Working paper, November 2026

Resurrecting Theology: When Historical Scholars Confront the Modern World Through AI

Cantay Caliskan,* Mija Aleksandraviciute, and Anna Fang

University of Rochester, Rochester, NY, USA
*Corresponding author. Email: cantay.caliskan@rochester.edu

> *"present of things past, memory; present of things present, sight; present of things future, expectation."*
>
> Augustine, *Confessions* XI.20.26, trans. E. B. Pusey

Abstract

Chatbots increasingly answer questions in the voice of religious figures, yet little is known about whose views such answers express. We asked six large language models to answer the World Values Survey and a Pew Global Attitudes questionnaire as each of 23 religious and intellectual figures associated with Judaism, Christianity, Islam, Buddhism and Hinduism. Two were frontier models accessed through commercial APIs (GPT-5.4 mini, DeepSeek-V3.2) and four were smaller open-weight models from laboratories in the traditions' regions (DictaLM 2.0, Jais 2, Krutrim-2, GLM-4). Personas were queried with and without passages retrieved from scholar-specific corpora, yielding 528,375 recorded responses that we compared with each other and with human survey data. In the main WVS retrieval panel, model identity accounts for 37% of item-level variance and tradition and scholar together for 17%, with the remainder reflecting model-scholar interaction and noise; two models portraying the same scholar disagree more than one model portraying two scholars. Documented disagreements between scholars point the expected way in about half of the model-contrast comparisons and the opposite way in about a sixth. Retrieval spreads personas apart but does not improve these contrasts, and its apparent reduction of the model's variance share is smaller and uncertain on items that every persona answered. The regional models show no detectable home-tradition advantage under this English-language setup. The models also place the same scholars very differently: without retrieval, DeepSeek puts the median scholar at the 85th percentile of WVS respondents on secular values and DictaLM at the 5th. What a resurrected scholar says depends, for now, more on the model and its pipeline than on the scholar.

Keywords: large language models, religion, persona simulation, retrieval-augmented generation, World Values Survey, Pew Global Attitudes, cultural alignment

1. Introduction

What if Thomas Aquinas could offer a view on gender roles, or Maimonides on the regulation of global finance? Questions of this kind were once a parlour game for historians of ideas. Large language models (LLMs) now answer them on request, in the first person and with the fluency of a seminar paper. Chatbots that speak in the voice of Jesus, answer questions about the Talmud or summarize hadith collections are already in public use (The Center for Unity 2026; RavGPT 2026; Islam&AI 2026), and they give users another route to religious texts and their interpretation. Whether an answer produced in a scholar's name tells us anything about that scholar, rather than about the model producing it, is an empirical question. It has received little systematic study.

This paper takes up that question for 23 religious and intellectual figures associated with Judaism, Christianity, Islam, Buddhism and Hinduism, from Adi Shankara in the eighth century to the 14th

Dalai Lama today. We ask six LLMs to answer, in each scholar's persona, the full questionnaire of the seventh wave of the World Values Survey (WVS; Haerpfer et al. 2024) and 87 items selected from the Spring 2024 Pew Global Attitudes Survey (Pew Research Center 2025). Personas are queried in two conditions. In the first the model receives only the scholar's name and tradition; in the second it also receives passages retrieved from a corpus assembled for that scholar, which consists mostly, though not entirely, of the scholar's own works in translation. Each question is asked five times. After parsing and validation, the 528,375 responses are compared across models, scholars and conditions, and with the answers of 97,220 WVS respondents in 66 countries and Pew respondents in 13 countries.

The study sits where two lines of research meet. The first treats LLMs as simulated survey respondents. Conditioning a model on a demographic backstory can reproduce the opinion patterns of real subgroups (Argyle et al. 2023), but the models' default answers resemble those of English-speaking and Protestant European publics (Santurkar et al. 2023; Durmus et al. 2023; Tao et al. 2024), and some apparent alignment disappears once answer order and labels are varied (Dominguez-Olmedo, Hardt, and Mendler-Dünner 2024). The second line studies religion in digital media and, more recently, in AI: how online tools redistribute religious authority (Campbell 2007, 2012), how LLMs represent religious groups (Abid, Farooqi, and Zou 2021; Plaza-del-Arco et al. 2024), and how reliably they answer questions of religious law (Atif et al. 2025). Persona simulation of religious authorities is where these lines meet, and where questions of fidelity also become questions of religious authority. A user who asks what al-Ghazālī would think about women's employment receives an answer that borrows the authority of a twelfth-century theologian. One motivation sometimes offered for such systems is that believers might consult foundational thinkers in this way without the mediation of clergy or commentators. We treat this as a possibility to be tested rather than a goal to be endorsed, and the evidence below bears directly on it.

The paper extends an earlier study (Caliskan et al. 2025), which simulated twelve Abrahamic scholars with three commercial models, grounded every persona in the scholar's texts through retrieval-augmented generation (RAG; Lewis et al. 2020), and compared their WVS answers with contemporary populations. That study found that the models agreed on average but differed in the medians and modes of their answers, and that Muslim personas sat closer to the global average while Christian personas resembled their local populations. Its design left three questions open. Because every persona was grounded in retrieved text, it could not separate what the retrieved passages contributed from what a model already associated with a famous name. Because all three models were commercial systems trained mainly on English-language data, it could not ask whether models built in a tradition's home region portray that tradition differently. And because it averaged raw answer codes across items with different scales, its comparisons mixed substance with scale length.

The present study addresses each of these. The panel now spans five traditions and twelve centuries, includes three women (Hannah Arendt, Rosemary Radford Ruether and Fatema Mernissi) and one living figure (the 14th Dalai Lama), and adds eleven Buddhist and Hindu scholars to the twelve Abrahamic ones. Personas are run with and without retrieval, so that persona-only and retrieval-augmented answers can be compared. The model panel is built on provenance. Two frontier models accessed through commercial APIs, GPT-5.4 mini (OpenAI 2026b) and DeepSeek-V3.2 (DeepSeek-AI 2025), sit beside four smaller models with open weights from laboratories rooted in the traditions' regions. These are DictaLM 2.0 from Israel (Shmidman et al. 2024), Jais 2 from the United Arab Emirates (Anwar et al. 2026), Krutrim-2 from India (Kallappa et al. 2025) and GLM-4-9B-0414 from China (Z.ai (Zhipu AI) 2025). Because the four regional models were run on all 23 scholars, we can test whether a model portrays its home tradition more distinctively than other models do. Finally, the measurement is rebuilt. Answers are normalized item by item, generation failures are kept apart from "don't know" answers, demographic questions are removed from the value analyses, and the data are tested for response artifacts before they are interpreted. Raw responses are not imputed; sparse missing cells are filled only for the variance decomposition, and those results

are checked on complete-case items.

Our central finding concerns the balance between model and scholar. With retrieval, the identity of the model accounts for 37% of the item-level variance in WVS answers and the identity of the tradition and scholar together for 17%; the remaining 46% is specific to particular model-scholar combinations or is noise. Two models portraying the same scholar also disagree more than one model portraying two different scholars. The personas are not empty, however. Documented disagreements between scholars, such as that between Mernissi and Ibn Taymiyya on gender, point the expected way in 5 of 5 contrast families when the differences are averaged over models (one-sided exact $p = 0.031$, the smallest value five families allow). Retrieval changes the response profiles and spreads the personas apart, but it does not improve the seven historical contrasts tested here, and its apparent reduction of the model's share of the variance is smaller and statistically uncertain on items that every persona answered. It sharpens the separation of traditions only for the two commercial models. The regional models showed no detectable home-tradition advantage in our tests. With retrieval, we found no average advantage of a scholar's reference state over an arbitrary country in closeness to the persona, while without retrieval the reference states ranked somewhat better than chance. The differences between models in where they place the scholars are large. Without retrieval, DeepSeek places the median scholar at the 85th percentile of WVS respondents on secular values, while DictaLM, which chooses the first answer option for most items, places the median scholar at the 5th. These are properties of the tested models and their serving pipelines; the design has no persona-free baseline, so they are not direct measurements of each model's default values.

The next section reviews the research on simulated respondents, cultural alignment and religion in AI. Section 3 lists the research questions and Section 4 the hypotheses. Sections 5 and 6 describe the data and the measurement, Section 7 reports the results, and Section 8 discusses what they imply for claims that AI can speak for religious authorities.

2. Literature review

2.1 Language models as survey respondents and personas

The idea that a language model can stand in for a person in a survey rests on the claim of "algorithmic fidelity" made by Argyle et al. (2023). Conditioned on demographic backstories, GPT-3 reproduced the voting and attitude patterns of American subgroups closely enough that the authors proposed "silicon samples" as a complement to human panels. Later work has been more cautious. On the OpinionQA benchmark, built from American opinion surveys, language models diverged substantially from all sixty demographic groups tested, and steering a model toward a group narrowed the gap only partly (Santurkar et al. 2023). Bisbee et al. (2024) found that simulated respondents recovered average feeling-thermometer scores in the American National Election Studies but understated their variance, distorted the relationships among variables and gave different answers to the same prompts a few months apart. Studies of the response process itself point the same way. Models do not reproduce the familiar effects of question wording and order on human respondents, yet they react to changes that leave humans unmoved (Tjuatja et al. 2024). Forced-choice formats can produce value profiles that change once the model is allowed to answer freely or the prompt is paraphrased (Röttger et al. 2024). Most directly relevant here, Dominguez-Olmedo, Hardt, and Mendler-Dünner (2024) showed that once the order and labels of answer options are randomized, the survey answers of many models move toward uniform noise, so that apparent alignment with a group can be an artifact of position preferences. We take these critiques as design constraints. They are the reason the present study repeats every question, keeps failures and abstentions apart from substantive answers, and tests whether a model's answers depend on the content of an item or only on the position of an option.

Whether a persona changes much at all is a separate question. Across a range of subjective annotation tasks, Hu and Collier (2024) found that persona variables explain little of the variance in

human judgments and that persona prompting improves model predictions only modestly where they do matter. Wang, Morgenstern, and Dickerson (2024) argued, with empirical tests, that models asked to speak for identity groups tend to reproduce how outsiders see a group and to flatten the diversity within it. A model asked to speak as a Hanbali jurist or a Zen abbot is in a similar position. Its portrait may draw more on what the training data say about such figures than on what the figures themselves wrote. Shanahan, McDonell, and Reynolds (2023) offer a useful interpretive frame: dialogue agents are better described as playing characters, or a superposition of characters, than as holding beliefs. On this view, the question is not whether a persona "believes" what Aquinas believed but how well the character it plays matches the documented positions of the person it is named after.

Historical figures raise problems that contemporary groups do not. Varnum et al. (2024) proposed "historical LLMs" trained on texts from a given period as instruments for studying past psychologies, and noted the sampling problems such models inherit from what happens to survive in the written record. Character-LLM trained agents to play specific historical people from reconstructed experiences and evaluated them by interview (Shao et al. 2023), and generative agents with memory and reflection showed that LLM characters can sustain believable individual and social behavior over time (Park et al. 2023). Our approach differs from these in two ways. We do not train or fine-tune any model; the personas are produced by prompting alone, with or without passages retrieved from the scholar's corpus, so that the comparison isolates what retrieved text adds to a model's prior. And we evaluate the personas with standardized survey instruments, which makes their answers comparable with each other and with contemporary populations, at the cost of asking pre-modern thinkers about matters they never considered.

Retrieval is the natural way to ground a persona in a corpus. RAG pairs a generator with a retriever over an external collection and improves performance on knowledge-intensive tasks (Lewis et al. 2020), and retrieval over a large corpus can substitute for much of the knowledge otherwise stored in parameters (Borgeaud et al. 2022). Retrieval has also done well when the aim is to inject knowledge. In the tasks evaluated by Ovadia et al. (2024), RAG outperformed unsupervised fine-tuning for both familiar and new facts, and in a related comparison its advantage was largest for less popular entities (Soudani, Kanoulas, and Hasibi 2024), which describes most of the scholars studied here outside the European canon. A recent benchmark for Islamic question answering found that agentic retrieval over Qur'anic evidence improves on standard RAG (Bhatia et al. 2026). These results concern factual recall. Whether retrieval makes a persona's value judgments more faithful to its source is a separate question, and one that standard evaluations, built around multiple criteria of accuracy and calibration (Liang et al. 2023), are not designed to answer.

2.2 Whose values: cultural alignment and regional models

When LLMs answer cross-cultural value surveys without guidance, their answers look most like those of a small set of societies. GlobalOpinionQA, built from Pew Global Attitudes and WVS items, found that a model's default answers resembled those of the United States and some European and South American countries, and that prompting it to answer as a given country could produce stereotyped answers (Durmus et al. 2023). Atari et al. (2023) reported that GPT's answers on psychological measures resemble those of WEIRD populations (Henrich, Heine, and Norenzayan 2010) and drift further from human answers as cultural distance from the United States grows. Tao et al. (2024) placed five GPT models on the Inglehart–Welzel cultural map using WVS items and found them closest to English-speaking and Protestant European countries; asking the model to answer as a person from a given country improved alignment for most countries. The cultural map behind this work summarizes cross-national variation along two dimensions, traditional versus secular-rational values and survival versus self-expression values (Inglehart and Welzel 2005), and evolutionary modernization theory links movement along them to rising existential security and the decline of traditional religiosity (Inglehart 2018). Welzel (2013) recast the second dimension as

emancipative values, which we use below. For a study of religious scholars this matters directly. If a model's default is a secular, self-expressive Western respondent, a persona has to move the model a long way to sound like a medieval theologian, and it may not move it far enough.

Where a model's training data come from, and in which language it is prompted, seem to matter for these answers. Models align more closely with a culture when prompted in its dominant language and when their pretraining data contain more of it (AlKhamissi et al. 2024). Probes of English models for moral norms across countries, benchmarked against both WVS and Pew data, found that cross-national norms are estimated less accurately than English ones (Ramezani and Xu 2023). Yet regional training is no guarantee of regional values: multilingual and even Arabic-specific models favored Western-associated entities in Arabic contexts (Naous et al. 2024). Efforts to localize models explicitly, through culturally grounded instruction data and feedback (Huang et al. 2024) or through fine-tuning on data seeded from the WVS (Li et al. 2024), report gains on cultural benchmarks. These efforts respond to a wider argument that AI's data, models and developers are concentrated in WEIRD societies, with consequences for whose values AI systems express (Mihalcea et al. 2025). They also connect to proposals for "pluralistic" alignment that asks models to represent a range of views rather than a single default (Sorensen et al. 2024).

The four open-weight models in our panel come from this movement toward regional and sovereign AI. DictaLM 2.0 adapts Mistral-7B to Hebrew through continued pretraining on Hebrew and English text (Shmidman et al. 2024). Jais 2 was trained from scratch with a large Arabic share, including dialects (Anwar et al. 2026; Inception and MBZUAI and Cerebras 2025). Krutrim-2 builds on Mistral-NeMo with English and 22 Indian languages (Kallappa et al. 2025). GLM-4-9B-0414 belongs to a family trained mainly on Chinese and English (Team GLM et al. 2024; Z.ai (Zhipu AI) 2025). None of these models was built to represent a religious tradition, and the link between a model's region and a tradition is loose: Israel's population is not the population of Spinoza's Amsterdam, and China is not Tibet. The pairing nonetheless allows an exploratory test, under the English-language conditions in which such models are often used, of a claim made for regional models, namely that they carry the cultural knowledge of their home region better than models trained elsewhere.

2.3 Religion, authority and AI

Research on digital religion has long asked how new media change who speaks for a tradition. Campbell (2007) distinguished four layers of religious authority, namely hierarchy, structure, ideology and text, and showed that the internet both challenges and reinforces each. Her later model of "networked religion" describes a religious life in which community, identity and authority are negotiated across online and offline settings (Campbell 2012). LLM chatbots extend this process in a new direction, because they do not only relay religious texts but generate new statements in a tradition's voice. Table 1 lists six systems that the earlier study reviewed, with their status as of late 2026. Some retrieve and summarize canonical texts, such as the Islam&AI resources for Qur'an and hadith (Islam&AI 2026) and IslamiCity's ChatILM (IslamiCity 2025). Others answer as a religious figure, such as AskJesus, which speaks in the voice of Jesus as portrayed in *The Urantia Book* (The Center for Unity 2026). Several have changed since. The RavGPT service described in the earlier study now redirects to RavChat, which presents itself as a study partner that shows its sources and declines to issue halakhic rulings (RavGPT 2026), and Chatbot Eli no longer appears to be maintained (Chatbot Eli 2026). Whatever their design, these systems make a tacit claim of fidelity to a tradition that is rarely tested.

Evidence on how LLMs treat religion has accumulated quickly. GPT-3 associated Muslims with violence far more often than other religious groups (Abid, Farooqi, and Zou 2021). An emotion-attribution study found that models portray the major Western religions with more nuance, reduce Hinduism and Buddhism to stereotypes, and refuse more often when prompted about Judaism

Table 1. Religious chatbots reviewed in the earlier study

System	Tradition	Basis and status (Nov 2026)
RavGPT / RavChat	Judaism	Jewish texts; now a study partner that cites sources
AskJesus	Christianity	ChatGPT in the voice of Jesus of *The Urantia Book*; active
Bible AI	Christianity	Bible search, Q&A and reading plans; active
Chatbot Eli	Christianity	gotquestions.org and Bible translations; last updated 2022
Islam&AI	Islam	Open Qur'an, hadith and tafsir resources for NLP; active
ChatILM	Islam	Assistant built on IslamiCity content; launched 2025

Sources: RavGPT (2026), The Center for Unity (2026), Bible AI (2026), Chatbot Eli (2026), Islam&AI (2026), and IslamiCity (2025).

and Islam (Plaza-del-Arco et al. 2024). Indian-specific stereotypes about caste and religion are reproduced more strongly than Western ones (Khandelwal et al. 2024), and models show human-like favoritism toward ingroups and hostility toward outgroups (Hu et al. 2024). On questions of Islamic law across the four Sunni schools, models differ widely in accuracy and in their willingness to abstain, and all perform worse in Arabic than in English (Atif et al. 2025). Two recent studies from a single consortium show that models mention religion far less often than people expect when answering everyday ethical dilemmas (Wingate et al. 2026) and respond unevenly to requests for help in changing faith, encouraging some conversions more than others (Israelsen et al. 2026). A qualitative study of chatbot answers to religious questions found that they simplify doctrine while promoting a pluralist, tolerance-centred morality (Tsuria and Tsuria 2024). On the demand side, field and online experiments in a Japanese Buddhist temple, a Taoist temple in Singapore and online Christian sermons found that robot and AI preachers were judged less credible and received smaller donations than human ones (Jackson et al. 2023).

These literatures leave an empirical question open. Work on simulated respondents has asked how well LLMs represent contemporary national and demographic groups, not historical individuals whose views are documented in a written corpus. Studies of cultural alignment have compared commercial models with national publics but have rarely tested whether regional models represent their home traditions better. Research on religion and AI has documented bias and unreliability in how models talk about religions, but has not measured how faithfully models speak as religious authorities, or how much of what they say comes from the authority's texts rather than from the model. Our earlier study took a first step for the Abrahamic traditions. The present study asks the question across five traditions and six models, with a design that compares the contributions of the model and the scholar and contrasts persona-only with retrieval-augmented answers.

3. Research questions

The study asks six questions. The first two concern what a persona contains, the next three what shapes it, and the last whether the measurements can be trusted. Each question is tied to one or more of the hypotheses stated in Section 4, and Section 6 gives the estimator and test used for each.

- **RQ1, model and scholar.** How much of the variation in persona answers is due to the model that produces the persona, and how much to the scholar it portrays? When two models portray the same scholar, do they agree more than one model portraying two different scholars? (H1,

H1b)

- **RQ2, distinctiveness and known differences.** Do personas from the same tradition resemble each other more than personas from different traditions, and do scholars with documented opposing positions differ in the expected direction on the survey items that bear on those positions? (H2a, H2b)
- **RQ3, retrieval.** What changes when the persona prompt is accompanied by passages retrieved from the scholar's corpus? We ask whether the personas become more differentiated, whether the model's share of the variance falls and whether the known-group contrasts improve. (H3)
- **RQ4, provenance.** Do the four models developed in a tradition's home region portray that tradition's scholars more distinctively, or closer to the home population, than other models do? (H4)
- **RQ5, presentism and place.** Where do personas sit relative to contemporary national populations? Are they closer to the average country than to their own reference states, more secular or more emancipative than those states, and does the scholar's era leave a trace in the answers? (H5)
- **RQ6, measurement.** How valid are the answers of each model and condition? Is DictaLM's frequent choice of the first response option a substantive stance or a preference for a position on the scale? Do the main results hold on a second, independently built instrument? (H6, H7)

RQ1, RQ2 and RQ5 extend the questions of an earlier study (Caliskan et al. 2025) to a larger panel and a fuller instrument. RQ3 and RQ4 are new and depend on the two features added in this design, the persona-only condition and the regional models. RQ6 is a precondition for interpreting the others.

4. Hypotheses

We state seven hypotheses, two of which have two parts. Four of them restate in testable form the hypotheses of our earlier study: H1 and H1b correspond to its cross-model hypotheses (its H1 and H4), H2 to its cross-tradition hypothesis (H2) and H5 to its hypothesis about global and regional proximity (H3). H3, H4, H6 and H7 are new and use the extended design.

The first hypothesis concerns the balance between the model and the scholar (RQ1). Persona prompting has a modest effect on the judgments of demographic personas (Hu and Collier 2024), and the answers of different models resemble different publics (Santurkar et al. 2023; Tao et al. 2024). If the same holds for religious personas, the identity of the model should matter more than the identity of the scholar. H1 predicts that model identity explains a larger share of the variance in item responses than scholar identity does, in every panel and under both conditions. H1b predicts, as a corollary, that the same scholar is portrayed differently by different models: the distance between two models' portraits of one scholar should exceed the distance between one model's portraits of two scholars.

The second hypothesis asks whether the personas nonetheless carry information about their subjects (RQ2). The traditions differ in their teachings on authority, gender and the relation between religion and the state, and even a model with strong habits of its own should register some of these differences if the persona has any effect. H2a predicts that personas cluster by tradition, so that the distance between personas from different traditions exceeds the distance between personas from the same tradition. Clustering of this kind could also arise from stereotypes about a tradition, so it shows distinctiveness rather than fidelity. H2b is a known-groups test and the closest the design comes to a test of fidelity. Scholars with documented opposing positions should differ in the expected direction on the items that bear on those positions. We test seven such contrasts, which fall into five families because two scholars appear twice. Three concern gender equality: Mernissi against Ibn Taymiyya, and Ruether against Aquinas and against Luther. One sets Spinoza against Maimonides on the authority of religion in public life, two set Vivekananda against Shankara and against Nisargadatta on social engagement, and one sets the 14th Dalai Lama against Tsongkhapa on

the relation between Buddhist teaching and modern science. In each pair the first scholar's writings take the more egalitarian, secular, socially engaged or science-friendly position (Appendix 1 gives the textual basis for each expected direction). The contrasts are coarse by design; they ask whether the direction of a well-known difference survives, not whether its size is right.

The third hypothesis concerns grounding (RQ3). Retrieval improves factual recall in question answering (Lewis et al. 2020; Ovadia et al. 2024), and passages from a scholar's corpus might pull a persona away from the model's usual answers and toward what is particular to that scholar. H3 predicts that RAG, compared with BASE, increases differentiation between scholars, lowers the model's share of the variance and improves known-group validity.

The fourth hypothesis concerns regional models (RQ4). If training on a region's language and data transfers cultural knowledge (AlKhamissi et al. 2024), a regional model should portray its home tradition's scholars more distinctively and closer to the home population than other models do. H4 predicts a home-tradition advantage for four pairings: Jais 2 with Islam, DictaLM 2.0 with Judaism, Krutrim-2 with Hinduism and GLM-4 with Buddhism. With four models and English prompts, the test is exploratory. No regional Christian model was available, so we add GPT-5.4 mini as the Christian pairing in a further exploratory version. The choice rests on evidence that earlier GPT models' default answers sit closest to English-speaking and Protestant European values (Tao et al. 2024), which does not establish the religious leanings of the model we use.

The fifth hypothesis concerns presentism (RQ5). A model asked to speak as a medieval scholar may still answer as a person of the present, and specifically as the secular, self-expressive respondent that commercial models have been found to resemble. H5 predicts that personas sit closer to the contemporary global average than to the populations of their own regions, that they lean toward the secular and emancipative poles relative to those populations, and that the scholar's era predicts traditionalism only weakly. This extends the earlier study's proximity test, which compared personas with global and local populations on five WVS items, to the full set of value items and to two summary indices.

