

EXPLAINABILITY AS A SHARĪ'AH OBLIGATION: A MAQĀSID GOVERNANCE MODEL FOR AI IN ISLAMIC FINANCE GROUNDED IN AL-SARAKHSĪ'S MU'ĀMALĀT

Muhammad Syafiq Bin Abdul Aziz^{1,a*}, Mohd Shahid Bin Mohd Noh^{2,b} and Muhamad Azhari Wahid^{3,c}

¹Department of Fiqh-usul and Applied Sciences, Academy of Islamic Studies, Universiti Malaya, Kuala Lumpur, Malaysia

²Department of Shariah, Economics and Governance, Academy of Islamic Studies, Universiti Malaya, Kuala Lumpur, Malaysia

³Faculty of Economics & Muamalat, Islamic Science University of Malaysia, Negeri Sembilan, Malaysia

^a md.syafiq.abdulaziz@gmail.com

^b shahid82@um.edu.my

^c azhariwahid@usim.edu.my

*Corresponding Author: md.syafiq.abdulaziz@gmail.com

Abstract: This study examines the governance gap that has emerged from the rapid adoption of artificial intelligence in Islamic finance, particularly in credit scoring, *ṣukūk* pricing, and *takāful* underwriting, where black-box machine learning models generate decisions whose opacity raises substantive concerns under the Sharī'ah principles of 'adl, *riḍā*, *amānah*, and contractual transparency, and introduces a form of uncertainty analogous to *gharar* that weakens Sharī'ah auditability. The research problem addressed is the absence of an operational governance framework that integrates Explainable AI (XAI) techniques with Islamic jurisprudential epistemology, since existing XAI tools such as LIME and SHAP were not designed with reference to *maqāsid al-sharī'ah* reasoning. The study adopts a qualitative doctrinal research design combined with conceptual framework development, drawing on *Imām al-Sarakhsī's Al-Mabsūṭ* as the primary classical source and supplementing the doctrinal analysis with two illustrative case studies on *ṣukūk* pricing and *takāful* underwriting. The main findings indicate that four jurisprudential principles reconstructed from *al-Sarakhsī's mu'āmalāt* theory, namely *bayān*, *daf' al-ḍarar*, *yaqīn*, and *maṣlahah*, can be operationalised into a three-layer interpretive structure comprising probabilistic *bayān*, justificatory *ta'līl*, and *maṣlahah*-based evaluation, which together form the proposed *Maqāsid-Explainability Governance Model (MEGM)*. The author's analysis suggests that layered interpretability can satisfy certain auditability requirements articulated by AAOIFI and IFSB that conventional XAI approaches, when applied in isolation, may not fully meet, and that explainability in Islamic finance should be understood not merely as a technical preference but as a fiduciary obligation. The study concludes that *al-Sarakhsī's mu'āmalāt* theory provides a working jurisprudential foundation for ethical AI governance, and recommends that Islamic financial institutions, Sharī'ah committees, and regulators adopt the MEGM as a structured tool for auditing automated decision-making systems.

Keywords: Explainable AI, *Maqāsid al-Sharī'ah*, *Al-Sarakhsī*, *Mu'āmalāt*, Sharī'ah Governance

1. Introduction

The integration of artificial intelligence into Islamic financial services has accelerated faster than the jurisprudential infrastructure required to govern it. Across the global Islamic finance industry, machine learning models are now being deployed to support credit scoring decisions, to assist in *ṣukūk* pricing, to assess underwriting risk in *takāful* contracts, and to flag potentially non-compliant transactions for Sharī'ah review (IFSB, 2020, 2024; AAOIFI, 2023; Shalhoob, 2025). Industry data confirm that AI is already embedded in Islamic banking operations: an IFSB survey reports identity verification (67 per cent), chatbots and virtual assistants (56 per cent), and digital-footprint analysis (44 per cent) as the most common applications (IFSB, 2025). The sector in which these tools operate is itself substantial and concentrated: global Islamic finance assets reached USD 3.88 trillion at the end of 2024, a 14.9 per cent year-on-year increase (IFSB, 2025; ICD–LSEG, 2025), with the Gulf Cooperation Council accounting for 53.1 per cent of those assets and the East Asia and Pacific region a further 21.9 per cent, with activity concentrated in jurisdictions with established Sharī'ah supervisory architectures (Bank Negara Malaysia, 2019). Within those jurisdictions, recent scholarship in this journal has identified persistent operational gaps in Sharī'ah audit practice across institutional sectors (Kamaruddin et al., 2024) and has examined how Sharī'ah committees interact with fintech innovation in Islamic banks (Mohd Haridan et al., 2023); the present study extends that line of inquiry from product-level review to algorithmic decision-making review. This trajectory raises a governance question that is jurisprudentially urgent: when a model whose internal logic is opaque produces a decision that affects a Muslim contracting party, can that decision satisfy the conditions of a valid *mu'āmalah* under classical Sharī'ah doctrine?

The concern is concrete. Opacity in algorithmic decision-making cuts against several foundational principles of *fiqh al-mu'āmalāt*. The requirement of *'adl* (justice) demands that the parties to a contract be treated equitably and on terms whose grounds are knowable; the principle of *riḍā'* (informed consent) presupposes that consent is given on the basis of adequate disclosure; the doctrine of *amānah* (fiduciary trust) imposes on the institution a duty to act with transparency toward those whose interests it manages; and the prohibition of *gharar* restricts the validity of contracts whose material terms are concealed in excessive uncertainty (al-Zuhaylī, 2002; Kamali, 2000, 2007; Al-Suwailem, 2006). When the institution itself cannot reconstruct why its model produced a particular decision, these principles are not merely strained but structurally compromised; the broader literature has documented this opacity as a feature of the machine-learning techniques themselves (Pasquale, 2015; Burrell, 2016).

A parallel literature in computer science, generally referred to as Explainable Artificial Intelligence or XAI, has developed in response to similar concerns in secular contexts (Arrieta et al., 2020; Guidotti et al., 2018). Techniques such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) generate post-hoc approximations of how a model arrived at a particular output, identifying the features that most influenced the decision (Ribeiro et al., 2016; Lundberg & Lee, 2017), and have been applied to financial tasks including fraud detection and lending (Awosika et al., 2024; Babaei et al., 2023). The conceptual status of such post-hoc methods is, however, contested: Rudin (2019) argues that high-stakes decisions should rely on inherently interpretable models rather than on post-hoc explanations of opaque ones, while methodological surveys note that the field still lacks agreed criteria for what makes an explanation adequate (Doshi-Velez & Kim, 2017; Lipton, 2018; Molnar, 2022; Saeed & Omlin, 2023). A related strand in AI ethics has converged on broadly shared principles of transparency and accountability (Floridi et al., 2018; Jobin et al., 2019), yet has also cautioned that high-level principles alone cannot determine concrete

obligations in a particular domain (Mittelstadt, 2019; Mittelstadt et al., 2016), and that a meaningful right to explanation requires institutionally specified standards rather than technical outputs alone (Selbst & Powles, 2017; Selbst & Barocas, 2018; Wachter et al., 2018). While these techniques and frameworks have meaningfully reduced technical opacity, they were not designed with reference to Islamic jurisprudential epistemology, and they do not address the normative question of whether an explanation is jurisprudentially adequate. A SHAP attribution that satisfies a data scientist may leave a Shari‘ah committee unconvinced, and the distance between the two is conceptual rather than terminological.

