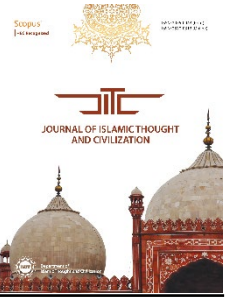
Journal of Islamic Thought and Civilization (JITC)
Volume 16 Issue 1, Spring 2026
ISSN$_{(P)}$: 2075-0943 ISSN$_{(E)}$: 2520-0313
Homepage: https://journals.umt.edu.pk/index.php/JITC



Title:	**Artificial Reasoning and Islamic Law: Mapping LLM Capabilities Across Juristic Ranks and Reasoning Frameworks**
Author (s):	Abdullah Omran[1], Said Hassan[2], and Wesam M. Hassan[3]
Affiliation (s):	[1]Indiana University Bloomington, United States of America [2]American University of Sharjah, United Arab Emirates [3]Ahram Canadian University, Egypt
DOI:	https://doi.org/10.32350/jitc.161.12
History:	Received: July 15, 2025, Revised: March 09, 2026, Accepted: March 10, 2026, Published: May 11, 2026
Citation:	Omran, Abdullah., Said Hassan, Wesam M. Hassan. "Artificial Reasoning and Islamic Law: Mapping LLM Capabilities Across Juristic Ranks and Reasoning Frameworks." *Journal of Islamic Thought and Civilization* 16, no. 1 (2026): 196–211. https://doi.org/10.32350/jitc.161.12
Copyright:	© The Authors
Licensing:	This article is open access and is distributed under the terms of Creative Commons Attribution 4.0 International License
Conflict of Interest:	Author(s) declared no conflict of interest

A publication of
Department of Islamic Thought and Civilization, School of Social Science and Humanities
University of Management and Technology, Lahore, Pakistan

Artificial Reasoning and Islamic Law: Mapping LLM Capabilities Across Juristic Ranks and Reasoning Frameworks

Abdullah Omran
Department of Middle Eastern Languages and Cultures,
Indiana University Bloomington,
United States of America

Said Hassan*
Department of Arabic and Translation Studies,
American University of Sharjah,
United Arab Emirates

Wesam M. Hassan
Faculty of Mass Communication,
Ahram Canadian University,
Egypt

Abstract

The growing use of Artificial Intelligence (AI) in religious inquiry raises a question about how these systems can contribute to Islamic legal reasoning. The current study aimed to examine whether Large Language Models (LLMs) can approximate the forms of analysis that Islamic jurisprudence expects from its practitioners. Islamic legal tradition has identified ranks of competence, from the independent *mujtahid* to the mufti and to the *muqallid*, each with expectations in terms of language, evidence, and judgment. The study assessed LLM performance in relation to these expectations by examining their training data, reasoning patterns, and results of empirical evaluations. Recent models show improved results on structured and rule-based tasks, such as inheritance calculations, where fine-tuning and retrieval-augmented methods improve accuracy and reduce hallucination rates. They can classify concepts, organize texts, and reproduce deductive patterns found in juristic writings. These strengths support LLMs as juristic research tools. Important limitations remain. LLMs cannot judge the strength of evidence, reconcile competing juristic views, or distinguish between equivocal and unequivocal proofs. They treat textual inputs without considering evidentiary hierarchy and therefore lack calibration for the assessment of proofs. Most critically, they cannot access what this study calls the *passive contextual element* in juristic reasoning that relies on awareness of intention, social custom, social conditions, and consequences of a ruling, none of which LLMs can reliably infer or evaluate. The study concluded that LLMs hold an intermediate position. They assist juristic research through retrieval, organization, and structured reasoning, yet they cannot assume the epistemic or ethical responsibilities that shape Islamic legal judgment.

***Keywords*:** artificial reasoning, fatwa, *fiqh,* generative Artificial Intelligence (AI), *ijtihad*, juristic reasoning, Large Language models (LLMs), *mujtahid*

Introduction

"Knowledge is part and parcel of religion; so, exercise caution from whom you are learning" said Muḥammad ibn Sīrīn (d. 110/729).[1] This widely circulated statement becomes newly relevant

*Correspondence concerning this manuscript should be addressed to Said Hassan, Associate Professor at Department of Arabic and Translation Studies, American University of Sharjah, United Arab Emirates at sfhassan@aus.edu

[1]Muḥammad ibn ʿAbd Allāh al-Khaṭīb al-Tabrīzī (d. 741/1340), *Mishkāt al-Maṣābīḥ* [The Niche of Lamps], ed. Muḥammad Nāṣir al-Dīn al-Albānī, 3rd ed., vol. 1 (Bayrūt: al-Maktab al-Islāmī, 1985), 90.

when rephrased: "…be mindful from *which* you are learning," with the word 'which' here designating learning from a non-human agent, namely an Artificial Intelligence (AI)-powered system, such as a Large Language Model (LLM). Ibn Sīrīn's insight emphasizes the learning source because an individual's faith depends on the reliability of religious information taken from a given source. This presupposes a level of competence in the individual who evaluates that source. Once the source's veracity is confirmed, the information received is granted a high degree of confidence. Individuals exhibit diverse competency levels across the Islamic sciences, particularly in the jurisprudential domain where fatwas are issued. Assessment criteria and credibility thresholds consequently vary. An expert mufti possesses the training to evaluate whether a legal opinion aligns with textual sources and the established juristic methodology; a non-expert relies on less rigorous standards. It is therefore essential to provide an accurate assessment of LLMs as they relate to the production of religious and legal knowledge. This is because this forms the basis of how Islamic knowledge is calibrated and presented in response to fatwa-related queries.

Current research on LLMs in Islamic domains focuses on fatwa question answering and inheritance calculations, yet this research did not assess if these models follow the epistemic structure of Islamic legal theory. Fatwa and Hajj datasets emphasize extraction rather than reasoning.[2] Recent benchmark evaluations show persistent hallucinations and weak performances on identifying the *ʿillah* (the ruling's operative/effective cause) and applying *qiyās* (analogical deduction).[3] These limitations leave an unresolved question about whether the reasoning of current LLMs aligns with the methodological expectations of Islamic legal theory. The current study addressed that gap by examining LLM reasoning alongside the established ranks of juristic competence and by evaluating the extent to which LLM reasoning methods resemble, support, or diverge from authoritative legal judgment.

The purpose of this study was to evaluate how contemporary LLMs relate to the established hierarchy of juristic reasoning and to determine the scope of their potential role in issuing fatwas. The study proceeded by outlining the structure and training of LLMs, examining classical juristic qualifications, and analyzing the reasoning patterns displayed by current models. This approach clarifies the limits and possible uses of LLMs in religious settings and highlights the areas of legal judgment that remain exclusive to qualified human jurists.

Methodologically, the study adopted a qualitative comparative approach. It first reviewed the architecture, training data, and reasoning mechanisms of contemporary LLMs described in recent AI research. It then examined classical juristic literature to outline the hierarchy of legal competence and the methodological requirements governing Islamic legal reasoning. Finally, it evaluated how current models perform in relation to these criteria by drawing on benchmark studies, empirical evaluations, and documented model behavior in legal tasks.

