Spektrum Iran | Vol. 39 | Issue 1 | Winter & Spring 2026 | pp. 159-190 |

Original Research Paper

Semantic sovereignty in large language models: Auditing the representation of persianate and Shi'i meanings across languages

Hamed Zarabi [1*], **Aysheh Mohammadzadeh**[2]

1 PhD Candidate in Applied Linguistics, Hakim Sabzevari University, Sabzevar, Iran

2 Assistant Professor, Department of English Language Teaching, Gonbad KavoosBranch, Islamic Azad University, Gonbad Kavoos, Iran

Received: Aug., 27, 2025 **Accepted:** Dec. 7, 2025

Abstract

Large language models increasingly translate, summarize, and explain traditions they have not inherited. This study develops semantic sovereignty as a framework for testing whether multilingual models preserve Persianate and Shi'i meanings or relocate them into dominant Anglophone interpretive regimes. A bounded cross-lingual mixed-method audit examined 120 purposively selected units from Rumi, Hafez, Ferdowsi, and Shi'i hermeneutics. Four model systems, blinded during expert scoring, were tested under Persian-only, English-only, cultural-context, and expert-context conditions, with two runs per prompt, yielding 3,840 outputs. A three-member disciplinary panel applied a 0–4 Semantic Preservation Scale and a multi-label error taxonomy; mutually exclusive percentages summarize each output's adjudicated dominant error. Across 16 model-by-domain cells, the mean score was 2.53/4.00 (SD = 0.38). Rumi was frequently psychologized into global spirituality, Hafez was over-clarified, Ferdowsi was depoliticized into leadership language, and Shi'i concepts were flattened into generic religious vocabulary. Expert-context prompts produced the strongest descriptive means, but displacement persisted. Because product identities, exact access dates, complete raw transcripts, and output-level scores were not retained, the analysis is descriptive and bounded rather than inferential or product-reproducible. Multilingual AI should therefore be evaluated not only for linguistic performance or cultural bias, but for hermeneutic accountability: preserving metaphor, doctrine, ambiguity, register, authority, and civilizational memory. The study contributes a culturally grounded audit framework for identifying subtle semantic loss across languages.

Keywords: semantic sovereignty, Persianate studies, large language models, Shi'i hermeneutics, cultural AI audit

* Corresponding Author ✉ hamed.zarabi@hsu.ac.ir https://orcid.org/0009-0009-9218-9341

How to Cite this Article:
Zarrabi, H., & Mohammadzadeh, A. (2026). Semantic sovereignty in large language models: Auditing the representation of persianate and Shi'i meanings across languages. *Spektrum Iran*, 39(1), 159-190.

https://doi.org/10.22034/spektrum.2026.587146.1069

© 2026 The Author(s). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

اشپکتروم ایران | دوره ۳۹ | شماره ۱ | زمستان و بهار ۲۰۲۶ | صص ۱۹۰-۱۵۹ |

مقاله پژوهشی

حاکمیت معنایی در مدل‌های زبانی بزرگ: ممیزی معنای ایرانی‌ـاسلامی و شیعی در میان زبان‌ها

حامد ضرابی[1*]، **عایشه محمدزاده**[2]

۱ دانشجوی دکتری زبان‌شناسی کاربردی، گروه زبان‌شناسی کاربردی، دانشگاه حکیم سبزواری، سبزوار، ایران

۲ استادیار، گروه آموزش زبان انگلیسی، واحد گنبد کاووس، دانشگاه آزاد اسلامی، گنبد کاووس، ایران

دریافت: ۱۴۰۴/۰۶/۰۵؛ **پذیرش:** ۱۴۰۴/۰۹/۱۶

چکیده:

مدل‌های زبانی بزرگ سنت‌هایی را ترجمه و تبیین می‌کنند که وارث تاریخی آن‌ها نیستند. این پژوهش «حاکمیت معنایی» را چارچوبی برای سنجش حفظ معانی ایرانی‌ـاسلامی و شیعی یا انتقال آن‌ها به تفسیر مسلط انگلیسی می‌داند. یک ممیزی ترکیبی و میان‌زبانی، ۱۲۰ واحد هدفمند از مولانا، حافظ، فردوسی و هرمنوتیک شیعی را بررسی کرد. چهار سامانه ناشناس برای ارزیابان، در وضعیت‌های فارسی، انگلیسی، فرهنگی و تخصصی آزموده شدند؛ اجرای دوگانه دستورها ۳۸۴۰ خروجی ایجاد کرد. هیئتی سه‌نفره، مقیاس صفر تا چهارِ حفظ معنایی و رده‌بندی چندبرچسبی خطاها را به کار برد؛ درصدهای متقابلاً انحصاری نیز خطای غالبِ داوری‌شده هر خروجی را خلاصه می‌کنند. میانگین ۱۶ سلول مدل‌ـحوزه ۲٫۵۳ از ۴٫۰۰ بود (انحراف معیار = ۰٫۳۸). مولانا غالباً به معنویت جهانی و روان‌شناختی تقلیل یافت، حافظ بیش از حد توضیح داده شد، فردوسی در زبان رهبری مدرن سیاست‌زدایی شد و مفاهیم شیعی به واژگان عمومی دینی فروکاسته شدند. زمینه تخصصی بالاترین میانگین‌های توصیفی را ایجاد کرد، اما جابه‌جایی باقی ماند. چون هویت محصولات، تاریخ‌های دقیق دسترسی، متن کامل خروجی‌ها و امتیازهای سطح خروجی حفظ نشده‌اند، تحلیل توصیفی و محدود است و ادعای استنباط آماری یا بازتولیدپذیری محصول‌محور ندارد. بنابراین، هوش مصنوعی چندزبانه باید افزون بر عملکرد زبانی و سوگیری فرهنگی، از منظر پاسخ‌گویی هرمنوتیکی، حفظ استعاره، آموزه، ابهام، لحن، اقتدار و حافظه تمدنی، ارزیابی شود. مطالعه چارچوبی فرهنگی برای شناسایی فقدان معنا میان زبان‌ها ارائه می‌کند.

واژه‌های کلیدی: آیین‌های فرهنگی، اینستاگرام، آیین‌مندی دیجیتال، جامعه شبکه‌ای، تحلیل مضمون

* نویسنده مسئول

https://orcid.org/0009-0009-9218-9341

hamed.zarabi@hsu.ac.ir

https://doi.org/10.22034/spektrum.2026.587146.1069

Spektrum Iran | Jg. 39 | Nr. 1 | Winter und Frühling-2026 | S. 159-190 |

Original-Forschungsarbeit

Semantische souveränität in großen Sprachmodellen: Prüfung persianischer und schiitischer Bedeutungen über Sprachgrenzen hinweg

Hamed Zarabi [1*], **Aysheh Mohammadzadeh**[2]

1 Doktorand der Angewandten Linguistik, Universität Hakim Sabzevari, Sabzevar, Iran

2 Assistenzprofessorin, Abteilung für Englischdidaktik, Zweigstelle Gonbad Kavoos, Islamische Azad-Universität, Gonbad Kavoos, Iran

Eingegangen: 27. Aug. 2025 **Angenommen:** 7. Dec. 2025

Zusammenfassung:

