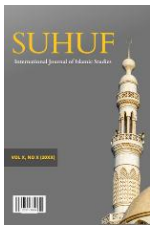
SUHUF

Vol. 38, No. 1, 2026

International Journal of Islamic Studies

journal homepage: https://journals2.ums.ac.id/index.php/suhuf/index



Can Artificial Intelligence Perform *Ijtihad*? Authority, Epistemology, and Legitimacy in Digital Islamic Jurisprudence

Alwy Ahmed Mohamed[a,b,1,*], Mowafg Masuwd[c,2], Muhammad Rehan Sabir[d,3], Mahmudulhassan[e,4]

[a] Universitas Muhammadiyah Surakarta, Surakarta, Indonesia
[b] The Aga Khan High School, Mombasa, Kenya
[c] University of Zawia, Zawiya, Libya
[d] Government College University Faisalabad, Punjab, Pakistan
[e] Islamic Arabic University, Dhaka, Bangladesh

[1] alwyahmed94@gmail.com; [2] masuwd@zu.edu.ly; [3] rehanzayer@gmail.com; [4] hassanfaruky@gmail.com
* Corresponding Author

https://doi.org/10.23917/suhuf.v38i1.16437

ARTICLE INFO

ARTICLE HISTORY
Received Month 02, 2026
Revised Month 04, 2026
Accepted Month 05, 2026

KEYWORDS
Algorithmic authority
Artificial intelligence
Digital fatwa
Islamic legal theory
Ijtihad
usul al-fiqh

ABSTRACT

Background: Artificial intelligence (AI) can rapidly search juristic sources, compare legal opinions, and generate structured arguments. Yet its fluency may obscure the distinction between computational capability and legitimate authority in Islamic jurisprudence. **Objective**: This study examines whether AI can independently perform *ijtihad* and identifies conditions for its responsible use without displacing human juristic authority. **Methodology**: A structured normative-doctrinal approach integrates classical *usul al-fiqh* and fatwa doctrine with scholarship on generative AI, algorithmic accountability, and AI governance. Analysis applies five criteria: source competence, methodological reasoning, contextual judgment, moral agency, and answerability. **Results**: Current AI systems satisfy only instrumental aspects of source competence and methodological reasoning. They can retrieve sources, map disagreements, identify analogies, and organize preliminary arguments, but cannot possess legally recognized intention, bear *taklif*, assume responsibility for harm, or receive juristic authorization. The study proposes a five-layer Augmented *ijtihad* Model: source retrieval, doctrinal mapping, inferential support, contextual and *maqasid* evaluation, and final authorization. **Conclusion**: AI can strengthen *ijtihad* but cannot function as an independent *mujtahid*. **Implication**: Islamic legal institutions should require verified databases, school-of-law tagging, uncertainty disclosure, qualified human review, audit trails, contestability, and post-deployment monitoring, positioning AI as accountable research infrastructure under scholarly supervision rather than autonomous religious authority.

This is an open-access article under the CC-BY license.

1. Introduction

Generative artificial intelligence has moved the problem of automated religious advice from speculation to ordinary practice. A user can now enter a question concerning worship, family relations, finance, medical treatment, or social conduct and receive a fluent answer within seconds. Such systems can summarize legal doctrines, retrieve scriptural passages, compare juristic positions, and imitate the argumentative form of a fatwa. These capabilities create genuine opportunities for access and research. They also create a distinctive epistemic risk: a grammatically confident response may appear authoritative even when its sources are incomplete, its attribution is wrong, its treatment of disagreement is selective, or its application to the questioner's circumstances is defective. Hallucination in natural-language generation is not merely a technical inconvenience in this setting. When the output is presented as religious law, factual unreliability can be converted into a claim of normative authority [1], [2].

The institutional significance of the problem is already visible. Research on automated fatwa systems has documented the use of case-based reasoning, text classification, and generative techniques to retrieve or formulate Islamic legal opinions [3]. Studies of AI and the mufti emphasize accessibility and efficiency while warning that context-sensitive judgment cannot be reduced to textual similarity [4]. Empirical work in Islamic finance likewise finds that Sharia advisers value AI's data-processing capacity but remain concerned about verification costs, automation of sacred judgment, and the weakening of scholars' own *istinbat* competencies [5]. Egypt's Dar al-Ifta has taken a more restrictive institutional position, permitting AI as a means in principle while rejecting reliance on current applications for fatwa because they do not reliably integrate authoritative sources, legal method, changing custom, and the questioner's circumstances [6]. These developments show that the relevant issue is no longer whether Islamic legal institutions will encounter AI. It is how they should classify its role.

The emerging literature approaches this issue through three partially connected conversations. The first develops Islamic ethical frameworks for AI. Virtue-based, *maqasid*-based, pluralist, and trusteeship approaches have shown that responsible AI cannot be assessed only through accuracy or utility; it must also be evaluated in relation to human dignity, moral formation, justice, welfare, and accountability before God [7], [8], [9]. Recent work has further examined the theological meanings embedded in technological systems and cautioned against reducing technology to a morally neutral means [10]. This scholarship corrects the cultural narrowness of universalized AI ethics, but it generally addresses AI governance at a broad ethical level rather than the specific legal architecture of *ijtihad*.

The second conversation concerns AI-assisted fatwa and Islamic legal reasoning. It identifies useful functions such as retrieval, classification, comparison, and preliminary drafting, while usually concluding that final authority should remain with human scholars [4], [11], [12]. This conclusion is broadly persuasive, but the formula "AI as a tool under human supervision" is under-specified. It does not tell institutions which tasks are delegable, how much review counts as meaningful supervision, what evidence must accompany an output, how disagreement between schools should be represented, or who bears responsibility when a technically assisted ruling causes harm. Without these specifications, "human in the loop" can become ceremonial approval rather than substantive juristic control.

