Islamic Digital Religion and Intelligent Media

Vol. 1, No. 1 (October 2026), 21-40
P-ISSN: XXXX-XXXX | E-ISSN: XXXX-XXXX
DOI: https://doi.org/10.xxxxx/drim.vXiX.XX

Provenance Before Authority: An Islamic Normative Framework for Transparent, Verifiable, and Accountable Generative AI-Mediated Da'wah

Yasirul Amri[1*], Ansori Hidayat[1], Irhamuddin[2], Muhammad Fadhil Hadziq[3]

[1] Universitas Islam Yasni, Bungo, Indonesia
[2] Al-Azhar University, Cairo, Egypt
[3] International University of Africa, Sudan.

[*]Correspondence: amriyasirul@gmail.com

Received: 19-09-2026	Revised: 07-10-2026	Accepted: 30-10-2026

Abstract. Generative artificial intelligence increasingly mediates how Muslims seek religious information, compare interpretations, and obtain practical guidance. Yet, fluent answers can appear authoritative without exposing their evidentiary lineage or accountable human source. This article develops a normative framework for evaluating generative AI-mediated da'wah through Islamic principles of tabayyun, amanah, epistemic restraint, scholarly referral, and authenticated transmission. The study uses normative-conceptual analysis that integrates Qur'anic norms, contemporary scholarship on Islamic epistemology and digital religion, recent empirical research on religious large language models, and international frameworks for trustworthy artificial intelligence. The analysis identifies a structural mismatch between probabilistic generation and the provenance-sensitive character of Islamic religious knowledge. The analysis synthesizes five requirements into the Islamic Religious AI Provenance Framework (IRAPF): source traceability, epistemic transparency, tabayyun affordance, authority boundary, and human accountability. The framework further distinguishes low-risk retrieval from medium-risk interpretation and high-risk personalized religious guidance, with progressively stronger verification and oversight requirements. The article concludes that generative AI may legitimately assist religious communication only when provenance precedes persuasive authority and when institutional arrangements preserve human responsibility for consequential religious guidance. This framework contributes a specifically Islamic normative vocabulary to digital religion scholarship while remaining interoperable with contemporary artificial intelligence governance principles.

Keywords: Generative AI; digital da'wah; religious authority; tabayyun; source provenance

Introduction

Generative artificial intelligence has moved from an experimental technology into an everyday interface for information seeking, writing, education, counseling, and decision support. Religious information is increasingly included within this transition because conversational systems can respond immediately to questions about worship, ethics, family life, spirituality, and Islamic law. Recent research already shows that Muslim users sometimes treat general-purpose systems as religious assistants, perceived scholars, or faith counselors, even when they remain concerned about misguidance and excessive reliance (Jahan et al., 2026). The change therefore concerns more than technological convenience, because conversational interfaces

Copyright © 2026 by Author(s)
This work is licensed under a **Creative Commons Attribution-NonCommercial 4.0 International License**.

reorganize how religious knowledge becomes accessible and credible. Digital religion scholarship has long demonstrated that media technologies do not simply transmit religion but participate in reorganizing practice, authority, identity, community, and interpretation (Campbell & Tsuria, 2021). Within Muslim contexts, social media influencers have likewise reconfigured traditional patterns of religious legitimacy through accessibility, affective communication, and platform visibility (Zaid et al., 2022). Generative AI intensifies this process because the interface can synthesize individualized religious language at conversational speed while obscuring the heterogeneous sources from which it statistically generates its output.

The central problem is not merely that large language models sometimes produce inaccurate religious answers. Contemporary technical research shows that generative models can produce factuality and faithfulness failures that remain linguistically plausible, while retrieval augmentation and other mitigation techniques reduce but do not eliminate these risks (Huang et al., 2025; Ji et al., 2023). In religious domains, the normative implications are sharper because accuracy alone does not exhaust the requirements of legitimate religious communication. Islamic traditions of knowledge have historically placed considerable weight upon identifiable teachers, authenticated transmission, source criticism, methodological qualification, and responsible interpretation. Brown (2017) shows that hadith scholarship developed elaborate procedures for evaluating both transmitted reports and the credibility of their transmission chains, while broader Islamic legal reasoning similarly distinguishes textual evidence, interpretive competence, juristic disagreement, and contextual application. A fluent answer may therefore be propositionally correct while remaining epistemically deficient when it conceals its source, collapses disagreement, exaggerates certainty, or implies authority beyond the competence of its producer.

Recent studies demonstrate that this distinction between persuasive credibility and justified authority is becoming empirically significant. Alam et al. (2025) found that Muslim participants strongly preferred scholarly guidance in principle yet favored AI-generated answers in blind comparisons in 81.3 percent of cases, while expert reviewers identified deficiencies in depth, accuracy, and contextual sensitivity. FiqhQA similarly found substantial variation across models, legal schools, languages, and abstention behavior when it tested large language models on Islamic jurisprudence questions (Atif et al., 2025). Other evaluations report methodological inconsistency, weak attribution, misquotation, and inadequate jurisprudential reasoning in AI-based Islamic guidance systems (Susanti et al., 2025; Ünal, 2024). These findings create a serious governance problem because the most rhetorically persuasive answer can become the most trusted answer even when its provenance and interpretive basis remain unclear. The emerging evidence therefore suggests a need to separate user-perceived authority from epistemically warranted authority when assessing generative AI in religious communication.

Existing Islamic scholarship on artificial intelligence has provided important foundations but remains fragmented across several analytical directions. Ethical studies commonly employ maqasid al-shariah, justice, accountability, virtue, human dignity, or moderation to assess artificial intelligence as a broad technological phenomenon (Elmahjub, 2023; Kannike & Fahm, 2025; Kosasih et al., 2024; Raquib et al., 2022; Ahmad et al., 2025). Jurisprudential studies increasingly ask whether artificial intelligence can perform ijtihad, issue fatwas, or participate in legal reasoning, and they generally retain final authority for qualified human scholars (Belhassen et al., 2025). Digital religion research, meanwhile, examines the migration of religious authority toward influencers, platforms, metrics, and algorithmic visibility (Hidayatullah, 2025; Maulana & Anoraga, 2025; Rachman et al., 2025; Siregar et al., 2025). These studies converge around concerns about trust and authority, yet they often address ethics, jurisprudence, platformization, and technical reliability as separate problems. A more

integrated model is needed to explain what conditions must be in place before generative output can responsibly participate in Islamic religious communication.

The research gap becomes especially visible around provenance. International artificial intelligence governance increasingly emphasizes transparency, traceability, accountability, human oversight, testing, and risk management, particularly for generative systems whose outputs may be difficult to verify (Batool et al., 2025; National Institute of Standards and Technology [NIST], 2024; OECD, 2024a; Haditia et al., 2026). Technical research also shows that explicit attribution to multiple sources can improve verifiability and reliability in long-form answer generation (Patel et al., 2024). However, provenance is rarely developed as an explicitly religious concept that connects source identity, interpretive tradition, uncertainty, authorization, and downstream responsibility. Islamic discussions often invoke tabayyun, sanad, or maqasid, but these concepts are often presented as general ethical reminders rather than operational criteria for evaluating an AI-mediated religious message. The unresolved problem is therefore how to translate Islamic epistemic norms into a governance architecture that can distinguish legitimate assistance from simulated authority.

