HalalLLM vs. KosherLLM: When Abrahamic Scholars Confront Modern Questions through AI

Cantay Caliskan[1], Mehmet Karabela[1], Xinran Wang[1], Yiheng Yao[1], and Keming Zhang[1]

University of Rochester, Rochester, NY, USA

Abstract. What if you could ask a 12th-century philosopher about modern society? This study explores that possibility by resurrecting twelve of the most influential scholars in Judaism, Christianity, and Islam through the architecture of large language models (LLMs). We instructed three state-of-the-art models—GPT-4 Turbo, Qwen-Plus, and DeepSeek-R1—to role-play these historical figures, grounding their perspectives in their authentic writings using a retrieval-augmented generation (RAG) pipeline that retrieves a median of eight citations per prompt. These AI-driven scholars then answered 290 questions from the World Values Survey, generating over 50,000 synthetic responses that we benchmarked against contemporary global attitudes. Our empirical analysis reveals that while the models appear similar on the surface, their underlying interpretive styles differ significantly. A repeated-measures ANOVA confirmed that although the models' *mean* outputs were statistically indistinguishable, their *median* ($F(2, 22) = 13.87$, $p < .001$) and *modal* ($F(2, 22) = 9.71$, $p < .001$) responses were not, confirming that a scholar's ideas are represented differently across models (Hypothesis 4). Post-hoc tests specified these differences, showing Qwen's median output was significantly higher than both DSR1 and GPT-4. Furthermore, our findings indicate that models sometimes project a universalized worldview, with Islamic scholars' profiles aligning closer to the global average ($t = -5.77$, $p < .0001$), while at other times capturing regional particularity, with Christian scholars' profiles aligning more with their local norms ($t = 17.86$, $p < .0001$). By prototyping a method to engage with classical religious tradition, this work opens new avenues for digital humanities and highlights the critical ethical and methodological challenges in creating AI systems that can faithfully preserve and transmit centuries of human thought.

Keywords: Large language models · AI and religion · Retrieval-augmented generation · World Values Survey · Abrahamic religions

1 Simulating Abrahamic Scholars with Large Language Models

What if Thomas Aquinas could offer a perspective on gender roles, or Maimonides on global financial regulation? Such questions, once the domain of pure speculation, have now become empirically tractable through modern advances in artificial intelligence. This study undertakes the systematic simulation of twelve of the most influential thinkers in Abrahamic history, moving beyond metaphor to ask how their worldviews—rooted in particular historical moments—might engage with the defining dilemmas of the present. We explore this possibility by using large language models (LLMs) as a vehicle for resurrecting these historically-grounded perspectives, placing them in direct conversation with contemporary global opinion.

To achieve this, we instruct three state-of-the-art architectures—GPT-4 Turbo, DeepSeek-R1, and Qwen-Plus—to role-play canonical scholars spanning the eleventh to twentieth centuries. Grounded in each scholar's authentic corpus via a retrieval-augmented generation (RAG) pipeline that retrieves a median of eight citations per prompt, our AI personae answered all 290 items from Wave 7 of the World Values Survey (WVS). Every question was posed ten times to capture stochastic variation, yielding $12 \times 3 \times 290 \times 5 = 52\,200$ synthetic answers. These responses were then benchmarked against 88,166 human observations from fifty-nine countries, enabling a direct, large-scale comparison between simulated historical thought and modern values.

This project is therefore motivated by a dual purpose. Intellectually, it pioneers a methodology for approximating historically-grounded value systems with quantifiable fidelity, extending research on value alignment and AI interpretability. Practically, it offers a prototype for a new mode of engagement with religious tradition, one where believers might consult the ethical insights of foundational thinkers without the mediation of modern clergy or commentators—potentially enriching personal decision-making while preserving doctrinal nuance.

The remainder of the paper reviews related work, describes data construction and pre-registered hypotheses, details the empirical strategy, presents results, and discusses implications for AI ethics, comparative religion,

and the design of spiritually informed conversational agents. Figure 1 provides a methodological overview of our research.

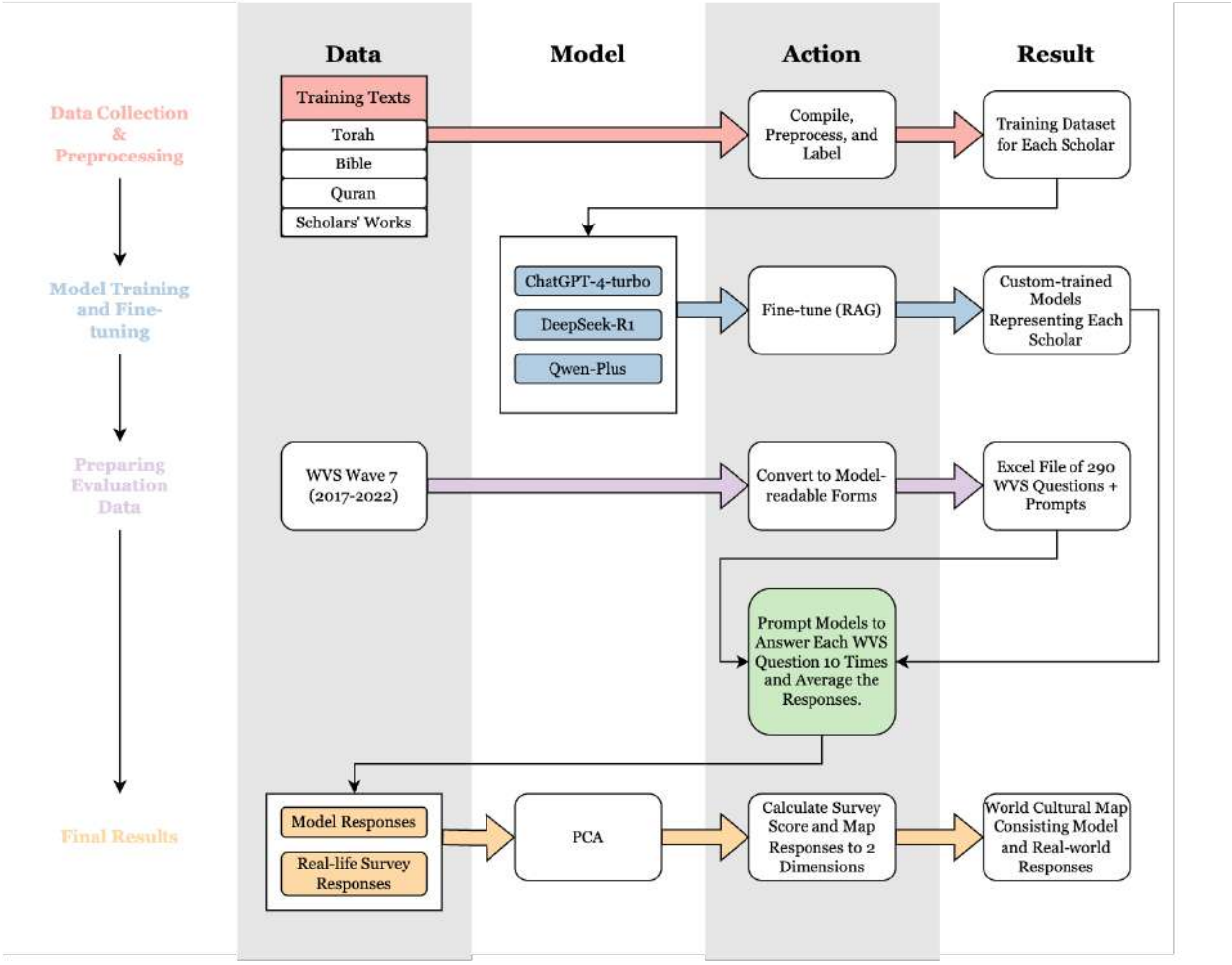

Fig. 1: Research Pipeline

2 The Current Landscape of Religious Chatbots

Having shown how custom-trained LLMs can approximate the voices of canonical scholars, we now turn to the wider ecosystem in which such simulations must be situated. Large language models already mediate a growing share of everyday encounters with sacred texts. By fusing generative capacities with vast digitized corpora, modern chatbots not only answer doctrinal queries and synthesize commentaries but also personalize devotional study in real time. In doing so, they reshape the very mechanisms through which religious authority is accessed, contested, and reproduced in a networked public sphere.

RavGPT integrates recent advances in natural language processing with the Sefaria digital library, a corpus that includes the Torah, Talmud, and *Shulḥan Arukh*. By mapping user queries to canonical passages and commentaries, the system seeks to democratize access to rabbinic literature while retaining contextual fidelity [5].

AskJesus, developed by The Center for Unity, employs ChatGPT to simulate conversational engagement with Christological teachings. The model's responses adopt the tone and doctrinal perspective of *The Urantia Book*, offering users an interactive companion that frames answers within that text's portrayal of Jesus' ministry [24].

Bible AI targets private devotional study by combining an AI-driven search engine with customizable reading plans and an auxiliary repository of more than 80,000 articles and one million videos. These supplementary materials are algorithmically juxtaposed with scriptural passages, creating a multidimensional entry point into Christian exegesis [3].

Chatbot Eli aggregates outputs from multiple LLMs to address roughly 20,000 frequently asked questions on faith, theology, and daily life. Its answers draw on `gotquestions.org` and standard Bible translations, and the system iteratively fine-tunes its parameters on user feedback to improve accuracy and stylistic coherence [6].

Islam&AI adopts an open-source strategy, applying the Universal Sentence Encoder to Quranic verses, hadith corpora, and historical records. With training data that exceed 700,000 hadith records and incorporate multiple multilingual translations and tafsir, the platform enables nuanced retrieval and interpretation across a broad spectrum of Islamic scholarship [13].

ChatILM, maintained by IslamiCity, employs advanced natural-language techniques to support interactive learning in Islamic theology, the Quran and hadith studies, and Muslim history. The interface is designed to accommodate diverse educational backgrounds while adhering to scholarly standards of citation and textual fidelity [14]. Table 1 summarizes existing large language model-based chatbots.

Table 1: Overview of religious chatbots and their features

Name	Year	Religion	Functionality	Base Model	Training Data
RavGPT	2023	Judaism	AI assistant for Torah study	Not specified	Jewish texts including Torah, Talmud, Shulhan Arukh, from Sefaria database
AskJesus	2023	Christianity	AI-powered Q&A on Jesus' teachings	ChatGPT	The Urantia Book, Part IV
Bible AI	2023	Christianity	AI-driven Bible search and study tools	Not specified	Bible
Chatbot Eli	2022	Christianity	Answers questions on faith and life	Not specified	`gotquestions.org`, the Bible
Islam&AI	2022	Islam	Provides Quran and Hadith Q&A using NLP methods	USE and NLP tools	Quran corpus, over 700 000 hadith records, and multiple tafsir datasets
ChatILM	2024	Islam	AI-driven Q&A on Islamic topics	Not specified	Not specified

3 Research Hypotheses

While recent advances illustrate the technical plausibility of large language models that emulate religious voices, scholarly analysis of such systems remains at a formative stage. The foregoing review has outlined the principal architectures and application domains that currently define this emerging landscape. To move from description to explanation, this study now articulates a set of theoretically motivated expectations about how different model configurations and scholarly personae should behave when tasked with generating value-laden responses. These expectations are formalized in the following research hypotheses, which will guide the ensuing empirical analysis.

H1: *Cross-LLM Model Comparison.* This hypothesis posits measurable differences among the outputs of Qwen-Plus, DeepSeek-R1, and GPT-4 Turbo when answering religious queries. Specifically, some models may produce responses that more closely align with contemporary scholarly consensus or prevailing religious interpretations. Testing this hypothesis enables a direct developer-to-developer comparison of faithfulness, accuracy, and topical relevance.

H2: *Cross-Religious Scholar Comparison.* This hypothesis predicts divergence in portrayals across religious traditions. By analyzing AI-generated responses attributed to key figures in Islam, Christianity, and Judaism, the study evaluates how each model negotiates theological, cultural, and historical distinctions—thereby revealing whether each tradition's intellectual and spiritual particularities are captured faithfully and accurately.

H3: *Historical-to-Contemporary Global and Regional Proximity and AI Projection.* This hypothesis anticipates systematic and varied patterns in the proximity of historical scholarly thought to both contemporary global norms and the specific attitudes of present-day populations in their historical regions. Comparisons—such as medieval Baghdad versus modern Iraq or medieval Cairo versus contemporary Egypt—test the model's capacity to integrate historical intellectual contexts with modern social evolution, revealing whether AI responses project these thinkers into a broadly "globalized" discourse or honor contemporary regional particularities.

H4: *Intra-Scholar, Cross-Model Comparison.* This hypothesis proposes that a single scholar's ideas will be represented differently across models. By querying the same scholar with Qwen-Plus, DeepSeek-R1, and GPT-4 Turbo, we conduct an AI-to-AI comparison of interpretive style, bias, and fidelity to the scholar's corpus, thereby revealing model-specific tendencies in reconstructing individual intellectual profiles.

4 Selection Criteria and Biographical Profiles of Twelve Abrahamic Scholars

The selection of scholars for this study was guided by a commitment to capturing the extensive diversity and depth of Abrahamic intellectual traditions. The panel includes twelve influential figures spanning over eight centuries, from the 12th-century philosopher Moses Maimonides to the late 20th-century feminist theologian Rosemary Radford Ruether. These scholars represent three Abrahamic traditions—Judaism, Christianity, and Islam—and cover diverse geographic regions from the Middle East and North Africa to Europe and North America. The intellectual spectrum includes philosophers and jurists such as Thomas Aquinas and Ibn Taymiyya, revolutionary thinkers like Martin Luther and Ali Shariati, and pioneering feminist voices including Hannah Arendt, Rosemary Radford Ruether, and Fatema Mernissi. Gender balance was thoughtfully addressed, ensuring the inclusion of three influential female scholars, thereby acknowledging the crucial yet historically underrepresented contributions of women. This meticulous selection provides a multidimensional foundation for analyzing how historical religious perspectives engage with contemporary ethical dilemmas. An overview of the scholars in our study is provided in Figure 2 and Figure 3.