Große Sprachmodelle übersetzen, verdichten und erläutern zunehmend Traditionen, deren historische Träger oder Erben sie nicht sind. Diese Studie entwickelt semantische Souveränität als Rahmen, um zu prüfen, ob multilinguale Modelle persianische und schiitische Bedeutungen bewahren oder in dominante anglophone Interpretationsordnungen verlagern. Ein begrenztes, sprachübergreifendes Mixed-Methods-Audit untersuchte 120 gezielt ausgewählte Einheiten aus Rumi, Hafis, Ferdowsi und der schiitischen Hermeneutik. Vier während der Expertenbewertung verblindete Modellsysteme wurden unter vier Bedingungen getestet: nur Persisch, nur Englisch, kultureller Kontext und Expertenkontext. Jeder Prompt wurde zweimal ausgeführt, wodurch 3.840 Ausgaben entstanden. Ein dreiköpfiges Fachgremium verwendete eine Skala der semantischen Bewahrung von 0 bis 4 sowie eine mehrdimensionale Fehlertaxonomie; die sich gegenseitig ausschließenden Prozentwerte fassen den jeweils adjudizierten dominanten Fehler zusammen. Über 16 Modell-Domänen-Zellen betrug der Mittelwert 2,53 von 4,00 Punkten (SD = 0,38). Rumi wurde häufig zu globaler, psychologisierter Spiritualität verengt, Hafis übermäßig vereindeutigt, Ferdowsi in moderne Führungssprache entpolitisiert und schiitische Begriffe auf allgemeines religiöses Vokabular reduziert. Expertenkontext erzeugte die höchsten deskriptiven Mittelwerte, doch semantische Verschiebung blieb bestehen. Da Produktidentitäten, genaue Zugriffsdaten, vollständige Rohprotokolle und Bewertungen auf Ausgabenebene nicht aufbewahrt wurden, ist die Analyse deskriptiv und begrenzt; sie beansprucht weder statistische Inferenz noch produktspezifische Reproduzierbarkeit. Multilinguale KI sollte daher nicht nur nach sprachlicher Leistung oder kultureller Verzerrung, sondern auch nach hermeneutischer Verantwortlichkeit beurteilt werden: nach der Bewahrung von Metapher, Doktrin, Mehrdeutigkeit, Register, Autorität und zivilisatorischem Gedächtnis. Die Studie bietet damit einen kultursensiblen Audit-Rahmen zur Erkennung subtiler Bedeutungsverluste zwischen Sprachen.

Schlüsselwörter: kulturelle rituale, Instagram, digitale ritualität, Netzwerkgesellschaft, thematische Analyse

* Korrespondierender Autor ✉ hamed.zarabi@hsu.ac.ir https://orcid.org/0009-0009-9218-9341

Wie dieser Artikel zu zitieren ist:

Zarrabi, H., & Mohammadzadeh, A. (2026). Semantic sovereignty in large language models: Auditing the representation of persianate and Shi'i meanings across languages. *Spektrum Iran*, 39(1), 159-190.

https://doi.org/10.22034/spektrum.2026.587146.1069

© 2026 The Author(s). Dieser Artikel ist unter der Creative-Commons-Lizenz Namensnennung 4.0 International (CC BY 4.0) lizenziert.

1. Introduction

Large language models increasingly act as unofficial translators of civilization, yet fluent paraphrase can conceal the relocation of culturally saturated meaning. Terms such as ʿerfān, velāyat, taʾwīl, rindī, farrah, and ʿadl are not detachable lexical units; they carry poetic, sacred, juridical, and political histories. The question is therefore not only whether LLMs translate Persian correctly, but whether they preserve the interpretive conditions through which Persianate and Shiʻi meanings remain intelligible (Bender & Koller, 2020; Tao et al., 2024).

Transformer-based systems were designed for sequence modelling rather than adjudicating inherited layers of meaning (Vaswani et al., 2017; Devlin et al., 2019). Although multilingual models show strong cross-lingual performance, training data, tokenization, alignment resources, and evaluation remain unequal (Qin et al., 2025; Xu et al., 2025). Persian is especially revealing: it is civilizationally dense but computationally uneven, and its benchmarks expose variable performance across reasoning, knowledge, and language-specific tasks (Abaskohi et al., 2024; Farahani et al., 2021). Research also places culturally unspecified LLM outputs near dominant Western or Anglophone value regions, so linguistic fluency cannot establish cultural adequacy (Durmus et al., 2023; Tao et al., 2024). Figure 1 visualizes this problem.

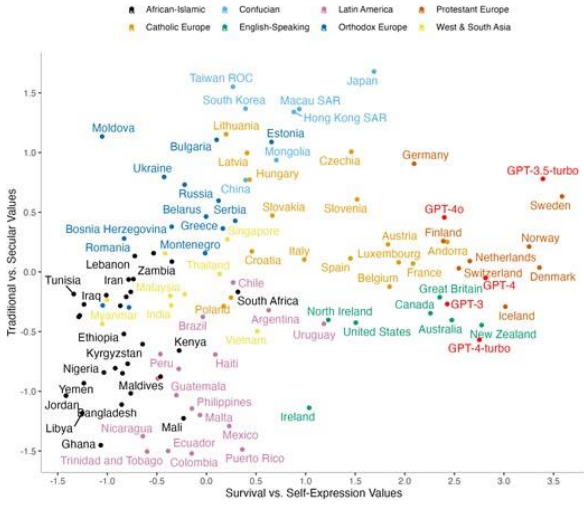

Figure 1. Mapping Cultural Bias and Semantic Alignment in Large Language Model Outputs

Opinion and value audits show that model responses privilege particular demographic and geopolitical distributions rather than a neutral human standpoint (Santurkar et al., 2023; Zhao, W., et al., 2024). For Persianate materials, this can become computational Orientalism: surface exoticism survives while doctrinal, ritual, and ethical grammars disappear (Noble, 2018; Said, 1978). The reception of Rumi already documents omission and domestication in English circulation (Azadibougar & Patton, 2015; Naghmeh-Abbaspour et al., 2021), while Shi'i concepts require attention to authorized interpretation and communal memory (Alak, 2023; Bender & Friedman, 2018). Semantic sovereignty names a tradition's claim not merely to appear in model outputs, but to remain interpretable on its own historical and normative terms.

This study audits four domains—Rumi's mystical semantics, Hafez's lyric ambiguity, Ferdowsi's epic-political vocabulary, and Shi'i hermeneutic concepts—across model and prompt conditions. It combines expert scoring with a culturally specific taxonomy of metaphor loss, doctrinal flattening, register shift, domestication, universalization, ambiguity closure, depoliticization, and authority erasure. Its contribution is to extend AI evaluation from bias and task accuracy to the preservation of civilizational meaning.

2. Review of the Related Literature

2.1. Technical Foundations, Accountability, and Evaluation

Self-attention enabled Transformers to model long-range token relations and supported BERT-style pre-training, autoregressive generation, and instruction-tuned systems (Vaswani et al., 2017; Devlin et al., 2019; Brown et al., 2020; Ouyang et al., 2022). Scale and fluency, however, do not establish grounded understanding. Bender and Koller (2020) distinguish linguistic form from meaning produced through communicative, social, embodied, and historical conditions. That distinction is decisive for Persianate and Shi'i materials, where a fluent explanation may detach ʿerfān, velāyat, taʾwīl, rindī, farrah, or shahādat from the disciplines that authorize their interpretation.

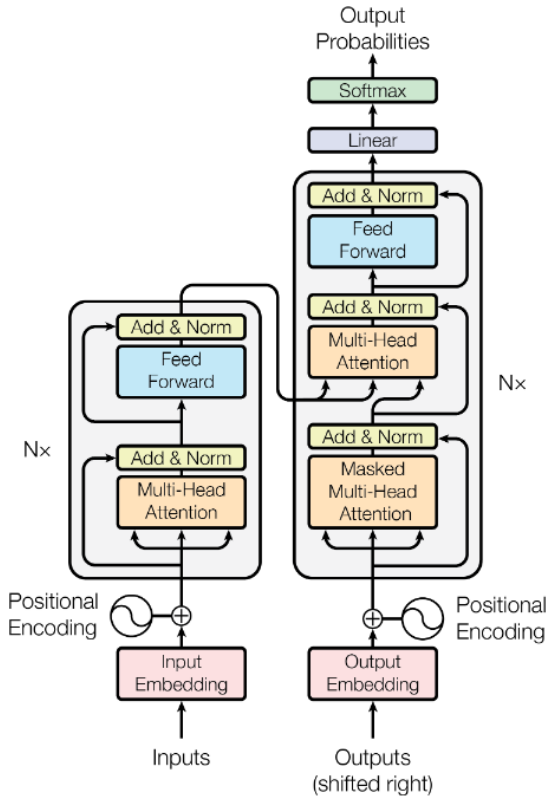

Figure 2. The Transformer Model Architecture (Vaswani et al., 2017)

Figure 2 anchors this distinction: attention-based sequence modelling does not itself supply sacred authority, poetic indeterminacy, doctrinal hierarchy, or civilizational memory.