2. What are Large Language Models (LLMs)?

LLMs are AI systems trained on massive amounts of text to understand and generate human language. They are built using the transformer architecture, an attention mechanism that allows the

[2]Ohoud Alyemny, Hend Al-Khalifa, and Abdulrahman Mirza, "A Data-Driven Exploration of a New Islamic Fatwas Dataset for Arabic NLP Tasks," *Data* 8, no. 10 (2023): 155, https://doi.org/10.3390/data8100155 ; Hayfa A. Aleid, and Aqil M. Azmi, "Hajj-FQA: A Benchmark Arabic Dataset for Developing Question-Answering Systems on Hajj Fatwas," *Journal of King Saud University – Computer and Information Sciences*, July 25, 2025, https://doi.org/10.1007/s44443-025-00128-w

[3]Ezieddin Elmahjub et al., "IslamicLegalBench: Evaluating LLMs' Knowledge and Reasoning of Islamic Law Across 1,200 Years of Islamic Pluralist Legal Traditions," *arXiv* preprint arXiv:2602.21226 (2026), https://arxiv.org/abs/2602.21226

model to focus on the relevant parts of the input when generating output.[4] This enhances the processing and generation of language more efficiently. While early LLMs were massive in size, the term now refers more broadly to models, large or small, that use these techniques to perform language tasks, such as answering questions, classifying text, or writing coherent paragraphs.[5] Elmahjub notes that AI may speed legal interpretation and support deeper analysis across large legal corpora, while also identifying new links between legal opinions. He also warns that several constraints limit this potential, including uneven Arabic and English data, incomplete digitized collections, the difficulty of processing unstructured texts, and the ethical and methodological demands of issuing fatwa.[6] Despite their strengths, LLMs come with real challenges. Because they are trained on large and often unfiltered datasets, they can reflect and even amplify biases found in the data. They may generate text that sounds correct but is factually wrong, and they can be used to produce harmful or misleading content.[7]

3. How Does a Mufti, a Muslim Jurist, and a Large Language Model (LLM) Reason?

This question presupposes a divergence of expertise within the same field. The three categories, muftis, Muslim jurists, and LLMs, differ in their command of information, both in quantity and diversity. The more extensive and principled the information they possess, the greater the potential for delivering better-informed assessments and more refined conclusions. Beyond just a body of subject-specific knowledge or a collection of texts, these information types encompass an awareness of legal literature genres and familiarity with relevant religious, social, and political contexts. This breadth of knowledge substantially enhances the robustness and quality of any judgments rendered, as it accommodates multiple layers of consideration that would otherwise escape those lacking in these fields, either partially or entirely. Ultimately, the reasoning process involves understanding, synthesizing, interpreting, evaluating, and rendering a judgment on a particular issue using this comprehensive information base. Accordingly, it is crucial to elucidate the distinctions and qualifications between jurists and muftis to demonstrate how the process of legal reasoning varies when employed by scholars of differing legal expertise.

3.1. Training[8]

Islamic legal scholarship has always been intimately bound to the purpose and evolving form of its textual productions. Juristic writings are deeply connected with methodological commitments, legal principles, and pedagogical imperatives. The organizational elements of Islamic jurisprudence (*fiqh*) literature are not just academic; they actively shape legal development and hierarchies of reasoning and authority.

Foundational epistemological principles, such as *sadd al-dharāʾiʿ* (preemptive restriction), *istishāb* (presumption of continuity), and *ʿurf* (customary practice) profoundly impact the structure

[4]Ashish Vaswani, et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems* 30 (2017): 5998–6008, https://doi.org/10.48550/arXiv.1706.03762

[5]Ibomoiye Domor Mienye, et al., "Large Language Models: An Overview of Foundational Architectures, Recent Trends, and a New Taxonomy," *Discover Applied Sciences* 7, no. 1027 (2025), https://doi.org/10.1007/s42452-025-07668-w

[6]Ezieddin Elmahjub, "From Manuscripts to Digital Corpus: Structuring Islamic Data Sources for the Future of AI Jurisprudence," *IslamicLaw.app* Blog, March 27, 2025, https://islamiclaw.blog/ (accessed February 22, 2026).

[7]Jay Alammar, and Maarten Grootendorst, *Hands-on Large Language Models: Language Understanding and Generation*, 1st ed. (Sebastopol, CA: O'Reilly Media, 2024), 25.

[8]Large part of this section is informed by: Haytham ibn Fahd Rūmī, *Al-Ṣiyāghah al-fiqhīyah fī al-ʿaṣr al-ḥadīth: dirāsah taʾṣīlīyah* [Contemporary Juristic Discourse: A Foundational Study], al-Ṭabʿah al-Ūlā (al-Riyāḍ: Dār al-Tadmurīyah, 2014).

and aims of juristic writings. The application of *sadd al-dharāʾiʿ*, especially in the Mālikī school, demands legal reasoning that incorporates societal implications and preventative logic. Juristic literature shaped by this doctrine, particularly *nawāzil* (contemporary cases) works like al-Wansharīsī's (d. 914/1508) *al-Miʿyār al-Muʿrib*, training jurists to assess real-life cases with foresight and ethical sensitivity.

Variation in the levels of *ijtihād* across the schools of law shapes the nature of juristic writing. The Ḥanafī and Mālikī schools, which allow layered forms of *ijtihād,* such as *tarjīḥ* (evidential preponderance) and *taḥqīq al-manāṭ* (context-tied legal reasoning), produced *mukhtasarāt* (concise abridgements of foundational texts) and commentaries that combine doctrine with evidential analysis, as seen in al-Marghīnānī's d. 593/1197) *al-Hidāyah*. The Shāfiʿī school, which is stricter in defining *ijtihād*, favors works with strong textual fidelity and structured logic, such as al-Nawawī's (d. 676/1277) *al-Majmūʿ*. Fears of juristic incoherence led to corrective compilations, such as al-Mawṣilī's (d. 683/1284) *al-Mukhtār li-l-Fatwā*, which preserve rulings and mark the limits of legitimate opinions. Across time periods, authorship shifts from early integrative reasoning in works, such as Mālik's (d. 179/795) *al-Muwaṭṭaʾ*, al-Shāfiʿī's (d. 204/820) *al-Risālah*, and Aḥmad ibn Ḥanbal's (d. 241/855) *al-Musnad* to more formalized pedagogy in Ibn Qudāmah's (d. 620/1223) *ʿUmdat al-Fiqh*, *al-Muqniʿ*, and *al-Mughnī*, as well as in al-Nawawī's progression from *al-Minhāj* to *Rawḍat al-Ṭālibīn*.

Writings address distinct audiences, which shapes format and complexity. For lay readers, Ibn Bālbān's (d. 1083/1672) *Mukhtaṣar al-Ifādāt* offers accessible rulings. For beginners, works, such as Ibn Abī Zayd's (d. 386/996) *al-Risālah al-Fiqhiyyah* and *al-Tuḥfah* cover the essentials. For jurists and muftīs, Ibn Rushd's (d. 595/1198) *Bidāyat al-Mujtahid* maps intra-madhhab divergences and chains of evidence. For judges, manuals, such as Ibn Farḥūn's (d. 799/1397) *Tabsirat al-Ḥukkām* and Ibn ʿĀṣim's (d. 829/1426) *Tuḥfat al-Ḥukkām* provide procedural guidance on testimony, judicial ethics, and evidence. *Mutūn* (foundational legal primers) and *mukhtasarāt* support memorization but risk oversimplification without commentary, which prompted scholars to produce expansive *shurūḥ* (explanations of foundational texts) like al-Bājī's (d. 474/1082) commentary on *al-Mudawwanah* and *ḥawāshī* (marginal glosses) like Ibn ʿĀbidīn's (d. 1252/1836) *Radd al-Muḥtār,* both aim to restore nuances and contexts.