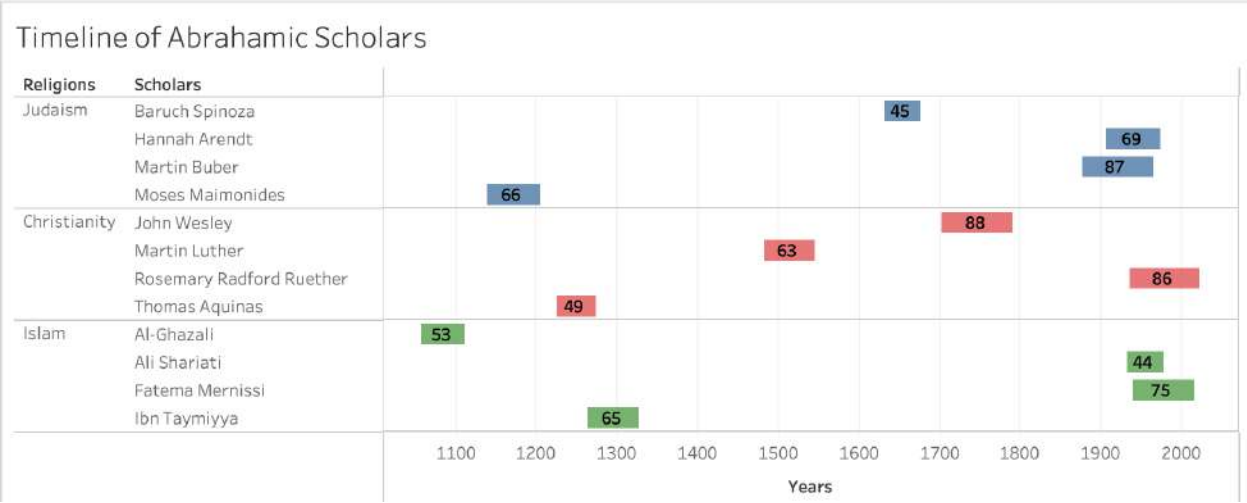

Fig. 2: Lifespans of the profiled Abrahamic scholars (birth–death years).

4.1 Jewish Thinkers

Moses Maimonides (1138–1204) Moses Maimonides (Rabbi Moses ben Maimon) exemplifies the synthesis of religious law and philosophy in medieval Jewish scholarship. He is widely regarded as "the greatest Jewish philosopher of the medieval period" [15]. As both a preeminent *halakhic* authority and a rationalist philosopher, Maimonides had a transformative impact on Judaism. His 14-volume law code, *Mishneh Torah*, established him as "the leading rabbinic authority of his time—and quite possibly of all time" [15]. Equally influential was his philosophical magnum opus, *The Guide of the Perplexed*, which sought to reconcile biblical faith with Aristotelian philosophy. Maimonides's influence extended beyond Judaism; subsequent thinkers in other traditions, including Aquinas in Christianity and Spinoza in the early modern period, engaged with his ideas. His inclusion ensures representation of medieval religious rationalism and the enduring legacy of a scholar who stands at the crossroads of Jewish law, philosophy, and medicine.

Baruch Spinoza (1632–1677) Baruch Spinoza's work represents a significant departure from traditional religious authority within the Jewish intellectual continuum. Although excommunicated from his Sephardic Jewish community, he is now recognized as "one of the most important philosophers—and certainly the most radical—of the early modern period" [21]. Spinoza's thought is distinguished by its naturalism; he conceived of God and Nature as one, challenged scriptural literalism, and laid the intellectual groundwork for Enlightenment critiques of religion. His writings, such as the *Ethics* and the *Theological-Political Treatise*, advance democratic, secular values and a "deep critique of the pretensions of Scripture and sectarian religion" [21]. In the context of this study, Spinoza's inclusion provides a critical, humanistic perspective from within the Jewish tradition. Historically,

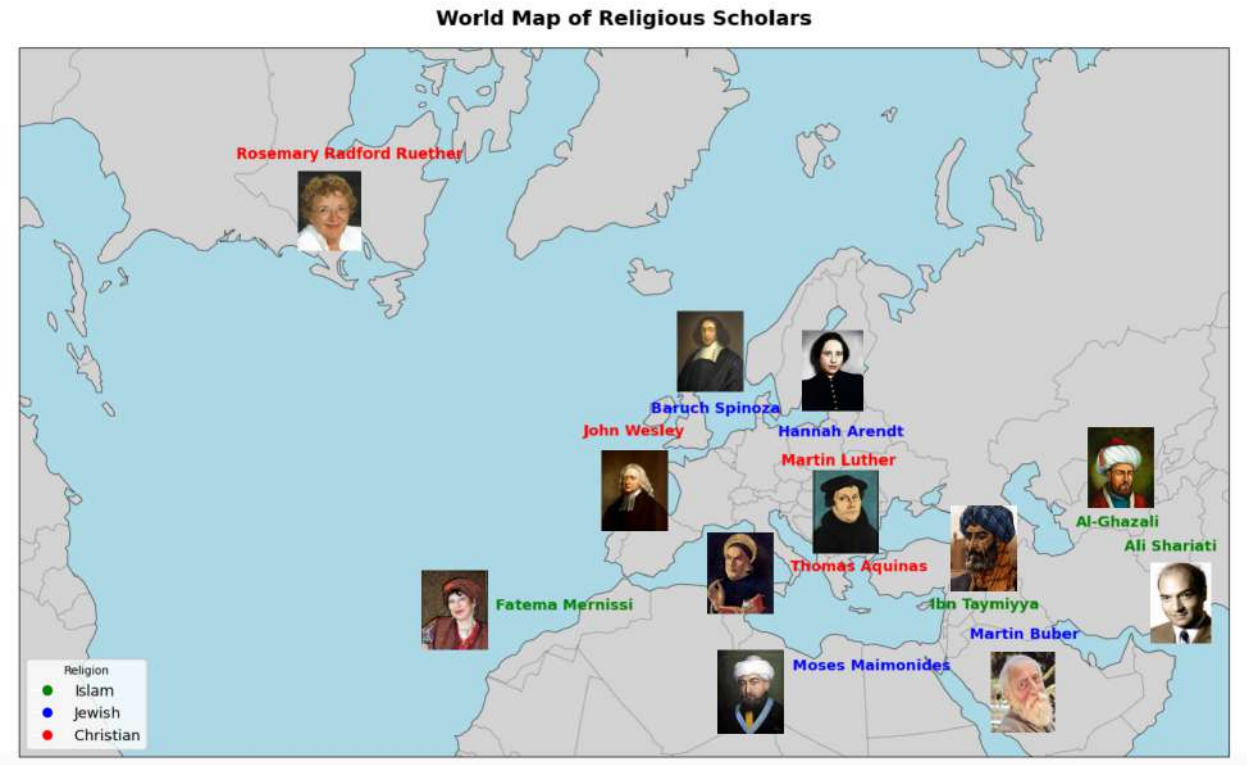

Fig. 3: Global distribution of the scholars' primary residences.

he illustrates the transition from medieval scholasticism to Enlightenment rationality. The enduring relevance of Spinoza is noted by scholars, who state that "of all the philosophers of the seventeenth century, Spinoza is among the most relevant today" [21], making him an ideal interlocutor for modern value questions that require reconciling faith with reason or tradition with liberalism.

Martin Buber (1878–1965) Martin Buber offers a 20th–century Jewish perspective centered on existential and dialogical philosophy. A prolific scholar and activist, Buber's work ranged from mystical writings on *Ḥasidism* to essays on politics and education. He is best known for his book *I and Thou* (1923), which introduced his philosophy of dialogue, distinguishing the "I–Thou" relationship from the "I–It" relationship. Buber became "a major twentieth-century figure in Jewish thought and the philosophy of religion," revered for connecting modern existentialist themes with biblical and Ḥasidic insights [29]. His inclusion adds a theologically rich yet non-dogmatic Jewish perspective. Buber's emphasis on dialogue, interpersonal ethics, and community addresses modern concerns about relational values and social cohesion. Furthermore, his life spanned Europe and Israel, and he engaged with Zionism and post-WWII reconciliation, thus contributing a voice attuned to modern Jewish identity, ethics, and politics.

Hannah Arendt (1906–1975) Although Hannah Arendt is more often identified as a political philosopher than a theologian, her Jewish heritage and reflections on moral questions make her a vital contributor to this cohort. Arendt was "one of the most influential political philosophers of the twentieth century" [16]. A student of Heidegger and Jaspers, Arendt fled Nazi Germany and later analyzed the nature of totalitarianism and the moral dilemmas of modernity. Her landmark works, including *The Origins of Totalitarianism*, *The Human Condition*, and *Eichmann in Jerusalem*, probe the capacity for evil, the fragility of human rights, and the importance of civic engagement. Notably, Arendt coined the phrase "the banality of evil" to describe the ordinary roots of wrongdoing, reflecting her concern with personal responsibility under authoritarian regimes. Including Arendt ensures the study benefits from a Jewish perspective on ethics, politics, and human dignity in the modern world, while also broadening the gender representation.

4.2 Christian Thinkers

Thomas Aquinas (c. 1225–1274) St. Thomas Aquinas is a central figure in Christian theology and Western philosophy. Educated in the Scholastic tradition, Aquinas is described as "the greatest figure of thirteenth-century Europe in the two preeminent sciences of the era, philosophy and theology" [20], epitomizing the high medieval synthesis of faith and reason. He systematized Christian doctrine in works like the *Summa Theologiae*, harmonizing Aristotelian philosophy with Church teachings. Aquinas is often seen as "the summit of the Christian tradition" after the Church Fathers and as "a foundational figure of modern thought" for his influence on ethics, metaphysics, and natural law [20]. His efforts to systematically rework Aristotelianism reshaped Western philosophy and provoked centuries of debate. The inclusion of Aquinas provides a medieval Catholic viewpoint, emphasizing rational inquiry in theology, virtue ethics, and the concept of a divinely ordained moral order.

Martin Luther (1483–1546) Martin Luther contributes the revolutionary voice of the Protestant Reformation. A German monk-turned-reformer, Luther is "the central figure of the Protestant Reformation" [10]. By posting his *Ninety-Five Theses* in 1517, he challenged Catholic doctrines and fundamentally altered the course of Christian history. Luther's theology—emphasizing salvation by faith alone and the authority of scripture over Church hierarchy—was radical for its time and remains foundational for Protestant denominations. Although primarily a theologian, Luther's ideas also carry philosophical import; scholars note that "the philosophical interest and impact of his ideas is significant" and comparable to figures like Augustine or Aquinas [10]. His inclusion ensures the representation of a Reformist perspective that champions individual conscience, the vernacular Bible, and critiques of institutional religious corruption.

John Wesley (1703–1791) John Wesley represents the evangelical and revivalist stream of Christian thought. An Anglican clergyman who, with his brother Charles, founded the Methodist movement, Wesley sparked an 18th-century revival that emphasized personal faith, holiness, and social service. He is included not only for founding a major Protestant tradition but also for his role in pioneering modern evangelical spirituality. Wesley traveled extensively, preaching to large crowds in open-air settings and organizing believers into small groups for discipleship. His Arminian theology, which stressed free will and sanctification, stood in contrast to the Calvinist predestinarianism of his day. Under Wesley's leadership, Methodism grew into a force for democratizing religion, taking the Gospel outside church walls to common people, including in the American colonies [27]. His concern for social issues—prison reform, abolition of slavery, and assistance to the poor—makes his perspective relevant to modern questions of social justice.

Rosemary Radford Ruether (1936–2022) Rosemary Radford Ruether provides a contemporary Christian voice and a critical feminist theological perspective. A Catholic theologian, Ruether was renowned as a "pioneer in feminist theology" whose work deeply challenged and transformed Christian thought from the late 20th century onward [11]. Her scholarship, exemplified by works such as *Sexism and God-Talk* (1983), critiques the patriarchal structures of the Church and reinterprets Christian doctrine to promote gender equality and social justice. By including Ruether, the study incorporates a perspective from liberation theology and feminist critique within Christianity. She represents the viewpoint of a woman interrogating her own religious tradition; for instance, Ruether argued that genuine Christian faith demands the full human dignity of women, famously asserting, "the critical principle of feminist theology is the promotion of the full humanity of women." Colleagues have noted that "though best noted as a pioneer in feminist theology, Dr. Ruether's work changed theology itself" [11], leaving an indelible mark on how churches address sexism and eco-justice.

4.3 Islamic Thinkers

Al-Ghazālī (1058–1111) Abū Ḥāmid al-Ghazālī is widely regarded as one of the most significant theologians in Islamic history. A Persian-born scholar, he achieved mastery in jurisprudence, theology (*kalām*), philosophy, and Sufism. Al-Ghazālī's authority in Islam is evidenced by his honorific title, "Ḥujjat al-Islām" (Proof of Islam) [9]. He was "one of the most important and influential thinkers in Islamic intellectual history," having "contributed immensely" to philosophy and the religious sciences [9]. His seminal work, *Iḥyā' 'ulūm al-dīn* (The Revival of the Religious Sciences), is credited with integrating Sufism into orthodox Sunni Islam [25]. In his *Tahāfut al-Falāsifa*

(The Incoherence of the Philosophers), Al-Ghazālī critiqued Avicenna and other Islamic peripatetic philosophers, defending orthodox faith against what he perceived as excessive rationalism. His work reshaped Islamic thought, prompting a turn away from Hellenistic philosophy toward a renewed grounding in revelation and spiritual experience.

