Islamic Digital Religion and Intelligent Media

Vol. 1, No. 1 (October 2026), 87-108
P-ISSN: XXXX-XXXX | E-ISSN: XXXX-XXXX
DOI: https://doi.org/10.xxxxx/drim.vXiX.XX

The Missing Madhhab: Generative AI and Religious Authority among Indonesian Muslim Users

OPEN ACCESS

Nurbaiti[1*], Aas Siti Sholichah[1], Muhammad Arinal Haq[2]

[1] Universitas PTIQ Jakarta, Indonesia
[2] University of Jordan, Amman, Jordan

[*]Correspondence: nurbaiti@ptiq.ac.id

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

Abstract. Generative artificial intelligence now answers everyday questions of Islamic law for Muslim users who once brought those questions to a teacher or a study circle. This study asks how such systems compress juristic plurality into what appears to be a single Islamic ruling, and under what conditions users recover the disagreement that has been removed. A multi-site qualitative design was used across four Indonesian urban clusters between May and July 2026. Thirty informants were interviewed, comprising twenty-two Muslim users aged eighteen to thirty-four with varying madhhab literacy and eight experts in fiqh, usul al-fiqh, fatwa, or pesantren education. Semi-structured interviews of fifty to eighty minutes were analysed through codebook thematic analysis with constant comparison, negative-case analysis, and reflexive memoing, yielding twenty-four codes and six themes. The analysis identifies a one-answer pathway in which single-answer formatting and fluent declarative wording invite users to infer doctrinal consensus, after which the expectation of disagreement recedes. Plurality proved recoverable but prompt-dependent, so that users who already knew what to ask obtained comparative answers while novices did not know what was missing. Verification was stratified by perceived stakes and by religious literacy, and human scholars re-entered as corrective infrastructure rather than being displaced. Negative cases show that homogenisation is contingent on interface defaults rather than technologically inevitable, which locates intervention in answer design and in teaching users to recognise when one answer conceals several. The title's framing of madhhab plurality as missing should be read in this qualified sense: the underlying juristic material was rarely absent from the system itself, and the problem identified here is the single-answer default presentation and the prompt-dependence of recovery, rather than an absence of the madhhab material from generative systems as such.

Keywords: generative artificial intelligence; ikhtilaf; madhhab literacy; religious authority; qualitative multi-site study

Introduction

A Muslim student in Makassar opens a chat window and asks whether nail polish invalidates ablution, and the reply arrives in under a second, in confident and well-ordered prose, carrying no indication that jurists have ever disagreed about it. The same question put to a teacher in a mosque study circle would ordinarily produce a different kind of exchange, one in which the answer is attached to a school of law, to a chain of reasoning, and to a person who can be asked a follow-up question. Generative systems have entered this space very quickly and are now consulted for worship, family matters, transactions, and the ordinary permissibility questions

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

of daily life. Recent work on Islamic religious authority documents how digital platforms rearrange who may speak for the tradition without dissolving institutional control altogether (Akmaliah, 2022; Raya, 2025). What generative systems add to that rearrangement is a speaker that produces new religious text on demand, in a single voice, and without the credentials that Islamic tradition has long required of anyone who issues a ruling.

The difficulty this creates is not reducible to factual error. Evaluation studies of large language models answering Islamic questions report uneven accuracy, unverifiable citations, and positions that shift across models and across the language of the prompt (Abdelaal et al., 2026; Bouchekif et al., 2025; Wahid, 2025). Yet a system can be propositionally correct within one school of law and still mislead, because the answer arrives stripped of the attribution that would tell a reader which school it belongs to. Islamic legal tradition treats disagreement among qualified jurists as an ordinary and legitimate feature of the law rather than as a defect to be resolved (Abu Baedah et al., 2026). An interface that returns one practical ruling per question is therefore not neutral with respect to that tradition, whatever the accuracy of the ruling it returns.

Scholarship on digital religion anticipated part of this problem well before generative models arrived. Two decades of work have argued that religious authority is renegotiated rather than abolished when religious life moves into networked media, and that legitimacy is contested through visibility, platform affordance, and institutional recognition (Campbell & Tsuria, 2022; Kołodziejska et al., 2023). Studies of Muslim publics have traced how influencers, preachers, and ordinary users reconfigure the sources they treat as authoritative (Whyte, 2022; Zaid et al., 2022). Indonesian research has shown that digital platforms make internal disagreement visible without displacing established organisations (Akmaliah, 2022), and that young Muslims assemble religious guidance across several overlapping media at once (Fakhruroji, 2025; Hefner, 2022). This literature explains how authority circulates, but it was built to describe the circulation of human speech rather than the generation of new religious text by a machine (Ramadhan et al., 2026).

A second and much larger literature approaches the same systems from the side of technical reliability. Surveys of hallucination in language generation establish that fluent output and factual grounding come apart in predictable ways (Huang et al., 2025; Ji et al., 2023), and work on generative search engines finds that citation and verifiability often fail to keep pace with the confidence of the prose (Liu et al., 2023). Studies of human reliance show that users lean on machine output more heavily when it is expressed without hedging, and that expressed uncertainty measurably changes how far users depend on an answer (Kim et al., 2024; Zhou et al., 2024). Research on automation bias and overreliance has developed interventions that slow users down at the moment of acceptance (Buçinca et al., 2021; Romeo & Conti, 2026; Vasconcelos et al., 2023). This work measures whether an answer is correct and whether users lean on it too readily, but it does not ask whether a community would accept the answer as authoritative.

These two literatures rarely address one another, and the gap between them is where the present study is located. More specifically, neither literature alone can answer whether a judgement that a generated fiqh answer looks correct is the same judgement a user needs in order to recognise that a position is contested: benchmark work on sectarian preference asks whether a model can be shown to favour one school, not whether a user infers consensus from the answer's form, and the reliance literature on hedging and fluency asks whether users over-trust an answer's confidence, not whether that confidence specifically conceals juristic plurality. The present study is positioned to answer this combined question because it examines how users themselves interpret the compressed form of a religious answer, which is the empirical object neither literature targets on its own. Benchmark work on Islamic content has begun to measure exactly the phenomenon at issue, since recent evaluations detect latent

sectarian preference and test whether models identify the school behind a position (Abdelaal et al., 2026; Maheshwari et al., 2026). Normative scholarship has separately asked whether a machine can perform ijtihad and what happens to the office of the mufti when rulings are generated rather than issued (A. Ahmed et al., 2026; S. Ahmed, 2024; Fazal, 2026). What neither body of work supplies is evidence from the users themselves about how a single generated answer comes to be read as the position of Islam, and about what interrupts that reading when it occurs. Indonesian studies of digital fatwa have flagged the proliferation of guidance detached from juristic method (Ali & Aljahsh, 2025), yet the interactional mechanism behind that detachment remains undocumented (Kurniawan et al., 2025).

Three further considerations make the Indonesian setting analytically useful for this question. Indonesia holds the largest Muslim population in the world and sustains a dense ecology of pesantren, mass organisations, and local teaching traditions whose fiqh is specific rather than generic, which means that a decontextualised ruling has something concrete to be decontextualised from. Generative tools have been adopted rapidly among Indonesian students and young professionals, so the encounter under study is ordinary rather than exotic (Kasneci et al., 2023). Work on Indonesian religious authority also shows that the relationship between online guidance and institutional teaching is actively negotiated rather than settled (Faizin et al., 2025; Raya, 2025), which makes it possible to observe users in the act of negotiating it.