This integration of textual form and juristic function shows a finely structured pedagogy. Jurists gain doctrinal mastery through exposure to texts suited to their epistemic status and institutional role. Each type of juristic writing rests on assumptions about the audience, authority, and legal reasoning. Legal authorship is tied to the legal authority within each school.[9] Al-Subkī (d. 756/1355) identified three tiers: the theorist jurist who analyzes the *fiqh* corpus, the mufti who issues practical rulings with attention to context, and the judge who combines the mufti's expertise with procedural knowledge that governs adjudication.[10] A more detailed hierarchy appears in al-Zuḥaylī's (d. 1436/2015) six levels: the independent jurist who establishes the school's methodology, the bounded-independent jurist who operates within that methodology, the restricted jurist who extends the school's logic into unaddressed areas, the jurist of preference who evaluates conflicting views, the jurist of fatwa who applies the school's rulings without full internal mastery, and the follower who adheres to established positions but does not have the capacity for independent reasoning.[11]

[9]Ahmed Fekry Ibrahim, "Rethinking the *Taqlīd* Hegemony," *Journal of the American Oriental Society* 136, no. 4 (2017), https://doi.org/10.7817/jameroriesoci.136.4.0801

[10]ʿAlī ibn ʿAbd al-Kāfī al-Subkī, *Fatāwā al-Subkī* [Fatwas of al-Subkī] (Bayrūt: Dār al-Maʿrifah, 1950).

[11]Wahbah al-Zuḥaylī, *Al-Fiqh al-Islāmī wa-Adillatuh* [Jurisprudence in Islam and Its Evidence], 2nd ed. (Dimashq: Dār al-Fikr, 1984), 1:46–48.

This hierarchy reflects varied reasoning capacities depending on knowledge and application. A mujtahid is one who deduces rulings from their sources while meeting conditions that include the mastery of Arabic, knowledge of Qurʾānic and hadith sciences, understanding of the objectives of the *Sharīʿah*, familiarity with consensus and secondary legal discussions, proficiency in analogical reasoning and its conditions, awareness of historical and social contexts, as well as recognition of one's own competence and moral integrity.[12] These requirements show that the mujtahid functions as an active epistemological agent. Each juristic category is shaped by the limitations of its expertise, which affects their authority and judgment. The restricted jurist understands the internal logic of the school, while the jurist of fatwa applies existing rulings without developing new analogies. Al-Nawawī and Ibn al-Ṣalāḥ similarly categorized mufti qualifications and allowed the possibility of domain-specific muftis with narrower competence.[13] Al-Nawawī, citing al-Juwaynī, held that even a skilled scholar of *uṣūl al-fiqh* (foundations of Islamic jurisprudence) cannot issue fatwas without practical expertise, since he may not fully grasp the incident at hand or the situational considerations that govern the application of rulings.[14]

Across the four Sunni schools, legal reasoning begins with classifying scriptural evidence as *qaṭʿī* (unequivocal), or *ẓannī* (equivocal) with regard to the chain of transmission and the meaning.[15] Unequivocal evidence, such as Qurʾānic verses with explicit rulings and also widely transmitted hadiths, yields definitive outcomes that require no further inference. Equivocal evidence, whether due to solitary transmission or semantic ambiguity, allows interpretive flexibility and the use of tools, such as *qiyās, istiḥsān* (juristic preference), and *maṣlaḥah mursalah* (public interest).[16] Each school aligns these tools to the strength of the proof. Some limit legal extension when evidence is unequivocal, while others permit analogy or juristic preference when the text leaves uncertainty. Only jurists with the authority to classify such texts and determine when equivocal evidence may be extended or overridden may employ these tools.[17] This hierarchy regulates who interprets equivocal texts and under which conditions.

The Ḥanafī school emphasizes rational deduction centered on *qiyas*, at times preferring it over solitary hadiths that lack corroboration. It builds rulings through identifying the *ʿillah* and through structured case evaluation, while permitting *istiḥsān* and *ʿurf* (custom) when consistent with the aims of *Sharīʿah*.[18] The Mālikī school gives substantial weight to local custom and public interest and

[12]Muḥammad Abū Zahra, *Uṣūl al-Fiqh* [Principles of Islamic Jurisprudence] (al-Qāhirah: Dār al-Fikr al-ʿArabī, 1973), 34–38.

[13]Abū-Zakarīyyā Muḥyi al-Dīn al-Nawawī, *Al-Majmūʿ Sharḥ al-Muhadhdhab* [The Compendium in the Explanation of the (Shāfiʿī) School] (Bayrūt: Dār al-Fikr, 1996); ʿUthmān ibn ʿAbd al-Raḥmān Ibn al-Ṣalāḥ, *Adab al-Muftī wa-al-Mustaftī* [Guidelines for the Jurisconsult and the Questioner], 1st ed. (Bayrūt: Maktabat al-ʿUlūm wa-al-Ḥikam, 1986).

[14]al-Nawawī, *Al-Majmūʿ Sharḥ al-Muhadhdhab*, vol. 1, 44.

[15]Muḥammad Sallām Madkūr, *Manāhij al-Ijtihād fī al-Islām fī al-Aḥkām al-Fiqhīyyah wa-al-ʿAqāʾidīyyah* [Methodologies of Ijtihād in Islam in Legal and Doctrinal Matters], 1st ed. (Kuwait: Jāmiʿat al-Kūwayt, 1973).

[16]Wael B. Hallaq, *A History of Islamic Legal Theories: An Introduction to Sunni Usul al-Fiqh* (Cambridge: Cambridge University Press, 1997).

[17]Abū al-Maʿālī ʿAbd al-Malik ibn ʿAbd Allāh al-Juwaynī (Imām al-Ḥaramayn), *al-Ghiyāthī: Ghiyāth al-Umam fī Ilyāth al-Ẓulam* [The Succor of Nations in Times of Tyranny], 1st ed. (Bayrūt: Dār al-Kutub al-ʿIlmiyyah, 1997).

[18]Muḥammad Sallām Madkūr. *Manāhij al-Ijtihād fī al-Islām fī al-Aḥkām al-Fiqhīyyah wa-al-ʿAqāʾidīyyah*.

favors inductive reasoning that seeks ethical coherence.[19] The Shāfiʿī school requires precision in textual grounding and allows *qiyās* only when a clear *ʿillah* is identified from valid texts and when the extension remains tied to Qurʾān, *Sunnah,* or consensus, avoiding tools that are judged as being speculative. The Ḥanbalī school prioritizes transmission by giving primacy to Qurʾān, authentic hadīth, and Companions' opinions, allowing analogy only when textual guidance is absent and under strict limits.[20]

Comparative evaluation of LLMs shows a structural asymmetry, since these models do not distinguish between *qaṭʿī* and *ẓannī* evidence (unless they are spelled out explicitly and textually). They also cannot access the institutional memory or ethical accountability that governs the use of *istiḥsān* or *maṣlaḥah*. The relational and contextual dimensions of legal judgment also remain inaccessible to current systems, including those which use retrieval or chain-of-thought prompting. This is a structural gap produced by the difference between statistical prediction and normatively grounded deliberation, and it frames the distinction explored in the following section.