Ibn Taymiyya (1263–1328) Taqī al-Dīn Aḥmad ibn Taymiyya, a Syrian Ḥanbali jurist and theologian, offers a contrasting Islamic perspective that is rigorously scripturalist and often polemical. Ibn Taymiyya is known for advocating a return to the foundational sources of Islam (the Qur'an and Sunnah) and for opposing what he considered to be unsanctioned innovations (*bid'ah*). Living during the Mongol invasions, he issued fatwas that articulated a puritanical reformism within Sunni Islam. His outspoken views against prevailing religious and political authorities led to his imprisonment on multiple occasions. Ibn Taymiyya's thought has had an enduring influence; in the modern era, he is frequently cited as an inspiration for revivalist movements, and his writings were a "major source of the Wahhābiyyah" reform of the 18th century [26]. His inclusion provides a voice representing orthodox rigor and reform, one that emphasizes strict monotheism, adherence to scripture, and moral austerity.

Ali Shariati (1933–1977) Ali Shariati introduces a modern, revolutionary Islamic voice from the Shī'a and Iranian context. A sociologist educated in Tehran and Paris, Shariati became "one of the key intellectual leaders of the Islamic Revolution (1979)" in Iran [1]. Often termed the "ideologue of the Iranian Revolution," Shariati synthesized Islamic principles with anti-colonial socialist thought. He drew on Western thinkers such as Marx and Fanon alongside Islamic sources to craft an ideology of liberation rooted in Shī'ite concepts of martyrdom and justice. His lectures and writings galvanized a generation of Iranian youth in the 1960s and 1970s. The inclusion of Shariati adds a perspective at the intersection of religion, politics, and social change. He represents the voice of an Islamic intellectual grappling with modernity, colonialism, and inequality. Moreover, Shariati's inclusion ensures the Shī'ite tradition is represented in the selection.

Fatema Mernissi (1940–2015) Fatema Mernissi, a Moroccan sociologist and pioneering Islamic feminist, is an indispensable addition to this group. She represents a gender-progressive interpretation of Islam and the experience of modern Muslim women. Mernissi has been described as "one of the founders of Islamic feminism," a scholar who challenged both Western stereotypes and indigenous patriarchal readings of Islam [8]. Her works, such as *Beyond the Veil* (1975) and *The Veil and the Male Elite* (1991), critically examine how male-dominated interpretations of scripture have been used to suppress women's rights. She excavated early Islamic history to argue for equality as an intrinsic Islamic value. The inclusion of Mernissi's thought incorporates a reformist voice that speaks to contemporary questions of gender, sexuality, and human rights in Muslim contexts. Mernissi also brings further geographic breadth with a North African perspective. Indeed, she was "an important and influential figure in Arab Islamic feminism," recognized for promoting the compatibility of Islam and feminism [8].

4.4 Broad Representation Across Time and Traditions

In summary, the scholars selected for this study collectively embody the richness and complexity inherent in Abrahamic thought across diverse historical, geographical, and theological landscapes. The panel spans pivotal historical junctures from medieval rationalism exemplified by Maimonides and Aquinas, through early-modern philosophical transformations represented by Spinoza and Luther, to contemporary theological debates driven by feminist critiques from Ruether and Mernissi and anti-colonial thought from Shariati. The careful inclusion of influential yet varied perspectives, from Buber's dialogical spirituality to Ibn Taymiyya's rigorous scripturalism, ensures the trained language models can authentically reflect the multifaceted nature of Abrahamic traditions. By deliberately encompassing diverse intellectual perspectives and ensuring representation of historically marginalized voices, particularly women and non-Western scholars, this selection offers a comprehensive foundation to critically and authentically address modern ethical and social questions, thereby effectively bridging historical wisdom with contemporary discourse.

5 Custom LLM Training and the Integration of World Values Survey

5.1 Introduction to Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) is an architectural framework that couples a language model's generative capacity with an external knowledge-retrieval mechanism [17]. Rather than relying exclusively on information embedded within the model parameters, RAG dynamically retrieves pertinent documents from an external corpus at inference time and incorporates them into the generation process. This procedure enables the model to ground its responses in current, verifiable, and domain-specific content.

The RAG architecture conventionally comprises two sequential stages: *retrieval* and *generation*. Upon receiving a user query, the retriever searches a vector database for the most semantically similar passages via dense embeddings and cosine similarity. These passages are supplied as context to a frozen language model, which then produces an informed, contextually rich response. Because the knowledge base can be updated independently of the language model, the approach facilitates continual incorporation of new material without necessitating model retraining.

5.2 Comparison Between Fine-Tuning and RAG

In preliminary system development, we investigated conventional fine-tuning by adapting a GPT-3.5-turbo variant with curated Islamic scholarly texts. This procedure entailed constructing prompt–completion pairs and fine-tuning via OpenAI's API. Although fine-tuning successfully reproduced specific scholarly tones, terminologies, and voices—such as that of Ibn Taymiyyah—it exhibited notable limitations.

Fine-tuning embeds information directly in model parameters, rendering updates cumbersome, costly, and inflexible; any new material or corrections require renewed fine-tuning. Moreover, fine-tuned models are susceptible to hallucination, often presenting fabricated content with unwarranted confidence—an unacceptable drawback in religious contexts where textual accuracy is essential [4].

RAG, by contrast, decouples knowledge storage from model parameters. Corpus updates are achieved simply by refreshing the document store, thereby permitting the rapid integration of newly available interpretations or commentaries. Empirical studies show that RAG architectures outperform fine-tuned models when confronted with recently introduced external knowledge [23]. Additional research indicates that RAG systems handle rare, domain-specific information more effectively and maintain superior factual accuracy [19]. Consequently, we adopted a RAG-based pipeline, which offers modularity, citation transparency, improved handling of specialized knowledge, and adherence to scholarly standards of evidence-based interpretation.

5.3 Retrieval-Augmented Generation for Sacred and Scholarly Texts

Our design decisions align with best practices in the literature. The seminal RAG study advises retrieving five to ten passages to balance retrieval noise against knowledge sufficiency [17]. Accordingly, we configured the system to retrieve the eight most semantically relevant text chunks for each query. These chunks—drawn from a custom corpus curated for this project—are supplied to the language model alongside the user request. This knowledge base includes the foundational scriptures of each tradition (the Torah, the Bible, and the Quran) as well as the collected works of the twelve profiled scholars, as detailed in the Appendix. Retrieving eight passages affords adequate contextual breadth for nuanced theological questions while limiting the inclusion of extraneous material, thereby optimizing performance and computational efficiency.

5.4 Evaluation Methodology Using the World Values Survey (WVS)

To evaluate model outputs systematically, we employed a structured questionnaire derived from World Values Survey (WVS) Wave 7 (2017–2022) [28]. The WVS, covering more than sixty nations, provides an authoritative dataset on values, beliefs, and cultural norms. Wave 7 includes responses from fifty-nine countries, collected primarily between 2018 and 2020, with some surveys extending into 2021–2023 owing to pandemic-related delays. This dataset offers a rigorous foundation for assessing socio-cultural and ethical stances at both global and regional scales.

Each language model answered 290 standardized WVS questions encompassing trust, religiosity, political preferences, happiness, moral values, education, gender roles, and personal responsibility. Representative items include:

- **Example 1**: "Do you think that having a feeling of responsibility is one of the most important qualities children should learn at home?" (Respond with *1* for Yes, *0* for No.)
- **Example 2**: "Taking all things together, would you say you are happy?" (Respond with *1* for Very happy, *2* for Rather happy, *3* for Not very happy, or *4* for Not at all happy.)

All models were prompted in the following standardized format:

> *I am an [Christian/Islamic/Jewish] scholar named XXX. Please respond to questions based solely on information extracted from my works. The response should contain only a single numerical value and nothing else.*

This instruction ensured that responses reflected the perspective of a specific historical scholar, minimizing the intrusion of contemporary biases. Responses were collected for subsequent comparative analysis.

Employing the WVS enables quantitative comparison of simulated attitudes and belief structures across models, scholars, and religions under a consistent empirical lens. By relying on a widely cited, peer-reviewed dataset, we align our evaluation with established social-science methodologies [12].

To enhance robustness, five response rounds per model were generated for each scholar. Each round comprised answers to the same set of questions, recorded on their respective Likert-type scales. Repeated sampling mitigates randomness and supports reliable cross-model and cross-scholar comparisons.

5.5 Data Cleaning and Preprocessing

All responses were numerically encoded and standardized to ensure consistency and compatibility for statistical analysis. Missing values, non-numeric responses, and anomalous entries were systematically addressed using a set of predefined imputation rules. When responses contained multiple numeric values, the value that appeared most frequently or aligned most closely with the intent of the question was selected. For responses where no meaningful numeric value was provided (e.g., "I don't know" or blank fields), neutral or average values were imputed, depending on the distribution of responses for the given question. For categorical variables, consistent default values were applied where information was absent, such as assuming the individual lives in a single-person household, has no prior employment, is unmarried, or speaks the most widely spoken language in their country. These procedures were designed to preserve the integrity of the dataset while minimizing bias introduced by missing or inconsistent data, following best practices for survey data and AI output evaluation [18].

6 Data Collection

6.1 Base LLMs

In this study, we anchor our retrieval-augmented generation (RAG) pipeline on three large language models—**Qwen-Plus**, **DeepSeek-R1**, and **GPT-4 Turbo**—selected for their complementary strengths in linguistic competence, multilingual reasoning, inference latency, and operating cost. Critically, the models originate from AI labs on different continents: Alibaba's DAMO Academy in East Asia, DeepSeek AI in East Asia, and OpenAI in North America—thereby broadening the geographic scope of our analysis. Because pre-training corpora often draw disproportionately on sources local to the developer's cultural and linguistic milieu, combining models from distinct regions increases the likelihood that our system captures a richer diversity of textual traditions and mitigates region-specific biases.

Qwen-Plus (72 B parameters), released in 2025 by Alibaba's DAMO Academy, is an open-weight model optimized for language understanding and generation. Fine-tuned variants attain state-of-the-art performance on multiple English benchmarks while retaining strong cross-lingual coverage, making the model well suited to retrieval-augmented religious dialogue tasks [2]. Its inference speed lies between GPT-4 Turbo and DeepSeek-R1, enabling real-time deployment without prohibitive computational overhead.

DeepSeek-R1 (≈ 671 B parameters) was introduced in 2024 by DeepSeek AI, a Hangzhou-based research lab spun out of a quantitative hedge fund. Emphasizing multi-step reasoning and multilingual comprehension, DeepSeek-R1 excels on complex, open-ended prompts in academic settings [7]. Although its extensive context processing yields the highest latency among the three models, the resulting interpretive depth is indispensable for nuanced theological or historical inquiries.

GPT-4 Turbo (≈ 1.76 T parameters), launched by OpenAI in late 2024, remains the benchmark for general natural-language processing tasks, including summarization, code synthesis, logical reasoning, and multilingual generation [22]. When paired with an optimized RAG workflow, GPT-4 Turbo delivers unmatched accuracy and fluency. Despite its higher operational cost, the model's reliability and breadth make it essential for applications demanding maximal precision.

Together, these three models provide a robust, geographically diverse foundation for our RAG system, enabling dynamic retrieval and generation of sophisticated religious content while carefully balancing accuracy, latency, and cost. An overview of these models is provided in Table 2.

Model	Year of Origin	Estimated # of Parameters	Strengths	Approx. Cost (per round per scholar)	Runtime Efficiency
Qwen-Plus	2025	∼72 billion	Excels in language understanding and generation; performs competitively on English benchmarks with fine-tuning.	$0.31	Moderate speed; slower than GPT-4 Turbo, faster than DeepSeek-R1
DeepSeek-R1	2024	∼671 billion	Strong reasoning and multi-step question answering in multilingual contexts; excels in open-ended academic prompts.	$0.90	Slowest among the three owing to heavier context-interpretation load
GPT-4 Turbo	2024	∼1.76 trillion	State-of-the-art performance across virtually all NLP tasks, including reasoning, summarization, coding, and multilingual generation.	$8.00	Fastest response time, especially when integrated with optimized RAG pipelines

Table 2: Comparison of Base Language Models Used in the RAG Pipeline

6.2 Collection of LLM Responses and World Values Survey

This study constructs a custom dataset of simulated survey responses generated by three large language models (LLMs)—Qwen-Plus, DeepSeek-R1, and GPT-4Turbo—and attributed to twelve historical scholars representing the Islamic, Jewish, and Christian traditions. For each scholar, the models produced answers to *290* cultural and value-oriented items taken verbatim from the World Values Survey (WVS) Wave 7 questionnaire [28]. The procedure yielded 36 spreadsheets (12×3 models). To bolster robustness, each model supplied 5 independent responses to every WVS question. To ensure answer independence, each question was presented to the model individually and in random order rather than sequentially.

Each scholar is therefore represented by three files—one per LLM—whose responses reflect that scholar's documented worldview and philosophical orientation. The prompting protocol used to elicit these answers is detailed in Section 5.4.

To contextualize the synthetic data, the analysis also employs the official WVS Wave 7 microdata (2017–2022), which comprise responses from more than fifty countries [28]. Incorporating these human observations allows a systematic comparison between the simulated historical scholars and contemporary populations, thereby grounding the study in empirically validated sociocultural measures [12].

7 Exploratory Visual Mapping of Aggregate Value Orientations

This section undertakes a visual mapping of the combined World Values Survey (WVS)–scholar corpus. Each observation's 290 attitudinal items is condensed into three feature blocks that capture distinct measures of central tendency: the mean block contains the arithmetic averages of each numerical item, the median block records median responses, and the mode block registers the most frequently chosen categories. For countries, these statistics reflect the aggregated preferences of survey respondents, whereas for the twelve Abrahamic scholars they summarize responses generated through a simulation protocol that queried bespoke large language models. An overview can be seen in Figure 4.

Within every block, the high-dimensional response matrix is projected into two dimensions via three complementary manifold learners—principal component analysis (PCA), uniform manifold approximation and projection (UMAP), and t-distributed stochastic neighbor embedding (t-SNE). The resulting embeddings are partitioned with K-means clustering set a priori to $k = 3$, mirroring the expectation that Islamic, Christian, and Jewish scholars should occupy distinct regions of the latent value space. Each cluster is color-coded according to the dominant religious affiliation of the scholars it contains—green for Islam, red for Christianity, and blue for Judaism—while

countries appear as circles and scholars as larger, annotated squares; convex hulls demarcate the spatial extents of country clusters to provide additional context. Additionally, the customized LLM responses are presented in a color-coded overview in the Appendix (Figures 5, 6, and 7).