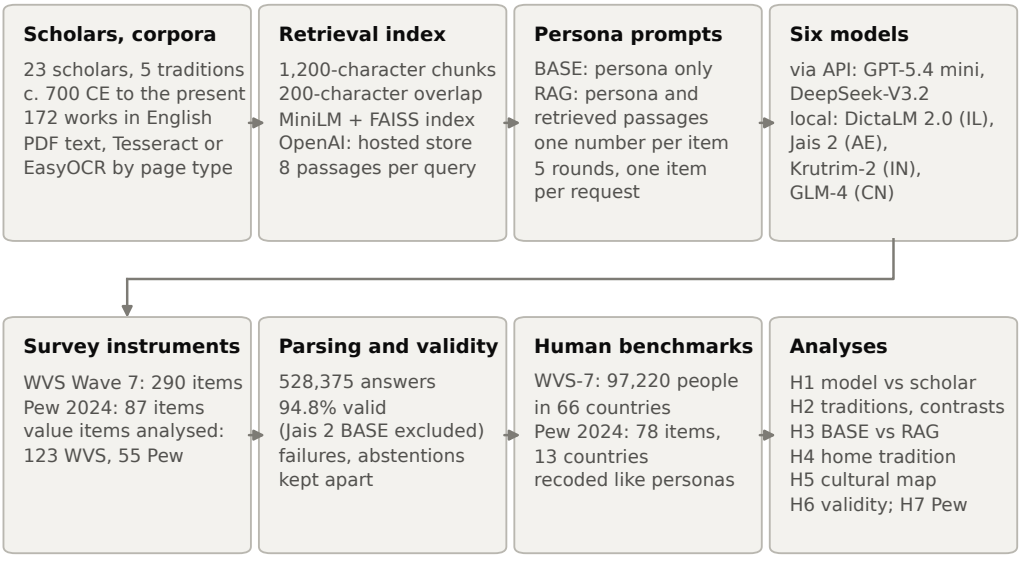
Two secondary hypotheses address the measurement itself (RQ6). H6 predicts that response validity, measured by failures, preferences for particular answer positions and invariance across rounds, varies by model and condition and is worst for the smaller open-weight models. It includes a specific test of DictaLM's tendency, noticed during data collection, to answer "1" to most items. If the tendency reflects a traditional stance, it should depend on whether "1" is the traditional or the secular end of an item's scale; if it reflects a position preference, it should not. H7 concerns the Pew items, a second instrument built independently of the WVS but answered by the same models and personas. It predicts that the variance comparison of H1, the tradition separation of H2a, the retrieval effects on dispersion and separation of H3 and the home-population comparison of H4 hold on that instrument. The known-group contrasts are defined on WVS items only and are not part of H7.

5. Data

The design is summarized in Figure 1. Corpora were assembled for each scholar, converted to clean text and indexed for retrieval. Each of six models then answered two surveys in each scholar's persona, with and without retrieved passages, five times over. The answers were parsed, validated and compared with human benchmarks. This section describes each component; Section 6 describes how the answers were measured and compared.

5.1 Scholars

The panel contains 23 scholars from 5 traditions (Table 2, Figure 2). The twelve Jewish, Christian and Muslim scholars are those of the earlier study (Caliskan et al. 2025): Maimonides, Spinoza, Buber and Arendt; Aquinas, Luther, Wesley and Ruether; al-Ghazālī, Ibn Taymiyya, Shariati and Mernissi. Six Buddhist scholars (Dōgen, Hakuin, Tsongkhapa, Ledi Sayadaw, Yin Shun and the 14th Dalai

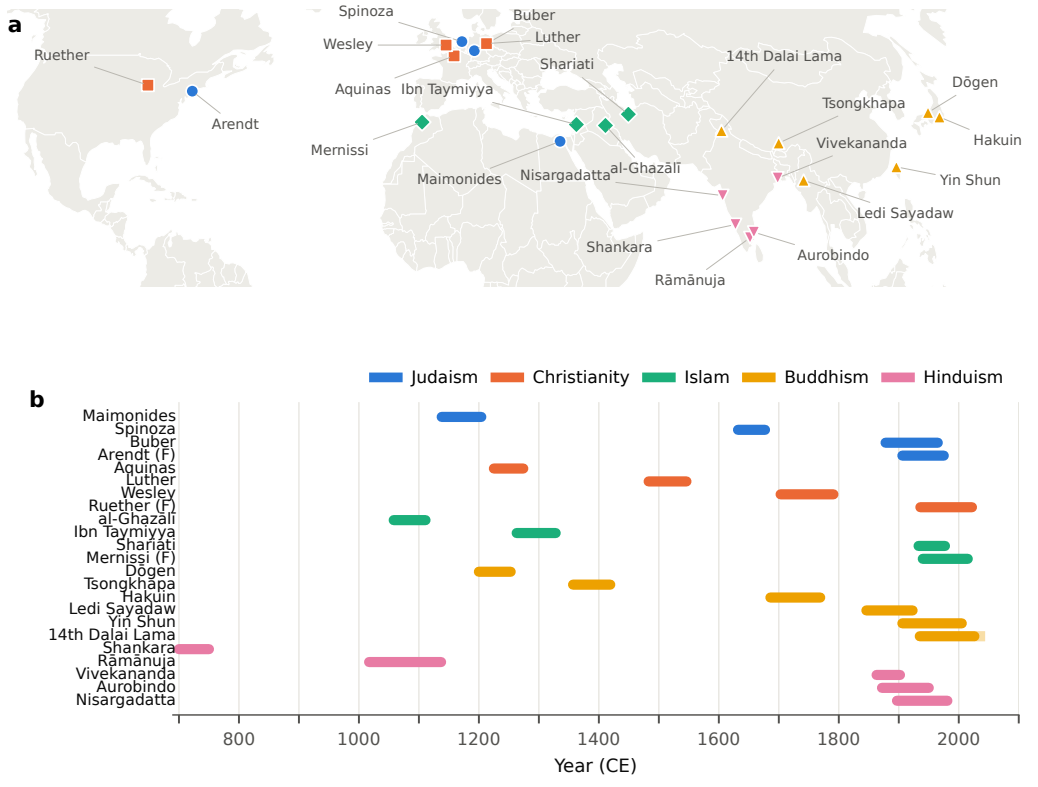


Note: Counts are responses after parsing. BASE prompts contain the persona only; RAG prompts add passages retrieved from the scholar's corpus.

Figure 1. Study design

Lama) and five Hindu scholars (Shankara, Rāmānuja, Vivekananda, Aurobindo and Nisargadatta) were added for this study. Their lives run from Shankara, born around 700, to Mernissi, born in 1940, and the 14th Dalai Lama is alive. We use "scholar" broadly. The panel consists of religious and intellectual figures associated with the five traditions: systematic theologians and jurists, mystics and monastic teachers, reformers, and modern intellectuals. Some of the last, such as Arendt and Mernissi, owe their standing to scholarship rather than to religious office, and their relation to their tradition is in part critical or cultural. It covers the Middle East and North Africa, Europe, North America and South and East Asia, although several scholars were active far from the modern state whose population we use as their reference (Figure 2a). Three of the 23 are women, all in the Abrahamic panels. Our candidate-generation and corpus-availability criteria, described next, produced no women in the Buddhist and Hindu panels; the one woman on the candidate lists, the Tibetan teacher Ani Lochen, did not pass the availability screen. Appendix 1 gives a short profile of each scholar.

The Buddhist and Hindu scholars were chosen by a procedure designed to limit the influence of our own familiarity with the European canon. We gave three LLMs (ChatGPT-5.2, Gemini Pro and DeepSeek) the same prompt asking for the ten scholars in each tradition whose preserved works were most extensive and influential, spanning centuries, regions, schools and, where possible, genders (Appendix 4). Each model received the prompt ten times, and we kept the names that appeared in all ten runs. Because the three models' stable lists overlapped on only about three names, we took their union for each tradition, which gave 11 Buddhist and 16 Hindu candidates. Expert consultation yielded four suggestions for each tradition, of which three Buddhist names and one Hindu name were new to the candidate pools. We then counted, for every candidate, the works available in English in accessible digital form. Candidates with fewer than about ten such works were removed where possible, which halved the Buddhist pool from 14 to 7, and the remaining candidates were ranked by availability, with additional weight given to the expert recommendations. The top six Buddhist and five Hindu scholars formed the final panel. The procedure favours scholars with large



Note: (a) Main place of activity. (b) Lifespans; (F) marks the three women and the faded segment the living scholar.

Figure 2. Scholars in space and time

translated corpora and is one reason for the panel's gender imbalance.

5.2 Corpora

Each scholar's retrieval corpus is a folder of English texts and translations. The project inventory lists 172 titles, from 4 for Buber to 19 for Luther (median 7.0; Appendix 2). A title in the inventory is a source volume, not a distinct work: collected and complete editions sit beside single works that they may contain, and in a few cases the same work appears under two titles or in two translations. The Abrahamic corpora were built for the earlier study. For the Buddhist and Hindu scholars, we collected scanned and born-digital PDFs and converted them to text with a hybrid procedure. We first compared native text extraction with PyMuPDF (Artifex Software, Inc. 2026), Tesseract (Smith 2007) and EasyOCR (JaidedAI 2024) on four test pages chosen to span the conditions in our collection. They were two difficult scans (from *Instruction for Chinese Women and Girls* and *Shōbōgenzō Zuimonki*), a clean born-digital page from *Karmayogin* and a page mixing English and Sanskrit from *Gītā Bhāṣya*. Native extraction was the most accurate on born-digital files. Tesseract was faster and stable on clean, text-dense scans, and EasyOCR produced fewer garbage characters on degraded and decorative pages. The final pipeline therefore attempts native extraction first, accepts it when the extracted text passes basic quality checks, and otherwise applies image normalization and contrast enhancement before sending clean pages to Tesseract and degraded ones to EasyOCR. Unicode normalization and removal of running headers, page numbers and other layout residue follow.

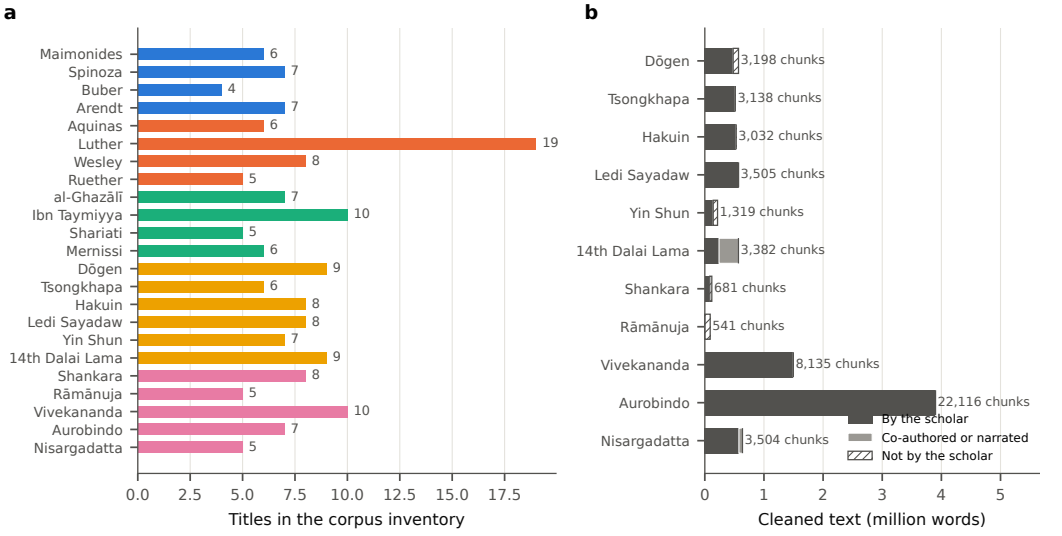
The cleaned Buddhist and Hindu corpora comprise 96 text files and 9.2 million words, split into

Table 2. The 23 scholars

Scholar	Dates	Place of activity	School or lineage	Sex	Works	WVS reference state(s)
Judaism						
Maimonides	1138–1204	Fustat	Rabbinic rationalism	M	6	Egypt
Spinoza	1632–1677	Amsterdam	Rationalist philosophy	M	7	Netherlands
Buber	1878–1965	Frankfurt	Dialogical philosophy, Hasidism	M	4	Germany
Arendt	1906–1975	New York	Political philosophy	F	7	Germany, United States
Christianity						
Aquinas	1225–1274	Paris	Scholasticism (Dominican)	M	6	–
Luther	1483–1546	Wittenberg	Lutheran Reformation	M	19	Germany
Wesley	1703–1791	London	Methodism	M	8	Great Britain
Ruether	1936–2022	Evanston	Feminist theology (Catholic)	F	5	United States
Islam						
al-Ghazālī	1058–1111	Baghdad	Ash'ari theology, Sufism	M	7	Iran, Iraq
Ibn Taymiyya	1263–1328	Damascus	Hanbali jurisprudence	M	10	Jordan, Lebanon
Shariati	1933–1977	Tehran	Shi'i revolutionary thought	M	5	Iran
Mernissi	1940–2015	Rabat	Islamic feminism, sociology	F	6	Morocco
Buddhism						
Dōgen	1200–1253	Eiheiji	Soto Zen	M	9	Japan
Hakuin	1686–1769	Hara	Rinzai Zen	M	8	Japan
Tsongkhapa	1357–1419	Lhasa	Gelug (Tibetan)	M	6	China
Ledi Sayadaw	1846–1923	Monywa	Theravada (Burmese)	M	8	Myanmar
Yin Shun	1906–2005	Hsinchu	Humanistic Buddhism	M	7	China, Taiwan
Dalai Lama	b. 1935	Dharamsala	Gelug (Tibetan)	M	9	China, India
Hinduism						
Shankara	c. 700–750	Sringeri	Advaita Vedanta	M	8	India
Rāmānuja	c. 1017–1137	Srirangam	Vishishtadvaita Vedanta	M	5	India
Vivekananda	1863–1902	Kolkata	Neo-Vedanta, Ramakrishna Mission	M	10	India
Aurobindo	1872–1950	Puducherry	Integral yoga	M	7	India
Nisargadatta	1897–1981	Mumbai	Advaita (Navnath)	M	5	India

Note: Works are the number of texts in each scholar's retrieval corpus (Appendix 2). Italy and France, the natural reference states for Aquinas, are not in WVS Wave 7.

52,551 retrieval chunks of 1,200 characters with a 200-character overlap. Their size varies by a factor of about forty, from 91.0 thousand words for Rāmānuja to 3.9 million for Aurobindo (Figure 3). The variation reflects what has been translated and digitized, and it means that retrieval draws on a far richer pool for some scholars than for others. We read the opening pages of every cleaned file and classified its authorship (Appendix 2). Most of the text, 86%, consists of the scholars' own works in translation, often with a translator's introduction; another 6% consists of Nisargadatta's talks as recorded and edited by visitors. The rest is of three kinds. Three books of the 14th Dalai Lama were written with or narrated by another author (*The Art of Happiness*, *The Book of Joy* and *Destructive Emotions*), and together they make up 58% of his corpus. One book in Nisargadatta's corpus is a disciple's account of conversations with him. In Shankara's *Atma Bodha* the translator's introduction precedes the text and fills 57% of the file. Finally, three files are not by the scholar at all. Rāmānuja's folder holds a modern monograph on his epistemology and a magazine issue on another figure, and none of his own commentaries. Yin Shun's holds a book of talks by another teacher introducing his thought (Bhikṣu Changtzu 2017), 34% of the corpus, and Dōgen's holds a later teacher's exposition of Sōtō Zen that quotes him. Of the inventory titles, two in the Abrahamic corpora are also not the scholar's own: *The Noble Lesson*, a medieval Waldensian poem, listed under Luther, and *Political Thought of Ibn Taymiyyah*, a secondary study (Khan 1973). The Abrahamic text



Note: (a) Titles in each scholar's corpus inventory. (b) Cleaned text and retrieval chunks for the Buddhist and Hindu corpora built for this study, by authorship of the text.

Figure 3. Corpus size by scholar

files were not available for the audit, so we could not check what they contain. Rather than rebuild the indexes, we report the corpora as they were used and repeat the main retrieval analyses without the affected scholars (Section 6).

In every pipeline except OpenAI's, which used the provider's hosted file search, the retrieval index embeds each chunk with all-MiniLM-L6-v2, which reads at most 256 word pieces (Appendix 5). A chunk of 1,200 characters averages 277 word pieces in these corpora, so 81% of chunks are cut short when embedded and about 8% of the text is not represented in the embeddings. The truncation affects which chunks are retrieved, not what the model reads: the full chunk is passed to the model.

5.3 Models

Table 3 describes the six models. GPT-5.4 mini (OpenAI 2026b, 2026a) and DeepSeek-V3.2, accessed through the `deepseek-chat` endpoint (DeepSeek-AI 2025), are frontier systems from the United States and China. Both were queried through their providers' APIs at the default sampling settings; DeepSeek-V3.2's weights are also public, but it was not run locally. The four open-weight models were chosen for their provenance. DictaLM 2.0 is a 7-billion-parameter adaptation of Mistral-7B to Hebrew from Dicta, the Israel Center for Text Analysis (Shmidman et al. 2024). Jais 2 8B Chat was trained from scratch on English and Arabic, including dialectal Arabic, by Inception, MBZUAI and Cerebras in the United Arab Emirates (Anwar et al. 2026; Inception and MBZUAI and Cerebras 2025). Krutrim-2, a 12-billion-parameter model from Ola Krutrim in India, is built on Mistral-NeMo and covers English and 22 Indian languages (Kallappa et al. 2025). GLM-4-9B-0414, from Zhipu AI in China, belongs to the GLM-4 family trained mainly on Chinese and English (Team GLM et al. 2024; Z.ai (Zhipu AI) 2025). The open-weight models ran locally through llama.cpp at temperature zero, in 4-bit quantized versions except for Krutrim-2, which ran in a 3-bit version. The quantization keeps local inference feasible but may cost the smaller models some accuracy. The two groups of models therefore differ in size, sampling temperature, prompt wording and retrieval implementation as well as in provenance, so what we call a model effect is the effect of a model and its pipeline together. We return to this point in the discussion.

Table 3. The six models

Model	Developer	Country	Release	Size	Quant.	Context	Access	Temp.	Home
GPT-5.4 mini	OpenAI	USA	Mar 2026	n.d.	–	400K	API	default	(Christianity)
DeepSeek-V3.2	DeepSeek	China	Dec 2025	685B MoE	–	128K	API	default	–
DictaLM 2.0	Dicta	Israel	Jul 2024	7B	Q4_K_M	n.s.	open weights	0	Judaism
GLM-4-9B-0414	Zhipu AI (Z.ai)	China	Apr 2025	9B	Q4_K_M	32K	open weights	0	Buddhism
Jais 2 8B Chat	Inception et al.	UAE	Dec 2025	8B	Q4_K_M	8K	open weights	0	Islam
Krutrim-2	Ola Krutrim	India	Feb 2025	12B	IQ3_M	128K	open weights	0	Hinduism

Note: Size is the parameter count reported by the developer (n.d., not disclosed). Context is the native window in tokens (n.s., not stated in the model card). Temp. is the sampling temperature in our runs; "default" means the provider default. The four open-weight models were run locally with llama.cpp; the quantization is that of the GGUF file used. Jais 2 was developed by Inception, MBZUAI and Cerebras. DeepSeek-V3.2 was accessed through the deepseek-chat endpoint. The Christian pairing for GPT-5.4 mini is exploratory.

5.4 Instruments

The first instrument is the full questionnaire of WVS Wave 7 (Haerpfer et al. 2024), 290 items covering social values, trust, religion, ethics, politics and demographics (Table 4; Appendix 3). The WVS has been the standard instrument for cross-national comparison of values since 1981, it was the instrument of the earlier study, and its microdata allow comparison with individual respondents in many countries. Its main limitation for our purpose is timing. Wave 7 was fielded between 2017 and 2022, before the models in our panel were trained.

The second instrument is a questionnaire of 87 items that we selected from the Pew Research Center's Spring 2024 Global Attitudes Survey (Pew Research Center 2025), which is closer in time to the models' training data and stronger on religion in public life and relations between religious groups. Three lists of countries play a part in this instrument. The selection frame was a list of 22 countries in the Middle East and in South and East Asia, where the non-Abrahamic traditions and the regional models are concentrated. They are Egypt, Israel, Jordan, Kuwait, Lebanon, the Palestinian territories, Tunisia and Turkey, and Bangladesh, China, India, Indonesia, Japan, Malaysia, Pakistan, the Philippines, Singapore, South Korea, Sri Lanka, Taiwan, Thailand and Vietnam. Only 13 of them were part of the Spring 2024 fieldwork: Bangladesh, India, Indonesia, Israel, Japan, Malaysia, the Philippines, Singapore, South Korea, Sri Lanka, Thailand, Tunisia and Turkey. These 13 countries are the human benchmark, and each of the 78 items with published toplines was asked in at least six of them. Within that frame we excluded items tied to particular political figures or recent events, such as the war in Ukraine. We kept items that elicit judgments about what is good, right or desirable, among them preferred systems of government, the role of religion in law and public life, gender equality, national identity, diversity, free expression and religious practice. The third list belongs to four items that ask about a respondent's own country, nationality or religion. For these, the persona was asked once for each country on a list associated with its tradition, 22 country versions in all, which enter only the validity diagnostics. We call this questionnaire Pew Set B. An earlier list of Pew-style items, which we call Set A, was used in the BASE runs of the two commercial models before Set B was finalized; GPT-5.4 mini answered 87 of its items and DeepSeek-V3.2 90. The two sets share almost no items. Set A therefore enters only the validity diagnostics, and comparisons on Pew between BASE and RAG are possible only for the regional models.

5.5 Collection and coverage

Each prompt presented the persona, one survey question and its answer codes, and asked for a single number. The persona opened with the sentence "I am a [tradition] scholar named [name]", followed by an instruction to answer from the scholar's works with a single number; the wording of that instruction varied slightly between model pipelines but not between conditions. In the RAG condition the prompt was preceded by passages retrieved from the scholar's corpus (Appendix 4

Table 4. Instruments and items

	WVS Wave 7	Pew Set B
Items asked	290	87
Values and norms	129	55
Perceptions of own country	95	21
Behaviour	29	5
Identity and belonging	6	4
Demographics	31	0
Repeated in another format	0	2
Value items analysed	123	55
Human benchmark	66 countries	13 countries
	97,220 respondents	78 items aligned

Note: Value items analysed excludes the six WVS postmaterialism items (nominal). Pew Set A, asked only of the commercial models without retrieval, had 87 items for GPT-5.4 mini and 90 for DeepSeek-V3.2 and enters only the validity diagnostics. The four identity items of Set B were expanded into country-specific versions.

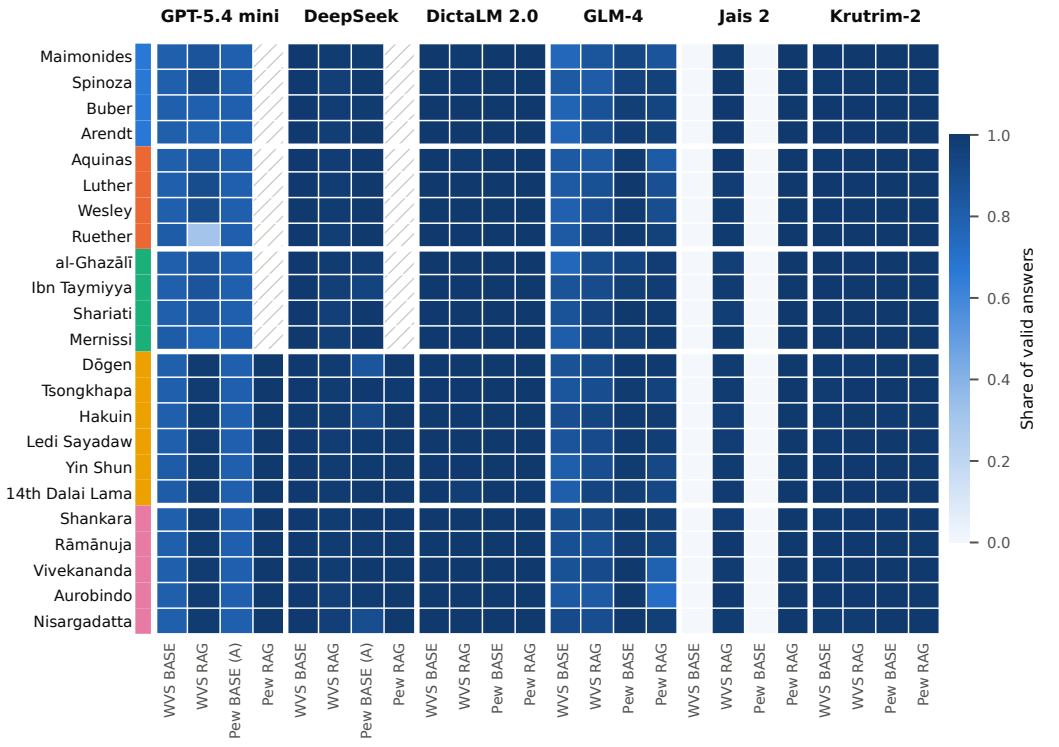
and Appendix 5). Every item was asked in five rounds, and every item was sent as a separate request with no conversation history, so an earlier question could not influence the answer to a later one. In the RAG runs and in all runs of the regional models, items were sent in a new random order in each round; the BASE runs of the two commercial models kept the same order in every round. Because requests were independent, the order matters only through possible caching or drift on the provider's side. Answer options were always shown in the questionnaire's order, never reversed.

The full design has 552 cells, one for each combination of six models, two conditions, two instruments and 23 scholars. We collected 528 of them in 528 files holding 528,375 responses (Figure 4). The missing cells are the Pew RAG runs of the two commercial models for the twelve Abrahamic scholars, which were not run because of their API cost. Jais 2 returned no usable answer in either BASE survey: all 45,295 of its BASE responses were empty, runs of dots or skipped prompts. With retrieved context the same model answered normally, so we report the BASE failure in Appendix 6 and use Jais 2 in the RAG analyses only. Across the remaining cells, 94.8% of non-demographic responses were valid answers. The median cell had 99.2% valid answers; the 53 cells under 80% belong to GLM-4, which often answered with decimals or lists of options, and to GPT-5.4 mini, which often declined to answer without retrieved context (Figure 5).