4. LLM and Legal Reasoning

LLMs do not perform reasoning in the juristic sense but perform statistical reproduction of reasoning-like patterns. Their outputs emerge from probabilistic associations learned from vast corpora, not through epistemic deliberation or normative judgment.[21] This distinction is critical when evaluating if LLMs can meaningfully approximate the reasoning of a mujtahid. Reasoning for LLMs refers to the model's capacity to perform complex text-generation tasks that require multi-step inference, not to conscious deliberation or ethical evaluation. Recent evaluation suites, including PLawBench, assess models on analogical mapping, statutory interpretation, and evidentiary reasoning. Frontier models achieve above 85% accuracy on formally bounded tasks but performance declines sharply when tasks become open-ended, hierarchically complex, or ethically charged, which are precisely the conditions under which Islamic legal reasoning operates.[22]

4.1. Training Datasets

ChatGPT's underlying models illustrate the scale and diversity of modern LLM training. Early deployments relied on GPT-3's 175 billion-parameter architecture, which was pre-trained on roughly 500 billion tokens drawn from filtered Common Crawl, curated web text (WebText-style corpora), large book collections, and English Wikipedia, amounting to tens of terabytes of text.[23] Subsequent generations, such as GPT-4 and later ChatGPT models scaled this paradigm by ingesting far larger

[19] ʿAmr Muḥammad ʿAfīfī, *Ḥarakat al-Naqd al-Fiqhī: Fuṣūl fī Tārīkh al-Ijtihād wa-Masīrat al-Naqd al-Fiqhī* [The Movement of Juristic Critique: Studies in the History of Ijtihād and the Trajectory of Juristic Criticism] (al-Riyāḍ: Markaz Namāʾ li-l-Buḥūth wa-l-Dirāsāt, 2025).

[20] ʿAfīfī, *Ḥarakat al-Naqd al-Fiqhī: Fuṣūl fī Tārīkh al-Ijtihād wa-Masīrat al-Naqd al-Fiqhī.*

[21] Jian-Qiao Zhu, and Thomas L. Griffiths, "Incoherent Probability Judgments in Large Language Models," arXiv preprint arXiv:2401.16646 (2024), https://doi.org/10.48550/arXiv.2401.16646.

[22] Neel Guha et al., *LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models*, arXiv preprint arXiv:2308.11462, 2023, https://arxiv.org/abs/2308.11462

[23] Tom B. Brown et al., "Language Models Are Few-Shot Learners," in *Advances in Neural Information Processing Systems* 33 (2020): 1877–1901.

mixtures of publicly available web data and licensed third-party corpora, then applying Reinforcement Learning from Human Feedback (RLHF) and related alignment techniques.[24]

Non-peer-reviewed reports indicate that GPT-5 continues the trend of extreme scaling, trained on roughly 0.5 quadrillion unfiltered tokens (i.e., hundreds of terabytes of raw text) which were later reduced to tens of trillions of filtered tokens, including substantial synthetic LLM-generated material and licensed publisher content. These reports also place the model's size in the low hundreds of billions of parameters.[25] Google's Gemini 3 Pro, according to its model card, adopts a similarly large-scale but explicitly multimodal approach, training a sparse Mixture-of-Experts' system across text, code, images, audio, and video, with a context window reaching one million tokens. It is further refined through instruction-tuning and RLHF-style preference training.[26]

The training data used for Arabic LLMs depends on which kinds of Arabic texts are available in digital form. Studies of Arabic LLMs report that most models are trained mainly on Modern Standard Arabic (MSA). These models draw on news articles, Wikipedia, web forums, and other edited text, while Classical Arabic and many spoken dialects appear much less often in large datasets.[27] This pattern has a clear effect: models perform well on formal MSA tasks but not as well with older writing styles and everyday speech used in different regions. Benchmark scores fall when models are tested on dialectal or pre-modern Arabic.[28]

Islamic material is entered into Arabic LLMs using more specialized collections because general web crawls do not contain enough reliable religious text. One major source is the OpenITI corpus, which gives machine-readable versions of many classical Islamic works, including *fiqh*, ḥadīth, and early commentaries.[29] Another pathway is the creation of models trained on legal and religious documents. AraLegal BERT is trained on a 4.5 GB collection of legal texts that includes legislation, court cases, and Islamic legal writings, and it performs better than general Arabic BERT models on legal tasks.[30]

Some systems focus on *fatwā* question answering. KAB, which stands for Knowledge Augmented BERT2BERT, uses hundreds of thousands of *fatwā* question and answer pairs taken from official online sources. This gives researchers a large, supervised dataset for Islamic legal question answering and chatbot design.[31] Recent surveys also note that newer Arabic-focused large

[24]Dimitris Karydas et al., "Training Methods for Large Language Models: Current Approaches and Challenges," *Technologies* 14, no. 2 (2026): 133, https://doi.org/10.3390/technologies14020133

[25]"What's in GPT-5's Dataset?" Benched.ai, https://benched.ai/guides/gpt-5-dataset

[26]"Google DeepMind and Google Cloud," "*Gemini 3 Pro Model Card"* Accessed March 2. https://deepmind.google/models/model-cards/gemini-3-1-pro/

[27]Shahad Al Khalifa et al., "The Landscape of Arabic Large Language Models," *Communications of the ACM* 68, no. 10 (2025): 54–61, https://doi.org/10.1145/3737453

[28]Abdessalam Bouchekif et al., "Assessing Large Language Models on Islamic Legal Reasoning: Evidence from Inheritance Law Evaluation," in *Proceedings of the Third Arabic Natural Language Processing Conference* (Suzhou, China: Association for Computational Linguistics, 2025), 246–257, https://doi.org/10.18653/v1/2025.arabicnlp-main.20

[29]Lorenz Nigst et al., OpenITI: A Machine-Readable Corpus of Islamicate Texts, Version 2021.1.4, dataset, 2020, https://doi.org/10.5281/zenodo.4513723

[30]Muhammad Al-Qurishi et al., "AraLegal-BERT: A Pretrained Language Model for Arabic Legal Text," in *Proceedings of the Natural Legal Language Processing Workshop 2022* (Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022), 338–344, https://doi.org/10.18653/v1/2022.nllp-1.31

[31]Saud S. Alotaibi et al., "KAB: Knowledge Augmented BERT2BERT Automated Questions-Answering System for Jurisprudential Legal Opinions," *International Journal of*

models, such as ALLaM, AceGPT, ArabianGPT, and Fanar are trained on expanded Arabic text collections that include Qurʾān, ḥadīth, *tafsīr*, and *fatwā* archives. Some of these models also use Retrieval Augmented Generation, which lets the models look up classical or modern Islamic sources while they answer questions. This is important because Islamic texts are detailed and context-sensitive and are not fully represented in standard training data.[32]

4.2. Reasoning Capabilities and Current Performance

LLMs now resemble reasoning agents, yet much of this apparent capability arises from patterned reproduction rather than the internal manipulation of propositions which a jurist or logician would recognize. Evidence describes this behavior as pattern-driven, such that accuracy is impaired and explanations become less reliable when tasks move away from familiar textual forms that the models encountered in training.[33] This suggests that successful reasoning often reflects the recognition of trained forms rather than a construction of new legal or logical structure. A broad survey of chain-of-thought prompting adds that step-by-step rationales can improve answers on some benchmarks but the very same rationales shift as a result of small prompt changes or distribution shifts. This instability indicates that what looks like deduction is frequently linguistic association rather than the application of stable rules.[34] The limitation becomes the clearest in causal reasoning. This is because current systems largely reuse causal phrases learned during training rather than forming new causal analyses, and that performance drops when models face fresh causal problems or counterfactuals.[35]

Arabic language models inherit these strengths and weaknesses with added linguistic complexity. The COLING AraDiCE benchmark shows that Arabic-centric systems outperform multilingual baselines for dialect and culturally-grounded tasks, especially where usage remains close to Modern Standard Arabic. This reflects better handling of regional expressions and sociolinguistic context.[36] Linguistic strength, however, does not guarantee deeper legal inference. A study of Saudi commercial court decisions found that a general-purpose model outperformed an

Computer Science and Network Security 22, no. 6 (2022): 346–356. https://doi.org/10.22937/IJCSNS.2022.22.6.44.