The literature situated at the intersection of these concerns remains thin. Recent work on AI ethics in Islamic finance has tended to address Shari‘ah compliance at the product level rather than at the level of algorithmic decision-making (Mohamed & Ali, 2018; Bin-Nashwan et al., 2022). The broader *maqāsid al-shari‘ah* literature has been productively renewed by Auda (2008), Kamali (2008), and El-Mesawi’s rendering of Ibn ‘Ashūr (2006), and has been applied to the architecture of Islamic finance by Laldin and Furqani (2013) and to its contemporary governance and fintech challenges by the contributors to Oseni and Ali (2019), but operational application to algorithmic governance has not yet been attempted. The classical sources of *fiqh al-mu‘āmalāt* remain underexploited in this respect; Imām al-Sarakhsī’s *Al-Mabsūt*, the most comprehensive Hanafi treatment of contractual jurisprudence, has been cited in passing in contemporary Islamic finance scholarship but has not been used as the principal framework for AI governance. This paper takes that step.

Three research objectives organise the inquiry. First, the study seeks to identify, from al-Sarakhsī’s *Al-Mabsūt*, the jurisprudential principles most directly relevant to the governance of automated decision-making, with particular attention to *bayān* (clarity), *daf‘ al-ḍarar* (harm prevention), *yaqīn* (legal certainty), and *maṣlahah* (public welfare). Second, the study seeks to operationalise these principles into a governance framework that interfaces with existing Explainable AI techniques, producing the Maqāsid-Explainability Governance Model (MEGM). Third, the study seeks to assess, through two illustrative case applications, whether the proposed model satisfies the auditability requirements articulated by AAOIFI and IFSB more fully than conventional XAI deployed in isolation.

The remainder of the paper proceeds as follows. The next section sets out the methodological commitments of the study. The third section develops the theoretical foundation by reconstructing the four governance principles from al-Sarakhsī’s *Al-Mabsūt*. The fourth section presents the Maqāsid-Explainability Governance Model in architectural detail. The fifth section applies the model to *ṣukūk* pricing and *takāful* underwriting through illustrative case analyses. The sixth section discusses implications for AAOIFI and IFSB alignment, Shari‘ah governance practice, and *maqāsid* theory more broadly. The paper closes with a summary of contributions, an acknowledgement of limitations, and three directions for future research.

2. Methodology

This study adopts a qualitative research design that combines doctrinal analysis of classical Islamic legal texts with conceptual framework development, broadly in the spirit of design science research (Hevner et al., 2004; Peffers et al., 2007), although the study stops short of the empirical artefact-evaluation that the design science paradigm requires and is therefore best characterised as a doctrinal-conceptual study rather than a full design science project. The doctrinal dimension of the methodology follows the conventions articulated by Hutchinson and Duncan (2012), in which the primary data of the research are authoritative texts whose

meanings are reconstructed through interpretive engagement rather than through empirical observation. The study makes no claim to empirical validation: its output is a conceptual artefact, and the case applications in Section 5 are illustrative demonstrations of the framework's internal logic rather than tests of its effectiveness in deployment.

The primary classical source is Imām al-Sarakhsī's *Al-Mabsūṭ*, consulted in the Dār al-Maʿrifah Beirut edition (1993) in thirty volumes. The Hanafī school is selected over the Shāfiʿī, Mālikī, and Ḥanbalī schools for two reasons. First, the Hanafī tradition offers the most developed treatment of contractual epistemology and the doctrine of *bayān*, which is directly relevant to the question of explanation in automated decision-making. Second, al-Sarakhsī's *Al-Mabsūṭ* is the most comprehensive single Hanafī treatment of *muʿāmalāt* in the classical period, integrating jurisprudential reasoning with a level of doctrinal detail that few other works of the period match (Nyazee, 2000; Hallaq, 2009). Reading is focused on the sections of *Al-Mabsūṭ* most directly relevant to the research questions, namely *Kitāb al-Buyūʿ* (the Book of Sales) and *Kitāb al-Istiḥsān* (Juristic Preference), with cross-references to *Kitāb al-Ijārāt* (Hire), *Kitāb al-Sharikah* (Partnership), and *Kitāb al-Rahn* (Pledge) where the argument requires. The four governing principles examined below, namely *bayān* (disclosure), *dafʿ al-ḍarar* (harm prevention), *yaqīn* (the security of an established position), and *maṣlaḥah* (the orientation of outcomes to legitimate interests), were selected because each corresponds to a distinct epistemic or normative precondition of a valid *muʿāmalah* that algorithmic opacity directly threatens. The broader values of *ʿadl* (justice) and *riḍāʿ* (consent), invoked in the introduction, operate at a higher level of abstraction and are, in al-Sarakhsī's treatment, realised through these more specific operative requirements rather than standing alongside them.

The analytical procedure unfolds in four stages. In the first stage, jurisprudential principles relevant to contractual disclosure, harm prevention, certainty, and welfare are extracted from the primary source through close reading. In the second stage, these principles are reconstructed in a form suitable for application to algorithmic decision-making, with explicit attention to the analogical reasoning required to extend a classical doctrine to a novel object. In the third stage, the reconstructed principles are integrated with established XAI techniques, principally LIME (Ribeiro et al., 2016) and SHAP (Lundberg & Lee, 2017), to construct the Maqāsid-Explainability Governance Model. In the fourth stage, the model is applied to two representative cases in contemporary Islamic finance, namely *ṣukūk* pricing and *takāful* underwriting, to assess its operational coherence and its alignment with established Sharīʿah audit standards.

Three limitations are acknowledged at the outset. First, the case studies presented are illustrative rather than empirical; the model awaits validation through field deployment in an Islamic financial institution. Second, the doctrinal reading of al-Sarakhsī advanced here is interpretive and admits of alternative readings within the Hanafī school and across the four Sunni schools more broadly. Third, the analogical extension of classical principles to algorithmic decision-making, while methodologically defensible within the Hanafī tradition of *takhrīj*, remains a constructive rather than a direct application, and the limits of the analogy are discussed explicitly in the theoretical section that follows.