This study therefore examines how Muslim users interpret AI-generated fiqh responses that omit or minimise madhhab distinctions, through what interactional mechanisms a single answer comes to be perceived as doctrinal consensus, what forms of literacy and verification resist that perception, and how Islamic scholars read the implications for ikhtilaf and religious authority. The unit of analysis is the episode in which a user brings one concrete fiqh question to a generative system and decides what to do with the answer. Evidence comes from thirty informants across four Indonesian urban clusters, drawn so that low, medium, and high madhhab literacy are all represented among users, and so that expert judgement can be compared against user experience of the same phenomenon.

The remainder of the article proceeds in the conventional order. The method section sets out the multi-site design, the sampling logic, the interview instrument, the analytic procedure, and the measures taken to establish trustworthiness. The results section reports six themes together with the coding framework and the point at which new codes ceased to appear. The discussion interprets those themes against the digital religion and machine reliability literatures and proposes the one-answer pathway as a way of joining them. The conclusion states what the study establishes and what should follow from it for answer design and for religious education.

Method

Research Design and Position of the Researchers

This study adopted a multi-site qualitative design because the object of inquiry is an interpretive process rather than a measurable quantity. The question is how a generated ruling acquires standing in a user's religious reasoning, and that acquisition happens in accounts, justifications, and recollected episodes that only interview material can reach. Multi-site designs are appropriate where a phenomenon is expected to appear in recognisably similar form across settings that nonetheless differ in institutional texture (Bayuo et al., 2025). Four Indonesian urban clusters were selected for that reason, and the design retained divergent and negative cases from the outset so that the resulting account would remain explanatory rather than deterministic. The research team combined training in Islamic legal studies with training in digital religion research, and this combination shaped both the interview guide and the coding. Reflexive memos recorded where the team's own madhhab commitments might have inclined

an interpretation, following current guidance on reflexivity in qualitative work (Olmos-Vega et al., 2023). No member of the team held a teaching or supervisory relationship with any informant.

Sites, Informants, and Sampling

Fieldwork was conducted between May and July 2026 in Jakarta and Depok, in Yogyakarta, in Surabaya and Sidoarjo, and in Makassar. The clusters were chosen to vary the institutional environment rather than to represent Indonesia statistically. Jakarta and Depok offer urban diversity with mixed institutional reference points, Yogyakarta combines high educational diversity with the visible coexistence of several religious orientations, Surabaya and Sidoarjo carry a strong pesantren ecology with its own fiqh pedagogy, and Makassar provides a comparison outside Java with different urban religious networks. Thirty informants were recruited in total, comprising twenty-two Muslim users of generative systems and eight expert informants, with the distribution across clusters shown in Table 1.

Table 1. Informant composition and analytic purpose of each site cluster

Cluster	Users	Experts	Total	Analytic purpose
Jakarta–Depok	6	2	8	Urban diversity and mixed institutional reference points.
Yogyakarta	6	2	8	High educational diversity and visible coexistence of religious orientations.
Surabaya–Sidoarjo	5	2	7	Strong pesantren ecology and local fiqh pedagogy.
Makassar	5	2	7	Non-Java comparison and varied urban religious networks.
All clusters	22	8	30	Thirty informants in total

Sampling proceeded in two stages that served different purposes. Maximum-variation purposive sampling was used first, so that the user group would span madhhab literacy, educational background, occupation, and frequency of use, an approach suited to studies that aim for conceptual range rather than statistical representation (Bouncken et al., 2026). Theoretical sampling followed once provisional categories had formed, and later recruitment deliberately sought informants whose accounts might contradict the emerging pattern. How many of the thirty informants were recruited specifically during this later, deliberately disconfirming-case-seeking phase, as distinct from the initial maximum-variation sampling, is not reported here; this bears on how the negative cases reported under the counter-homogenisation theme should be weighted, since a negative case that was deliberately sought carries a different interpretive status than one that emerged incidentally within a maximum-variation sample. Users were eligible if they were Muslim adults aged eighteen to thirty-five, had used a generative system for Islamic questions at least five times in the previous six months, and could recount one concrete fiqh episode in detail. Experts were eligible with at least five years of work in fiqh, usul al-fiqh, Islamic law, fatwa, pesantren education, or digital religious education. The realised user sample ranged in age from eighteen to thirty-four and was evenly divided by gender, with seven users of low madhhab literacy, eight of medium literacy, and seven of high literacy. The procedure used to classify informants into these three literacy levels, whether a structured instrument, a brief screening interview, self-report, or researcher judgement made during recruitment, is not described here; because madhhab literacy functions as a central stratifying variable in the analysis, this is noted as a limitation on how firmly the literacy-stratified claims can be read. Expert informants ranged from thirty-four to fifty-two years of age and included lecturers in fiqh and Islamic family law, a pesantren teacher, a

professor of usul al-fiqh, a pesantren curriculum coordinator, a fatwa researcher, and a digital da'wah educator.

Instrument and Data Collection

Data were gathered through semi-structured interviews of fifty to eighty minutes, conducted in Indonesian and audio-recorded with consent. The guide contained ten open questions and was built so that abstract judgements would be anchored in one concrete episode rather than in general opinion. Informants first described their usage ecology, including frequency, device, platform, and motivation. They were then asked to reconstruct a specific fiqh question they had put to a generative system, together with the wording of the prompt and the shape of the response. Subsequent questions addressed how certain the answer felt and why, whether any school of law or disagreement was named, what the informant did before accepting or acting on the answer, and whether the experience had changed how they think about difference in fiqh. A live prompt task followed, in which informants were asked what happens when they request a comparison across schools, and this task produced observational material alongside recollection. This close correspondence between the guide's structure and the one-answer pathway that later organises the findings is more accurately read as a feature of the semi-structured design, in which a mechanism already suspected by the research team was probed directly through targeted follow-up questions, rather than as an independently emergent discovery arrived at without that prior direction. The guide closed with questions about conflict with local practice, about the circumstances under which a human scholar is preferred, and about any experience that ran against the idea that these systems simplify. The anchor topics informants chose were themselves varied, ranging from qunut in the dawn prayer and the units of Tarawih to professional zakat, cryptocurrency trading, marriage guardianship, and the determination of prayer times during flights.

Analytic Procedure

Recordings were transcribed in full and analysed through codebook thematic analysis, a form of pattern-based analysis that retains an explicit and revisable coding structure while treating interpretation as an active practice of the analyst (Braun & Clarke, 2021a, 2021c). This analysis follows the codebook tradition specifically, including double-coding a subset for calibration; Braun and Clarke's own reflexive-TA writings, cited here for the point about interpretation as an active analytic practice, are among the sources that distinguish reflexive thematic analysis from codebook or coding-reliability approaches precisely because the two rest on different assumptions about procedures like inter-coder calibration. The citation here is accordingly to Braun and Clarke's general point about analyst interpretation rather than an endorsement of reflexive TA's position on calibration, and the codebook approach with its calibration check is the tradition actually followed throughout. Coding began inductively on the first batch of interviews and generated eighteen first-order codes. Constant comparison was then applied across informant groups, across sites, and across literacy levels, so that each new account was read against the categories already formed. The codebook was consolidated to twenty-four codes organised under six themes, and every addition, merger, and exclusion was recorded in a memo. A quarter of the transcripts were double-coded for calibration, and disagreements were discussed and resolved by returning to the transcript rather than by treating agreement as evidence of correctness. The number of coding disagreements that arose within this double-coded quarter, and their distribution across the six themes, is not separately reported here, nor is an inter-rater agreement statistic such as a percentage or kappa figure for the double-coded subset; this is noted as a limitation, since a theme that generated more disagreement during calibration would warrant more caution in how firmly its boundaries are drawn than a theme on which coders agreed readily. Negative-case analysis ran throughout and was treated as a test

of the emerging claim rather than as a residual category. Analysis was conducted on the Indonesian transcripts, and only the excerpts selected for presentation were translated, with the original wording retained alongside the translation so that readers can inspect the basis of each interpretation. Whether these translated excerpts were checked by a second bilingual team member or back-translated is not reported here; this is noted as a limitation, since translation choices, for example in rendering modal or evidential particles in the Indonesian quotations, can affect how certain or hedged an informant's statement appears in English.