5.6 Human benchmarks

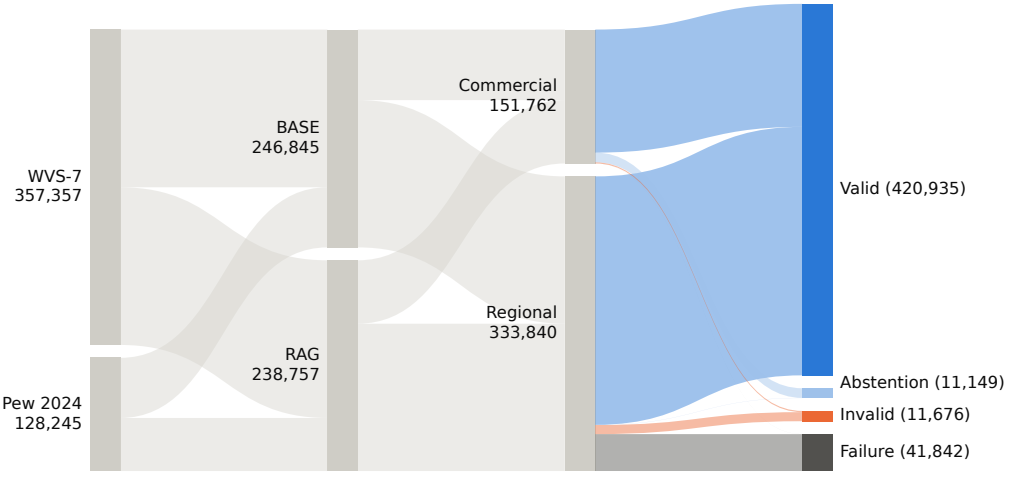
For the WVS we use the final cross-national release of Wave 7 (Haerpfer et al. 2024), with 97,220 respondents in 66 countries and territories. Country distributions and means are weighted with the WVS design weights. Each scholar is assigned one or two modern reference states, chosen to match the place where the scholar lived or the population that most directly inherits the scholar's tradition (Table 2). Several assignments are compromises. Tibet is not surveyed, so Tsongkhapa's reference is China; Syria is absent from Wave 7, so Ibn Taymiyya's references are Jordan and Lebanon; and Italy and France, where Aquinas taught, are not in Wave 7, so he has no reference state. No Jewish scholar has Israel as a WVS reference state, because Israel is not in Wave 7. For Pew Set B we use the country toplines published by Pew, which we matched to our items by question wording and aligned option by option. Toplines were available for 78 of the 87 items and for 13 countries, including Israel, India, Japan, Sri Lanka and Thailand, the Pew reference states for the Jewish, Hindu and several Buddhist scholars.

6. Measurement and methods



Note: Shading gives the share of valid answers among non-demographic items. Hatched cells were not collected. "A" marks cells that used Pew Set A.

Figure 4. Coverage and validity by cell



Note: Flows from instrument and condition to model and to the status assigned by the parser. Demographic items are excluded.

Figure 5. Where the responses go

6.1 Transforming raw answers to scores

All raw answers were re-parsed with a single set of rules, applied in a fixed order and identical for every model. Chat-template tokens and surrounding whitespace were removed first. Empty output, runs of dots and skipped prompts were coded as generation failures. Text expressing ignorance or refusal, and numeric codes for "don't know" or "not applicable" where the item offered them, were coded as abstentions. A leading number followed by the end of the answer, a line break or a sentence of explanation was accepted, which recovers answers such as "3. Never. As a Muslim scholar, I believe…". Answers that listed several options, gave a range, repeated strings of digits or contained more than one candidate number were coded as invalid, as were decimals on ordinal and categorical items and numbers outside the item's codes. Text such as a country name was invalid unless the item asked for it. Each response thus has one of four statuses: valid, abstention, failure or invalid. The rules recover answers the original collection scripts had dropped, notably ten-point answers in the DeepSeek BASE files, and they discard answers the scripts had accepted, such as "331" read from a string of digits. Among the regional models, re-parsing changed 0.8% of DictaLM's and 1.8% of GLM-4's previously recorded values; most of the change for GLM-4 is in answers now coded invalid (Appendix 6).

We index models by $m \in \mathcal{M}$, conditions by $c \in \{\text{BASE}, \text{RAG}\}$, scholars by $s \in \mathcal{S}$ with $|\mathcal{S}| = 23$, the tradition of scholar s by $\tau(s)$, items by i and rounds by $r = 1, \ldots, 5$. Let y_{mcsir} be a valid answer code and V_{mcsi} the set of rounds with a valid answer. The primary unit of analysis is the modal answer

$$\tilde{y}_{mcsi} = \arg\max_{k} \sum_{r \in V_{mcsi}} \mathbf{1}[y_{mcsir} = k], \tag{1}$$

defined when $|V_{mcsi}| \geq 1$; cells without a valid round are missing. When several codes share the highest count, the code closest to the median of the valid answers is taken, and if two codes are equally close, the lower one. Ties of any kind occur in 1.7% of the 80,314 persona-item cells, and 0.4% are settled by the final rule; they are most frequent for DeepSeek (7.2%), which was sampled at the provider's default temperature. Round-level answers are used only to measure stability. Each modal answer is rescaled to the unit interval over the item's codes $[a_i, b_i]$,

$$x_{mcsi} = \frac{\tilde{y}_{mcsi} - a_i}{b_i - a_i} \in [0, 1], \tag{2}$$

so that a four-point and a ten-point item contribute on the same scale; binary items coded 0 and 1 by the prompt were mapped to the WVS codes before rescaling. Raw responses are never imputed; sparse missing cells are filled only inside the variance decomposition described below, and every result that depends on the fill-in is also reported on complete-case items.

We classified every item as a value or norm, a perception of the respondent's own country, a behaviour, an identity or a demographic fact (Appendix 3). The hypothesis tests use the 123 WVS and 55 Pew value items, which form the item set $\mathcal{I}$ of each instrument. The six WVS postmaterialism items, which ask for a first and second choice among unordered goals, are nominal and are excluded, as are two Pew items that repeat another item in a different format. Questions about the state of a country ("How common is corruption here?") or a respondent's own conduct ("Did you vote?") have no natural answer for a persona and are excluded; demographic items serve only as a consistency check.

A persona's profile is the vector $\mathbf{x}_{mcs} = (x_{mcsi})_{i \in \mathcal{I}}$. The distance between two profiles p and q is the mean absolute difference over the items both have answered,

$$d(p, q) = \frac{1}{|\mathcal{I}_{pq}|} \sum_{i \in \mathcal{I}_{pq}} |x_{pi} - x_{qi}|, \qquad |\mathcal{I}_{pq}| \geq 10, \tag{3}$$

Table 5. Units, tests and corrections

Question	Unit	Test	Resampling, n	Correction
RQ1 (H1)	item	mean η^2 shares; share of draws with $\bar{\eta}^2_M > \bar{\eta}^2_T + \bar{\eta}^2_S$	items (2,000); scholars (300)	none; five panels reported
RQ1 (H1b)	scholar	one-sided Wilcoxon signed rank	$n = 23$ scholars	none
RQ2 (H2a)	model × condition	permutation of tradition labels	5,000 relabelings	Holm within instrument and condition
RQ2 (H2b)	contrast; contrast family	exact one-sided Wilcoxon on model means	$n = 7$ contrasts; $n = 5$ families	none
RQ3 (H3)	model; item; contrast	bootstrap of dispersion change; Wilcoxon on matched items; exact Wilcoxon on paired contrast change	items (1,000); $n = 5$ models; $n = 7$ contrasts	none
RQ4 (H4)	model × scholar	exact permutation of home assignments	24 (regional) or 120 assignments	both tails reported
RQ5 (H5)	model × scholar; scholar; model	Wilcoxon on cells, scholar means and model means; crossed bootstrap	scholars and models (4,000)	none
RQ6 (H6, H7)	response; item set	descriptive shares; code-1 comparison	–	–

where $\mathcal{I}_{pq}$ is the set of shared items. In practice the overlap is large: within a model and condition, two WVS personas share at least 84 value items and two Pew personas at least 35. Because the shared set still differs between pairs, we recomputed every separation ratio on the items answered by all personas of the model (between 64 and 123 WVS items). The two sets of distances correlate at $r \geq 0.85$, and no separation ratio changes by more than 0.08.

Two summary indices locate personas and countries on the cultural map. We use Welzel's (2013) secular values and emancipative values indices, the index-based versions of the traditional–secular and survival–self-expression dimensions of Inglehart and Welzel (2005). Secular values average four components (defiance of authority, disbelief, relativism and scepticism), each built from three WVS items; emancipative values average autonomy, equality, choice and voice. We computed the indices from their component items with the published recoding rules and checked that our recodes reproduce the component variables distributed with the WVS file. Persona answers were recoded with the same rules, and a persona's index is computed when at least one item of every component is available. Country scores are means over respondents weighted with the WVS design weight. We write $\mathbf{w}_{mcs} = (\mathrm{SEC}_{mcs}, \mathrm{EMA}_{mcs})$ for a persona and $\mathbf{w}_j$ for country j.

6.2 Estimation and tests by research question

Table 5 summarizes, for each research question, the unit of analysis, the test, the resampling unit, the number of independent units and the family within which p-values are corrected. Because the same scholars recur across models and the same models across scholars, no test treats model-by-scholar cells as independent observations unless a dependence-aware version is reported beside it. All analyses were run in Python with `numpy`, `pandas` and `scipy`; random seeds are fixed in the code.

For RQ1, consider a panel of models $\mathcal{M}$ in one condition. Each item i gives a model-by-scholar table of modal answers x_{msi} with grand mean $\bar{x}_{\cdot\cdot i}$, model means $\bar{x}_{m\cdot i}$, scholar means $\bar{x}_{\cdot si}$ and tradition means $\bar{x}_{\cdot ti}$ over the n_t scholars of tradition t. The sums of squares are

$$\begin{aligned} SS^{\mathrm{tot}}_i &= \textstyle\sum_m \sum_s (x_{msi} - \bar{x}_{\cdot\cdot i})^2, \qquad SS^M_i = |\mathcal{S}| \sum_m (\bar{x}_{m\cdot i} - \bar{x}_{\cdot\cdot i})^2, \\ SS^T_i &= |\mathcal{M}| \textstyle\sum_t n_t (\bar{x}_{\cdot ti} - \bar{x}_{\cdot\cdot i})^2, \qquad SS^S_i = |\mathcal{M}| \textstyle\sum_s (\bar{x}_{\cdot si} - \bar{x}_{\cdot\cdot i})^2 - SS^T_i, \end{aligned} \tag{4}$$

and the shares $\eta^2_{k,i} = SS^k_i / SS^{\mathrm{tot}}_i$ for $k \in \{M, T, S\}$ are averaged over items, $\bar{\eta}^2_k = |\mathcal{I}|^{-1} \sum_i \eta^2_{k,i}$. The remainder, $1 - \bar{\eta}^2_M - \bar{\eta}^2_T - \bar{\eta}^2_S$, is the model-by-scholar interaction together with noise in the modal

answers, since each cell holds one modal answer. H1 compares $\bar{\eta}^2_M$ with the combined scholar share $\bar{\eta}^2_T + \bar{\eta}^2_S$; it is a comparison of main effects and makes no claim about the remainder. Uncertainty comes from 2,000 bootstrap resamples of items and, for the main WVS panel, 300 resamples of scholars, and we report the share of resamples in which the model share exceeds the scholar share. Where an item's table has missing cells, they are filled with that model's mean on the item if no more than 15% of the table is missing, and the item is dropped otherwise. Because filling with model means can inflate $\bar{\eta}^2_M$, every comparison is repeated on complete-case items. We estimate the decomposition for five panels: all six models with retrieval, the five models with usable BASE data, the four regional models with retrieval on each instrument, and the three regional models with BASE data on Pew. Because the design is fully crossed and nearly balanced at the level of the modal answer, we use this direct decomposition rather than a mixed-effects model.

For H1b we define two distances for each scholar. D_s^{cross} is the mean distance between two models' portraits of the scholar, and D_s^{within} the mean distance between one model's portrait of the scholar and the same model's portraits of the other scholars:

$$D_s^{\text{cross}} = \binom{|\mathcal{M}|}{2}^{-1} \sum_{m<m'} d(\mathbf{x}_{ms}, \mathbf{x}_{m's}), \qquad D_s^{\text{within}} = \frac{1}{|\mathcal{M}|(|\mathcal{S}|-1)} \sum_{m} \sum_{s' \neq s} d(\mathbf{x}_{ms}, \mathbf{x}_{ms'}). \tag{5}$$

A one-sided Wilcoxon signed-rank test is applied to $D_s^{\text{cross}} - D_s^{\text{within}}$ across the 23 scholars, so that the scholar is the unit of inference.

Tradition separation (RQ2, H2a) is measured within each model and condition. Let $\bar{d}^{\neq}_{mc}$ and $\bar{d}^{=}_{mc}$ be the mean distances over pairs of personas of model m in condition c from different and from the same tradition. The separation ratio is

$$R_{mc} = \bar{d}^{\neq}_{mc} / \bar{d}^{=}_{mc}, \tag{6}$$

with $R > 1$ indicating clustering by tradition. Its one-sided p-value is $\left(1 + \#\{b : R^{(b)}_{mc} \geq R_{mc}\}\right)/(B+1)$, where $R^{(b)}$ is the ratio after the bth of $B = 5{,}000$ random permutations of the tradition labels among scholars. Holm's correction is applied within each instrument and condition, a family of three to six models.

For H2b each contrast k has an expected-higher scholar h_k, an expected-lower scholar l_k and a domain of items J_k (Appendix 7). The three gender contrasts use five gender-equality items. The contrast between Spinoza and Maimonides uses four items on the importance of God and religion, religious exclusivism and the authority of religion over science. The Vedanta contrasts use three items on the value of work and social duty, and the Tibetan contrast five items on science and religion. With x^*_{msj} equal to x_{msj} or $1 - x_{msj}$ according to the item's direction, the domain score and the contrast difference are

$$G_{msk} = \frac{1}{|J_k|} \sum_{j \in J_k} x^*_{msj}, \qquad \Delta_{mk} = G_{mh_k k} - G_{ml_k k}. \tag{7}$$

A difference $\Delta_{mk} > 0$ is in the expected direction. We report the shares of the model-contrast pairs that are positive, negative and tied. Since the pairs from one contrast are not independent, the test averages over models, $\bar{\Delta}_k = |\mathcal{M}|^{-1} \sum_m \Delta_{mk}$, and applies an exact one-sided Wilcoxon signed-rank test to the seven contrast means. The contrasts also share scholars, since Ruether and Vivekananda each appear twice. We therefore repeat the test on five contrast families, averaging the two Ruether contrasts and the two Vivekananda contrasts; with five families the smallest attainable one-sided p is $1/32 \approx 0.031$.

For RQ3, the comparison of conditions uses the models that have both: five models on the WVS and the three regional models with BASE data on Pew. Dispersion is the mean pairwise distance among a model's personas,

$$V_{mc} = \binom{|\mathcal{S}|}{2}^{-1} \sum_{s<s'} d(\mathbf{x}_{mcs}, \mathbf{x}_{mcs'}), \qquad \Delta V_m = V_{m,\mathrm{RAG}} - V_{m,\mathrm{BASE}}, \tag{8}$$

and its 95% interval comes from 1,000 bootstrap resamples of items. The shift of a persona between conditions is $d(\mathbf{x}_{m,\mathrm{BASE},s}, \mathbf{x}_{m,\mathrm{RAG},s})$. We compare the separation ratios $R_{m,\mathrm{RAG}}$ and $R_{m,\mathrm{BASE}}$ model by model. The variance decomposition is estimated in matched panels that contain the same models in both conditions, both with fill-in and on complete-case items; on the latter the item-wise change in $\eta^2_{M,i}$ is tested with a one-sided Wilcoxon test. For the known-group contrasts the change is $\Delta^{\mathrm{RAG}}_{mk} - \Delta^{\mathrm{BASE}}_{mk}$. Averaged over the five models it gives one value per contrast, which we test with an exact one-sided Wilcoxon test over the seven contrasts and, as above, over the five families. The 35 model-contrast pairs are reported descriptively only.

For RQ4 we use the tradition margin. Within a model and condition, a persona's margin is its mean distance to personas of other traditions minus its mean distance to the other personas of its own tradition:

$$g_{ms} = \frac{1}{|\{s' : \tau(s') \neq \tau(s)\}|} \sum_{\tau(s') \neq \tau(s)} d(\mathbf{x}_{ms}, \mathbf{x}_{ms'}) - \frac{1}{|\{s' \neq s : \tau(s') = \tau(s)\}|} \sum_{s' \neq s,\ \tau(s') = \tau(s)} d(\mathbf{x}_{ms}, \mathbf{x}_{ms'}). \tag{9}$$

A large margin means the model portrays the scholar as clearly a member of his or her tradition. Let a home map π assign a tradition to each regional model. We estimate by least squares

$$g_{ms} = \alpha + \mu_m + \lambda_{\tau(s)} + \gamma\, \mathbf{1}[\tau(s) = \pi(m)] + \varepsilon_{ms}. \tag{10}$$

The estimate $\hat{\gamma}$ for the actual map π_0 is compared with its values $\hat{\gamma}_\pi$ under all bijections π of the four home traditions to the four regional models, so that the test asks whether the actual pairing stands out among the $4! = 24$ possible ones. The upper and lower exact p-values are $p_+ = \#\{\pi : \hat{\gamma}_\pi \geq \hat{\gamma}_{\pi_0}\}/24$ and $p_- = \#\{\pi : \hat{\gamma}_\pi \leq \hat{\gamma}_{\pi_0}\}/24$, and the two-sided value is $\min(1, 2\min(p_+, p_-))$, whose smallest attainable value is $2/24 \approx 0.083$. H4 predicts an advantage, so p_+ is the confirmatory test; a result in the other tail is reported as exploratory. With GPT-5.4 mini and Christianity added there are $5! = 120$ assignments. A second outcome measures closeness to the home population. For each persona the C WVS countries are ordered by $d(\mathbf{x}_{ms}, \mathbf{z}_j)$, where $\mathbf{z}_j$ is the vector of weighted country means on the same items, and the reference state j^* receives the percentile rank

$$\rho_{ms} = \frac{\mathrm{rank}(j^*) - 1}{C - 1}, \tag{11}$$

averaged over reference states when a scholar has two. A value of zero means the reference state is the closest country, and one half is what an arbitrary country would achieve. The same permutation test is applied to ρ, with $C = 66$ on the WVS and $C = 13$ on Pew.

For RQ5, let Σ be the covariance matrix of the country points $\mathbf{w}_j$, so that the Mahalanobis distance is $D(\mathbf{a}, \mathbf{b}) = \sqrt{(\mathbf{a} - \mathbf{b})^\top \Sigma^{-1} (\mathbf{a} - \mathbf{b})}$. Let $\bar{\mathbf{w}}$ be the mean of all countries and $\mathbf{w}_{\ell(s)}$ the mean of scholar s's reference states. For each persona we form

$$\delta^{GL}_{ms} = D(\mathbf{w}_{ms}, \bar{\mathbf{w}}) - D(\mathbf{w}_{ms}, \mathbf{w}_{\ell(s)}), \qquad e_{ms} = \mathrm{SEC}_{ms} - \mathrm{SEC}_{\ell(s)}, \tag{12}$$

so that $\delta^{GL} < 0$ means the persona is closer to the global than to the local point and $e > 0$ that it is more secular than its reference states; the emancipative difference is defined in the same way.

Personas enter if both indices could be computed for them and for at least one reference state, which excludes Aquinas (no reference state) and Maimonides (Egypt lacks the secular-values index). We test these differences in three ways. The first is a Wilcoxon signed-rank test over the model-by-scholar cells, which treats cells as independent. The second averages over models and tests the 21 scholar means; the third averages over scholars and applies a sign test to the model means. Finally, a crossed bootstrap resamples scholars and models independently and with replacement (4,000 draws) and gives a 95% interval for the grand mean that allows for both sources of dependence. The same four analyses are applied to the item-level comparison, which replaces the indices by the full profile, $d(\mathbf{x}_{ms}, \bar{\mathbf{z}}) - d(\mathbf{x}_{ms}, \mathbf{z}_{\ell(s)})$, and to the reference-state rank $\rho_{ms} - 0.5$. Because a persona is a single respondent rather than a population, we also place each persona's index scores among individual WVS respondents, as weighted percentiles of the respondent-level indices distributed with the WVS file, with which our country scores correlate at $r = 1.000$. The era gradient is the Pearson correlation, across scholars, between year of birth and the secular-values score averaged over models.

For RQ6 we tabulate the four statuses, the shares of valid answers at the first and the last option of each scale, and the share of persona-item cells answered identically in all five rounds. The test of DictaLM's preference for "1" needs items on which "1" is clearly the traditional or clearly the secular answer. We identify them from the country data: for each WVS item whose lowest code is 1 and which is observed in at least 30 countries, let κ_i be the correlation across countries between the item mean and the country's secular-values score. Code 1 is the traditional pole if $\kappa_i > 0.3$ (higher codes, more secular) and the secular pole if $\kappa_i < -0.3$. This yields 73 items on which "1" is the traditional answer and 26 on which it is the secular answer, and for each model and condition we compare

$$P_{\text{trad}} = \Pr(y = 1 \mid \kappa_i > 0.3), \qquad P_{\text{sec}} = \Pr(y = 1 \mid \kappa_i < -0.3) \tag{13}$$

with the same quantities for WVS respondents. A preference that tracks a traditional stance implies $P_{\text{trad}} \gg P_{\text{sec}}$; a preference for the first position implies that both are high. Because the options were never presented in reversed order, the comparison is indirect and cannot fully separate position from content. H7 repeats the H1, H2a, H3 and H4 analyses on Pew Set B.

The audit of the Buddhist and Hindu corpora (Section 5) found material not written by the scholar in some of them. We therefore repeat the main RAG analyses (H1, H2a, H3 dispersion, H4 margin and H5) twice. The first repetition drops the 3 scholars whose corpora contain a quarter or more of text not by the scholar (Shankara, Rāmānuja and Yin Shun). The second also drops the 14th Dalai Lama, 58% of whose corpus consists of co-authored or narrated books.

7. Results

Table 6 summarizes the tests; the subsections below give the evidence in the order of the hypotheses.

7.1 H1: model and scholar

Model identity explains more of the variation in persona answers than scholar identity in every panel we examined (Figure 6; Table 13 in Appendix 7). On the WVS value items with retrieval, the model accounts for 37% of the item-level variance (95% interval 33 to 40%), the tradition for 6.2% and the individual scholar within a tradition for 11.1%. Taken together, the two scholar components are about half the size of the model component. The remaining 46% is the largest single component. It combines the model-by-scholar interaction, the part of each portrait specific to one model's reading of one scholar, with noise in the modal answers. The comparison of main effects therefore does not imply that most of the variance is due to the model, or that the scholar carries no information. Without retrieval the imbalance is larger: across the five models with usable BASE data, the model accounts for 55% and the scholar for 14%. The pattern holds among the four regional models alone (38% against 21%) and on Pew Set B (38% against 21% with retrieval). In all five panels the model

Table 6. Summary of hypothesis tests

	Prediction	Verdict
H1	Model explains more variance than scholar	Supported
H1b	Same scholar differs across models	Supported
H2a	Personas cluster by tradition	Mostly supported
H2b	Known contrasts in expected direction	Partly supported
H3	Retrieval increases differentiation	Supported
	Retrieval lowers the model's share	Uncertain
	Retrieval improves known-group validity	Not supported
H4	Home-tradition advantage, regional models	Not supported
H5	Closer to global than local	Not sustained
	More secular than reference states	Not supported
H6	Validity varies by model and condition	Supported
H7	H1, H2a, H3 and H4 patterns hold on Pew	Supported

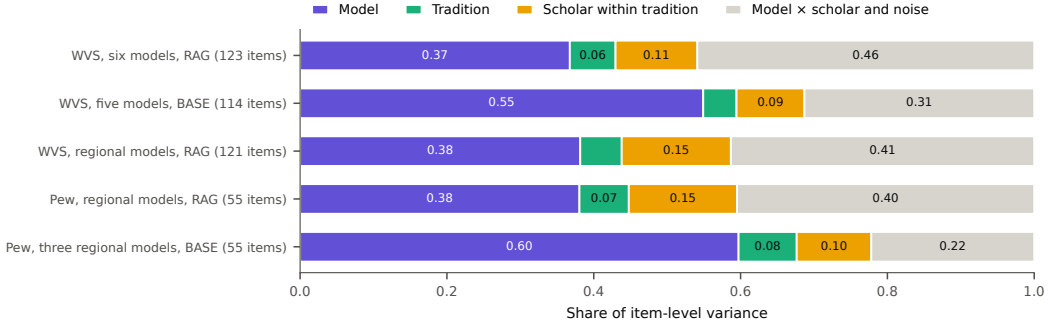
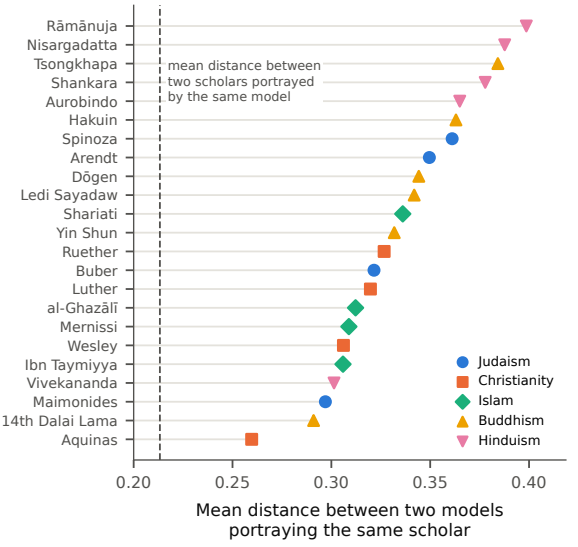


Figure 6. Sources of variance in persona answers

share exceeded the combined scholar share in every one of 2,000 bootstrap resamples of items, and on the WVS with retrieval also in every one of 300 resamples of scholars. The comparison survives the robustness checks in Appendix 8: mean rather than modal answers, dropping single scholars, splitting the panel into Abrahamic and non-Abrahamic traditions, using complete-case items only and dropping DictaLM, the model with the most extreme answers. The smallest model share in these variants is 0.28 and the largest scholar share 0.24.