This article addresses that problem through three research questions that remain suitable for normative analysis. First, what epistemic problems arise when generative AI communicates Islamic religious knowledge without transparent source provenance? Second, how can Islamic principles of tabayyun, amanah, authenticated transmission, scholarly referral, and epistemic restraint be translated into normative requirements for AI-mediated da'wah? Third, what governance framework can distinguish informational assistance from religious authority across different risk levels? The unit of analysis is not an individual user or a specific AI product, because the study instead analyzes normative claims about legitimate religious communication and the governance conditions attached to those claims. The article therefore treats generative AI as a mediated communication system whose legitimacy depends upon both informational performance and institutional responsibility.

The study proposes the Islamic Religious AI Provenance Framework, abbreviated as IRAPF, as its principal theoretical contribution. Several of these five requirements run structurally parallel to pillars already established in secular AI-governance instruments (e.g., transparency and explainability in the NIST AI RMF, human oversight and accountability in the EU AI Act and OECD principles, cited below); the framework's distinctiveness lies less in introducing new governance categories than in grounding each requirement in a specific Islamic epistemic warrant (tabayyun, sanad, amanah) that determines how the requirement is applied to religious content specifically, rather than in the requirement categories themselves being unprecedented. IRAPF is organized around five interdependent requirements: source traceability, epistemic transparency, tabayyun affordance, authority boundary, and human accountability. These requirements are not presented as replacements for established Islamic jurisprudence or scholarly institutions. They instead function as a normative interface between Islamic epistemic principles and contemporary systems of automated religious communication. The framework is designed to be technologically neutral enough to apply across chatbots, search assistants, retrieval-augmented systems, institutional da'wah applications, and future multimodal religious agents. It also remains compatible with broader governance approaches that emphasize transparency, risk proportionality, human oversight, and accountability (NIST, 2023, 2024; OECD, 2024b; UNESCO, 2023).

The remainder of the article follows the IMRaD structure required for an analytically transparent normative study. The method section explains the source domains, inclusion logic, normative coding procedure, and validity safeguards used to construct the framework. The results section identifies the provenance problem, formulates the five IRAPF requirements, and develops a three-tier model that differentiates lower-risk retrieval from consequential religious guidance. The discussion then situates the framework within digital religion, Islamic

communication ethics, and artificial intelligence governance while clarifying its theoretical and institutional implications. The conclusion summarizes the normative argument and identifies empirical, technical, and comparative studies that can test the framework across Muslim-majority and Muslim-minority contexts.

Method

Research Design and Methodological Orientation

This study employs normative-conceptual analysis rather than empirical fieldwork because its principal objective is to formulate standards for legitimate AI-mediated religious communication. Normative analysis evaluates what ought to count as responsible practice by relating concrete technological conditions to explicit ethical, epistemic, and institutional principles. The conceptual component then organizes those principles into a coherent model that can guide subsequent empirical testing and system design. This orientation is appropriate because the core questions concern legitimacy, responsibility, verification, and authority rather than prevalence or statistical association. Qualitative analytical procedures were informed by contemporary thematic analysis guidance, particularly the emphasis on transparent coding, conceptual coherence, and reflexive interpretation (Braun & Clarke, 2021). This guidance was adapted rather than applied directly: Braun and Clarke's reflexive thematic analysis was developed for coding empirical qualitative data such as interview transcripts, and what carries over to the present normative-conceptual synthesis is the general commitment to transparent, revisable coding categories, not the full reflexive-TA procedure, which presupposes a corpus of first-person empirical accounts that this study does not have. The overall research design also follows the principle that conceptual frameworks should make their selection criteria, analytical categories, and inferential transitions visible enough for other researchers to challenge or reproduce the reasoning.

Source Domains and Inclusion Strategy

The analytical corpus combined four source domains that perform different functions within the argument. The first domain consisted of Qur'anic normative principles relevant to verification, responsibility, knowledge claims, trustworthy communication, and referral to qualified knowledge, including Qur'an 49:6, 17:36, 4:58, and 16:43. The second domain consisted of recent Islamic scholarship on artificial intelligence, digital fatwa, religious authority, digital da'wah, maqasid, and Islamic technology ethics published mainly between 2021 and 2026 (Kurniawan et al., 2025). The third domain consisted of recent technical studies evaluating hallucination, source attribution, model reliability, abstention, and user trust in large language models. The fourth domain consisted of international governance instruments from NIST, UNESCO, OECD, the European Union, and related scholarship on artificial intelligence governance. Recent peer-reviewed sources were prioritized, while books were limited to publications within the previous decade, consistent with the recency requirements for this article. This recency requirement applies to the technical, governance, and digital-religion source domains described above, where currency of evidence matters for claims about model behaviour and policy instruments; it does not apply to the Qur'anic and classical normative material drawn on for the framework's Islamic grounding, where source age carries no comparable requirement.

Table 1. Analytical source domains used to construct the normative framework.

Source domain	Primary analytical function	Illustrative materials
Islamic normative sources	Identify verification, responsibility, expertise, and restraint principles	Qur'an 49:6; 17:36; 4:58; 16:43; authenticated transmission scholarship
Digital religion and Islamic AI scholarship	Map authority shifts and domain-specific religious risks	Recent studies on digital da'wah, algorithmic authority, AI fatwa, and Muslim GenAI use
Technical reliability research	Identify system-level limitations affecting epistemic trust	Hallucination, attribution, retrieval grounding, abstention, and trust calibration research
AI governance frameworks	Translate principles into organizational controls and proportional safeguards	NIST AI RMF, NIST GAI Profile, OECD principles, UNESCO guidance, and the EU AI Act

Selection and Analytical Relevance

Sources were retained when they contributed directly to at least one of four questions: how religious authority is formed in digital environments, how Islamic epistemic traditions regulate trustworthy knowledge transmission, how generative AI produces or obscures uncertainty, and how contemporary governance frameworks assign accountability for automated systems. Studies dealing only with historical corpus computation without implications for living religious communication were excluded because the journal scope prioritizes mediated religious life. This exclusion rests on editorial scope rather than on a conceptual criterion internal to constructing IRAPF; readers reproducing this corpus-construction logic outside this journal's specific scope should treat it as a publication-fit decision, distinct from the analytical inclusion criteria (thematic relevance to the four guiding questions) used elsewhere in this section. Materials were also excluded when they duplicated a stronger source without adding conceptual or empirical value. The resulting corpus was not treated as a statistically representative sample because normative research does not infer population parameters from documents. Instead, the corpus served as a purposive body of evidence for conceptual saturation across Islamic epistemology, digital religion, technical reliability, and governance. Because the search and screening process that bounded this corpus is not reported in step-by-step form above, the saturation claim should be read as an assessment internal to the sources actually consulted, rather than as a claim that has been verified against the full corpus of potentially relevant material.

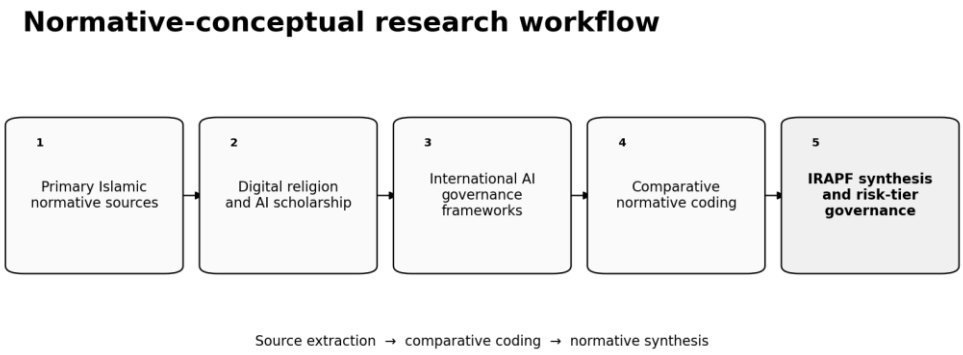


Figure 1. Normative-conceptual workflow connecting Islamic sources, current scholarship, governance concepts, and framework synthesis.

