Interdisciplinary Explorations in Research Journal (IERJ)
Vol. 4, Number 3 (Sept-Dec, 2026), pp. 972-981
E-ISSN: 3032-1069

HUMAN RESPONSIBILITY IN THE PRODUCTION OF AI-GENERATED ISLAMIC RELIGIOUS CONTENT

M. Hasby
Fakultas Dakwah dan Ilmu Komunikasi UIN Antasari Banjarmasin, Indonesia
E-mail: mhasby21@gmail.com

Received 03-10-2026| Received in revised form 25-10-2026 | Accepted 10-11-2026

Abstract
Generative AI now writes answers about prayer, inheritance, and divorce, drafts sermons, and summarizes tafsir for large numbers of Muslim users. The practice is controversial because the output reads like religious guidance while no accountable person may stand behind it. Indonesian bodies such as Nahdlatul Ulama and the Indonesian Ulema Council have responded with warnings, yet machine-made religious content keeps growing, so the question of who answers for it is now urgent. Research on the topic clusters around AI-generated fatwas, digital religious authority, and general AI ethics. Islamic studies stress ijtihad, intention, and moral agency, while AI ethics describes a responsibility gap, but the two bodies of work rarely meet at the level of the production chain. This article asks who bears responsibility at each stage of producing AI-generated Islamic content, what weakens that responsibility, and how it can be rebuilt. We treat responsibility as a duty that must be assigned to identifiable humans. We used a qualitative normative-conceptual design with content analysis of Islamic primary sources, classical works on the ethics of the mufti, Indonesian institutional positions, and scholarship published mainly between 2021 and 2026. Each production stage was coded for actors, duties, and risks, and read through maqasid al-shari'ah and the AI responsibility literature. Islamic norms already place accountability on human speakers. Opacity, missing moral agency, platform scale, and thin institutional rules spread that accountability thin. A five-duty model of authorship, verification, disclosure, traceability, and remedy can restore it, with stricter review for higher-risk content. Machines can assist, but responsibility for religious speech stays with people. We recommend that religious councils, publishers, platforms, and developers adopt risk-tiered review and disclosure rules, and that Indonesian regulators align AI ethics guidance with them.
Keywords: *AI-generated religious content; human responsibility; fatwa authority; maqasid al-shari'ah; Indonesian Islam*

Abstrak
AI generatif kini menulis jawaban tentang salat, warisan, dan perceraian, menyusun draf khotbah, dan merangkum tafsir bagi banyak pengguna Muslim. Praktik ini diperdebatkan karena hasilnya terbaca seperti tuntunan agama, sementara tidak ada pihak yang jelas bertanggung jawab di baliknya. Lembaga Indonesia seperti Nahdlatul Ulama dan Majelis Ulama Indonesia telah menanggapinya dengan peringatan, tetapi konten keagamaan buatan mesin terus bertambah, sehingga pertanyaan tentang siapa yang bertanggung jawab kini mendesak. Penelitian tentang topik ini mengelompok pada fatwa buatan AI, otoritas keagamaan digital, dan etika AI secara umum. Kajian Islam menekankan ijtihad, niat, dan kecakapan moral, sedangkan etika AI menggambarkan celah tanggung jawab, tetapi keduanya jarang bertemu pada tingkat rantai produksi. Artikel ini menanyakan siapa yang menanggung

tanggung jawab pada setiap tahap produksi konten Islam buatan AI, apa yang melemahkannya, dan bagaimana tanggung jawab itu dapat dibangun kembali. Tanggung jawab diperlakukan sebagai kewajiban yang harus dilekatkan pada manusia yang dapat diidentifikasi. Kami memakai desain normatif-konseptual kualitatif dengan analisis isi terhadap sumber primer Islam, karya klasik tentang etika mufti, posisi lembaga di Indonesia, dan literatur yang terbit terutama pada 2021 hingga 2026. Setiap tahap produksi dikodekan menurut pelaku, kewajiban, dan risikonya, lalu ditafsirkan melalui maqasid syariah dan literatur tanggung jawab AI. Norma Islam sudah menempatkan akuntabilitas pada penutur manusia. Opasitas model, ketiadaan kecakapan moral, skala platform, dan aturan kelembagaan yang tipis membuat akuntabilitas itu menyebar dan melemah. Model lima kewajiban berupa kepenulisan, verifikasi, keterbukaan, keterlacakan, dan pemulihan dapat memulihkannya, dengan peninjauan yang lebih ketat untuk konten berisiko tinggi. Mesin dapat membantu, tetapi tanggung jawab atas ucapan keagamaan tetap berada pada manusia. Kami merekomendasikan agar majelis keagamaan, penerbit, platform, dan pengembang menerapkan aturan peninjauan dan keterbukaan berjenjang menurut risiko, serta agar regulator Indonesia menyelaraskan pedoman etika AI dengan aturan tersebut.
Kata kunci: konten keagamaan buatan AI; tanggung jawab manusia; otoritas fatwa; maqasid syariah; Islam Indonesia

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

1. Introduction

Large language models have become an everyday source of Islamic content. People ask chatbots about prayer, inheritance, and divorce, preachers use them to draft sermons, and content creators use them to produce posts and videos at a pace no scholar could match. The trend is controversial because the output sounds confident and learned while no accountable person may stand behind it. Nahdlatul Ulama's 2023 national assembly ruled that asking AI about religion is permissible but that treating its answers as guidance to be practiced is forbidden, because their correctness is not assured (NU Online, 2023a). The Indonesian Ulema Council (MUI) has said that AI cannot be a source of religious rulings and can only accompany scholars (detikHikmah, 2025; NU Online, 2025a). The emergency is that production already runs ahead of any rule that says who must answer for what is produced.