3. Al-Sarakhsī's Theory Of Mu'āmalāt And The Jurisprudential Foundations Of Algorithmic Governance

3.1 Al-Sarakhsī and Al-Mabsūṭ in Context

Shams al-Dīn Abū Bakr Muḥammad ibn Aḥmad al-Sarakhsī (d. circa 483/1090) is widely regarded as one of the most authoritative Hanafī jurists of the classical period. His magnum opus, *Al-Mabsūṭ*, was composed under extraordinary circumstances; according to the traditional account, much of the work was dictated to his students from a well in Uzjand where he had been confined for political reasons (Hallaq, 2009; Calder, 1993). The work runs to thirty volumes in modern editions and offers the most systematic treatment of *mu'āmalāt* in the Hanafī corpus, integrating the rulings of earlier authorities such as Muḥammad ibn al-Ḥasan al-Shaybānī with al-Sarakhsī's own jurisprudential synthesis. Where al-Sarakhsī's positions can be triangulated with other classical Hanafī authorities such as al-Kāsānī (1986) and al-Marghīnānī (2006), the convergence is noted in support of the broader Hanafī position.

For the purposes of this study, three features of *Al-Mabsūṭ* are particularly relevant. First, al-Sarakhsī treats *mu'āmalāt* as a coherent jurisprudential domain in which contractual validity depends on a configuration of epistemic and normative conditions rather than on a single criterion. Second, his treatment of disclosure (*bayān*), harm avoidance (*daf' al-darar*), certainty (*yaqīn*), and welfare (*maṣlaḥah*) is consistently linked to the institutional context within which contracting takes place, which makes his framework unusually well suited to governance questions in which institutions, not only individuals, are the relevant agents. Third, the Hanafī school's acceptance of *istiḥsān* as a recognised juristic method opens space for the kind of analogical reasoning by which classical principles can be extended to contemporary objects such as algorithmic decision-making (Nyazee, 2000).

3.2 Bayān: The Obligation of Clarity in Contractual Relations

Within al-Sarakhsī's framework, *bayān* refers to the obligation that contractual parties make material aspects of the transaction sufficiently clear to one another. A terminological clarification is in order. In *uṣūl al-fiqh*, *bayān* is a technical term for the clarification of legal indicants (al-Sarakhsī classifies its modes at length in his *Uṣūl*), and the present study uses *bayān* in the applied, contractual sense that the same jurist develops in the law of sale, where the duty of disclosure is at its most demanding. The principle is grounded in the Qur'anic injunction concerning the writing of debts:

"O you who have believed, when you contract a debt for a specified term, write it down... and let not the scribe refuse to write as Allah has taught him." (Qur'an 2:282)

This verse, often called *āyat al-mudāyanah* (the verse of contracting debt), is the longest in the Qur'an and is treated by al-Sarakhsī and other Hanafī authorities as the foundational textual warrant for documentary and disclosure obligations in *mu'āmalāt* (al-Sarakhsī, 1993, Vol. 12, Kitāb al-Buyū'). The principle of *bayān* extends beyond mere writing to encompass the substantive disclosure of the terms, conditions, and material features of the contract such that the parties' consent is genuinely informed. Where *bayān* is defective, the contract is exposed to the defect of *gharar* and may be rendered invalid or rescindable.

Al-Sarakhsī gives this obligation its sharpest expression in his treatment of the *murābahah* (cost-plus resale), where the seller is held to a heightened standard of candour regarding the basis on which the goods were acquired. He writes:

وَإِذَا اشْتَرَى شَيْئًا بِنَسِيئَةٍ فَلَيْسَ لَهُ أَنْ يَبِيعَهُ مُرَابَحَةً حَتَّى يَنْبَيِّنَ أَنَّهُ اشْتَرَاهُ بِنَسِيئَةٍ؛ لِأَنَّ بَيْعَ الْمُرَابَحَةِ بَيْعُ أَمَانَةٍ تَنْفِي عَنْهُ كُلُّ تَهْمَةٍ وَجَنَائَةٍ، وَيُتَحَرَّزُ فِيهِ مِنْ كُلِّ كَذِبٍ، وَفِي مَعَارِيضِ الْكَلَامِ شُبُهَةٌ فَلَا يَجُوزُ اسْتِعْمَالُهَا فِي بَيْعِ الْمُرَابَحَةِ.

“If a person has bought something on deferred terms (nasī’a), he may not sell it on a murābahah basis until he discloses that he bought it on deferred terms; for the murābahah sale is a sale of trust (bay‘ amāna), one from which every suspicion and wrongdoing is excluded, and in which every falsehood must be scrupulously guarded against. Even equivocal speech (ma‘ārīd al-kalām) carries an element of doubt, so it may not be resorted to in a murābahah sale.” (al-Sarakhsī, 1993, Vol. 13, p. 78, *Kitāb al-Buyū‘, Bāb al-Murābahah*)

Carried over to algorithmic decision-making, the principle of *bayān* asks more of the institution than the technical demand for explainability does. The institution deploying an automated decision system is, in effect, the disclosing party to every contract that the system mediates. If the institution cannot itself reconstruct why a particular decision was reached, then the disclosure required by *bayān* has not in fact occurred, and the validity of the resulting contract is open to question. It must be conceded that *murābahah* is a sale of trust (*bay‘ amāna*) carrying disclosure duties heavier than those of an ordinary bargained sale (*bay‘ al-musāwama*), in which the seller need not reveal cost. The claim advanced here is therefore not that every contract imposes *murābahah*-grade disclosure, but that an institution which deploys a decision system on behalf of, and with material consequences for, those whose interests it manages occupies a position analogous to the trustee in a *bay‘ amāna*, and is held by parity of reasoning to the heightened standard that such a position attracts.

3.3 Daf‘ al-Ḍarar: The Principle of Harm Prevention

The principle of *daf‘ al-ḍarar*, the avoidance and removal of harm, is one of the five foundational maxims (*al-qawā‘id al-kullīyyah al-khams*) of Islamic jurisprudence (al-Zuhaylī, 2002; Heinrichs, 2002). Its textual basis is the well-known prophetic tradition:

“Lā ḍarar wa lā ḍirār” (There shall be no harm, nor reciprocation of harm). (Ibn Mājah, *Sunan*, no. 2340; al-Dāraqutnī, *Sunan*, no. 4540)

Al-Sarakhsī treats this maxim as operative across the entire field of *mu‘āmalāt*, generating affirmative duties to prevent reasonably foreseeable harm to contractual parties and to third parties affected by the transaction (al-Sarakhsī, 1993). Crucially for the present argument, the duty of *daf‘ al-ḍarar* is not exhausted by avoiding intentional harm; it includes the obligation to avoid harm that arises from negligence in the design, oversight, or execution of the institution’s practices.

The principle takes concrete operational form in the *murābahah* chapter, where al-Sarakhsī makes the buyer’s need to repel harm the explicit operative cause for the option to rescind when material information has been withheld at the moment of consent:

وَهَذَا لِأَنَّ الْمُشْتَرِيَّ إِنَّمَا التَّزَمَ رَبْحًا بِنَاءً عَلَى خَبَرِهِ أَنَّهُ اشْتَرَاهُ لِنَفْسِهِ بِكَذَا مِنَ التَّمَنِ، فَلَوْ عَلِمَ أَنَّهُ اشْتَرَاهُ بِالنَّسِيئَةِ لَمْ يَرْغَبْ فِي شِرَائِهِ بِالنَّقْدِ بِذَلِكَ الْقَدْرِ مِنَ التَّمَنِ، فَضَلًّا مِنْ أَنْ يُعْطِيَهُ عَلَى ذَلِكَ رَبْحًا، فَلِحَاجَتِهِ إِلَى دَفْعِ الضَّرَرِ أَتَيْنَا لَهُ الْخِيَارَ.