The study did not treat a fixed number of interviews as a threshold of adequacy. Recent methodological work argues that saturation claims should rest on the interpretive yield of further data rather than on an arithmetic rule (Braun & Clarke, 2021b; Hennink & Kaiser, 2022). Code emergence was therefore tracked batch by batch and is reported in Table 2. Summing the new codes reported batch by batch in Table 2 (18 + 6 + 2 + 1 + 0) gives twenty-seven first-order codes, three more than the twenty-four codes reported in the final codebook in Table 3. This difference reflects codes that were later merged or dropped during consolidation; the manuscript does not separately report which codes these were or whether the zero-new-code stopping point in the final batch was assessed against the raw or the already-consolidated code list, which is noted here as an unreported detail rather than resolved with figures not available. The final three interviews produced no new first-order code and served instead to elaborate existing categories and to confirm the negative cases already identified, which is the conceptual stopping point the team adopted. Figure 1 summarises the full sequence from site selection through to theme construction.

Table 2. First-order code emergence across interview batches

Batch	Interviews	New codes	Interpretive note
Interviews 1–8	8	18	Initial code generation; strong emergence of single-answer and verification patterns.
Interviews 9–16	8	6	New codes mainly concerned local-practice mismatch, prompt dependency, and institutional invisibility.
Interviews 17–24	8	2	Only two materially new codes: recovered plurality and accountable authority.
Interviews 25–27	3	1	One negative-case code added: AI can foreground plurality when explicitly framed.
Interviews 28–30	3	0	No new first-order code; interviews elaborated existing categories and confirmed negative cases.

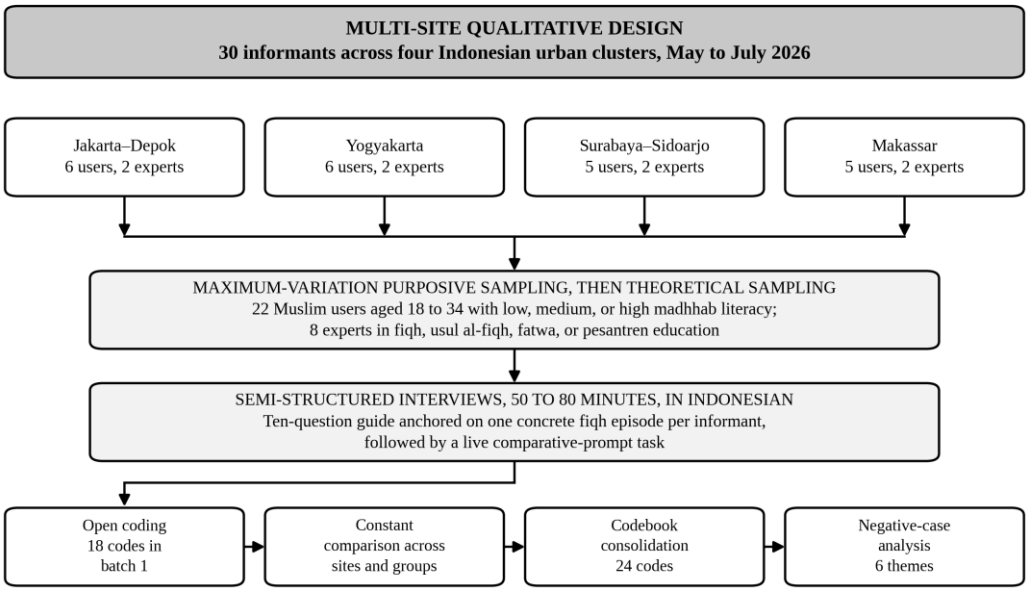


Figure 1. Multi-site design and analytic sequence from site selection to theme construction

Trustworthiness and Ethics

Credibility rests on maximum-variation sampling across four locations and two informant groups, and on the deliberate presentation of convergent and divergent accounts side by side. Dependability rests on the iterative codebook with its memoed revisions and on the double-coded subset. Confirmability rests on reflexive memoing and on an audit trail that separates what an informant claimed, what the analyst interpreted, and what the study infers theoretically, a separation that recent syntheses treat as the core of trustworthy qualitative reporting (S. K. Ahmed, 2024). A single worked example illustrates this separation using material already presented above: an informant's account that a confident, declarative tone made an answer feel like settled law is the claim; the observation that the same formulation recurred nearly verbatim across informants in different cities is the analyst's interpretation that the impression is produced by the interface rather than by any particular question; and the conclusion that single-answer formatting therefore functions as a distinct causal mechanism, named here as the one-answer pathway, is the theoretical inference drawn from that interpretation. This manuscript does not otherwise set out such an example explicitly. Transferability is supported through thick description of location, literacy, and use context, and the study does not extend its Indonesian findings to Muslim populations elsewhere. Preliminary theme summaries were returned to a subset of informants for comment, and those comments were treated as further data rather than as verification of the analysis, in line with current argument about the limits of member checking (Motulsky, 2021). Reporting follows the Consolidated Criteria for Reporting Qualitative Research (Tong et al., 2007). Regarding the subset of informants to whom preliminary theme summaries were returned for comment described above, its size and composition, including how many user versus expert informants it comprised and across which literacy levels, is not reported here; because member-checking responses were treated as further analytic data rather than as verification, the subset's composition could itself have shaped which theme refinements or negative cases surfaced at this stage, which is noted as an unreported detail bearing on the analysis. Participation was voluntary and informed, no incentive was offered, and informants are identified throughout by coded labels in the ranges U01 to U22 and E01 to E08 so that no personal identifier appears in the analysis or in the quoted material.

Results

The thirty interviews produced twenty-four first-order codes organised under six themes, set out in full in Table 3. The account below moves from the mechanism that produces an apparent consensus, through the conditions under which plurality returns, to the ways in which human religious authority re-enters the process. Excerpts are given in the original Indonesian followed by an English rendering, and each is attributed to a coded informant.

Table 3. Coding framework: twenty-four first-order codes under six themes

Code	Label	Definition	Theme
C01	Fluent certainty	AI wording appears polished, confident, or final	T1
C02	Single-answer default	Initial response gives one practical ruling	T1
C03	Consensus inference	User interprets one answer as "the Islamic ruling"	T1
C04	Convenience shortcut	Speed/ease reduces motivation to consult plural sources	T1
C05	Madhhab omission	School identity absent from initial response	T2
C06	Plurality-on-request	Multiple schools appear after a follow-up prompt	T2
C07	Prompt dependency	Output diversity depends on how user frames the question	T2
C08	Novice query gap	User does not know what distinctions to ask for	T2
C09	Local-practice mismatch	AI answer differs from community/pesantren practice	T3
C10	Institutional invisibility	Local religious institutions/sources not named	T3
C11	Global-generic framing	Answer feels universally applicable or detached from place	T3
C12	Valid-but-decontextualized	Answer may be valid in one school but context is hidden	T3
C13	Low-stakes acceptance	User accepts routine guidance without verification	T4
C14	High-stakes escalation	User seeks human/institutional confirmation for consequential issues	T4
C15	Literacy-based skepticism	Religious education triggers caution toward one-answer outputs	T4
C16	Source triangulation	User compares AI with other sources or institutions	T4
C17	AI as first stop	AI is used for initial orientation, terminology, or quick map	T5
C18	Human confirmation	Teacher/scholar consulted after AI	T5
C19	Accountable authority	Human authority valued for responsibility/context	T5
C20	AI not mufti	Explicit boundary that AI should not issue final religious judgment	T5
C21	Comparative prompting	User asks for multiple madhhab positions	T6
C22	Source-demanding prompt	User requests dalil, references, provenance	T6
C23	Recovered plurality	Follow-up interaction restores awareness of disagreement	T6
C24	Negative case	Evidence that AI initially presents plurality or supports ikhtilaf literacy	T6

The One-Answer Effect

Users with low and medium madhhab literacy consistently described the first response as a polished and declarative statement that gave no sign of juristic disagreement. The mechanism they described was not misinformation of any kind. A position that is defensible within one school was formatted as a universal rule, and the confidence of the formatting did the work that an explicit claim of consensus would otherwise have done. Several informants reported that the thought of an alternative position simply did not arise until later.