Normative Coding and Synthesis Procedure

The analysis proceeded through four linked stages that are summarized in Figure 1. During the first stage, normative propositions were extracted from Islamic sources and classified under verification, trustworthy transmission, epistemic humility, expertise, responsibility, and protection from harm. During the second stage, contemporary AI and digital religion literature was coded for recurrent problems involving opacity, false attribution, hallucination, overtrust, authority displacement, platform mediation, and accountability gaps. During the third stage, these two sets of categories were compared with governance concepts such as transparency, explainability, human oversight, auditability, risk proportionality, and organizational accountability. During the fourth stage, convergent categories were synthesized into operational requirements that remained recognizably Islamic while also being intelligible within contemporary AI governance discourse. The framework was refined until every proposed requirement could be justified from both a normative source and a documented technological risk.

Validity and Interpretive Safeguards

Three procedures were used to strengthen analytical validity within the limits of normative research. First, source triangulation required every major framework component to be supported by Islamic normative reasoning, contemporary religious communication scholarship, and current artificial intelligence research rather than by a single disciplinary literature. As implemented, this check required at least one supporting source from each of the three literatures to be attached to a component, which is a co-citation form of triangulation rather than the stronger methodological sense in which independent sources corroborate the same substantive conclusion; the latter, more demanding form was not systematically verified here and is noted as a limitation on this safeguard. Second, rival interpretation testing examined whether a less restrictive interpretation could preserve legitimate religious assistance without unnecessarily prohibiting beneficial technologies. Third, boundary testing applied the emerging framework to hypothetical low-risk, medium-risk, and high-risk religious use cases to determine whether the framework could distinguish informational assistance from consequential guidance. These safeguards do not transform normative reasoning into empirical proof, but they reduce arbitrary theorizing by exposing each normative claim to competing disciplinary considerations.

Ethical Scope and Researcher Position

No human participants, private records, or personally identifiable data were collected, so institutional review procedures for human subjects were not applicable to this document-based study. The analysis nevertheless treats religious communication as a sensitive domain because inaccurate guidance can affect worship, family decisions, financial conduct, identity, and spiritual well-being. The researcher therefore adopts a precautionary interpretive stance that does not assume technology is intrinsically harmful and does not treat traditional authority as technologically static. The evaluative question remains whether specific forms of mediation preserve verifiability, qualified interpretation, and accountable responsibility. This balanced stance is necessary because both technological rejection and uncritical adoption can obscure the actual governance choices available to Muslim institutions and developers.

Results

Epistemic Asymmetry Between Fluency and Traceable Religious Authority

Generative AI creates an epistemic asymmetry between rhetorical fluency and traceable religious authority. Large language models produce outputs through probabilistic prediction across training patterns and contextual prompts, while users encounter those outputs as coherent conversational statements. This architecture allows a response to sound complete even when the system cannot expose the full evidentiary pathway through which the statement was generated. Technical literature describes this problem through hallucination, unfaithfulness, opacity, and unverifiable synthesis (Heersmink et al., 2024; Huang et al., 2025). Religious research adds a more specific concern because persuasive fluency can be misrecognized as scholarly competence. Alam et al. (2025) demonstrate that users can prefer AI answers while experts identify substantive weaknesses, and Atif et al. (2025) show that models vary considerably across madhhab-sensitive questions and abstention behavior. The normative problem is therefore not limited to factual error, because the deeper issue is the presentation of epistemic confidence without equivalent visibility into source quality, interpretive method, and accountable authority.

Islamic Normative Foundations of Provenance Before Authority

Islamic norms provide a distinctive vocabulary for converting provenance from a technical convenience into a religious obligation. Qur'an 49:6 establishes a verification orientation before acting upon consequential information, while Qur'an 17:36 warns against following claims without knowledge. Qur'an 16:43 establishes referral to people of knowledge when the inquirer lacks competence, and Qur'an 4:58 connects entrusted responsibility with its proper discharge. These principles do not map mechanically onto software architecture, yet together they support a normative pattern in which information should be verifiable, uncertainty should not be concealed, expertise should be recognized, and entrusted communicative power should remain accountable. The hadith tradition's historical attention to transmission and narrator evaluation reinforces this orientation by demonstrating that Islamic epistemic legitimacy has never depended exclusively upon the surface plausibility of a statement (Brown, 2017). The resulting principle can be expressed succinctly as provenance before authority, meaning that persuasive religious guidance should not claim or imply authority beyond the traceability of its evidentiary and institutional basis.

The Five Normative Requirements of the IRAPF Model

Five normative requirements are necessary to operationalize this principle in generative AI-mediated da'wah. Source traceability requires the system to identify the relevant Qur'anic verses, hadith materials, juristic sources, institutional fatwas, or scholarly publications that

substantively support an answer rather than merely listing decorative citations. Epistemic transparency requires the system to distinguish direct textual evidence from interpretation, summarize meaningful disagreement, identify the legal or theological tradition being represented, and disclose material uncertainty. Tabayyun affordance requires the interface to make verification practically possible through accessible links, quotations within lawful limits, source metadata, and prompts encouraging confirmation when consequences are significant. Authority boundary requires the system to state when it is providing informational synthesis rather than a binding or personalized religious ruling. Human accountability requires a named institution, qualified reviewer, or governance process to retain responsibility for high-stakes deployment rather than transferring responsibility to the user through generic disclaimers.

These five requirements form the Islamic Religious AI Provenance Framework shown in Figure 2. The framework is intentionally relational because no single requirement can independently create legitimate authority. A system may cite sources yet remain misleading when citations do not support the generated claim, which means traceability requires epistemic transparency. A system may disclose uncertainty yet still encourage users to treat it as a substitute mufti, which means transparency requires an authority boundary. A system may include a disclaimer yet leave the user unable to verify anything, which means authority boundaries require tabayyun affordances. Finally, a well-designed interface can still fail institutionally when no accountable human or organization owns the consequences of deployment. The framework therefore treats legitimacy as an interaction between source quality, interpretive honesty, user verification capacity, role limitation, and organizational responsibility.