Together, these visualizations furnish an initial cartography of shared value orientations and enable a preliminary assessment of whether the hypothesized tripartite religious structure manifests consistently across manifold-learning techniques and measures of central tendency. More detailed statistical tests—conducted pairwise within and across religious traditions—are presented in subsequent sections to evaluate these mapped patterns rigorously.

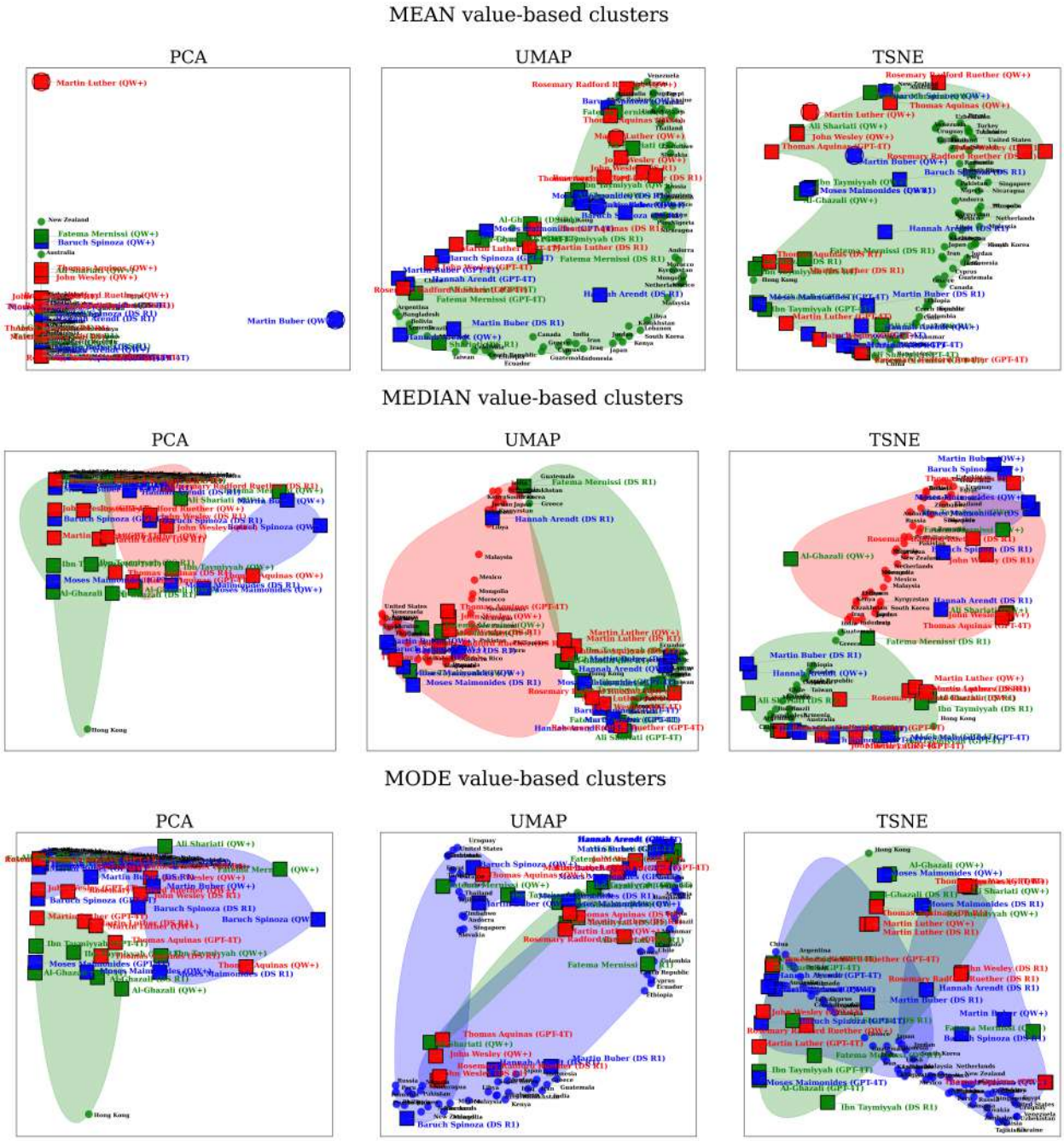

Fig. 4: Visual mapping of Abrahamic scholars within the World Values Survey dataset. *Note*: Larger square markers denote scholar locations, while smaller circular markers represent typical country responses. Country points are colored according to the predominant scholar affiliation within each cluster.

8 Empirical Strategy

This study embeds twelve large-language-model (LLM) personae—four each from Christianity, Judaism, and Islam, in a $3{:}1$ male–female ratio—within the value space defined by the *World Values Survey* (WVS) Wave 7 (2017–2022). Each persona answers all 290 WVS items, ensuring strict item comparability with human respondents and enabling a coherent test of four hypotheses.

A two-pronged data strategy is employed. Synthetic responses are generated by Qwen-Plus, DeepSeek-R1, and GPT-4 Turbo for every scholar, yielding a $290{\times}36$ matrix (12 scholars $\times$ 3 models). The human benchmark comprises roughly 80 000 Wave 7 respondents from more than 50 countries. Items are scored so that larger integers indicate stronger endorsement.

H1 (cross-LLM comparison). To test for systematic differences in model outputs, we employ an ordered logistic regression. For World Values Survey (WVS) item j and response category threshold k, the cumulative logit model is specified as:

$$\operatorname{logit}\left[\Pr\left(Y_{ij} \leq k\right)\right] = \theta_{kj} - \left(\beta_1 \, \mathbf{1}_{\mathsf{Qwen},i} + \beta_2 \, \mathbf{1}_{\mathsf{DeepSeek},i}\right). \quad (1)$$

In this model, the indicator variables for Qwen-Plus and DeepSeek-R1 are coded with GPT-4 Turbo as the reference category. To account for the non-independence of multiple responses to the same survey item, standard errors are clustered at the item level.

H2 (cross-religious scholar comparison). To compare outputs across religious traditions, we regress scholar-level aggregate scores on religion dummies via Ordinary Least Squares (OLS). The analysis is performed independently for several output metrics, including the mean, median, and mode of each scholar's responses. For a given output metric d, the model is specified as:

$$F_{id} = \alpha_d + \delta_{1d}\mathbf{1}_{\mathsf{Christian},i} + \delta_{2d}\mathbf{1}_{\mathsf{Jewish},i} + \varepsilon_{id}. \quad (2)$$

Here, F_{id} is the aggregate score for scholar i, and the model uses Islam as the reference group. The point estimates for the coefficients are derived from a model using heteroscedasticity-consistent standard errors (HC1). To address small-sample uncertainty, 95% confidence intervals are constructed using a percentile bootstrap method, based on the empirical distribution of coefficients from 10,000 replications.

H3 (historical vs. contemporary region). This hypothesis is tested by comparing each scholar's proximity to global versus contemporary regional norms. The analysis represents each scholar i using a five-dimensional feature vector, $\mathbf{x}_i$, comprising the mean responses to the first five World Values Survey questions.

The local country centroid, $\boldsymbol{\mu}_{c(i)}$, and the global centroid, $\boldsymbol{\mu}_{\mathrm{G}}$, are computed from country-level aggregate data. The common covariance matrix, $\boldsymbol{\Sigma}$, is calculated from these country centroids and is regularized to ensure it is invertible. The Mahalanobis distances to these centroids are calculated as:

$$D_i^{\mathrm{global}} = \left\|\boldsymbol{\Sigma}^{\dagger/2}\big(\mathbf{x}_i - \boldsymbol{\mu}_{\mathrm{G}}\big)\right\|, \qquad D_i^{\mathrm{local}} = \left\|\boldsymbol{\Sigma}^{\dagger/2}\big(\mathbf{x}_i - \boldsymbol{\mu}_{c(i)}\big)\right\|, \quad (3)$$

where $\boldsymbol{\Sigma}^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of the regularized covariance matrix. A paired t-test on the difference $D_i^{\mathrm{global}} - D_i^{\mathrm{local}}$ is then performed to evaluate relative proximity. This test is applied at three distinct levels of aggregation: for each scholar individually, for each of the three religious traditions, and for all scholars combined.

H4 (intra-scholar, cross-model). To test for differences in how models represent a single scholar, we first compute scholar-level aggregate scores. For each scholar i and each model m, we calculate the mean of the `output_mean`, `output_median`, and `output_mode` across all survey items. Let this aggregated score be denoted by Y_{im}.

A repeated-measures ANOVA is then applied to these scores. This test evaluates whether there are systematic differences in the outputs across the models, treating the model as a within-subject factor. The underlying linear model can be expressed as:

$$Y_{im} = \mu + \pi_i + \tau_m + \varepsilon_{im}, \quad (4)$$

where μ is the grand mean, π_i is the effect for scholar i, and τ_m is the effect for model m. The ANOVA tests the null hypothesis that the model effects are equal ($H_0 : \tau_{\mathsf{GPT4}} = \tau_{\mathsf{Qwen}} = \tau_{\mathsf{DSR1}}$).

Following a significant ANOVA result, a post-hoc analysis using Tukey's Honestly Significant Difference (HSD) test is performed to identify which specific pairs of models produce significantly different outputs.

9 Results

To test Hypothesis 1, we fitted an ordered logistic regression model to predict the response category based on the large language model (LLM) used. Standard errors were clustered by item to account for the non-independence of responses. In line with the hypothesis, GPT-4 Turbo was treated as the reference category against which Qwen-Plus and DeepSeek-R1 were compared. The overview of the results is shown in Table 3.

The analysis revealed no statistically significant differences between the models. The coefficient for Qwen-Plus was not significant ($\beta = -0.01$, SE = 0.06, $p = .833$), nor was the coefficient for DeepSeek-R1 ($\beta = 0.02$, SE = 0.05, $p = .749$). For both comparisons, the 95% confidence intervals contained zero, indicating that their performance on this task could not be statistically distinguished from the reference model, GPT-4 Turbo.

An examination of the marginal expected responses further underscored the lack of practical difference between the models. The average predicted response category for GPT-4 Turbo was 5.02, compared to 5.00 for Qwen-Plus and 5.04 for DeepSeek-R1. These marginal differences are negligible, suggesting a high degree of similarity in the models' outputs.

In summary, *the results do not support Hypothesis 1*. We found no statistical evidence to suggest that the outputs of Qwen-Plus, DeepSeek-R1, and GPT-4 Turbo differ systematically when answering the selected religious queries.

Table 3: Model Coefficients and Marginal Expected Responses by LLMs

LLM	**Coefficient**	**Std. Error**	z**-value**	p**-value**	**Expected Response**
GPT-4 Turbo	— (Reference)	—	—	—	5.02
Qwen-Plus	-0.013	0.062	-0.21	.833	5.00
DeepSeek-R1	0.017	0.054	0.32	.749	5.04

Note: Coefficients for Qwen-Plus and DeepSeek-R1 are relative to GPT-4 Turbo. Model fit statistics: Log-Likelihood = -20168, AIC = 40470, BIC = 40970, N = 9,986. Standard errors are clustered by item. The numerous model threshold parameters are omitted for clarity.

Table 4 shows the results of the test for Hypothesis 2, and presents ordinary least squares estimates from regressing each aggregate output metric on religion dummies for Christian and Jewish scholars, with Islam serving as the reference category. The intercept $\hat{\alpha}$ denotes the mean score for Islamic scholars, whereas $\hat{\delta}_{\text{Christian}}$ and $\hat{\delta}_{\text{Jewish}}$ capture average deviations from this baseline. Ninety-five percent percentile-bootstrap confidence intervals (CI) accompany the point estimates. The column labeled N reports the number of scholar–model observations included in each regression.

Table 4: Cross-religion Scholar Comparison

	Islam baseline		Christian Δ		Jewish Δ	
Score	$\hat{\alpha}$	N	$\hat{\delta}_{\text{C}}$	95% CI	$\hat{\delta}_{\text{J}}$	95% CI
`Output (Mean)`	12.57	10 450	3.12	[–3.67, 10.42]	85.20	[–3.48, 259.64]
`Output(Median)`	12.04	10 450	0.40	[–5.22, 5.96]	2.17	[–3.89, 8.08]
`Output (Mode)`	11.85	10 450	0.45	[–5.17, 5.98]	1.17	[–4.68, 6.90]

The confidence intervals for all religion differentials encompass zero, indicating that none of the estimated gaps is statistically distinguishable from the Islamic baseline at the 5% level. Although the point estimates are uniformly positive—suggesting, in isolation, higher scores for Christian and especially Jewish scholars—the wide uncertainty bands permit both negative and positive effects. The particularly broad interval for the Jewish coefficient in `output_mean` (–3 to 260) signals pronounced sampling variability, plausibly driven by unequal scholar counts or substantial within-tradition heterogeneity. Collectively, these findings reveal no significant cross-religious divergence in the aggregate metrics and, consequently, offer *no empirical support for Hypothesis 2*.

There is some empirical support for Hypothesis 3. In our framework, we represent each scholar's synthetic output as a five-dimensional vector $\mathbf{x}_i = (\overline{Q}_1, \ldots, \overline{Q}_5)$, where $\overline{Q}_k$ is the mean AI response on question k. We then locate $\mathbf{x}_i$ relative to two normative centroids:

1. **Global centroid** $\boldsymbol{\mu}_{\mathrm{G}}$: the average of all present-day country vectors, summarizing the cross-national "average worldview."
2. **Local centroid** $\boldsymbol{\mu}_{c(i)}$: the average of the scholar's corresponding present-day country vector only, capturing the contemporary regional norm—e.g., modern Iraq for Al-Ghazali.

