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics and Society*, ACM, 2019, pp. 195-200.
61. Institute of Internal Auditors. *The IIA's three lines model: An update of the three lines of defense*. The IIA, 2020.AOIFI.