“The buyer assumed the markup only on the strength of [the seller’s] report that he had bought the item for himself at such-and-such a price; had he known that the purchase was on deferred terms, he would not have wished to buy it for cash at that figure, let alone yield a markup on top of it. So, by reason of his need to repel harm (daf‘ al-ḍarar), we established for him the option [to rescind].” (al-Sarakhsī, 1993, Vol. 13, p. 78, *Kitāb al-Buyū‘, Bāb al-Murābahah*)

Translated into the algorithmic setting, *daf‘ al-ḍarar* forbids more than deliberate injury. Deploying a decision system whose outputs may systematically disadvantage protected categories of contracting parties, or whose error profile is unknown, is itself a harm-by-

negligence the institution is bound to prevent. The principle therefore requires not only explanation of individual decisions but ongoing monitoring of aggregate model behaviour for patterns that would constitute *ḍarar*.

3.4 *Yaqīn*: The Principle of Legal Certainty

The principle of *yaqīn* is articulated in the celebrated juristic maxim:

“*Al-yaqīn lā yuzāl bi-sh-shakk*” (*Certainty is not displaced by doubt*). (*al-Suyūṭī, al-Ashbāh wa al-Nazā’ir*; cf. *al-Sarakhsī, 1993, Vol. 1*)

Within al-Sarakhsī’s framework, *yaqīn* establishes a presumption of stability for legal states once they are constituted, and it requires that this presumption be displaced only by evidence rising to the level of certainty rather than mere doubt or conjecture (al-Suyūṭī, n.d.; Nyazee, 2000). Al-Sarakhsī applies this standard explicitly in the law of sale, refusing to allow a binding contractual claim to rest on any epistemic basis weaker than certainty:

وَأِنْ كَانَ مُخْتَلَفًا لَا يَبِيعُ الْبَاقِيَ مِنْهُ مَرَابَحَةً؛ لِأَنَّ انْقِسَامَ الثَّمَنِ عَلَى الْأَجْنَاسِ الْمُخْتَلِفَةِ بِاعْتِبَارِ الْقِيَمَةِ، وَطَرِيقُ مَعْرِفَتِهَا الْحَزْرُ وَالظَّنُّ، فَلَمْ يَكُنْ حِصَّةُ كُلِّ جَنْسٍ مِنَ الثَّمَنِ مَعْلُومًا يَقِينًا لِيَبِيعَهُ مَرَابَحَةً عَلَيْهِ.

“*If [the goods purchased together] are of different kinds, he does not sell the remainder on a murābahah basis; for the apportionment of price across different kinds proceeds by reference to value, and the path to knowing value is estimation and conjecture (al-ḥazr wa al-ẓann), so the share of each kind in the price is not known with certainty (ma’lūman yaqīnan) such that he could sell it on a murābahah basis.*” (al-Sarakhsī, 1993, Vol. 13, p. 81, *Kitāb al-Buyū’, Bāb al-Murābahah*)

The transfer to the algorithmic setting must be made with care, since Islamic commercial law does not in general require certainty: most *mu’āmalāt* proceed lawfully on preponderant probability (*ẓann ghālib*), and the judgement of a human underwriter or pricer has always been probabilistic. The maxim does not demand that a decision be certain; it demands that an established position not be displaced on a basis weaker than the doubt that purports to displace it. Read in this way, the standard that transfers to algorithmic decision-making is not statistical certainty but evidential traceability: where a model output adversely alters a contracting party’s established position, the institution must be able to give an interrogable account of the basis on which it did so, rather than appeal to an output it cannot itself reconstruct. So understood, *yaqīn* does not duplicate *bayān* but complements it: the one concerns the security of an established legal position, the other the adequacy of disclosure at the point of contracting.

3.5 *Maṣlaḥah*: The Principle of Public Welfare

The principle of *maṣlaḥah* refers to the protection and promotion of legitimate human interests, ordered in the classical literature according to the hierarchy of *ḍarūriyyāt* (necessities), *ḥājīyyāt* (needs), and *taḥsīniyyāt* (improvements) (al-Shāṭibī, 1997; Auda, 2008; Asutay, 2012). Several Qur’anic verses anchor the principle textually, among them:

“*Allah intends for you ease and does not intend for you hardship.*” (*Qur’an 2:185*)

“*We have not sent you except as a mercy to the worlds.*” (*Qur’an 21:107*)

Within Hanafī methodology, *maṣlaḥah* is closely associated with the doctrine of *istiḥsān*, which permits the juristic departure from strict analogical reasoning where such departure better realises the underlying objectives of the Sharī’ah (al-Sarakhsī, 1993; Kamali, 2008). Anchoring this methodology in the same Qur’anic principle of *yusr* cited above, al-Sarakhsī defines *istiḥsān* as a welfare-realising departure from strict analogy:

الِاسْتِحْسَانُ تَرْكُ الْقِيَاسِ وَالْأَخْذُ بِمَا هُوَ أَوْفَقُ لِلنَّاسِ ... وَخَاصِلُ هَذِهِ الْعِبَارَاتِ أَنَّهُ تَرْكُ الْعُسْرِ لِلْيُسْرِ وَهُوَ أَصْلٌ فِي الدِّينِ.

“Istiḥsān is the abandonment of strict analogy (qiyās) and the adoption of what is more conducive to people’s welfare ... the substance of these formulations is that it is the abandonment of hardship in favour of facility (tark al-‘usr li-l-yusr), and this is a foundational principle in religion.” (al-Sarakhsī, 1993, Vol. 10, p. 145, *Kitāb al-Istiḥsān*)

Finally, *maṣlaḥah* adds a further test: an AI system in Islamic finance must answer not only for its technical accuracy but for how well it advances legitimate interests at the appropriate level of the *maqāṣid* hierarchy.

3.6 From Classical Principles to Algorithmic Governance

The application of these four classical principles to algorithmic decision-making is, by its nature, analogical rather than direct. Al-Sarakhsī did not write about machine learning models, and any argument that purports to derive specific algorithmic governance rules from *Al-Mabsūṭ* must be defended on methodological grounds. The defence advanced here rests on the Hanafi tradition of *takhrīj al-furū‘ ‘alā al-uṣūl*, the disciplined extension of classical principles to novel cases through reasoned analogy (Nyazee, 2000). The principles invoked here are themselves abstract: *bayān* is the obligation to make material aspects of a transaction sufficiently clear; *daḥḥ al-ḍarar* is the obligation to avoid foreseeable harm; *yaqīn* is the obligation to ground adverse decisions in evidence rising to the level of certainty; and *maṣlaḥah* is the obligation to advance legitimate interests in the appropriate order of priority. These obligations do not depend on the technology of the decision instrument; they apply to any institutional means of arriving at a decision that affects a contracting party. The argument that follows therefore treats the AI system as a contemporary instance of a familiar jurisprudential object, namely an institutional decision-making practice with consequences for the validity of resulting contracts.