Intuitively, if a scholar's AI-generated profile is closer to $\boldsymbol{\mu}_{\mathrm{G}}$ than to $\boldsymbol{\mu}_{c(i)}$, the model is projecting that thinker into a broadly "globalized" discourse rather than the specific regional attitudes of their historical homeland. Formally, we quantify this proximity by computing the Mahalanobis distances (defined in the Empirical Strategy section). A paired t-test on the differences $\Delta_i = D_i^{\mathrm{global}} - D_i^{\mathrm{local}}$ then evaluates, for each scholar, whether the AI response is statistically closer to the global norm ($\Delta_i < 0$) or to the modern regional norm ($\Delta_i > 0$).

As Table 5 shows, individual-scholar t-tests exhibit pronounced heterogeneity. Three Islamic figures (Al-Ghazali, Ali Shariati, Moses Maimonides) yield significantly negative t-values ($p < 0.0001$), indicating their AI profiles lie nearer the global centroid than their own country's centroid. By contrast, Baruch Spinoza and Fatema Mernissi (Judaism) and Rosemary Radford Ruether (Christianity) produce strongly positive t-statistics ($p < 0.0001$), signaling a tighter fit with their modern regional norms. Aggregating by tradition confirms these patterns: Islamic scholars overall remain significantly closer to the global average ($t = -5.77,\ p < 0.0001$), Christian scholars align more with their local centroid ($t = 17.86,\ p < 0.0001$), and Jewish scholars show no reliable bias ($t = -1.25,\ p = 0.2132$). Finally, pooling all valid scholar–model pairs yields a modest but significant global bias ($t = -2.52,\ p = 0.0117$), suggesting that—depending on both the thinker and the tradition—the LLMs sometimes universalize medieval voices, while at other times they honor contemporary regional particularities.

Table 5: Paired t-tests on $D_{\mathrm{global}} - D_{\mathrm{local}}$

Group	n	t	p
Per scholar			
Al-Ghazali	806	-40.627	< 0.0001
Ali Shariati	784	-5.488	< 0.0001
Baruch Spinoza	543	26.761	< 0.0001
Fatema Mernissi	499	31.301	< 0.0001
Moses Maimonides	523	-37.907	< 0.0001
Rosemary Radford Ruether	290	17.864	< 0.0001
By religion			
Islam	2,089	-5.774	< 0.0001
Christianity	290	17.864	< 0.0001
Judaism	1,066	-1.246	0.2132
All scholars combined			
All scholars	3,445	-2.523	0.0117

The statistical results directly support Hypothesis 4. The significant differences found in the `median` and `mode` of the outputs demonstrate that the AI models represent the same scholar's ideas differently. This confirms that the models have distinct 'interpretive styles' and 'model-specific tendencies,' as the typical and most frequent responses they generate are not the same. Even though the overall average (`mean`) response was similar, the differences in the `median` and `mode` reveal a more nuanced, underlying variation in how each model reconstructs an intellectual profile.

To further specify the variations identified by the ANOVA, a post-hoc analysis using Tukey's Honestly Significant Difference (HSD) test was performed. This test facilitated pairwise comparisons to quantify the mean differences between the models for the `output_median` and `output_mode` features. The analysis of median out-

puts revealed significant divergences in the models' typical responses. A substantial and statistically significant difference was observed between Qwen and GPT-4, where Qwen's median output was, on average, 7.02 units higher ($p < .001$). A significant difference was also identified between Qwen and DSR1, with Qwen's median output being 3.86 units higher ($p = .027$). The comparison between DSR1 and GPT-4, however, did not yield a statistically significant difference ($p = .080$), which suggests their typical outputs are statistically comparable. These findings demonstrate that Qwen generates a significantly elevated median response compared to both DSR1 and GPT-4, quantitatively confirming a distinct interpretive tendency.

The analysis of modal output, which reflects the most frequent response for a given scholar, isolated a more specific discrepancy. The pairwise comparison indicated a highly significant difference between Qwen and GPT-4, where Qwen's modal output was 5.81 units higher on average ($p < .001$). In contrast, the differences between DSR1 and Qwen ($p = .106$) and between DSR1 and GPT-4 ($p = .101$) were not found to be statistically significant. This outcome suggests that while DSR1's most frequent response is not statistically distinguishable from the other models, Qwen and GPT-4 generate significantly different common themes from one another. This provides strong quantitative evidence of model-specific bias in the representation of a scholar's intellectual profile, thereby substantiating the primary hypothesis that individual models exhibit unique interpretive styles. The results associated with Hypothesis 4 are provided in Table 6.

Table 6: Repeated-Measures ANOVA and Tukey HSD Post-Hoc Tests for Summary Features

A. Repeated-Measures ANOVA

Feature	F	Num DF	Den DF	p
output_mean	1.2170	2	22	0.3153
output_median	13.8712	2	22	0.0001
output_mode	9.7129	2	22	0.0009

B. Tukey HSD for `output_median`

Group1	Group2	Mean Diff.	p-adj	Lower	Upper	Reject
DSR1	GPT4	-3.1658	0.0799	-6.6401	0.3085	False
DSR1	Qwen	3.8550	0.0270	0.3807	7.3293	True
GPT4	Qwen	7.0208	0.0001	3.5465	10.4951	True

C. Tukey HSD for `output_mode`

Group1	Group2	Mean Diff.	p-adj	Lower	Upper	Reject
DSR1	GPT4	-2.9239	0.1008	-6.3049	0.4571	False
DSR1	Qwen	2.8895	0.1059	-0.4915	6.2705	False
GPT4	Qwen	5.8133	0.0005	2.4323	9.1943	True

10 Discussion

This study's empirical analysis reveals a complex and multifaceted relationship between large language models and the simulation of historical religious thought. The findings indicate that while different models exhibit a high degree of surface-level similarity, significant underlying variations emerge upon closer inspection, confirming some of the study's central hypotheses while challenging others.

An initial cross-model comparison (Hypothesis 1) using an ordered logistic regression found no statistically significant differences between the outputs of GPT-4 Turbo, Qwen-Plus, and DeepSeek-R1. With GPT-4 Turbo as the reference, the coefficients for Qwen-Plus ($\beta = -0.01$, $p = .833$) and DeepSeek-R1 ($\beta = 0.02$, $p = .749$) were

negligible, suggesting a high degree of parity in their aggregate responses. Similarly, the study found no empirical support for systematic differences across religious traditions (Hypothesis 2). Although point estimates suggested higher scores for Christian and Jewish scholars relative to their Islamic counterparts, the wide confidence intervals rendered these differences statistically insignificant, indicating substantial intra-tradition heterogeneity.

More nuanced results emerged from the historical-to-contemporary regional comparison (Hypothesis 3). A paired t-test on the Mahalanobis distances between a scholar's simulated profile and both global and regional value centroids revealed significant heterogeneity. The profiles of Islamic scholars like Al-Ghazali ($t = -40.63$, $p < .0001$) aligned more closely with the global average, suggesting the models project a universalized worldview. In contrast, Christian scholars, such as Rosemary Radford Ruether ($t = 17.86$, $p < .0001$), demonstrated a significantly tighter fit with their modern regional norms. This suggests that the models' capacity to capture regional specificity is inconsistent and may be dependent on the tradition being simulated.

The most compelling evidence arose from the intra-scholar, cross-model comparison (Hypothesis 4). A repeated-measures ANOVA confirmed that while the models produced similar mean outputs, there were highly significant differences in their median ($F(2, 22) = 13.87, p < .001$) and modal ($F(2, 22) = 9.71, p < .001$) responses. A post-hoc Tukey HSD test specified these variations, revealing that Qwen's median output was significantly higher than both DSR1 (mean diff. = 3.86, $p = .027$) and GPT-4 (mean diff. = 7.02, $p < .001$). This finding confirms that different models possess distinct interpretive styles, biases, and tendencies when reconstructing the intellectual profile of a single scholar.

Looking forward, these results carry significant academic and policy repercussions. For academia, they demonstrate that LLMs can serve as powerful tools for digital humanities, enabling new forms of textual analysis and historical simulation. However, the model-specific biases uncovered here underscore the critical need for methodological vigilance and cross-platform validation in future research. For policymakers and developers, these findings highlight the ethical complexities of deploying AI in religious contexts. The tendency of models to either universalize or regionalize historical thought without a clear, predictable pattern poses a challenge for creating systems that are both globally relevant and culturally sensitive. As AI-driven religious chatbots become more prevalent, ensuring they represent traditions with fidelity and nuance is not merely a technical challenge but a profound social responsibility. Future work must focus on developing more transparent and controllable architectures that allow for greater fidelity to source material and mitigate the subtle but significant biases revealed in this study.

Bibliography

[1] Ervand Abrahamian. Ali shari'ati: Ideologue of the iranian revolution. *MERIP Reports*, 12(102):24–28, 1982.

[2] Alibaba DAMO Academy. Qwen-plus: Advancing open-weight language understanding. `https://github.com/QwenLM/Qwen-Plus`, 2025. Accessed: 2025-04-08.

[3] Bible AI. Bible ai – advanced bible search and study tools. `https://bibleai.com/`, 2023. Accessed: 2024-04-06.

[4] Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Euan Rutherford, Kathleen Millican, et al. Improving language models by retrieving from trillions of tokens. *Nature*, 2022.

[5] BrainyBlaze. Ravgpt – advancing torah study through artificial intelligence. `https://ravgpt.ai/`, 2023. Accessed: 2024-04-06.

[6] Chatbot Eli. Chatbot eli – your questions about god answered. `https://www.chatboteli.com/`, 2022. Accessed: 2024-04-06.

[7] DeepSeek AI. Deepseek-r1: Multilingual reasoning at scale. `https://github.com/deepseek-ai`, 2024. Accessed: 2025-04-08.

[8] Emily Evans. Remembering fatima mernissi, a pioneer in islamic feminism. Brandeis University News, January 2016.

[9] Frank Griffel. Al-ghazali (c. 1058—1111).

[10] Michael A. G. Haykin. Martin luther. The Stanford Encyclopedia of Philosophy (Fall 2020 Edition), 2020.

[11] Howard University School of Divinity. Husd community remembers the life and legacy of dr. rosemary radford ruether, May 2022.

[12] Ronald Inglehart. *Cultural Evolution: People's Motivations are Changing, and Reshaping the World*. Cambridge University Press, 2021.

[13] IslamAndAI. Quran-nlp project repository. `https://github.com/islamAndAi/QURAN-NLP`, 2022. Accessed: 2024-04-06.

[14] IslamiCity. Chatilm – enhancing islamic knowledge with ai. `https://chatilm.islamicity.org/`, 2024. Accessed: 2025-04-08.

[15] Joel L. Kraemer. Maimonides. The Stanford Encyclopedia of Philosophy (Summer 2024 Edition), 2024.

[16] Julia Kristeva. Hannah arendt. The Stanford Encyclopedia of Philosophy (Spring 2024 Edition), 2024.

[17] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Kelvin Guu, Douwe Kiela, Angela Fan, Edouard Grave, Armand Joulin, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems*, 2020.

[18] Percy Liang, Rishi Bommasani, et al. Holistic evaluation of language models. arXiv preprint arXiv:2211.09110, 2022.

[19] Yang Liu et al. Fine-tuning vs. retrieval-augmented generation for less popular knowledge. arXiv preprint arXiv:2403.01432, 2024.

[20] Ralph McInerny and John O'Callaghan. Saint thomas aquinas. The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), 2022.

[21] Steven Nadler. Baruch spinoza. The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), 2023.

[22] OpenAI. Gpt-4 turbo: High-efficiency reasoning and generation. `https://openai.com/research/gpt-4-turbo`, 2024. Accessed: 2025-04-08.

[23] Dimitris Papailiopoulos et al. Fine-tuning or retrieval? comparing knowledge injection in llms. In *Proceedings of EMNLP 2024*, 2024.

[24] The Center for Unity. Askjesus chatbot – explore jesus' teachings through ai. `https://www.askjesuschatbot.com/`, 2023. Accessed: 2024-04-06.

[25] The Editors of Encyclopaedia Britannica. al-ghazali. *Encyclopædia Britannica*, May 2024.

[26] The Editors of Encyclopaedia Britannica. Ibn taymiyyah. *Encyclopædia Britannica*, June 2024.

[27] The Editors of Encyclopaedia Britannica. John wesley. *Encyclopædia Britannica*, May 2024.

[28] World Values Survey Association. World values survey wave 7 (2017-2022) official aggregate v.3.0. `https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp`, 2022. Accessed: 2025-04-08.

[29] Michael Zank and Zachary Braiterman. Martin buber. The Stanford Encyclopedia of Philosophy (Fall 2020 Edition), 2020.