[32]Ummar Abbas et al., "Fanar: An Arabic-Centric Multimodal Generative AI Platform," arXiv preprint arXiv:2501.13944, 2025, https://doi.org/10.48550/arXiv.2501.13944

[33]Nagraj Naidu, and Omar F. El Gayar, "A Review of Reasoning in Artificial Agents Using Large Language Models," *in Proceedings of the 58th Hawaii International Conference on System Sciences* (HICSS 58) (2025). https://scholar.dsu.edu/bispapers/440 (accessed March 3, 2026)

[34]Zheng Chu et al., "Navigate through Enigmatic Labyrinth: A Survey of Chain-of-Thought Reasoning," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (August 11–16, 2024), 1173–1203, https://aclanthology.org/2024.acl-long.65.pdf

[35]Haoang Chi et al., "Unveiling Causal Reasoning in Large Language Models: Reality or Mirage?" in *Advances in Neural Information Processing Systems* 37 (Proceedings of the 38th Conference on Neural Information Processing Systems, NeurIPS 2024) (Neural Information Processing Systems Foundation, 2024), https://proceedings.neurips.cc/paper_files/paper/2024/file/af2bb2b2280d36f8842e440b4e275152-Paper-Conference.pdf

[36]Basel Mousi et al., "AraDiCE: Benchmarks for Dialectal and Cultural Capabilities in LLMs," in *Proceedings of the 31st International Conference on Computational Linguistics* (Abu Dhabi, UAE: Association for Computational Linguistics, 2025), 4186–4218, https://aclanthology.org/2025.coling-main.283/

Arabic-specialized system by a wide margin, which points to the importance of model architecture and reasoning depth for legal analytics, beyond surface-level language alignment.[37]

The contrast between patterned competence and genuine inference becomes sharpest in Islamic inheritance law. Here the rules are explicit and arithmetic, and frontier models perform very well on standardized multiple-choice scenarios drawn from classical sources. A 2025 ArabicNLP evaluation reports accuracies over 90% for leading frontier systems.[38] The same benchmark and its shared-task overview also show that Arabic-adapted open models cluster below 50%, typically due to misapplication of foundational constraints, errors in identifying heirs, and confusion about the hierarchy that governs exclusion or redistribution. These results suggest that when *fiqh* problems resemble stable templates with limited branching, models can exploit pattern regularity. As soon as the legal chain requires school-specific maxims or the consideration of exceptions, performance drops. This is consistent with the broader observation that these systems depend on the recognition of familiar textual configurations rather than the construction of new juristic reasoning on demand.

Retrieval-augmented generation offers a practical way to stabilize answers and reduce unsupported claims in Islamic domains. A *fatwa* system that uses a two-stage retrieval pipeline with a re-ranker shows measurable reductions in hallucinations and improvements in answer stability due to anchoring responses in authoritative sources, such as Dar al-Iftāʾ archives. This has the effect of turning freeform generation into a citation-guided synthesis, which benefits tasks that require verifiable grounding.[39] In the ArabicNLP shared tasks, a retrieval-augmented pipeline reached high accuracy on general Islamic knowledge, thereby demonstrating the value of coupling generation with evidence selection when the goal is correctness rather than stylistic fluency.[40] These gains matter for production settings because they make outputs easier to audit and curate. They also create a workflow where scholars can quickly trace the basis of an answer back to recognized sources and flag any unsupported steps.

The lesson from these evaluations is straightforward. LLMs are valuable assistants for Islamic studies when tasks are regular, rules are explicit, and evidence is plentiful. They help researchers find passages, normalize terminology across sources, and compute structured outcomes, such as inheritance shares. They are less reliable when a question requires abductive reasoning concerning causes, analogical transfer that depends on an identified cause, or probabilistic assessment in cases of uncertainty. In those zones, the same properties that make models fluent also make them unstable. Retrieval, fine-tuning, and careful prompting improve reliability but do not replace the methodical work a jurist performs when weighing texts, exceptions, and principles. The prudent stance is therefore to deploy these systems as accelerators and synthesizers, with domain experts setting the frame, checking the steps, and making the rulings.

[37]Adel Ammar et al., "Prediction of Arabic Legal Rulings Using Large Language Models," *Electronics* 13, no. 4 (2024): 764, https://doi.org/10.3390/electronics13040764

[38]Abdessalam Bouchekif et al., "QIAS 2025: Overview of the Shared Task on Islamic Inheritance Reasoning and Knowledge Assessment," in *Proceedings of the Third Arabic Natural Language Processing Conference: Shared Tasks* (Suzhou, China: Association for Computational Linguistics, 2025), 851–860, https://doi.org/10.18653/v1/2025.arabicnlp-sharedtasks.117

[39]Marryam Yahya Mohammed et al., "Aftina: Enhancing Stability and Preventing Hallucination in AI-Based Islamic Fatwa Generation Using LLMs and RAG," *Neural Computing and Applications* 37 (2025): 20957–20982, https://doi.org/10.1007/s00521-025-11229-y

[40]Mohamed Samy et al., "Tokenizers United at QIAS-2025: RAG-Enhanced Question Answering for Islamic Studies by Integrating Semantic Retrieval with Generative Reasoning," in *Proceedings of the Third Arabic Natural Language Processing Conference: Shared Tasks* (Suzhou, China: Association for Computational Linguistics, 2025), 960–965, https://doi.org/10.18653/v1/2025.arabicnlp-sharedtasks.133

5. LLMs and Islamic Legal Reasoning

LLMs reproduce legal reasoning by generating responses that follow patterns in the legal texts and styles they have encountered. When given a legal problem, a model may supply a textbook definition that appears to apply legal rules to facts, yet its method does not rely on first principles. The model identifies scenarios that resemble those in its training data and responds in a similar way. Researchers describe this as pattern matching. Others list it as fact finding, rule interpretation, legal qualification, and decision-making as core reasoning tasks, and GPT-4 is reported to simulate these tasks convincingly.[41]

Domain specific LLMs offer more reliable performance. Generalist models often hallucinate when prompts require knowledge beyond their training data, and they may generate fabricated details. Deductive reasoning works most consistently for clear Islamic legal questions, since explicit Qurʾānic verses and widely transmitted hadiths appear often in training. Inductive pattern matching also functions well for established *fiqh* categories. Abductive, causal, and analogical reasoning that depends on identifying the operative cause remains less reliable because it requires conceptual and contextual precision that statistical methods cannot capture. Models trained on classical Arabic legal corpora perform better on structured tasks but the epistemological gap between their output and genuine Islamic legal reasoning remains large.

Three types of challenges prevent LLMs from performing real juristic reasoning. Text-based challenges arise from the lack of comprehensive and well annotated corpora. Knowledge-based challenges stem from the discipline's demands, including mastery in the knowledge of hadith transmitters, evidentiary evaluation, and intra school methodology. Even if a model internalized these fields, it would still need to adjust to each school's practice and to the circumstances of each question. Reasoning-based challenges are the most serious because Islamic legal reasoning requires intentional judgment and entails shouldering the responsibility towards the questioner, which a statistical system cannot supply.