The distances show the same imbalance at the level of individual portraits, in line with H1b. Two models portraying the same scholar differ on average by 0.33 on the unit scale, while one model portraying two different scholars differs by 0.21. The first distance exceeds the second for all 23 scholars (Wilcoxon $p < 0.001$). The two hypotheses are closely linked, since a large model effect implies different portraits of the same scholar, but the distances make the size of the problem concrete. GPT-5.4 mini's Aquinas is closer to GPT-5.4 mini's Ibn Taymiyya (distance 0.08) than to DeepSeek's Aquinas (0.17). Cross-model disagreement varies by scholar (Figure 7). It is lowest for Aquinas (0.26), Maimonides and the 14th Dalai Lama and highest for Rāmānuja (0.40), Nisargadatta, Tsongkhapa and Shankara. On average the models disagree more about the Buddhist and Hindu



Note: Mean absolute difference between the profiles of the same scholar from two models (WVS, RAG). The dashed line is the mean difference between two scholars portrayed by the same model.

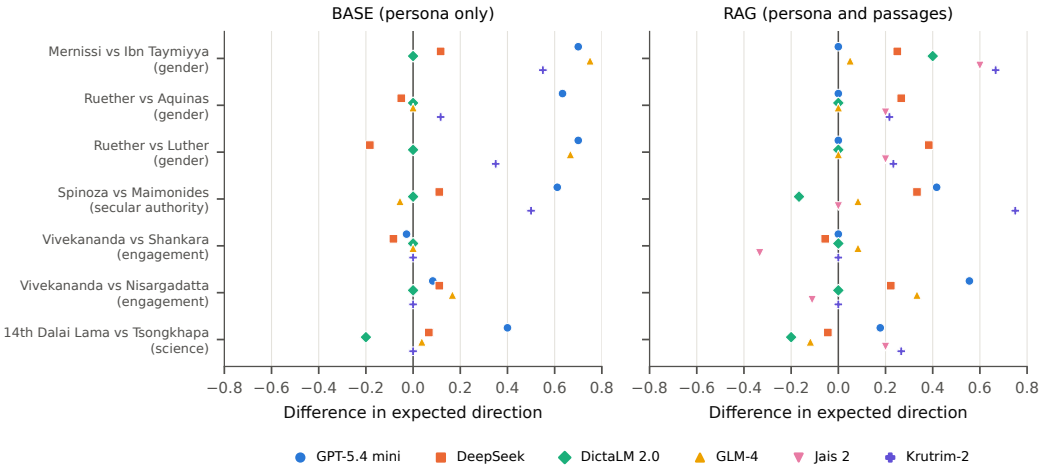
Figure 7. Disagreement between models by scholar

scholars than about the Jewish, Christian and Muslim ones (0.35 against 0.32; Mann–Whitney p = 0.011). Among the eleven Buddhist and Hindu scholars, disagreement does not track corpus size (Spearman ρ = −0.31, p = 0.35). Rāmānuja, the scholar on whom the models disagree most, is also the one whose corpus contains none of his own writings.

7.2 H2: traditions and known groups

Personas cluster by tradition in most models, as H2a predicts, but the clustering is modest. With retrieval, the ratio of between-tradition to within-tradition distance on the WVS ranges from 0.97 to 1.47, and it exceeds the permutation distribution after Holm correction for 5 of the six models (Table 14). GPT-5.4 mini separates the traditions most clearly (1.47), followed by DeepSeek (1.26); DictaLM does not separate them at all (0.97, Holm-adjusted p = 0.69). A ratio of 1.2 means that two scholars from different traditions are on average 20% further apart than two scholars from the same tradition, a real but not a sharp distinction, and one that stereotyped portraits of each tradition would also produce. All four regional models also separate the traditions on Pew Set B (ratios 1.15 to 1.24).

The known-group contrasts give partial support for H2b. Across six models and seven contrasts with retrieval, 52% of the differences point in the expected direction, 17% point the other way and 31% are ties (Figure 8). Averaged over models, 6 of the 7 contrasts are in the expected direction (exact one-sided Wilcoxon p = 0.023). When the two Ruether contrasts and the two Vivekananda contrasts are each merged, 5 of the 5 contrast families are (p = 0.031, the smallest attainable value). Without retrieval the counts are the same (6 of 7 contrasts and 5 of 5 families). The personas therefore carry some information about who disagrees with whom, but the information is uneven. The contrast between Mernissi and Ibn Taymiyya on gender equality is recovered by 83% of the models and that between Spinoza and Maimonides on religious authority by 67%. The contrasts between Ruether and Aquinas or Luther, between the Dalai Lama and Tsongkhapa and between Vivekananda and Nisargadatta are recovered by half of the models, and that between Vivekananda and Shankara on social engagement by 17%. The ties come mostly from two models with retrieval: DictaLM, which gives identical answers to most items for most scholars, and GPT-5.4 mini, each with 4 of 7 contrasts



Note: Difference in domain score between the expected-higher and the expected-lower scholar (WVS). Points to the right of zero are in the expected direction; points on the zero line are ties.

Figure 8. Known-group contrasts by model

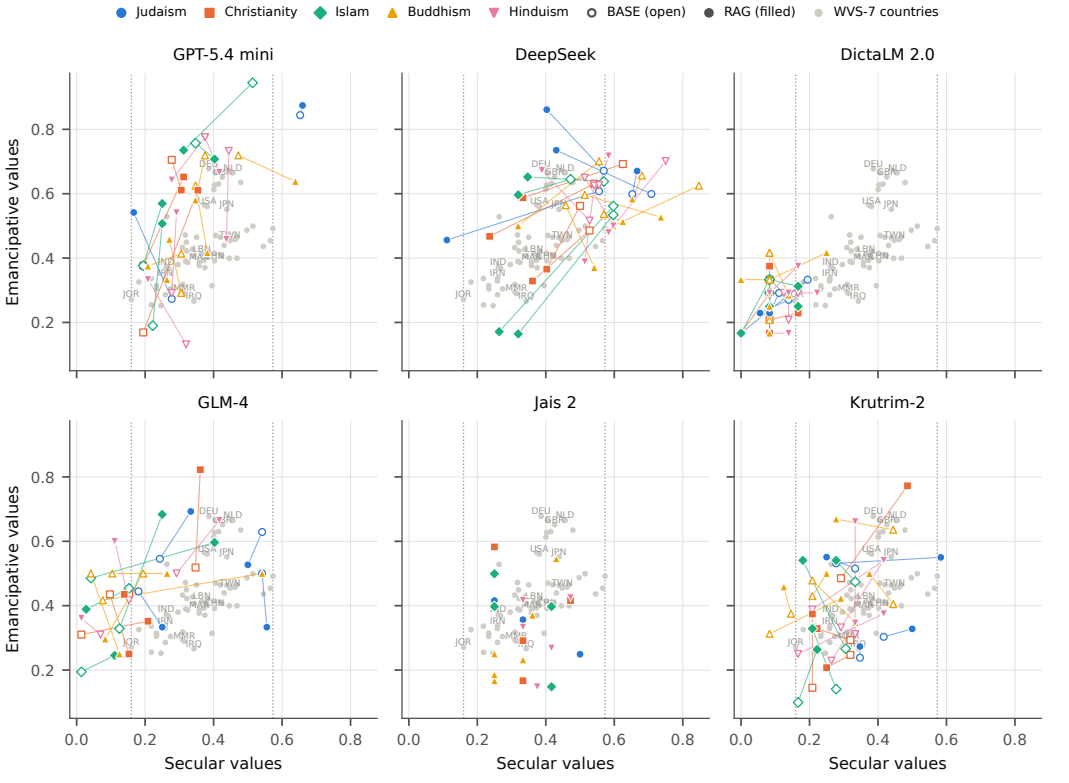
tied. GPT-5.4 mini's three gender contrasts, all clearly in the expected direction without retrieval, become exact ties with it (Table 15).

7.3 H3: what retrieval changes

Consistent with H3, adding retrieved passages increases the dispersion of a model's personas; this greater differentiation does not, however, carry over to the known-group contrasts. On the WVS, the mean distance among a model's personas rises with retrieval for 4 of the five models with both conditions, and the bootstrap interval excludes zero for all 4. The increase is largest for DeepSeek (from 0.18 to 0.32) and smallest for DictaLM (from 0.05 to 0.11). GPT-5.4 mini, whose BASE personas were already the most dispersed, does not change (difference −0.003, interval −0.039 to 0.029). Dispersion also rises for all three regional models on Pew. The personas move a long way between conditions, by 0.22 on average on the WVS, about the size of the typical difference between two scholars portrayed by the same model.

Whether retrieval also lowers the model's share of the variance depends on how missing cells are handled. In matched panels with the same models in both conditions and with sparse cells filled, the model share falls from 0.55 to 0.34 on the WVS and from 0.60 to 0.36 on Pew. The scholar's share rises from 0.14 to 0.21 and from 0.18 to 0.26. On the 43 WVS items that every persona answered in both conditions, the model share falls less, from 0.45 to 0.36, and the item-wise difference is not significant ($p = 0.22$). For the 23 complete Pew items it falls from 0.44 to 0.37 ($p = 0.07$). The direction is the same throughout, but on complete-case items the decline is smaller and could be due to chance.

Retrieval does not consistently move personas toward their traditions. Tradition separation rises with retrieval for the two commercial models only (GPT-5.4 mini from 1.08 to 1.47, DeepSeek from 1.06 to 1.26). It falls for the three regional models with both conditions, most for GLM-4 (from 1.56 to 1.15), and on Pew it falls for all three. Without retrieval, the smaller models seem to answer largely from the tradition named in the prompt, which separates the traditions cleanly but says little about the individual scholar; retrieved passages break up these blocks. The known-group contrasts do not improve either. Averaged over the five models with both conditions, the difference in the expected direction grows with retrieval for 3 of the 7 contrasts and 2 of the 5 families (exact



Note: Secular and emancipative values (Welzel 2013). Lines join each scholar's BASE (open) and RAG (filled) positions. Grey points are WVS Wave 7 countries, with reference states labelled; dotted lines mark the least and most secular country.

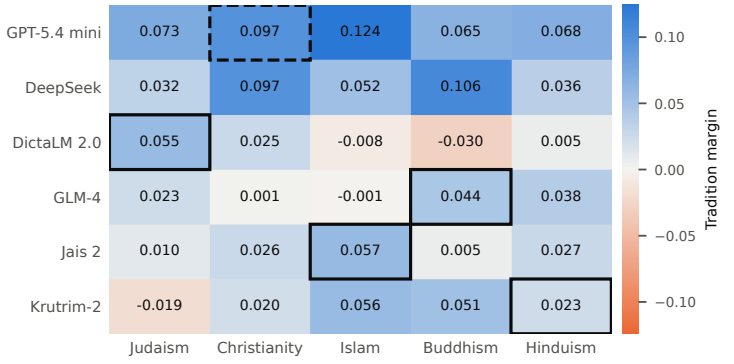
Figure 9. Personas on the cultural map

one-sided $p = 0.71$ and 0.78); the mean change per contrast is −0.028. Figure 9 shows how far and in which directions the personas move on the cultural map. For DeepSeek, retrieval lowers the average secular score from 0.58 to 0.45. For GPT-5.4 mini and GLM-4 the moves are large but go in all directions, and DictaLM's personas stay within a small region at the traditional, low-emancipation corner of the map.

7.4 H4: regional models and home traditions

Turning to H4, we find no evidence of a systematic home-tradition advantage. The tradition margin measures how clearly a model places a scholar within his or her tradition, and three of the four regional models do record their largest margin for their home tradition: DictaLM for Judaism (0.055), Jais 2 for Islam (0.057) and GLM-4 for Buddhism (0.044). Krutrim-2 does not; its Hindu margin (0.023) is smaller than its Islamic and Buddhist margins (Figure 10). The home coefficient, 0.029, is matched or exceeded by three of the 24 possible assignments of home traditions to the regional models, the actual one included (exact $p_+ = 0.125$; two-sided 0.25). With four models the test cannot distinguish this pattern from chance. In the exploratory version with GPT-5.4 mini as the Christian pairing, $p_+ = 0.033$ over 120 assignments (two-sided 0.067). GPT-5.4 mini's Christian margin is above its average across traditions, but its Islamic margin is higher still, and with no regional Christian model to compare we do not read this as evidence of a Christian affinity.

On closeness to the home population the estimate points the other way, although not significantly.



Note: Mean distance to personas of other traditions minus mean distance to personas of the same tradition (WVS, RAG). Outlined cells are the home pairings; the dashed outline marks the exploratory pairing.

Figure 10. Tradition margin by model and tradition

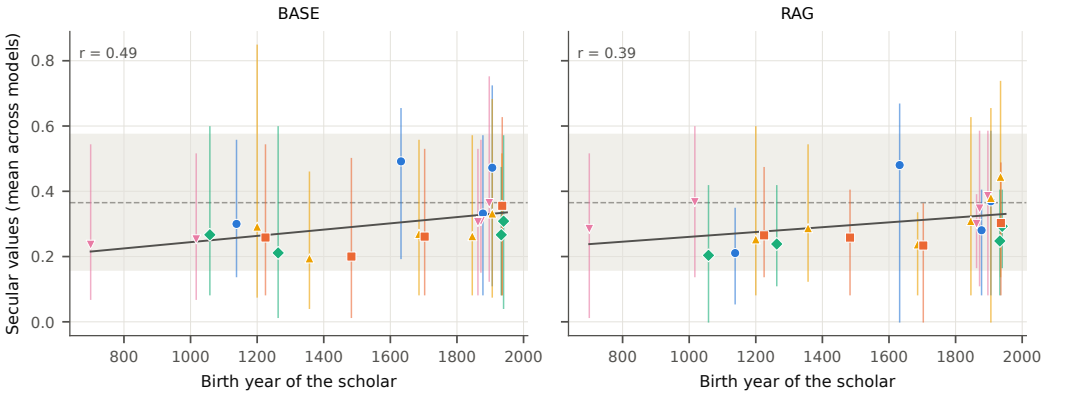
When the WVS countries are ranked by their distance to a persona, the regional models place their home scholars' reference states further down the ranking than other models do (coefficient 0.20 on the zero-to-one rank scale). The actual pairing is the least favourable of the 24 possible assignments, which gives a one-sided p of 0.042 in the direction H4 did not predict and a two-sided p of 0.083. We treat this home disadvantage as an exploratory observation. On Pew the coefficient has the same sign and is not significant (one-sided $p = 0.12$). Because no Jewish scholar has Israel as a WVS reference state, the WVS version of this test asks DictaLM to be close to Egypt, the Netherlands, Germany and the United States; the Pew version, which uses Israel, gives the same answer.

7.5 H5: personas and present-day populations

H5 receives little support once the dependence between personas is taken into account, and the pattern differs from the one the earlier study reported. On the two indices, personas with retrieval are closer to the mean of all countries than to their own reference states in 64% of the 119 model-scholar cells (Wilcoxon $p = 0.03$ treating cells as independent), and without retrieval in 57% ($p = 0.66$). The cells are not independent, however. Averaged over models, 7 of the 21 scholars are closer to their reference states than to the global mean ($p = 0.10$), and the crossed bootstrap interval for the mean difference, −0.53 to 0.13, includes zero. Across the full set of value items the global and local distances are almost identical in both conditions (mean 0.381 and 0.381 with retrieval).

The rank of the reference state is more informative, and here the two conditions differ. Among the 66 WVS countries ordered by closeness to a persona, the reference state sits on average at percentile 49 with retrieval. We detect no advantage over the 50 that an arbitrary country would achieve ($p = 0.57$ over cells; crossed bootstrap interval for the difference from one half −0.10 to 0.09). This is a failure to find localization, not evidence that the reference state is no better a match than any other country. Without retrieval the reference state sits at percentile 42, somewhat better than chance over cells ($p = 0.011$) and across scholars (only 5 of 22 scholars have a mean rank worse than one half, $p = 0.04$). The crossed interval, −0.20 to 0.05, is nonetheless wide enough to include zero.

Contrary to the second part of H5, the personas are not uniformly modern: on average they are slightly less secular than their reference populations (mean difference −0.037 with retrieval and −0.044 without) and no more emancipative. Over cells these differences reach $p = 0.03$ and 0.03, but neither the scholar-level tests ($p = 0.14$ and 0.05) nor the crossed bootstrap (with retrieval, −0.14 to 0.07) excludes zero. The average hides large differences between models (Figure 9). Placed among individual WVS respondents, DeepSeek's median BASE persona sits at the 85th percentile of secular



Note: Points are means over models; vertical lines span the model range. The shaded band is the range of WVS Wave 7 countries and the dashed line their mean.

Figure 11. Secular values by birth year

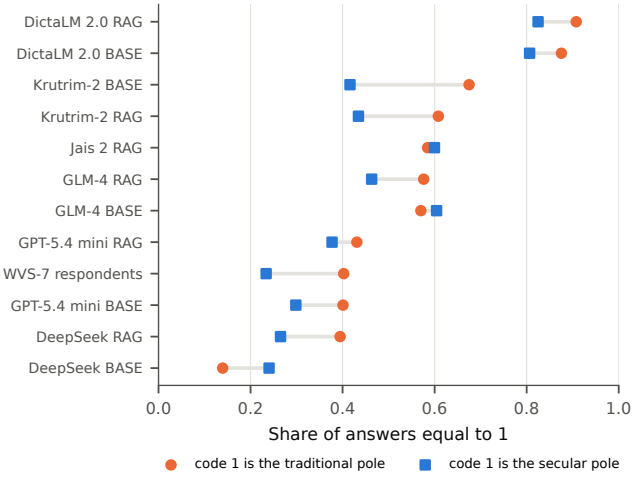
values, and 17% of its personas are more secular than 95% of respondents. DictaLM's median persona sits at the 5th percentile, and 74% of its personas are less secular than 95% of respondents; GLM-4 places 35% of its BASE personas there. GPT-5.4 mini's median persona sits near the middle (49th percentile). Whether a twelfth-century theologian comes out as a modern secularist or as more traditional than nearly every living respondent depends mainly on which model is asked.

The scholar's era leaves a modest trace in the answers (Figure 11). Averaged over models, scholars born later score as more secular, with a correlation of 0.49 without retrieval ($p = 0.02$) and 0.39 with it ($p = 0.07$). The slope is about 0.01 on the unit scale per century of birth. Over the twelve centuries between Shankara and Mernissi it adds up to 0.12, a quarter of the 0.48 that separates the average DeepSeek persona from the average DictaLM persona.

7.6 H6: response validity

Validity differs sharply across models and conditions (Table 12). Failures and invalid answers are concentrated in the smaller open-weight models, as H6 predicts, but the largest abstention rate belongs to GPT-5.4 mini. Jais 2 failed completely without retrieved context. GLM-4 gave invalid answers, mostly strings of repeated digits (41% of its invalid answers) and decimals (32%), to 17% of WVS items without retrieval and 11% with it. GPT-5.4 mini abstained on 20% of WVS items without retrieval, usually by giving the don't-know code, and on 8% with retrieval. The other models answered nearly everything. With retrieval, the local models gave the same answer in all five rounds for 98% to 99% of persona-item cells on the WVS, as expected at temperature zero. DeepSeek, sampled at the provider's default temperature, did so for only 21% without retrieval and 52% with it.

The most unusual response pattern is DictaLM's. It chose the first option of the scale for 78% of its valid WVS answers without retrieval and 82% with it. The comparison in Figure 12 is consistent with a strong preference for response code 1. On items where "1" is the traditional answer, DictaLM chose it 91% of the time with retrieval; on items where "1" is the secular answer, it still chose it 82% of the time. For WVS respondents the corresponding figures are 40% and 23%. The preference carries a small directional component, and since the options were never shown in reversed order the two item sets differ in content as well as in the meaning of "1". Even so, much of DictaLM's apparent traditionalism, and with it much of its extreme position on the cultural map, appears to come from where the options sit on the scale rather than from what they say. Krutrim-2 without retrieval shows the opposite profile: its choice of "1" depends on direction more strongly than that



Note: Share of valid answers equal to 1 on WVS items where code 1 is the traditional pole and where it is the secular pole.

Figure 12. When does a model answer "1"?

of human respondents (67% and 42%, against 40% and 23%). On the demographic items used as a consistency check, personas reported the scholar's sex correctly in 88% of model-condition-scholar cells and a religion consistent with the scholar's tradition in 90%, with DictaLM again lowest (67%).

7.7 H7: a second instrument

The Pew items, selected and fielded independently of the WVS, reproduce the patterns that H7 covers. Because the same models, prompts and corpora answer both instruments, this is a check on measurement rather than an independent replication, and it does not extend to the known-group contrasts, whose domains are defined on WVS items. Model identity outweighs scholar identity on Pew Set B with and without retrieval, and all four regional models separate the traditions with retrieval. Retrieval increases dispersion for all three models with both conditions and reduces tradition separation for all three, and the regional models' personas are no closer to their home populations than other models' personas are. Pew also allows a check on the commercial models for the eleven Hindu and Buddhist scholars, for whom both commercial Pew RAG runs exist. With the commercial and regional models together, the model share on Pew is 0.45 and the scholar share 0.17, close to the values for the same scholars and models on the WVS (0.47 and 0.13).

7.8 Sensitivity to corpus composition

The corpus audit raises the question of whether the retrieval results depend on the scholars whose corpora contain material not written by them. Dropping Shankara, Rāmānuja and Yin Shun changes little (Table 17 in Appendix 8). In the six-model WVS panel the model share becomes 0.37 and the scholar share 0.18, against 0.37 and 0.17 for all scholars. Tradition separation remains significant after Holm correction for 5 of the six models on the WVS and 4 of four on Pew. Dispersion still rises with retrieval for 4 of five models, and the home coefficient for the tradition margin stays small (0.035, p_+ = 0.125). Dropping the 14th Dalai Lama as well gives the same picture. The retrieval analyses for the Abrahamic scholars could not be checked in this way, because their text files were not available for the audit.

8. Discussion

We asked how faithfully LLM personas reproduce the value positions of religious scholars and what shapes those portrayals. The personas recover some documented differences between scholars, but model identity accounts for more of the variation in their answers than scholar identity, and two models' versions of one scholar differ more than one model's versions of two scholars. Traditions are separable in most models, and well-known disagreements between scholars point the expected way more often than the opposite way, although the evidence on the contrasts rests on a handful of comparisons. Retrieval mainly increases the differentiation among personas; on the contrasts that bear on historical fidelity its effect is nil or slightly negative. Provenance shows no detectable effect, and era leaves only a small trace.

These results revise the picture from our earlier study (Caliskan et al. 2025). That study concluded that models sometimes universalize and sometimes localize, with Muslim personas close to the global average and Christian personas close to their local populations. With the full set of value items rather than five, and with reference states ranked among all countries rather than compared with a single global mean, we no longer detect localization when the personas are grounded in retrieved passages. The apparent proximity to the global mean also does not survive an analysis that treats scholars and models, rather than their combinations, as the units. What remains is a model-specific placement on the cultural map that has little to do with where or when the scholar lived. The difference also illustrates a general point about the measurement of persona fidelity: averages over a few items can produce patterns that a fuller instrument does not sustain, and a single global benchmark can make proximity look meaningful when the benchmark is close to everything.

8.1 Who speaks when a model speaks for a scholar

The dominance of the model over the scholar fits the view of dialogue agents as role players rather than believers (Shanahan, McDonell, and Reynolds 2023) and the finding that persona variables explain little of the variance in model judgments (Hu and Collier 2024). A persona prompt tells the model whom to play, and the model continues to constrain the response profile after it has been told. How it does so differs sharply between models. Without retrieval, DeepSeek placed the median scholar among the most secular sixth of living WVS respondents, and DictaLM placed most of its scholars among the least secular twentieth, largely, it seems, because it prefers the first answer option. GPT-5.4 mini placed its personas, on average, near the middle of the world's distribution. These placements say little about Ibn Taymiyya or Ruether and a good deal about the models and the pipelines that served them. Because the design contains no persona-free baseline, we cannot say how much of each placement is a stable value orientation of the model. What we observe is the combined effect of a model, its prompt, its sampling settings and, for the local models, its quantization and serving template.

