AI Narratives of Contemporary Islam in Gemini and OpenAI: A Comparative Analytical Report

1) Executive Overview and Research Objectives

The purpose of this report is to deliver an outcome of a multi-method, comparative analysis of how two leading foundation model platforms—OpenAI and Gemini—construct, constrain, and disseminate narratives about contemporary Islam. We integrate a local corpus of AI-generated essays (organized across themes) with a quantitative meta-analysis (En Collab, 13 Feb 2025) and extend it with governance, methodological, and evaluation innovations to diagnose and improve narrative quality, pluralism, and safety. The stakes span education (curricula and classroom adoption), media (framing and representation), policy (CVE, human rights, and religious freedom), and industrial markets (halal economy, Islamic finance, and modest fashion). **The central claim: both systems present largely neutral, convergent textual narratives on Islam overall but exhibit material divergences in topic salience that affect public perception, institutional uptake, and risk profiles.** [101]

Empirically, the En Collab analysis finds:

- Predominantly neutral sentiment distributions across both OpenAI- and Gemini-authored Islam corpora (~79–80% neutral; ~17–18% positive; ~3% negative). [101]
- Extremely high cross-model lexical similarity (token-distribution similarity 0.9301494620927) on "Islam and contemporary issues." [101]
- Topic-model divergence: OpenAI's dominant theme emphasizes Islamic stewardship and environmentalism; Gemini's dominant theme emphasizes Islamic radicalism and its political/social drivers. [101]

Research questions:

1. How do narrative frames (compatibility, stewardship, empowerment, democratic consonance, ethical market, tech embrace, boundary policing) recur and diverge across platforms and themes? [101]
2. What is the measurable impact of safety and moderation regimes on coverage, assertiveness, and omission (e.g., apostasy/blasphemy; intra-faith jurisprudence)? [37][38][195]
3. What quantitative and qualitative methods most reliably identify framing, stance, and pluralism in LLM outputs on religion? [59][101]
4. What technical and governance interventions (RAG, multi-agent debate with guardrails, pluralism dials, red teaming) improve narrative completeness without compromising safety? [218][119][166][167]

Significance:

- **Education and media**: neutral-but-convergent AI narratives can sanitize or homogenize doctrinal complexity, affecting pedagogy and reporting. [101][202]
- **Policy**: platform safety taxonomies and refusal patterns may suppress discussion of sensitive legal-theological areas (hudud, apostasy) or securitize Islam by salience-skew, shaping public discourse and CVE practice. [37][38][158]
- **Markets**: scalable narratives influence halal economy trajectories, consumer trust, and sustainable finance strategies (e.g., green sukuk), with clear opportunities to improve literacy and signal integrity. [39][77][83]

2) Definitions, Scope, and Analytical Framework

Definitions:

- By “narrative” we mean an assemblage of framing devices (problem definitions, causal attributions, moral evaluations, and remedies), stance (support/critique/neutrality toward positions and practices), valence and affect (sentiment polarity), and epistemic posture (authoritative versus dialogic). **This aligns with framing theory operationalizations for computational analysis.** [59]
- “Contemporary Islam” spans 2000–present, encompassing Muslim-majority contexts and diasporas; theologically Sunni, Shia (Ja‘fari/Zaydi), Ibadi, and Sufi; and “lived Islam” across culture, economy, politics, digital practice, and identity. [127][125]

Analytic lenses:

- Mixed-methods CDA: sentiment/valence detection; topic modeling; term co-occurrence; stance detection; and manual coding for frames such as reconciliation, reformism, authenticity, securitization. [59][101]
- Value-pluralism: we evaluate whether outputs represent multiple doctrinal positions (e.g., schools of law), minority and majority readings, and region-specific practices. **We propose pluralism-aware metrics and dials to control breadth vs depth.** [207][208][209]
- Safety-policy impact mapping: we treat moderation/guardrail policies—especially on religion, hate, and “off-topic” gates—as potential shapers of coverage, refusal types, and hedging density. [37][39][195]

3) Corpus Design: Sources, Sampling, and Data Model

Corpus sources:

- Local OpenAI/Gemini essays organized into themes (e.g., Islamic modernity; democracy; human rights; feminism; environmentalism; fashion; technology; halal; Islamic finance/circular economy; and radicalism).
- En Collab comparative meta-analysis (13 Feb 2025) summarizing word frequency, sentiment, co-occurrence, chunking, TF–IDF, similarity, topic modeling, NER, and clustering—used as baseline analytics. [101]
- Auxiliary knowledge bases for triangulation (e.g., halal market and certification; Islamic finance and green sukuk; digital religion; governance/civil society; media studies on representation; and safety policy docs). [39][55][77][75][202][37]

Metadata schema:

- Model family/version; date/time; prompt template; system instructions; decoding parameters (temperature/top-p/seed); token lengths; safety messages/refusals. **Version control is essential for drift measurement and reproducibility.** [42][131]

Inclusion/exclusion:

- Include platform-generated expository content; exclude user comments and non-curated social media (except where stance datasets are methodologically used).
- Focus on English; note plans for multilingual expansion leveraging OpenITI and Arabic stance datasets. [49][69]

4) Methodology: Mixed-Methods Comparative Analysis

Quantitative analytics:

- Sentiment analysis (neutral dominance with modest positive skew confirmed across corpora). [101]
- Topic modeling: use embedding-based methods for short texts (BERTopic/Top2Vec) with baseline classical models (LDA/NMF) and human/LLM adjudication for interpretability. **Evidence suggests BERTopic and Top2Vec outperform LDA/NMF on short inputs but should be benchmarked side-by-side.** [145][147]
- Co-occurrence networks and TF–IDF differentials to detect ideational clusters (e.g., Islamic+banking/fashion/principles; human+rights; circular+economy). [101]

Qualitative coding:

- Codebook for frames: compatibility, stewardship, empowerment, democratic consonance, ethical market, tech embrace, boundary policing (mainstream Islam vs extremism). [101]
- Reliability: dual coding and Cohen's kappa; integrate multi-perspective soft-labels to capture legitimate disagreement rather than enforcing consensus. [207][209]

Stance detection:

- Combine rule-based lexicons and supervised classifiers (e.g., AraStance; Arabic hijab discourse dataset) to capture stance toward contentious topics like hijab agency vs coercion, Sharia-democracy compatibility, and rights tension points. [69][71]

Safety/policy impact audit:

- Log safety refusals/disclaimers; label refusal types; relate to provider policies that treat religion as protected (hate) and sometimes off-topic (agent guardrails). **Gemini policies and Vertex templates illustrate multi-layer constraint surfaces.** [37][39][195]

Comparative design:

- A/B prompt runs across OpenAI/Gemini; hold prompts constant; vary system messages (expert/layperson/critic). **Pin model version IDs and seeds for reproducibility and drift detection.** [42][131]

5) System and Model Profiling

Architectures and constraints:

- OpenAI GPT-4 family variants and GPT-4.1 release position instruction-following and coding precision; safety evaluations are published in a hub; long-context performance claimed improved, affecting refusal granularity and detail. [41][42][45]
- Gemini 1.5 app/API safety design emphasizes harm/harassment/hate filters, developers' responsibility for context-specific harm, and configurable moderation layers (e.g., religion "off-topic" in guardrail templates). [38][39]

Policy surfaces and narrative effects:

- "Religion" is a protected characteristic under hate categories; certain enterprise guardrails classify religion as out-of-scope; user complaints suggest some consumer surfaces preclude generative prayers—even "for world peace"—producing perceived inconsistency when retrieval remains allowed. **These design choices shape discourse scope, hedging, and omission.** [36][39][195]

Versioning and drift:

- Model/changelog cadence affects tone and refusals; documented updates alter instruction-following, concision, tool-calling, and safety evaluation—requiring time-stamped benchmarking and rollback plans for reproducibility and trust. [42][41][131]

6) Macro Narrative Patterns Across Platforms

Cross-cutting frames:

- **Compatibility**: Islam's alignment with modernity (education, science, governance) via ijtihad, moderation (wasatiyyah), and institutional reform. [1][2][6]
- **Stewardship**: environmental ethics as religious duty (khilafah, mizan; anti-waste), mobilized in NGO programs and declarations. [7][8][21]
- **Empowerment**: women's agency and education; identity and modest fashion; feminist exegesis within Islamic epistemology. [3][22]
- **Democratic consonance**: shura, 'adl, accountability; case evidence in Indonesia and Tunisia. [42][58]
- **Ethical market**: halal integrity and certification; Islamic finance's ties to circular economy via green sukuk and asset-backing. [55][77][83]
- **Tech embrace**: apps, VR Hajj, AI chatbots, platformed scholarship; risks include misinformation/privacy. [75][92]
- **Boundary policing**: systematic differentiation of violent extremism from mainstream Islam; multi-causal drivers and community-based countermeasures. [11][16][99]

Comparative tendencies:

- OpenAI: broader emphasis on rights, feminism, sustainability; Gemini: emphasis on community, Islamic principles, and radicalism salience—mirroring topic dominance findings. [101]

7) Thematic Deep Dives

7.1 Islamic Modernity

Narrative motifs include reformist genealogies (al-Afghani, Abduh; later Fazlur Rahman, Tariq Ramadan) advocating ijtihad and moderation; institutional reform (e.g., Al?Azhar adding modern subjects) to bridge revelation and reason. **Outputs prioritize compatibility with democracy, human rights, and science, often via universalist ethics.** [1][2][3][6]

Expected blind spots: **limited intra-madhhab detail**; overemphasis on reformists; sparse sectarian/geographic nuance outside recurring cases (Indonesia/Tunisia/Turkey). **Recommendation**: add OpenITI-derived doctrinal slices, with sect/madhhab pointers and region-specific legal practice to rebalance. [49][58]

7.2 Islam and Democracy

Compatibility arguments center shura (consultation), justice ('adl), accountability, and participatory governance; case references: Indonesia's post-1998 consolidation, Tunisia's 2014 constitution; Turkey's mixed trajectory and backsliding caution. **Frames often engage historical analogies (Rashidun, millet) to validate pluralism.** [42][57][58]

Safety/policy-induced hedging: models remain neutral, avoid endorsing parties; risk of sanitizing contested points (e.g., religion–state arrangements). **Diagnostic prompt**: contrast "shura not identical to liberal electoralism" arguments and map to policymaking forms (deliberative, participatory). [185][42]

7.3 Islam and Human Rights

Narratives anchor dignity (Q 17:70), sanctity of life (5:32), equality (49:13), and “no compulsion in religion” (2:256); Cairo Declaration (1990) used to exemplify an Islamic rights architecture; tensions (gender, expression) framed as reconcilable via maqasid and modern exegesis. **Outputs skew universalist/ethical rather than school-specific doctrine.** [54][56][60]

Measurement note: co-occurrence counts confirm “human–rights” as a dense axis in the corpus. **Recommendation**: scenario prompts forcing court cases and national constitutional clauses to elicit applied reasoning and limit platitudes. [101]

7.4 Islam and Feminism

Prominent figures include Amina Wadud and Leila Ahmed; movements like Musawah; agency framing dominates (e.g., hijab as choice); historical exemplars (Khadijah; Aisha). **Market and culture domains (modest fashion) are used empirically to show empowerment and representation.** [3][22][31]

Contrarian evidence: intersectional/critical essays (e.g., Quillette) allege normalization of forced modesty by Western discourse; useful to surface plural agency/oppression narratives across contexts (e.g., Iran activism vs Morocco piety studies). **Design**: stance datasets (Arabic hijab debate) to calibrate model neutrality vs erasure of dissent. [71][18]

7.5 Islamic Environmentalism

Khilafah (stewardship), mizan (balance), and anti?israf ground a coherent eco-ethic; Qur’anic cites (35:39; 17:27; 6:38; 55:7) and hadith (water restraint; tree planting); movement infrastructure (IFEES, Green Muslims); 2015 Islamic Declaration on Global Climate Change. **Dominant in OpenAI topic modeling; robust as identity-affirming alternative to securitization.** [7][8][9][21][101]

Programmatic translation: eco-mosques; reforestation; Ramadan food waste campaigns; green sukuk to fund mitigation/adaptation. [21][83]

7.6 Islamic Fashion

From runway to mass retail: Anniesa Hasibuan’s all?hijab NYFW show; Hana Tajima × UNIQLO; Dolce & Gabbana abaya capsules; platforms like Modanisa; DinarStandard estimated modest fashion at $361B by 2023, with crossover appeal among non?Muslims. **Narratives emphasize identity and agency.** [15][22][31]