LLMs are therefore treated here as civilizational mediators rather than neutral translation tools. They can preserve communicative surface while relocating meaning across cultural regimes, a risk that ordinary measures of grammaticality or lexical similarity do not capture (Bender & Koller, 2020).

2.2. Documentation, Bias, and Semantic Harm

Responsible-AI scholarship rejects datasets and models as neutral artifacts. Data statements, model cards, and datasheets document collection, intended use, exclusion, demographic imbalance, and

institutional power (Bender & Friedman, 2018; Mitchell et al., 2019; Gebru et al., 2021). Such documentation is necessary but insufficient: recording Persian data does not show whether outputs distinguish mystical love from therapeutic self-love, taʾwīl from generic interpretation, Hafezian ambiguity from confusion, or Ferdowsian kingship from modern leadership.

Figure 3 illustrates why assumptions acquired during pre-training may persist through fine-tuning and shape later translation, explanation, and summarization.

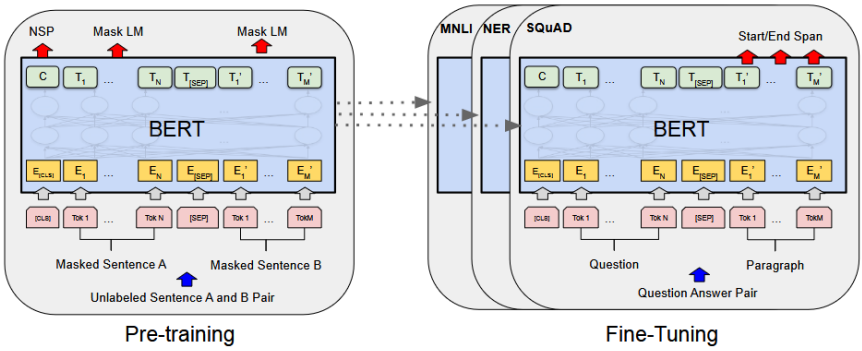

Figure 3. Overall Pre-Training and Fine-Tuning Procedures for BERT (Devlin et al., 2019)

Bias research demonstrates that computational systems reproduce asymmetries of race, gender, religion, geography, class, and language (Buolamwini & Gebru, 2018; Noble, 2018; Blodgett et al., 2020; Mehrabi et al., 2021). Persianate displacement is often harder to detect because the output may be polite, elegant, and superficially accurate. Replacing a dense poetic or juridical term with a generic modern equivalent can weaken interpretive sovereignty without producing an obvious stereotype or hallucination.

Semantic displacement is thus a distinct harm: competent linguistic mediation systematically thins the interpretive conditions of a tradition.

2.3. Evaluation Beyond Accuracy

LLM evaluation now includes robustness, calibration, fairness, reasoning, factuality, safety, and alignment (Srivastava et al., 2023; Liang et al., 2023; Chang et al., 2024), but these frameworks rarely test sacred

register, literary ambiguity, doctrinal hierarchy, or civilizational memory. A system may succeed on Persian reading comprehension yet close Hafez's ambiguity, summarize the Shāhnāmeh while erasing its political theology, present a de-Islamized Rumi, or translate Shi'i terminology without authority relations. Semantic audit therefore asks not simply whether an output matches a reference, but whether metaphor, doctrine, register, ambiguity, and cultural difference survive cross-lingual mediation.

2.4. Multilingual LLMs and Persian NLP

Multilingual support remains uneven across corpora, tokenization, alignment, benchmarks, safety testing, and downstream performance (Qin et al., 2025; Xu et al., 2025; Zhu et al., 2024). A model may generate fluent Persian while organizing its explanations through English-centered concepts. Persian exposes the weakness of a simple high-resource/low-resource distinction: historically it is a transregional literary, philosophical, bureaucratic, and religious language, yet computationally it remains unevenly represented.

Persian-specific resources such as ParsBERT and ParsiNLU address script, morphology, orthographic variation, clitics, informal registers, and reasoning tasks (Farahani et al., 2021; Khashabi et al., 2021). These foundations measure functional competence, but not whether culturally dense concepts retain their poetic, theological, or political architecture across languages.

2.5. Persian Benchmarking and Civilizational Competence

Persian LLM benchmarking finds substantial task variation and reports that English prompts can sometimes outperform Persian prompts (Abaskohi et al., 2024). This creates a central paradox: English mediation may improve procedural structure while increasing semantic domestication. Figure 4 accordingly supports a domain-specific audit rather than a global claim that a model is simply good or bad at Persian.

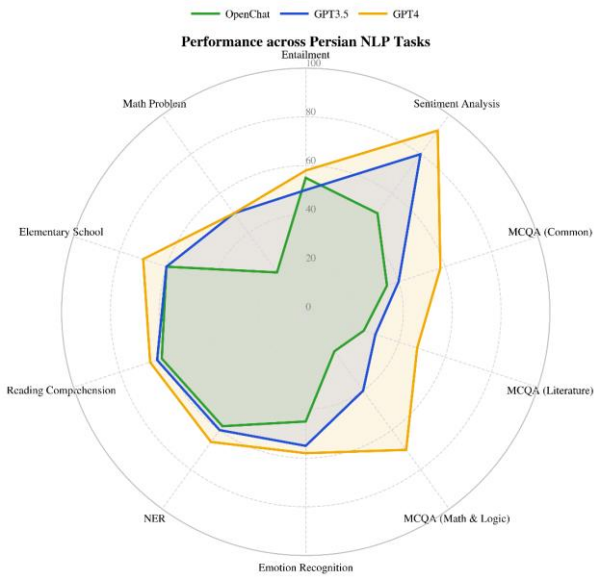

Figure 4. Performance of GPT-3.5, GPT-4, and OpenChat-3.5 across Persian NLP Tasks (Abaskohi et al., 2024)

A model may answer factual questions in Persian while failing to preserve a Hafezian metaphor, a Qur'anic allusion, or the doctrinal field of a Shi'i term. Current benchmarks are therefore a foundation for, not a substitute for, semantic accountability.

Persian should also be understood as a civilizational testbed. Across empire, court, mysticism, epic, philosophy, devotion, and transregional circulation, its archive repeatedly tests the limits of translation (Lewis, 2015; Ingenito, 2018; Kia, 2020). Semantic sovereignty does not claim that Persianate meaning is ineffable; it requires evaluation criteria appropriate to the tradition's own semantic structures.

2.6. Persianate Literature and Hermeneutic Loss

2.6.1. Rumi and Domestication

Rumi's global reception shows how translation can detach poetry from Islamic, Qur'anic, Persian, and Sufi matrices while retaining attractive themes of love and healing (Azadibougar & Patton, 2015; Naghmeh-Abbaspour et al., 2021). LLMs may automate this domesticated archive, substituting self-discovery and inner peace for prophetic inheritance, Sufi

discipline, annihilation, and subsistence. Reading Rumi as ethical and pedagogical tradition rather than inspirational content clarifies what such outputs lose (Akbari, 2025).

2.6.2. Hafez and Ambiguity

Hafez places a different pressure on models: his ghazals depend on double address, irony, allusion, and unresolved plurality. Terms such as mey, rind, sāqī, pīr-e moghān, and kharābāt move among erotic, mystical, social, and anti-hypocritical readings (Yarshater, 1988; Ingenito, 2018). Because LLM helpfulness rewards clear explanation, a confident single reading may be less faithful than an explicit preservation of uncertainty.

Technical explainability often treats clarity as a virtue (Zhao, H., et al., 2024), whereas hermeneutics understands meaning as historically situated and contested (Ricoeur, 1981; Gadamer, 2004). A Hafez audit must therefore test whether ambiguity is preserved as meaning rather than solved as noise.

2.6.3. Ferdowsi and Civilizational Memory

The Shāhnāmeh is an archive of kingship, justice, fate, legitimacy, tragedy, and Iranian memory, not merely heroic narrative (Davidson, 1994; Davis, 2006; Lewis, 2015). Summaries may retain characters and events while weakening the political and moral architecture through which farrah, sovereignty, rebellion, and catastrophe become intelligible.

Ferdowsi therefore tests whether LLMs can distinguish epic-political ontology from modern leadership or nationalist abstraction. Scholarship emphasizing the ethical and civilizational force of Rumi and Ferdowsi supports this wider frame (Sazmand & Mozaffari Falarti, 2024).