This has a direct implication for claims that AI can speak for religious authorities. The religious chatbots reviewed above make an implicit promise that the answer reflects the tradition or the figure in whose name it is given. Our results suggest that such answers should be read first as products of a particular model and pipeline, modified by the persona, and only secondarily as statements of the persona's subject. The idea that believers could consult foundational thinkers directly through AI, without the mediation of clergy or commentators, is not supported by this evidence. The mediation does not disappear. It moves from human interpreters who can be named, questioned and held to a tradition's standards, to a model whose interpretive habits are largely invisible to the user and change when the model is updated. In Campbell's (2007) terms, the persona borrows the authority of the text while the authority of the interpreter passes, unacknowledged, to the developer.

The disagreement between models is largest for the Hindu and Buddhist scholars and smallest for Aquinas, Maimonides and the Dalai Lama. One possible explanation, which our data cannot test directly, is the depth of each figure's footprint in English-language training data. Where much has

been written about a scholar, models may converge on a shared portrait, and where little has been written, each model fills the gap differently. The size of our own retrieval corpora does not account for the difference, but their composition may contribute in one case: the models disagree most about Rāmānuja, whose corpus contains a study of his thought but none of his own commentaries. The pattern matches the evidence that non-Western religions are represented more thinly and more stereotypically in LLMs (Plaza-del-Arco et al. 2024) and the argument that AI's data are skewed toward WEIRD societies (Mihalcea et al. 2025). It also means that the scholars for whom an AI interlocutor might be most useful to non-specialists, those whose works are least accessible in English, are those for whom current models are least consistent.

8.2 What grounding can and cannot do

Adding retrieved passages changed the personas' response profiles and spread them apart. In the full panels the model's share of the variance fell and the scholar's rose, but on items that every persona answered the decline is smaller and could be due to chance. The added differentiation did not translate into better recovery of known disagreements, and for the regional models it weakened rather than strengthened the separation of traditions. One reading is that without retrieval the smaller models answer largely from the tradition named in the prompt, producing a stereotyped "Buddhist" or "Muslim" block of answers, which retrieved passages break up. The passages, however, rarely bear directly on a survey question about, say, whether university education is more important for a boy than for a girl. A passage from Shankara's commentary on the *Bhagavad Gītā*, retrieved for such a question, can change the model's answer without telling it what Shankara would have said. We did not evaluate the retrieved passages themselves, so this explanation remains to be tested at the level of passages. The finding matters for the design of religious AI, where RAG is often presented as a guarantee of faithfulness (Bhatia et al. 2026). Corpus grounding can make the evidence available to a persona more explicit, but on our evidence it was insufficient to establish fidelity on value questions that the corpus does not address. Our corpus audit adds a practical caution: a corpus assembled under a scholar's name can contain studies, introductions and co-authored books, and in one case nothing by the scholar at all. Our main results do not depend on these corpora, but a system that retrieves from such material and answers in the first person attributes other authors' words to the scholar.

8.3 Regional models and the limits of provenance

The regional models did not portray their home traditions more faithfully in our tests. Three of the four placed their home tradition's scholars most clearly within that tradition, which is suggestive, but the pattern could arise by chance with four models, and no model's personas were closer to its home population. This is consistent with evidence that regional training does not by itself carry regional values (Naous et al. 2024). Several features of our design work against finding an advantage. The prompts and corpora were in English, while the regional models' distinctive training data are in Hebrew, Arabic, Indian languages or Chinese (AlKhamissi et al. 2024). The open-weight models were small and quantized, and the pairing of a national model with a religious tradition is itself loose; a Hebrew-language model encodes contemporary Israeli culture rather than the worlds of Maimonides or Spinoza. The result therefore does not show that regional models cannot represent their traditions. It shows that provenance alone, under the English-language conditions in which such models are often used, did not improve the fidelity measures used here. For debates about sovereign AI, the implication is that claims of cultural representativeness should be tested against the populations and traditions they invoke rather than assumed from where a model was built.

8.4 Limitations

The first group of limitations concerns what the design can identify. There is no persona-free or tradition-only baseline, so a model's characteristic answers are inferred from its personas rather than

measured directly. Nor can the comparison of retrieval with no retrieval separate the effect of the scholar's passages from the effect of receiving any passages at all; an irrelevant-text or shuffled-corpus control would be needed for that. Answer options were never presented in reversed order, so position preferences could only be diagnosed indirectly. The known-group contrasts are few and coarse, two scholars appear in two contrasts each, and specialists would draw some of the contrasts differently. A fuller validation would need expert ratings of how each scholar would answer each item, supported by passages from the scholars' works. The panel of 23 scholars is small for statistical comparisons of traditions, and with five or six models any generalization to other models should be made with care.

A second group concerns the instruments and the corpora. Some survey items ask premodern figures about institutions and technologies that did not exist in their lifetimes. We chose the WVS and Pew items because they allow comparison with living populations, not because they capture what the scholars cared about. Prompts and corpora were in English. The corpora reflect what has been translated and digitized, and our candidate-generation and corpus-availability criteria produced no women in the Buddhist and Hindu panels. The audit of the Buddhist and Hindu corpora found secondary and co-authored material and, for Rāmānuja, no primary text. The retrieval analyses do not change when the affected scholars are dropped, but the Abrahamic text files could not be audited in the same way, and the passages actually retrieved for each answer were not logged. The panel also includes a living person, whose views are documented in ways the others' are not and who could object to how they are portrayed.

A third group concerns the models and how they were served. The split between the commercial and the regional models is also a split by size and pipeline. The commercial models were sampled at their providers' default temperatures and the local models at temperature zero, so round-to-round variation is not comparable across the two groups; using modal answers limits but does not remove this difference. Local inference used 4-bit quantization, and 3-bit for Krutrim-2. Formatting problems, among them residual chat-template tokens in DictaLM's answers, empty BASE outputs from Jais 2 and malformed strings from GLM-4, suggest that serving configurations interact with model weights. Reruns with the official templates would show how much of these models' behaviour is due to serving. A total serving failure is not evidence that a model lacks relevant knowledge. On the commercial side, cost limited the Pew RAG runs to the Hindu and Buddhist scholars, and the BASE runs used a fixed question order and an earlier Pew questionnaire. Commercial models also change behind stable names, so the placements we report for GPT-5.4 mini and DeepSeek-V3.2 describe those versions at the time of collection. Finally, our measures summarize answers on survey scales, which may understate differences that would appear in open-ended answers and overstate those that come from scale use, as the case of DictaLM shows.

8.5 Ethics and future work

Simulating religious figures raises ethical questions beyond measurement. Some figures in the panel are revered within their traditions, while others remain contested, and putting words in their mouths, especially on questions they never addressed, risks misrepresenting them to people who trust the output. The risk is greatest when outputs are presented as authoritative, as some religious chatbots present them. The living scholar raises questions of consent and reputation, and the modern corpora raise questions of copyright, which is why we describe rather than distribute them. We see persona simulation as a research instrument for understanding what models carry about religious traditions, not as a substitute for the traditions' own interpreters.

Four extensions would strengthen the evidence. Specialist ratings of how each scholar would answer each item would allow a direct test of fidelity rather than the indirect tests used here. Prompting in the scholars' own languages, with corpora in those languages, would give the regional models a fairer test. Persona-free baselines for each model, together with irrelevant-passage controls, would measure what the persona and the retrieved text each add. And open-ended answers, coded

by specialists, would show whether the patterns we find on survey scales hold when models explain their reasoning in the scholar's voice.

8.6 Conclusion

Large language models can be made to speak in the voice of Maimonides, al-Ghazālī or Shankara, and with access to texts associated with them they speak in voices that differ from one another. Across six models, two instruments and 23 scholars, model identity explained more variation in survey responses than the combined tradition and scholar main effects. Retrieval made many personas more distinct, but the contrasts used for validation were recovered no better with it than without, and models built in a tradition's home region showed no detectable advantage in portraying it. These survey responses support stronger conclusions about differences among the tested model pipelines than about the historical figures themselves. For users of religious chatbots, the results counsel caution in treating a persona's answer as the view of the thinker it names. For researchers, they point to the next validation tasks: verifying the authorship of corpora and the passages actually retrieved, checking serving configurations, and comparing responses with item-by-item assessments by specialists.

Data and code availability

The item-level persona responses, the item catalog, the corpus manifest, the collection and analysis code and the derived benchmark files are available from the corresponding author on request, and we will deposit them in a public repository when the paper is published. The World Values Survey Wave 7 data are available from the World Values Survey Association (Haerpfer et al. 2024), and the Pew Spring 2024 toplines from the Pew Research Center (Pew Research Center 2025). The retrieval corpora include works under copyright and are not redistributed; Appendix 2 lists the inventory titles and the cleaned files with their authorship status, word counts and chunk counts.

Funding

The scholars received no external funding for this research and were self-funded.

Competing interests

The authors declare no competing interests.

Ethics

The study recruited no participants. It used publicly released survey data and responses generated by language models.

Author contributions

C.C. conceived the study and designed the analysis. A.F. and M.A. assembled and cleaned the corpora, built the retrieval pipelines and collected the model responses. C.C., A.F. and M.A. wrote the manuscript. All authors reviewed and approved the final manuscript.

References

Abid, Abubakar, Maheen Farooqi, and James Zou. 2021. Large language models associate Muslims with violence. *Nature Machine Intelligence* 3:461–463. https://doi.org/10.1038/s42256-021-00359-2.

Abrahamian, Ervand. 1982. Ali Shari'ati: ideologue of the Iranian revolution. *MERIP Reports,* no. 102 (January). https://merip.org/1982/01/ali-shariati-ideologue-of-the-iranian-revolution/.

AlKhamissi, Badr, Muhammad ElNokrashy, Mai AlKhamissi, and Mona Diab. 2024. Investigating cultural alignment of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* edited by Lun-Wei Ku, Andre Martins, and Vivek Srikumar, 12404–12422. Bangkok, Thailand: Association for Computational Linguistics, August. https://doi.org/10.18653/v1/2024.acl-long.671. https://aclanthology.org/2024.acl-long.671/.

Anwar, Mohamed, Abed Alhakim Freihat, George Ibrahim, Mostafa Awad, Abdelrahman Sadallah, Gurpreet Gosal, Gokulakrishnan Ramakrishnan, Sarath Chandran, Biswajit Mishra, Rituraj Joshi, et al. 2026. *Jais 2: a family of Arabic-centric open large language models.* Earlier released as an IFM technical report dated 9 December 2025. arXiv: 2608.13580 [cs.CL]. https://arxiv.org/abs/2608.13580.

Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. Out of one, many: using language models to simulate human samples. *Political Analysis* 31 (3): 337–351. https://doi.org/10.1017/pan.2023.2.

Artifex Software, Inc. 2026. *PyMuPDF: high-performance Python library for data extraction, analysis, conversion and manipulation of PDF documents.* GitHub repository pymupdf/PyMuPDF. Version 1.27.x series; GNU AGPL v3 or commercial license; accessed 7 November 2026. https://github.com/pymupdf/PyMuPDF.

Atari, Mohammad, Mona J. Xue, Peter S. Park, Damián E. Blasi, and Joseph Henrich. 2023. *Which humans?* PsyArXiv preprint. https://psyarxiv.com/5b26t.

Atif, Farah, Nursultan Askarbekuly, Kareem Darwish, and Monojit Choudhury. 2025. Sacred or synthetic? evaluating LLM reliability and abstention for religious questions. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,* 8:217–226. https://doi.org/10.1609/aies.v8i1.36543. https://ojs.aaai.org/index.php/AIES/article/view/36543.

Augustine. 1838. *The Confessions of S. Augustine.* Translated by Edward Bouverie Pusey. Library of the Fathers 1. Oxford: J. H. Parker.

Bhatia, Gagan, Hamdy Mubarak, Mustafa Jarrar, George Mikros, Fadi Zaraket, Mahmoud Alhirthani, Mutaz al-Khatib, et al. 2026. From RAG to agentic RAG for faithful Islamic question answering. In *Findings of the Association for Computational Linguistics: ACL 2026,* edited by Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens, 26469–26488. San Diego, California, United States: Association for Computational Linguistics, July. https://doi.org/10.18653/v1/2026.findings-acl.1317. https://aclanthology.org/2026.findings-acl.1317/.

Bhikṣu Changtzu. 2017. *Introducing Venerable Yinshun to the West.* Hsinchu: Yinshun Cultural and Educational Foundation.

Bible AI. 2026. *Bible AI.* Website. Accessed 7 November 2026. https://bibleai.com.

Bingenheimer, Marcus. 2009. Writing history of Buddhist thought in the twentieth century: Yinshun (1906–2005) in the context of Chinese Buddhist historiography. *Journal of Global Buddhism* 10:255–290. https://www.globalbuddhism.org/article/download/1152/987.

Bisbee, James, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson. 2024. Synthetic replacements for human survey data? The perils of large language models. *Political Analysis* 32 (4): 401–416. https://doi.org/10.1017/pan.2024.5.

Borgeaud, Sebastian, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, et al. 2022. Improving language models by retrieving from trillions of tokens. In *Proceedings of the 39th International Conference on Machine Learning,* edited by Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, 162:2206–2240. Proceedings of Machine Learning Research. PMLR. https://proceedings.mlr.press/v162/borgeaud22a.html.

Braun, Erik. 2013. *The birth of insight: meditation, modern Buddhism, and the Burmese monk Ledi Sayadaw.* Chicago: University of Chicago Press.

Caliskan, Cantay, Mehmet Karabela, Xinran Wang, Yiheng Yao, and Keming Zhang. 2025. HalalLLM vs. KosherLLM: when Abrahamic scholars confront modern questions through AI. Unpublished manuscript, University of Rochester.

Campbell, Heidi A. 2007. Who's got the power? Religious authority and the Internet. *Journal of Computer-Mediated Communication* 12, no. 3 (April). https://doi.org/10.1111/j.1083-6101.2007.00362.x.

———. 2012. Understanding the relationship between religion online and offline in a networked society. *Journal of the American Academy of Religion* 80 (1): 64–93. https://relbib.de/Record/164665482X.

Chatbot Eli. 2026. *Chatbot Eli.* Website. Site footer reads 2020–2022; accessed 7 November 2026. https://chatboteli.com.

d'Entrèves, Maurizio Passerin. 2026. Hannah Arendt. In *The Stanford Encyclopedia of Philosophy,* Fall 2026, edited by Edward N. Zalta and Uri Nodelman. First published 27 July 2006; substantive revision 12 February 2024. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2026/entries/arendt/.

Dalai Lama. 2005. *The universe in a single atom: the convergence of science and spirituality.* New York: Morgan Road Books.

Dalal, Neil. 2021. Śaṅkara. In *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta and Uri Nodelman. First published 4 October 2021. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/shankara/.

DeepSeek-AI. 2025. *DeepSeek-V3.2: pushing the frontier of open large language models.* arXiv: 2512.02556 [cs.CL]. https://arxiv.org/abs/2512.02556.

Dominguez-Olmedo, Ricardo, Moritz Hardt, and Celestine Mendler-Dünner. 2024. Questioning the survey responses of large language models. In *Advances in Neural Information Processing Systems,* 37:45850–45878. https://proceedings.neurips.cc/paper_files/paper/2024/hash/515c62809e0a29729d7eec26e2916fc0-Abstract.html.

Doniger, Wendy, and Wilfried Huchzermeyer. 2024. *Sri Aurobindo.* Encyclopædia Britannica. Last revised 2 April 2024; accessed 7 November 2026. https://www.britannica.com/biography/Sri-Aurobindo.

Douze, Matthijs, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2024. *The Faiss library.* arXiv: 2401.08281 [cs.LG]. https://arxiv.org/abs/2401.08281.

Durmus, Esin, Karina Nguyen, Thomas I. Liao, Nicholas Schiefer, Amanda Askell, Anton Bakhtin, Carol Chen, et al. 2023. *Towards measuring the representation of subjective global opinions in language models.* Version 2, 12 April 2024. arXiv: 2306.16388 [cs.CL]. https://arxiv.org/abs/2306.16388.

Griffel, Frank. 2026. Al-Ghazali. In *The Stanford Encyclopedia of Philosophy,* Fall 2026, edited by Edward N. Zalta and Uri Nodelman. First published 14 August 2007; substantive revision 15 January 2026. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2026/entries/al-ghazali/.

Haerpfer, Christian W., Ronald Inglehart, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Jaime Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, and Bi Puranen. 2024. *World Values Survey Wave 7 (2017–2022) cross-national data-set.* Dataset, version 6.0.0. https://doi.org/10.14281/18241.24.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33 (2–3): 61–83. https://doi.org/10.1017/S0140525X0999152X.

Howard University School of Divinity. 2022. *Remembering Dr. Rosemary Radford Ruether.* Howard University School of Divinity website. Undated tribute page with statements by Alice Ogden Bellis, Renee K. Harrison, Harold Dean Trulear, Ronald Hopson, and Cheryl J. Sanders; accessed 7 November 2026. https://divinity.howard.edu/remembering-dr-rosemary-radford-ruether/.

Hu, Tiancheng, and Nigel Collier. 2024. Quantifying the persona effect in LLM simulations. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* edited by Lun-Wei Ku, Andre Martins, and Vivek Srikumar, 10289–10307. Bangkok, Thailand: Association for Computational Linguistics, August. https://doi.org/10.18653/v1/2024.acl-long.554. https://aclanthology.org/2024.acl-long.554/.

Hu, Tiancheng, Yara Kyrychenko, Steve Rathje, Nigel Collier, Sander van der Linden, and Jon Roozenbeek. 2024. Generative language models exhibit social identity biases. Published online December 2024, *Nature Computational Science,* https://doi.org/10.1038/s43588-024-00741-1.

Huang, Huang, Fei Yu, Jianqing Zhu, Xuening Sun, Hao Cheng, Dingjie Song, Zhihong Chen, et al. 2024. AceGPT, localizing large language models in Arabic. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers),* edited by Kevin Duh, Helena Gomez, and Steven Bethard, 8139–8163. Mexico City, Mexico: Association for Computational Linguistics, June. https://doi.org/10.18653/v1/2024.naacl-long.450. https://aclanthology.org/2024.naacl-long.450/.

Inception and MBZUAI and Cerebras. 2025. *Jais-2-8B-Chat.* Hugging Face model card inceptionai/Jais-2-8B-Chat. Apache-2.0 license; accessed 7 November 2026. https://huggingface.co/inceptionai/Jais-2-8B-Chat.

Inglehart, Ronald, and Christian Welzel. 2005. *Modernization, cultural change, and democracy: the human development sequence.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790881. https://www.cambridge.org/core/books/abs/modernization-cultural-change-and-democracy/introduction/C217BBFF5237D90224237C704E182011.

Inglehart, Ronald F. 2018. *Cultural evolution: people's motivations are changing, and reshaping the world.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108613880. https://www.cambridge.org/core/books/cultural-evolution/global-cultural-patterns/E93E7B2545810AB094A68A70BCD217E2.

Islam&AI. 2026. *QURAN-NLP: Quran, hadith, translations, tafaseer, corpus linguistics.* GitHub repository islamAndAi/QURAN-NLP. Apache-2.0 (code); accessed 7 November 2026. https://github.com/islamAndAi/QURAN-NLP.

IslamiCity. 2025. *ChatILM: enhancing Islamic knowledge with AI.* IslamiCity.org. Published 24 July 2025; accessed 7 November 2026. https://islamicity.org/100779/chatilm-enhancing-islamic-knowledge-with-ai/.

Israelsen, Brett, Sheryl Carty, Josh Coates, Nancy Fulda, Julie Park, and Pete Whiting. 2026. *When AI takes sides on questions of faith: persistent asymmetries in AI-mediated faith guidance.* Version 2, 28 May 2026. arXiv: 2605.22975. https://arxiv.org/abs/2605.22975.

Jackson, Joshua Conrad, Kai Chi Yam, Pok Man Tang, Ting Liu, and Azim Shariff. 2023. Exposure to robot preachers undermines religious commitment. *Journal of Experimental Psychology: General* 152 (12): 3344–3358. https://doi.org/10.1037/xge0001443.

JaidedAI. 2024. *EasyOCR: ready-to-use OCR with 80+ supported languages.* GitHub repository JaidedAI/EasyOCR. Version 1.7.2 (24 September 2024); Apache-2.0 license. https://github.com/JaidedAI/EasyOCR.

Johnson, Jeff, Matthijs Douze, and Hervé Jégou. 2021. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data* 7 (3): 535–547. https://doi.org/10.1109/TBDATA.2019.2921572.

Kallappa, Aditya, Guo Xiang, Jay Piplodiya, Manoj Guduru, Neel Rachamalla, Palash Kamble, Souvik Rana, et al. 2025. *Krutrim-2 LLM.* Hugging Face model card krutrim-ai-labs/Krutrim-2-instruct; code at https://github.com/ola-krutrim/Krutrim-2-12B. Krutrim Community License 1.0; accessed 7 November 2026. https://huggingface.co/krutrim-ai-labs/Krutrim-2-instruct.

Khan, Qamar-ud-Din. 1973. *The political thought of Ibn Taymiyah.* Islamabad: Islamic Research Institute.

Khandelwal, Khyati, Manuel Tonneau, Andrew M. Bean, Hannah Rose Kirk, and Scott A. Hale. 2024. Indian-BhED: a dataset for measuring India-centric biases in large language models. In *Proceedings of the International Conference on Information Technology for Social Good (GoodIT '24).* Bremen, Germany: Association for Computing Machinery. https://doi.org/10.1145/3677525.3678666.

Laoust, Henri. 2024. *Ibn Taymiyyah.* Encyclopædia Britannica. Last revised 14 March 2024; accessed 7 November 2026. https://www.britannica.com/biography/Ibn-Taymiyyah.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, et al. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems,* 33:9459–9474. https://papers.nips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Li, Cheng, Mengzhuo Chen, Jindong Wang, Sunayana Sitaram, and Xing Xie. 2024. CultureLLM: incorporating cultural differences into large language models. In *Advances in Neural Information Processing Systems,* 37:84799–84838. https://papers.neurips.cc/paper_files/paper/2024/hash/9a16935bf54c4af233e25d998b7f4a2c-Abstract-Conference.html.

Liang, Percy, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, et al. 2023. Holistic evaluation of language models. *Transactions on Machine Learning Research,* https://openreview.net/forum?id=iO4LZibEqW.

Lopez, Donald S. 2024. *14th Dalai Lama.* Encyclopædia Britannica. Last revised 29 April 2024; accessed 7 November 2026. https://www.britannica.com/biography/Dalai-Lama-14th.

Luther, Martin. 1962. The estate of marriage. In *Luther's Works, Volume 45: The Christian in Society II,* edited by Walther I. Brandt. Philadelphia: Muhlenberg Press.

Mihalcea, Rada, Oana Ignat, Longju Bai, Angana Borah, Luis Chiruzzo, Zhijing Jin, Claude Kwizera, Joan Nwatu, Soujanya Poria, and Thamar Solorio. 2025. Why AI is WEIRD and shouldn't be this way: towards AI for everyone, with everyone, by everyone. In *Proceedings of the AAAI Conference on Artificial Intelligence,* 39:28657–28670. https://doi.org/10.1609/aaai.v39i27.35092. https://ojs.aaai.org/index.php/AAAI/article/view/35092.

miriam cooke. 2015. Fatima Mernissi 1940–2015. *Review of Middle East Studies* 49 (2): 217–219. https://doi.org/10.1017/rms.2016.33. https://www.cambridge.org/core/journals/review-of-middle-east-studies/article/fatima-mernissi-19402015/E5581CDCCAEA56AAA95756EBB1FFD4A6.

Nadler, Steven. 2026. Baruch Spinoza. In *The Stanford Encyclopedia of Philosophy,* Fall 2026, edited by Edward N. Zalta and Uri Nodelman. First published 29 June 2001; substantive revision 8 November 2023. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2026/entries/spinoza/.

Nagatomo, Shigenori. 2024. Japanese Zen Buddhist philosophy. In *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta and Uri Nodelman. First published 28 June 2006; substantive revision 7 March 2024. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/japanese-zen/.

Namgyal Monastery Institute of Buddhist Studies. 2026. *The Dalai Lamas.* Website. Accessed 7 November 2026. https://namgyal.org/pages/dalailamas.

Naous, Tarek, Michael J. Ryan, Alan Ritter, and Wei Xu. 2024. Having beer after prayer? Measuring cultural bias in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* 16366–16393. Bangkok, Thailand: Association for Computational Linguistics, August. https://doi.org/10.18653/v1/2024.acl-long.862. https://aclanthology.org/2024.acl-long.862/.

Nisargadatta Maharaj. 1973. *I Am That: talks with Sri Nisargadatta Maharaj.* Translated by Maurice Frydman. Bombay: Chetana.

OpenAI. 2026a. *GPT-5.4 mini (model ID gpt-5.4-mini; snapshot gpt-5.4-mini-2026-03-17).* OpenAI API documentation. Accessed 7 November 2026. https://developers.openai.com/api/docs/models/gpt-5.4-mini.

———. 2026b. *Introducing GPT-5.4 mini and nano.* OpenAI announcement, March. https://openai.com/index/introducing-gpt-5-4-mini-and-nano/.