5.1. Passive Contextual Elements

Due to the spatial-temporal and situational nature of fatwa, it relies on factors that lie beyond the technicalities of textual information. These include the mufti's acquaintance with the questioner's intention, understanding the prevalent social norms, and gauging the immediate or foreseeable consequences. Furthermore, it also includes taking into consideration the degree of necessity, the level of customary allowance, and the readiness of the inquirer to understand and implement the fatwa. Dispensing fatwa depends on determinants which are not available in the toolkit of LLMs unless such information is fed as prompting caveats to elicit a specific answer. This presumes the questioner is aware of the contextual factors a mufti must consider before engaging a question and providing a thoughtful, well-informed response, which is not always the case, if at all. Consider, for instance, the following statements by classical jurists, quoted and paraphrased in full, respectively.

> Whenever a customary practice evolves, it should be acknowledged and adopted; conversely, practices that have gone out of use ought to be abandoned. Scholars must resist an uncritical attachment to the textual rulings recorded in classical works they have studied throughout their lives. Instead, when an individual from a different region requests a legal opinion, the jurisconsult should first ascertain the prevailing norms of that individual's locale, apply those norms in formulating the ruling, and issue the

[41] Li Zhang et al., "Thinking Longer, Not Always Smarter: Evaluating LLM Capabilities in Hierarchical Legal Reasoning," *arXiv* preprint arXiv:2510.08710 (2025), https://doi.org/10.48550/arXiv.2510.08710

> fatwa accordingly, regardless of the practitioner's own regional custom or the established precedents in authoritative texts.[42]

The other states:

> Questions may arise between qualified jurists (mujtahids) or lay students (muqallids) in four configurations: scholar to scholar, student to student, scholar to student, and student to scholar. Peer exchanges serve to confirm understanding, resolve uncertainties, and refine reasoning. When a scholar questions a student, the aim is to assess comprehension or draw on the student's insights; when a student queries a jurist, the exchange is the primary avenue for acquiring legal knowledge. In the first three cases, the respondent is generally obliged to answer if he possesses the relevant knowledge. In the fourth case, the duty to respond is qualified: an answer is required if the jurist is certain of the matter, the questioner is capable of understanding, and the inquiry does not demand undue hardship. If the question is speculative, beyond the student's capacity, or excessively technical, a full reply is not obligatory.[43]

These two statements point to a form of situated knowledge rooted in socio-epistemic context and practical awareness of circumstances required for issuing religious verdicts. This traditional framework has shifted with technologies from radio to online fatwa platforms, yet such platforms still recognize the contextual nature of fatwa production. The passive recognition of the underlying context makes tailored rulings difficult for LLMs, especially when custom-based responses are needed. However, there are ways to compensate for some missing information. LLM-driven systems can become more reliable, factual, and context aware through methods, such as Retrieval Augmented Generation (RAG) and Explainable AI (XAI). These approaches reduce hallucination, supply real-world information, clarify reasoning, and add domain specific rules.

RAG allows the model to consult an external knowledge base during a query. It retrieves relevant passages and uses them to form an answer grounded in verified evidence. For religious systems, the model can draw Qur'anic verses, hadith reports, or prior fatwas and summarize them rather than relying only on its internal memory. This reduces hallucination and increases trust. One caution is that the system remains dependent on the quality of the retriever. If irrelevant material is retrieved, the model may incorporate it and reduce the accuracy of the response.

However, this is significantly important that RAG's source-based response places substantial weight on the inquirer because it leaves the question of source verification and relevance to the user's judgment. In a way, it resembles a hypothetical case where the mufti reports several sources to the individual and lets him decide which to choose for the decision-making process. So, even when the system provides a source-based answer, the response still lacks the reasoning capabilities to judge the evidentiality relevance of the sources which is derived mainly from the semantic relevance of the text. Since the model draws its material from online repositories, it also inherits the biases created by online visibility. Websites that attract more traffic or publish large volumes of content, such as Islamweb, are more likely to appear in the retrieved results. Their prominence on the internet makes them frequent candidates in RAG-assisted outputs, not because they are more authoritative, but because they are more visible.

XAI for LLM reasoning is yet another method to tackle reasoning deficiencies. Explainable AI refers to techniques that make the AI's decision process more transparent. For LLMs, which are black-box neural nets, this is challenging. However, one straightforward approach is to have the model generate explanations or reasoning steps in natural language, often called "chain-of-thought". Rather than merely stating the answer, the model can be prompted to first outline its reasoning and

[42]Aḥmad ibn Idrīs Qarāfī and Qāsim ibn al-Shāṭṭ, *Al-Furūq* [The Differences Separating Legal Rules], vol. 1 (Dār Iḥyāʾ al-Kutub al-ʿArabīyah, 1929), 129.

[43]Ibrāhīm ibn Mūsá ibn Muḥammad al-Shāṭibī, *Al-Muwāfaqāt* [The Reconciliations], 2nd ed., vol. 5 (al-Qāhirah: Dār Ibn ʿAffān, 2006).

then give a conclusion. Chain-of-thought prompting augments the model's answer with an explanation to ensure that the reasoning process is transparent for users.[44] This process is valuable because it allows for the identification of errors, such as wrongly cited principles.

Although XAI cannot supply the passive contextual information a mufti relies on, such as intention, social conditions, or local custom, it can help clarify how the model handled the textual evidence it was given. A mufti evaluates not only the sources but also the circumstances surrounding the questioner. XAI cannot recreate this sensitive contextual awareness but it can show which sources shaped the model's answer and how the model weighed them. This supports scholarly oversight by making the model's internal steps visible, which helps reviewers trace how different pieces of evidence were used and determine whether the reasoning aligns with accepted interpretive methods.

6. Conclusion

As this study has shown, LLMs neither fit neatly into a single juristic rank nor fully replicate any human role within the Islamic legal hierarchy. Instead, they occupy an intermediate space, capable of simulating certain analytical tasks yet lacking the deeper judgment, context-sensitivity, and ethical responsibility that underpin genuine *ijtihād* and *fatwa* issuance. LLM capabilities have been mapped onto the four principal tiers of juristic competence. These are as follows:

1. Theorist-Jurist (Analytical Corpus), Partial Simulation: LLMs operate in the domain of syntactic and semantic pattern recognition. This gives LLM the ability to simulate aspects of the theorist-jurist's work, such as recognizing legal principles, parsing maxims, and classifying legal concepts found in their training data. Their strength lies in summarizing and emulating the structure of juristic texts. As Spaić and Jovanović demonstrated, GPT-4 can perform key legal reasoning functions (fact-finding, interpretation, classification, and justification) in a textually convincing manner, though this ability is derived from statistical pattern-matching rather than principled interpretation.[45]

2. Mufti (Contextual Responsa), Inadequate Simulation: LLMs are unable to assume the epistemic, ethical, or contextual burden required of a mufti. While they can mimic fatwa forms and produce rulings resembling real-world religious opinions, they lack the ability to weigh *maqāṣid* (objectives of *Shariʿah*), consider situational variables, or exercise moral discretion. Local norms, lived experiences, personal histories, and political realities remain beyond the scope of text-trained models. Without a mechanism to perceive or interact with such elements, the LLM's response cannot achieve the normative intentionality required for a genuine fatwa.