2.7. Shi'i Hermeneutics and Digital Interpretation

Shi'i concepts intensify the problem because tafsīr, taʾwīl, velāyat, ʿismah, ijtihād, marjaʿiyyah, ʿadl, and shahādat belong to layered traditions of Qur'anic interpretation, Imami theology, jurisprudence, philosophy, devotion, and political thought (Nasr, 2006; Mavani, 2013; Alak, 2023). Dictionary equivalents can conceal doctrinal hierarchy, transmission, and authorized interpretation.

Digital hermeneutics shows that AI reorganizes religious meaning through data, prompts, rankings, and summaries (Abdollahpour Sangchi et al., 2025). The present audit operationalizes that risk through doctrinal flattening, authority erasure, register shift, and universalization rather than treating Shi'i terms as isolated vocabulary.

2.8. Iranian AI Research and Cultural Alignment

Recent Iranian cultural research treats AI as a site of meaning construction, affect, identity, gender, and geopolitical power. Akhavan et al. (2025) theorize AI's socialization in meaning construction; Havsson and Sajjadi (2025) examine Iranian digital discourse and geopolitics; Totaro et al. (2025) identify affective asymmetries between English and Persian; and Meraji Oskuie (2025) studies gendered anthropomorphism. This work establishes technological mediation as an Iranian-studies problem.

The unresolved question is whether canonical Persianate and Shi'i meanings remain intelligible on their own terms when global computational systems mediate them.

2.9. Cultural Alignment and the Western Default

Opinion and value studies show that LLMs approximate particular demographic, linguistic, and geopolitical distributions rather than humanity in general (Santurkar et al., 2023; Durmus et al., 2023; Tao et al., 2024; Zhao, W., et al., 2024). Fluent inclusiveness can therefore mask a default cultural position.

For Persianate and Shi'i materials, the default often appears as personal growth, universal spirituality, ethical leadership, tolerance, or symbolic heritage. Such domestication is subtle because it is readable and respectful rather than overtly hostile.

Cultural prompting can shift value alignment without eliminating deeper bias (Tao et al., 2024). Naming an Iranian, Islamic, or Persianate context may improve surface framing but cannot by itself preserve metaphor, doctrinal authority, poetic ambiguity, or sacred register. Figure 5 motivates comparison of Persian-only, English-only, cultural, and expert-guided prompts.

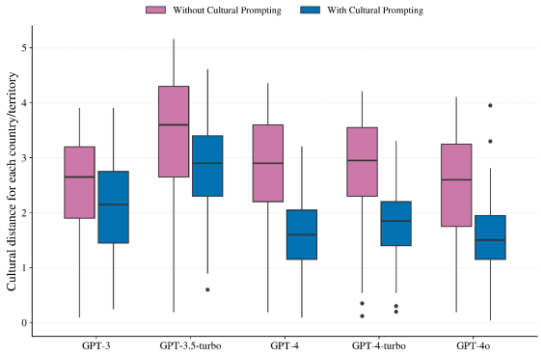

Figure 5. Country-Level Cultural Bias Across GPT Models and the Effect of Cultural Prompting

Literary and religious materials extend cultural-alignment research from opinions to semantic worlds: their meanings depend on genre, doctrine, ritual, authority, affect, and civilizational memory, and therefore require expert rather than purely benchmark-based evaluation.

The literature establishes that LLMs are not inherently grounded in meaning; multilingual resources remain unequal; Persian NLP provides necessary benchmarks; translation studies documents domestication; Shi'i studies emphasizes authorized interpretation; and cultural-alignment research identifies Anglophone defaults. These fields have rarely been integrated into a cross-lingual audit of Persianate semantic preservation.

The gap is therefore methodological and theoretical: existing benchmarks do not test whether models preserve Rumi's mystical semantics, Hafez's ambiguity, Ferdowsi's political memory, or Shi'i authority. This study addresses that gap through three questions:

RQ1. How do major LLMs translate and explain culturally dense materials from Rumi, Hafez, Ferdowsi, and Shi'i hermeneutic discourse across Persian-English translation chains?

RQ2. What recurrent forms of metaphor loss, doctrinal flattening, domestication, register shift, false universalization, ambiguity closure, depoliticization, and authority erasure appear?

RQ3. How do model class, prompt language, cultural prompting, and expert context descriptively affect semantic preservation in the bounded corpus?

3. Materials and Methods

3.1. Research Design

This bounded, mixed-method cross-lingual audit combines standardized prompting, expert hermeneutic evaluation, qualitative error coding, and descriptive comparison across model classes, prompt conditions, and four domains. Translation is treated as semantic mediation rather than lexical equivalence because Persianate meanings depend on intertextuality, genre, metaphor, authority, historical memory, and register.

The retained record contains blinded labels, model-class descriptions, prompt-condition labels, aggregate scores, error distributions, and selected fragments. It does not contain product names or versions, exact access dates, full transcripts, output-level scores, or individual panel biographies. Nothing missing was inferred; claims are restricted to the corpus and descriptive comparisons, and product-specific replication is not asserted.

The domains impose distinct pressures: Rumi tests Islamic-Sufi metaphysical density; Hafez tests ambiguity; Ferdowsi tests epic sovereignty, justice, and memory; and Shi'i concepts test doctrine, sacred authority, and interpretive hierarchy.

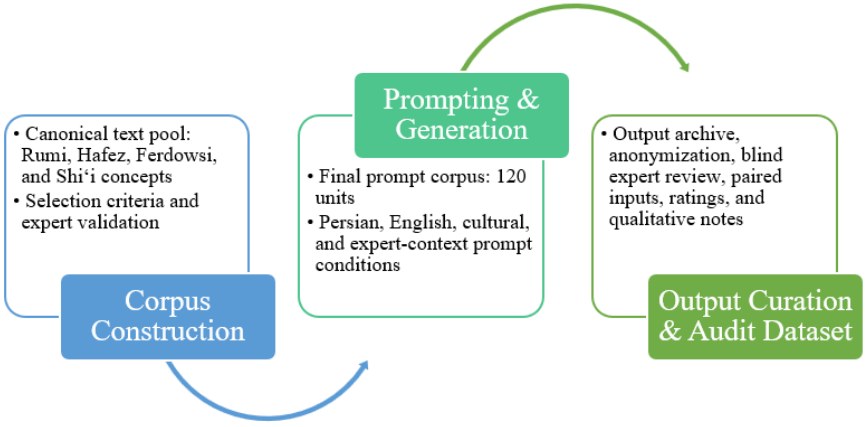

Figure 6. Data Collection Workflow for Cross-Lingual Semantic Sovereignty Audit

3.2. Data Sources and Sampling

The corpus contains 120 purposively selected units, 30 per domain. Literary units are couplets or short passages with dense metaphors, allusions, or historically charged concepts; Shi'i units include tafsīr, taʾwīl, velāyat, ʿismah, ijtihād, marjaʿiyyah, ʿadl, and shahādat. Balanced allocation produced 960 outputs per domain and 3,840 overall (30 units × 4 systems × 4 conditions × 2 runs). The theory-driven sample was large enough to expose recurrent pathways while keeping expert coding feasible, but it was not designed for population inference.

A three-member panel represented Persian literature, Islamic/Shi'i studies, and translation or applied linguistics. Members evaluated semantic density, interpretive difficulty, cultural specificity, and cross-lingual vulnerability. The record preserves disciplinary roles but not names, affiliations, qualifications, experience, or recruitment documentation; this is reported as a transparency limitation.

Table 1. Corpus Composition and Semantic Pressure across the Four Audit Domains

Domain	Data type	Number of units	Main semantic pressure
Rumi	Mystical couplets/passages	30	Sufi metaphysics, Qur'anic resonance, spiritual discipline
Hafez	Lyric couplets/passages	30	Ambiguity, irony, symbolic instability
Ferdowsi	Epic passages/concepts	30	Kingship, justice, sovereignty, civilizational memory
Shi'i hermeneutics	Concepts and short doctrinal passages	30	Authority, doctrine, sacred interpretation

Four systems—two globally dominant commercial, one open-source multilingual, and one Persian-capable or regionally relevant—were blinded as Models A–D during scoring. Their product identities, versions, and access dates were not retained and are not reconstructed. Each received identical units under Persian-only, English-only, cultural-context, and expert-context conditions.