The urgency lies in what religious content does. A wrong recipe costs a meal, but a wrong ruling on worship, marriage, or finance changes how a believer acts before God and toward others. Islamic ethics treats speaking about religion without knowledge as a grave matter (Q 7:33; 16:116), and the Prophetic warning against attributing falsehood to him is well known. Indonesia adds a practical layer: a large and well-connected Muslim public, a strong culture of scholarly transmission (*sanad*), and platforms where AI-assisted content spreads without institutional filters (Raya,

2025; Samudra et al., 2024). If responsibility cannot be located, harm cannot be corrected.

Global scholarship has begun to map the field. Ahmed (2024), Usmonov (2025), and Yiğit and Şimşek (2025) examine AI-generated fatwas and find limits in contextual reasoning and inconsistencies against official rulings, while Ab Rahim et al. (2025) propose guidelines that keep authority and accountability with human scholars. Atallah (2026) places AI within a wider account of digital mediation and fatwa authority. In AI ethics, Matthias (2004) and Santoni de Sio and Mecacci (2021) describe responsibility gaps, and Floridi and Cowls (2019) add accountability and explicability to a shared set of principles. The gap is that Islamic studies work centers on the single act of issuing a fatwa, while AI ethics seldom engages Islamic accounts of responsibility. Neither follows content through the full chain of design, prompting, review, publication, and correction.

Indonesian work approaches the topic through digital authority and institutional statements. Studies describe online muftis, preachers, and algorithmic visibility (Hamdani, 2023; Raya, 2025; Samudra et al., 2024), the digital transmission of Nahdlatul Ulama's *bahtsul masa'il* (Yahya & Junaidi, 2021), and the MUI fatwa on conduct in social media (Shuhufi et al., 2022). Recent institutional statements address AI directly (NU Online, 2023a, 2025a, 2025b). What is missing is a systematic account that turns these positions into duties for specific actors and connects them to Indonesian regulatory instruments, such as the 2023 circular on AI ethics.

The problem is a responsibility gap in a domain where responsibility is central. Islamic law holds the human speaker accountable, yet AI-generated religious content passes through several hands and no one feels fully answerable. Three questions guide the study. First, what do Islamic sources, institutional positions, and AI ethics say about who is responsible for religious content that a machine helps to produce? Second, what factors weaken that responsibility in practice? Third, what model can restore it across the production chain?

Two frames guide the analysis. In Islamic thought, religious speech is a trust (*amanah*), and the mufti is described as one who signs on behalf of God (Ibn Qayyim al-Jawziyyah, 1991), subject to conditions of knowledge, integrity, and caution set out in works on the ethics of the mufti (Ibn al-Salah, 1986). Qur'anic injunctions to verify news (Q 49:6) and to avoid speaking without knowledge (Q 17:36) extend this duty to

anyone who transmits religious claims. Maqasid al-shari'ah, especially the protection of religion and intellect (Auda, 2008; Ibn Ashur, 2006), supplies the purpose that rules should serve. In positive law, Indonesia has no statute on AI-generated religious content. Relevant instruments include MUI Fatwa 24/2017 on conduct in social media, the Electronic Information and Transactions Law as amended in 2024, and the 2023 ministerial circular on AI ethics.

The study has three objectives: to map responsibility at each stage of producing AI-generated Islamic content, to identify the factors that weaken it, and to propose a model that restores it.

The contribution is threefold. Conceptually, the article moves the question from whether AI may issue fatwas to who must answer for AI-assisted religious content, and it shows that Islamic norms already contain an answer. Analytically, it offers a stage-by-stage account of responsibility, so duties can be assigned to developers, prompters, reviewers, publishers, and councils. Practically, it proposes the Accountable Human Chain model, which religious councils and platforms can adopt without waiting for new legislation and can scale to the risk of the content.

2. Literature Review

Research on religion and technology has moved from asking whether religion survives digital life to asking how authority is rebuilt within it. Hjarvard's (2008, 2011) theory of mediatization treats media as institutions that reshape religion, and Campbell and Tsuria (2021) show how religious communities negotiate digital tools. For Islam, Bunt (2003, 2018) documented cyber-Islamic environments and online fatwas, Šisler (2011) examined online counselling and identity, and Skovgaard-Petersen (1997) showed that print had already changed how fatwas circulate. Studies of Indonesian Islam add a national angle: the authority of ulama organizations, the role of pesantren, and the reach of social media preachers make the country a useful site for studying how religious authority is negotiated online (Hamdani, 2023; Raya, 2025; Yahya & Junaidi, 2021). Generative AI raises a new question, because the system produces new religious text rather than only carrying human-authored text. The shift matters for accountability, since a text with no human author is hard to hold to account. We group recent work into three trends: the factors that shape the quality and authority of AI religious content, the processes through which it is produced and checked, and its impact on communities.

The first trend identifies factors behind quality and authority. Technical studies explain why fluent output can be wrong: language models predict plausible text without a grounded notion of truth, and they can fabricate sources (Bender et al., 2021; Ji et al., 2023). Islamic studies add that AI lacks ijtihad, intention (*niyya*), and moral accountability (Ahmed, 2024; Usmonov, 2025), and comparisons with official fatwas report inconsistencies and sectarian blind spots (Yiğit & Şimşek, 2025). Atallah (2026) adds epistemic opacity and bias to the list. Platform conditions matter as well, since visibility metrics and engagement shape whose religious voice spreads (Samudra et al., 2024; Wahid, 2024), and religious influencers compete with trained scholars (Hamdeh, 2020; Triantoro et al., 2023). Demand for fast answers pushes users toward whatever responds first (Hamdani, 2023). A further factor is the corpus itself. Models learn from text that overrepresents some languages, schools, and regions, so their religious answers may reflect whichever tradition is best represented online, which fits reports of sectarian blind spots (Yiğit & Şimşek, 2025). Ali and Aljahsh (2025) describe the proliferation of anomalous digital fatwas and its societal effects.