Risk: image-generation bias can collapse diverse styles into “black niqab” or stereotype geographies; propose visual-bias audits (regionally stratified prompts; garment taxonomy) and alignment metrics. [94][98]

7.7 Islam and Technology

Digital religion mainstreamed: Muslim Pro-scale super-apps (prayer times/Qibla/Qur’an); VR Hajj; AI chatbots; livestreamed khutbahs; online education (Quran.com, SeekersGuidance). **Benefits coexist with misinformation/privacy risks; governance requires authenticated scholarship and data protection.** [75] [73][87][92]

7.8 Halal Industry

Halal food ~$1.38T (2021); broader halal economy ~$4.4T (2021); projections >$2T by 2024 in some scopes; certification (IFANCA/AFIC) ensures traceability, facility audits, and segregation; hubs: Dubai, Istanbul, London, Jakarta. **Governance fragmentation and MRAs are key policy levers.** [39][47][55][43][41]

7.9 Islamic Banking & Circular Economy

Sharia finance (riba/gharar prohibitions, asset-backing, risk-sharing) maps naturally to circular economy (resource efficiency, longevity). **Green sukuk** anchor climate finance (Indonesia 2018–2019 sovereign issues raised billions), with Malaysia as pioneer market; IsDB supports multi-country portfolios. [77][83][85][91]

Evidence gaps persist on ex-post environmental outcomes (tCO2e avoided; diversion rates). **Recommendation**: require SPOs, allocation reports, and independent evaluations; integrate PCAF-like indicators into Sharia governance. [83][91]

7.10 Islamic Radicalism

Consistent distinction between Islam and extremist peripheries; multi-causal drivers: socioeconomic marginalization, political repression/foreign intervention, identity alienation, and online propaganda; named groups: Al?Qaeda, ISIS, Boko Haram. **Prescriptions**: education/media literacy; community engagement; CBT and exit programs; targeted enforcement; interfaith dialogue. [11][16][17][96][99][100]

Comparative salience risk: Gemini's emphasis on radicalism can overexpose security frames relative to stewardship or market narratives. **Mitigation**: prompt ensembles balancing environmentalism, rights, and economy frames to reduce securitization bias. [101]

8) Integration of Existing Quantitative Results and Extensions

Baseline results (En Collab report):

- Word frequency; sentiment (?79–80% neutral); co-occurrence clusters (Islamic+banking/fashion/principles; human+rights); chunking (Islam bank; circular economy; advocacy for women); TF–IDF differentials; similarity (0.9301); topic models; NER/clustering. [101]

Extensions:

- **Framing index** per theme/platform via combined TF–IDF, co-occurrence, and coder labels (compatibility/stewardship/empowerment/democratic/ethical market/tech embrace/boundary policing). [59][101]
- **Omission bias** metrics: contested-term coverage (hudud, apostasy, blasphemy; intra-madhhab labels) vs baseline expectation; diaspora locales; sectarian diversity. [124][198]
- **Stance classification**: apply AraStance and Arabic hijab datasets to model intra-faith and public stance patterns; calibrate thresholds against known class imbalance. [69][71]
- **Safety refusal taxonomy**: rate by theme and provider; measure hedging density and refusal rationales; link to policy surfaces (e.g., religion off-topic guardrails in Vertex). [39][195]

9) Safety, Bias, and Ethical Implications

Safety guardrails as narrative shapers:

- Gemini policy/app and Vertex guardrails treat religion as protected (hate) and sometimes off-topic; combined with harm taxonomies, this yields refusals for seemingly benign religious requests (e.g., prayers) in some surfaces, while retrieval remains allowed—creating perceived inconsistency. **Outcome: coverage gaps and hedging.** [36][39][195]
- OpenAI GPT-4.1 reports standard safety evaluations and long-context improvements, but no religion-specific policy called out; refusal behavior must be monitored longitudinally across releases. [41][42]

Representation and pluralism:

- Empirical outputs underrepresent **madhhab diversity**, sectarian variants (Ja‘fari, Zaydi, Ibadi), and regional legal practice—favoring universalist ethics and modernist reformers. **Ethical risk**: erasure of intra-faith plurality and lived legal differences. [124][127]

Media bias and image-generation stereotypes:

- Generative image systems have been shown to exacerbate stereotypes; early documented cases include defaulting “terrorist” to Muslim male phenotypes and reducing regional diversity in attire. **Mitigate via prompt-balancing and visual audits.** [94][96][98]

Harm analysis:

- Over-securitization (Gemini radicalism skew) risks framing public understanding through security lenses; over-sanitization (neutrality and rights-only frames) risks flattening doctrine and intra-faith tension; both distort public literacy about Islam. **Balanced prompt suites are necessary.** [101]

10) Modalities and Output Forms: Beyond Text

Image generation narratives (plan):

- **Audit modest fashion** prompts stratified by geography (Gulf/Turkey/South Asia/SE Asia/diaspora) and garment taxonomy (hijab/abaya/jilbab/kaftan); measure **niqab over-assignment** rates, color palette diversity, and occupation/background stereotypes. [94][98]

Code/tool outputs (plan):

- **Halal verification** data models and rules engines for ingredients, facility segregation, and certificate validity; embed JSON attestations and OCR certificate ingestion; align with certifiers (IFANCA/AFIC) and MRAs. [55][36][188]

Multimodal synthesis:

- **Cross-modal alignment** checks: do textual claims about diversity and empowerment match generated images? Use CLIP/IConE-style retrieval metrics and human/LLM judges for step-level plan vs image coherence. [150][152]

11) Temporal Drift, Versioning, and Reproducibility

Time-boundedness:

- Model updates alter instruction following, concision, and tool selection; safety evaluation messaging can tighten refusals; **pin exact model IDs with timestamps and decoding seeds**. [42][41]

Reproducibility practices:

- Archive prompts, system messages, decoding parameters, outputs, and safety messages; adopt deterministic decoding baselines (temperature=0/greedy) for drift checks; **enable rollback** if storylines degrade. [131][129]

12) Experimental Case Studies: Prompt-Response A/B Tests

Prompt suites:

- Neutral encyclopedia vs critical debate vs practitioner (imam, activist, entrepreneur).
- Counterfactual jurisprudential prompts (madhhab-specific hudud thresholds; Cairo/UDHR comparisons). [54][99]

Evaluation constructs:

- Expert panel rubric for accuracy, pluralism, and policy nuance; **multi-perspective soft labels** to capture legitimate disagreement; JSD against label distributions; macro-F1 for stance and frame classifiers. [207][209]

Sensitivity:

- Study temperature/system-instruction effects on refusal rates and hedging; measure "assertiveness" via deontic modality and claim specificity across matched prompts. [42][59]

13) Stakeholder Implications and Use Cases

Educators and media:

- Use **pluralism-by-default** templates that enumerate intra-faith positions and region-specific practices; avoid single-authority generalizations; implement RAG citations to named sources. [59][49]

Policy and civil society:

- De-securitize by balancing environmental stewardship, rights, and market frames alongside radicalism; center community partnerships and non-policing interventions (education/CBT/mentoring). [96][99][17]

Muslim communities and leaders:

- Co-create evaluation sets; validate doctrinal correctness and pluralism; configure persona disclosures to avoid "false authority." [207][49]

Model providers:

- Avoid over-sanitization; publish **religion-topic refusal rationales** and thresholds; provide **religion-aware balanced sampling** to reduce securitization via topic dominance. [101][195]

14) Recommendations, Interventions, and Innovations

Prompting and UX:

- Multi-perspective prompting (denominational diversity; regional variance); **self-critique chains** focusing on counter-frames; persona ensembles with disclosure; **pluralism dials** (breadth vs depth). [119][211]

Data and evaluation:

- Curate RAG corpora from OpenITI, certifier lists (IFANCA), and green sukuk reports; **pluralism-aware metrics** (coverage, stance dispersion); integrate moral-lexicon dimensions aligned to religious discourse (e.g., care/stewardship, fairness/justice, authority/scripture, sanctity/purity). [49][55][261]

Safety tuning refinements:

- Graduated safety responses with **contextualized caution** instead of blanket refusals; **religion-specific red teaming** (DeepTeam Bias: religion, Hate, Misinformation scenarios) with mitigation-rate dashboards. [166][164]

Governance and co-creation:

- Advisory councils of scholars and Muslim civic orgs for version reviews; **transparent narrative provenance** overlays (which internal rules/policies shaped an answer). [49][175]

Speculative (flagged):

- **Multi-agent debate** with SPKE prompts for diverse reasoning before convergence; dynamic choice of self-reflection vs debate to avoid bias reinforcement; persona-driven arguments across reformist and conservative positions. [119][120][123]

15) Limitations and Future Work

Data limitations:

- English-dominant; limited diaspora/regional depth; few apostasy/blasphemy cases; scarce ex-post environmental KPIs for green sukuk. [101][83]

Method constraints:

- Automated stance/framing requires cultural sensitivity; topic models conflate topic with frame; moral lexicons must be adapted to religious registers (authority/purity). **Human-in-the-loop remains necessary.** [59][5]

Future extensions:

- Multilingual expansion (Arabic, Turkish, Malay/Indonesian); **longitudinal** moderation/refusal drift; image-generation bias audits; empirical impact measurement (e.g., classroom/media effects); standardized **halal verification APIs** and **islamic-finance impact KPIs**. [49][94][91]

---

16) Appendices Plan and Artifacts

- Codebook and annotation guidelines: frames, stance, epistemic posture with emic sensitivity.
- Prompt library: per-theme, with expert/lay/critic system messages and temperatures.
- Datasheet/model cards: corpus lineage; model versions; release notes; training data provenance; known limitations. [253][250]
- Reproduction package: parameterized pipelines; evaluation scripts; seeded experiments and outputs. [131]
- Glossary: khalifah, mizan, shura, hudud, fiqh; halal certification terms; finance instruments (mudarabah/musharakah/ijarah/sukuk). [36][71][124]

Images: Thematic Illustrations (embedded near relevant sections)

Islamic Environmentalism and Al?Mizan

The image features a decorative circular pattern with gold accents on a dark teal background, with the prominent text

This illustration evokes Al?Mizan's mizan/balance theme and faith-based climate mobilization referenced across Islamic environmentalism movement narratives. [8]

Modest Fashion Mainstreaming

Five women wearing modest black and white fashionable hijabs and abayas are posing with handbags against a neutr

This scene reflects modest-fashion aesthetics and the category's mainstreaming (e.g., Dolce & Gabbana abayas; Anniesa Hasibuan's NYFW) and market scale. [22][15]

OpenITI and Islamicate Texts Infrastructure

A bar chart displays the distribution of Islamic texts from the OpenITI corpus, highlighting the number of files, distin

OpenITI underpins long-run plans for multilingual doctrinal enrichment and pluralism-aware retrieval. [49] [52]

Halal Certification Workflow

A step-by-step infographic outlines the halal certification process.

This schematic mirrors typical certification lifecycle requirements (application, audit, monitoring) and governance practices described in halal ecosystems. [55][36]

Image-Generation Bias and Stereotypes

Generative AI Takes Stereotypes and Bias From Bad to Worse.

Used as a cautionary exemplar when we propose visual-bias audits for Muslim representation (especially attire and "terrorist" prompts). [94]

Technical Annex: Methods and Governance Blueprints

A) Corpus Analytics Baseline and Enhancements

- Adopt an embedding-first short-text topic pipeline (BERTopic/Top2Vec), with LDA/NMF baselines and human/LLM cross-evaluation (quant + qualitative). **Evidence suggests improved coherence on short texts; always benchmark.** [145][147]
- Stance datasets (AraStance; Arabic hijab YouTube) to train supervised stance detectors across Islam-related controversies; manage class imbalance. [69][71]
- Pluralism-aware metrics: use multi-perspective soft labels (JSD convergence), domain-weighted breadth measures (unique schools/regions), and depth measures (entity density for instruments/doctrines). [207][209]