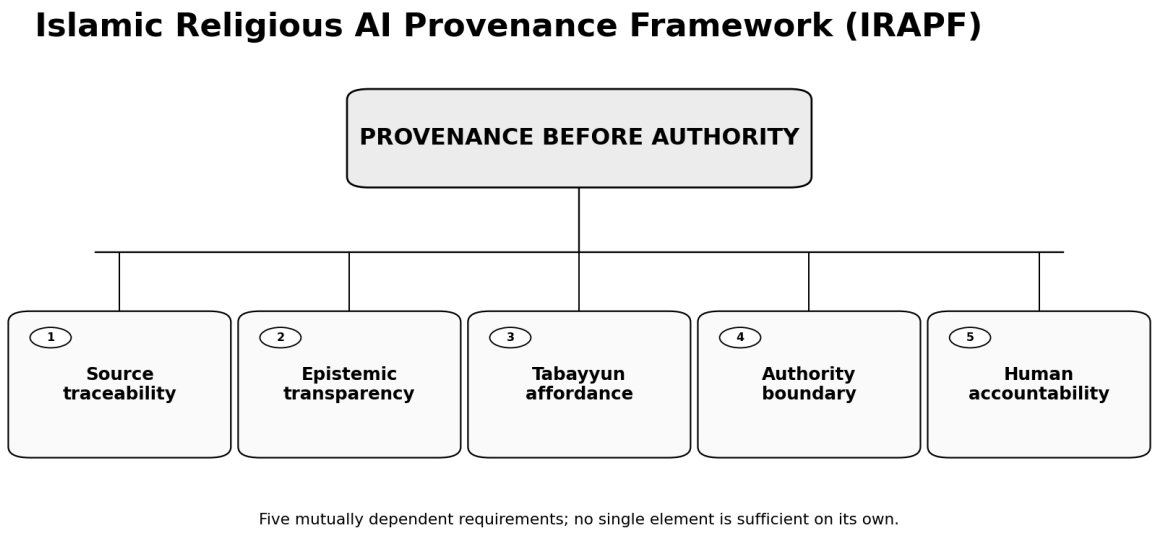


Figure 2. The IRAPF model places provenance before religious authority through five mutually dependent normative requirements.

Table 2. Operational requirements within the Islamic Religious AI Provenance Framework.

IRAPF requirement	Normative purpose	System expectation	Primary failure prevented
Source traceability	Connect claims with identifiable evidence	Provide supporting source metadata and claim-level attribution	Fabricated or decorative sourcing

IRAPF requirement	Normative purpose	System expectation	Primary failure prevented
Epistemic transparency	Prevent false certainty and hidden interpretation	Separate evidence, interpretation, disagreement, and uncertainty	Persuasive overconfidence
Tabayyun affordance	Make verification practically achievable	Enable users to inspect and cross-check consequential claims	Passive dependence on interface authority
Authority boundary	Distinguish assistance from juristic or spiritual authority	State task limits and escalate high-risk guidance	Substitution of AI for qualified scholarship
Human accountability	Keep responsibility addressable and correctable	Assign institutional owners, review procedures, and correction channels	Responsibility gaps after harmful outputs

Risk-Proportionate Governance of AI-Mediated Religious Guidance

Provenance requirements should increase with the religious consequences of the task. Tier One consists of retrieval and educational orientation, such as locating verses, identifying the general definition of a ritual term, listing widely recognized scholarly institutions, or summarizing an established topic for further study. These tasks still require accurate sourcing, but ordinary citation and uncertainty disclosure may provide proportionate safeguards. Tier Two consists of interpretive synthesis, such as comparing madhhab positions, explaining contested theological arguments, summarizing contemporary fatwas, or connecting scriptural principles with new social questions. These tasks require stronger disclosure of interpretive plurality, explicit identification of source traditions, and structured verification prompts. Tier Three consists of personalized rulings and high-stakes spiritual guidance, including individualized fatwa-like advice, marriage or divorce consequences, complex financial permissibility, accusations affecting religious standing, or guidance involving severe psychological or spiritual vulnerability. These contexts require qualified human oversight, conservative abstention defaults, auditable records, and clearly accountable institutional ownership.

The risk-tier model responds to a limitation in binary debates about whether artificial intelligence may or may not provide Islamic guidance. Existing empirical studies show that AI can sometimes perform useful retrieval, organization, translation, and preliminary comparison even while it remains unreliable for autonomous juristic judgment (Belhassen et al., 2025; Jan, 2026; Susanti et al., 2025). A binary prohibition would ignore these beneficial functions, while unrestricted use would ignore the growing evidence of confident error and user overtrust. Risk proportionality provides a more defensible approach because it connects governance intensity to the consequence of epistemic failure. This approach also aligns with broader artificial intelligence governance frameworks that scale controls according to foreseeable harm and system context (NIST, 2024; OECD, 2025). The three-tier structure itself follows an already-established risk-proportionality logic shared with the EU AI Act's minimal/limited/high-risk categories and with NIST's risk-based controls; what is specific to IRAPF is the substantive criterion used to assign a task to a tier, namely the religious consequence of epistemic failure (ritual, interpretive, or personal-guidance stakes), rather than the tiering architecture as such.

Abstention as Epistemic Restraint and Responsible Communication

Abstention should be recognized as a positive religious communication behavior in some circumstances. Contemporary model evaluation often rewards systems for producing an answer, but this incentive can encourage confident guessing when reliable knowledge is unavailable. FiqhQA demonstrates that models differ substantially in their willingness to abstain, and lower error can sometimes accompany more conservative refusal behavior (Atif et al., 2025). Islamic epistemic restraint provides normative support for this design choice because claiming knowledge without adequate basis is not equivalent to helpfulness. Under IRAPF, abstention is appropriate when a system cannot identify reliable sources, cannot determine the relevant interpretive tradition, cannot resolve materially conflicting evidence, or receives a personalized question whose consequences require contextual facts beyond the model's competence. The preferred response in such situations is not silence alone but structured referral to qualified human expertise together with a transparent explanation of why automated guidance would be unreliable.

Plurality, Madhhab Sensitivity, and Interpretive Provenance

Plurality within Islamic religious knowledge constitutes an additional provenance requirement. Generative systems frequently optimize for a single coherent answer, yet many religious questions legitimately permit more than one juristic or theological position. Compression into one response can therefore create a false impression of consensus even when the model has not produced a factual hallucination. IRAPF treats pluralism as part of provenance because users should know which interpretive tradition supports a recommendation and whether other recognized positions materially differ. This requirement becomes especially important when general-purpose models silently privilege the most common material in their training data. Recent benchmarks indicate that model performance and doctrinal preferences can vary across schools and languages, which means apparently neutral output can reproduce hidden distributional biases (Abdelaal et al., 2026; Atif et al., 2025). A provenance-aware system should therefore label school-specific positions when relevant, avoid presenting contested matters as universal rules, and identify when its source base is insufficiently balanced for comparative religious guidance (Mahendra et al., 2026).

Accountability Beyond Generic Disclaimers

Generic disclaimers are insufficient substitutes for accountable system design. Many religious applications state that users should consult scholars or verify answers independently, yet such statements can become ritualized warnings that do not alter the persuasive structure of the interface. If a chatbot produces confident personalized guidance, presents itself through anthropomorphic language, and hides its evidentiary uncertainty, a disclaimer placed elsewhere cannot fully correct the authority signal communicated by the interaction. IRAPF therefore treats responsibility as an architectural rather than merely textual issue. Institutions should document the sources used for retrieval, the scope of approved tasks, the escalation conditions for human review, the treatment of contested interpretations, and the procedures for correcting harmful outputs. This finding closely parallels contemporary governance research that defines trustworthy AI through organizational controls rather than isolated statements of ethical intention (Batool et al., 2025; NIST, 2023).

Discussion

The results answer the first research question by showing that the primary epistemic danger of generative AI-mediated religious communication is a decoupling of credibility from provenance. Previous studies of digital Islam already demonstrate that religious authority can become linked to visibility, accessibility, aesthetics, and networked influence rather than

exclusively to conventional scholarly credentials (Hidayatullah, 2025; Maulana & Anoraga, 2025; Zaid et al., 2022; Ramadhan et al., 2026). Generative AI introduces a qualitatively different layer because the communicative actor is not merely a new preacher using digital media but a synthetic system capable of producing context-responsive religious language at scale. The interface can simulate individualized attention while lacking the embodied scholarly formation, accountable institutional membership, and transparent interpretive commitments normally associated with religious advice. This does not mean users are irrational when they find AI useful, because clarity, accessibility, and organization are legitimate communicative values. The problem arises when these values become proxies for religious authority despite weak evidence about source quality and interpretive reliability.