The third conversation concerns AI governance. International frameworks converge around transparency, fairness, safety, accountability, privacy, and human oversight [13], [14]. NIST translates these values into organizational functions of governing, mapping, measuring, and managing AI risk [15], while its generative-AI profile emphasizes confabulation, information integrity, data privacy, and human-AI configuration risks [16]. Yet high-level principles do not automatically produce accountable practice. Shows that principles are weak when institutions lack

shared professional duties, translation mechanisms, and enforceable responsibility. For Islamic legal institutions, the missing link is a jurisprudential account of what may be delegated and what must remain attached to a qualified and answerable juristic subject [17].

This article addresses that gap by distinguishing computational competence from normative standing. Computational competence concerns what a system can functionally produce: retrieval, classification, prediction, comparison, or draft reasoning. Normative standing concerns whether the producer is entitled to issue a legal-religious judgment and can be called to account for doing so. The distinction matters because accurate assistance does not itself create authority, just as a searchable legal database does not become a judge and a medical imaging system does not become the physician responsible for treatment. In Islamic jurisprudence, the distinction is sharper because *ijtihad* is embedded in a normative order of knowledge, intention, responsibility, and recognized scholarly practice.

Accordingly, the article asks three questions:

1. Which legal, epistemic, and ethical conditions make *ijtihad* an authoritative juristic act rather than a sequence of information-processing operations?
2. Which functions within an *ijtihad* workflow may be delegated to AI, and which functions remain non-delegable?
3. What institutional safeguards are required for legitimate AI-supported Islamic legal reasoning?

The article makes three contributions. First, it develops a non-equivalence test that separates a system's capacity to reproduce parts of legal reasoning from its standing to perform *jtihad*. Second, it disaggregates *ijtihad* into five layers and specifies a decreasing gradient of permissible delegation. Third, it converts general calls for human oversight into an operational governance framework built around provenance, pluralism, qualified review, auditability, contestability, and continuing responsibility. The resulting Augmented *ijtihad* Model retains the central conclusion that AI cannot operate as an independent *mujtahid*, while showing in concrete terms how it can strengthen the research environment in which accountable jurists work.

2. Conceptual Foundations

2.1 *Ijtihad*, fatwa, and juristic authority

Ijtihad is commonly described as the exertion of qualified juristic effort to determine a Sharia ruling when the answer is not simply given by an unambiguous and dispositive text. This definition contains both an activity and a status. The activity includes identifying relevant sources, evaluating transmission and language, reconciling evidence, reasoning analogically, considering custom, and assessing consequences. The status concerns the qualifications and recognized responsibility of the person who undertakes that activity. Classical legal theory therefore does not treat every successful inference as *ijtihad* in the authoritative sense. It asks who reasons, from which sources, through which methods, within which school or interpretive tradition, and under what duties [18], [19], [20].

A fatwa is related to but not identical with *ijtihad*. It is a non-binding juristic response to a question, issued by a mufti and addressed to a *mustafti*. It may report an established school position, select among recognized views, or involve fresh reasoning. Its form is dialogical and contextual: the quality of the ruling depends partly on how accurately the question and surrounding facts are understood. The classical literature on fatwa therefore attends not only to substantive doctrine but also to the ethics of the mufti, the formulation of the question, the distinction between fatwa and judicial decision, and the consequences of applying a general rule to a particular case [21], [22]. A digital system that produces legal information may participate in this communicative chain without acquiring the office or status of the mufti.

Juristic authority is consequently relational. It is not a property generated by the sophistication of an answer alone. It emerges from demonstrated learning, fidelity to legal method, ethical reliability, recognition by competent communities, and answerability for the act of speaking in the name of the law. Hallaq's account of authority and continuity in Islamic law shows how juristic change occurred through disciplined engagement with inherited doctrine rather than unbounded individual creativity [19]. The categories of *mujtahid mutlaq*, school-affiliated *mujtahid*, and more limited juristic roles illustrate graded competence, but none treats authority as self-certifying. This matters for AI because performance benchmarks can measure some outputs, while juristic authorization remains an institutional and normative judgment.

2.2 Constitutive conditions and supporting competencies

The requirements associated with *ijtihad* can be organized into two groups. Supporting competencies include knowledge of the *Quran* and *Sunnah*, Arabic language, legal theory, consensus, analogy, substantive doctrine, and the ability to identify and reconcile evidence. These competencies enable juristic reasoning and can be assisted by external tools. Libraries, indexes, concordances, databases, and collaborative research have always extended the jurist's capacity without becoming jurists themselves.

Constitutive conditions, by contrast, concern the kind of accountable act that *ijtihad* is. They include intention (niyyah), moral integrity (adalah), responsibility under the law (taklif), prudential awareness of context, and acceptance of responsibility for the ruling's justification and foreseeable effects. Not every school states these conditions in identical terminology or assigns them the same doctrinal location. The shared point is narrower: legitimate juristic speech is attributable to an answerable subject. The jurist must be capable of explaining why a source was selected, why one interpretation was preferred, how relevant disagreement was treated, and why a general norm properly applies to the facts presented.

The distinction between supporting and constitutive conditions prevents two opposite mistakes. The first is technological exceptionalism, which assumes that because machines are novel, none of their functions can enter juristic practice. Many research operations are instrumentally delegable. The second is functional reductionism, which assumes that reproducing enough operations produces the normative office itself. A system may outperform a scholar in corpus search and still lack the standing to authorize a ruling. The legal significance of AI therefore cannot be decided by a single comparison between "human intelligence" and "machine intelligence." It requires task-level analysis.

2.3 AI as a sociotechnical arrangement

For this analysis, "AI" does not denote a self-contained artificial mind. It denotes a sociotechnical arrangement comprising a model, training or retrieval data, system instructions, interface design, developers, deployers, reviewers, and an institution that determines how outputs will be used. This framing is essential because bias and authority enter at several points: corpus selection may privilege one school; metadata may erase the difference between binding text and later commentary; interface design may present one answer as settled; and institutional workflow may pressure reviewers to approve outputs quickly. Treating only the model as the object of analysis would hide the human decisions that configure its apparent authority [23].