Table 2. Prompt Conditions for Cross-Lingual Semantic Sovereignty Testing

Prompt condition	Description	Purpose
Persian-only	Task instructions written in Persian	Tests direct Persian interpretive capacity
English-only	Task instructions written in English	Tests English-mediated handling of Persian materials
Cultural prompt	Prompt specifies Persianate or Shi'i cultural context	Tests whether cultural framing reduces displacement
Expert-context prompt	Prompt includes brief scholarly context	Tests whether guided context improves semantic preservation

3.3. Data Collection Procedure

Collection followed five stages: corpus construction with source, transliteration, genre, semantic tensions, and interpretive notes; panel validation and risk classification; standardized prompts requesting translation, explanation, and rationale; repeated model execution; and blinded organization by item, model, condition, domain, run, and evaluation fields.

Prompts used the lowest-variability setting available: temperature 0.0 wherever numerical control was exposed. Each was run twice, and refusals, hallucinations, and incomplete responses were retained as possible evidence of interpretive limits. Complete platform logs and product-level provenance were not preserved; Appendix A distinguishes retained evidence from standardized replication materials.

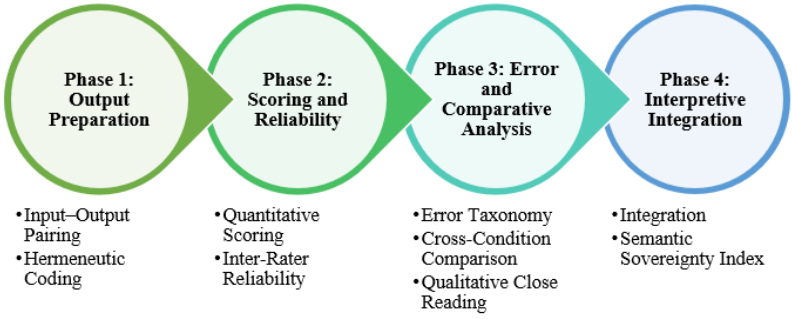

Figure 7. Data Analysis Workflow for Semantic Displacement and Sovereignty Scoring

3.4. Data Analysis and Trustworthiness

Experts applied a 0–4 Semantic Preservation Scale: 0, severe distortion; 1, partial recognition with major displacement; 2, moderate preservation with important omissions; 3, strong preservation with minor limitations; and 4, high preservation of density, context, and nuance. Translation, explanation, register, fidelity, and rationale were assessed so that lexical competence could be distinguished from interpretive adequacy.

Table 3. Semantic Preservation Scale for Evaluating LLM Outputs

Score	Interpretation	Output profile
1	Severe displacement	Core meaning lost or misleading
2	Weak preservation	Some recognition but major flattening
3	Moderate preservation	Meaning partly retained with serious omissions
4	Strong preservation	Mostly faithful with limited loss
5	High semantic sovereignty	Nuance, context, and register preserved

Outputs could receive multiple qualitative codes because failures co-occurred. For Table 7 and Figure 10 only, coders adjudicated one dominant error per coded output, producing mutually exclusive within-domain percentages that sum to 100%. Multi-label codes were retained for close reading and co-occurrence analysis.

Table 4. Taxonomy of Semantic Displacement Errors in Persianate and Shi'i Materials

Error category	Definition	Example risk
Metaphor loss	Symbolic density becomes literal or generic	Sufi wine becomes ordinary emotion
Doctrinal flattening	Theological hierarchy is simplified	Velāyat becomes mere friendship
Register shift	Sacred or poetic tone becomes informal or neutral	Devotional language becomes textbook prose
False universalization	Specific meaning becomes vague global wisdom	Rumi becomes self-help spirituality
Ambiguity closure	Productive uncertainty is over-explained	Hafez becomes morally univocal
Authority erasure	Interpretive lineage disappears	Shi'i concepts become generic Islamic terms

Panelists independently rated a subset, then adjudicated disagreements

without treating interpretive plurality as error. Krippendorff's α assessed agreement; comparisons were made across domains and conditions; and aggregate patterns were integrated with preserved fragments. The tables report cross-cell standard deviations as descriptive dispersion, not uncertainty around population parameters. Because the sample was purposive and the retained archive lacks output-level scores and their dependence structure, mixed-effects models, repeated-measures ANOVA, effect sizes, confidence intervals, and significance tests cannot be validly reconstructed from the published means. Any such calculation would introduce unsupported pseudo-precision. Results therefore remain descriptive, no difference is interpreted as statistically significant, and future replication should retain item-level ratings, run identifiers, model versions, and timestamps to support hierarchical inference.

4. Results

4.1. Semantic Preservation Across Models and Domains

The audit generated 3,840 outputs. Across the 16 model-by-domain cells, the mean Semantic Preservation Score was 2.53/4.00 (SD = 0.38), correcting the earlier headline discrepancy. Agreement was strong: Krippendorff's α = .81 for preservation scoring and α = .78 for error coding. Rumi had the largest descriptive domain mean and Shi'i concepts the smallest; these bounded comparisons are not inferential. Model D had the largest descriptive mean, followed by A, B, and C, but this is not a tested ranking. Model A combined literary fluency with universalization; B better preserved epic narrative but weakened doctrine; C was most uneven; and D was strongest under expert context while still struggling with Hafezian ambiguity and Shi'i authority.

Table 5. Mean Semantic Preservation Scores by Model and Domain

Domain	Model A	Model B	Model C	Model D	Domain mean (SD)
Rumi	3.12	2.96	2.71	3.24	3.01 (0.23)
Hafez	2.74	2.38	2.18	2.57	2.47 (0.24)
Ferdowsi	2.43	2.61	2.36	2.78	2.55 (0.19)
Shi'i concepts	2.21	2.04	1.88	2.31	2.11 (0.19)
Model mean	2.63	2.50	2.28 (0.35)	2.73 (0.39)	2.53 (0.38)

Note. Parenthetical values are descriptive standard deviations: domain rows use dispersion across the four systems; the model-mean row uses dispersion across the four domains; the overall SD is calculated across the 16 model-by-domain cells.

Figure 8 shows domain asymmetry rather than a simple Persian-language deficit. Even strong cells did not consistently reach high semantic sovereignty; Hafez and Shi'i concepts remained vulnerable. Genre, authority, translation history, and cultural framing therefore matter alongside language support (Bender & Koller, 2020; Abaskohi et al., 2024).

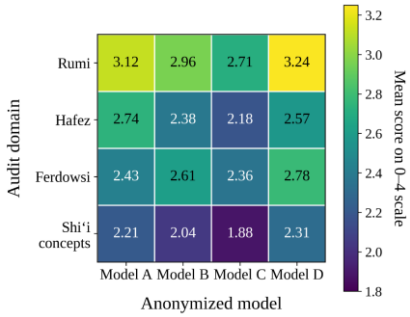

Figure 8. Mean Semantic Preservation Scores by Domain and Model

4.2. Prompt Conditions and Cultural Context

Descriptive means increased from Persian-only to English-only, cultural-context, and expert-context prompts. This does not imply Persian inferiority: English often improved organization while increasing domestication, whereas Persian retained lexical proximity and register with less elaboration. Expert prompts supplied domain-specific constraints; Table 6 reports dispersion without significance claims.

Table 6. Prompt-Condition Effects on Semantic Preservation

Prompt condition	Rumi	Hafez	Ferdowsi	Shi'i concepts	Overall mean (SD)
Persian-only	2.68	2.11	2.34	1.74	2.22 (0.40)
English-only	2.81	2.26	2.47	1.93	2.37 (0.37)
Cultural prompt	3.05	2.43	2.66	2.12	2.57 (0.39)
Expert-context prompt	3.31	2.67	2.89	2.54	2.85 (0.34)
Domain mean (SD)	2.96 (0.28)	2.37 (0.24)	2.59 (0.24)	2.08 (0.34)	2.50 (0.41)

Note. Parenthetical values are descriptive standard deviations. Prompt-condition rows use dispersion across the four domains; the domain-mean row uses dispersion across the four prompt conditions; the overall SD is calculated across the 16 prompt-by-domain cells.