> "Karena bahasanya rapi dan yakin, kesan pertama saya adalah memang itulah hukum Islamnya. Saya tidak langsung terpikir bahwa bisa ada pendapat lain." (U02)
>
> *Because the language was neat and confident, my first impression was that this simply was the Islamic ruling. It did not immediately occur to me that there could be another position.*

The same formulation recurred almost verbatim from informants in different cities, which suggests that the impression is produced by the interface rather than by any particular question.

Expert informants named the same mechanism but separated two things that users tend to run together. An answer may be propositionally accurate and still fail a different test, because accuracy concerns whether a statement is defensible while epistemic integrity concerns whether the answer discloses its own standing.

> "Ketika sistem tidak menandai status khilafiyah, tidak memberi atribusi, dan menulis dengan nada deklaratif, pengguna mudah mengira kesimpulan itu mewakili Islam secara umum. Kesan konsensus terbentuk pada level penyajian sebelum pengguna memeriksa isi lebih dalam." (E04)
>
> *When the system does not flag that an issue is disputed, gives no attribution, and writes in a declarative tone, users readily assume the conclusion represents Islam in general. The impression of consensus is formed at the level of presentation, before the user examines the content more deeply.*

This last observation locates the phenomenon with unusual precision. The impression of consensus forms at the level of presentation, which means it forms before any question of content has been opened. Convenience reinforced the effect, since informants who valued speed reported little motivation to seek a second source once a usable answer had appeared.

Plurality Becomes Prompt-Dependent

Juristic plurality was rarely absent from the underlying system, but it was conditional on how the user framed the request. Informants who asked explicitly for the four Sunni schools, for evidence, or for comparative reasoning described answers that changed character entirely. This produced an asymmetry that runs against the ordinary assumption that better tools serve less expert users better.

> "Ketika saya menulis 'menurut empat mazhab', jawabannya berubah total dan menjadi lebih banyak. Itu membuat saya sadar bahwa jawaban awal sebenarnya hanya salah satu cara menyederhanakan." (U12)
>
> *When I wrote 'according to the four schools', the answer changed completely and became much fuller. That made me realise the initial answer was only one way of simplifying.*

Users with prior madhhab literacy knew which distinctions to request, whereas novices did not know that anything had been left out. An expert informant pressed this point into a criticism of prompting as a remedy, since a solution that depends on the user already knowing the shape of the missing material cannot help the users who most need it.

> "Prompt dapat menjadi mekanisme koreksi. Ketika pengguna meminta 'bandingkan empat mazhab, sebutkan sumber dan titik perbedaan', sistem biasanya membuka lebih banyak kemungkinan. Tetapi menjadikan prompt sebagai solusi utama memindahkan beban ke pengguna yang mungkin justru tidak tahu apa yang hilang." (E05)
>
> *Prompting can act as a corrective mechanism. When a user asks to compare the four schools and to name the sources and the points of difference, the system usually opens up more possibilities. But making the prompt the main solution shifts the burden onto the user who may be precisely the one who does not know what is missing.*

The live prompt task carried out during the interviews made the asymmetry observable rather than merely reported. Informants were asked to put a comparative request to the system while the interviewer watched, and the difference between a bare question and a request naming the four schools was visible within a single exchange. Low-literacy informants were able to perform the comparative prompt once it had been dictated to them, and several expressed surprise at the volume of material that appeared. What they could not do without help was arrive at the wording themselves, and two of them said afterwards that they would not have known the request was available to make. Whether this researcher-observed live-task material was coded under the same codes as retrospective, recalled-episode material, or tracked

separately, is not stated here; this bears in particular on how confidently the novice query gap code (C08) is evidenced, since observed behaviour under researcher prompting and self-report of an unobserved past episode are different kinds of evidence, and the manuscript does not report whether either code rests disproportionately on one type over the other.

Decontextualisation of Local Fiqh

A third pattern concerns the loss of place. Practices associated with pesantren teaching, with local scholars, or with organisational guidance were rarely visible in an answer unless the user had written them into the prompt. Informants were careful about what they were claiming here, and several insisted that the problem was not that the system had said something false.

> "Saya pernah menemukan jawaban AI yang berbeda dari praktik yang biasa diajarkan di lingkungan pesantren. Setelah ditanya lagi, ternyata ada dasar mazhab lain. Jadi yang hilang bukan selalu kebenarannya, tetapi konteks kenapa komunitas kami memilih satu pendapat." (U03)
>
> *I once found an AI answer that differed from the practice usually taught in the pesantren environment. When I asked further, it turned out there was a basis in another school. So what went missing was not always the truth of it, but the context of why our community chose one position.*

The analysis gives this pattern the name valid-but-decontextualised guidance. An answer can be defensible inside one juristic tradition while appearing falsely universal once it has been detached from that tradition and from the institutions that maintain it. What the user loses is not the ruling but the reason a particular community holds it, and that reason is precisely what makes the ruling intelligible as something other than an arbitrary preference (Mahendra et al., 2026).

Verification Is Stratified by Stakes and by Literacy

Verification behaviour divided along two axes at once. Questions perceived as low in consequence were commonly accepted as convenient orientation with no further checking, while family, financial, or legally consequential matters triggered escalation to a scholar or an institution. Cutting across this, informants with pesantren or formal Islamic studies backgrounds showed markedly stronger scepticism toward single-answer output and were more likely to demand evidence and school attribution before doing anything with the answer.

> "Saya biasanya meminta sumber, posisi mazhab, dan alasan. Setelah itu saya bandingkan dengan kitab, situs lembaga keagamaan, atau guru yang saya percaya, terutama sebelum menjadikannya dasar tindakan." (U15)
>
> *I usually ask for the source, the position of the school, and the reasoning. After that I compare it against the classical texts, the websites of religious institutions, or a teacher I trust, especially before I make it the basis of an action.*

Homogenisation was therefore most pronounced where low stakes, high convenience, and low madhhab literacy occurred together, and it was weakest where any one of those conditions failed. An expert informant refused the suggestion that an accurate answer settles the matter, and in doing so restated the distinction that runs through the whole dataset.

> "Tidak. Sebuah jawaban bisa saja sesuai dengan satu pendapat yang valid tetapi tetap problematis jika dipresentasikan tanpa atribusi sehingga pluralitas hilang. Jadi akurasi proposisional berbeda dari integritas epistemik dan keterbukaan terhadap ikhtilaf." (E06)
>
> *No. An answer may accord with one valid position and still be problematic if it is presented without attribution, so that plurality disappears. Propositional accuracy is therefore different from epistemic integrity and openness to ikhtilaf.*

Human Authority Re-enters as Corrective Infrastructure

Human religious authority did not recede in these accounts. It re-entered at specific points, namely where guidance met ambiguity, where it collided with local practice, or where a decision carried consequences for other people. Users positioned the system as a first stop, a map, or an aid to vocabulary, while scholars were valued for asking about circumstances, for carrying responsibility, and for being able to explain why a disagreement exists at all (Noh et al., 2026).