The analogical extension defended here on classical grounds is reinforced by the contemporary systems approach to *maqāṣid* advanced by Auda (2008), which treats the Sharī‘ah as an open, multi-dimensional, and purposeful system rather than a closed catalogue of fixed-object rulings. On Auda’s account, the four governance principles reconstructed above are expected to operate concurrently and to engage new social and technological environments, including those that classical jurists did not anticipate, and the reorientation of *maqāṣid* theory from a protective to a developmental register further supports the constructive deployment of these principles in the design of algorithmic governance instruments rather than their defensive invocation alone (Auda, 2008, esp. chs. 5–6). The MEGM’s layered architecture is therefore not only consistent with the Hanafi method of *takhrīj* but with a contemporary *maqāṣidī* methodology in which several principles constrain one another rather than displacing one another in sequence.

4. The Maqāṣid-Explainability Governance Model (MEGM)

4.1 Architecture of the Model

The Maqāṣid-Explainability Governance Model proposed in this study is constructed as a layered governance instrument that sits above existing Explainable AI techniques rather than replacing them. The base of the model consists of the underlying machine learning system whose decisions are subject to governance, typically a supervised classifier or regressor used for credit scoring, *ṣukūk* pricing, *takāful* underwriting, or related applications. Above this technical base sits an Explainability Layer, which generates post-hoc explanations of individual

decisions through established techniques such as LIME and SHAP (Ribeiro et al., 2016; Lundberg & Lee, 2017). The Explainability Layer produces feature attributions but does not, by itself, generate jurisprudential adequacy.

This reliance on post-hoc explanation invites an objection. Rudin (2019) argues that high-stakes decisions should employ models that are interpretable by design rather than opaque models wrapped in post-hoc approximations, since attributions of the kind produced by LIME and SHAP may be unstable or unfaithful to the model they purport to explain. The objection is well taken, and the MEGM accommodates it in two ways. Where an interpretable-by-design model is feasible, the framework prefers it and applies the three interpretive layers directly to the model's own logic rather than to a surrogate. Where institutional constraints make an opaque model unavoidable, the fidelity of the explanation becomes itself an object of the Sharī'ah audit: an attribution whose stability cannot be demonstrated is, for the purposes of *bayān*, a defective disclosure, and the *yaqīn* layer requires that such fragility be recorded rather than concealed. The post-hoc explanation is therefore not assumed to be reliable; its reliability is part of what the framework is designed to interrogate.

Above the Explainability Layer, the MEGM introduces three interpretive layers that translate technical explanations into jurisprudentially auditable form. The first, Probabilistic *Bayān*, renders feature attributions in the form of disclosure statements that meet the Hanafi standard for material clarity in contracting. The second, Justificatory *Ta'līl*, identifies the *'illah* (operative cause) of the decision and assesses whether the cause is jurisprudentially admissible. The third, *Maṣlaḥah*-Based Evaluation, situates the decision within the *maqāṣid* hierarchy and evaluates its contribution to or against legitimate interests. Above the three interpretive layers sits the Sharī'ah Audit Layer, which provides documentary outputs suitable for review by a Sharī'ah supervisory board.

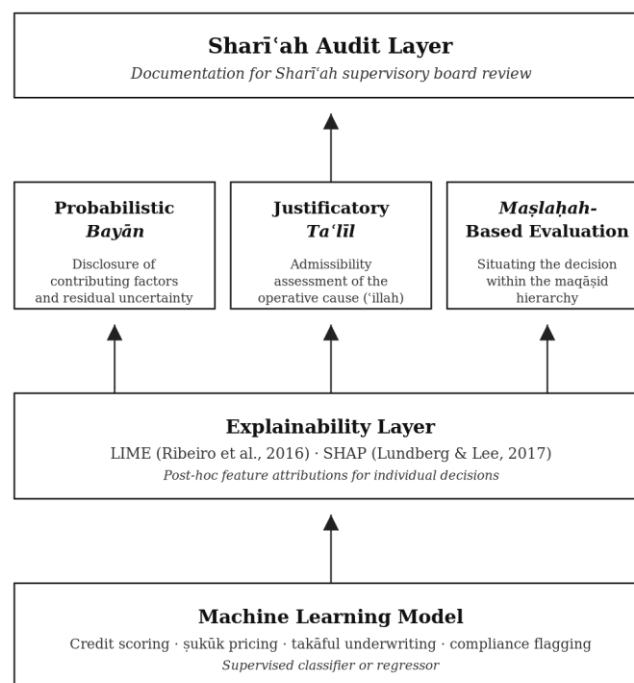


Figure 1. The Maqāṣid-Explainability Governance Model (MEGM)

4.2 Probabilistic Bayān

The Probabilistic *Bayān* layer takes as input the feature attributions produced by the Explainability Layer and produces as output a structured disclosure statement that satisfies the Hanafī standard of material clarity. The disclosure statement identifies, in non-technical language accessible to the contracting party, the factors that most influenced the decision, the direction and approximate magnitude of their influence, and the residual uncertainty associated with the model’s output. The criterion for adequate *bayān* is not perfect transparency, which is unattainable in any consequential decision, but sufficient transparency that the contracting party can form a reasoned judgement about whether to consent to the transaction on the disclosed terms.

4.3 Justificatory Taʿlīl

The Justificatory *Taʿlīl* layer addresses a question that technical XAI does not: are the operative reasons for the decision jurisprudentially admissible? Drawing on the Hanafī tradition of cause-identification in legal reasoning, this layer maps the feature attributions onto candidate *ʿilal* (operative causes) and assesses each for admissibility. Causes that are admissible include factors directly relevant to risk, contractual capacity, and material circumstance. Causes that are inadmissible include factors that operate as proxies for protected characteristics, factors that are unrelated to the legitimate basis of the contract, and factors whose use would conflict with the principle of *dafʿ al-ḍarar*. The output of this layer is a justificatory statement that either ratifies the decision or flags it for human review.

4.4 Maṣlaḥah-Based Evaluation

The *Maṣlaḥah*-Based Evaluation layer assesses the decision against the *maqāṣid* hierarchy. A decision that supports the protection of *ḍarūriyyāt* (necessities), such as decisions concerning livelihood or shelter, is evaluated against a more demanding standard than a decision that affects *taḥsīniyyāt* (improvements). The output of this layer is an evaluative score that situates the decision within the *maqāṣid* framework and identifies any tension between the decision’s technical optimum and its jurisprudential propriety.