Large language models generate text through probabilistic relationships learned from data. Their fluency does not guarantee source fidelity, doctrinal consistency, or stable reasoning across rephrased prompts. They may attribute statements to the wrong jurist, merge incompatible positions, omit minority views, or produce a plausible citation that does not exist [1], [2]. Retrieval-augmented generation and constrained corpora can reduce these risks, but they do not eliminate the normative choices involved in selecting the corpus, ranking sources, resolving contradictions, and determining

acceptable uncertainty. The system's outputs should therefore be treated as defeasible research artifacts rather than presumptively authoritative conclusions.

2.4 Three category errors

Three category errors recur in debates about AI and *ijtihad*. The functional-equivalence error moves from "AI can perform operation X" to "AI possesses the status of the professional who sometimes performs X." It ignores that professional authority comprises duties, relationships, and institutional recognition beyond task execution. The output-authority error infers legitimacy from surface quality. A coherent answer may still be sourced badly, contextually inappropriate, or outside the system's authorized domain. The nominal-oversight error assumes that any human approval preserves human authority. Oversight is meaningful only when the reviewer is qualified, receives adequate evidence, has sufficient time, can reject the output, and remains identifiable as the person responsible for the final judgment. These errors establish the need for a method that evaluates both functional capability and normative allocation.

3. Method

3.1 Research design

The study uses a structured normative-doctrinal method combined with conceptual analysis and institutional design. Doctrinal research is appropriate because the principal question concerns the content and implications of juristic categories, not the frequency of attitudes in a population. The method identifies authoritative concepts, tests their coherence when applied to a new technological setting, and formulates a normative framework [24]. The analysis is comparative in a limited sense: it places Islamic doctrines of authority and responsibility in dialogue with contemporary accounts of AI capability and governance without treating the two normative systems as interchangeable.

This is not a systematic literature review and does not claim exhaustive coverage. Sources were selected purposively according to three criteria. First, doctrinal authority guided the selection of classical and modern works on *usul al-fiqh*, *ijtihad*, fatwa, custom, *maqasid*, and juristic authority. Second, direct topical relevance guided the selection of peer-reviewed studies on AI, fatwa, Islamic legal reasoning, Islamic AI ethics, generative-model limitations, and algorithmic accountability. Third, institutional authority guided the selection of governance instruments and contemporary fatwa positions, including UNESCO's Recommendation on the Ethics of Artificial Intelligence, the NIST AI Risk Management Framework and Generative AI Profile, and the 2025 position of Dar al-Ifta al-Misriyyah.

The temporal boundary for technology-specific and governance materials was 2019 to 10 July 2026, with earlier sources retained when conceptually foundational. Classical works were not restricted by date. Sources were located through publisher platforms, academic databases, backward citation from directly relevant studies, and institutional websites. Inclusion required a direct contribution to at least one analytical dimension: juristic competence, authority, contextual judgment, moral agency, accountability, AI reliability, or governance. Materials concerned only with AI applications unrelated to legal-religious reasoning were excluded.

3.2 Analytical framework

The analysis proceeded through four stages. First, doctrinal sources were read to derive the conditions that distinguish research assistance from authoritative *ijtihad*. Second, the AI-supported legal workflow was functionally decomposed so that different operations could be assessed separately. Third, each operation was subjected to a non-equivalence test: functional capability, epistemic reliability, normative significance, attributable responsibility, and institutional

authorization. Fourth, the resulting distinctions were translated into a governance model. Table 1 specifies the analytical questions.

Table 1. Analytical dimensions for assessing AI-supported *ijtihad*

Dimension	Core question	Indicative evidence	Jurisprudential significance
Source competence	Can the system identify, authenticate, rank, and cite relevant sources?	Provenance, citation accuracy, corpus coverage, source hierarchy	Necessary research support; insufficient for authority
Methodological reasoning	Can it apply recognized interpretive and analogical procedures consistently?	Reproducibility, rule traceability, treatment of conflicting evidence	Supports inference but remains school- and problem-dependent
Contextual judgment	Can it establish facts, custom, consequences, and the questioner's circumstances?	Fact elicitation, domain expertise, sensitivity to time and place	Central to the application of a rule and to fatwa validity
Moral agency	Can it form legally relevant intention and exercise ethically accountable judgment?	Capacity for commitment, conscientious choice, and reason-giving	Constitutive of attributable juristic action
Answerability and authorization	Can it bear *taklif*, accept correction, and be recognized as responsible for the ruling?	Named responsibility, liability, review, institutional mandate	Constitutive of legitimate juristic authority

3.3 Interpretive safeguards and limitations

Three safeguards were used to reduce selective interpretation. First, the study distinguishes descriptive claims about current AI systems from normative claims about what institutions should permit. A future system may change the descriptive assessment without automatically changing the normative structure of authorization. Second, rival considerations were retained. AI's speed, breadth, and consistency were evaluated alongside hallucination, bias, opacity, and automation dependence; *maqasid* was treated as a disciplined juristic mode rather than a free-standing appeal to benefit. Third, conclusions were tested against a negative case: even if an AI output were correct, would correctness alone confer the office, responsibility, and authority of the *mujtahid*? The answer to this question determines whether capability is necessary, sufficient, or merely supportive.

The study has three limitations. It does not empirically benchmark a particular model. It gives primary attention to *Sunni usul* and fatwa doctrine, while using work on Shi'i *ijtihad* to identify shared and divergent concerns [12]. It also analyzes current AI as institutionally deployed technology rather than settling the metaphysical possibility of future machine consciousness. The argument is intentionally narrower: present systems are not recognized bearers of juristic duty, and no plausible performance metric by itself supplies that status.

4. Results and Discussion

4.1 Computational competence is not normative standing

The doctrinal analysis supports a two-part finding. AI can perform a growing number of operations used within juristic research, but performance of those operations does not confer juristic standing. At the source level, a constrained system can search large corpora, retrieve parallel passages, identify recurring terms, locate prior fatwas, and organize conflicting opinions. At the analytical level, it can extract candidate legal causes, compare fact patterns, generate possible analogies, and test the consistency of a draft. These are significant capabilities. They may reduce search costs, expose overlooked evidence, and improve the documentation of collective deliberation.