> "Saya nyaman memakai AI untuk orientasi awal, tetapi untuk keputusan yang punya konsekuensi ibadah atau keluarga saya masih lebih tenang jika ada konfirmasi manusia." (U05)
>
> *I am comfortable using AI for initial orientation, but for decisions that carry consequences for worship or for the family I am still more at ease if there is human confirmation.*

The resulting arrangement is hybrid rather than substitutive. An expert informant went further and described a use in which the system supports rather than erodes the tradition, since a generative tool can introduce the vocabulary of ikhtilaf to someone who has never encountered it. That benefit was conditional, and the condition was that the space for comparison be opened deliberately by the system or by the user.

Counter-Homogenisation through Prompt Literacy

Negative cases qualified the central claim and were retained for that purpose. Several high-literacy users and expert informants reported that a generative system could display multiple schools quickly when asked to, in some cases making comparative fiqh more accessible than an ordinary web search would. These accounts locate the problem in the default shape of the answer rather than in the technology as such.

> "Saya tidak melihat AI selalu menghomogenkan. Jika pertanyaan dibuat spesifik, ia justru bisa memperlihatkan perbedaan mazhab lebih cepat daripada pencarian biasa. Jadi masalahnya ada pada desain jawaban default dan literasi pengguna, bukan semata-mata pada teknologinya." (U10)
>
> *I do not see AI as always homogenising. If the question is made specific, it can actually show the differences between schools faster than an ordinary search. So the problem lies in the design of the default answer and in user literacy, not in the technology by itself.*

Expert informants translated the same observation into a specification for answer design, and one of them set out what a religiously adequate reply would have to do. The list is instructive because every item is a disclosure requirement rather than a constraint on content.

> "Jawaban sebaiknya menandai jika isu diperselisihkan, menyebutkan posisi mazhab atau otoritas, memberi sumber yang dapat diverifikasi, menyatakan ketidakpastian, dan mendorong konsultasi manusia ketika risikonya tinggi. Pendidikan pengguna perlu mengajarkan bukan hanya cara bertanya, tetapi cara mengenali kapan satu jawaban menyembunyikan pluralitas." (E01)
>
> *An answer should flag that an issue is disputed, name the position of the school or the authority, give verifiable sources, state uncertainty, and encourage human consultation when the risk is high. User education needs to teach not only how to ask, but how to recognise when one answer is concealing plurality (Fatmawati, Khoiriyah, et al., 2026).*

Table 4. Themes, defining codes, and the informants whose accounts anchor each theme

Theme	Defining codes	Anchoring informants
T1. The one-answer effect	C01 fluent certainty; C02 single-answer default; C03 consensus inference; C04 convenience shortcut	U02, U06, U09, E04
T2. Plurality becomes prompt-dependent	C05 madhhab omission; C06 plurality-on-request; C07 prompt dependency; C08 novice query gap	U12, U17, U20, E05

Theme	Defining codes	Anchoring informants
T3. Decontextualisation of local fiqh	C09 local-practice mismatch; C10 institutional invisibility; C11 global-generic framing; C12 valid-but-decontextualised	U03, U13, U20, E03
T4. Verification stratified by stakes and literacy	C13 low-stakes acceptance; C14 high-stakes escalation; C15 literacy-based skepticism; C16 source triangulation	U04, U15, U22, E06
T5. Human authority as corrective infrastructure	C17 AI as first stop; C18 human confirmation; C19 accountable authority; C20 AI not mufti	U05, U11, U19, E02
T6. Counter-homogenisation through prompt literacy	C21 comparative prompting; C22 source-demanding prompt; C23 recovered plurality; C24 negative case	U10, U17, U22, E01

The Pathway That Joins the Themes

Read together, the six themes describe a contingent interaction rather than a fixed technological outcome. The strongest pathway runs from single-answer formatting and fluent certainty, through low madhhab literacy and the pull of convenience, to a perceived doctrinal consensus, after which the salience of ikhtilaf declines. Figure 2 sets out that pathway together with the conditions that amplify it and the mechanisms that interrupt it. Each interrupter identified in the data breaks the sequence at a different point, which matters for intervention because it means there is no single place where the process must be caught.

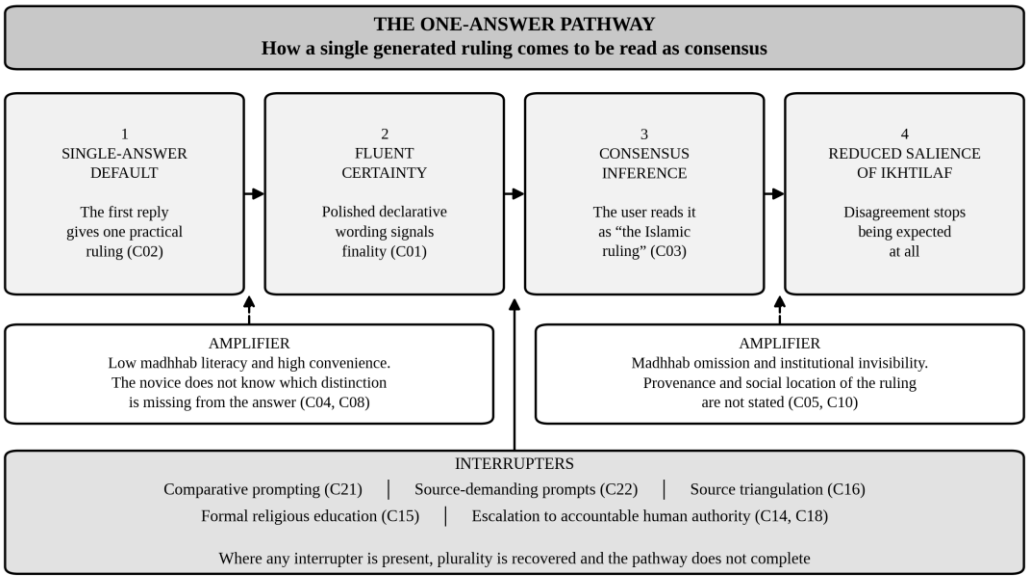


Figure 2. The one-answer pathway, its amplifying conditions, and the mechanisms that interrupt it

Code emergence supports the claim that the six themes exhaust the pattern available in this dataset. Eighteen codes appeared in the first batch of eight interviews, six more in the second, and only two in the third. A single further code was added in the fourth batch, which captured the negative case in which a system foregrounds plurality when explicitly framed to do so. The final three interviews added no new first-order code and were used to elaborate existing categories and to confirm the counterexamples already recorded.

Discussion

How a Single Answer Becomes a Consensus

The first research question asked how Muslim users interpret generated fiqh responses that omit madhhab distinctions, and the evidence answers it at the level of form rather than content. Users did not reason their way to the conclusion that a position was unanimous. They inherited that conclusion from the shape of the reply, because a single practical ruling written in confident prose carries an implicit claim about the state of the law that no sentence in the reply ever makes. This is why the distinction drawn by expert informants between propositional accuracy and epistemic integrity does real analytical work. This distinction was articulated nearly verbatim by informant E06, who contrasted propositional accuracy with epistemic integrity and openness to ikhtilaf directly, and was echoed in substance by informant E04's account of how an impression of consensus forms at the level of presentation; the present analysis adopts and generalises this emically derived distinction across the dataset rather than deriving it independently from the literature, and its contribution lies in that generalisation and in connecting it to the one-answer pathway construct. An evaluation that scores an Islamic answer as correct can be entirely right and still miss what these users experienced, since the failure lies in what the answer declines to disclose about its own provenance. Benchmark studies have begun to register this, and recent evaluations of Islamic legal reasoning report that models frequently reach a defensible ruling without identifying the school that grounds it (Bouchekif et al., 2025; Wahid, 2025). The present study shows what that omission does once a human being reads it.