This distinction extends digital religion theory by separating mediated visibility from generated epistemic authority. Platform studies usually explain how algorithms prioritize some religious actors over others, while recent Indonesian research shows that platform logics can reconstruct da‘wah authority through engagement metrics, cross-network exposure, and influencer practices (Rachman et al., 2025; Siregar et al., 2025; Tabaika et al., 2025). Generative AI does not simply amplify a human actor who already possesses a biography and institutional location. It composes a synthetic answer from probabilistic patterns and then presents that answer through a conversational persona whose apparent confidence may exceed its actual epistemic grounding. IRAPF therefore adds provenance as a missing dimension in theories of mediated religious authority. Concretely, the platform-visibility account in Rachman et al. (2025) explains how engagement metrics and cross-network exposure redistribute authority among human religious actors who already hold a biography and institutional position; it does not address a speaker that holds neither, which is the specific condition under which provenance, rather than visibility, becomes the operative variable. Authority in intelligent environments should be analyzed through at least three questions: who or what speaks, from which identifiable knowledge lineage the content is derived, and who remains answerable when the output causes religious or social consequences.

The second research question is answered through the operational translation of tabayyun, amanah, scholarly referral, and authenticated transmission into system requirements. Tabayyun becomes more than an instruction for users to fact-check because the system itself should provide verifiable pathways. Amanah becomes more than a virtue attributed to developers because institutional design should specify who holds responsibility for approved sources, model updates, escalation rules, and correction procedures. Scholarly referral becomes more than a disclaimer because the system should recognize task boundaries and actively route high-risk questions toward qualified experts. Authenticated transmission becomes more than a historical analogy to isnad because contemporary provenance mechanisms can record source identity, versioning, institutional endorsement, and the relationship between generated claims and supporting evidence. These translations preserve the normative orientation of Islamic epistemology without pretending that modern machine learning can literally reproduce classical scholarly institutions.

The framework also clarifies why source citation alone is not enough. Technical research on multi-source attribution shows that citation and evidence linkage can improve verifiability, but citations can still be incomplete, irrelevant, or strategically decorative (Patel et al., 2024). Within Islamic discourse, a verse or hadith can be accurately quoted while the surrounding interpretation remains contested, contextually inappropriate, or detached from juristic method. Provenance must therefore include interpretive metadata as well as bibliographic metadata. A strong religious AI system should identify whether a statement represents broad consensus, a school-specific position, a contemporary institutional fatwa, or a contested scholarly interpretation. It should also distinguish between direct evidence and generated explanation. This requirement is especially important in multilingual environments

because recent benchmarks show substantial performance differences across languages and doctrinal categories (Atif et al., 2025; Abdelaal et al., 2026).

The third research question concerns governance, and the risk-tier architecture provides a middle position between technological enthusiasm and categorical rejection. Islamic AI ethics research increasingly argues that technology should be assessed according to justice, welfare, harm prevention, dignity, and accountable human agency rather than according to novelty alone (Elmahjub, 2023; Kannike & Fahm, 2025; Raquib et al., 2022; Tawakkal et al., 2026). IRAPF complements these broad ethical principles by focusing on the communication chain through which religious claims become persuasive. Tier One applications may responsibly improve access when they retrieve authenticated materials and preserve source visibility. Tier Two applications require more careful representation of disagreement because interpretive synthesis can erase distinctions among legal schools or theological traditions. Tier Three applications require the strongest controls because individualized religious rulings can shape marriages, contracts, worship obligations, reputational judgments, and vulnerable users' spiritual decisions. The framework therefore asks institutions to govern not only what the model can technically generate but what the system should be permitted to represent as guidance.

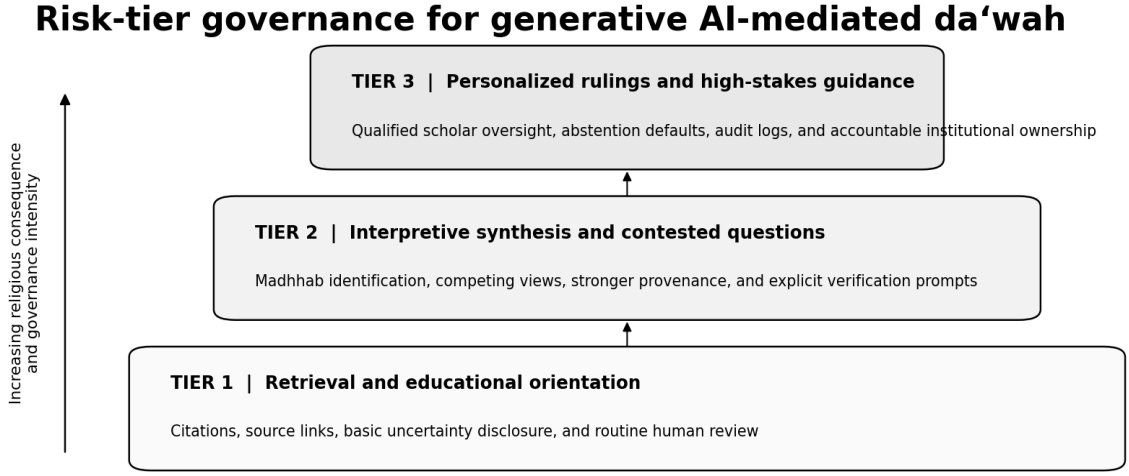


Figure 3. Risk-tier governance increases provenance, abstention, expert oversight, and institutional accountability as religious consequences become more serious.

This risk-sensitive approach aligns with international artificial intelligence governance without reducing Islamic ethics to imported regulatory language. NIST frameworks emphasize governance, mapping, measurement, management, transparency, and human oversight, while OECD principles emphasize explainability, robustness, and accountability (NIST, 2023, 2024; OECD, 2024a, 2024b). UNESCO similarly advocates human-centered governance and institutional validation for generative artificial intelligence in sensitive educational contexts (UNESCO, 2023). IRAPF shares these concerns but adds a specifically Islamic account of why provenance matters in religious communication. For an ordinary information service, transparency may primarily support user autonomy and auditability. For Islamic religious guidance, provenance additionally supports fidelity to recognized sources, awareness of legitimate interpretive plurality, and the moral responsibility attached to speaking about religious obligations. The framework therefore demonstrates how a faith-based normative vocabulary can contribute to pluralistic AI governance rather than merely receive norms developed elsewhere.

The discussion also reveals an important difference between trust and trustworthiness. Heersmink et al. (2024) distinguish phenomenological trust in language model interactions from the epistemic conditions that would justify such trust. The same distinction is essential in religious contexts because users may reasonably appreciate an interface while lacking sufficient information to evaluate its scholarly reliability. Alam et al. (2025) provide a concrete example of this divergence, where user preference for AI answers exceeded expert assessment of their quality. IRAPF therefore avoids making user satisfaction the final measure of alignment. A religiously aligned system should sometimes frustrate immediate user preference by disclosing uncertainty, presenting competing views, declining personalized rulings, or directing the user toward qualified scholars. Such friction can be normatively valuable because responsible guidance should protect users from persuasive overconfidence rather than merely optimize conversational acceptance.