The second trend concerns process. Ab Rahim et al. (2025) propose guidelines that require human oversight, explainability, and accountability in AI-assisted fatwa issuance. Usmonov (2025) concludes that AI cannot fully replace scholars but may work under supervision in areas such as Islamic finance, halal certification, and preliminary guidance, and Ahmed (2024) describes a partnership between AI and human judgment. Sati et al. (2025) trace how the digital transformation of tafsir affects legal derivation, and Hastuti et al. (2026) show how the concept of *tabayyun* and the MUI fatwa can be built into a fact-verification design. In Indonesia, Nahdlatul Ulama's assembly both restricted reliance on AI answers and recommended that the organization build an AI whose content is supplied by people with fatwa authority (NU Online, 2023a, 2023b). MUI has announced guidance on AI for religious references (NU Online, 2025b). These process studies share a cautious stance, but they differ on how much AI should do. Some see room for AI in preliminary guidance and retrieval, others restrict it to support for scholars, and the Indonesian positions are stricter on reliance than on use. The difference shows that the debate is about conditions of use, not a simple permission or ban.

The third trend examines impact. Digital mediation can widen access to religious guidance, but it also risks fragmentation, erosion of institutional credibility, and diffusion of accountability (Atallah, 2026; Wahid, 2024). Problems of credibility

predate AI: Yucel and Albayrak (2021) found weaknesses in online fatwas on organ donation, and Maravia et al. (2021) analyzed contested religious advice during the pandemic. Ismail and Seedat (2017) show that online fatwa spaces rebuild authority through framing and rhetoric. MUI leaders now warn that AI should not displace scholars or the chain of scholarly transmission (NU Online, 2025a). Read together, these studies suggest that the harm of AI religious content lies less in single errors than in the loss of a visible, answerable speaker. Digital fatwa research also points to benefits that a responsibility model should not erase: easier access for people far from scholars, quicker response to new questions, and wider reach for moderate voices (Ahmed, 2024; Atallah, 2026). Any model must therefore protect accountability without shutting down useful assistance.

The literature has strengths, but five limits stand out. First, it centers on the fatwa, while most AI religious content is sermon drafts, tafsir summaries, devotional texts, and social media posts that never claim to be fatwas. Second, many studies are conceptual or test a handful of prompts, so their evidence is thin. Third, guidelines remain general and rarely assign duties to developers, publishers, and platforms. Fourth, Indonesian institutional positions are known mostly through news reports and have not been analyzed systematically. Fifth, AI ethics speaks of responsibility gaps and the problem of many hands (Santoni de Sio & Mecacci, 2021; Thompson, 1980) but seldom engages Islamic accounts of accountability, while Islamic studies seldom engage the many-hands problem. Reliance on mainly Sunni sources, which Atallah (2026) acknowledges, adds a further limit. Finally, almost no study treats the people who prompt AI as responsible agents, even though they decide what to ask, what to keep, and what to publish.

This article responds in three ways. It widens the object from the fatwa to AI-generated Islamic content of all kinds. It follows that content through five production stages and asks who owes what at each. And it joins Islamic accountability norms with the AI responsibility literature to propose a model with assigned duties and risk tiers. The novelty lies in the combination: a stage-by-stage account of responsibility, paired with a model that lets institutions apply stricter duties where the content is riskier. It also gives the later sections a clear test: a model is useful if it assigns a duty at every stage and keeps useful assistance available.

3. Method

This study uses a qualitative, normative-conceptual design. We ask *what* Islamic and institutional sources say about responsibility for religious content, *why* that responsibility weakens when a machine helps produce the content, and *how* it can be rebuilt. A conceptual design suits the question because the problem lies in how duties are defined and assigned, which output tests alone cannot show. The design is conceptual by choice: it aims to build and justify a model, and empirical testing is left for later work. Table 1 summarizes the design.

Table 1. Research design

Element	Specification	Reason
Approach	Qualitative, normative-conceptual	The question concerns how duties are defined and assigned
Orientation	Content analysis of normative texts, institutional positions, and scholarship	Links what sources say to how AI content is produced
Theoretical lens	Maqasid al-shari'ah (Auda, 2008; Ibn Ashur, 2006) and AI responsibility-gap literature	Joins the purpose of religious rules with accountability in AI
Corpus	Qur'anic and Prophetic texts, classical adab al-mufti works, MUI and NU positions, scholarship from 2021 to 2026	Covers text, institution, and recent research
Analytical unit	Five production stages coded on four dimensions of responsibility	Locates where responsibility fades (Figure 1)
Technique	Directed content analysis with a deductive grid and inductive codes	Keeps the grid open to unexpected matters
Quality control	Triangulation, audit trail, negative case search	Reduces selective reading

The unit of analysis is the production stage. We distinguish five stages: design and training, prompting and generation, review and verification, publication and dissemination, and reception and correction. Each stage is coded on four dimensions of responsibility: role (the duty attached to a position), epistemic (the duty to know and verify), moral (answerability before God and people), and institutional (the duty to set and enforce rules). Figure 1 shows the chain and the coding dimensions. The four dimensions are not exclusive, and one duty can sit on more than one of them, but separating them helps show whether a gap concerns who must act, what they must know, whom they answer to, or what rules back them. Using stages rather than single acts lets us locate where responsibility fades and compare duties across actors.