Appendix

A. Corpus of Abrahamic Scholar Texts for LLM Customization

Judaism				
Author	**Gender**	**Time Frame**	**Number of Works**	**Works Used**
Moses Maimonides	Male	1138–1204	6	– The Guide for the Perplexed – Mishneh Torah – The Preservation of Youth: Essays on Health – The Laws of the Hebrews Relating to the Poor and the Stranger – The Eight Chapters of Maimonides on Ethics (Shemonah Perakim) – On the Elucidation of Some Symptoms and the Response to Them
Baruch Spinoza	Male	1632–1677	7	– A Short Treatise on God, Man and His Well-Being – The Principles of Cartesian Philosophy – Tractatus Theologico-Politicus – Epistolae – Spinoza Complete Works – The Collected Works of Spinoza – The Vatican Manuscript of Spinoza's Ethica
Martin Buber	Male	1878–1965	4	– I and You – Eclipses of God – Two Types of Faith – The Writings of Martin Buber
Hannah Arendt	Female	1906–1975	7	– Human Condition – Origins of Totalitarianism – Between Past and Future – On Revolution – On Violence – Crises of the Republic – Love and St Augustine

Christianity				
Author	**Gender**	**Time Frame**	**Number of Works**	**Works Used**
Thomas Aquinas	Male	1225–1274	6	– Of God and His Creatures – Summa Theologica – Disputed Questions on Truth
Martin Luther	Male	1483–1546	19	– A manual of the book of Psalms – A Treatise on Good Works – An Explanation of Luther's Small Catechism – An Open Letter on Translating – Commentary on Genesis, Vol. 1 – Commentary on Genesis, Vol. 2 – Commentary on the Epistle to the Galatians – Concerning Christian Liberty – Epistle Sermons, Vol. 2 – Epistle Sermons, Vol. 3 – Lessons in the Small Catechism of Dr. Martin Luther – Luther Little Instruction Book – Martin Luther's Large Catechism – Selections from the Table Talk of Martin Luther – The Book of Vagabonds and Beggars – The Epistles of St. Peter and St. Jude Preached and Explained – The Smalcald Articles – Works of Martin Luther (Volume I) – Works of Martin Luther (Volume II)
John Wesley	Male	1703–1791	15	– The works of the Rev. John Wesley, Vol. 01 – The works of the Rev. John Wesley, Vol. 02 – The works of the Rev. John Wesley, Vol. 03 – The works of the Rev. John Wesley, Vol. 04 – The works of the Rev. John Wesley, Vol. 05 – The works of the Rev. John Wesley, Vol. 06 – The works of the Rev. John Wesley, Vol. 07 – The works of the Rev. John Wesley, Vol. 08 – The works of the Rev. John Wesley, Vol. 09 – The works of the Rev. John Wesley, Vol. 10 – The works of the Rev. John Wesley, Vol. 11 – The works of the Rev. John Wesley, Vol. 12 – The works of the Rev. John Wesley, Vol. 13 – The works of the Rev. John Wesley, Vol. 14 – Thoughts upon slavery
Rosemary Radford Ruether	Female	1936–2022	5	– Goddesses and the divine feminine: a Western religious history – Christianity and Social Systems : Historical Constructions and Ethical Challenges – America, Amerikkka : Elect Nation and Imperial Violence – Sexism and God-Talk – Ecofeminism – The Challenge to Theology

Islam				
Author	**Gender**	**Time Frame**	**Number of Works**	**Works Used**
Ghazali	Male	1058–1111	10	– Revival of Religious Learnings – Incoherence of Philosophers – Alchemy of Happiness – Al Ghazali on Patience and Thankfulness – On Disciplining the Soul – Dear Son – Deliverance from Error – Counsel for Kings – The Niche for Lights – Beginning of Guidance
Ibn Taymiyya	Male	1263–1328	23	– The Friends of Allah & The Friends of Shaytan – Gardens of Purification – The Lofty Virtues of Ibn Taymiyyah – Diseases of the Hearts & Their Cures – Seeking Forgiveness – Ibn Taymiyyah Expounds on Islam – Sharh Al-Aqeedat-il-Wasitiyah – Kitab Al-Iman – Book of Emaan – The Relief From Distress – Purify My Heart – The Principle Of Love And Desire – The Political Shariyah – The Goodly Word - Al Kalim al Tayyib – The Religious And Moral Doctrine Of Jihad – Enjoining Good, Forbidding Evil – Fatwas of Muslim Women – The Nature of Fasting – Ibn Taymeeyah's Letters From Prison – Ibn Taymiyah's Essay on The Jinn – The Right Way – Response to Those Who Say ... – Madinan Way

Author	Gender	Time Frame	Number of Works	Works Used
Ali Shariati	Male	1933–1977	25	– Hajj – Pilgrimage – Where Shall we Begin? – Abu-Dhar – Fatemeh is Fatemeh – A Message to the Enlightened Thinkers – Islamology – Expectations from the Muslim Woman – A Visage of Prophet Mohammad – Mission of a Free Thinker – A Glance at Tomorrow's History – An Approach to Understanding Islam – Reflections of a Concerned Muslim – Philosophy of Supplication – The Free Man and Freedom of the Man – Cultural Resource Extraction – Reflections of Humanity – Mohammad Iqbal – Selection or Election – Horr – Red Shiism vs. Black Shiism – Jihad and Martyrdom – One, Followed with Indefinite Zeros – Art Awaiting the Savior – Martyrdom, Arise and Bear Witness – Norouz; Declaration of Iranians' Livelihood
Fatema Mernissi	Female	1940–2015	7	– Beyond the Veil – The Veil and the Male Elite – Forgotten Queens of Islam – Women's Rebellion and Islamic Memory – Dreams of Trespass – Scheherazade Goes West – Islam and Democracy

B. World Values Survey Items

Below is a list of questions posed to respondents in a select group of countries during Wave 7 (2017–2022) of the World Values Survey.

1. How important is your family in your life? Rate on a scale from 1 ("Very important") to 4 ("Not at all important").
2. How important are your friends in your life? (1–4 as above.)
3. How important is leisure time in your life? (1–4 as above.)
4. How important is politics in your life? (1–4 as above.)
5. How important is work in your life? (1–4 as above.)
6. How important is religion in your life? (1–4 as above.)
7. Do you think good manners are an especially important quality for children to learn at home? (1 = yes; 0 = no)
8. Do you believe that independence is an essential quality that children should be encouraged to learn at home? (1 = yes; 0 = no)
9. Would you consider hard work to be an especially important value for children to develop at home? (1 = yes; 0 = no)

10. Do you think that having a feeling of responsibility is one of the most important qualities children should learn at home? (1 = yes; 0 = no)
11. Is imagination an especially valuable trait for children to develop at home? (1 = yes; 0 = no)
12. Do you believe that tolerance and respect for other people should be prioritized as an essential value for children to learn at home? (1 = yes; 0 = no)
13. Would you consider thrift, saving money, and taking care of things to be an important lesson for children to learn at home? (1 = yes; 0 = no)
14. Do you think that determination and perseverance are crucial values for children to develop at home? (1 = yes; 0 = no)
15. Do you consider religious faith to be one of the five most important qualities for children to be encouraged to learn at home? (1 = yes; 0 = no)
16. Should not being selfish (unselfishness) be emphasized as a key value for children to learn at home? (1 = yes; 0 = no)
17. Do you believe that obedience is an especially important quality that children should be encouraged to develop at home? (1 = yes; 0 = no)
18. Do you not want to have people who are drug addicts as neighbors? (1 = yes; 0 = no)
19. Do you not want to have people of a different race as neighbors? (1 = yes; 0 = no)
20. Do you not want to have people who have AIDS as neighbors? (1 = yes; 0 = no)
21. Do you not want to have immigrants or foreign workers as neighbors? (1 = yes; 0 = no)
22. Do you not want to have homosexuals as neighbors? (1 = yes; 0 = no)
23. Do you not want to have people of a different religion as neighbors? (1 = yes; 0 = no)
24. Do you not want to have people who are heavy drinkers as neighbors? (1 = yes; 0 = no)
25. Do you not want to have unmarried couples living together as neighbors? (1 = yes; 0 = no)
26. Do you not want to have people who speak a different language as neighbors? (1 = yes; 0 = no)
27. Has one of your main goals in life been to make your parents proud? (1 = Strongly agree, . . . , 4 = Strongly disagree)
28. Do you believe that when a mother works for pay, the children suffer? (1 = Strongly agree, . . . , 4 = Strongly disagree)
29. Do you agree that, on the whole, men make better political leaders than women do? (1 = Strongly agree, . . . , 4 = Strongly disagree)
30. Do you think a university education is more important for a boy than for a girl? (1 = Strongly agree, . . . , 4 = Strongly disagree)
31. Do you agree that, on the whole, men make better business executives than women do? (1 = Strongly agree, . . . , 4 = Strongly disagree)
32. Do you believe that being a housewife or homeworker is just as fulfilling as working for pay? (1 = Strongly agree, . . . , 4 = Strongly disagree)
33. Do you agree that when jobs are scarce, men should have more right to a job than women? (1 = Strongly agree, . . . , 5 = Strongly disagree)
34. Do you believe that when jobs are scarce, employers should give priority to people of this country over immigrants? (1 = Strongly agree, . . . , 5 = Strongly disagree)
35. Do you think that if a woman earns more money than her husband, it is almost certain to cause problems? (1 = Strongly agree, . . . , 5 = Strongly disagree)
36. Do you agree that homosexual couples are as good parents as other couples? (1 = Strongly agree, . . . , 5 = Strongly disagree)
37. Do you agree that it is a duty towards society to have children? (1 = Strongly agree, . . . , 5 = Strongly disagree)
38. Do you agree that adult children have the duty to provide long-term care for their parents? (1 = Strongly agree, . . . , 5 = Strongly disagree)
39. Do you agree that people who don't work turn lazy? (1 = Strongly agree, . . . , 5 = Strongly disagree)
40. Do you agree that work is a duty towards society? (1 = Strongly agree, . . . , 5 = Strongly disagree)
41. Do you agree that work should always come first, even if it means less spare time? (1 = Strongly agree, . . . , 5 = Strongly disagree)
42. There are three basic attitudes toward society: (1) radical revolution; (2) gradual reform; (3) defend the present society. Which reflects your view?

43. Do you think "Less importance placed on work" would be good, bad, or neutral? (1 = Good; 2 = Don't mind; 3 = Bad)
44. Do you think "More emphasis on the development of technology" would be good, bad, or neutral? (1 = Good; 2 = Don't mind; 3 = Bad)
45. Do you think "Greater respect for authority" would be good, bad, or neutral? (1 = Good; 2 = Don't mind; 3 = Bad)
46. Taking all things together, would you say you are happy? (1 = Very happy; 2 = Rather happy; 3 = Not very happy; 4 = Not at all happy)
47. All in all, how would you describe your state of health? (1 = Very good; . . . ; 5 = Very poor)
48. Some people feel they have no real control over their lives; others feel they do. On a scale from 1 to 10, how much freedom do you feel you have?
49. How satisfied are you with your life as a whole? (1 = Completely dissatisfied; . . . ; 10 = Completely satisfied)
50. How satisfied are you with the financial situation of your household? (1 = Completely dissatisfied; . . . ; 10 = Completely satisfied)
51. Over the last 12 months, how often have you or your family gone without enough food to eat? (1 = Often; 2 = Sometimes; 3 = Rarely; 4 = Never)
52. Over the last 12 months, how often have you or your family felt unsafe because of crime in your home? (1 = Often; 2 = Sometimes; 3 = Rarely; 4 = Never)
53. Over the last 12 months, how often have you or your family gone without medicine or medical treatment you needed? (1 = Often; 2 = Sometimes; 3 = Rarely; 4 = Never)
54. Over the last 12 months, how often have you or your family gone without a cash income? (1 = Often; 2 = Sometimes; 3 = Rarely; 4 = Never)
55. Over the last 12 months, how often have you or your family gone without safe shelter? (1 = Often; 2 = Sometimes; 3 = Rarely; 4 = Never)
56. Comparing your standard of living with your parents' at your age, are you (1) better off, (2) worse off, or (3) about the same?
57. Generally speaking, would you say most people can be trusted (1) or that you need to be very careful (2)?
58. Do you trust people from your family completely, somewhat, not very much, or not at all? (1–4 scale)
59. Do you trust people from your neighborhood completely, somewhat, not very much, or not at all? (1–4)
60. Do you trust people you know personally completely, somewhat, not very much, or not at all? (1–4)
61. Do you trust people you meet for the first time completely, somewhat, not very much, or not at all? (1–4)
62. Do you trust people of another religion completely, somewhat, not very much, or not at all? (1–4)
63. Do you trust people of another nationality completely, somewhat, not very much, or not at all? (1–4)
64. How much confidence do you have in the churches? (1 = A great deal; . . . ; 4 = None at all)
65. How much confidence do you have in the armed forces? (1–4)
66. How much confidence do you have in the press? (1–4)
67. How much confidence do you have in television? (1–4)
68. How much confidence do you have in labor unions? (1–4)
69. How much confidence do you have in the police? (1–4)
70. How much confidence do you have in the courts? (1–4)
71. How much confidence do you have in the government? (1–4)
72. How much confidence do you have in political parties? (1–4)
73. How much confidence do you have in parliament? (1–4)
74. How much confidence do you have in the civil service? (1–4)
75. How much confidence do you have in universities? (1–4)
76. How much confidence do you have in elections? (1–4)
77. How much confidence do you have in major companies? (1–4)
78. How much confidence do you have in banks? (1–4)
79. How much confidence do you have in environmental organizations? (1–4)
80. How much confidence do you have in women's organizations? (1–4)
81. How much confidence do you have in charitable/humanitarian organizations? (1–4)
82. How much confidence do you have in the European Union? (1–4)
83. How much confidence do you have in the United Nations? (1–4)
84. How much confidence do you have in the International Monetary Fund (IMF)? (1–4)