The second research question concerned the interactional mechanism, and the one-answer pathway names it. Work on machine reliability has established that fluency and grounding come apart (Huang et al., 2025; Ji et al., 2023), and work on human reliance has established that unhedged output draws heavier dependence than hedged output does (Kim et al., 2024; Zhou et al., 2024). The religious case adds a step that neither literature anticipates. In an ordinary factual domain, an unhedged wrong answer misleads about a fact. In fiqh, an unhedged answer misleads about the structure of the law itself, because it implies that the question has one settled resolution when the tradition treats disagreement among qualified jurists as normal (Abu Baedah et al., 2026). The harm is therefore not confined to the episode at hand. Users in this study described a slow decline in the expectation that disagreement exists at all, and that decline is a change in religious literacy rather than a single error.

Prompt Dependence and the Inversion of Who Is Served

The third research question asked what resists homogenisation, and the answer is uncomfortable for any remedy built on user behaviour. Plurality was recoverable in almost every case, but recovery required the user to ask for it in terms the user could only supply if the relevant distinction was already known. This inverts the usual promise of an accessible tool. The user with pesantren training extracts comparative fiqh in a single follow-up, while the user with no such training receives a confident answer and has no reason to suspect that anything is absent. Research on AI literacy has mapped the competencies that separate these two users (Ng et al., 2024), and work on generative search has shown that conversational interfaces can narrow the range of views a user encounters relative to conventional search (Sharma et al., 2024). The Indonesian material adds that the narrowing is felt most sharply by those with the least capacity to notice it, which makes prompt literacy a poor foundation for policy even though it is an effective individual practice. Figure 3 sets out the resulting distribution across the two axes. With twenty-two user informants in total, any count-based breakdown across the four stakes-by-literacy cells in Figure 3 would involve very small cell sizes; whether the figure reports actual informant counts for each cell or is a conceptual illustration of the pattern

described in the text is not stated here, and the caption and in-text description would benefit from making this explicit so that the figure is not read as supporting a quantitative distribution the qualitative design was not built to establish.

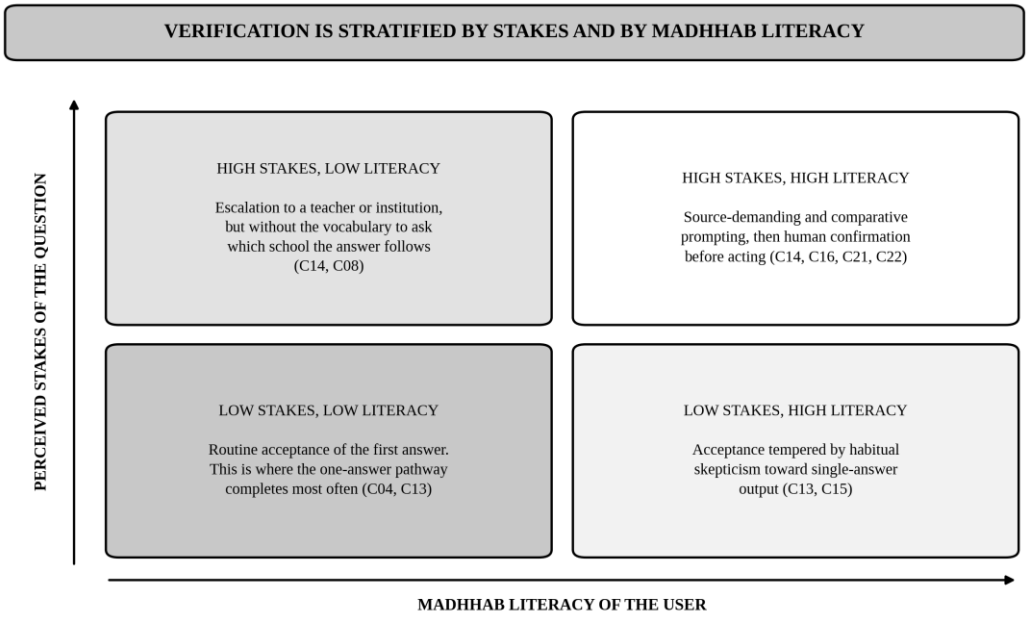


Figure 3. Verification practice stratified by perceived stakes and by madhhab literacy

Reading the four positions together clarifies where intervention would do most good. The cell in which low stakes meet low literacy is where the pathway completes most often, and it is also the cell that receives the least attention, precisely because the questions involved look trivial. Yet the accumulation of small accepted answers is what shapes a user's background sense of how settled the law is. The cell in which high stakes meet low literacy is different and in some ways more troubling, since these users do escalate to a teacher but arrive without the vocabulary to ask which school the machine answer followed. Their escalation is therefore less informative than it appears, and the human scholar is asked to adjudicate an answer whose provenance nobody in the conversation can state.

A Hybrid Authority Ecology Rather Than a Substitution

The fourth research question asked how scholars read the implications for religious authority, and their accounts do not support a displacement narrative. Human authority re-entered as corrective infrastructure at determinate points, namely ambiguity, collision with local practice, and consequence for others. This is consistent with two decades of digital religion scholarship arguing that authority is renegotiated rather than lost when religious life moves into mediated environments (Campbell & Tsuria, 2022; Kołodziejska et al., 2023), and with Indonesian work showing that established organisations retain the power to validate even as platforms multiply the voices in circulation (Akmaliah, 2022; Raya, 2025). What the present study adds is the trigger condition. As summarised here, the renegotiation literature cited above establishes that authority is re-sought under mediation in general, but it is not described in this manuscript as having already specified accountability for consequences to others as a discrete trigger; on that basis, the contribution is best read as a specific empirical instance of a trigger condition within an already-established renegotiation mechanism, rather than as a previously undescribed theoretical mechanism in its own right. Users did not consult a scholar because they doubted the machine in general. They consulted a scholar when a decision would touch another person,

which suggests that the operative criterion is accountability rather than accuracy (Tawakkal et al., 2026).

That criterion has a counterpart in the juristic tradition itself. The office of the mufti carries responsibility for a ruling in a way that a generated text cannot, and recent normative scholarship has argued that this responsibility, rather than the informational content of a fatwa, is what resists automation (A. Ahmed et al., 2026; S. Ahmed, 2024; Fazal, 2026). The users in this study reached the same conclusion without the vocabulary to state it, since they described wanting someone who could be asked again, who would enquire about their circumstances, and who would bear the weight of having advised them. Studies of anomalous digital fatwa have warned that guidance detached from this structure proliferates easily (Ali & Aljahsh, 2025), and the accounts collected here show the mechanism by which ordinary users nonetheless keep pulling that structure back into the process.