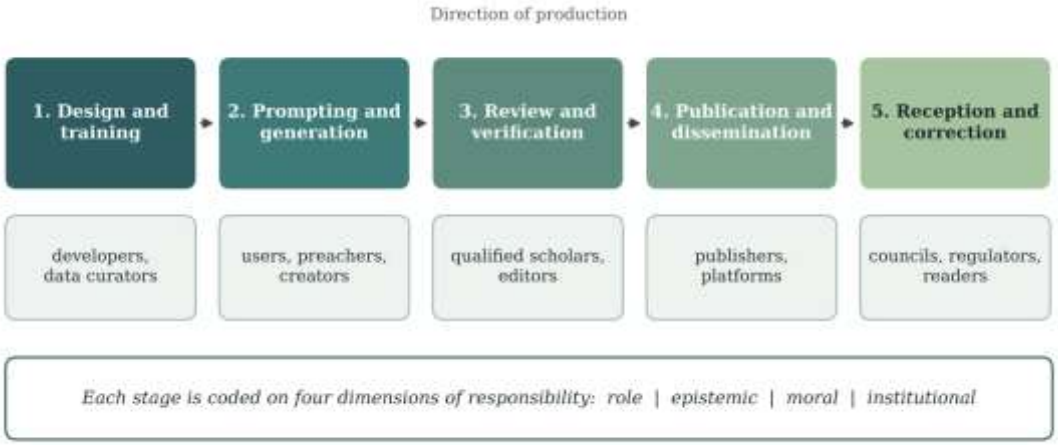

Figure 1. Unit of analysis: five production stages of AI-generated Islamic content

Primary sources are normative texts. They include the Qur'anic verses on speaking without knowledge, verifying news, and trusts, together with the Prophetic reports on attributing falsehood to the Prophet and on issuing fatwas without knowledge. They also include classical works on the ethics of the mufti (Ibn al-Salah, 1986; Ibn Qayyim al-Jawziyyah, 1991), MUI Fatwa 24/2017, the Electronic Information and Transactions Law as amended, and the 2023 ministerial circular on AI ethics. Institutional positions of Nahdlatul Ulama and MUI were taken from published news reports of their national assemblies and statements. We treat the Qur'anic and Prophetic texts as normative sources whose meaning is read through the classical works cited, not as texts that we interpret independently. Secondary sources are peer-reviewed studies on AI-generated fatwas, digital religious authority, and AI ethics, published mainly between 2021 and 2026.

Data collection followed a structured search and screening procedure. We searched Scopus, Web of Science, Google Scholar, and the Garuda portal using English and Indonesian terms such as AI-generated fatwa, Islamic chatbot, digital fatwa authority, responsibility gap, *kecerdasan buatan*, and *fatwa*. We kept studies that addressed religious content produced or assisted by AI, or that supplied a framework for responsibility, and excluded work on AI in Islamic finance or education that did not discuss religious authority or accountability. Institutional statements were retrieved from NU Online and detikHikmah and were read as reported positions, not as formal fatwa texts.

Analysis used directed qualitative content analysis. We built a deductive grid from the five stages and four dimensions, then added inductive codes when sources raised matters the grid had not anticipated, such as the distinction between content types and the role of *sanad*. Coded material was grouped under three dimensions: descriptive (what sources say about each stage), critical (which factors weaken responsibility), and transformative (what model would restore it). Interpretation used maqasid al-shari'ah to test whether a duty protects religion and intellect, and the AI responsibility literature to test whether a duty closes a gap. Triangulation across Islamic texts, institutional statements, and independent studies, an audit trail of coding decisions, and a search for negative cases guarded against selective reading. We did not interview practitioners or run tests on AI systems, so the findings are conceptual and should be tested empirically.

4. Results

4.1. Descriptive: Text and Practice

The first finding is that Islamic and institutional sources already assign responsibility for religious speech to identifiable humans, but they do so mainly for the act of issuing a fatwa, and they say little about the other stages through which AI-assisted content is produced.

In short, the texts are clear about who answers for a ruling and quiet about who answers for a machine-drafted sermon, summary, or post.

Table 2 sets out the stages. In classical sources, the mufti must be knowledgeable, trustworthy, and careful, and speaks as one who signs on behalf of God (Ibn al-Salah, 1986; Ibn Qayyim al-Jawziyyah, 1991). The Qur'an forbids attributing to God what

one does not know (Q 7:33; 16:116), tells believers to verify reports (Q 49:6), and forbids pursuing what one has no knowledge of (Q 17:36). These verses address speakers, not tools, so they bind whoever prompts, edits, or publishes AI output. Institutional positions follow the same logic. Nahdlatul Ulama allows asking AI but forbids practicing from its answers because correctness is not guaranteed (NU Online, 2023a), and MUI leaders say AI lacks the human consciousness that issuing a fatwa requires and cannot replace scholars or *sanad* (detikHikmah, 2025; NU Online, 2025a).

Table 2. Responsibility across production stages

Stage	Main actors	Grounds in Islamic and institutional sources	Typical gap
Design and training	Developers, data curators, model providers	Cooperation only in righteousness (Q 5:2); explicability and accountability principles (Floridi & Cowls, 2019)	Training sources and religious-domain testing not disclosed
Prompting and generation	Users, preachers, content creators	Duty to know before speaking (Q 17:36; Q 7:33)	Output attributed to the tool rather than to the prompter
Review and verification	Qualified scholars, editors, fatwa bodies	Conditions of the mufti (Ibn al-Salah, 1986); verification of reports (Q 49:6); supervised use (Ab Rahim et al., 2025; Usmonov, 2025)	Review absent or nominal; no standard for what review means
Publication and dissemination	Publishers, platforms, institutions	Responsibility of whoever transmits (Q 49:6); MUI Fatwa 24/2017 on social media conduct	AI involvement not labeled; amplification by engagement
Reception and correction	Councils, regulators, readers	NU rule against practicing from AI answers (NU Online, 2023a); MUI stress on scholars and sanad (NU Online, 2025a)	No channel for redress and no record of authorship

Studies of practice point to the same boundaries. Ahmed (2024) notes that chatbots offer accessibility, consistency, and volume but lack empathy and nuance. Yiğit and Şimşek (2025) report inconsistencies and sectarian blind spots when AI answers family-related religious questions are compared with official fatwas. Ab Rahim et al. (2025) and Usmonov (2025) recommend supervised use, with ultimate authority and accountability kept with human scholars. In Indonesia, the 2023 Nahdlatul Ulama assembly also recommended building an AI supplied with content by authorized people (NU Online, 2023b), and MUI announced that it would prepare guidance on AI for religious references (NU Online, 2025b). Taken together, the descriptive evidence shows agreement on principle and absence of procedure. Existing legal instruments, including MUI Fatwa 24/2017 and the 2023 circular on AI ethics, speak to online conduct and AI ethics in general terms and do not name religious content as a category. The gap is clearest at the middle and end of the chain. Review is the stage where classical doctrine speaks most directly, yet it is the stage most easily skipped when content is produced quickly. Reception and correction receive almost no attention in the institutional statements, although Ibn Qayyim al-Jawziyyah (1991) wrote at length on how rulings change with circumstances.