Ovadia, Oded, Menachem Brief, Moshik Mishaeli, and Oren Elisha. 2024. Fine-tuning or retrieval? comparing knowledge injection in LLMs. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing,* edited by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, 237–250. Miami, Florida, USA: Association for Computational Linguistics, November. https://doi.org/10.18653/v1/2024.emnlp-main.15. https://aclanthology.org/2024.emnlp-main.15/.

Paravizian, Saja. 2026. *Al-Ghazālī (c. 1056–1111).* Internet Encyclopedia of Philosophy (ISSN 2161-0002). Undated entry; accessed 7 November 2026. https://iep.utm.edu/al-ghazali/.

Park, Joon Sung, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23).* 22 pages. San Francisco, CA, USA: Association for Computing Machinery. https://doi.org/10.1145/3586183.3606763.

Pasnau, Robert. 2026. Thomas Aquinas. In *The Stanford Encyclopedia of Philosophy,* Fall 2026, edited by Edward N. Zalta and Uri Nodelman. First published 7 December 2022. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2026/entries/aquinas/.

Pew Research Center. 2025. *Spring 2024 survey data.* Dataset, Global Attitudes Survey (35 countries; fieldwork 5 January–22 May 2024). Updated 14 August 2026 with an erratum on the Mexico questionnaire. https://doi.org/10.58094/dwf6-w702. https://www.pewresearch.org/dataset/spring-2024-survey-data/.

Plaza-del-Arco, Flor Miriam, Amanda Cercas Curry, Susanna Paoli, Alba Cercas Curry, and Dirk Hovy. 2024. Divine LLaMAs: bias, stereotypes, stigmatization, and emotion representation of religion in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024,* edited by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, 4346–4366. Miami, Florida, USA: Association for Computational Linguistics, November. https://doi.org/10.18653/v1/2024.findings-emnlp.251. https://aclanthology.org/2024.findings-emnlp.251/.

Ramezani, Aida, and Yang Xu. 2023. Knowledge of cultural moral norms in large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* edited by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, 428–446. Toronto, Canada: Association for Computational Linguistics, July. https://doi.org/10.18653/v1/2023.acl-long.26. https://aclanthology.org/2023.acl-long.26/.

Ranganathan, Shyam. 2026. *Rāmānuja (c. 1017–c. 1137).* Internet Encyclopedia of Philosophy (ISSN 2161-0002). Undated entry; accessed 7 November 2026. https://iep.utm.edu/ramanuja/.

RavGPT. 2026. *RavGPT.* Website. Ravgpt.ai now redirects to rav.chat ("RavChat"); accessed 7 November 2026. https://ravgpt.ai.

Reimers, Nils, and Iryna Gurevych. 2019. Sentence-BERT: sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP),* edited by Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, 3982–3992. Hong Kong, China: Association for Computational Linguistics, November. https://doi.org/10.18653/v1/D19-1410. https://aclanthology.org/D19-1410/.

Röttger, Paul, Valentin Hofmann, Valentina Pyatkin, Musashi Hinck, Hannah Kirk, Hinrich Schuetze, and Dirk Hovy. 2024. Political compass or spinning arrow? Towards more meaningful evaluations for values and opinions in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* edited by Lun-Wei Ku, Andre Martins, and Vivek Srikumar, 15295–15311. Bangkok, Thailand: Association for Computational Linguistics, August. https://doi.org/10.18653/v1/2024.acl-long.816. https://aclanthology.org/2024.acl-long.816/.

Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? In *Proceedings of the 40th International Conference on Machine Learning,* edited by Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, 202:29971–30004. Proceedings of Machine Learning Research. PMLR. https://proceedings.mlr.press/v202/santurkar23a.html.

Seeskin, Kenneth. 2026. Maimonides. In *The Stanford Encyclopedia of Philosophy,* Fall 2026, edited by Edward N. Zalta and Uri Nodelman. First published 24 January 2006; substantive revision 18 November 2024. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2026/entries/maimonides/.

Shanahan, Murray, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature* 623 (7987): 493–498. https://doi.org/10.1038/s41586-023-06647-8.

Shao, Yunfan, Linyang Li, Junqi Dai, and Xipeng Qiu. 2023. Character-LLM: a trainable agent for role-playing. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing,* edited by Houda Bouamor, Juan Pino, and Kalika Bali, 13153–13187. Singapore: Association for Computational Linguistics, December. https://aclanthology.org/2023.emnlp-main.814/.

Shmidman, Shaltiel, Avi Shmidman, Amir D. N. Cohen, and Moshe Koppel. 2024. *Adapting LLMs to Hebrew: unveiling DictaLM 2.0 with enhanced vocabulary and instruction capabilities.* arXiv: 2407.07080 [cs.CL]. https://arxiv.org/abs/2407.07080.

Smith, Ray. 2007. An overview of the Tesseract OCR engine. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007),* 629–633. IEEE Computer Society. https://doi.org/10.1109/ICDAR.2007.56.

Sorensen, Taylor, Jared Moore, Jillian Fisher, Mitchell L. Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, et al. 2024. Position: a roadmap to pluralistic alignment. In *Proceedings of the 41st International Conference on Machine Learning,* edited by Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, 235:46280–46302. Proceedings of Machine Learning Research. PMLR. https://proceedings.mlr.press/v235/sorensen24a.html.

Soudani, Heydar, Evangelos Kanoulas, and Faegheh Hasibi. 2024. Fine tuning vs. retrieval augmented generation for less popular knowledge. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region (SIGIR-AP '24),* 12–22. Tokyo, Japan: Association for Computing Machinery. https://doi.org/10.1145/3673791.3698415. https://arxiv.org/abs/2403.01432.

Sounds True. 2026. *Sri Nisargadatta Maharaj.* Publisher's author page. Accessed 7 November 2026. https://www.soundstrue.com/blogs/authors/sri-nisargadatta-maharaj.

Sparham, Gareth, and Chandra Chiara Ehm. 2024. Tsongkhapa. In *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta and Uri Nodelman. First published 18 July 2011; substantive revision 17 April 2024. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/tsongkhapa/.

Stern, Robert, and Volker Leppin. 2026. Martin Luther. In *The Stanford Encyclopedia of Philosophy,* Fall 2026, edited by Edward N. Zalta and Uri Nodelman. First published 22 July 2020; substantive revision 16 January 2025. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2026/entries/luther/.

Tao, Yan, Olga Viberg, Ryan S. Baker, and René F. Kizilcec. 2024. Cultural bias and cultural alignment of large language models. *PNAS Nexus* 3 (9): pgae346. https://doi.org/10.1093/pnasnexus/pgae346.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, et al. 2024. *ChatGLM: a family of large language models from GLM-130B to GLM-4 all tools.* arXiv: 2406.12793 [cs.CL]. https://arxiv.org/abs/2406.12793.

The Center for Unity. 2026. *AskJesus chatbot.* Website. Accessed 7 November 2026. https://askjesuschatbot.com.

The Editors of Encyclopædia Britannica. 2024a. *Dōgen.* Encyclopædia Britannica. Last revised 28 February 2024; accessed 7 November 2026. https://www.britannica.com/biography/Dogen.

———. 2024b. *John Wesley.* Encyclopædia Britannica. Last revised 15 April 2024; accessed 7 November 2026. https://www.britannica.com/biography/John-Wesley.

———. 2026. *Vivekananda.* Encyclopædia Britannica. Accessed 7 November 2026. https://www.britannica.com/biography/Vivekananda.

Thomas Aquinas. 1947. *Summa theologica.* Translated by Fathers of the English Dominican Province. New York: Benziger Brothers.

Tjuatja, Lindia, Valerie Chen, Tongshuang Wu, Ameet Talwalkar, and Graham Neubig. 2024. Do LLMs exhibit human-like response biases? A case study in survey design. *Transactions of the Association for Computational Linguistics* 12:1011–1026. https://doi.org/10.1162/tacl_a_00685. https://aclanthology.org/2024.tacl-1.56/.

Tsuria, Ruth, and Yossi Tsuria. 2024. Artificial intelligence's understanding of religion: investigating the moralistic approaches presented by generative artificial intelligence tools. *Religions* 15 (3): 375. https://doi.org/10.3390/rel15030375.

Varnum, Michael E. W., Nicolas Baumard, Mohammad Atari, and Kurt Gray. 2024. Large language models based on historical text could offer informative tools for behavioral science. *Proceedings of the National Academy of Sciences of the United States of America* 121:e2407639121. https://doi.org/10.1073/pnas.2407639121.

Wang, Angelina, Jamie Morgenstern, and John P. Dickerson. 2024. *Large language models that replace human participants can harmfully misportray and flatten identity groups.* Version 3, 3 February 2025. arXiv: 2402.01908 [cs.CY]. https://arxiv.org/abs/2402.01908.

Wang, Wenhui, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. MiniLM: deep self-attention distillation for task-agnostic compression of pre-trained transformers. In *Advances in Neural Information Processing Systems,* 33:5776–5788. https://papers.nips.cc/paper/2020/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Watt, W. Montgomery. 2026. *Al-Ghazālī.* Encyclopædia Britannica. Last revised 1 January 2026; accessed 7 November 2026. https://www.britannica.com/biography/al-Ghazali.

Welzel, Christian. 2013. *Freedom rising: human empowerment and the quest for emancipation.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139540919. https://www.cambridge.org/core/books/freedom-rising/introduction/55D068262C8B08BA7C4E3BC56CB87E5F.

Wingate, David, Sheryl Carty, Joshua Coates, Daniel Feldman, Nancy Fulda, Larry Howell, Brett Israelson, et al. 2026. *Omissive bias in religious representation: benchmarking LLM answers to everyday ethical decision-making.* arXiv: 2605.24319. https://arxiv.org/abs/2605.24319.

Z.ai (Zhipu AI). 2025. *GLM-4-9B-0414.* Hugging Face model card zai-org/GLM-4-9B-0414. Released 14 April 2025; MIT license; accessed 7 November 2026. https://huggingface.co/zai-org/GLM-4-9B-0414.

Zank, Michael, and Zachary Braiterman. 2026. Martin Buber. In *The Stanford Encyclopedia of Philosophy,* Spring 2026, edited by Edward N. Zalta and Uri Nodelman. First published 20 April 2004; substantive revision 28 July 2020. Superseded by a 12 May 2026 revision by Zachary Braiterman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2026/entries/buber/.

Appendix 1. Scholar profiles

This appendix summarizes who the scholars were and why each was included. It also records the positions on which the known-group contrasts of H2b rest; the last subsection sets out the expected direction of each contrast and the items used to test it.

Appendix 1.1 Judaism

Moses Maimonides (1138–1204), born in Córdoba and active for most of his life in Fustat, combined religious law and philosophy. His fourteen-book code, the *Mishneh Torah*, made him the leading halakhic authority of his age, and his *Guide of the Perplexed* sought to reconcile biblical faith with Aristotelian philosophy (Seeskin 2026). Later readers in other traditions, Aquinas among them, engaged with his work. His rationalism works inside a framework in which revealed law governs communal life, which is why he serves as the expected-lower pole in the contrast with Spinoza on the authority of religion.

Baruch Spinoza (1632–1677) broke with that framework. Expelled from Amsterdam's Sephardic community, he identified God with Nature, read scripture as a historical document and argued in the *Theological-Political Treatise* for freedom of thought and a state not governed by religious authorities (Nadler 2026). Spinoza provides a radical, humanistic voice from within the Jewish intellectual tradition and the expected-higher pole in the secular-authority contrast.

Martin Buber (1878–1965) brought a dialogical philosophy of religion to the twentieth century. *I and Thou* (1923) distinguished the relation of mutual address from the instrumental relation to things, and his writings on Hasidism, education and politics made him a major figure in modern Jewish thought (Zank and Braiterman 2026). He taught in Germany and later in Jerusalem, where he argued for Arab–Jewish reconciliation within the Zionist movement.

Hannah Arendt (1906–1975) is better known as a political philosopher than as a religious thinker, but her Jewish identity and her writing on the moral catastrophes of the twentieth century make her an important member of the panel (d'Entrèves 2026). A student of Heidegger and Jaspers who fled Nazi Germany, she analysed totalitarianism, the conditions of political action and, in *Eichmann in Jerusalem*, what she called the banality of evil, the capacity of thoughtless functionaries to commit monstrous crimes (d'Entrèves 2026). She is also one of the three women in the panel.

Appendix 1.2 Christianity

Thomas Aquinas (c. 1225–1274) is the central figure of medieval Scholasticism. His *Summa Theologiae* and *Summa contra Gentiles* systematized Christian doctrine with the tools of Aristotelian philosophy, and his accounts of natural law and virtue remain foundational for Catholic moral thought (Pasnau 2026). His view of the created order, including the place of women within it, makes him the expected-lower pole in one of the gender contrasts.

Martin Luther (1483–1546) began the Protestant Reformation. His theology of justification by faith and of the authority of scripture over the institutional church reshaped Western Christianity, and his ideas carry philosophical weight comparable to that of Augustine and Aquinas (Stern and Leppin 2026). His writings on marriage and the household, such as *The Estate of Marriage* in our corpus, defend a gendered order of family life, which places him on the expected-lower side of the second gender contrast.

John Wesley (1703–1791), an Anglican priest, led the evangelical revival from which Methodism grew (The Editors of Encyclopædia Britannica 2024b). His Arminian theology stressed free will and the pursuit of holiness, and his open-air preaching and small-group discipline carried religion to working people in Britain and the American colonies. He also wrote against slavery and on the duties that wealth imposes, and his writings are among the panel's few eighteenth-century texts on social reform.

Rosemary Radford Ruether (1936–2022) was a Catholic theologian whose *Sexism and God-Talk* (1983) became a founding text of feminist theology. Her work criticized the patriarchal structures of church teaching and joined feminist critique with liberation theology and ecological ethics; colleagues credited her with changing theology itself, not only feminist theology (Howard University School of Divinity 2022). She is the expected-higher pole in the two gender contrasts.

Appendix 1.3 Islam

Abū Ḥāmid al-Ghazālī (1058–1111) mastered jurisprudence, theology, philosophy and Sufism, and later tradition gave him the title Ḥujjat al-Islām, the Proof of Islam (Paravizian 2026). His *Incoherence of the Philosophers* attacked the Aristotelian philosophers of his day, and his *Revival of the Religious Sciences* is credited with making Sufism an accepted part of Sunni orthodoxy (Watt 2026; Griffel 2026).

Taqī al-Dīn Ibn Taymiyya (1263–1328) was a Hanbali jurist and theologian in Damascus who called for a return to the Qur'an and the practice of the early community and opposed what he regarded as unwarranted innovations. His polemics against the religious and political authorities of his time led to repeated imprisonment. His writings later inspired the Wahhābī reform movement and remain a reference point for modern revivalists (Laoust 2024). His strict scripturalism places him on the expected-lower side of the gender contrast with Mernissi.

Ali Shariati (1933–1977), an Iranian sociologist trained in Tehran and Paris, combined Shi'i ideas of martyrdom and justice with anticolonial and Marxian social theory. His lectures shaped the generation that made the revolution of 1979, and he is often described as its ideologue (Abrahamian 1982). Among the Muslim scholars he is the only Shi'i and the only revolutionary.

Fatema Mernissi (1940–2015), a Moroccan sociologist, argued in *Beyond the Veil* (1975) and *The Veil and the Male Elite* (1991) that the subordination of women in Muslim societies rests on

male-dominated readings of scripture and history rather than on Islam itself. Her work helped to launch the wave of Islamic feminism of the late twentieth century (miriam cooke 2015). She is the expected-higher pole in the gender contrast with Ibn Taymiyya.

Appendix 1.4 Buddhism

Dōgen (1200–1253) founded the Sōtō school of Zen in Japan after training in China. His *Shōbōgenzō* treats time, being and practice with great philosophical subtlety, and his teaching that practice and awakening are one made seated meditation the centre of practice (The Editors of Encyclopædia Britannica 2024a; Nagatomo 2024). His work anchors the monastic and contemplative strand of East Asian Mahayana in the panel.

Hakuin Ekaku (1686–1769) revived the Rinzai school of Japanese Zen and systematized its curriculum of kōan practice, which Rinzai training still follows (Nagatomo 2024). Besides formal treatises he wrote works of practical advice, among them *Orategama* and *Yasen kanna*, and left a large body of painting and calligraphy, carrying Zen teaching beyond the monastery.

Tsongkhapa (1357–1419) founded the Gelug school of Tibetan Buddhism, from which the line of the Dalai Lamas descends. His *Great Treatise on the Stages of the Path* and its companion, the *Great Exposition of Tantra*, integrate Indian Madhyamaka philosophy with monastic discipline and remain foundational texts of the Gelug tradition (Sparham and Ehm 2024). As a scholastic master working entirely within a traditional monastic and scriptural framework, he is the expected-lower pole in the science contrast with the 14th Dalai Lama.

Ledi Sayadaw (1846–1923) was a Burmese Theravāda monk who wrote widely on Abhidhamma philosophy and meditation under British colonial rule. By teaching insight meditation to lay people rather than to monks alone, he began a lineage that passed through Saya Thetgyi and U Ba Khin to S. N. Goenka and the modern vipassanā movement (Braun 2013). He is the panel's Theravāda scholar and its witness to Buddhist reform under colonial rule.

Yin Shun (1906–2005) was a Chinese monk and scholar who returned to Indian sources, especially the early Āgamas and Madhyamaka, to build what he called a Buddhism for the human world. His writings, which combine critical historical method with traditional commentary, shaped humanistic Buddhism in Taiwan, notably through his student Cheng Yen, founder of the Tzu Chi Foundation (Bingenheimer 2009). His writings stand for the reformist and socially engaged strand of modern Chinese Buddhism.

Tenzin Gyatso, the 14th Dalai Lama (born 1935), is the only living scholar in the panel. A leading Tibetan Buddhist teacher in the Gelug tradition and, until 2011, Tibet's political leader in exile in Dharamsala, he received the Nobel Peace Prize in 1989; the formal head of the Gelug school is the Ganden Tripa (Lopez 2024; Namgyal Monastery Institute of Buddhist Studies 2026). His books and public teachings address ethics, compassion and consciousness for secular as well as Buddhist audiences, and his long dialogue with neuroscientists and physicists makes him the expected-higher pole in the science contrast with Tsongkhapa.

Appendix 1.5 Hinduism

Adi Shankara (conventionally c. 700–750, although other datings place him later) systematized Advaita Vedānta in commentaries on the *Brahma Sūtras*, the principal *Upaniṣads* and the *Bhagavad Gītā*. He taught that the self and ultimate reality (*Brahman*) are identical and that the world of multiplicity is an appearance, and tradition credits him with founding monastic centres (*maṭhas*) that remain active, an attribution that rests on hagiographies written centuries later (Dalal 2021). His emphasis on knowledge and renunciation as the path to liberation, rather than on action in the world, makes him an expected-lower pole in the engagement contrasts.

Rāmānuja (c. 1017–1137) developed Viśiṣṭādvaita, qualified non-dualism, in argument with Shankara's Advaita; the traditional dates are uncertain, and recent scholarship places him somewhat

later (Ranganathan 2026). For Rāmānuja the self and the world are real and form the body of a personal God, and devotion and surrender to that God are the highest path. His theism and his emphasis on devotion make him Shankara's counterpart within Vedānta.

Swami Vivekananda (1863–1902), the leading disciple of Ramakrishna, presented Vedānta to Western audiences at the World's Parliament of Religions in Chicago in 1893 (The Editors of Encyclopædia Britannica 2026). Through the Ramakrishna Mission, founded in 1897, he joined contemplative practice to organized service to the poor and argued that work for others is itself a spiritual discipline. This socially engaged reading of Vedānta makes him the expected-higher pole in both engagement contrasts.

Sri Aurobindo (1872–1950), educated at Cambridge, led the radical wing of the nationalist movement in Bengal. He began the practice of yoga in 1905, spent a year in prison after his arrest in 1908 and, after his acquittal, withdrew from politics and settled in Pondicherry in 1910 (Doniger and Huchzermeyer 2024). There he developed Integral Yoga, described in *The Life Divine* and *The Synthesis of Yoga*, which treats cosmic evolution as a progressive manifestation of the divine. His world-affirming philosophy sets him apart from the renunciant strands of Hindu thought.

Nisargadatta Maharaj (1897–1981), a Mumbai shopkeeper in the Navnath lineage, taught a spare form of Advaita in dialogues with visitors from around the world, collected in *I Am That* (Nisargadatta Maharaj 1973; Sounds True 2026). He rejected ritual and metaphysical elaboration in favour of direct attention to the nature of awareness. His inward teaching, which treats the person and the world as appearances in consciousness, makes him the second expected-lower pole in the engagement contrasts.

Appendix 1.6 Expected directions of the known-group contrasts

The contrasts test the direction of a difference on a small set of WVS items, so the expected direction must be read off positions the scholars actually took, not off their era. The three gender contrasts use five items on whether children suffer when a mother works for pay (Q28), whether men make better political leaders (Q29), whether university education matters more for a boy (Q30), whether men should have more right to a job when jobs are scarce (Q33) and whether a wife's higher earnings cause problems (Q35). Mernissi argued that the subordination of women rests on male readings of scripture and history (miriam cooke 2015), and Ruether's *Sexism and God-Talk* is a sustained critique of patriarchal theology (Howard University School of Divinity 2022). On the other side, Ibn Taymiyya worked within the rules of classical Hanbali law on the legal standing of women, Aquinas held that woman is by nature subject to man (Thomas Aquinas 1947, I, q. 92, a. 1, ad 2), and Luther's *The Estate of Marriage* defends a household ordered by the husband (Luther 1962). None of the five items was a question these authors faced in modern form, and the contrasts ask only whether the more egalitarian author scores higher.

The contrast between Spinoza and Maimonides uses the importance of religion (Q6) and of God (Q164, reversed), whether religion is always right when it conflicts with science (Q169) and whether one's own religion is the only acceptable one (Q170). Spinoza's *Theological-Political Treatise* argues for freedom of thought and against the authority of religious leaders in the state (Nadler 2026), while Maimonides's code makes revealed law the framework of communal life (Seeskin 2026). The Vedanta contrasts use the importance of work (Q5), whether work is a duty towards society (Q40) and whether work should come first (Q41), all reversed so that higher scores mean more weight on work. Vivekananda presented work for others as a spiritual discipline and founded the Ramakrishna Mission to organize it (The Editors of Encyclopædia Britannica 2026); Shankara taught that liberation comes through knowledge and renunciation rather than action (Dalal 2021), and Nisargadatta's teaching directs attention away from the world altogether. These three items are only a rough proxy for social engagement, and the weakest of our domains.

The science contrast uses four items on the benefits of science and technology (Q158, Q159,

Q163 and Q160, the last reversed) and the item on conflicts between science and religion (Q169). Here the expectation rests mainly on one side. The 14th Dalai Lama has written that Buddhism must abandon traditional claims that scientific analysis conclusively shows to be false (Dalai Lama 2005, 3) and has held a long series of dialogues with scientists (Lopez 2024). Tsongkhapa gave reasoning a high place in his philosophy (Sparham and Ehm 2024), but he wrote before modern science existed and within a traditional cosmology. The contrast therefore asks whether personas register an explicit modern endorsement of science, not whether Tsongkhapa opposed it, and it should be read as a weak test of fidelity.

Appendix 2. Corpus inventory

Table 7 lists the titles in the project's corpus inventory, which records the source volumes collected for each scholar. All are English texts or translations. A title is counted once whether it is a single work or a collected edition, so the counts are numbers of source volumes rather than of distinct works. Superscript letters mark titles that are not the scholar's own writing (a), that are the same work as another listed title under a different title or translation (b), that are collected or complete editions which may contain other listed titles (c), that were co-authored or narrated by another author (d), or whose traditional attribution to the scholar is doubted by modern scholarship (e). For the Abrahamic scholars the inventory is the only record available to us. For the Buddhist and Hindu scholars, the files that were actually cleaned and indexed are listed in Table 8; they do not match the inventory for every scholar, and for Rāmānuja none of the inventory titles is among them.