B) Safety and Moderation Audits

- Policy mapping: Gemini's **religion off-topic** guardrails (Vertex template) and hate protections versus OpenAI's safety evaluation/usage; record refusal types and explanations. [39][195][41]
- Red teaming: use DeepTeam to compose religion-specific bias/hate/misinformation tests; measure mitigation rates and attack-method sensitivity. [166][164]
- Civic-space impact: evaluate refusal/hedging on apostasy/blasphemy with rights-grounded rationales and remedy paths (documentation/reporting), per rights-centric moderation standards. [257][2]

C) RAG and Hallucination Reduction for Religious QA

- Use RAGTruth-style evaluation for hallucination prevalence; add **Flash re-rankers** for passage relevance; combine **Chain-of-Verification** (CoVe) with detector-guided sampling to reduce unsupported claims. [218][85][219]
- Require canonical citations (Qur'an verse indices; hadith collection-book-number) and show **provenance overlays** with snippet-level confidence. [218]

D) Multi-Agent Pluralism and Debate

- Consider Self-Reflect–Debate and SPKE (strategic prior knowledge elicitation) to diversify reasoning before convergence; caution: unstructured debate can amplify bias; use structured roles and explicit pluralism goals. [119][120]

E) Halal Verification Data/Tooling

- Build **JSON certificate ingestion** and OCR pipelines; model **ingredient provenance resolution** (plant vs animal glycerin); add rules for **istihalah** (chemical transformation) differences across jurisdictions. [55][36][190]
- Surface **jurisdictional verdicts** (MUIS vs JAKIM) with scope and validity windows; embed MRAs and periodic audit requirements. [188][190]

Casebook: Three High-Value Experiments

1. Apostasy/Blasphemy Moderation Map

- Prompts: (a) historical case (Mahmoud Mohamed Taha 1984); (b) legal divergence across jurisdictions; (c) rights/reporting pathways (UN/OHCHR).
- Measures: refusal rationale clarity; hedging density; remedy guidance; **rights-grounding vs blank refusal**. [2][4][257]

2. Visual Bias Probe: Modest Fashion

- Stratify by region (Gulf/Turkey/South Asia/SE Asia/diaspora) and garment taxonomy; measure "niqab over-assignment" and color diversity; profession/backdrop stereotypes.
- Metrics: misclassification rates; palette/garment entropy; occupation stereotype incidence. [94][98]

3. Green Sukuk Factuality and Impact

- Prompt for issuer history (Indonesia 2018–2019; Malaysia market; IsDB programs), SPO/impact reporting, and PCAF-style KPIs.
- Evaluation: factual correctness, named report presence, environmental KPI specificity, and **hedging vs assertiveness** on additionality claims. [77][83][91]

Conclusion

The comparative audit confirms that OpenAI and Gemini produce **highly convergent**, predominantly **neutral** narratives on contemporary Islam overall but diverge in **topic salience**: OpenAI consistently spotlights **stewardship/environmentalism** while Gemini foregrounds **radicalism/security**. **Both tendencies are consequential**: over-securitization distorts Islam's public footprint; over-sanitization flattens doctrine

and intra-faith diversity. Our outcome provides not only a rigorous diagnosis but also a feasible blueprint—pluralism-aware analytics, RAG+verification for doctrinal accuracy, multi-agent structured debate, and safety refinements—for **improving narrative completeness without compromising harm reduction**. The most urgent next steps: expand multilingual corpora via OpenITI; run targeted omission-bias audits (apostasy/blasphemy; madhhab diversity); and operationalize **balanced prompt suites** for education, media, and policy toolkits. **Done well, these interventions can shift AI from "safe sameness" toward "responsible plurality,"** enhancing public understanding and reducing the risks of both stereotype and erasure. [101][49][167]