4.2. Critical: Factor Analysis

The second finding is that four factors weaken responsibility in practice: opacity and fabrication in the models, the absence of moral agency in the tool, platform scale and incentives, and thin institutional rules for review and remedy.

Put differently, the sources assign responsibility, but the way AI content is produced makes it hard for anyone to pick it up.

The first factor is opacity and fabrication. Language models generate fluent text without a grounded notion of truth, and they can invent sources or misattribute them (Bender et al., 2021; Ji et al., 2023). In religious content, an invented hadith reference or a misquoted verse looks authoritative, and readers seldom check it. This undermines the duty of verification (Q 49:6) at the point where the reader is least able to perform it. The second factor is the absence of moral agency. Ijtihad and fatwa are acts of a responsible agent with intention and accountability before God and people, which AI lacks (Ahmed, 2024; Usmonov, 2025). This does not make responsibility disappear. It means that responsibility must be located in the humans around the tool,

and the many-hands problem makes that hard: developers, prompters, editors, and platforms each contribute a little, and each can point to the others (Santoni de Sio & Mecacci, 2021; Thompson, 1980). The third factor is platform scale and incentives. Engagement metrics reward speed, emotional resonance, and volume rather than scholarly care (Samudra et al., 2024; Wahid, 2024), and AI lowers the cost of producing religious content to nearly nothing (Raya, 2025). A creator who posts AI-drafted material under a personal brand can reach more people than a council can review. The fourth factor is thin institutional rules. Nahdlatul Ulama's ruling and MUI's statements set limits on reliance, but they are not review procedures, and no instrument we found requires labeling, archiving, or correction channels for AI religious content (Ab Rahim et al., 2025; Atallah, 2026).

These factors compound one another. A fabricated citation that no one reviewed can be amplified by a platform, shared by readers who trust the brand of the poster, and left uncorrected because there is no channel for complaint and no record of who prompted or approved it. Responsibility is not denied at any single point. It simply never lands anywhere. In Indonesia the combination is especially visible, because social media preachers with large followings can publish AI-assisted material under their own names, which makes the speaker identifiable but not necessarily the reviewer.

User behavior adds a pressure that sits beside the four factors. Readers often treat a polished answer as settled, which is the habit that the Nahdlatul Ulama ruling targets by forbidding practice on the basis of AI answers (NU Online, 2023a). Studies of online fatwas describe a similar pattern of fatwa shopping, in which people choose the answer they prefer (Wahid, 2024). AI makes this easier, because a user can ask again until the answer agrees with what was wanted.

4.3. Transformative: Implications and Model

The third finding is that the four weaknesses can be addressed by a five-duty model, which we call the Accountable Human Chain (AHC), under which every piece of AI-assisted religious content has a named human owner, a proportionate review, a visible label, a traceable record, and a channel for correction.

In plain terms, the model does not ask whether a machine may speak about religion. It asks who stands behind what is said and what they must do before and after it is published.

Table 3 sets out the five duties. Authorship requires that a named person or institution adopts the content and its claims as its own, which applies the principle that religious speech is a trust (*amanah*) and cannot be left with a tool. Verification requires review by a qualified person in proportion to risk, which applies *tabayyun* and the classical conditions for those who speak on religion. Disclosure requires that AI involvement be stated, in line with the principle of explicability (Floridi & Cowls, 2019) and Atallah's (2026) call for transparency. Traceability requires records of prompts, sources, versions, and reviewers, which functions as a modern counterpart of *sanad* by making the path of a claim visible. Remedy requires a channel through which errors can be reported, corrected, and retracted, with a duty to respond. Two features keep the model workable. First, duties attach to roles, not to technology, so the model does not depend on any particular AI system. A mosque that uses a chatbot to draft Friday sermons and a media company that generates short videos owe the same five duties. Second, the duties can be met at different scales. A small mosque committee can keep a shared log and name its imam as reviewer, while a large platform can build review queues and public correction pages.

Table 3. The Accountable Human Chain (AHC) model

Duty	What it requires	Main bearer	Basis
Authorship	A named person or institution adopts the content and its claims	Publisher, preacher, council	Religious speech as a trust (amanah); Q 7:33
Verification	Review by a qualified person in proportion to risk	Reviewers, editors, fatwa bodies	Tabayyun (Q 49:6); conditions of the mufti
Disclosure	AI involvement is stated to readers	Publisher, platform	Explicability (Floridi & Cowls, 2019); transparency (Atallah, 2026)
Traceability	Records of prompts, sources, versions, and reviewers	Publisher, platform, developer	Modern counterpart of sanad; archiving (Atallah, 2026)

Duty	What it requires	Main bearer	Basis
Remedy	Channel to report, correct, and retract errors, with a duty to respond	Councils, publishers, platforms	Accountability principle; protection of religion and intellect

The model scales duties to risk, as Table 4 shows. Informational content, such as prayer times or definitions, needs spot checks and disclosure. Explanatory content, such as tafsir summaries or sermon drafts, needs review by a qualified person before release. Advisory or fatwa-like content, such as rulings on worship, family, or finance, should be issued only by a qualified mufti or recognized fatwa body, with AI limited to retrieval and drafting support. This tiering matches the positions of Ab Rahim et al. (2025) and Usmonov (2025) and respects the Nahdlatul Ulama and MUI limits on treating AI answers as rulings. It also answers the maqasid concern: the protection of religion and intellect calls for the strictest care where error is most costly.