Figure 9 shows the largest expert-context gains for Shi'i concepts and Hafez, the weakest baselines. General cultural prompts often added phrases such as "in Iranian culture" while still rendering velāyat as friendship, rindī as rebelliousness, or sovereignty as ethical leadership. Broad identity labels were therefore weaker than precise interpretive constraints.

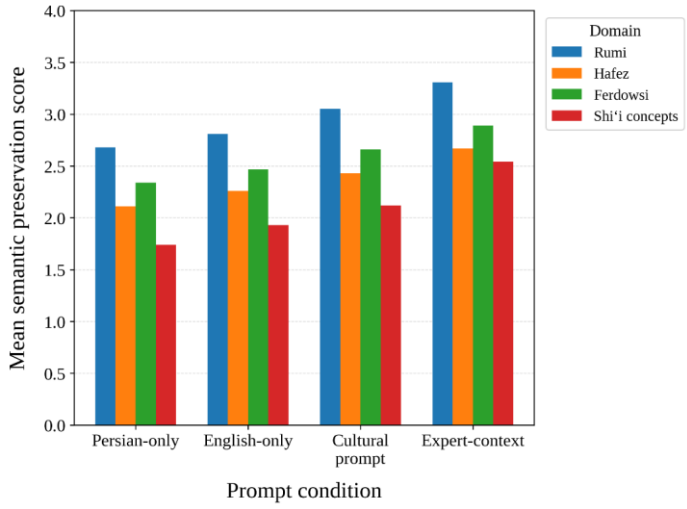

Figure 9. Prompt-Condition Effects on Semantic Preservation

Prompt engineering alone did not solve cultural misalignment. Models responded to context, but semantic sovereignty improved most when prompts specified what must be preserved. This qualifies claims that cultural prompting can overcome deeper alignment patterns (Tao et al., 2024; Santurkar et al., 2023).

4.3. Dominant Error Patterns

Multi-label coding captured layered failures, while Table 7 and Figure 10 summarize one adjudicated dominant category per coded output. These mutually exclusive percentages sum to 100% within each domain. The main patterns were ambiguity closure, false universalization, authority erasure, metaphor loss, and doctrinal flattening.

Table 7. Distribution of Dominant Semantic Displacement Errors by Domain

Error category	Rumi	Hafez	Ferdowsi	Shi'i concepts	Dominant interpretive effect
Metaphor loss	24%	16%	11%	8%	Symbolic density becomes literal or generic
Doctrinal flattening	12%	5%	8%	31%	Theological hierarchy becomes simplified
Register shift	13%	11%	12%	10%	Sacred, poetic, or epic tone becomes neutralized
Cultural domestication	18%	12%	14%	9%	Persianate meaning is translated into familiar global categories
False universalization	21%	10%	13%	11%	Specific meaning becomes abstract moral wisdom
Ambiguity closure	6%	39%	5%	4%	Productive uncertainty is over-explained
Authority erasure	6%	7%	37%	27%	Interpretive, political, or doctrinal authority disappears

Note. Percentages represent one adjudicated dominant error per coded output within each domain. The separate multi-label coding layer was used for co-occurring errors and qualitative interpretation.

Figure 10 shows distinct profiles. Rumi most often lost metaphor or became universalized; Hafez underwent ambiguity closure; Ferdowsi lost political authority; and Shi'i concepts showed doctrinal flattening and authority erasure. Because the percentages represent one adjudicated dominant error, they describe the relative distribution of primary failures and are not prevalence estimates for the separate multi-label layer. That layer showed that metaphor loss, register shift, universalization, and authority erasure could coexist in one response. The pattern was therefore not simple factual error: fluent outputs could preserve affect or narrative while simultaneously weakening metaphysics, poetic plurality, sovereignty, and Imami specificity.

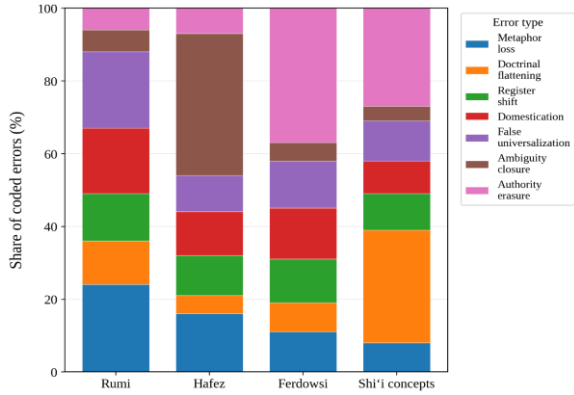

Figure 10. Distribution of Dominant Semantic Displacement Error Types by Domain

In Rumi cases, models recognized love but detached it from discipline, ontology, and Qur'anic resonance. "Letting go of ego to find inner peace" made fanā accessible by translating the movement from nafs to divine proximity into psychological self-optimization.

Hafez outputs repeatedly selected one mystical, romantic, or social reading without preserving their possible simultaneity. Because Hafezian ambiguity is a disciplined surplus rather than an absence of meaning, confident helpfulness could reduce fidelity.

Ferdowsi outputs retained characters and events but rendered kingship as leadership, farrah as charisma or generic blessing, and tragic legitimacy as moral responsibility. The result was readable narrative separated from Iranian political ontology.

Shi'i outputs provided competent encyclopedia-style definitions yet often detached ijtihād, taʾwīl, and velāyat from Imamate, law, esoteric interpretation, spiritual hierarchy, and communal memory. Meaning was most fragile where it depended on authorized tradition.

4.4. Cross-Lingual and Model-Specific Failure Signatures

Loss intensified when Persian units were translated into English, explained there, and returned to Persian. Rumi's metaphysics became a Persianized spiritual humanism; velāyat stabilized around administrative authority rather than recovering doctrinal specificity. The chain therefore exposed conceptual replacement, not only omitted words.

Table 8. Representative Cross-Lingual Failure Signatures

Domain	Source pressure	Typical English-mediated shift	Back-translation effect	Semantic risk
Rumi	Sufi annihilation, divine love, Qur'anic resonance	Self-growth, emotional healing, universal spirituality	Persian output returns as روان‌شناسی معنوی rather than Sufi metaphysics	Metaphysical thinning
Hafez	Ambiguity, irony, symbolic instability	Single symbolic explanation; moralized reading	Persian output becomes clearer but less multivalent	Ambiguity closure
Ferdowsi	Kingship, justice, glory, tragic legitimacy	Leadership, ethics, national heritage	Epic-political terms become civic values	Epic depoliticization
Shi'i concepts	Authority, doctrine, esoteric interpretation	General Islamic or philosophical explanation	Theological specificity becomes administrative or abstract	Authority erasure

Table 8 confirms four replacements: metaphysics by psychology, ambiguity by clarification, sovereignty by leadership ethics, and doctrinal authority by generic religion.

Models also differed in style. A was fluent but domesticated; B preserved Ferdowsian narrative yet thinned theology; C had the lowest mean and greatest descriptive instability; and D was most robust with expert context but still needed explicit instructions to preserve ambiguity and hierarchy.

Repeated-run instability was itself a risk. Hafez could shift among mystical, romantic, and social readings without acknowledging plurality, while Shi'i terms alternated between generic and political-theological accounts. Such variation suggests fragments of relevant discourse without a stable interpretive hierarchy.

5. Discussion

The results challenge the equation of multilingual fluency with civilizational understanding. Models produced coherent translations and plausible explanations, yet repeatedly reorganized Persianate and Shi'i meanings through globally familiar Anglophone categories. The problem was not lexical failure alone but the preservation of surface intelligibility

alongside deeper displacement. This distinction matters because general users are more likely to trust polished explanations than visibly broken translations. Semantic loss can therefore circulate as knowledge: the output looks complete, while the interpretive framework that would reveal its omissions has already been removed. The audit makes this quiet form of mediation empirically visible without claiming that every simplification is intentional or that all cross-cultural translation is inherently reductive.