The principle of authority boundary is particularly important because contemporary interfaces encourage anthropomorphic interpretation. Users may address a chatbot as though it understands intention, sincerity, lived circumstances, and moral responsibility, even though the system generates outputs from statistical relationships within data and prompts. Indonesian ethnographic research on AI ethics shows that Muslim actors already distinguish computational capability from niyyah, moral agency, and human responsibility (Wong-A-Foe, 2025). Islamic jurisprudential studies likewise resist treating computational fluency as autonomous ijtihad because qualified juristic reasoning depends upon contextual understanding, methodological competence, and accountable judgment (Belhassen et al., 2025). IRAPF translates this insight into interface governance: the system may assist with retrieval and structured comparison, but it should not communicate as though probabilistic generation possesses independent religious standing.

Human accountability must therefore remain visible at the organizational level. A mosque, university, fatwa institution, educational platform, or technology company that deploys religious AI should identify who approves source collections, who validates retrieval pipelines, who reviews high-risk outputs, and who responds when errors are discovered. This requirement is stronger than the common practice of shifting responsibility entirely to the user through terms of service. AI governance scholarship increasingly emphasizes that accountable systems require governance artifacts, documentation, lifecycle controls, and clearly assigned stakeholders (Batool et al., 2025). The same logic becomes more compelling when the application represents sacred texts or religious rulings. Institutional ownership does not guarantee correctness, but it creates an addressable chain of responsibility through which errors can be investigated, corrected, and learned from (Noh et al., 2026).

IRAPF also has implications for the design of retrieval-augmented generation systems. Retrieval can reduce some hallucination risks by grounding outputs in curated documents, yet current technical surveys caution that retrieval does not automatically eliminate unfaithfulness or unsupported synthesis (Huang et al., 2025). Religious applications should therefore separate retrieval confidence from interpretive confidence. A system may be highly confident that a passage appears in an authenticated source while remaining uncertain about how that passage applies to a user's circumstances. Interface design should make these levels visible, especially where multiple juristic positions exist. Source panels, school labels, confidence explanations, and explicit statements of unresolved disagreement can transform provenance from invisible backend infrastructure into an observable part of religious communication.

The framework further suggests that Muslim institutions should develop provenance registries rather than relying exclusively on general-purpose model behavior. Such registries could identify approved Qur'anic editions, hadith databases, recognized commentaries, legal manuals, institutional fatwas, and contemporary scholarly publications together with metadata concerning language, school, date, and authority status. Version control would allow

institutions to record when a source set changes, while audit logs could show which materials supported a consequential answer. These mechanisms would not reproduce classical isnad in a literal technological form, and claiming such equivalence would oversimplify a rich scholarly tradition. They would instead express the underlying normative commitment that religious claims should remain connected to identifiable and assessable sources. This design direction is consistent with recent Islamic chatbot research calling for expert-grounded systems rather than unbounded general-purpose generation (Susanti et al., 2025).

The growing problem of synthetic religious disinformation further strengthens the need for provenance-aware governance. Hidayah et al. (2026) document how generative AI can imitate clerical voices, sacred symbols, and emotionally resonant religious narratives in ways that hijack familiar markers of authority. When synthetic content can reproduce the appearance of a preacher or institution, users need mechanisms that authenticate not only textual claims but also the origin of audio, video, and institutional endorsements. IRAPF can therefore be extended beyond text chatbots toward multimodal religious media by treating provenance as a cross-format requirement. Future implementations could integrate content credentials, institutional signatures, source registries, and public correction records so that authenticity becomes technically inspectable rather than socially assumed.

The framework has direct implications for da‘wah practice. Digital da‘wah has often been evaluated through reach, engagement, clarity, and audience relevance, yet AI-mediated da‘wah requires additional criteria because automation can scale messages faster than human verification processes. A responsible institution should define what the system may answer, which materials it may use, how uncertainty should be presented, when a scholar must intervene, and how users can contest or report problematic guidance. These measures do not weaken da‘wah effectiveness, because reliable provenance can itself become a communicative value that strengthens trust over time. The approach also avoids framing technology as an external threat to religion, since the framework explicitly permits beneficial assistance where the system’s role remains proportionate and transparent (Fatmawati, Khoiriyah, et al., 2026).

A provenance-centered approach also creates an institutional memory that ordinary conversational systems often lack. Religious guidance becomes difficult to audit when each answer is transient, model versions change silently, and source collections evolve without visible records. Institutions deploying generative AI should therefore preserve versioned documentation showing which model, retrieval corpus, prompt policy, and review standard governed consequential outputs at a particular time. Such documentation can support correction when an error is reported and can reveal whether a problematic answer resulted from a defective source, an outdated institutional ruling, a retrieval failure, or an unsupported model inference. This function closely resembles broader lifecycle governance expectations within contemporary AI risk management, but it has an additional religious value because institutions should be able to explain how entrusted religious communication was produced and supervised. Provenance therefore operates not only at the level of individual citations but also at the level of institutional memory and accountable historical record.

The framework additionally protects interpretive plurality from being flattened by optimization for conversational simplicity. Muslim users may belong to different legal schools, theological orientations, institutional traditions, and minority contexts, while general-purpose models frequently answer from an implicit composite that does not disclose whose methodology is being followed. A system that presents one synthesized position without attribution can unintentionally transform statistical predominance into normative dominance. IRAPF responds by requiring systems to reveal when answers are school-specific, when established disagreement exists, and when the available evidence does not justify prioritizing one position. This approach does not require every answer to become an exhaustive jurisprudential encyclopedia, because proportionality still applies. It does require the system to

avoid manufacturing consensus through omission whenever disagreement would materially affect a user's understanding or conduct. In this sense, provenance also functions as a safeguard for intra-Muslim epistemic diversity.

The same logic has implications for model procurement and institutional partnerships. Muslim organizations increasingly depend upon commercial foundation models whose training data, update cycles, moderation policies, and system prompts remain partly inaccessible. Responsible deployment therefore requires governance at the application layer even when full model transparency is unavailable. Institutions can constrain approved tasks, maintain curated retrieval corpora, test citation entailment, monitor outputs, record escalation events, and publish clear statements about the limits of the underlying model. Where a vendor cannot support necessary auditability for high-stakes religious use, the appropriate response may be to restrict the system to lower-risk retrieval or educational functions. This approach makes procurement itself part of religious accountability because institutional leaders remain responsible for the technological conditions under which guidance is communicated to their communities.

Several limitations should qualify the claims advanced in this study. Normative-conceptual analysis cannot establish how Muslim users will actually respond to provenance features, whether stronger source transparency will reduce overtrust, or which interface designs best communicate juristic plurality. The framework also draws heavily upon Sunni-oriented scholarship and contemporary Indonesian discussions, although the underlying principles could be examined comparatively across Shi'i traditions, minority Muslim contexts, and different institutional cultures. Furthermore, the rapid evolution of models means that technical limitations documented in 2024 or 2025 may change, although improvements in accuracy would not remove the normative questions of authority and accountability. These limitations support empirical validation rather than weakening the need for a clear normative benchmark.

Future research should therefore test IRAPF through experiments, audits, and participatory design. Controlled studies could compare ordinary AI answers with answers that include source traceability, uncertainty disclosure, madhhab labels, and scholar referral to measure effects on trust calibration and verification behavior. Technical audits could test whether cited sources genuinely entail generated claims across Arabic, Indonesian, and English. Institutional studies could examine how fatwa bodies, pesantren, universities, and da'wah organizations assign responsibility when deploying AI tools. Comparative research could also evaluate whether the provenance-before-authority principle travels across religious traditions that possess different models of sacred authority. These empirical directions would transform the present normative framework into a testable research program for intelligent digital religion (Fatmawati, Musyarofah, et al., 2026).