85. How much confidence do you have in the International Criminal Court (ICC)? (1–4)
86. How much confidence do you have in NATO? (1–4)
87. How much confidence do you have in the World Bank? (1–4)
88. How much confidence do you have in the World Health Organization (WHO)? (1–4)
89. How much confidence do you have in the World Trade Organization (WTO)? (1–4)
90. Should international organizations prioritize effectiveness (1) or democracy (10)? (scale 1–10)
91. Five countries have permanent UN Security Council seats. Which is *not* a member? (1 = France; 2 = China; 3 = India)
92. Where are the IMF headquarters located? (1 = Washington, DC; 2 = London; 3 = Geneva)
93. Which problem does Amnesty International primarily deal with? (1 = Climate change; 2 = Human rights; 3 = Destruction of historic monuments)
94. Are you an active (2), inactive (1), or non-member (0) of a church or religious organization?
95. ... of a sport or recreational organization?
96. ... of an art, music, or educational organization?
97. ... of a labor union?
98. ... of a political party?
99. ... of an environmental organization?
100. ... of a professional association?
101. ... of a humanitarian or charitable organization?
102. ... of a consumer organization?
103. ... of a self-help or mutual aid group?
104. ... of a women's group?
105. ... of any other type of organization?
106. On a scale from 1 ("Incomes should be made more equal") to 10 ("Greater incentives for individual effort"), where do you place yourself?
107. On a scale from 1 (private ownership) to 10 (government ownership) of business and industry, where do you place yourself?
108. On a scale from 1 (government responsibility) to 10 (personal responsibility), where do you place yourself?
109. On a scale from 1 (competition is good) to 10 (competition is harmful), where do you place yourself?
110. On a scale from 1 (hard work brings a better life) to 10 (success is luck/connect ions), where do you place yourself?
111. Which statement comes closer to your own view on environment vs. growth? (1 = Protect environment; 2 = Promote growth)
112. How would you rate corruption in your country on a scale from 1 (none) to 10 (abundant)?
113. Among state authorities, how many are involved in corruption? (1 = none; 2 = few; 3 = most; 4 = all)
114. Among business executives, how many are involved in corruption? (1–4)
115. Among local authorities, how many are involved in corruption? (1–4)
116. Among civil service providers (police, judiciary, civil servants, doctors, teachers), how many are involved in corruption? (1–4)
117. Among journalists and media, how many are involved in corruption? (1–4)
118. How often do ordinary people in your neighborhood have to pay a bribe, give a gift, or do a favor to get local services? (1 = never; 2 = rarely; 3 = frequently; 4 = always)
119. "On the whole, women are less corrupt than men." Do you (1) strongly agree, (2) agree, (3) disagree, (4) strongly disagree, or (0) hard to say?
120. How high is the risk of being held accountable for giving or receiving a bribe? (1–10)
121. How would you evaluate the impact of immigrants on your country's development? (5 = Very good; ...; 1 = Very bad)
122. Do you agree that immigration fills job vacancies? (2 = Agree; 1 = Hard to say; 0 = Disagree)
123. Do you agree that immigration strengthens cultural diversity? (2 = Agree; 1 = Hard to say; 0 = Disagree)
124. Do you agree that immigration increases the crime rate? (2 = Agree; 1 = Hard to say; 0 = Disagree)
125. Do you agree that immigration offers asylum to political refugees? (2 = Agree; 1 = Hard to say; 0 = Disagree)
126. Do you agree that immigration increases the risk of terrorism? (2 = Agree; 1 = Hard to say; 0 = Disagree)
127. Do you agree that immigration offers people from poor countries a better living? (2 = Agree; 1 = Hard to say; 0 = Disagree)

128. Do you agree that immigration increases unemployment? (2 = Agree; 1 = Hard to say; 0 = Disagree)
129. Do you agree that immigration leads to social conflict? (2 = Agree; 1 = Hard to say; 0 = Disagree)
130. What should government policy be on allowing foreigners to work? (1 = Open door; 2 = Jobs available; 3 = Strict limits; 4 = Prohibit)
131. How secure do you feel these days? (1 = Very secure; . . . ; 4 = Not at all secure)
132. How often do robberies occur in your neighborhood? (1 = Very frequently; . . . ; 4 = Not at all frequently)
133. How often does alcohol consumption in the streets occur? (1–4)
134. How often do police or military interfere with people's private lives? (1–4)
135. How often does racist behavior occur? (1–4)
136. How often do drug sales occur? (1–4)
137. How often do street violence and fights occur? (1–4)
138. How often does sexual harassment occur? (1–4)
139. Have you avoided carrying much money for security reasons? (1 = Yes; 2 = No)
140. Have you preferred not to go out at night for security reasons? (1 = Yes; 2 = No)
141. Have you carried a weapon for security reasons? (1 = Yes; 2 = No)
142. To what degree are you worried about losing your job? (1 = Very much; . . . ; 4 = Not at all)
143. To what degree are you worried about not being able to give your children a good education? (1–4)
144. Have you personally been the victim of a crime in the last year? (1 = Yes; 2 = No)
145. Has someone in your immediate family been the victim of a crime in the last year? (1 = Yes; 2 = No)
146. To what degree are you worried about a war involving your country? (1–4)
147. To what degree are you worried about a terrorist attack? (1–4)
148. To what degree are you worried about a civil war? (1–4)
149. If you had to choose between freedom (1) and equality (2), which is more important?
150. If you had to choose between freedom (1) and security (2), which is more important?
151. If war came to your country, would you be willing to fight for it? (1 = Yes; 2 = No)
152. Which goal is most important for your country in the next ten years? (1 = High economic growth; 2 = Strong defense; 3 = More say in decisions; 4 = Beautify environment)
153. Which is second-most important? (same coding)
154. If you had to choose, which is most important? (1 = Maintaining order; 2 = More say in government; 3 = Fighting inflation; 4 = Protecting freedom of speech)
155. Which is next-most important? (same coding)
156. Which is most important? (1 = Stable economy; 2 = Progress to a more humane society; 3 = Ideas count more than money; 4 = Fighting crime)
157. Which is next-most important? (same coding)
158. To what extent do you agree that science and technology make our lives healthier, easier, and more comfortable? (1 = Completely disagree; . . . ; 10 = Completely agree)
159. To what extent do you agree that because of science and technology there will be more opportunities for the next generation? (1–10)
160. To what extent do you agree that we depend too much on science and not enough on faith? (1–10)
161. To what extent do you agree that science breaks down people's ideas of right and wrong? (1–10)
162. To what extent do you agree that it is not important to know about science in your daily life? (1–10)
163. All things considered, is the world better off or worse off because of science and technology? (1 = Much worse off; . . . ; 10 = Much better off)
164. How important is God in your life? (1 = Not at all; . . . ; 10 = Very important)
165. Do you believe in God? (1 = Yes; 2 = No)
166. Do you believe in life after death? (1 = Yes; 2 = No)
167. Do you believe in hell? (1 = Yes; 2 = No)
168. Do you believe in heaven? (1 = Yes; 2 = No)
169. When science and religion conflict, is religion always right? (1 = Strongly agree; . . . ; 4 = Strongly disagree)
170. The only acceptable religion is your religion. (1 = Strongly agree; . . . ; 4 = Strongly disagree)
171. Apart from weddings and funerals, how often do you attend religious services? (1 = More than once a week; . . . ; 7 = Never)
172. Apart from weddings and funerals, how often do you pray? (1 = Several times a day; . . . ; 8 = Never)

173. Independently of whether you attend services, would you say you are (1) a religious person, (2) not a religious person, or (3) an atheist?
174. The basic meaning of religion is: (1) to follow religious norms and ceremonies; (2) to do good to other people.
175. The basic meaning of religion is: (1) to make sense of life after death; (2) to make sense of life in this world.
176. "Nowadays one often has trouble deciding which moral rules are right to follow." Do you agree? (1 = Completely agree; . . . ; 10 = Completely disagree)
177. To what extent do you think claiming government benefits to which you are not entitled is justifiable? (1 = Never justifiable; . . . ; 10 = Always justifiable)
178. To what extent do you think avoiding a fare on public transport is justifiable? (1–10)
179. To what extent do you think stealing property is justifiable? (1–10)
180. To what extent do you think cheating on taxes is justifiable? (1–10)
181. To what extent do you think accepting a bribe in the course of duty is justifiable? (1–10)
182. To what extent do you think homosexuality is justifiable? (1–10)
183. To what extent do you think prostitution is justifiable? (1–10)
184. To what extent do you think abortion is justifiable? (1–10)
185. To what extent do you think divorce is justifiable? (1–10)
186. To what extent do you think sex before marriage is justifiable? (1–10)
187. To what extent do you think suicide is justifiable? (1–10)
188. To what extent do you think euthanasia is justifiable? (1–10)
189. To what extent do you think it is justifiable for a man to beat his wife? (1–10)
190. To what extent do you think it is justifiable for parents to beat their children? (1–10)
191. To what extent do you think violence against other people is justifiable? (1–10)
192. To what extent do you think terrorism as a political, ideological, or religious means is justifiable? (1–10)
193. To what extent do you think having casual sex is justifiable? (1–10)
194. To what extent do you think political violence is justifiable? (1–10)
195. To what extent do you think the death penalty is justifiable? (1–10)
196. Should the government have the right to keep people under video surveillance in public areas? (1 = Definitely should; . . . ; 4 = Definitely should not)
197. Should the government have the right to monitor all emails and Internet traffic? (1–4 as above)
198. Should the government have the right to collect information about anyone living in the country without their knowledge? (1–4 as above)
199. How interested are you in politics? (1 = Very interested; . . . ; 4 = Not at all interested)
200. When you get together with friends, would you say you discuss political matters (1) frequently, (2) occasionally, or (3) never?
201. How often do you use a daily newspaper for information? (1 = Daily; 2 = Weekly; 3 = Monthly; 4 = Less than monthly; 5 = Never)
202. How often do you watch TV news for information? (1–5)
203. How often do you listen to radio news for information? (1–5)
204. How often do you use a mobile phone for information? (1–5)
205. How often do you use email for information? (1–5)
206. How often do you use the Internet for information? (1–5)
207. How often do you use social media for information? (1–5)
208. How often do you talk with friends or colleagues to obtain information? (1–5)
209. Have you ever signed a petition? (1 = Have done; 2 = Might do; 3 = Would never do)
210. Have you ever joined in boycotts? (1–3)
211. Have you ever attended peaceful demonstrations? (1–3)
212. Have you ever joined strikes? (1–3)
213. Have you ever donated to a group or campaign? (1–3)
214. Have you ever contacted a government official? (1 = Have done; 2 = Might do; 3 = Would never do)
215. Have you ever encouraged others to take action about political issues? (1–3)
216. Have you ever encouraged others to vote? (1–3)
217. Have you ever searched for political information online? (1 = Have done; 2 = Might do; 3 = Would never do; –3 = Not applicable)
218. Have you ever signed an electronic petition? (1–3; –3 = Not applicable)

219. Have you ever encouraged others to take political action via social media? (1–3; –3 = Not applicable)
220. Have you ever organized political activities or protests via social media? (1–3; –3 = Not applicable)
221. Do you vote in local-level elections? (1 = Always; 2 = Usually; 3 = Never; 4 = Not allowed)
222. Do you vote in national-level elections? (1–4)
223. If there were a national election tomorrow, for which party would you vote? (numbered list)
224. How often are votes counted fairly? (1 = Very often; . . . ; 4 = Not at all)
225. How often are opposition candidates prevented from running? (1–4)
226. How often does TV news favor the governing party? (1–4)
227. How often are voters bribed? (1–4)
228. How often do journalists provide fair coverage of elections? (1–4)
229. How often are election officials fair? (1–4)
230. How often do wealthy individuals buy elections? (1–4)
231. How often are voters threatened with violence at the polls? (1–4)
232. How often are voters offered a genuine choice in elections? (1–4)
233. How often do women have equal opportunities to run for office? (1–4)
234. "How important would you say having honest elections is for you?" (1 = Very important; . . . ; 4 = Not at all important)
235. How do you feel about having a strong leader who bypasses parliament and elections? (1 = Very good; . . . ; 4 = Very bad)
236. How do you feel about experts, not the government, making decisions for the country? (1–4)
237. How do you feel about the army ruling the country? (1–4)
238. How do you feel about a democratic political system? (1–4)
239. How do you feel about a system governed by religious law with no parties or elections? (1–4)
240. "In political matters, people often talk about 'the left' and 'the right.' How would you place your views?" (1 = Left; 10 = Right)
241. How essential is it for democracy that governments tax the rich and subsidize the poor? (1 = Not at all essential; . . . ; 10 = Definitely essential; 0 if against democracy)
242. How essential is it for democracy that religious authorities interpret the laws? (1–10; 0 if against democracy)
243. How essential is it for democracy that people choose their leaders in free elections? (1–10; 0 if against democracy)
244. How essential is it for democracy that people receive state aid for unemployment? (1–10; 0 if against democracy)
245. How essential is it for democracy that the army takes over when the government is incompetent? (1–10; 0 if against democracy)
246. How essential is it for democracy that civil rights protect people from state oppression? (1–10; 0 if against democracy)
247. How essential is it for democracy that the state makes people's incomes equal? (1–10; 0 if against democracy)
248. How essential is it for democracy that people obey their rulers? (1–10; 0 if against democracy)
249. How essential is it for democracy that women have the same rights as men? (1–10; 0 if against democracy)
250. How important is it for you to live in a country governed democratically? (1 = Not at all important; . . . ; 10 = Absolutely important)
251. How democratically is this country being governed today? (1 = Not at all democratic; . . . ; 10 = Completely democratic)
252. On a scale from 1 to 10, how satisfied are you with how the political system is functioning?
253. How much respect is there for individual human rights nowadays in this country? (1 = A great deal; 2 = Fairly much; 3 = Not much; 4 = None at all)
254. How proud are you to be [country's nationality]? (1 = Very proud; . . . ; 5 = Not at all / not your nationality)
255. How close do you feel to your village, town, or city? (1 = Very close; . . . ; 4 = Not close at all)
256. How close do you feel to your county, region, or district? (1–4)
257. How close do you feel to your country? (1–4)
258. How close do you feel to your continent? (1–4)
259. How close do you feel to the world? (1–4)
260. What is the respondent's sex? (1 = Male; 2 = Female)
261. What is your year of birth? (_____)

262. Based on your year of birth, how old are you? (_____ years)
263. Were you born in this country or are you an immigrant? (1 = Born here; 2 = Immigrant)
264. Is your mother an immigrant? (1 = No; 2 = Yes)
265. Is your father an immigrant? (1 = No; 2 = Yes)
266. Respondent's country of birth: _____; ISO code: _____
267. Mother's country of birth: _____; ISO code: _____
268. Father's country of birth: _____; ISO code: _____
269. Are you a citizen of this country? (1 = Yes; 2 = No)
270. Including yourself, how many people live here regularly as household members? (_____)
271. Do you live with your own parents or parents-in-law? (1 = No; 2 = Own parents; 3 = Parents-in-law; 4 = Both)
272. What language do you normally speak at home? (1–5, customize as needed)
273. What is your current marital status? (1 = Married; 2 = Living together; 3 = Divorced; 4 = Separated; 5 = Widowed; 6 = Single)
274. Do you have any children? (0 if none; otherwise number: _____)
275. What is the highest educational level you have attained? (0 = ISCED 0; . . . ; 8 = ISCED 8; –3 = Not applicable)
276. What is the highest educational level your spouse has attained? (same coding)
277. What is the highest educational level your mother has attained? (same coding)
278. What is the highest educational level your father has attained? (same coding)
279. Are you currently employed? (1 = Full-time; 2 = Part-time; 3 = Self-employed; 4 = Retired; 5 = Housewife; 6 = Student; 7 = Unemployed; 8 = Other)
280. Is your spouse currently employed? (same coding)
281. To which occupational group do you belong? (0 = Never had a job; 1 = Professional; 2 = Higher administrative; 3 = Clerical; 4 = Sales; 5 = Service; 6 = Skilled worker; 7 = Semi-skilled worker; 8 = Unskilled worker; 9 = Farm worker; 10 = Farm proprietor)
282. To which occupational group does your spouse belong? (same coding)
283. When you were 14, to which occupational group did your father belong? (same coding)
284. Are you working for a (1) government/public institution, (2) private business/industry, or (3) private non-profit organization?
285. Are you the chief wage-earner in your household? (1 = Yes; 2 = No)
286. During the past year, how would you describe your family's financial situation? (1 = Saved money; 2 = Just got by; 3 = Spent some savings; 4 = Spent savings and borrowed money)
287. People sometimes describe themselves as belonging to a social class. How would you describe yourself? (1 = Upper class; 2 = Upper middle class; 3 = Lower middle class; 4 = Working class; 5 = Lower class)
288. On an income scale where 1 represents the lowest income group and 10 the highest, which group does your household fall into? (1–10)
289. Do you belong to a religion or religious denomination? (0 = No; 1 = Roman Catholic; 2 = Protestant; 3 = Orthodox; 4 = Jewish; 5 = Muslim; 6 = Hindu; 7 = Buddhist; 8 = Other)
290. What is your ethnic group? (1 = White; 2 = Black; 3 = South Asian; 4 = East Asian; 5 = Arabic/Central Asian; 6 = Other)

C. Responses to World Values Survey Questions By Custom-Trained LLMs of Abrahamic Scholars

The pairwise plots below illustrate the central-tendency measures (mean, median, and mode) of the responses generated by the custom-trained language models for the World Values Survey questions. In these plots, blue shading denotes a response value of 0, while red shading denotes a response value of 10 or greater.