Their significance must nevertheless be stated precisely. A retrieved passage is not yet an interpretation; a similarity is not yet a legally valid *illah*; a list of consequences is not yet a *maqasid* judgment; and a generated answer is not yet a fatwa. Each transition requires a decision about relevance, hierarchy, meaning, and application. Those decisions are shaped by methodological commitments and by a factual account of the case. Because AI outputs depend on corpus design and statistical association, their apparent comprehensiveness can conceal systematic absence. A corpus may overrepresent contemporary online fatwas, one language, one jurisdiction, or one school while underrepresenting manuscript traditions, minority positions, or tacit institutional practice.

This point changes the standard of evaluation. The relevant question is not whether an AI system produces the same conclusion as a scholar in a sample of cases. Agreement may result from memorization, data dominance, or a shared omission. The stronger test asks whether the system can disclose the evidentiary path, represent material disagreement, identify uncertainty, and remain within its authorized function. In high-stakes domains, a correct answer produced through an unreliable process is not sufficient because the process must support review, correction, and responsibility.

Current generative models are particularly vulnerable to the authority effect of fluent language. Users commonly treat detail and confidence as signals of competence, although the model's linguistic confidence is not calibrated to juristic correctness. A low-probability or minority view may be presented in the same rhetorical register as a consensus position. A fabricated citation may resemble a classical bibliographic entry. Therefore, the first governance requirement is epistemic humility by design: the interface must make the system's derivative and defeasible role visible rather than borrowing the voice of a mufti.

4.2 Intention and moral agency: relevant but not the sole argument

The absence of *niyyah* and moral agency remains an important reason why AI cannot independently perform *ijtihad*. Juristic reasoning is not merely the production of a semantically acceptable proposition. It is an attributable act undertaken for reasons by a person who can orient the act toward a lawful purpose, recognize duties, and respond to criticism. AI systems do not presently form legally cognizable intentions or experience the obligations their outputs purport to state. They optimize generated responses under externally supplied objectives.

The argument should not, however, depend entirely on a contested philosophical claim that machines can never understand or intend. Such a claim is broader than the legal issue and vulnerable to speculative counterexamples. The stronger jurisprudential claim is institutional: no present AI system is a bearer of *taklif*, a member of a juristic profession, or an officeholder authorized to speak as a mufti. It cannot be sanctioned, repent, undertake continuing education, disclose a conflict of interest in the human sense, or accept professional responsibility. Even if future systems display more sophisticated agency-like behavior, a separate normative decision would still be required before a religious-legal community could recognize them as bearers of authority.

This distinction explains why tools can be integrated without anthropomorphism. A concordance does not require intention to assist interpretation. A retrieval model does not require piety to locate evidence. Intention becomes decisive when the system is presented as the author of the normative judgment rather than as an instrument used by an accountable jurist. Thus, the absence of *niyyah* does not prohibit all AI participation; it marks the boundary between instrumental contribution and attributable legal-religious action.

4.3 Context, custom, and the limits of textual abundance

Fatwa practice demonstrates why more text does not automatically produce better judgment. The mufti must often determine what the questioner is actually asking, whether relevant facts have been omitted, which technical expertise is required, and how custom changes the meaning or effect of a transaction. Al-Qarafi's distinction among fatwa, adjudication, and political authority warns against transferring a ruling across institutional contexts without attending to the function in which it operates [22]. Ibn al-Qayyim likewise connected changes in fatwa to changes in time, place, circumstances, intentions, and customs [25]. These are not supplementary details appended after textual reasoning; they condition the legal characterization of the case.

AI can support contextual inquiry by generating structured questions, flagging missing variables, retrieving domain standards, and comparing similar cases. It cannot be presumed to possess the lived and professional understanding required to decide which facts are credible, ethically salient, or socially consequential. In medical matters, for example, factual characterization may require a clinician. In Islamic finance, the nominal contract may differ from the transaction's economic substance. In family disputes, power imbalance or coercion may be invisible in the initial question. An automated answer based on the first prompt can therefore convert incomplete facts into false legal certainty.

Context also includes juristic pluralism. Islamic law contains recognized disagreement within and across schools. A model trained on undifferentiated material can silently combine premises from incompatible methodologies. Responsible systems must preserve rather than flatten this plurality. They should tag the school, genre, date, jurisdiction, and authority level of sources; distinguish a transmitted school position from a contemporary preference; and explain when an answer depends on a disputed principle. The objective is not algorithmic neutrality, which is unattainable, but traceable positionality.

4.4 Maqasid reasoning cannot be reduced to outcome optimization

Maqasid sharia provides a necessary lens for evaluating the purposes and consequences of AI-supported legal practice. AI may advance access to knowledge, consistency of documentation, preservation of institutional memory, and detection of overlooked harm. These benefits can support protection of religion, life, intellect, family, property, and human dignity. Islamic AI-ethics scholarship rightly argues that technological development should be assessed against substantive accounts of welfare and virtue rather than treated as an end in itself [7], [8].

Yet *maqasid* must not become a label attached to an algorithmic utility calculation. The identification of a legitimate benefit, its rank, its relationship to texts, and the distribution of its gains and burdens are juristic questions. A system optimized for efficiency may increase the number of answered questions while weakening intellectual formation, centralizing interpretive power, or excluding underrepresented schools. A system optimized for user satisfaction may return the easiest view rather than the best-supported one. A *maqasid* analysis must therefore examine the practice created by the technology, not only the immediate convenience of the output.

The analysis yields a precautionary rule: the higher the potential impact on protected interests and the more contested the doctrinal pathway, the less authority may be delegated. Low-stakes educational retrieval can permit substantial automation with clear disclaimers. Personalized rulings in medicine, family law, inheritance, or complex finance require independent domain evidence, qualified juristic deliberation, and explicit authorization. This risk-sensitive approach is compatible with contemporary AI governance while remaining grounded in the internal logic of *maqasid*.