Conclusion

This article argues that the central governance problem in generative AI-mediated da'wah is not simply whether artificial intelligence can produce religiously plausible sentences, but whether persuasive output remains connected to verifiable sources, transparent interpretation, bounded authority, and accountable human institutions. The normative analysis identifies a structural mismatch between probabilistic generation and the provenance-sensitive character of Islamic religious knowledge. Islamic principles of tabayyun, amanah, scholarly referral, epistemic restraint, and authenticated transmission provide a coherent basis for evaluating this mismatch without requiring technological rejection. These principles support the broader proposition that provenance should precede authority whenever automated systems communicate consequential religious guidance.

The study answers its research questions through the Islamic Religious AI Provenance Framework. First, source opacity, confident uncertainty, collapsed interpretive plurality, and

responsibility gaps explain why fluent AI output can become epistemically misleading even when portions of the content are factually correct. Second, the five IRAPF requirements translate Islamic norms into operational expectations concerning source traceability, epistemic transparency, tabayyun affordance, authority boundary, and human accountability. Third, the three-tier governance model distinguishes lower-risk retrieval from interpretive synthesis and high-stakes personalized guidance, requiring progressively stronger verification, abstention, expert review, and institutional ownership. This proportional structure provides a more useful governance approach than categorical acceptance or rejection of religious artificial intelligence.

The framework also opens a practical research agenda for Muslim institutions and technology developers. Future systems should be evaluated not only through answer accuracy but also through citation entailment, source lineage, uncertainty communication, representation of disagreement, escalation behavior, and accountability mechanisms. Empirical research should test whether provenance features improve trust calibration among different Muslim publics and whether scholars consider the resulting outputs more defensible. The broader implication is that intelligent religious media can support access and education while remaining subordinate to transparent knowledge lineages and accountable human judgment. Such an approach preserves technological usefulness without allowing synthetic fluency to become an unexamined substitute for religious authority.

Declaration of Generative AI Use

Generative artificial intelligence was used in the preparation of this manuscript. Anthropic's Claude typeset the manuscript into the journal template, which covered the placement of headings, tables, figures, and the reference list.

Authors' Contribution Statement

Yasirul Amri: conceptualization, methodology, formulation of the IRAPF framework, writing of the original draft, and project administration.
Ansori Hidayat: literature search and screening, data curation, analysis of digital religion and AI governance sources, and writing (review and editing).
Irhamuddin: analysis of Qur'anic and classical Islamic sources, validation of the normative foundations, and writing (review and editing).
Muhammad Fadhil Hadziq: investigation of empirical studies on religious large language models, validation of the risk tier model, and writing (review and editing).
All authors contributed to the discussion of findings, read, and approved the final manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

Abdelaal, A., Al Haffar, M. N., Fawzi, M., & Magdy, W. (2026). IslamicMMLU: A benchmark for evaluating LLMs on Islamic knowledge. arXiv. https://arxiv.org/abs/2603.23750

Ahmad, Y. S., Abdulrasaq, Y., Ali, A. M., & Ali, F. (2025). Integrating Islamic ethical principles into AI education and character building in digital society. *Tunjuk Ajar: Journal of Education and Culture, 1*(2), 130–146. https://doi.org/10.64929/ta.v1i2.38

Alam, S. M., Abdulhai, M., & Salehi, N. (2025). Blind faith? User preference and expert assessment of AI-generated religious content. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 2410–2438). ACM. https://doi.org/10.1145/3715275.3732162

Anugrah, P., Ilyas, A. R., Kurniati, & Mustafa, Z. (2025). The limits of artificial intelligence in judicial reasoning: Jurisprudential analysis of Islamic legal authority. Jurnal Kajian Islam Interdisipliner, 10(2), 139–162. https://doi.org/10.14421/4nhp7791

Atif, F., Askarbekuly, N., Darwish, K., & Choudhury, M. (2025). Sacred or synthetic? Evaluating LLM reliability and abstention for religious questions. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 8(1), 217–226. https://doi.org/10.1609/aies.v8i1.36543

Aykul, A., & Alam, M. N. (2024). ChatGPT'nin İslâm ceza hukuku özelinde değerlendirilmesi. Dinbilimleri Akademik Araştırma Dergisi, 24(3), 193–212. https://doi.org/10.33415/daad.1582902

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. AI and Ethics, 5, 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Belhassen, Y., Ali, A., & Hachicha, S. (2025). Algorithmic ijtihad and the reconfiguration of religious authority: Evaluating the epistemic and ethical boundaries of generative AI in Islamic jurisprudence. International Jornal of Noesantara Islamic Studies, 2(2), 77–87. https://doi.org/10.70177/ijonis.v2i2.3287

Braun, V., & Clarke, V. (2021). Thematic analysis: A practical guide. SAGE.

Brown, J. A. C. (2017). Hadith: Muhammad's legacy in the medieval and modern world (2nd ed.). Oneworld Publications.

Campbell, H. A., & Tsuria, R. (Eds.). (2021). Digital religion: Understanding religious practice in digital media (2nd ed.). Routledge.

Effendi, E., Waliyajati, R. S. R., & Truna, D. S. (2026). Reconfiguring religion in the digital era: A conceptual framework for the mediatization of religion in Indonesia. Abrahamic Religions: Jurnal Studi Agama-Agama, 6(2). https://doi.org/10.22373/arj.v6i2.32996

Elmahjub, E. (2023). Artificial intelligence in Islamic ethics: Towards pluralist ethical benchmarking for AI. Philosophy & Technology, 36, 73. https://doi.org/10.1007/s13347-023-00668-x

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union.

Fatimah, N., Murjazin, & Kharis, M. A. (2026). Algorithmic ulama: Generative AI, religious authority, and the reconfiguration of da'wah communication ethics in Indonesian Islam. SUHU: Journal of Sufism and Humanities, 2(1). https://doi.org/10.69548/suhu.v2i1.94

Fatmawati, F., Khoiriyah, N., Arbi, A., & Minangsih, K. (2026). Digital da'wah management and anti-hoax literacy: A multicultural local wisdom reading: Manajemen dakwah digital dan literasi anti-hoaks: Pembacaan kearifan lokal multikultural. Alhadharah: Jurnal Ilmu Dakwah, 25(1), 121–134. https://doi.org/10.18592/alhadharah.v25i1.20548

Fatmawati, F., Musyarofah, U., & Minangsih, K. (2026). Digital da'wah content in Islamic religious education: Critical reception and curriculum integration in Banten senior high schools. Reslaj: Religion Education Social Laa Roiba Journal, 8(7), 900–918. https://doi.org/10.47467/reslaj.v8i7.13058

Firdaus, M. A., Syihabuddin, M., & Fuady, Z. (2025). Islam and artificial intelligence: Perspectives from traditionalist and modernist Muslim communities in Indonesia. MIQOT: Jurnal Ilmu-ilmu Keislaman, 49(1).