Table 7. Titles in the corpus inventory

Scholar	No.	Titles
Judaism		
Moses Maimonides	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	7	A Short Treatise on God, Man and His Well-Being; The Principles of Cartesian Philosophy; Tractatus Theologico-Politicus; Epistolae; Spinoza Complete Works[c]; The Collected Works of Spinoza[c]; The Vatican Manuscript of Spinoza's Ethica
Martin Buber	4	I and You [I and Thou]; Eclipses of God; Two Types of Faith; The Writings of Martin Buber[c]
Hannah Arendt	7	Human Condition; Origins of Totalitarianism; Between Past and Future; On Revolution; On Violence; Crises of the Republic; Love and St Augustine
Christianity		
Thomas Aquinas	6	Of God and His Creatures; Summa Theologica; Disputed Questions on Truth; Catena Aurea; Selected Writings; St Thomas on Politics and Ethics
Martin Luther	19	A manual of the book of Psalms; A Treatise on Good Works; An Open Letter to The Christian Nobility; Disputation on the power and efficacy of indulgences; The Freedom of a Christian; The Babylonian Captivity of the Church; The Bondage of the Will; A Commentary on St. Paul's Epistle to the Galatians; Whether Soldiers, Too, Can Be Saved; A Treatise on Christian Liberty[b]; Luther's Commentary on Genesis; The Estate of Marriage; The Adoration of the Sacrament; The Table Talk or Familiar Discourse of Martin Luther; The Noble Lesson[a]; Commentary on the Sermon on the Mount; Commentary on St. John's Gospel; Commentary on St. Peter & St. Jude; The Works of Martin Luther[c]
John Wesley	8	Advice to a People Called Methodists; A Plain Account of Christian Perfection; A Word to a Smuggler; Characteristics of a Methodist; Sermons on Several Occasions; Selections from the Writings of John Wesley[c]; Thoughts Upon Slavery; Wesley's 52 Standard Sermons
Rosemary Radford Ruether	5	Sexism and God-Talk; Gaia and God; Faith and Fratricide; To Change the World; Women Healing Earth
Islam		

Scholar	No.	Titles
Al-Ghazālī	7	The Alchemy of Happiness; The Incoherence of the Philosophers; The Niche of Lights; The Book of Counsel for Kings; The Aims of the Philosophers; The Book of Knowledge; The Mysteries of Almsgiving
Ibn Taymiyya	10	Essays on the Creed; Enjoining Right and Forbidding Wrong; Public Duties in Islam; Ibn Taymiyya Against the Greek Logicians; The Criterion Between the Allies of the Merciful and the Allies of the Devil; The Difference Between the Advice, Enjoining Good and Forbidding Evil; The Upright Way; The Goodly Word; The Removal of Blame; Political Thought of Ibn Taymiyyah[a]
Ali Shariati	5	The Hajj; Marxism and Other Western Fallacies; Religion vs Religion; Shariati on Shariati and the Muslim Woman; The Sociology of Islam
Fatema Mernissi	6	Beyond the Veil; Islam and Democracy; Islam and Human Rights; Muslim Women and Sexuality; The Veil and the Male Elite; Women and Islam[b]
Buddhism		
Dōgen	9	Moon in a Dewdrop; Rational Zen; Shobogenzo; Shobogenzo Zuimonki; A Primer of Sōtō Zen; Dōgen's Pure Standards for the Zen Community; Dōgen's Extensive Record; Enlightenment Unfolds; Treasury of the True Dharma Eye
Hakuin Ekaku	8	Essential Teachings of Zen Master Hakuin; Hakuin on Kensho; Orategama; Sound of One Hand; The Religious Art of Zen Master Hakuin; The Zen Master Hakuin; Wild Ivy; Zen Words for the Heart
Tsongkhapa	6	The Great Treatise on the Stages of the Path to Enlightenment; Ocean of Eloquence; In Praise of Dependent Origination; Essence of Eloquence; The Great Treatise on the Stages of Mantra; Tsong-kha-pa's Final Exposition of Wisdom
Ledi Sayadaw	8	A Manual of the Excellent Man; Manual of Buddhism; The Manuals of Buddhism; Bodhipakkhiya Dipani; Magganga Dipani; Niyama Dipani; Patthana Dipani; The Requisites of Enlightenment
Yin Shun	7	Chinese Madhyamaka; For a World in Crisis; Introducing Venerable Yinshun to the West[a]; Synopsis on the Practice of Wisdom in Buddhism (Part 1); Synopsis on the Practice of Wisdom in Buddhism (Part 2); The Three Essentials for Learning the Buddha's Teachings; Zhongguan Jinlun
The 14th Dalai Lama	9	An Open Heart; Beyond Religion; Ethics for the New Millennium; How to Practice; Kindness, Clarity, and Insight; My Spiritual Journey; The Art of Happiness[d]; The Universe in a Single Atom; Toward a True Kinship of Faiths
Hinduism		
Adi Shankara	8	Aparokshanubhuti[e]; Atma Bodha[e]; Bhaja Govindam[e]; Commentary on the Bhagavad Gita; Commentary on the Brahma Sutras; Vivekachudamani[e]; The Crest-Jewel of Wisdom[b,e]; Upadesa Sahasri
Rāmānuja	5	Bhagavad Gita Bhasya; Vedartha Sangraha; Sri Bhasya; Vedanta Sara; Vedanta Dipa
Swami Vivekananda	10	Addresses on Bhakti Yoga; Addresses on Jnana Yoga; Addresses on Karma Yoga; Addresses on Practical Vedanta; Complete Works of Swami Vivekananda[c]; Lectures from Colombo to Almora; Lectures and Discourses; Raja-Yoga; The East and the West; The Yoga Aphorisms of Patanjali
Sri Aurobindo	7	Essays on the Gita; Integral Yoga; Savitri; The Human Cycle; The Ideal of Human Unity; The Life Divine; The Synthesis of Yoga
Nisargadatta Maharaj	5	Consciousness and the Absolute; I Am That; Nectar of Immortality; Prior to Consciousness; Seeds of Consciousness

In Table 8, authorship is classified from the opening pages of each file, words are whitespace-separated tokens and chunks are the retrieval chunks of 1,200 characters with a 200-character overlap.

Table 8. Cleaned Buddhist and Hindu corpus files

File (title)	Authorship	Words	Chunks
Dōgen			
Shobogenzo vol. 3	by the scholar	92,752	523
Shinji Shobogenzo	by the scholar	91,686	515
Zen is Eternal Life	not by the scholar (study)	90,438	502
Shobogenzo vol. 1	by the scholar	90,196	503

File (title)	Authorship	Words	Chunks
Shobogenzo vol. 4	by the scholar	89,022	499
Eihei Koroku (Dogen's Extensive Record)	by the scholar	54,844	314
Shobogenzo vol. 2	by the scholar	48,247	270
Gakudo Yojinshu	by the scholar	6,225	36
Tenzo Kyokun	by the scholar	5,263	29
Fukanzazengi	by the scholar	1,064	7
Hakuin Ekaku			
Poison Blossoms from a Thicket of Thorn (Keiso dokuzui)	by the scholar	338,493	1,922
Hakuin's Precious Mirror Cave	by the scholar	113,478	658
The Essential Teachings of Zen Master Hakuin	by the scholar	40,036	228
Orategama	by the scholar	14,709	83
Zen Words for the Heart	by the scholar	12,496	73
Yasenkanna	by the scholar	12,013	68
Tsongkhapa			
Ocean of Reasoning	by the scholar	202,849	1,235
Golden Garland of Eloquence vol. 1	by the scholar	200,761	1,229
The Great Treatise on the Stages of the Path to Enlightenment	by the scholar	105,103	631
The Great Exposition of Secret Mantra (excerpt)	by the scholar	5,538	34
Splendor of an Autumn Moon (excerpt)	by the scholar	1,494	9
Ledi Sayadaw			
The Manuals of Buddhism	by the scholar	226,426	1,383
The Manual of Dhamma	by the scholar	168,737	1,044
Manual of Light	by the scholar	49,573	305
A Manual of the Excellent Man	by the scholar	45,366	275
The Requisites of Enlightenment (Bodhipakkhiya Dipani)	by the scholar	29,935	190
Manual of Insight (Vipassana Dipani)	by the scholar	21,520	131
The Noble Eightfold Path and its Factors Explained (Magganga Dipani)	by the scholar	18,497	109
Manual of Mindfulness of Breathing	by the scholar	10,535	68
Yin Shun			
Introducing Venerable Yinshun to the West	not by the scholar (study)	73,347	438
Repenting for Karmic Hindrances	by the scholar	18,198	112
Synopsis on the Practice of Wisdom in Buddhism (Part 2)	by the scholar	10,934	70
Ananda's Faults	by the scholar	10,485	65
Synopsis on the Practice of Wisdom in Buddhism (Part 1)	by the scholar	10,431	66
The Three Essentials for Learning the Buddha's Teachings	by the scholar	9,591	62
The Buddhist Perspective of Wealth	by the scholar	8,509	50
Discourse on Practicing and Studying the Buddha-dharma	by the scholar	8,146	51
The Development of the Bodhi Vow	by the scholar	8,144	50
Recollecting Buddhas	by the scholar	7,698	48
The Ultimate Meaning of Learning the Buddha-dharma	by the scholar	7,353	44
Do Not Misunderstand Buddhism	by the scholar	7,140	44
Benefiting Others and Oneself	by the scholar	6,263	40
Virtuous Practices of Lay Buddhists	by the scholar	5,519	34
The Core Tenet of the Buddha-dharma is Great Compassion	by the scholar	5,457	34
Confidence and Its Cultivation	by the scholar	5,283	33
How to Universally Apply Buddhism in Daily Life	by the scholar	4,399	28
An Explanation on the Uniqueness of the Law of Cause	by the scholar	3,371	20
The Significance of Life and Death	by the scholar	3,028	18
Attaining No Physical Suffering from No Psychological Suffering	by the scholar	1,937	12

File (title)	Authorship	Words	Chunks
The 14th Dalai Lama			
Destructive Emotions	co-authored or narrated	153,624	924
The Art of Happiness	co-authored or narrated	87,348	517
The Book of Joy	co-authored or narrated	87,220	493
Ancient Wisdom; Modern World (Ethics for the New Millennium)	by the scholar	60,627	354
teachings on Shantideva's Guide to the Bodhisattva's Way of Life	by the scholar	49,917	303
Illuminating the Path to Enlightenment	by the scholar	49,179	303
Stages of Meditation	by the scholar	38,200	228
Opening the Eye of New Awareness	by the scholar	28,842	178
Eight Verses for Training the Mind	by the scholar	7,352	44
Essence of the Heart Sutra	by the scholar	3,595	22
Five Point Peace Plan for Tibet	by the scholar	2,620	16
Adi Shankara			
Atma Bodha (Self-Knowledge)	by the scholar; long translator's introduction	61,332	361
Commentary on the Bhagavad Gita	by the scholar	35,471	199
Upadesa Sahasri	by the scholar	21,714	121
Rāmānuja			
The Theory of Knowledge (on Ramanuja's epistemology)	not by the scholar (study)	83,052	495
Anjali (magazine issue April 2007)	not by the scholar (unrelated)	7,969	46
Swami Vivekananda			
The Complete Works of Swami Vivekananda	by the scholar	1,496,850	8,135
Sri Aurobindo			
Record of Yoga	by the scholar	441,646	2,618
The Synthesis of Yoga	by the scholar	344,987	1,980
Letters on Himself and the Ashram	by the scholar	271,378	1,476
Letters on Yoga IV	by the scholar	257,122	1,413
The Human Cycle	by the scholar	249,608	1,460
Letters on Poetry and Art	by the scholar	239,153	1,324
Isha Upanishad	by the scholar	212,836	1,225
The Mother	by the scholar	210,935	1,134
Hymns to the Mystic Fire	by the scholar	204,625	1,106
The Secret of the Veda	by the scholar	193,546	1,092
Essays in Philosophy and Yoga	by the scholar	193,286	1,102
Savitri	by the scholar	190,640	1,081
The Renaissance in India	by the scholar	168,830	987
Vedic and Philological Studies	by the scholar	158,260	891
Karmayogin	by the scholar	152,415	884
Essays Divine and Human	by the scholar	147,615	843
The Future Poetry	by the scholar	142,366	808
Kena and Other Upanishads	by the scholar	121,943	692
Nisargadatta Maharaj			
I Am That	recorded talks, edited	184,832	994
Meditations with Nisargadatta Maharaj	recorded talks, edited	91,203	510
Seeds of Consciousness	recorded talks, edited	62,882	340
Pointers from Nisargadatta Maharaj	co-authored or narrated	60,208	350
Prior to Consciousness	recorded talks, edited	59,864	330
The Nectar of Immortality	recorded talks, edited	43,214	243
The Experience of Nothingness	recorded talks, edited	35,112	194
Consciousness and the Absolute	recorded talks, edited	31,799	177
Nisargadatta Gita	recorded talks, edited	28,690	155
Gleanings from Nisargadatta	recorded talks, edited	27,696	152
Self Knowledge and Self Realisation	by the scholar	9,184	49
Be Free (sayings)	recorded talks, edited	1,871	10

Appendix 3. Items and item classes

Table 9 and Figure 13 show how the 290 WVS items divide into modules and classes. We classed as value items those that ask for a judgment of what is good, right, important or desirable, including the postmaterialism and science items and the political-culture items on preferred forms of government. Perception items ask about conditions in the respondent's country or neighbourhood, behaviour items about the respondent's own conduct, identity items about national and local belonging, and demographic items about the respondent's characteristics. In Pew Set B, the 21 perception items are those on how common discrimination and conflict between groups are (Q1, Q3 and Q11 to Q14), on how well the country performs on equality, tolerance, work and the law (Q17 to Q23), and on other conditions in the respondent's country (Q2, Q4, Q5, Q41, Q46, Q75, Q83 and Q86). The items on the causes of economic inequality (Q78 to Q82) ask for a judgment and are value items. Items about the respondent's own religious practice (Q60 to Q63 and Q73) are behaviour, and two items that repeat another item in a different format (Q16 and Q47) are left out. The six WVS postmaterialism items (Q152 to Q157) ask for first and second choices among unordered goals and are not scaled. Tables 10 and 11 list the two Pew questionnaires. Set B was used in every RAG run and in every run of the regional models; Set A was used only in the BASE runs of GPT-5.4 mini (87 items) and DeepSeek-V3.2 (90 items).

Table 9. WVS Wave 7 modules

Module	Range	Items	Value items
Social values, attitudes and stereotypes	Q1–Q45	45	45
Happiness and well-being	Q46–Q56	11	0
Social capital, trust and membership	Q57–Q105	49	7
Economic values	Q106–Q111	6	6
Perceptions of corruption	Q112–Q120	9	0
Migration	Q121–Q130	10	10
Security	Q131–Q151	21	0
Postmaterialist index	Q152–Q157	6	6
Science and technology	Q158–Q163	6	6
Religious values	Q164–Q175	12	10
Ethical values and norms	Q176–Q198	23	23
Political interest and participation	Q199–Q234	36	0
Political culture and regimes	Q235–Q259	25	16
Demography	Q260–Q290	31	0

Table 10. Pew Set B items

Code	Question	Class
Q10_GOV_DIRECT_DEMO	Direct democracy (citizens vote directly): good or bad way of governing?	value
Q11_CONFLICT_RELIGION	How strong are conflicts between people of different religions?	perception
Q12_CONFLICT_RACE	How strong are conflicts between people of different ethnic/racial backgrounds?	perception
Q13_CONFLICT_URBAN_RURAL	How strong are conflicts between urban and rural residents?	perception
Q14_CONFLICT_POLITICAL	How strong are conflicts between supporters of different political parties?	perception
Q15_DIVERSITY_4PT	Does diversity of religions/ethnicities/cultures make the country better or worse?	value

Code	Question	Class
Q16_DIVERSITY_3PT	Does diversity of religions/ethnicities/cultures make the country better or worse?	duplicate
Q17_PERF_GENDER_EQ	Is country doing well ensuring men and women have equal rights?	perception
Q18_PERF_EQUAL_CHANCE	Is country doing well ensuring everyone has an equal chance of success?	perception
Q19_PERF_TOLERANCE	Is country doing well being tolerant of different races and religions?	perception
Q1_DISCRIM_RELIGION	Is discrimination against people based on their religion a problem in this country?	perception
Q20_PERF_IMMIGRANTS	Is country doing well taking in immigrants and refugees?	perception
Q21_PERF_ELDERS	Is country doing well respecting elders?	perception
Q22_PERF_WORK	Is country doing well in working hard?	perception
Q23_PERF_LAW	Is country doing well in respecting its laws?	perception
Q24_VAL_GENDER_EQ	Is ensuring men and women have equal rights an essential value?	value
Q25_VAL_EQUAL_CHANCE	Is ensuring everyone has an equal chance of success an essential value?	value
Q26_VAL_TOLERANCE	Is being tolerant of different races and religions an essential value?	value
Q27_VAL_IMMIGRANTS	Is taking in immigrants and refugees an essential value?	value
Q28_VAL_ELDERS	Is respecting elders an essential value?	value
Q29_VAL_WORK	Is working hard an essential value?	value
Q2_UNEQUAL_GENDER	Is unequal rights for men and women a problem in this country?	perception
Q30_VAL_LAW	Is respecting the laws of this country an essential value?	value
Q31_CHILD_GAY	How would you feel if your child came out as gay or lesbian?	value
Q32_CHILD_INTERRACIAL	How would you feel if your child married someone of a different race/ethnicity?	value
Q33_CHILD_INTERFAITH	How would you feel if your child married someone with different religious beliefs?	value
Q34_CHILD_MOVE	How would you feel if your child moved far from home for a job?	value
Q35_CHILD_POLICE	How would you feel if your child became a police officer?	value
Q36_CHILD_TEACHER	How would you feel if your child became a schoolteacher?	value
Q37_AGE_MARRY	What is the best age to get married?	value
Q38_AGE_FIRSTCHILD	What is the best age to have your first child?	value
Q39_AGE_BUY_HOME	What is the best age to buy a home?	value
Q3_DISCRIM_RACE	Is discrimination based on race/ethnicity a problem in this country?	perception
Q40_AGE_RETIRE	What is the best age to retire?	value
Q41_CHINA_ECON_INFLUENCE	How much influence does China have on economic conditions in this country?	perception
Q42_OPINION_INDIA	Opinion of India?	value
Q43_OPINION_RUSSIA	Opinion of Russia?	value
Q44_OPINION_CHINA	Opinion of China?	value
Q45_US_DEMOCRACY	Is democracy in the United States a good example for other countries?	value
Q46_RELIGIOUS_TEXT_ACTUAL	How much influence does the dominant religious text have on laws TODAY?	perception
Q47_RELIGIOUS_TEXT_COMPOSITE	How much influence should the dominant religious text have on laws?	duplicate
Q48_RELIGIOUS_TEXT_SHOULD	How much influence should the dominant religious text have on laws?	value
Q49_LEADER_STANDS_RELIGION	How important is it for the leader to stand up for your religious beliefs?	value
Q4_FREEDOM_SPEECH	How free are people to say what they want in this country?	perception
Q50_LEADER_RELIGIOUS	How important is it for the leader to have strong religious beliefs (different from yours)?	value
Q51_LEADER_SAME_RELIGION	How important is it for the leader to share your religious beliefs?	value
Q52_NATIONAL_LANGUAGE	How important is speaking the national language for being truly Pakistani?	identity
Q53_NATIONAL_CUSTOMS	How important is sharing national customs/traditions for being truly Pakistani?	identity
Q54_BORN_HERE	How important is being born in this country for being truly Pakistani?	identity
Q55_RELIGION_NATIONALITY	How important is being Buddhist for being truly Chinese?	identity
Q56_RELIGION_SUPERSTITION	Does religion encourage superstitious thinking or not?	value
Q57_RELIGION_TOLERANCE	Does religion encourage tolerance or intolerance?	value
Q58_RELIGION_SOCIETY	Does religion mostly help or hurt society?	value
Q59_SPIRITUAL_BELIEF	Is there something spiritual beyond the natural world?	value
Q5_ECON_SITUATION	How would you describe the current economic situation in this country?	perception
Q60_PRACTICE_INCENSE	Do you light incense or candles for spiritual/religious reasons?	behavior

Code	Question	Class
Q61_PRACTICE_FORTUNE	Do you consult fortune tellers, horoscopes, or other ways to see the future?	behavior
Q62_PRACTICE_SYMBOLS	Do you wear religious items or symbols?	behavior
Q63_PRACTICE_FAST	Do you fast during holy times?	behavior
Q64_BELIEF_REINCARNATION	Do you believe in reincarnation?	value
Q65_BELIEF_ANCESTORS	Do you believe that spirits of ancestors can help or harm you?	value
Q66_BELIEF_MAGIC	Do you believe spells, curses, or magic can influence people's lives?	value
Q67_SPIRIT_ANIMALS	Do you believe animals can have spirits or spiritual energies?	value
Q68_SPIRIT_OBJECTS	Do you believe certain objects (crystals, jewels, stones) can have spirits?	value
Q69_SPIRIT_NATURE	Do you believe parts of nature (mountains, rivers, trees) can have spirits?	value
Q6_GOV_MILITARY	Military rule: would it be a good or bad way of governing?	value
Q70_LIFE_AFTER_DEATH	Do you believe in life after death?	value
Q71_BELIEF_GOD	Do you believe in God?	value
Q72_PEOPLE_HELPFUL	Do most people try to help others or just look out for themselves?	value
Q73_PRAYER_FREQ	How often do you pray (outside of religious services)?	behavior
Q74_RELIGION_IMPORTANCE	How important is religion in your life?	value
Q75_CHILDREN_FUTURE	When children today grow up, will they be better or worse off financially than their parents?	perception
Q76_TRUST_GENERAL	Can most people be trusted?	value
Q77_ECON_REFORM	Does the economic system need to be reformed?	value
Q78_INEQ_RICH_INFLUENCE	Does rich people's political influence lead to economic inequality?	value
Q79_INEQ_EDUCATION	Do problems with the education system lead to economic inequality?	value
Q7_GOV_EXPERTS	Expert rule (not elected officials): good or bad way of governing?	value
Q80_INEQ_DISCRIM	Does discrimination against racial/ethnic minorities lead to economic inequality?	value
Q81_INEQ_BIRTH_OPP	Does being born with more opportunities lead to economic inequality?	value
Q82_INEQ_WORK_HARD	Does people working harder than others lead to economic inequality?	value
Q83_FREEDOM_SPEECH2	How free are people to say what they want?	perception
Q84_INTERNET_FREEDOM	How important is internet access without government censorship?	value
Q85_PRESS_FREEDOM	How important is a free press without government censorship?	value
Q86_MISINFORMATION	How big a problem is made-up news and misinformation?	perception
Q87_QURAN_LAWS	Should laws follow the Quran strictly, follow Islamic values, or not be influenced by the Quran?	value
Q8_GOV_STRONGLEADER	Strong leader without parliament/courts: good or bad way of governing?	value
Q9_GOV_REP_DEMOCRACY	Elected-representative democracy: good or bad way of governing?	value

Table 11. Pew Set A items

Code	Question
A001	Is discrimination against people based on their religion a problem in this country?
A002	How common is violence against women in this country?
A003	How common is violence against children in this country?
A004	How common is terrorism in this country?
A005	How common is violent crime in this country?
A006	How common is discrimination against ethnic and racial groups in this country?
A007	How common is discrimination against religious groups in this country?
A008	How common is discrimination against women in this country?
A009	How common is violence against transgender people in this country?
A010	Which system of government would you prefer?
A011	Which statement comes closer to your view? Representative democracy is a good way to govern this country, or is not a good way to govern this country?
A012	Which statement comes closer to your view? The government should take steps to prevent people from saying things offensive to minorities, or should protect people's right to say things offensive to minorities?

Code	Question
A013	Which statement comes closer to your view? The government should take steps to restrict false information online, even if it limits freedom of information, or should protect freedom of information online, even if it means false information can be published?
A014	Which statement comes closer to your view? The government should be able to prevent the dissemination of false information, or freedom of information should be protected even if false information is disseminated?
A015	Which statement comes closer to your view? The government should censor information that is offensive to religious beliefs, or freedom of information should be protected even if it offends religious beliefs?
A016	Which statement comes closer to your view? The government should be able to prevent criticism of government policies, or freedom of information should be protected even if it includes criticism of government policies?
A017	Which statement comes closer to your view? The government should be able to prevent criticism of religion, or freedom of information should be protected even if it includes criticism of religion?
A018	Which statement comes closer to your view? The government should be able to prevent criticism of minorities, or freedom of information should be protected even if it includes criticism of minorities?
A019	Which statement comes closer to your view? People should be free to criticize political leaders, or people should not be allowed to criticize political leaders?
A020	Which statement comes closer to your view? It is more important that everyone be free to pursue their life goals without interference from the state, or it is more important that the state play an active role in society so no one is in need?
A021	Which statement comes closer to your view? Being a member of our nation means respecting our laws and institutions, being born in this country, having family background from this country, or following the dominant religion in this country?
A022	Which statement comes closer to your view? Homosexuality should be accepted by society, or homosexuality should not be accepted by society?
A023	Which statement comes closer to your view? Our people are not perfect, but our culture is superior to others, or our people are no better than others?
A024	Which statement comes closer to your view? Children today will be better off financially than their parents, or worse off financially than their parents?
A025	Which statement comes closer to your view? It is more important to have a strong economy, or it is more important to have a good democracy?
A026	Which statement comes closer to your view? Men make better political leaders than women, or women make just as good political leaders as men?
A027	Which statement comes closer to your view? It is more important to maintain law and order in this country, or it is more important to protect the rights of free speech?
A028	Which statement comes closer to your view? Most people can be trusted, or you can't be too careful in dealing with people?
A029	Which statement comes closer to your view? Religion should play a very important role, somewhat important role, not too important role, or no role in politics?
A030	Which statement comes closer to your view? The growing diversity of our country makes it a better place to live, or does not make it a better place to live?
A031	Which statement comes closer to your view? We need to accept people from different backgrounds in our country to be more competitive in the global economy, or we should preserve our way of life from too much outside influence?
A032	Which statement comes closer to your view? Democracy is preferable to any other kind of government, in some circumstances a non-democratic government can be preferable, or for someone like me it doesn't matter what kind of government we have?
A033	Which statement comes closer to your view? Elected officials care what ordinary people think, or elected officials do not care what ordinary people think?
A034	Which statement comes closer to your view? Most politicians are corrupt, or not most politicians are corrupt?
A035	Which statement comes closer to your view? The justice system treats everyone fairly, or the justice system does not treat everyone fairly?
A036	Which statement comes closer to your view? Ordinary people can influence government decisions, or ordinary people cannot influence government decisions?
A037	Which statement comes closer to your view? The media are free to report the news, or the media are not free to report the news?
A038	Which statement comes closer to your view? People are free to practice their religion, or people are not free to practice their religion?