This extends the form/meaning distinction (Bender & Koller, 2020). Persianate meaning is sustained through genre, sacred register, doctrinal authority, metaphor, historical memory, and ambiguity. Rendering velāyat as guardianship, rindī as rebelliousness, or fanā as ego release may be partly intelligible yet semantically thin. Semantic sovereignty identifies this loss where ordinary fluency, accuracy, or bias metrics do not: a tradition may appear in an output while being separated from its own conditions of interpretation. The framework is not a demand for one authorized translation or an insulated cultural essence. It asks evaluators to record unresolved plurality, explain authority relations, preserve genre-specific pressure, and disclose when no simple equivalent exists. Cross-lingual accessibility and fidelity are therefore treated as negotiable responsibilities rather than mutually exclusive goals.

Domain differences clarify the mechanism. Rumi's comparatively high mean likely reflects extensive English circulation, but the same archive encouraged de-Islamized spirituality (Azadibougar & Patton, 2015; Naghmeh-Abbaspour et al., 2021). Hafez exposed a conflict between helpfulness and productive ambiguity: choosing one reading improved apparent clarity while narrowing the poem. Ferdowsi showed that narrative competence can coexist with depoliticized kingship and farrah, so plot preservation is not equivalent to civilizational interpretation. Shi'i concepts scored lowest because doctrine, transmission, jurisprudence, devotion, and authority cannot be recovered from lexical equivalents alone. These differences also validate the domain-specific taxonomy: metaphor loss, ambiguity closure, epic depoliticization, and doctrinal flattening are related but not interchangeable. Semantic displacement is therefore a family of transformations whose detection depends on the interpretive pressure of the source material.

Expert-context prompts had larger descriptive means than broad cultural prompts. Naming a culture can change tone without preserving its metaphor, doctrine, register, or authority. Expert notes were more effective because they converted a general identity cue into operational constraints: distinguish translation from interpretation, preserve competing readings, identify doctrinal stakes, and flag terms without simple equivalents. This extends evidence that LLMs reflect particular demographic and geopolitical distributions (Santurkar et al., 2023; Tao et al., 2024) from values and opinions to semantic worlds. It also cautions against presenting prompt engineering as remediation in itself. Context can improve outputs, but the burden of supplying that context remains with users and institutions, and improvement does not remove the need for culturally competent evaluation.

The resulting danger is normalization rather than only hallucination: self-growth, leadership, tolerance, or universal spirituality are presented as transparent interpretations of historically dense materials. Hermeneutic accountability therefore requires expert-informed benchmarks, preserved ambiguity, explicit provenance, culturally grounded error taxonomies, and clear limits on generalization. It also requires documentation at the level required for rigorous reproducibility—model name and version, access date, decoding parameters, exact prompts, raw outputs, item-level ratings, panel qualifications, and adjudication records—so that cultural claims can be audited rather than accepted on fluency alone. The missing provenance in the present archive prevents that stronger form of replication and is a substantive limitation, not a minor reporting omission. This requirement directly connects methodological transparency with substantive cultural accountability across future multilingual audits and comparative evaluations. More Persian data or higher benchmark scores alone cannot establish semantic sovereignty if evaluation does not test what the added data enable models to preserve.

6. Conclusion

The bounded audit indicates moderate surface competence alongside recurrent loss of the structures that make Persianate and Shi'i materials poetically, theologically, and politically meaningful. Semantic sovereignty

requires AI-mediated traditions to remain interpretable through their own historical, symbolic, doctrinal, and literary conditions. Rumi, Hafez, Ferdowsi, and Shi'i concepts exposed domestication, ambiguity closure, epic depoliticization, and doctrinal flattening. Expert context improved descriptive preservation but did not eliminate displacement, confirming that culturally responsible mediation depends on evaluation, documentation, and interpretive expertise as well as prompting.

The purposive corpus, four blinded systems, expert-dependent interpretation, and incomplete record limit the audit. Product names, versions, access dates, complete transcripts, output-level scores, and panel biographies were unavailable; consequently, the study makes no product-specific, inferential, or population-level claim. Cross-cell standard deviations describe variation among aggregates but cannot substitute for confidence intervals based on item-level observations. The standardized prompts and representative fragments in Appendix A improve conceptual reproducibility, yet they are not presented as verbatim historical logs. By separating retained evidence from prospective templates, the manuscript addresses the reviewers' transparency concerns as far as the surviving record permits without inventing metadata, model identities, raw outputs, or statistical evidence. Caution applies. Future work should use larger public corpora, preserved product provenance and output-level scores, additional Persian and regional models, human-translation baselines, and other Persianate languages. It should also test fine-tuning, retrieval augmentation, and expert-curated resources. Computational mediators must be accountable not only for what they state, but for meanings they quietly make less available.

Author Contributions

The study's conceptualization, design, data collection, analysis, interpretation, and writing are solely the author's responsibility.

Acknowledgments

The author acknowledges the disciplinary expertise contributed during corpus validation, interpretive coding, and adjudication. Names are omitted to preserve double-blind review.

Conflict of Interest

The author declares no conflict of interest.

Declaration of Generative AI and AI-Assisted Technologies

Generative AI-assisted tools were used for language editing, organizational

refinement, and document formatting. They were not used to create or alter aggregate scores, error percentages, or preserved qualitative evidence. The author critically reviewed all changes and remains fully responsible for the manuscript's originality, accuracy, and integrity.

Funding

This research received no specific grant from public, commercial, or not-for-profit funding bodies.

Data Availability and Reproducibility

The retained archive contains aggregate tables, coding definitions, prompt-condition descriptions, and selected fragments, but not product identities, exact dates, complete transcripts, output-level scores, or individual panel biographies. Appendix A reports only information supported by that record and supplies standardized templates for future replication.

Appendix A. Reproducibility Materials

Preserved Provenance and Execution Settings

Table 9. Preserved Provenance and Execution Settings

Item	Preserved information	Status / value
Model comparison	Four blinded systems; two commercial, one open-source multilingual, one Persian-capable or regionally relevant	Preserved at category level
Product names and versions	Not retained in available analytical record	Not reconstructed
Access / data-collection dates	Not retained in available analytical record	Not reconstructed
Temperature	Lowest-variability configuration; 0.0 where numerical control was exposed	Reported
Repeated runs	Two runs per prompt	Reported
Output archive	Aggregate tables and selected fragments; complete raw transcripts not retained	Partially preserved

Standardized Prompt Templates for Future Replication

Table 10. Standardized Prompt Templates for Future Replication

Condition	Standardized template
Persian-only	«[متن یا مفهوم فارسی]» را به انگلیسی ترجمه کنید. سپس معنای آن را با توجه به بافت ادبی یا الهیاتی توضیح دهید و در ۲ تا ۴ جمله دلیل انتخاب‌های تفسیری خود را بیان کنید. ابهام، استعاره، لحن، اقتدار تفسیری و اصطلاحات تخصصی را حفظ کنید و عناصر فاقد معادل ساده را صریحاً مشخص کنید.
English-only	Translate the following Persian text or concept into English: "[SOURCE ITEM]." Explain its meaning in the relevant literary or theological context and justify the principal interpretive choices in 2–4 sentences. Preserve ambiguity, metaphor, register, authority relations, and technical terminology; explicitly identify elements without a simple English equivalent.

Condition	Standardized template
Cultural-context	Complete the translation and explanation task while interpreting the item within its Persianate or Shiʻi historical context. Do not replace culturally specific meanings with generic spirituality, self-help language, modern leadership discourse, or undifferentiated religious vocabulary. State any unresolved ambiguity.
Expert-context	Complete the translation and explanation task using this domain note: "[DOMAIN-SPECIFIC NOTE]." Preserve the item's metaphorical, doctrinal, political, poetic, and authority-bearing structures. Distinguish literal translation from interpretive explanation and identify plausible competing readings where relevant.