3. *Mujtahid al-Muqayyad* / *Mujtahid al-Tarjīḥ* – Unsatisfactory Emulation: These jurists study competing intra-*madhhab* views, reconcile contradictory precedents, and identify the stronger opinion. LLMs lack meta-legal judgment, the ability to rank views based on evidence strength, school coherence, or historical development. They may present multiple positions (if such multiplicity appears in the data) but they do not evaluate them. They do not prioritize *qaṭʿī* over *ẓannī*, nor do they show systematic awareness of textual gradation or the reliability of chain-of-transmission (*isnād*) reliability. This reflects the absence of epistemic calibration within the

[44]Wei Jie Yeo et al., "How Interpretable Are Reasoning Explanations from Prompting Large Language Models?" in *Findings of the Association for Computational Linguistics: NAACL 2024* (Mexico City, Mexico: Association for Computational Linguistics, 2024), 2148–2164, https://doi.org/10.18653/v1/2024.findings-naacl.138

[45]Bojan Spaić, and Miodrag Jovanović, "Artificial Reason and Artificial Intelligence: The Legal Reasoning Capabilities of GPT-4," *Anali Pravnog fakulteta u Beogradu* 72, no. 3 (2024): 383–422, https://doi.org/10.51204/Anali_PFBU_24302A

model, since LLMs treat all memorized texts as equally plausible unless explicitly prompted otherwise.

4. *Muqallid*/Lay Interpreter, Surpassing but Non-Independent: In terms of information retrieval and recall, LLMs far exceed the lay interpreter or muqallid. They can summarize hundreds of works, reference classical positions, and present plausible legal opinions rapidly. However, their outputs remain dependent: they do not originate new *ijtihād*, nor do they verify their answers through source-critical methods.

In conclusion, this study showed that LLMs occupy an intermediate place in Islamic legal reasoning. They can support research by gathering material, organizing texts, and handling structured tasks, which allows scholars to focus on interpretation and context. However, they still have definite limits. LLMs do not judge evidence, do not weigh competing views, and do not recognize the difference between equivocal and unequivocal proofs. They also cannot access intention, custom, or social conditions, which are essential for fatwa production and for the practice of ijtihad. For these reasons, they cannot take the role of a mufti or a mujtahid. Their proper use is as supervised tools with substantially growing potential to assist with retrieval, classification, and preliminary analysis while leaving authoritative judgments to qualified jurists. Continued evaluation of their performance on specific juristic cases, including inheritance issues, ritual questions, and transactional matters, would help clarify where they can assist and where human expertise is required.

Authors' Contribution

Abdullah Omran: conceptualization, formal analysis, methodology, writing—original draft. **Said Hassan:** conceptualization, methodology, project administration, review and editing original draft. **Wesam Hassan**: conceptualization, and LLM technical discussion.

Conflict of Interest

The authors of the manuscript have no financial or non-financial conflict of interest in the subject matter or materials discussed in this manuscript.

Data Availability Statement

The data produced and analyzed in this study are available from the corresponding author upon request. Persistent links or DOIs to primary sources are provided in the reference list wherever possible. All materials have been curated in accordance with ethical standards and copyright regulations, ensuring transparency, reproducibility, and responsible data use.

Funding Details

No funding has been received for this research.

Generative AI Disclosure Statement

In preparing this manuscript, authors used AI tools only for limited editorial support. The tools were used to correct grammar, improve clarity, and assist with online search for relevant sources. All core ideas, arguments, structure, and interpretations in this manuscript are our own and were not generated by AI.

Bibliography

Abbas, Ummar, et al. "Fanar: An Arabic-Centric Multimodal Generative AI Platform." arXiv preprint arXiv:2501.13944 (2025). https://doi.org/10.48550/arXiv.2501.13944

Abū Zahra, Muḥammad. *Uṣūl al-Fiqh* [Principles of Islamic Jurisprudence]. al-Qāhirah: Dār al-Fikr al-ʿArabī, 1973.

ʿAfīfī, ʿAmr Muḥammad. *Ḥarakat al-naqd al-fiqhī: Fuṣūl fī tārīkh al-ijtihād wa-masīrat al-naqd al-fiqhī* [The Movement of Juristic Critique: Studies in the History of Ijtihād and the Trajectory of Juristic Criticism]. al-Riyāḍ: Markaz Namāʾ li-al-Buḥūth wa-al-Dirāsāt, 2025.

Alammar, Jay., and Maarten Grootendorst. *Hands-On Large Language Models: Language Understanding and Generation*. 1st ed. Sebastopol, California: O'Reilly Media, 2024.

Ammar, Adel, et al. "Prediction of Arabic Legal Rulings Using Large Language Models." *Electronics* 13, no. 4 (2024): 764. https://doi.org/10.3390/electronics13040764.

Aleid, Hayfa A., and Aqil M. Azmi. "Hajj-FQA: A Benchmark Arabic Dataset for Developing Question-Answering Systems on Hajj Fatwas." *Journal of King Saud University – Computer and Information Sciences* (July 25, 2025). https://doi.org/10.1007/s44443-025-00128-w

Benched.ai. "What's in GPT-5's Dataset?" Accessed February 25, 2026. https://benched.ai/guides/gpt-5-dataset

Bouchekif, Abdessalam, et al. "Assessing Large Language Models on Islamic Legal Reasoning: Evidence from Inheritance Law Evaluation." In *Proceedings of the Third Arabic Natural Language Processing Conference*, 246–257. Suzhou, China: Association for Computational Linguistics, 2025. https://doi.org/10.18653/v1/2025.arabicnlp-main.20.

Bouchekif, Abdessalam, et al. "QIAS 2025: Overview of the Shared Task on Islamic Inheritance Reasoning and Knowledge Assessment." In *Proceedings of the Third Arabic Natural Language Processing Conference: Shared Tasks*, 851–860. Suzhou, China: Association for Computational Linguistics, 2025. https://doi.org/10.18653/v1/2025.arabicnlp-sharedtasks.117.

Brown, Tom B., et al. "Language Models Are Few-Shot Learners." In *Advances in Neural Information Processing Systems* 33 (2020): 1877–1901.

Chi, Haoang, et al. "Unveiling Causal Reasoning in Large Language Models: Reality or Mirage?" In *Advances in Neural Information Processing Systems* 37. Neural Information Processing Systems Foundation, 2024. https://proceedings.neurips.cc/paper_files/paper/2024/file/af2bb2b2280d36f8842e440b4e275152-Paper-Conference.pdf.

Chu, Zheng, et al. "Navigate through Enigmatic Labyrinth: A Survey of Chain-of-Thought Reasoning." In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1173–1203. August 11–16, 2024. https://aclanthology.org/2024.acl-long.65.pdf.

Elmahjub, Ezieddin, et al. "IslamicLegalBench: Evaluating LLMs' Knowledge and Reasoning of Islamic Law Across 1,200 Years of Islamic Pluralist Legal Traditions." *arXiv* preprint arXiv:2602.21226 (2026). https://arxiv.org/abs/2602.21226.

Elmahjub, Ezieddin. "From Manuscripts to Digital Corpus: Structuring Islamic Data Sources for the Future of AI Jurisprudence." *IslamicLaw.app Blog*, March 27, 2025. https://islamiclaw.blog/.

Google DeepMind and Google Cloud. *Gemini 3 Pro Model Card.* 2025. Accessed February 25, 2026. https://ai.google.dev/gemini-api/docs/models

Guha, Neel, et al. *LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models*. *arXiv* preprint arXiv:2308.11462 (2023). https://arxiv.org/abs/2308.11462.

Hallaq, Wael B. *A History of Islamic Legal Theories: An Introduction to Sunni Uṣūl al-Fiqh*. Cambridge: Cambridge University Press, 1997.

Ibn al-Ṣalāḥ, ʿUthmān ibn ʿAbd al-Raḥmān. *Adab al-muftī wa-al-mustaftī* [Guidelines for the Jurisconsult and the Questioner]. 1st ed. Bayrūt: Maktabat al-ʿUlūm wa-al-Ḥikam, 1986.