Corroboration from the Model Side

Evidence generated on the model side corroborates what these informants described from the user side, which strengthens the account by triangulating two independent kinds of data. Benchmark work covering thousands of Islamic knowledge items reports accuracy that varies widely across models and includes explicit tasks for detecting whether a school of law has been identified (Abdelaal et al., 2026). Evaluations built specifically to probe latent sectarian preference find that the position a model adopts can shift with the language of the prompt and with the nationality the user claims (Maheshwari et al., 2026), which is a model-side statement of the prompt dependence that users in this study experienced as a change in the answer after a follow-up request. Assessments of juristic reasoning across ranks of authority report that models handle settled material more reliably than contested material (Omran et al., 2026), and a close reading of chatbot answers on waqf found positions that were defensible but presented without the apparatus that would let a learner locate them (Susanti et al., 2025). Work on abstention adds the complementary finding that systems seldom decline a religious question even where declining would be the responsible move (Atif et al., 2025). The present study explains why that reluctance matters, since a system that does not decline and does not attribute leaves the user with nothing to alert them.

Implications for Answer Design and Religious Education

Two consequences for answer design follow directly from this account. These design implications are suggested by informants' retrospective preferences and by one expert informant's specification (E01) rather than demonstrated by any test of whether implementing them would change users' inferences about consensus; they are accordingly better read as design directions that merit testing, consistent with the audit study proposed in the Conclusion, rather than as consequences established by the present findings. The first of these concerns disclosure rather than content. Every item in the specification offered by expert informant E01 asks the system to say something about its own answer rather than to change the answer, which means the intervention is compatible with systems that already produce defensible rulings. Marking an issue as disputed, naming the school, supplying a verifiable source, and stating uncertainty are all disclosure moves, and experimental work indicates that expressed uncertainty does shift reliance in the intended direction (Kim et al., 2024; Zhou et al., 2024). The second consequence concerns friction at the moment of acceptance. Work on overreliance shows that interventions which interrupt acceptance at the moment of decision outperform those that supply more information afterwards (Buçinca et al., 2021; Vasconcelos et al., 2023). For religious guidance, a friction point placed where a user is about to act on a ruling would reach exactly the high-stakes cell identified in this study (Haditia et al., 2026).

For religious education the implication is narrower than the usual call for digital literacy. The competence that mattered in these data was not scepticism toward technology in general, which several low-literacy users possessed without benefit, but the ability to recognise that a given question belongs to a disputed class. That is a fiqh competence before it is a digital one. Curricula that teach the map of ikhtilaf on common worship and transaction questions would give users the pattern-recognition that prompting presupposes, and pesantren and university programmes in Indonesia are already positioned to teach it (Faizin et al., 2025). The contribution of the present study to that debate is to specify what users need to recognise rather than what they need to distrust (Ahmad et al., 2025; Fatmawati, Musyarofah, et al., 2026).

Contribution to Digital Religion Scholarship

The theoretical contribution of this study is to separate two things that digital religion scholarship has tended to treat together. More precisely, the claim is not that the existing circulation framework is inadequate to describe generative systems in general, but that it does not, on its own, specify a mechanism for how a single synthesised answer produces an inference of consensus in the way that this study's one-answer pathway describes; the contribution is this more specific mechanism rather than a wholesale replacement of circulation-based accounts of authority. That literature has explained with some precision how mediation redistributes visibility and how new religious actors acquire standing through platforms (Becker, 2021; Bunt, 2018; Whyte, 2022). Its central concern has been the circulation of religious speech produced by people. Generative systems introduce a second process that runs alongside circulation, namely the production of new religious text whose default form carries an implicit claim about the state of the law. The one-answer pathway names that second process and shows that it operates before any question of credibility arises, since the user has drawn a conclusion about consensus from the shape of the reply rather than from any judgement about its source. This sits close to the argument that conversational interfaces reframe information seeking itself by returning a synthesised answer where a list of sources once invited comparison (Shah & Bender, 2022). The religious case sharpens that argument, because in fiqh the list of sources was not merely a convenience for the user but a representation of how the tradition holds its own disagreements. Concretely, Shah and Bender's general account would predict that returning a synthesised answer in place of a source list mainly costs the user an opportunity for convenient comparison; it does not on its own predict that the user would infer the matter is doctrinally settled, since ordinary informational synthesis does not usually carry a claim about consensus. The religious case sharpens the argument specifically by showing that synthesis here produces exactly that additional, mistaken inference of consensus, which is the prediction the general account does not generate.

Limitations

Four limitations qualify the conclusions that are drawn above. The design is cross-sectional and rests on recollected episodes alongside one live prompt task, so the account of how expectations decline over time is an interpretation of retrospective reports rather than an observed trajectory. The sample is urban and Indonesian, drawn from four clusters that were chosen for institutional contrast rather than for representativeness, and the findings should not be extended to rural settings or to other national contexts without replication. Interviews were conducted in Indonesian and only selected excerpts were translated, which means readers depend on the research team for the fidelity of the rendering even though the original wording is supplied. The study also did not audit the systems themselves, so claims about what a model does are claims about what informants reported and demonstrated rather than about model behaviour measured under controlled conditions. Pairing this design with a benchmark audit in Indonesian and Javanese would address that gap directly.

Conclusion

This study examined how generative artificial intelligence compresses juristic plurality in everyday Islamic legal inquiry, drawing on interviews with thirty informants across four Indonesian urban clusters. The analysis supports four conclusions about that process. First, the impression of doctrinal consensus is produced by the form of the answer rather than by its content, since a single practical ruling written in fluent declarative prose carries an implicit claim about the state of the law that the text never states and the user never examines. Second, juristic plurality remains recoverable but has become prompt-dependent, which inverts the expected distribution of benefit because the users who can retrieve comparative fiqh are those who already knew the distinction existed. Third, verification is stratified jointly by perceived stakes and by madhhab literacy, so homogenisation concentrates where low consequence, high convenience, and low literacy meet. Fourth, human religious authority is not displaced but returns as corrective infrastructure at the point where a decision acquires consequences for other people, which indicates that users are tracking accountability rather than accuracy.

Taken together these conclusions reframe the problem that religious literacy programmes are usually asked to solve. The risk identified here is not that users trust machines too much in general, since most informants were measured about what they would act on. The risk is that a particular class of question stops looking disputed, and that shift happens quietly, through the accumulation of small confident answers about matters that seemed too minor to check. Naming the one-answer pathway makes that process available for intervention, and the analysis shows that it can be interrupted at four separate points rather than only at the point of acceptance.

Three lines of work would advance this agenda most directly. Audit studies should test whether disclosure features change what users infer about consensus, using the specification that expert informants articulated as their outcome measure rather than accuracy alone. Longitudinal designs should follow users across a period of months to establish whether the decline in the expected salience of ikhtilaf that informants described is observable rather than merely recalled. Comparative work outside Indonesia and outside Indonesian-language prompting should establish how far the pathway holds where the institutional ecology of madhhab teaching differs, since the mechanism proposed here depends on there being a local tradition from which an answer can be detached.

Acknowledgement

The authors thank the thirty informants who gave their time in Jakarta, Depok, Yogyakarta, Surabaya, Sidoarjo, and Makassar, and the pesantren and university colleagues who commented on the interview guide before fieldwork began.

Declaration of Generative AI Use

Generative artificial intelligence was used in the preparation of this manuscript. Anthropic's Claude was used only to format the manuscript according to the journal template and to improve the language. The research design, fieldwork, coding decisions, thematic interpretation, and all substantive argumentation are the responsibility of the authors. The authors reviewed and edited the full manuscript and take full responsibility for its content.

Authors' Contribution Statement

Nurbaiti: conceptualization, research design, development of the interview guide, codebook thematic analysis, and writing of the original draft.
Aas Siti Sholichah: fieldwork coordination across the site clusters, conducting interviews, data curation, and writing (review and editing).