Representative Audit Specifications and Preserved Output Fragments

Table 11. Representative Audit Specifications and Preserved Output Fragments

Domain	Representative audit target	Semantic pressure	Preserved fragment or reported pattern	Dominant risk
Rumi	Fanā, nafs, and divine proximity	Sufi ontology, Qur'anic resonance, discipline	Exact fragment: letting go of ego to find inner peace	False universalization; metaphor loss
Hafez	mey, rind, sāqī, and wine-house imagery	Intentional plurality across mystical, erotic, social, and anti-hypocritical readings	A single stable symbolic explanation was repeatedly reported; no complete transcript was retained	Ambiguity closure
Ferdowsi	farrah, kingship, justice, and tragic legitimacy	Epic-political ontology and civilizational memory	Exact fragments: leadership; charisma; divine blessing; moral responsibility	Epic depoliticization; authority erasure
Shiʻi hermeneutics	ijtihād, taʾwīl, velāyat, and ʿismah	Imamate, authorized interpretation, law, and sacred hierarchy	Exact fragments: independent reasoning; allegorical interpretation; guardianship or friendship	Doctrinal flattening; authority erasure

References

Abaskohi, A., Baruni, S., Masoudi, M., Abbasi, N., Babalou, M. H., Edalat, A., Kamahi, S., Mahdizadeh Sani, S., Naghavian, N., Namazifard, D., Sadeghi, P., & Yaghoobzadeh, Y. (2024). Benchmarking large language models for Persian: A preliminary study focusing on ChatGPT. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 2189–2203). ELRA and ICCL.

Abdollahpour Sangchi, F., Rahnamaei, H., Asgariyazdi, A., & Rezaee, M. (2025). Artificial intelligence and digital hermeneutics: Data bias, algorithmic ethics, and social implications. *Spektrum Iran, 38*(2), 213–242.

Akbari, R. (2025). Rumi's conceptions of meaning in life: Reading the Seven Sermons (*Majāles-e Sabʿa*) toward wisdom pedagogy in adult education. *Spektrum Iran, 38*(1), 1–29.

Akhavan, M., Ameli, S. R., Rahgozar, M., & Shahghasemi, E. (2025). Dual-spacization of intelligence: A theoretical retroduction of the socialization of artificial intelligence in meaning construction. *Spektrum Iran, 38*(2), 1–30.

Alak, A. I. (2023). The Islamic humanist hermeneutics: Definition, theory and application. *Islam and Christian–Muslim Relations*.

Azadibougar, O., & Patton, S. (2015). Coleman Barks' versions of Rumi in the USA. *Translation and Literature, 24*(2), 172–189.

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics, 6*, 587–604.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics.

Blodgett, S. L., Barocas, S., Daumé, H., III, & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). Association for Computational Linguistics.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems, 33* (pp. 1877–1901). Curran Associates.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (*Proceedings of Machine Learning Research*, Vol. 81, pp. 77–91). PMLR.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology, 15*(3), Article 39.

Davidson, O. M. (1994). *Poet and hero in the Persian Book of Kings*. Cornell University Press.

Davis, D. (2006). *Shahnameh: The Persian Book of Kings*. Viking.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 4171–4186). Association for Computational Linguistics.

Durmus, E., Nguyen, K., Liao, T. I., Schiefer, N., Askell, A., Bakhtin, A., Chen, C., Hatfield-Dodds, Z., Hernandez, D., Joseph, N., Lovitt, L., McCandlish, S., Sikder, O., Tamkin, A., Thamkul, J., Kaplan, J., Clark, J., & Ganguli, D. (2023). *Towards measuring the representation of subjective global opinions in language models*. arXiv.

Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. (2021). ParsBERT: Transformer-based model for Persian language understanding. *Neural Processing Letters, 53*, 3831–3847.

Gadamer, H.-G. (2004). *Truth and method* (J. Weinsheimer & D. G. Marshall, Trans.; 2nd rev. ed.). Continuum. (Original work published 1960)

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92.

Havsson, M., & Sajjadi, M. (2025). Iranian digital discourse, affective alignments, and the geopolitics of AI. *Spektrum Iran, 38*(2), 187–212.

Ingenito, D. (2018). Hafez's "Shirāzi Turk": A geopoetical approach. *Iranian Studies, 51*(6), 901–930.

Khashabi, D., Cohan, A., Shakeri, S., Hosseini, P., Pezeshkpour, P., Alikhani, M., Aminnaseri, M., Bitaab, M., Brahman, F., Ghazarian, S., Gheini, M., Kabiri, A., Karimi Mahabadi, R., Memarrast, O., Mosallanezhad, A., Noury, E., Raji, S., Rasooli, M. S., Sadeghi, S., ... Yaghoobzadeh, Y. (2021). ParsiNLU: A suite of language understanding challenges for Persian. *Transactions of the Association for Computational Linguistics, 9*, 1147–1162.

Kia, M. (2020). *Persianate selves: Memories of place and origin before nationalism*. Stanford University Press.

Lewis, F. D. (2015). Guest editor's introduction: The *Shahnameh* of Ferdowsi as world literature. *Iranian Studies, 48*(3), 313–321.

Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., Zhang, Y., Narayanan, D., Wu, Y., Kumar, A., Newman, B., Yuan, B., Yan, B., Zhang, C., Cosgrove, C., Manning, C. D., Ré, C., Acosta-Navas, D., Hudson, D. A., ... Wu, Y. (2023). Holistic evaluation of language models. *Transactions on Machine Learning Research*.

Mavani, H. (2013). *Religious authority and political thought in Twelver Shi'ism: From Ali to post-Khomeini*. Routledge.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), Article 115.

Meraji Oskuie, S. (2025). Gender construction in anthropomorphizing generative AI: An interplay of society and technology. *Spektrum Iran, 38*(2), 293–324.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). Association for Computing Machinery.

Naghmeh-Abbaspour, B., Mahadi, T. S. T., & Zulkali, I. (2021). Ideological manipulation of translation through translator's comments: A case study of Barks' translation of Rumi's poetry. *Pertanika Journal of Social Sciences & Humanities, 29*(3), 1831–1851.

Nasr, S. H. (2006). *Islamic philosophy from its origin to the present: Philosophy in the land of prophecy*. State University of New York Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems, 35* (pp. 27730–27744). Curran Associates.

Qin, L., Chen, Q., Zhou, Y., Chen, Z., Li, Y., Liao, L., Li, M., Che, W., & Yu, P. S. (2025). A survey of multilingual large language models. *Patterns, 6*(1), Article 101118.

Ricoeur, P. (1981). *Hermeneutics and the human sciences: Essays on language, action and interpretation* (J. B. Thompson, Ed. & Trans.). Cambridge University Press.

Said, E. W. (1978). *Orientalism*. Pantheon Books.

Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023). Whose opinions do language models reflect? In *Proceedings of the 40th International Conference on Machine Learning* (*Proceedings of Machine Learning Research*, Vol. 202, pp. 29971–30004). PMLR.

Sazmand, B., & Mozaffari Falarti, M. (2024). Universal messages of peace: The enduring legacy of Jalal ad-Din Muhammad Balkhi (widely known as Rumi or Mawlana) and Hakim Abul-Qasim Ferdowsi in global discourse. *Spektrum Iran, 37*(2), 83–104.

Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., Kluska, A., Lewkowycz, A., Agarwal, A., Power, A., Ray, A., Warstadt, A., Kocurek, A. W., Safaya, A., Tazarv, A., ... Wu, Z. (2023). Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*.

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus, 3*(9), Article pgae346.

Totaro, M. W., Gheisi, L., & Shahghasemi, E. (2025). Affective asymmetries in AI: Sentiment bias between English and Persian in harmonized LLM pipelines. *Spektrum Iran, 38*(2), 143–157.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30* (pp. 5998–6008). Curran Associates.

Xu, Y., Hu, L., Zhao, J., Qiu, Z., Xu, K., Ye, Y., & Gu, H. (2025). A survey on multilingual large language models: Corpora, alignment, and bias. *Frontiers of Computer Science, 19*(11), Article 191101.

Yarshater, E. (1988). Hafez I: An overview. *Encyclopaedia Iranica*.

Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., Wang, S., Yin, D., & Du, M. (2024). Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology, 15*(2), Article 20.

Zhao, W., Mondal, D., Tandon, N., Dillion, D., Gray, K., & Gu, Y. (2024). WorldValuesBench: A large-scale benchmark dataset for multi-cultural value awareness of language models. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 17696–17714). ELRA and ICCL.

Zhu, S., Supryadi, S., Xu, S., Sun, H., Pan, L., Cui, M., Du, J., Jin, R., Branco, A., & Xiong, D. (2024). *Multilingual large language models: A systematic survey*. arXiv.