Haditia, M., Abubakar, I. Y., Rokimin, & Fauzi, A. (2026). Design and validation of a quality assurance framework for AI educational technology in low-resource settings. *Educational Technology in Developing Countries, 1*(1), 83–102. https://doi.org/10.64929/etdc.v1i1.68

Heersmink, R., de Rooij, B., Clavel Vázquez, M. J., & Colombo, M. (2024). A phenomenology and epistemology of large language models: Transparency, trust, and trustworthiness. Ethics and Information Technology, 26, 41. https://doi.org/10.1007/s10676-024-09777-3

Hidayah, N., Murfianti, F., Purnama, F. Y., & Fadhilah, N. (2026). Sacred symbols, synthetic lies: The hijacking of religious authority in AI-generated religious disinformation in Indonesia. ESENSIA: Jurnal Ilmu-Ilmu Ushuluddin, 27(1). https://doi.org/10.14421/esensia.v27i1.7344

Hidayatullah, R. (2025). Otoritas keagamaan digital: Pembentukan otoritas Islam baru di ruang digital. Ushuluna: Jurnal Ilmu Ushuluddin, 10(2). https://doi.org/10.15408/ushuluna.v10i2.42831

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems, 43(2). https://doi.org/10.1145/3703155

Imtiaz, M., & Shafique, H. M. M. (2025). Large language models as Islamic guidance tools: Trust, limitations, and ethical boundaries. Al-Irfan: Research Journal of Islamic Studies, 10(19). https://doi.org/10.58932/MULB0063

Jahan, I., Kamruzzaman, S., Hossen, K. M. N., Nurain, N., & Al Islam, A. B. M. A. (2026). From mind mapping to dream interpretation: Understanding Muslim users' engagement with GenAI tools in religious context. In Human-Computer Interaction (pp. 353–372). Springer. https://doi.org/10.1007/978-3-032-29586-6_23

Jan, A. (2026). Digital authority and Islamic legal reasoning: A comparative study of AI-generated and human-issued fatāwā. Online Journal of Research in Islamic Studies, 13(1). https://doi.org/10.22452/ris.vol13no1.1

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), Article 248. https://doi.org/10.1145/3571730

Kannike, U. M. M., & Fahm, A. O. (2025). Exploring the ethical governance of artificial intelligence from an Islamic ethical perspective. Jurnal Fiqh, 22(1), 134–161. https://doi.org/10.22452/fiqh.vol22no1.5

Kosasih, E., Islamy, M. R. F., & Wiwaha, R. S. (2024). Artificial intelligence in the era of Society 5.0: Compromising technological innovation through the wasathiyyah approach within the framework of Islamic law. Al-Istinbath: Jurnal Hukum Islam, 9(2). https://doi.org/10.29240/jhi.v9i2.9596

Kurniawan, E., Ahmad, H., & Muttaqin, M. Z. (2025). From content creators to zakat payers: The fatwa of Indonesian Ulema Council and the rise of digital fiqh. *Islamic Law and Social Issues in Society, 1*(2), 99–115. https://doi.org/10.64929/ilsiis.v1i2.9

Mahendra, B. A., Munandar, A., & Nurzuliana, Z. (2026). Mahar as esteem, not price: Mediating classical fiqh, the Compilation of Islamic Law, and custom in Indonesia's

plural legal order. *PFL: Plural Family Law Review, 1*(1), 59–80. https://www.ejournaltuah.id/index.php/pfl/article/view/13

Maulana, M. D., & Anoraga, B. (2025). Building algorithmic religious authority through different religious networks. Jurnal Dakwah Risalah, 36(2). https://doi.org/10.24014/jdr.v36i2.36050

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce.

National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). U.S. Department of Commerce.

Nawi, A., Yaakob, M. F. M., Ren, C. C., Khamis, N. Y., & Tamuri, A. H. (2021). A preliminary survey of Muslim experts' views on artificial intelligence. Islamiyyat: The International Journal of Islamic Studies, 43(2).

Noh, M. A. C., Basri, M., & Panaemalae, A. (2026). Ethical leadership: Reframing moral complexity and power through a philosophical anthropological perspective in organizational practice. *Leadership in Muslim Societies, 1*(1), 19–34. https://doi.org/10.64929/lms.v1i1.57

Noordin, M. F., & Zailan, S. N. S. (2025). AI from an Islamic perspective. Journal of Information Systems and Digital Technologies, 7(2), 102–114. https://doi.org/10.31436/jisdt.v7i2.630

OECD. (2024a). AI, data governance and privacy. OECD Publishing.

OECD. (2024b). OECD AI principles: Updated recommendations on trustworthy artificial intelligence. OECD.

OECD. (2025). Governing with artificial intelligence. OECD Publishing.

Patel, N., Subramanian, S., Garg, S., Banerjee, P., & Misra, A. (2024). Towards improved multi-source attribution for long-form answer generation. In Proceedings of NAACL-HLT 2024 (pp. 3906–3919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.216

Rachman, A., Saumantri, T., & Hidayatulloh, T. (2025). Transformation of religious authority in the digital era: A post-normal times analysis of the phenomenon of social media da'wah. Jurnal Ilmu Dakwah, 45(1). https://doi.org/10.21580/jid.v45.1.25644

Ramadhan, M. A. I., Syam, A. R., & Barkah, S. A. M. (2026). Customay judges in digital space: Authority contestations, living fiqh, and legal pluralism in Bugis customary inheritance dispute resolution via WhatsApp groups. *Tuah Customary Law Review, 1*(1), 1–20. https://ejournal.tuah.or.id/index.php/tclr/article/view/74

Raquib, A., Channa, B., Zubair, T., & Qadir, J. (2022). Islamic virtue-based ethics for artificial intelligence. Discover Artificial Intelligence, 2, 11. https://doi.org/10.1007/s44163-022-00028-2

Rizqullah, M. R., Purwarianti, A., & Aji, A. F. (2023). QASiNa: Religious domain question answering using Sirah Nabawiyah. arXiv. https://arxiv.org/abs/2310.08102

Siregar, K. T., Zhakypkazy, I., & Uyuni, B. (2025). Indonesian muballigh and religious authority: From sanad to algorithmic transformation of Islamic knowledge transmission in the digital era. Jurnal Indo-Islamika, 15(2). https://doi.org/10.15408/jii.v15i2.47052

Susanti, N. F. H., Tanjung, H., Supriyanto, T., Athief, F. H. N., & Furqoni, A. (2025). ChatGPT's validity and reliability in Islamic legal studies: A critical assessment of waqf for early learners. Suhuf: International Journal of Islamic Studies, 37(2). https://doi.org/10.23917/suhuf.v37i2.11821

Tabaika, M. A., Barizi, A., & Arif, Y. M. (2025). Digital da'wah and the reconstruction of Islamic authority. Al-Balagh: Jurnal Dakwah dan Komunikasi, 10(2). https://doi.org/10.22515/albalagh.v10i2.12116

Tawakkal, M. G., Kurniawan, M. I., & Faathima. (2026). Cash waqf linked sukuk and economic empowerment: A grounded theory from Indonesia. *Ethiconomics: Journal of Islamic Banking and Finance, 1*(01), 1–20. https://doi.org/10.64929/ethiconomics.v1i01.2

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO.

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

Ünal, S. (2024). İbadetlerle ilgili fetvalar ve yapay zekâ uygulamaları: Karşılaştırmalı bir analiz. Dinbilimleri Akademik Araştırma Dergisi, 24(3).

Wong-A-Foe, D. (2025). Between code and creed: Islamic ethical in-betweenness on AI in Indonesia. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 8(3), 2939–2941. https://doi.org/10.1609/aies.v8i3.36800

Zaid, B., Fedtke, J., Shin, D. D., El Kadoussi, A., & Ibahrine, M. (2022). Digital Islam and Muslim millennials: How social media influencers reimagine religious authority and Islamic practices. Religions, 13(4), 335. https://doi.org/10.3390/rel13040335