Code	Question
A039	Which statement comes closer to your view? The courts treat people equally, or the courts do not treat people equally?
A040	Which statement comes closer to your view? Everyone has an opportunity to get ahead, or not everyone has an opportunity to get ahead?
A041	Which statement comes closer to your view? It is important for older people to live with their adult children, or it is not important?
A042	Which statement comes closer to your view? People should be able to marry someone of a different religion, or should not be able to?
A043	Which statement comes closer to your view? People should be able to marry someone of a different ethnicity, or should not be able to?
A044	Which statement comes closer to your view? Religion is very important in my life, somewhat important, not too important, or not at all important?
A045	How often do you pray?
A046	Which statement comes closer to your view? There is only one true religion, or many religions can lead to eternal life?
A047	Which statement comes closer to your view? Religion should influence laws, or religion should not influence laws?
A048	Which statement comes closer to your view? It is necessary to believe in God in order to be moral and have good values, or it is not necessary?
A049	Which statement comes closer to your view? Women should obey their husbands, or wives should have equal say?
A050	Which statement comes closer to your view? Parents should choose whom their children marry, or children should choose whom they marry?
A051	Which statement comes closer to your view? It is very important to have children in order to have a fulfilling life, or it is not very important?
A052	Which statement comes closer to your view? Society is better off if men and women have different roles, or society is just as well off if men and women have the same roles?
A053	Which statement comes closer to your view? It is important to teach boys and girls the same things, or it is more important to teach boys and girls different things?
A054	Which statement comes closer to your view? It is more important for a husband to provide for the family than to care for children, or husbands should do both equally?
A055	Which statement comes closer to your view? It is acceptable for married women to work outside the home, or it is not acceptable?
A056	Which statement comes closer to your view? Religious leaders should have no influence over political decisions, or religious leaders should influence political decisions?
A057	Which statement comes closer to your view? People of many religions can live together in peace, or people of different religions are naturally in conflict?
A058	Which statement comes closer to your view? Immigration is good because it brings in new ideas and cultures, or immigration is bad because it threatens our way of life?
A059	Which statement comes closer to your view? Our country should be open to refugees, or should not be open to refugees?
A060	Which statement comes closer to your view? International trade is good for our country, or international trade is bad for our country?
A061	Which statement comes closer to your view? Climate change is a very serious problem, or climate change is not a very serious problem?
A062	Which statement comes closer to your view? Economic inequality is a very big problem, or economic inequality is not a very big problem?
A063	Which statement comes closer to your view? Rich people get rich because they have more opportunities, or rich people get rich because they work harder?
A064	Which statement comes closer to your view? Poor people are poor because they lack opportunities, or poor people are poor because they do not work hard enough?
A065	Which statement comes closer to your view? It is the responsibility of the government to reduce the gap between the rich and the poor, or the government should not be responsible?
A066	Which statement comes closer to your view? A good life is one with freedom to choose one's own path, or a good life is one with fulfilling duties to family and society?
A067	Which statement comes closer to your view? Respect for elders is one of the most important values children should learn, or it is not one of the most important values?
A068	Which statement comes closer to your view? Individual freedom is more important than social harmony, or social harmony is more important than individual freedom?
A069	Which statement comes closer to your view? It is better to be self-reliant than to depend on others, or interdependence and mutual support are better than self-reliance alone?

Code	Question
A070	Which statement comes closer to your view? A person's first duty is to themselves, or a person's first duty is to their family and community?
A071	Which statement comes closer to your view? Tradition should be preserved even if it slows social change, or social change should be embraced even if it challenges tradition?
A072	Which statement comes closer to your view? Society should emphasize obedience and discipline, or society should emphasize independence and creativity?
A073	Which statement comes closer to your view? The state should promote a single national culture, or the state should protect cultural diversity?
A074	Which statement comes closer to your view? It is important to maintain social hierarchies, or it is more important to reduce social hierarchies and promote equality?
A075	Which statement comes closer to your view? Citizens should obey authority even when they disagree, or citizens should challenge authority when it is unjust?
A076	Which statement comes closer to your view? Education should primarily teach respect and discipline, or education should primarily teach critical thinking and self-expression?
A077	Which statement comes closer to your view? It is better to avoid conflict and maintain harmony, or it is better to confront problems openly even if harmony is disrupted?
A078	Which statement comes closer to your view? Public criticism of leaders is harmful to social order, or public criticism of leaders is necessary for accountability?
A079	Which statement comes closer to your view? The individual should sacrifice personal interests for the good of the community, or the community should protect the individual's right to pursue personal interests?
A080	Which statement comes closer to your view? Religious diversity weakens national unity, or religious diversity strengthens or does not weaken national unity?
A081	Which statement comes closer to your view? People should be judged mainly by their social roles and duties, or people should be judged mainly by their individual choices and achievements?
A082	Which statement comes closer to your view? A good society is one in which everyone knows their place and responsibilities, or a good society is one in which individuals can redefine their own roles?
A083	Which statement comes closer to your view? Loyalty to family should outweigh loyalty to abstract laws or institutions, or loyalty to laws and institutions should outweigh loyalty to family?
A084	Which statement comes closer to your view? Elder generations should have more say in family and community decisions, or younger generations should have an equal say?
A085	Which statement comes closer to your view? It is more important to preserve national traditions than to adopt global norms, or it is more important to adopt global norms even if traditions change?
A086	Which statement comes closer to your view? Social harmony should sometimes take priority over complete honesty, or complete honesty should always take priority even if harmony suffers?
A087	Which statement comes closer to your view? Leaders should be chosen for their moral character more than for popular support, or leaders should be chosen by popular support even if moral character varies?
A088	Which statement comes closer to your view? It is more important to maintain inherited cultural values than to encourage experimentation and innovation, or it is more important to encourage experimentation and innovation even if inherited values change?
A089	Which statement comes closer to your view? A stable society is more important than a highly free society, or a highly free society is more important than a highly stable society?
A090	Which statement comes closer to your view? Moral education is more important than technical education, or technical education is more important than moral education?

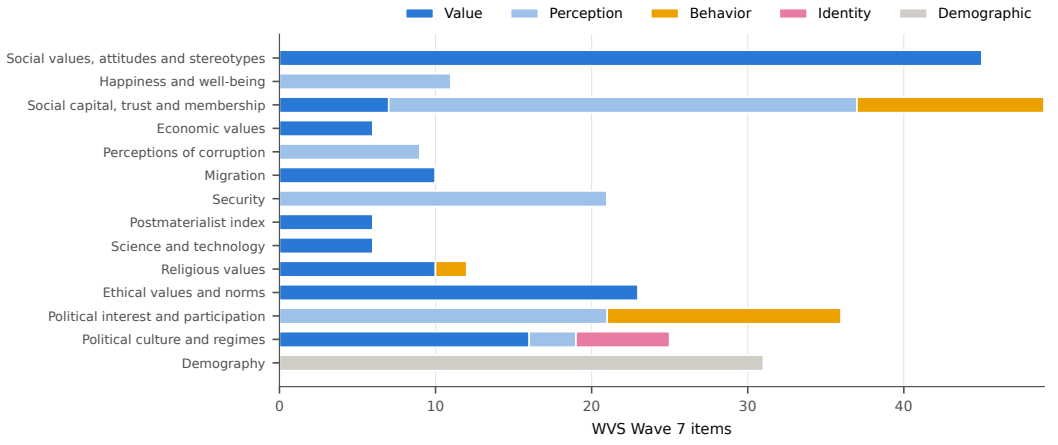


Figure 13. WVS items by module and class

Appendix 4. Prompts

The candidate lists for the Buddhist and Hindu panels were generated with the following prompt, given ten times to each of ChatGPT-5.2, Gemini Pro and DeepSeek, with the tradition inserted in brackets.

```
I'm conducting a study that reconstructs the intellectual perspectives of
historical religious scholars using large language models grounded in their
surviving primary writings. To identify suitable figures with sufficiently
large textual corpora for citation-based AI modeling, list the ten scholars
in [religion] whose preserved works are most extensive and influential,
ensuring the group collectively spans multiple centuries (ideally medieval
through modern), diverse geographic regions, distinct intellectual
traditions, and gender representation where historically available. Output
only ten TOP names. Pretend you do not recall anything from our last
conversations related to this question.
```

The survey prompts for GPT-5.4 mini and DeepSeek-V3.2 took the following form, with the tradition, the scholar's name, the question and its answer codes inserted.

```
I am a [religion] scholar named [scholar]. Please respond to questions
based solely on information extracted from my works. The response should
contain only a single numerical value and nothing else.

Question:
[question text]

Response coding:
[codes and labels]
```

The GLM-4 pipeline appended the sentence "Respond with ONLY the number. No words, no punctuation, just the integer." because the model otherwise tended to explain its answer. The Krutrim-2 pipeline used a shorter form:

```
I am a [religion] scholar named [scholar]. Based on my works, answer with
only a number.
Question: [question text]
Coding: [codes and labels]
Answer:
```

The prompts for DictaLM 2.0 and Jais 2 are documented in the pipeline scripts of the replication materials. In the RAG condition every pipeline prefixed the prompt with the retrieved passages, each labelled with its source file and chunk number, under the heading "Retrieved context". The OpenAI pipeline instead attached the scholar's hosted vector store to the request through the file-search tool, so the passages were selected and inserted by the API. In the BASE condition the same prompts were sent without any retrieved context. For the four Pew items about a respondent's own country, nationality or religion, the placeholder in the question was replaced by each country on the list associated with the persona's tradition, and the persona answered each version separately.

Appendix 5. Retrieval implementation

All retrieval pipelines share four components: a document collection, an embedding and indexing step, retrieval at query time and generation. The cleaned texts of each scholar form a separate collection. For every pipeline except OpenAI's, the whitespace-normalized text of each file was split into chunks of 1,200 characters with a 200-character overlap, and each chunk was embedded with the `all-MiniLM-L6-v2` sentence-transformer (Reimers and Gurevych 2019; Wang et al. 2020). The embeddings were L2-normalized and stored in an exact inner-product FAISS index (`IndexFlatIP`), so that similarity is the cosine of the angle between query and chunk (Johnson, Douze, and Jégou 2021; Douze et al. 2024); there is one index per scholar. At query time the survey question is embedded with the same model and the most similar chunks are retrieved, eight in the DeepSeek, DictaLM, GLM-4 and Jais pipelines and one in the archived Krutrim-2 pipeline. Their text, labelled with source file and chunk number, is placed before the prompt. The embedding model reads at most 256 word pieces. Measured with its own tokenizer, a chunk averages 277 word pieces including the two special tokens, 81% of chunks exceed the limit, and 8.1% of all word pieces fall beyond it. Truncation therefore affects which chunks are retrieved, mostly through the last sentences of a chunk, but the generating model always receives the full chunk. Retrieval parameters were left at their defaults across models to avoid tuning choices that could favour one model over another.

The OpenAI pipeline used OpenAI's hosted vector store and file-search tool, which chunk, embed and retrieve documents inside the API. Each scholar's documents were uploaded to a separate store, and every request attached that scholar's store. The number of passages retrieved and the chunking rules are set by the API rather than by us, which is the main implementation difference between the OpenAI arm and the others. The passages actually returned by the file-search tool were not logged.

The four open-weight models were served locally with `llama.cpp` (`llama-server`) from quantized GGUF files: DictaLM 2.0, GLM-4-9B-0414 and Jais 2 8B Chat at Q4_K_M and Krutrim-2 at IQ3_M. Requests were sent to the local completion endpoint with temperature zero and a cap on generated tokens (two tokens for GLM-4 and 50 for Krutrim-2 in the archived scripts). DeepSeek-V3.2 was queried through the `deepseek-chat` endpoint of the DeepSeek API and GPT-5.4 mini through the OpenAI Responses API; neither request set a temperature, so both used the providers' defaults. The pipelines were run in Python 3.12, and outputs were written to one file per model, condition, instrument and scholar, with the model name, round, question order, item identifier, country context, raw response and the number extracted by the collection script. The collection and analysis scripts are available from the corresponding author.

Appendix 6. Validity diagnostics

Table 12 reports, for every model, condition and instrument, the shares of valid answers, abstentions, invalid answers and failures among non-demographic items, the share of valid answers at the lowest and highest code, and the share of persona-item cells answered identically in all five rounds.

Jais 2 without retrieval is the clearest case of failure. All 45,295 of its BASE responses, on both instruments, were empty strings, runs of dots or prompts marked as skipped by the pipeline. With

retrieved context the same model, served in the same way, answered 98% of WVS items validly. We could not establish why the model produced nothing without context, and we exclude these cells from all analyses.

The parsing rules matter most for two models. DictaLM's raw answers carried chat-template tokens in 97% of cases (for example "[/INST]1"), and 2.4% were a number followed by an unrelated continuation; once the tokens are stripped, nearly all are clean numbers, and 1.5% remain invalid. GLM-4's invalid answers, 11.8% of its responses, were mostly strings of repeated digits (41%), decimals on ordinal scales (32%), numbers outside the item's codes (11%), ranges (8%) and lists of all options (7%). The original collection scripts had read some of these strings as numbers. GPT-5.4 mini's abstentions without retrieval were mostly the don't-know code (76% of its abstentions) rather than refusals in words.

The persona-consistency check uses two demographic items, the respondent's sex (Q260) and religious denomination (Q289). Across model-condition-scholar cells, the modal answer matched the scholar's sex in 88% of cases (94% for the three women and 87% for the men) and gave a denomination consistent with the scholar's tradition in 90%, ranging from 82% for the Jewish scholars to 98% for the Buddhist ones.

Table 12. Response validity by cell

Survey	Condition	Model	Valid	Abstain	Invalid	Failure	First option	Last option	Invariant
WVS	BASE	GPT-5.4 mini	0.80	0.20	0.00	0.00	0.29	0.23	0.90
WVS	BASE	DeepSeek-V3.2	0.99	0.00	0.00	0.00	0.14	0.15	0.21
WVS	BASE	DictaLM 2.0	1.00	0.00	0.00	0.00	0.78	0.08	0.98
WVS	BASE	GLM-4-9B-0414	0.83	0.00	0.17	0.00	0.49	0.19	1.00
WVS	BASE	Jais 2 8B Chat	0.00	0.00	0.00	1.00	–	–	–
WVS	BASE	Krutrim-2	0.99	0.00	0.00	0.00	0.54	0.22	0.99
WVS	RAG	GPT-5.4 mini	0.89	0.08	0.03	0.00	0.39	0.34	0.71
WVS	RAG	DeepSeek-V3.2	0.98	0.01	0.01	0.00	0.33	0.32	0.52
WVS	RAG	DictaLM 2.0	0.99	0.00	0.01	0.00	0.82	0.12	0.98
WVS	RAG	GLM-4-9B-0414	0.89	0.00	0.11	0.00	0.55	0.26	0.98
WVS	RAG	Jais 2 8B Chat	0.98	0.00	0.01	0.00	0.53	0.33	0.99
WVS	RAG	Krutrim-2	0.99	0.00	0.00	0.00	0.54	0.29	0.98
Pew	BASE	GPT-5.4 mini	0.79	0.21	0.00	0.00	0.48	0.49	0.90
Pew	BASE	DeepSeek-V3.2	0.98	0.02	0.00	0.00	0.18	0.69	0.56
Pew	BASE	DictaLM 2.0	1.00	0.00	0.00	0.00	0.74	0.15	0.97
Pew	BASE	GLM-4-9B-0414	0.98	0.00	0.02	0.00	0.08	0.41	1.00
Pew	BASE	Jais 2 8B Chat	0.00	0.00	0.00	1.00	–	–	–
Pew	BASE	Krutrim-2	1.00	0.00	0.00	0.00	0.13	0.46	1.00
Pew	RAG	GPT-5.4 mini	1.00	0.00	0.00	0.00	0.28	0.27	0.43
Pew	RAG	DeepSeek-V3.2	1.00	0.00	0.00	0.00	0.38	0.42	0.69
Pew	RAG	DictaLM 2.0	1.00	0.00	0.00	0.00	0.63	0.30	0.95
Pew	RAG	GLM-4-9B-0414	0.93	0.00	0.07	0.00	0.14	0.28	0.99
Pew	RAG	Jais 2 8B Chat	1.00	0.00	0.00	0.00	0.11	0.56	1.00
Pew	RAG	Krutrim-2	1.00	0.00	0.00	0.00	0.13	0.43	0.97

Note: Shares of non-demographic responses. First and last option are shares of valid answers at the lowest and highest code. Invariant is the share of scholar–item cells with the same answer in all five rounds.

Table 13. Variance components of persona answers

Panel	Model	Tradition	Scholar within tradition	Model × scholar	Items
WVS, six models, RAG	0.37 [0.33, 0.40]	0.06 [0.05, 0.08]	0.11 [0.10, 0.12]	0.46	123
WVS, five models, BASE	0.55 [0.50, 0.59]	0.05 [0.03, 0.06]	0.09 [0.08, 0.11]	0.31	114
WVS, four regional models, RAG	0.38 [0.34, 0.42]	0.06 [0.04, 0.07]	0.15 [0.14, 0.16]	0.41	121
Pew Set B, four regional models, RAG	0.38 [0.33, 0.43]	0.07 [0.04, 0.10]	0.15 [0.13, 0.17]	0.40	55
Pew Set B, three regional models, BASE	0.60 [0.52, 0.68]	0.08 [0.05, 0.12]	0.10 [0.08, 0.12]	0.22	55

Note: Mean share of item variance (η^2) with 95% intervals from 2,000 item bootstrap resamples. The model × scholar column is the remainder, which with one modal answer per cell is the interaction.

Appendix 7. Full results

Table 14 gives the tradition separation ratio and its permutation test for every model, condition and instrument. Table 15 gives the known-group differences for every model and contrast, with the WVS items that define each domain. Table 13 gives the variance components for all five panels with their intervals.

Table 14. Tradition separation tests

Survey	Cond.	Model	Ratio	p	p_{Holm}
WVS	BASE	GPT-5.4 mini	1.082	0.149	0.188
WVS	BASE	DeepSeek-V3.2	1.060	0.041	0.124
WVS	BASE	DictaLM 2.0	1.065	0.094	0.188
WVS	BASE	GLM-4-9B-0414	1.560	<0.001	<0.001
WVS	BASE	Krutrim-2	1.363	<0.001	<0.001
WVS	RAG	GPT-5.4 mini	1.470	<0.001	0.001
WVS	RAG	DeepSeek-V3.2	1.264	<0.001	0.001
WVS	RAG	DictaLM 2.0	0.975	0.687	0.687
WVS	RAG	GLM-4-9B-0414	1.149	<0.001	0.001
WVS	RAG	Jais 2 8B Chat	1.089	0.002	0.003
WVS	RAG	Krutrim-2	1.181	<0.001	0.001
Pew	BASE	DictaLM 2.0	1.402	<0.001	<0.001
Pew	BASE	GLM-4-9B-0414	1.651	<0.001	<0.001
Pew	BASE	Krutrim-2	1.562	<0.001	<0.001
Pew	RAG	DictaLM 2.0	1.221	0.004	0.004
Pew	RAG	GLM-4-9B-0414	1.150	<0.001	0.002
Pew	RAG	Jais 2 8B Chat	1.165	<0.001	<0.001
Pew	RAG	Krutrim-2	1.237	<0.001	<0.001

Note: Ratio of mean between-tradition to mean within-tradition profile distance; one-sided permutation p from 5,000 relabelings of traditions; Holm adjustment within survey and condition.

Table 15. Known-group contrasts by model

		BASE					RAG					
Contrast	Domain	GPT	DeepSeek	Dicta	GLM	Krutrim	GPT	DeepSeek	Dicta	GLM	Jais	Krutrim
Dalai Lama > Tsongkhapa	science	0.40	0.07	−0.20	0.04	0.00	0.18	−0.04	−0.20	−0.12	0.20	0.27
Mernissi > Ibn Taymiyya	gender	0.70	0.12	0.00	0.75	0.55	0.00	0.25	0.40	0.05	0.60	0.67
Ruether > Aquinas	gender	0.63	−0.05	0.00	0.00	0.12	0.00	0.27	0.00	0.00	0.20	0.22
Ruether > Luther	gender	0.70	−0.18	0.00	0.67	0.35	0.00	0.38	0.00	0.00	0.20	0.23
Spinoza > Maimonides	secular authority	0.61	0.11	0.00	−0.06	0.50	0.42	0.33	−0.17	0.08	0.00	0.75
Vivekananda > Nisargadatta	engagement	0.08	0.11	0.00	0.17	0.00	0.56	0.22	0.00	0.33	−0.11	0.00
Vivekananda > Shankara	engagement	−0.03	−0.08	0.00	0.00	0.00	0.00	−0.06	0.00	0.08	−0.33	0.00

Note: Difference in normalized domain scores, expected-higher minus expected-lower scholar (WVS). Gender: Q28, Q29, Q30, Q33, Q35; secular authority: Q6, Q164 (reversed), Q169, Q170; engagement: Q5, Q40, Q41 (all reversed); science: Q158, Q159, Q160 (reversed), Q163, Q169.

Appendix 8. Robustness

Table 16 repeats the variance decomposition for the WVS with retrieval under alternative specifications. The model share exceeds the scholar share in every variant. Using mean rather than modal answers changes neither share appreciably. Dropping the living scholar or Arendt, whose religious identity is the least theological in the panel, has no visible effect. The model share is higher among the Hindu and Buddhist scholars than among the Abrahamic ones, in line with the larger cross-model disagreement on the non-Abrahamic scholars reported in the main text. Dropping DictaLM, whose position preference makes its personas the most uniform, lowers the model share to 0.28 and raises the scholar share to 0.24, the closest the two come. Using only items with no missing cells, instead of filling up to 15% of missing cells with the model's item mean, leaves 53 items and gives a model share of 0.38 against a scholar share of 0.17.

The comparison of commercial and regional models on Pew is restricted to the Hindu and Buddhist scholars, for whom the commercial Pew RAG runs exist. On the WVS, restricting all six models to the same eleven scholars gives a model share of 0.47 and a scholar share of 0.13. The model share is higher than in the full panel, consistent with the larger cross-model disagreement on these scholars, and close to the Pew value for the same subset. The fixed question order of the commercial BASE runs could inflate their round-to-round stability. GPT-5.4 mini's BASE answers were identical across rounds for 90% of cells, against 71% in its randomized RAG runs, but DeepSeek's BASE answers were the least stable of any cell (21%), so order alone does not determine stability. Because every item was sent as a separate request, the order of presentation could not carry over from one item to the next. Using modal answers reduces the influence of individual unstable responses, but instability can still change the modal answer, and it can produce ties; ties occur in 1.7% of persona–item cells overall and in 7.2% of DeepSeek's.

Filling sparse missing cells with model means could inflate the model share in the variance decomposition. The comparison of conditions is therefore repeated on items that every persona answered in both conditions: 43 WVS items, where the model share is 0.45 without and 0.36 with retrieval, and 23 Pew items (0.44 and 0.37). The share of missing cells is small in both conditions (2.8% and 2.3% on the WVS).

Table 17 repeats the main retrieval analyses without the scholars whose corpora contain a substantial share of text not written by them (Section 5 and Appendix 2). The first restriction drops Shankara, Rāmānuja and Yin Shun, whose corpora contain a quarter or more of secondary, unrelated or editorial text; the second also drops the 14th Dalai Lama, whose corpus consists mostly of co-authored or narrated books. The separation tests use 2,000 rather than 5,000 relabelings. None of the conclusions changes.

Table 16. Robustness of the variance shares

Specification	Model	Scholar	Items
Main specification (modal answers, all scholars)	0.367	0.174	123
Mean instead of modal answers	0.388	0.173	123
Without the 14th Dalai Lama	0.373	0.172	122
Without Arendt	0.368	0.174	123
Hindu and Buddhist scholars only	0.474	0.133	121
Jewish, Christian and Muslim scholars only	0.414	0.162	118
Without DictaLM 2.0	0.277	0.236	120
Complete cases only (no fill-in)	0.381	0.170	53

Note: WVS value items, RAG condition, six models. Scholar share combines the tradition and within-tradition components.

Table 17. Retrieval results without the affected corpora

	H1 shares		H2a separation			H3	H4	H5
Scholars	Model	Scholar	WVS sig.	Pew sig.	WVS ratio	Dispersion up	Home $\hat{\gamma}$ (p_*)	More secular
All scholars	0.37	0.17	5 of 6	4 of 4	0.97–1.47	4 of 5	0.029 (0.12)	39%
Without Shankara, Rāmānuja and Yin Shun	0.37	0.18	5 of 6	4 of 4	1.00–1.54	4 of 5	0.035 (0.12)	37%
Also without the 14th Dalai Lama	0.37	0.18	6 of 6	4 of 4	1.11–1.71	4 of 5	0.037 (0.12)	36%

Note: RAG condition. H1: six models, WVS. H2a: number of models with Holm-adjusted permutation $p < 0.05$ (2,000 relabelings) and range of ratios across the six WVS models. H3: models whose WVS dispersion rises from BASE to RAG. H4: home coefficient for the tradition margin with its exact upper-tail p over 24 assignments. H5: share of model-scholar cells more secular than the reference states.