Ibrahim, Ahmed Fekry. "Rethinking the Taqlīd Hegemony." *Journal of the American Oriental Society* 136, no. 4 (2017). https://doi.org/10.7817/jameroriesoci.136.4.0801.

Al-Juwaynī, Abū al-Maʿālī ʿAbd al-Malik ibn ʿAbd Allāh. *al-Ghiyāthī: Ghiyāth al-umam fī ilyāth al-ẓulam* [The Succor of Nations in Times of Tyranny]. Bayrūt: Dār al-Kutub al-ʿIlmiyyah, 1997.

Karydas, Dimitris, Dimosthenis Margaritis, and Helen C. Leligou. "Training Methods for Large Language Models: Current Approaches and Challenges." *Technologies* 14, no. 2 (2026): 133. https://doi.org/10.3390/technologies14020133.

Al Khalifa, Shahad, et al. "The Landscape of Arabic Large Language Models." *Communications of the ACM* 68, no. 10 (2025): 54–61. https://doi.org/10.1145/3737453.

Ma, Yusong, et al. "A Survey of Retrieval-Augmented Generation (RAG) for Large Language Models." In *2025 International Conference on Trustworthy Big Data and Artificial Intelligence (ICTBAI)*, 7–13. Piscataway, NJ: IEEE, 2025. https://doi.ieeecomputersociety.org/10.1109/ICTBAI68361.2025.00008

Madkūr, Muḥammad Sallām. *Manāhij al-ijtihād fī al-islām fī al-aḥkām al-fiqhiyyah wa-al-ʿaqīdiyyah* [Methodologies of Ijtihād in Islam in Legal and Doctrinal Matters]. 1st ed. Kuwait: Jāmiʿat al-Kuwayt, 1973.

Mienye, Ibomoiye Domor, et al. "Large Language Models: An Overview of Foundational Architectures, Recent Trends, and a New Taxonomy." *Discover Applied Sciences* 7, no. 1027 (2025). https://doi.org/10.1007/s42452-025-07668-w.

Mohammed, Marryam Yahya, et al. "Aftina: Enhancing Stability and Preventing Hallucination in AI-Based Islamic Fatwa Generation Using LLMs and RAG." *Neural Computing and Applications* 37 (2025): 20957–20982. https://doi.org/10.1007/s00521-025-11229-y.

Mousi, Basel, et al. "AraDiCE: Benchmarks for Dialectal and Cultural Capabilities in LLMs." In *Proceedings of the 31st International Conference on Computational Linguistics*, 4186–4218. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2025. https://aclanthology.org/2025.coling-main.283/.

Naidu, Nagraj, and Omar F. El Gayar. "A Review of Reasoning in Artificial Agents Using Large Language Models." In *Proceedings of the 58th Hawaii International Conference on System Sciences*. 2025. https://scholar.dsu.edu/bispapers/440.

Al-Nawawī, Abū-Zakarīyā Muḥyi'd-Dīn. *Al-Majmūʿ Sharḥ al-Muhadhdhab* [The Compendium in the Explanation of the (Shāfiʿī) School]. Bayrūt: Dār al-Fikr, 1996.

Nigst, Lorenz, et al. *OpenITI: A Machine-Readable Corpus of Islamicate Texts*. Version 2021.1.4. Dataset. 2020. https://doi.org/10.5281/zenodo.4513723.

Alotaibi, Saud S., et al. "KAB: Knowledge Augmented BERT2BERT Automated Questions-Answering System for Jurisprudential Legal Opinions." *International Journal of Computer Science and Network Security* 22, no. 6 (2022): 346-356. https://doi.org/10.22937/IJCSNS.2022.22.6.44

Al-Qarāfī, Aḥmad ibn Idrīs., and Qāsim ibn al-Shāṭṭ. *Al-Furūq* [The Differences Separating Legal Rules]. Vol. 1. Bayrūt: Dār Iḥyāʾ al-Kutub al-ʿArabīyah, 1929.

Al-Qurishi, Muhammad., Sarah Alqaseemi, and Riad Souissi. "AraLegal-BERT: A Pretrained Language Model for Arabic Legal Text." In *Proceedings of the Natural Legal Language*

Processing Workshop 2022, 338–344. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. https://doi.org/10.18653/v1/2022.nllp-1.31.

Rūmī, Haytham ibn Fahd. *al-Ṣiyāghah al-fiqhiyyah fī al-ʿaṣr al-Hadīth: Dirāsah taʾṣīliyyah* [Contemporary Juristic Discourse: A Foundational Study]. al-Riyāḍ: Dār al-Tadmuriyyah, 2014.

Samy, Mohamed, et al. "Tokenizers United at QIAS-2025: RAG-Enhanced Question Answering for Islamic Studies by Integrating Semantic Retrieval with Generative Reasoning." In *Proceedings of the Third Arabic Natural Language Processing Conference: Shared Tasks*, 960–965. Suzhou, China: Association for Computational Linguistics, 2025. https://doi.org/10.18653/v1/2025.arabicnlp-sharedtasks.133.

Al-Shāṭibī, Ibrāhīm ibn Mūsá ibn Muḥammad. *al-Muwāfaqāt* [The Reconciliations]. 2nd ed. Vol. 5. al-Qāhirah: Dār Ibn ʿAffān, 2006.

Spaić, Bojan, and Miodrag Jovanović. "Artificial Reason and Artificial Intelligence: The Legal Reasoning Capabilities of GPT-4." *Anali Pravnog fakulteta u Beogradu* 72, no. 3 (2024): 383–422. https://doi.org/10.51204/Anali_PFBU_24302A.

Al-Subkī, Alī ibn ʿAbd al-Kāfī. *Fatāwā al-Subkī* [Fatwas of al-Subkī]. Bayrūt: Dār al-Maʿrifah, 1950.

Al-Tabrīzī, Muḥammad ibn ʿAbd Allāh al-Khaṭīb. *Mishkāt al-Maṣābīḥ* [The Niche of Lamps]. Edited by Muḥammad Nāṣir al-Dīn al-Albānī. 3rd ed. Bayrūt: al-Maktab al-Islāmī, 1985.

Vaswani, Ashish, et al. "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30 (2017): 5998–6008. https://doi.org/10.48550/arXiv.1706.03762.

Yeo, Wei Jie, et al. "How Interpretable Are Reasoning Explanations from Prompting Large Language Models?" In *Findings of the Association for Computational Linguistics: NAACL 2024*, 2148–2164. Mexico City, Mexico: Association for Computational Linguistics, 2024. https://doi.org/10.18653/v1/2024.findings-naacl.138.

Alyemny, Ohoud., Hend Al-Khalifa, and Abdulrahman Mirza. "A Data-Driven Exploration of a New Islamic Fatwas Dataset for Arabic NLP Tasks." *Data* 8, no. 10 (2023): 155. https://doi.org/10.3390/data8100155.

Zhang, Li, et al. "Thinking Longer, Not Always Smarter: Evaluating LLM Capabilities in Hierarchical Legal Reasoning." *arXiv* preprint arXiv:2510.08710 (2025). https://doi.org/10.48550/arXiv.2510.08710.

Zhu, Jian-Qiao., and Thomas L. Griffiths. "Incoherent Probability Judgments in Large Language Models." *arXiv* preprint arXiv:2401.16646 (2024). https://doi.org/10.48550/arXiv.2401.16646.

Al-Zuḥaylī, Wahbah. *al-Fiqh al-islāmī wa-adillatuh* [Islamic Jurisprudence and Its Evidence]. 2nd ed. Dimashq: Dār al-Fikr, 1984.