4.5 The authority gap

The decisive limitation of autonomous AI *ijtihad* is an authority gap. When an output is wrong, who must explain the error, correct the ruling, notify affected users, and bear responsibility for foreseeable harm? The model cannot perform these duties. Responsibility can easily become distributed among developers, data providers, platform owners, and nominal reviewers until no actor effectively owns the decision. General AI ethics principles do not solve this problem unless an institution allocates duties and enforcement mechanisms [17].

Islamic legal institutions already possess an accountability structure that can be adapted rather than displaced. A named mufti or deliberative council can own the final judgment; subject experts can verify the factual characterization; technical teams can be responsible for system validation and security; and institutional leadership can establish complaint, correction, and disclosure procedures. AI should be located inside this chain as a documented source of assistance. It should never occupy the final position simply because its output is efficient or because a human clicked "approve."

The negative-case test confirms this conclusion. Suppose a system produces a correct ruling with accurate citations and a persuasive explanation. The success shows that the system can contribute useful epistemic work. It does not show that the system formed an accountable intention, acquired juristic office, or became answerable for the ruling. Correctness is necessary for a good fatwa but not sufficient for authority. Conversely, a qualified jurist can make a good-faith error without ceasing to be the responsible author of the fatwa. The asymmetry reveals that authority is a normative relationship, not a reward automatically generated by output quality.

5. The Augmented *ijtihad* Model

5.1 Five layers of participation

The binary question "Can AI perform *ijtihad*?" becomes more tractable when the workflow is divided into five layers. Each layer has a different epistemic function, risk profile, and allocation of authority.

Layer 1: Source retrieval and provenance. AI searches approved corpora, retrieves relevant materials, and supplies verifiable citations. Automation is permissible when every material proposition can be traced to its source and the system distinguishes *Quran*, hadith, classical doctrine, institutional fatwa, statute, and secondary scholarship.

Layer 2: Doctrinal mapping. AI organizes positions by school, period, jurisdiction, and authority level; identifies areas of agreement and disagreement; and detects potentially conflicting evidence. The system may assist substantially, but a scholar must verify classifications and omissions.

Layer 3: Inferential support. AI generates candidate legal characterizations, analogies, counterarguments, and structured drafts. These outputs are hypotheses for deliberation, not findings. Qualified reviewers must reconstruct the inference from the cited materials and reject reasoning that merely resembles an accepted pattern.

Layer 4: Contextual and maqasid evaluation. Human jurists and relevant domain experts establish the facts, assess custom, weigh consequences, and determine how doctrinal options apply. AI may organize evidence or simulate scenarios, but it cannot control the evaluative frame.

Layer 5: Authorization and accountability. A named qualified jurist or authorized collective body adopts, modifies, or rejects the proposed ruling; records the reasons; communicates the answer; and remains responsible for correction. This layer is non-delegable under present conditions.

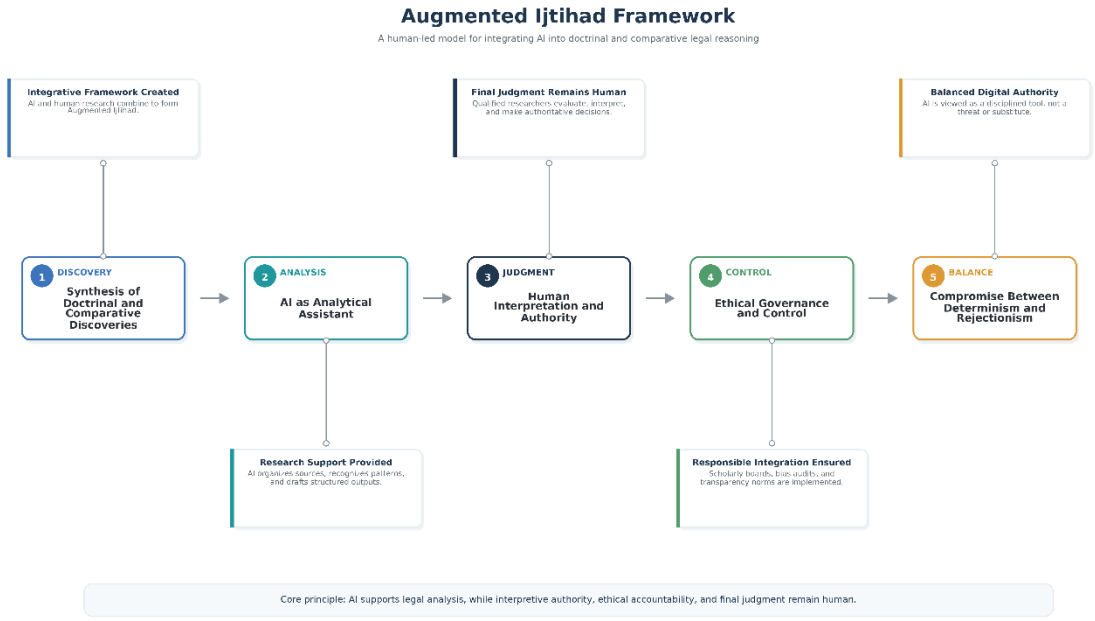


Figure 1. The Augmented *ijtihad* Model: a decreasing gradient of delegation from source retrieval to final authorization.

The model rests on a non-cumulative principle: success at lower layers does not aggregate into authority at the highest layer. Perfect retrieval plus strong doctrinal mapping plus plausible inference remains assistance until a qualified actor performs contextual judgment and accepts responsibility. Table 2 translates this principle into a delegation matrix.