4.5 Governance Protocol

Operationally, the MEGM is invoked at three points in the lifecycle of an automated decision. At the point of model deployment, the institution conducts a baseline audit using each of the three interpretive layers, producing a model governance dossier that is reviewed and signed off by the Sharīʿah supervisory board. At the point of individual decision-making, the model generates, alongside its primary output, the layered MEGM explanation, which is retained as part of the contract record. At the point of periodic review, the institution conducts an aggregate audit of model behaviour across decisions, with particular attention to patterns that would constitute *ḍarar* at the aggregate level. The protocol thus integrates ex ante design review, ex tempore decision audit, and ex post outcome monitoring within a single governance instrument, paralleling broader frameworks of algorithmic accountability that structure review across ex ante, in-process, and ex post moments (Diakopoulos, 2016).

5. Illustrative Application Of The MEGM To Şukūk Pricing And Takāful Underwriting

5.1 Case Application One: Şukūk Pricing

Consider an Islamic investment bank that uses a gradient boosting model to assist in the pricing of a new *şukūk* issuance. The model takes as inputs the issuer's financial characteristics, prevailing market conditions, the structure of the underlying asset, and historical pricing data from comparable issuances. The model produces as output a recommended yield range. A SHAP analysis of a representative pricing decision identifies the five features that most influenced the recommendation: issuer credit rating, asset class, sector volatility, issuance tenor, and the credit spread of comparable conventional issuances. The feature attributions, taken on their own, raise no immediate concern.

Applied to this decision, the MEGM produces three distinct outputs. The Probabilistic *Bayān* layer translates the SHAP attributions into a disclosure statement that identifies the five contributing factors in plain language, with their directional influence and approximate magnitude. The Justificatory *Ta'īl* layer examines each contributing factor for jurisprudential admissibility. Four of the five features are admissible without qualification; the fifth, the credit spread of comparable conventional issuances, is flagged as requiring scrutiny on the ground that systematic anchoring of *şukūk* pricing to conventional bond markets risks importing the interest-rate logic of *ribā*-based instruments into a Sharī'ah-compliant pricing process. The *Maşlahah*-Based Evaluation layer notes that the decision affects the public capital market and therefore engages *maşāliḥ 'āmmah* at the level of *ḥājiyyāt*. The composite MEGM output recommends that the model's recommendation be accepted in part but that the conventional spread feature be reweighted or replaced in subsequent model iterations.

Table 1. SHAP feature attributions for a representative *şukūk* pricing decision and the corresponding MEGM assessment at each interpretive layer

Feature	SHAP attribution	Probabilistic Bayān (Layer 1)	Justificatory Ta'īl (Layer 2)	Maşlahah-Based Evaluation (Layer 3)
Issuer credit rating	+0.34	Disclosed as the dominant upward driver of the recommended yield	Admissible; a recognised direct measure of issuer risk	Serves investor protection (<i>ḥājiyyāt</i>)
Asset class	+0.21	Disclosed as a structural determinant of pricing	Admissible; reflects the contractual basis of the issuance	Supports orderly capital-market function (<i>ḥājiyyāt</i>)
Sector volatility	+0.18	Disclosed as a sectoral risk premium	Admissible; a legitimate risk-pricing factor	Supports prudent risk allocation (<i>ḥājiyyāt</i>)
Issuance tenor	+0.12	Disclosed as a maturity adjustment	Admissible; reflects time to maturity, not interest	Ordinary term structure; neutral to <i>maşlahah</i>
Credit spread of comparable conventional issuances	+0.27	Disclosed as reliance on a conventional-market benchmark	Flagged; risks importing <i>ribā</i> -based interest-rate logic into a Sharī'ah-compliant process	Potential conflict at <i>darūriyyāt</i> (integrity of Islamic finance)

Source: Authors' synthesised illustration. SHAP attributions are illustrative and directional, presented for expository purposes.

5.2 Case Application Two: Takāful Underwriting

Consider a family *takāful* operator that uses a machine learning model to classify applicants into risk categories for the purposes of contribution calculation. The model takes as inputs demographic data, health declarations, lifestyle factors, and historical claims data. The SHAP attribution for a representative classification identifies age, declared chronic conditions, occupational risk, and postal code as the four most influential features. The first three are jurisprudentially uncontroversial. The fourth, postal code, is flagged by the Justificatory *Taʿlīl* layer as a candidate proxy variable.

The MEGM analysis proceeds further. Postal code may correlate with relevant underwriting factors such as access to healthcare and environmental quality, but it may also correlate with ethnicity, socioeconomic status, and religious affiliation, none of which are admissible bases for differential treatment (cf. Barocas & Selbst, 2016) in a *takāful* arrangement grounded in *taʿāwun* (mutual assistance). The *Maṣlaḥah*-Based Evaluation layer further notes that *takāful* is constituted on a cooperative rather than a commercial logic, and that pricing decisions that systematically disadvantage protected groups undermine the doctrinal foundation of the instrument itself. The composite MEGM output recommends that postal code be removed from the model and replaced by direct measurements of the underlying factors it was serving as a proxy for. The case illustrates how the model can detect a jurisprudential concern that the underlying technical XAI alone would not have surfaced.

Table 2. SHAP feature attributions for a representative *takāful* underwriting decision and the corresponding MEGM assessment at each interpretive layer

Feature	SHAP attribution	Probabilistic Bayān (Layer 1)	Justificatory Taʿlīl (Layer 2)	Maṣlaḥah-Based Evaluation (Layer 3)
Applicant age	+0.33	Disclosed as the leading driver of the risk classification	Admissible; a recognised actuarial basis	Supports fair contribution pricing (ḥājiyyāt)
Declared chronic conditions	+0.29	Disclosed as a direct health-risk factor	Admissible; materially relevant to the risk pooled	Supports the integrity of the risk pool (ḥājiyyāt)
Occupational risk	+0.16	Disclosed as an occupation-class loading	Admissible; a recognised hazard-class factor	Supports equitable contribution (ḥājiyyāt)
Postal code	+0.14	Disclosed but ambiguous; a geographic proxy variable	Flagged; candidate proxy for ethnicity, socioeconomic status, or religion, none admissible under <i>taʿāwun</i>	Conflict at ḍarūriyyāt; undermines the cooperative basis of <i>takāful</i>

Source: Authors' synthesised illustration. SHAP attributions are illustrative and directional, presented for expository purposes.