Table 4. Risk tiers for AI-assisted religious content

Tier	Examples	Minimum duties	Role of AI
1. Informational	Prayer times, definitions, event notices	Authorship, disclosure, spot checks	Drafting and retrieval
2. Explanatory	Tafsir summaries, sermon drafts, study notes	All five duties; review by a qualified person before release	Drafting under review
3. Advisory or fatwa-like	Rulings on worship, family, finance	All five duties; issued only by a qualified mufti or recognized fatwa body	Retrieval and comparison support only

The implications are practical. Religious councils can adopt the model as an internal standard and as a certification basis for publishers and platforms. Publishers

and preachers can apply it as a workflow. Developers can support it by documenting training sources and refusing to present output as a fatwa. Regulators can reference it when drafting rules, for example by adding religious content to the guidance on AI ethics. The model also names what it does not do. It does not certify that AI output is correct, and it does not replace the judgment of a qualified scholar. It makes the chain of answerability visible, so that when an error occurs, someone can be asked to explain and correct it.

5. Discussion

The study set out to find who is responsible at each stage of producing AI-generated Islamic content, what weakens that responsibility, and how it can be rebuilt. The main findings are that Islamic and institutional sources assign responsibility to humans but mainly for fatwas, that four factors spread responsibility thin, and that the AHC model restores it through five duties scaled to risk. The overall lesson is that the problem is not a missing doctrine. It is a missing structure that connects an existing doctrine to a new production process. It also suggests that debates framed around permission and prohibition leave a practical space unaddressed, namely how AI-assisted content should be handled when it is allowed.

The Indonesian context explains why that structure matters. A large Muslim public, strong religious organizations, and active social media make Indonesia a place where AI religious content will spread quickly. Nahdlatul Ulama and MUI have set clear limits on treating AI as a source of rulings, and both stress the role of scholars and *sanad* (NU Online, 2023a, 2025a). Yet limits on reliance are not procedures for production. A limit tells a reader what not to do, while a procedure tells a publisher what it must do. The AHC model is meant to supply the second kind of rule beside the first, and it fits the existing strengths of Indonesian religious institutions, which already run councils, review committees, and certification schemes. Strong religious organizations are an advantage, because they can issue standards that carry moral weight even without legal force. Another contextual point is language and ownership. Much AI-generated Indonesian religious content comes from models developed abroad and trained on mixed sources, so local religious bodies have limited influence over how the tools behave. Nahdlatul Ulama's recommendation to build an AI supplied with authoritative content (NU Online, 2023b) can be read as a response to

that dependence, and it makes the traceability duty especially relevant for any locally built system.

Theoretically, the findings link two literatures. In AI ethics, the responsibility gap describes situations where no one can be held answerable for what a learning system does (Matthias, 2004; Santoni de Sio & Mecacci, 2021). Islamic accountability norms offer a way to close it for religious speech: because the Qur'an addresses speakers and holds each person answerable for what they say (Q 17:36), the tool's lack of agency moves responsibility onto the people who adopt its output rather than dissolving it. Maqasid al-shari'ah adds purpose. If religion and intellect are to be protected (Auda, 2008; Ibn Ashur, 2006), then a system that lets unreviewed, fluent error circulate as religious guidance defeats the objective even if each person involved acted in good faith. The framework also explains why disclosure matters morally and not only technically. A reader who knows that content was machine-assisted can adjust trust and seek verification, which respects the reader's own duty to verify (Q 49:6). Hiding AI involvement removes that opportunity and shifts a burden the reader cannot bear.

A further insight concerns the framing of the debate. Much discussion asks whether AI can issue fatwas, and the answer in the sources is no, because a fatwa requires a responsible human agent. That question, however, covers only a small part of the content AI produces. The more useful question is who answers for AI-assisted content that is not framed as a fatwa but still shapes belief and practice. The classical tradition offers wisdom here: the mufti was cautious because he signed on behalf of God, and caution rose with the stakes. Risk tiering applies the same instinct to a new setting. Seen this way, the classical insistence that the mufti consider circumstances before answering is not an obstacle to technology. It is a standard that technology must meet. Where a system cannot consider the questioner's situation, a human must.

Compared with earlier work, the findings agree that AI cannot replace scholars and that human oversight is essential (Ab Rahim et al., 2025; Ahmed, 2024; Usmonov, 2025; Yiğit & Şimşek, 2025). They extend Atallah's (2026) accountability mechanisms, namely identity verification, AI disclosure, archiving, and traceability, by assigning them to specific stages and actors and by adding duties for developers and users, as well as a risk-based scale. They also respond to the limit noted in the literature that guidelines stay general. The findings differ from approaches that treat the question as a yes or no ruling on AI, since the sources support a graded answer: permitted

assistance with fixed human accountability. A further difference is scope. Studies that test chatbots on questions of family law or inheritance show what goes wrong in advisory answers, yet our tiers suggest that lower-risk content deserves attention too, since errors in a sermon draft or a study note can circulate widely under a trusted name.

The implications run in three directions. For religious councils, the model offers a way to turn existing statements into standards that publishers and platforms can follow. For regulators, it suggests that guidance on AI ethics should name religious content and require disclosure and correction channels. For universities and pesantren, it points to training for reviewers who must evaluate AI output and not only write original texts. The model should also be read as a minimum. Councils that already apply stricter internal rules can keep them, and the duties here are meant to be a shared floor that publishers and platforms can be asked to meet. Two cautions apply. Review duties must not become a gatekeeping device that excludes legitimate voices, and disclosure rules must be written so that they do not stigmatize every use of AI, such as translation or search. Both are matters of design that councils can refine in practice. A final implication concerns research. The model yields testable propositions: that content with named reviewers contains fewer fabricated references, that labeled content is checked more often by readers, and that councils with correction channels resolve errors faster. Indonesia's mix of state guidance and community authority makes it a good place to try them.