Table 2. Delegation matrix for AI-supported *ijtihad*

Juristic task	Appropriate AI role	Primary risk	Delegation status	Required human control
Corpus search and citation retrieval	Automate within an approved corpus	Fabricated or decontextualized citation	Conditionally delegable	Source verification and corpus governance
Mapping school positions	Classify and compare	Flattening disagreement; school misattribution	Assistive	Jurist validates taxonomy and material omissions
Identifying candidate *illah* or analogy	Generate hypotheses and counterexamples	False equivalence; hidden premise	Restricted assistance	Independent reconstruction of the analogy
Drafting a preliminary memorandum	Structure evidence and arguments	Automation bias; rhetorical overconfidence	Restricted assistance	Line-by-line substantive review
Establishing facts and custom	Generate questions; organize expert evidence	Incomplete facts; privacy; cultural misreading	Human-led	Direct fact finding and relevant professional expertise
Maqasid and consequence assessment	Scenario support	Utility reductionism; distributional blindness	Human-led	Deliberative weighing by qualified jurists and experts
Selecting and issuing the ruling	None beyond documentation support	Illegitimate authority; accountability gap	Non-delegable	Named jurist or authorized council decides and signs
Correction, appeal, and institutional learning	Track cases and flag patterns	Diffusion of responsibility	Human-owned	Formal review, notification, and remediation process

5.2 From "human in the loop" to qualified juristic control

The phrase "human in the loop" is too weak for high-stakes religious-legal decisions. A reviewer may lack the relevant training, receive only the model's summary, face pressure to process a large queue, or be unable to inspect proprietary system behavior. Such a person is present but not in control. Qualified juristic control requires five conditions: competence in the relevant doctrine; access to the underlying sources and system record; sufficient time for independent judgment; power to reject or suspend the output; and personal or institutional attribution of the final decision.

Review intensity should be proportionate to risk. A system that suggests keywords for a literature search requires limited oversight. A system that drafts an inheritance allocation, evaluates the permissibility of a financial product, or responds to a medical-fatwa request requires specialist review and may require collective *ijtihad*. The reviewer should not merely check whether the answer looks plausible. The reviewer must examine the decisive evidence, alternative views, factual assumptions, and foreseeable effects.

Collective review is particularly appropriate where the issue combines legal doctrine with medicine, finance, engineering, data science, or public policy. AI can improve collective *ijtihad* by maintaining an evidence record and exposing disagreements, but it should not collapse expert roles into a single generated synthesis. The jurist remains responsible for legal characterization; the domain expert remains responsible for technical facts; the institution remains responsible for workflow integrity.

6. Governance Requirements for Digital Islamic Jurisprudence

6.1 Corpus provenance and source hierarchy

An AI-supported fatwa system should operate on a governed corpus rather than an indiscriminate collection of internet text. Each source should carry metadata concerning author, title, edition, language, school, genre, date, jurisdiction, and authority level. Translations should be linked to the source language. Hadith material should include the edition and relevant grading information. Institutional fatwas should be distinguished from binding law, judicial decisions, and private scholarly opinions.

The system must show the sources used for each material step, not merely append a generic reading list. Retrieval logs should record the query, returned passages, ranking, model version, and any human modification. If the system cannot retrieve an authoritative basis, it should abstain or refer the matter to a scholar rather than fill the gap with fluent generalization. This requirement responds directly to generative-model confabulation and aligns with the NIST emphasis on information integrity and traceability [16].

6.2 Pluralism, uncertainty, and abstention

Islamic legal disagreement should be represented as structured knowledge, not noise to be averaged away. A system should disclose when a question has multiple recognized answers, identify the methodological basis of each, and avoid presenting majority frequency in the corpus as equivalent to consensus. It should not convert a confidence score into a claim of legal certainty. Model confidence reflects prediction behavior, not the normative weight of evidence.

Abstention is therefore a core design capability. The system should decline to provide a substantive recommendation when facts are incomplete, sources conflict beyond its authorized scope, the issue requires confidential counseling, or the consequences are high. Referral is not system failure; it is evidence that the institution has designed boundaries. A platform that always produces an answer structurally rewards overclaiming.

6.3 Auditability, contestability, and correction

Every AI-assisted ruling should be reproducible at the level necessary for institutional review. The record should include user-supplied facts, clarifying questions, retrieved sources, generated intermediate outputs, human edits, reviewer identity, and the final reasons adopted. Because models and corpora change, the version used must be recorded. Sensitive information should be minimized, protected, and retained only as long as necessary.

Affected users need a route to challenge an answer and request human review. Institutions should publish correction procedures and notify users when a material system error affects prior outputs. Repeated errors should trigger root-cause analysis rather than isolated editing. The objective is not only to identify which sentence was wrong but also to determine whether the cause lay in corpus coverage, retrieval ranking, prompting, model behavior, interface design, or inadequate review.

6.4 Bias, security, and continuing validation

Bias auditing in this domain must extend beyond demographic parity. Relevant questions include whether the corpus disproportionately represents one school or region; whether Arabic, English, and local-language queries receive comparable source quality; whether gendered or minority positions are systematically omitted; and whether contemporary state fatwas are presented as timeless doctrine. Testing should use adversarial questions, ambiguous fact patterns, minority-school positions, fabricated citations, and prompt-injection attempts.

Security is part of jurisprudential integrity. A manipulated corpus or compromised retrieval layer can silently alter religious advice. Access controls, change approval, source checksums, incident response, and independent technical review are therefore not merely information-technology concerns. They protect the reliability of the evidentiary environment on which the jurist depends.

Model updates require revalidation. An institution should not assume that a newer general-purpose model is safer for juristic use. Performance must be tested against a maintained evaluation set covering source accuracy, school attribution, conflicting evidence, contextual elicitation, abstention, and high-risk referral. Table 3 Deployment should be suspended when a material change cannot be evaluated.

Table 3. Minimum governance controls for AI-supported fatwa and *ijtihad* systems

Governance domain	Minimum control	Evidence of implementation
Scope	Written list of permitted, restricted, and prohibited uses	Public system card and internal authorization policy
Corpus	Approved sources with provenance and hierarchy metadata	Corpus register, edition record, change log
Output	Citation-linked reasoning, school labels, uncertainty, and abstention	User interface and stored decision record
Human review	Qualified reviewer with access, time, and rejection authority	Reviewer assignment, sign-off, and reasons
Accountability	Named jurist or council owns the final ruling	Signature, institutional mandate, correction duty
Contestability	Accessible route for appeal and human reconsideration	Complaint protocol and response-time standard
Privacy and security	Data minimization, access control, and incident response	Privacy assessment, security audit, incident log
Bias and pluralism	Testing across schools, languages, regions, and sensitive cases	Evaluation report and remediation record

Governance domain	Minimum control	Evidence of implementation
Change management	Revalidation after model, prompt, or corpus updates	Version register, approval record, rollback plan
Monitoring	Periodic error review and user-harm assessment	Post-deployment audit and public transparency report

These requirements translate the objectives of trustworthy AI into duties recognizable within Islamic legal institutions. Their purpose is not to certify a machine as a *mujtahid* but to preserve the epistemic conditions under which jurists can use complex tools without losing control of the judgment.