5.3 Comparative Observations

Read together, the two cases demonstrate that the MEGM generalises across structurally different applications without modification to its layered architecture. The *ṣukūk* case is a pricing (regression) task and the *takāful* case a classification task; yet in both the same pipeline applies, with the Probabilistic Bayān layer rendering the SHAP attributions as plain-language disclosure, the Justificatory *Taʿlīl* layer testing each feature for the admissibility of its operative

cause, and the Maṣlaḥah-Based Evaluation layer situating the decision within the maqāṣid hierarchy. The two cases nonetheless surface different categories of jurisprudential concern. In the ṣukūk case the flagged feature implicates the substantive validity of the pricing process, since anchoring to a conventional benchmark risks importing ribā-based logic; in the takāful case the flagged feature implicates distributive fairness and the cooperative (taʿāwun) foundation of the instrument through proxy discrimination. In both, the concern is generated by the Taʿlīl layer and escalated by the Maṣlaḥah layer, and in both it would have remained latent in the raw SHAP output, which on its own flags nothing. The model thus operates consistently across both ḥājiyyāt-level and ḍarūriyyāt-level concerns, which supports the claim that it provides a general rather than application-specific auditing structure for automated decision-making in Islamic finance.

6. Discussion

6.1 Alignment with AAOIFI and IFSB Standards

The MEGM produces governance outputs that align with several of the requirements articulated in the AAOIFI Governance Standards and the IFSB Sharīʿah Governance framework. The contractual instruments to which the model is applied, including ṣukūk and takāful, are themselves governed by the AAOIFI Sharīʿah Standards (AAOIFI, 2017), which the model's interpretive layers are designed to operationalise at the point of decision rather than merely at the point of product approval. AAOIFI Governance Standard No. 1 requires that Sharīʿah supervisory boards have access to documentation sufficient to discharge their oversight function (AAOIFI, 2023). The Sharīʿah Audit Layer of the MEGM produces documentation of just this kind, organised around the four jurisprudential principles that govern *muʿāmalāt*; the approach is consistent with broader work on ethics-based AI auditing (Mökander & Floridi, 2021). IFSB-10 and the more recent IFSB-25 on Sharīʿah governance disclosure require ongoing monitoring of Sharīʿah compliance across the institution's activities (IFSB, 2023). The aggregate audit component of the MEGM governance protocol responds directly to this requirement.

6.2 Implications for Sharīʿah Governance Practice

Adoption of the MEGM would carry several practical implications for Sharīʿah governance practice. Sharīʿah supervisory boards would receive structured, comparable documentation across all algorithmic decisions, allowing them to identify patterns that individual decisions would not reveal. Internal Sharīʿah auditors would gain a clearer mandate to engage with the technical operations of the institution, supported by a framework that translates between jurisprudential reasoning and technical output. Regulators would have a basis for examining the governance of algorithmic decision-making in Islamic financial institutions in a manner that is consistent with the existing Sharīʿah supervisory architecture rather than parallel to it.

6.3 Theoretical Implications

Beyond its operational implications, the study advances two theoretical contributions. First, it demonstrates that classical Hanafī *muʿāmalāt* theory, properly reconstructed, can serve as an operational foundation for contemporary governance instruments rather than as a purely historical reference. Second, it repositions al-Sarakhsī within contemporary Islamic finance

scholarship as a working theoretical resource for the regulation of automated decision-making, complementing the more familiar engagement with al-Shāṭibī and Ibn ‘Āshūr in the *maqāṣid* literature. Both contributions extend the analytical reach of classical *fiqh* into a domain in which it has not previously been deployed.

7. Conclusion

This study has argued that explainability in Islamic financial AI is owed as a fiduciary duty rather than offered as a technical convenience, and that the duty is already latent in classical *mu‘āmalāt* theory. Drawing on Imām al-Sarakhsī’s *Al-Mabsūt*, four jurisprudential principles, namely *bayān*, *daf‘ al-ḍarar*, *yaqīn*, and *maṣlaḥah*, have been reconstructed and operationalised into the Maqāṣid-Explainability Governance Model. Application of the model to *ṣukūk* pricing and *takāful* underwriting suggests that layered interpretability is capable of satisfying auditability requirements that conventional XAI, deployed in isolation, may not fully meet. The model repositions al-Sarakhsī’s *mu‘āmalāt* theory as an active jurisprudential foundation for ethical AI governance rather than as a historical reference.

Three limitations should be acknowledged. The case studies presented are illustrative rather than empirical, and the model awaits validation through field deployment in an Islamic financial institution. The doctrinal reading of al-Sarakhsī advanced here is interpretive and admits of alternative readings. The analogical extension of classical principles to algorithmic decision-making, though methodologically defensible within the Hanafī tradition of *takhrīj*, remains a constructive rather than a direct application.

These limitations in turn mark out where the work should go next. First, empirical validation of the MEGM through pilot deployment in collaboration with Islamic financial institutions and their Sharī‘ah supervisory boards would strengthen the model’s practical foundation. Second, extension of the framework to additional AI use cases in Islamic finance, including *zakat* calculation, *waqf* management, and Islamic microfinance, would test its generalisability. Third, comparative analysis with the maqāṣidī frameworks of al-Shāṭibī, Ibn ‘Āshūr, and Ibn Taymiyyah would situate al-Sarakhsī’s contribution within a broader classical conversation and may produce complementary governance principles.

8. Acknowledgements

The authors thank the Academy of Islamic Studies, Universiti Malaya, and the Faculty of Economics and Muamalat, Universiti Sains Islam Malaysia, for institutional support.

References

- AAOIFI. (2017). *Shari‘ah standards*. Accounting and Auditing Organization for Islamic Financial Institutions.
- AAOIFI. (2023). *Governance Standards for Islamic Financial Institutions*. Accounting and Auditing Organization for Islamic Financial Institutions.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>

- Asutay, M. (2012). Conceptualising and locating the social failure of Islamic finance: Aspirations of Islamic moral economy vs. the realities of Islamic finance. *Asian and African Area Studies*, 11(2), 93–113.
- Auda, J. (2008). *Maqasid al-Shariah as philosophy of Islamic law: A systems approach*. International Institute of Islamic Thought.
- Awosika, T., Shukla, R. M., & Pranggono, B. (2024). Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection. *IEEE Access*, 12, 64551–64560.
- Babaei, G., Giudici, P., & Raffinetti, E. (2023). Explainable fintech lending. *Journal of Economics and Business*, 125, 106126.
- Bank Negara Malaysia. (2019). *Shariah governance policy document*. Bank Negara Malaysia.
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
- Bin-Nashwan, S. A., Al-Daihani, M., Abdul-Jabbar, H., & Al-Ttaffi, L. H. A. (2022). Social solidarity amid the COVID-19 outbreak: Fundraising campaigns and donors' attitudes. *International Journal of Sociology and Social Policy*, 42(3/4), 232–247. <https://doi.org/10.1108/IJSSP-05-2020-0173>
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
- Calder, N. (1993). *Studies in early Muslim jurisprudence*. Clarendon Press.
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. <https://doi.org/10.1145/2844110>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People: An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), Article 93. <https://doi.org/10.1145/3236009>
- Hallaq, W. B. (2009). *Sharī'a: Theory, practice, transformations*. Cambridge University Press.
- Heinrichs, W. P. (2002). Qawā'id as a genre of legal literature. In B. G. Weiss (Ed.), *Studies in Islamic legal theory* (pp. 365–384). Brill.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hutchinson, T., & Duncan, N. (2012). Defining and describing what we do: Doctrinal legal research. *Deakin Law Review*, 17(1), 83–119. <https://doi.org/10.21153/dlr2012vol17no1art70>
- Ibn 'Āshūr, M. T. (2006). *Treatise on maqāṣid al-sharī'ah* (M. E. El-Mesawi, Trans.). International Institute of Islamic Thought.
- ICD–LSEG. (2025). *Islamic finance development report 2025: 50 years of exponential growth*. Islamic Corporation for the Development of the Private Sector & London Stock Exchange Group.
- IFSB. (2020). *Fintech and Islamic finance: Current landscape and future outlook*. Islamic Financial Services Board.