6. Conclusion

This article has argued that responsibility for AI-generated Islamic content stays with humans, and that the problem is how to make that responsibility visible and enforceable across the production chain. Reading five stages on four dimensions, we found that Islamic and institutional sources assign accountability to speakers but address mainly the fatwa, that opacity, missing moral agency, platform scale, and thin institutional rules weaken accountability, and that the five-duty Accountable Human Chain model can restore it. Each duty rests on norms the tradition already contains, which is why the proposal is modest in its means and substantial in its effect. Each duty is simple enough to explain to a mosque committee or a content team. The findings also correct a common framing. The question is not whether AI is permitted

to speak about religion, since the sources support assisted use under conditions. The question is whether people accept answerability for what is said in their name.

Theoretically, the study joins Islamic accountability norms with the responsibility-gap literature in AI ethics, showing that a tool's lack of moral agency relocates responsibility onto the humans who adopt its output. It also extends the discussion from the fatwa to the wider range of AI-assisted religious content. A further theoretical point is that the model treats responsibility as a set of duties that can be audited, not as a feeling of blame. That makes it usable by institutions that must decide in advance who will check, who will sign, and who will respond to complaints. Methodologically, the five-stage, four-dimension grid offers a way to compare sources and to locate where duties fade. Its limits are plain. The study is conceptual, draws institutional positions from published news reports, and did not test AI systems or interview practitioners. The findings therefore describe what duties the sources support, not how often they are met.

Practically, religious councils, publishers, platforms, and developers can adopt the model as a workflow and a certification basis, and regulators can align AI ethics guidance with it. Future research should test the model empirically, for instance by auditing a sample of AI religious content for fabricated citations and unlabeled AI use, interviewing preachers and editors about their review practices, and comparing the model's tiers with how councils actually treat different content. Comparative work with other Muslim-majority countries would show which duties travel well, and researchers should include non-Sunni settings so that the model does not rest on one tradition. Pilot use by a council, a university, or a media outlet would show where the duties are too heavy or too light, and the tiers could then be adjusted. Developers have a part to play as well, by documenting data sources and by designing systems that decline to present themselves as muftis. Machines will keep writing about religion. The question that remains is whether a person, by name, answers for what they write.

7. References

[Author's note, delete before submission: every entry should be checked against Scopus, Crossref, or the publisher page, and recent Scopus-indexed journal articles (2022 to 2025) should be added to raise their share of the list.]

Ab Rahim, S. F., Ab Rahman, M. F., Thaidi, H. A. A., Azimi, N. N. M. A. N. M., & Jailani, M. R. (2025). Artificial intelligence for fatwa issuance: Guidelines and

ethical considerations. *Journal of Fatwa Management and Research, 30*(1), 76-100. https://doi.org/10.33102/jfatwa.vol30no1.654

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). https://doi.org/10.25159/2957-9163/14400

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

Atallah, F. A. (2026). Digital mediation and fatwa authority in contemporary Islam: A critical Islamic legal and media-theoretical framework. *Religions, 17*(3), Article 350. https://doi.org/10.3390/rel17030350

Auda, J. (2008). *Maqasid al-shariah as philosophy of Islamic law: A systems approach*. International Institute of Islamic Thought.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. https://doi.org/10.1145/3442188.3445922

Bunt, G. R. (2003). *Islam in the digital age: E-jihad, online fatwas and cyber Islamic environments*. Pluto Press.

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.). (2021). *Digital religion: Understanding religious practice in digital media* (2nd ed.). Routledge.

detikHikmah. (2025, August 20). Ketua MUI: AI tak bisa dijadikan rujukan hukum agama. https://www.detik.com/hikmah/khazanah/d-8070039/ketua-mui-ai-tak-bisa-dijadikan-rujukan-hukum-agama

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Hamdani, A. R. (2023). Fatwa in the digital age: Online muftī, social media, and alternative religious authority. *Hikmatuna: Journal for Integrative Islamic Studies, 9*(1), 53-63. https://doi.org/10.28918/hikmatuna.v9i1.966

Hamdeh, E. (2020). Shaykh Google as Ḥāfiẓ al-ʿAṣr: The internet, traditional ʿulamā', and self learning. *American Journal of Islam and Society, 37*(1-2), 67-102. https://doi.org/10.35632/ajis.v37i1-2.851

Hastuti, H., Mubarak, A. I., & Andika, R. (2026). Implementasi konsep tabayyun dan fatwa MUI dalam perancangan sistem verifikasi fakta era kecerdasan buatan. *Teknosains: Media Informasi Sains dan Teknologi, 20*(2), 265-273. https://doi.org/10.24252/teknosains.v20i2.70591

Hjarvard, S. (2008). The mediatization of religion: A theory of the media as agents of religious change. *Northern Lights: Film & Media Studies Yearbook, 6*, 9-26. https://doi.org/10.1386/nl.6.1.9_1

Hjarvard, S. (2011). The mediatisation of religion: Theorising religion, media and social change. *Culture and Religion, 12*(2), 119-135. https://doi.org/10.1080/14755610.2011.579719

Ibn al-Ṣalāḥ, ʿU. (1986). *Adab al-muftī wa-l-mustaftī* (M. ʿA. ʿA. al-Qādir, Ed.). Maktabat al-ʿUlūm wa-l-Ḥikam.

Ibn Ashur, M. al-T. (2006). *Treatise on maqasid al-shari'ah* (M. E.-T. El-Mesawi, Trans.). International Institute of Islamic Thought.

Ibn Qayyim al-Jawziyyah, M. (1991). *Iʿlām al-muwaqqiʿīn ʿan rabb al-ʿālamīn* (M. ʿA. Ibrāhīm, Ed.). Dār al-Kutub al-ʿIlmiyyah.