7. Theoretical and Practical Implications

7.1 Reframing the debate

The principal theoretical implication is that the debate should move from machine personhood to the institutional allocation of epistemic labor and authority. Asking whether AI "reasons" in the abstract invites an all-or-nothing answer and encourages anthropomorphic comparisons. Asking which operation is performed, how it is validated, and who owns the resulting decision yields a more durable framework. It remains applicable as systems improve because it does not equate new capability with automatic authorization.

This reframing extends Islamic AI-ethics scholarship in a specific direction. *Maqasid*, virtue, pluralism, and trusteeship frameworks identify values by which technology should be judged [7], [8], [9]. The present analysis adds an institutional question: which actor has the right and duty to convert those values into a ruling? It also extends fatwa-and-AI scholarship by specifying the content of the often-repeated recommendation that AI should assist rather than replace the mufti [4]. Assistance is legitimate only when the workflow preserves provenance, independent judgment, and answerability.

The analysis also contributes to philosophy of technology by separating output competence from normative standing. Technologies routinely redistribute cognitive work, but the redistribution of work need not entail the transfer of office. The risk arises when institutions allow the system's scale and rhetorical confidence to redefine what counts as authoritative knowledge. The Augmented *ijtihad* Model resists that drift by making the boundary explicit and auditable.

7.2 Implications for fatwa institutions

Fatwa institutions should begin with workflow mapping rather than purchasing a general-purpose chatbot. They should identify repetitive, evidence-heavy tasks where automation can add value and then prohibit direct public deployment beyond validated scope. A prudent sequence is: internal corpus search; doctrinal comparison; draft memoranda for scholars; controlled pilot use; independent evaluation; and only then limited user-facing functionality. High-stakes personalized questions should route to human consultation.

Institutions should also invest in juristic digital competence. Scholars need enough technical understanding to interrogate retrieval, recognize model failure, and resist automation bias. Technical teams need enough legal-method training to understand why school boundaries, genre, edition, and source hierarchy matter. Governance cannot be outsourced entirely to either group.

The model supports collective *ijtihad*. A shared platform can preserve submissions from physicians, economists, engineers, and jurists; map areas of agreement; record objections; and connect the final ruling to its evidentiary trail. Such use strengthens rather than replaces deliberation. It is most valuable when the technology makes disagreement visible and reviewable instead of producing a premature synthetic answer.

7.3 Implications for developers and regulators

Developers should avoid product language that personifies the system as a mufti, scholar, or autonomous religious authority. Interface design should disclose that outputs are generated assistance, identify the responsible institution, and show when human review has occurred. Benchmarks should test citation fidelity, doctrinal attribution, counter-view coverage, abstention, and referral, not only semantic similarity to a reference answer.

Regulators and accreditation bodies can treat public-facing automated fatwa systems as high-impact advisory systems even when their outputs are formally non-binding. Religious advice can influence marriage, finance, health, social participation, and personal conduct. Risk classification should consider real reliance, not only legal enforceability. Minimum requirements should include accountable sponsorship, documented scope, privacy protection, correction mechanisms, and a prohibition on falsely representing generated content as reviewed by a scholar.

7.4 Future research

The framework requires empirical testing. Future studies should construct multilingual benchmark sets containing clear cases, disputed cases, incomplete facts, conflicting school positions, and adversarial prompts. Evaluation should compare open and proprietary models, retrieval architectures, and review protocols. The outcome measure should not be a single accuracy score. It should include provenance accuracy, issue spotting, representation of disagreement, abstention quality, reviewer workload, and correction performance.

Field research is also necessary. Ethnographic and interview studies can examine how muftis actually use AI, where automation bias enters, how institutional hierarchies affect review, and whether AI changes the intellectual formation of junior scholars. Comparative research should examine Sunni and Shi'i theories of authority, different national fatwa institutions, and minority-Muslim contexts. Finally, user studies should investigate when people treat an automated response as authoritative and which disclosures genuinely reduce misplaced trust.

8. Conclusion

Artificial intelligence can perform valuable parts of the research that surrounds *ijtihad*, but it cannot presently perform *ijtihad* as an independent bearer of Islamic legal authority. The reason is not a simple claim that machines are weak. Current systems can exceed individual scholars in search speed, textual scale, and pattern detection. The decisive point is that juristic authority is constituted by more than performance. It requires disciplined method, contextual judgment, attributable intention, responsibility under a normative order, and authorization within a community of legal practice.

The Augmented *ijtihad* Model makes this distinction operational. AI may retrieve and organize sources, map disagreement, propose candidate analogies, and structure preliminary arguments. Its role must narrow as the process moves toward factual characterization, *maqasid* evaluation, and final judgment. Authorization remains with a named qualified jurist or collective body that can explain, correct, and accept responsibility for the ruling.

Responsible integration therefore depends less on declaring AI "ethical" than on designing institutions that preserve juristic control. Governed corpora, source-level traceability, school-of-law tagging, uncertainty disclosure, meaningful abstention, qualified review, audit trails, contestability, security, and continuous validation are minimum conditions. Under those conditions, AI can deepen access to evidence and strengthen collective deliberation. Without them, fluent automation risks converting technical convenience into counterfeit authority. The future of digital Islamic jurisprudence lies neither in rejecting computation nor in personifying it, but in delegating analytical labor without delegating the moral and legal responsibility that makes *ijtihad* legitimate.

Funding and competing interests: The declarations applicable to any reuse or resubmission of this editorially upgraded manuscript must be confirmed by all authors before submission. The published version should be consulted for disclosures attached to the article of record.