- IFSB. (2023). *IFSB-25: Disclosures to promote transparency and market discipline for institutions offering Islamic financial services*. Islamic Financial Services Board.
- IFSB. (2024). *Islamic Financial Services Industry Stability Report 2024*. Islamic Financial Services Board.
- IFSB. (2025). *Islamic financial services industry stability report 2025*. Islamic Financial Services Board.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kamali, M. H. (2000). *Islamic commercial law: An analysis of futures and options*. Islamic Texts Society.
- Kamali, M. H. (2007). *Principles of Islamic jurisprudence* (3rd ed.). Islamic Texts Society.
- Kamali, M. H. (2008). *Maqasid al-shariah made simple*. International Institute of Islamic Thought.
- Kamaruddin, M. I. H., Shafii, Z., Hanefah, M. M., Salleh, S., & Zakaria, N. (2024). Issues and challenges in Shariah audit practices in Malaysian Shariah-based sectors. *The Journal of Muamalat and Islamic Finance Research*, 21(1), 82–100. <https://doi.org/10.33102/jmifr.564>
- al-Kāsānī, ‘A. al-D. (1986). *Badā’i ‘al-ṣanā’i fī tartīb al-sharā’i* (2nd ed., 7 vols.). Dār al-Kutub al-‘Ilmiyyah.
- Laldin, M. A., & Furqani, H. (2013). Developing Islamic finance in the framework of maqāṣid al-sharī‘ah: Understanding the ends (maqāṣid) and the means (wasā’il). *International Journal of Islamic and Middle Eastern Finance and Management*, 6(4), 278–289. <https://doi.org/10.1108/IMEFM-05-2013-0057>
- Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36–43. <https://doi.org/10.1145/3233231>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.
- al-Marghīnānī, B. al-D. (2006). *Al-Hidāyah sharḥ Bidāyat al-mubtadī* (M. A’zamī, Ed.). Dār al-Salām.
- Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. <https://doi.org/10.1177/2053951716679679>
- Mohamed, H., & Ali, H. (2018). *Blockchain, fintech, and Islamic finance: Building the future in the new Islamic digital economy*. De Gruyter.
- Mohd Haridan, N., Sheikh Hassan, A. F., Mohammed Shah, S., & Mustafa, H. (2023). Financial innovation in Islamic banks: Evidence on the interaction between Shariah board and FinTech. *Journal of Islamic Accounting and Business Research*, 14(6), 911–930. <https://doi.org/10.1108/JIABR-11-2022-0305>
- Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds and Machines*, 31(2), 323–327. <https://doi.org/10.1007/s11023-021-09557-8>
- Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Independently published.
- Nyazee, I. A. K. (2000). *Islamic jurisprudence (Uṣūl al-Fiqh)*. International Institute of Islamic Thought.

- Oseni, U. A., & Ali, S. N. (Eds.). (2019). *Fintech in Islamic finance: Theory and practice*. Routledge.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). <https://doi.org/10.1145/2939672.2939778>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Information Fusion*, 99, 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- al-Sarakhsī, S. (1993). *Al-Mabsūṭ* (30 vols.). Dār al-Maʿrifah.
- Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, 87(3), 1085–1139.
- Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, 7(4), 233–242. <https://doi.org/10.1093/idpl/ix022>
- Shalhoob, H. (2025). The role of AI in enhancing Shariah compliance: Efficiency and transparency in Islamic finance. *Journal of Infrastructure, Policy and Development*, 9(1), Article 11239. <https://doi.org/10.24294/jipd11239>
- al-Shāṭibī, I. (1997). *Al-Muwāfaqāt fī uṣūl al-sharīʿah* (M. Āl Salmān, Ed.; 4 vols.). Dār Ibn ʿAffān.
- Al-Suwailm, S. (2006). *Hedging in Islamic finance* (Occasional Paper No. 10). Islamic Research and Training Institute, Islamic Development Bank.
- al-Suyūṭī, J. al-D. (n.d.). *Al-Ashbāh wa al-Nazāʾir*. Dār al-Kutub al-ʿIlmiyyah.
- Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.
- al-Zuhaylī, W. (2002). *Al-Fiqh al-Islāmī wa adillatuh* (4th ed.). Dār al-Fikr.

Appendix A: Sharīʿah-Audit Implementation Checklist

The following checklist operationalises the Maqāṣid-Explainability Governance Model for use by Sharīʿah supervisory boards, internal Sharīʿah auditors, and regulators reviewing an Islamic financial institution’s deployment of an automated decision system. The checklist is organised by the three interpretive layers of the MEGM and may be applied at the point of model deployment, at the point of individual decision-making, and at periodic aggregate review.

Layer 1: Probabilistic Bayān

1. Does the institution generate, for each automated decision affecting a contracting party, a disclosure statement identifying the principal contributing factors?
2. Are these factors expressed in language reasonably accessible to the contracting party?

3. Is the directional influence of each factor disclosed, with an indication of approximate magnitude?
4. Is the residual uncertainty of the model's output disclosed?
5. Is the disclosure statement retained as part of the contract record for the duration of regulatory retention requirements?

Layer 2: Justificatory Ta'īl

6. For each significant contributing factor, has the operative cause (*'illah*) been identified?
7. Has each identified cause been assessed for jurisprudential admissibility?
8. Have proxy variables that may stand in for protected characteristics been identified and either removed or replaced with direct measurements?
9. Where a contributing factor has been flagged as questionable, has the decision been routed to human Sharī'ah review prior to finalisation?

Layer 3: Maṣlaḥah-Based Evaluation

10. Has the decision been situated within the *maqāṣid* hierarchy of *ḍarūriyyāt*, *ḥājīyyāt*, and *taḥsīniyyāt*?
11. Where the decision affects *ḍarūriyyāt*, has the more demanding standard of review been applied?
12. Has the aggregate behaviour of the model been monitored for patterns that would constitute *ḍarar* at the institutional or societal level?
13. Where such patterns have been detected, has the model been modified or its deployment limited?

Documentation Requirements

14. Has the institution maintained a model governance dossier that is reviewed and signed off by the Sharī'ah supervisory board at deployment and at each material model revision?
15. Are the outputs of the three interpretive layers retained in a form auditable by AAOIFI Governance Standards and IFSB Sharī'ah governance requirements?