Muhammad Arinal Haq: fiqh and usul al-fiqh analysis, validation of the codes and themes, negative-case analysis, 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:2603.23750). arXiv. https://arxiv.org/abs/2603.23750

Abu Baedah, S. S., Musolin, M. H., & Nurlaelah. (2026). Dialectics of fiqh al-ikhtilāf: A comparative madhhab analysis of legal pluralism in contemporary Muslim societies. *Mazahibuna: Jurnal Perbandingan Mazhab, 8*(1), 19–38. https://doi.org/10.24252/mazahibuna.vi.64995

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

Ahmed, A., Masuwd, M., Sabir, M. R., & Mahmudulhassan, M. (2026). Can artificial intelligence perform ijtihad? Authority, epistemology, and legitimacy in digital Islamic jurisprudence. *Suhuf: International Journal of Islamic Studies, 38*(1), 154–169. https://doi.org/10.23917/suhuf.v38i1.16437

Ahmed, S. (2024). Fatwā and AI: A literature-based assessment on how AI impacts the role of the muftī. *Journal for Islamic Studies, 42*(1–2), Article 14400. https://doi.org/10.25159/2957-9163/14400

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health, 2*, Article 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Akmaliah, W. (2022). Reclaiming moderate Islam in Nahdlatul Ulama: Challenging the dominant religious authority in digital platform. *Journal of Indonesian Islam, 16*(1), 223–248. https://doi.org/10.15642/JIIS.2022.16.1.223-248

Ali, A. M. M., & Aljahsh, M. A. I. (2025). The proliferation of anomalous digital fatwas: A critical examination of jurisprudential challenges and societal impacts in the digital age. *AHKAM: Jurnal Ilmu Syariah, 25*(1), 37–52. https://doi.org/10.15408/ajis.v25i1.37905

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

Bayuo, J., Wong, F. K. Y., Li, Y., Lu, W., & Wong, A. K. C. (2025). Framing the multi-centre qualitative research design as a novel methodology for nursing and healthcare research: Reflections and a methodological discussion. *Journal of Advanced Nursing, 81*(8), 5196–5207. https://doi.org/10.1111/jan.16548

Becker, C. (2021). The production of Salafi spaces in computer-mediated environments: A social theory perspective on “digital religion.” *Journal of Religion, Media and Digital Culture, 10*(1), 30–49. https://doi.org/10.1163/21659214-bja10033

Bouchekif, A., Rashwani, S., Sbahi, H., Gaben, S., Al Khatib, M., & Ghaly, M. (2025). Assessing large language models on Islamic legal reasoning: Evidence from inheritance law evaluation. In *Proceedings of the Third Arabic Natural Language Processing Conference* (pp. 246–257). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.arabicnlp-main.20

Bouncken, R. B., Czakon, W., & Schmitt, F. (2026). Purposeful sampling and saturation in qualitative research methodologies: Recommendations and review. *Review of Managerial Science, 20*(2), 579–615. https://doi.org/10.1007/s11846-025-00881-2

Braun, V., & Clarke, V. (2021a). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Braun, V., & Clarke, V. (2021b). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health, 13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

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

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Bunt, G. R. (2018). *Hashtag Islam: How cyber-Islamic environments are transforming religious authority*. University of North Carolina Press.

Campbell, H. A., & Tsuria, R. (Eds.). (2022). *Digital religion: Understanding religious practice in digital media* (2nd ed.). Routledge.

Faizin, N., Maarif, A. S., & Hanafi, Y. (2025). Considering religious moderation in Islamic law through AI ChatGPT and Bahsul Masail of Nahdlatul Ulama. *Law, Innovation and Technology, 17*(1), 271–288. https://doi.org/10.1080/17579961.2025.2469351

Fakhruroji, M. (2025). Socially distanced, digitally engaged: Mediatization of religious practices of Indonesian digital natives during the COVID-19 pandemic. *Journal of Religion, Media and Digital Culture, 14*(1), 23–44. https://doi.org/10.1163/21659214-bja10125

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

Fazal, M. A. (2026). AI and Islamic jurisprudence: Re-reading contemporary fatwās for ethical possibilities in techno-social forms of life. *Islamic Studies Review, 5*(1), 91–116. https://doi.org/10.56529/isr.v5i1.540

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

Hefner, C.-M. (2022). Morality, religious authority, and the digital edge: Indonesian Muslim schoolgirls online. *American Ethnologist, 49*(3), 359–373. https://doi.org/10.1111/amet.13088

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine, 292*, Article 114523. https://doi.org/10.1016/j.socscimed.2021.114523

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), 1–55. https://doi.org/10.1145/3703155

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), 1–38. https://doi.org/10.1145/3571730

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, S. S. Y., Liao, Q. V., Vorvoreanu, M., Ballard, S., & Vaughan, J. W. (2024). "I'm not sure, but…": Examining the impact of large language models' uncertainty expression on user reliance and trust. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 822–835). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658941

Kołodziejska, M., Fajfer, Ł., Hall, D., & Radde-Antweiler, K. (2023). Religious media settlers in times of deep mediatization. *Religion, 53*(2), 199–223. https://doi.org/10.1080/0048721X.2022.2083032

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

Liu, N. F., Zhang, T., & Liang, P. (2023). Evaluating verifiability in generative search engines. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 7001–7025). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.467

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

Maheshwari, A., Gajkeshwar, A., Sharma, K., & Patel, V. (2026). *SectEval: Evaluating the latent sectarian preferences of large language models* (arXiv:2603.12768). arXiv. https://arxiv.org/abs/2603.12768

Motulsky, S. L. (2021). Is member checking the gold standard of quality in qualitative research? *Qualitative Psychology, 8*(3), 389–406. https://doi.org/10.1037/qup0000215

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

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

Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2023). A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Medical Teacher, 45*(3), 241–251. https://doi.org/10.1080/0142159X.2022.2057287

Omran, A., Hassan, S., & Hassan, W. M. (2026). Artificial reasoning and Islamic law: Mapping LLM capabilities across juristic ranks and reasoning frameworks. *Journal of Islamic Thought and Civilization, 16*(1), 196–211. https://doi.org/10.32350/jitc.161.12

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

Raya, M. K. F. (2025). Digital Islam: New space for authority and religious commodification among Islamic preachers in contemporary Indonesia. *Contemporary Islam, 19*(1), 161–194. https://doi.org/10.1007/s11562-024-00570-z

Romeo, G., & Conti, D. (2026). Exploring automation bias in human-AI collaboration: A review and implications for explainable AI. *AI & Society, 41*(1), 259–278. https://doi.org/10.1007/s00146-025-02422-7

Shah, C., & Bender, E. M. (2022). Situating search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (pp. 221–232). Association for Computing Machinery. https://doi.org/10.1145/3498366.3505816

Sharma, N., Liao, Q. V., & Xiao, Z. (2024). Generative echo chamber? Effect of LLM-powered search systems on diverse information seeking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642459

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), 496–518. https://doi.org/10.23917/suhuf.v37i2.11821

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

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349–357. https://doi.org/10.1093/intqhc/mzm042

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on AI systems during decision-

making. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), 1–38. https://doi.org/10.1145/3579605

Wahid, S. H. (2025). Artificial intelligence in Islamic guidance: Assessing chatbot performance and jurisprudential adherence. *Journal of Digital Islamicate Research, 3*(1), 33–96. https://doi.org/10.1163/27732363-20250003

Whyte, S. A. (2022). Islamic religious authority in cyberspace: A qualitative study of Muslim religious actors in Australia. *Religions, 13*(1), Article 69. https://doi.org/10.3390/rel13010069

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), Article 335. https://doi.org/10.3390/rel13040335

Zhou, K., Hwang, J. D., Ren, X., & Sap, M. (2024). Relying on the unreliable: The impact of language models' reluctance to express uncertainty. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 3623–3643). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.198