Ethics approval: This normative-doctrinal study did not involve human participants, animals, or personally identifiable research data; ethical approval was not required.

Data availability: No empirical dataset was generated or analyzed. The legal, scholarly, and governance sources used in the analysis are identified in the reference list.

Author contributions: Authorship and author order are preserved from the published article. A contributor-role statement must be confirmed by each author before any reuse or resubmission.

Declaration of Generative AI: Generative AI was used to assist literature mapping, structural revision, language editing, and draft preparation. All claims, interpretations, references, and disclosure statements remain subject to independent verification and approval by the authors.

References

[1] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, pp. 610–623. doi: https://doi.org/10.1145/3442188.3445922.

[2] Z. Ji *et al.*, "Survey of hallucination in natural language generation," *ACM Comput. Surv.*, vol. 55, no. 12, p. 248, doi: https://doi.org/10.1145/3571730.

[3] A. A. Munshi, W. H. AlSabban, A. T. Farag, O. E. Rakha, A. A. Al Sallab, and M. Alotaibi, "Automated Islamic jurisprudence legal opinions generation using artificial intelligence," *Pertanika J. Sci. Technol.*, vol. 30, no. 2, pp. 1135–1156, doi: https://doi.org/10.47836/pjst.30.2.16.

[4] S. Ahmed, "Fatwa and AI: A literature-based assessment on how AI impacts the role of the mufti," *J. Islam. Stud.*, vol. 42, no. 1–2, doi: https://doi.org/10.25159/2957-9163/14400.

[5] A. Priantina, M. M. Uula, Aufa, and E. Herindar, "AI in fatwa formulation: Transforming Sharia-compliant finance," *J. Cent. Bank. Law Institutions*, vol. 4, no. 3, pp. 595–634, doi: https://doi.org/10.21098/jcli.v4i3.446.

[6] A. Hilmi, M. Najib, H. Bisri, and H. Suhendar, "Methodology of Islamic Economic Fatwas Issuance: Comparing Egypt's Dar al-Ifta' and Indonesia's National Sharia Board," *J. Huk. Islam*, vol. 23, no. 2, 2025, doi: https://doi.org/10.28918/jhi.v23i2.08.

[7] A. Raquib, B. Channa, T. Zubair, and J. Qadir, "Islamic virtue-based ethics for artificial intelligence," *Discov. Artif. Intell.*, vol. 2, p. 11, doi: https://doi.org/10.1007/s44163-022-00028-2.

[8] E. Elmahjub, "Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI," *Philos. Technol.*, vol. 36, p. 73, doi: https://doi.org/10.1007/s13347-023-00668-x.

[9] F. Ali, K. Bouzoubaa, F. Gelli, B. Hamzi, and S. Khan, "Islamic ethics and AI: An evaluation of existing approaches to AI using trusteeship ethics," *Philos. Technol.*, vol. 38, p. 120, doi: https://doi.org/10.1007/s13347-025-00922-4.

[10] M. G. Abdelnour, "Artificial intelligence and the Islamic theology of technology: From 'means' to 'meanings' and from 'minds' to "hearts," *Religions*, vol. 16, no. 6, p. 796, doi: https://doi.org/10.3390/rel16060796.

[11] A. Polat, S. Khan, and U. Bashir, "An inquiry into the application of artificial intelligence on

fatwa," in *Digital transformation in Islamic finance: A critical and analytical view*, Y. Suzuki and M. D. Miah, Eds., Routledge, pp. 273–286. doi: https://doi.org/10.4324/9781003262169-18.

[12] H. Latifi, "Challenges of using artificial intelligence in the process of Shi'i *Ijtihad*," *Religions*, vol. 15, no. 5, p. 541, doi: https://doi.org/10.3390/rel15050541.

[13] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 389–399, doi: https://doi.org/10.1038/s42256-019-0088-2.

[14] U.N.E.S.C.O., *Recommendation on the ethics of artificial intelligence*. UNESCO. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137

[15] E. Tabassi, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," *National Institute of Standards and Technology*. doi: https://doi.org/10.6028/NIST.AI.100-1.

[16] C. Autio *et al.*, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," *National Institute of Standards and Technology*. doi: https://doi.org/10.6028/NIST.AI.600-1.

[17] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nat. Mach. Intell.*, vol. 1, pp. 501–507, doi: https://doi.org/10.1038/s42256-019-0114-4.

[18] M. H. Kamali, *Principles of Islamic jurisprudence*, 3rd ed. Islamic Texts Society. doi: https://doi.org/10.22219/jes.v2i2.5109.

[19] W. B. Hallaq, *Authority, continuity, and change in Islamic law*. Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511495557.

[20] A. H. Al-Ghazali, "Al-Mustasfa min ilm al-usul." [Online]. Available: https://archive.org/details/in.ernet.dli.2015.432464/mode/2up

[21] D. P. Muhammad Masud, Brinkley Messick, *Islamic legal interpretation: Muftis and their fatwas*. Harvard University Press. [Online]. Available: https://nes.princeton.edu/publications/islamic-legal-interpretation-muftis-and-their-fatwas

[22] S. D. Al-Qarafi, "Al-Ihkam fi tamyiz al-fatawa an al-ahkam wa tasarrufat al-qadi wa al-imam." [Online]. Available: https://archive.org/details/McGillLibrary-115354-522/page/n1/mode/2up

[23] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, pp. 59–68. doi: https://doi.org/10.1145/3287560.3287598.

[24] T. Hutchinson and N. Duncan, "Defining and describing what we do: Doctrinal legal research," *Deakin Law Rev.*, vol. 17, no. 1, pp. 83–119, doi: https://doi.org/10.21153/dlr2012vol17no1art70.

[25] I. al-Jawziyyah, "I'lam al-muwaqqi'in an rabb al-alamin," *Dar al-Kutub al-Ilmiyyah*. [Online]. Available: https://ia600700.us.archive.org/23/items/Ibnu_Qayyim/Panduan Hukum Islam (I'lamul Muwaqi'in) I-IV.pdf