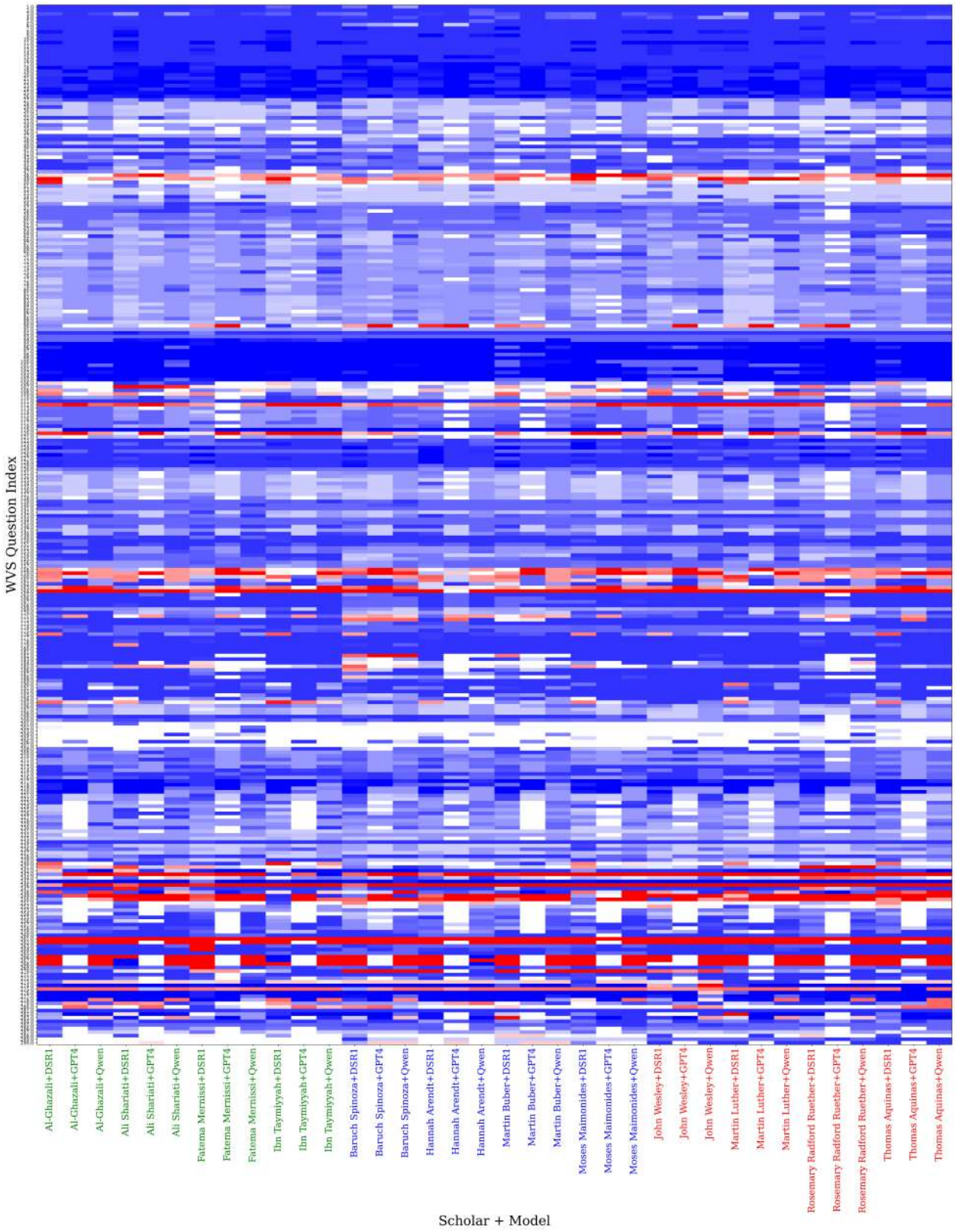

Fig. 5: Mean response values produced by the custom-trained large language models for World Values Survey items. *Note:* The color gradient ranges from blue (response value of 0) to red (response value of 10 or greater). White indicates responses outside the acceptable range.

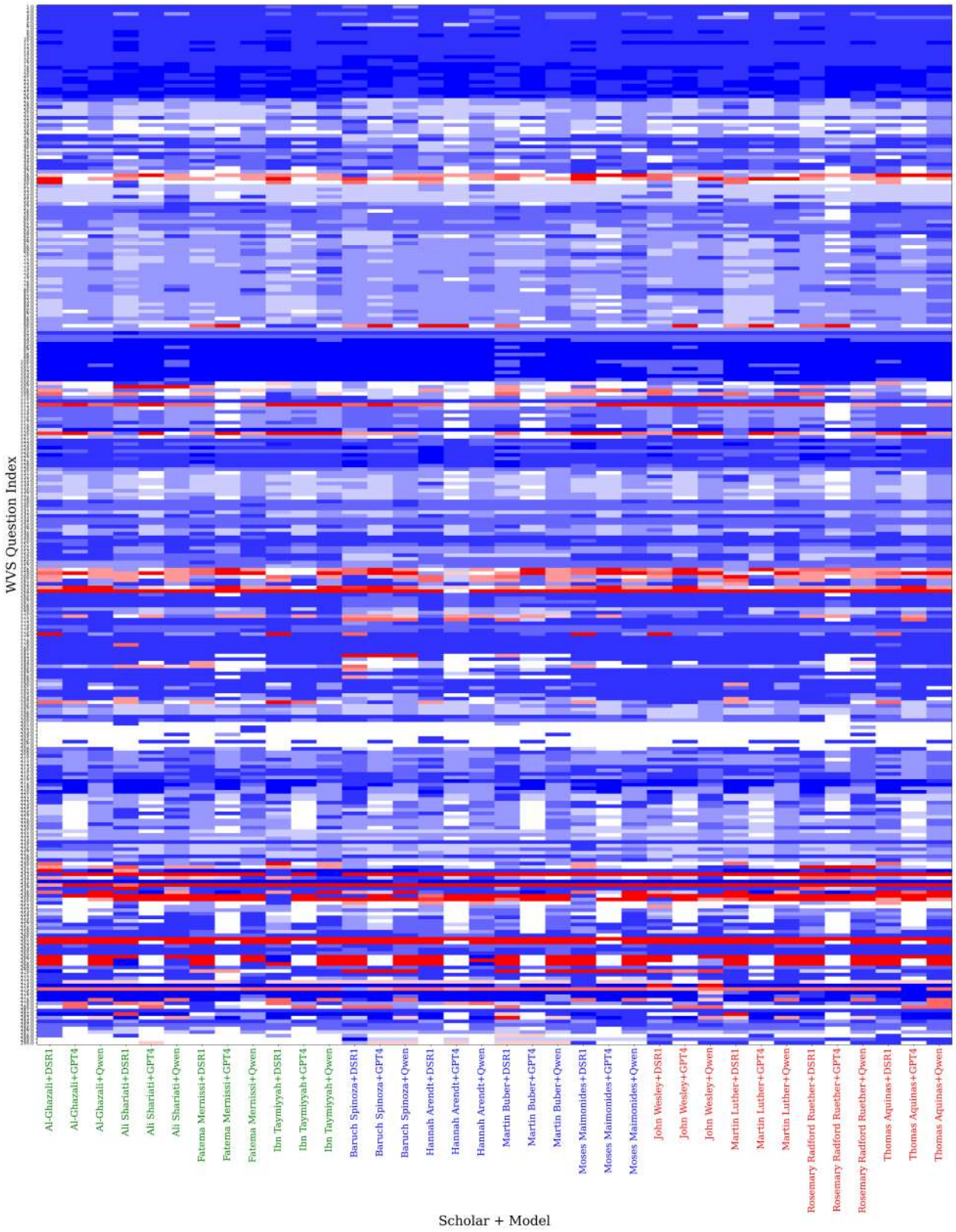

Fig. 6: Median response values produced by the custom-trained large language models for World Values Survey items. *Note:* The color gradient ranges from blue (response value of 0) to red (response value of 10 or greater). White indicates responses outside the acceptable range.

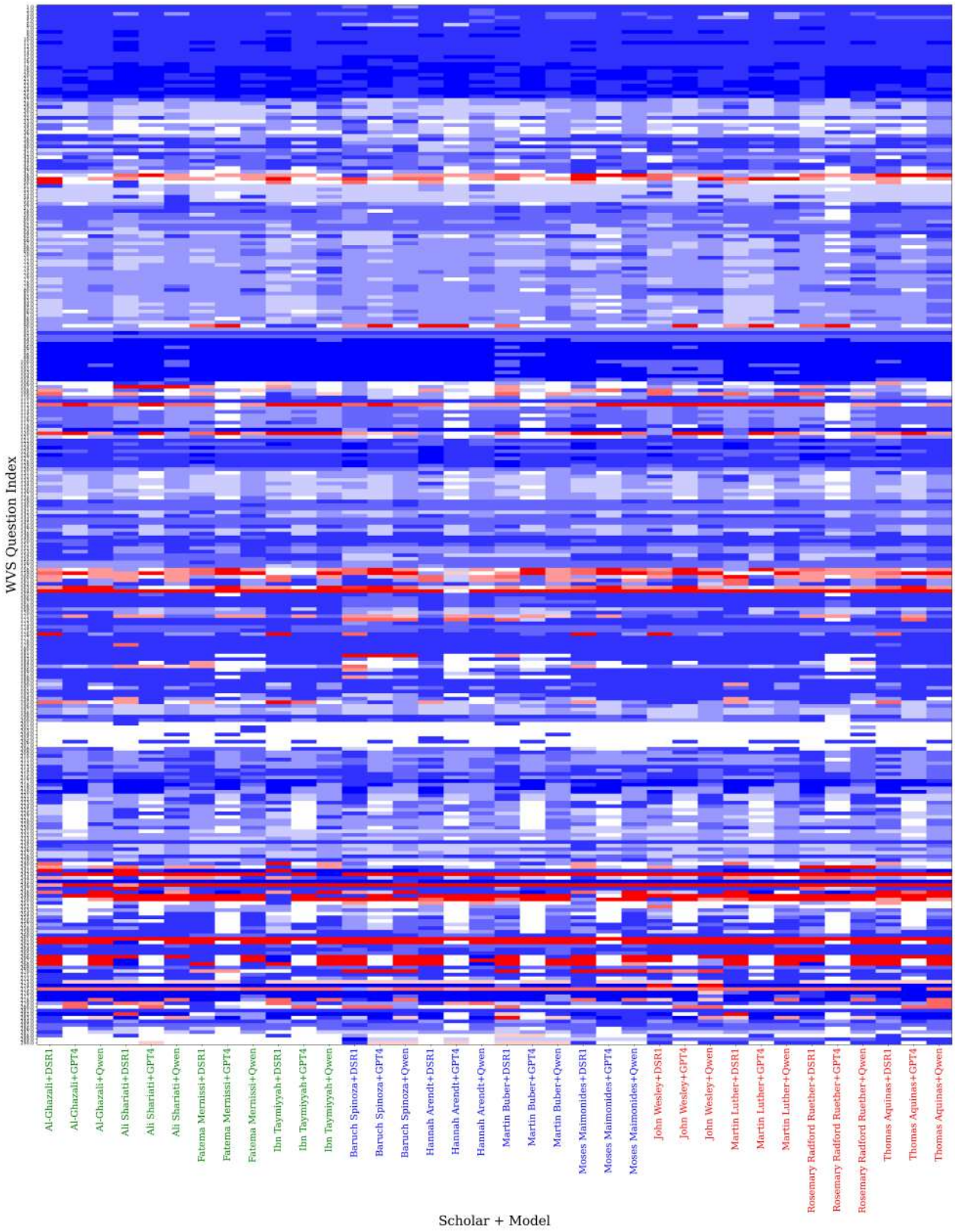

Fig. 7: Mode response values produced by the custom-trained large language models for World Values Survey items. *Note:* The color gradient ranges from blue (response value of 0) to red (response value of 10 or greater). White indicates responses outside the acceptable range.