Ismail, F., & Seedat, F. (2017). Gender asymmetry and mutual sexual relations in online legal interpretation: Beyond the dissonance through the fatwas of askimam.org. *Journal of Gender and Religion in Africa, 23*(2), 95-124.

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

Majelis Ulama Indonesia. (2017). *Fatwa Nomor 24 Tahun 2017 tentang Hukum dan Pedoman Bermuamalah melalui Media Sosial*. MUI.

Maravia, U., Bekzhanova, Z., Ali, M., & Alibri, R. (2021). British Muslims caught amidst FOGs: A discourse analysis of religious advice and authority. *Religions, 12*(2), Article 140. https://doi.org/10.3390/rel12020140

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*(3), 175-183. https://doi.org/10.1007/s10676-004-3422-1

Menteri Komunikasi dan Informatika Republik Indonesia. (2023). *Surat Edaran Nomor 9 Tahun 2023 tentang Etika Kecerdasan Artifisial*.

NU Online. (2023a, September 19). Munas NU 2023: Bertanya ke AI boleh, tapi haram dijadikan pedoman untuk diamalkan. https://nu.or.id/nasional/munas-nu-2023-bertanya-ke-ai-boleh-tapi-haram-dijadikan-pedoman-untuk-diamalkan-HvEJ3

NU Online. (2023b). Munas NU 2023 rekomendasikan PBNU bikin AI. https://www.nu.or.id/nasional/munas-nu-2023-rekomendasikan-pbnu-bikin-ai-mluw3

NU Online. (2025a, November 4). MUI: Teknologi AI hanya pendamping, tak bisa gantikan peran ulama dan sanad keilmuan. https://www.nu.or.id/nasional/mui-teknologi-ai-hanya-pendamping-tak-bisa-gantikan-peran-ulama-dan-sanad-keilmuan-iEFOG

NU Online. (2025b, November 18). MUI siapkan pedoman pemanfaatan AI untuk rujukan keagamaan. https://nu.or.id/nasional/mui-siapkan-pedoman-pemanfaatan-ai-untuk-rujukan-keagamaan-DHf7z

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

Samudra, I., Laini, Z. W., Yanti, J., Amin, S., & Pakpahan, F. N. (2024). Algorithmic authority and public trust: A bibliometric analysis of YouTube's recommendation system in shaping Islamic religious authority (2013-2023). *Cognoscere: Jurnal Komunikasi dan Media Pendidikan, 2*, 35-50. https://doi.org/10.61292/cognoscere.286

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology, 34*, 1057-1084. https://doi.org/10.1007/s13347-021-00450-x

Sati, A., Halim, A., Nasution, A. H., & Ridwan, M. (2025). The digital transformation of tafsir and its implications for Islamic legal derivation in the contemporary era. *MILRev: Metro Islamic Law Review, 4*(1), 389-415. https://doi.org/10.32332/milrev.v4i1.10425

Shuhufi, M., Fatmawati, F., & Qadaruddin, M. (2022). Islamic law and social media: Analyzing the fatwa of Indonesian Ulama Council regarding interaction on digital platforms. *Samarah: Jurnal Hukum Keluarga dan Hukum Islam, 6*(2), 823-843. https://doi.org/10.22373/sjhk.v6i2.15011

Šisler, V. (2011). Cyber counsellors: Online fatwas, arbitration tribunals and the construction of Muslim identity in the UK. *Information, Communication & Society, 14*(8), 1136-1159. https://doi.org/10.1080/1369118X.2011.579140

Skovgaard-Petersen, J. (1997). Fatwas in print. In M. K. Masud, B. Messick, & D. S. Powers (Eds.), *Islamic legal interpretation: Muftis and their fatwas* (pp. 281-297). Harvard University Press.

Stępniak, K. (2023). Communicating the sacred in religious advertising in light of the mediatization of religion theory and research on digital religion. *Church, Communication and Culture, 8*(2), 285-307. https://doi.org/10.1080/23753234.2023.2244537

Thompson, D. F. (1980). Moral responsibility of public officials: The problem of many hands. *American Political Science Review, 74*(4), 905-916.

Triantoro, D. A., Husna, F., Syahrani, R. A., Syam, F., & Fitri, A. (2023). Adaptation of popular culture in digital fatwa on social media. *Islam Realitas: Journal of Islamic and Social Studies, 9*(2), 169-185. https://doi.org/10.30983/islam_realitas.v9i2.7607

Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik, sebagaimana diubah dengan Undang-Undang Nomor 19 Tahun 2016 dan Undang-Undang Nomor 1 Tahun 2024 (Indonesia).

Usmonov, M. (2025). From human scholars to AI fatwas: Media, ethics, and the limits of AI in Islamic religious communication. *Al-i'lam: Journal of Contemporary Islamic Communication and Media, 5*(1). https://doi.org/10.33102/jcicom.vol5no1.125

Wahid, S. H. (2024). Research insights on online fatwas: A comprehensive systematic literature review. *Journal of Fatwa Management and Research, 29*(1), 23-46. https://doi.org/10.33102/jfatwa.vol29no1.570

Yahya, I., & Junaidi, A. A. (2021). Authority of fatwa in digital era: Methodology and transmission of Nahdlatul Ulama's bahtsul masail in Indonesia. *Webology, 18*(2), 654-674. https://doi.org/10.14704/WEB/V18I2/WEB18345

Yiğit, Y., & Şimşek, V. (2025). Content analysis of AI's responses to family-related religious questions: A comparative study with the fatwas of the Supreme Council of Religious Affairs. *Şırnak University Journal of Divinity Faculty, 37*, 58-73. https://doi.org/10.35415/sirnakifd.1695901

Yucel, S., & Albayrak, I. (2021). Are online fatwas credible? A critical analysis of fifty fatwas on organ donation and transplantation. *Religions, 12*(10), Article 902. https://doi.org/10.3390/rel12100902