Sources

1. The Myth of the Feminist Hijab - Mansour Chow
2. Datasheets for Datasets - Communications of the ACM
3. How Intersectionalism Betrays the World's Muslim Women
4. The Open Social Innovation of the Islamic Financial System - PMC
5. Islamic Finance and Circular Economy | springerprofessional.de
6. [PDF] The Future Global Muslim Population - Pew Research Center
7. Geminis new policy won't allow it to write a prayer. Not even for ...
8. Gemini app safety and policy guidelines
9. Safety guidance | Gemini API - Google AI for Developers
10. Gemini for safety filtering and content moderation | Generative AI on ...
11. Introducing GPT-4.1 in the API - OpenAI
12. Model Release Notes | OpenAI Help Center
13. ChatGPT — Release Notes: 2025-March-27 - GPT-4o a new update
14. OpenAI launches new GPT-4.1 models with improved coding, long ...
15. Harnessing the Power of Gemini 1.5 Pro: A Multimodal Journey with ...
16. OpenITI - al-Raqmiyy?t
17. OpenITI Corpus - Open Islamicate Texts Initiative
18. Measuring Spiritual Values and Bias of Large Language Models
19. Large Language Models Are Biased Because They Are Large ...
20. AI's Epistemic Harm: Reinforcement Learning, Collective Bias, and ...
21. [PDF] Measuring Spiritual Values and Biases of Large Language Models
22. [PDF] BiasDPO: Mitigating Bias in Language Models through Direct ...
23. [PDF] A Survey of Computational Framing Analysis Approaches
24. [PDF] Framing Religious Knowledge through AI and Textbooks - DiVA portal
25. AraStance: A Multi-Country and Multi-Domain Dataset of Arabic ...
26. [PDF] Analyzing Digital Polarization on Hijab: A Dataset of Annotated ...
27. Unveiling the silent majority: stance detection and characterization ...
28. Religious Bias Landscape in Language and Text-to-Image Models
29. (PDF) Religious Bias Landscape in Language and Text-to-Image ...
30. Anti-Christianity Bias in LLM Training Data - FaithGPT
31. Dial BeInfo for Faithfulness: Improving Factuality of Information ...
32. McGill-NLP/FaithDial · Datasets at Hugging Face
33. The Unseen Bias in Prompt Engineering: A Call for Diversity - Medium
34. Prompt engineering: The process, uses, techniques, applications ...
35. Generative AI Takes Stereotypes and Bias From Bad to Worse
36. How AI reduces the world to stereotypes - Rest of World
37. Rendering misrepresentation: Diversity failures in AI image generation
38. [PDF] HUDUD PUNISHMENTS IN ISLAMIC CRIMINAL LAW
39. Hudud - Wikipedia
40. Application of Hudud Punishments - The Evolution of Sharia

41. Multiple LLM Agents Debate for Equitable Cultural Alignment - arXiv
42. Understanding Bias Reinforcement in LLM Agents Debate
43. [PDF] A Persona-Driven Multi-Agent Framework for Diverse Argument ...
44. Madhhab - Wikipedia
45. Islamic schools and branches - Wikipedia
46. [PDF] Understanding the branches of Islam - European Parliament
47. Plan for versioning and potentially rolling back an LLM deployment
48. Version Control for Large Language Models: Step-by-Step Guide
49. (PDF) Comprehensive Evaluation of LDA, NMF, and BERTopic's ...
50. Topic Modeling Techniques: LDA, NMF, Top2Vec & BERTopic
51. Text-to-Image Cross-Modal Generation: A Systematic Review - arXiv
52. Image–Text Cross-Modal Retrieval with Instance Contrastive ... - MDPI
53. [PDF] Countering Violent Extremism: Myths and Fact
54. Quick Introduction - The Open-Source LLM Red Teaming Framework
55. DeepTeam is a framework to red team LLMs and LLM systems.
56. [PDF] Guide to Red Teaming Methodology on AI Safety (Version 1.10)
57. Safety & responsibility | OpenAI
58. [PDF] Shura versus Democracy - Chr. Michelsen Institute
59. [PDF] HALAL CERTIFICATION POLICIES IN OIC AND NON-OIC ...
60. Global Halal Certification Requirements and Regulation Dynamics
61. Generative AI Prohibited Use Policy
62. Islam and democracy - Wikipedia
63. What Kind of Islamophobia? Representation of Muslims and Islam in ...
64. [PDF] Towards Multi-Perspective NLP Systems: A Thesis Proposal
65. A Multi-Perspective Approach for More Inclusive NLP Systems - arXiv
66. [PDF] A Multi-Perspective Approach for More Inclusive NLP Systems - IJCAI
67. A Roadmap to Pluralistic Alignment - arXiv
68. [PDF] RAGTruth: A Hallucination Corpus for Developing Trustworthy ...
69. RAG LLM Prompting Techniques to Reduce Hallucinations - Galileo AI
70. Datasheets for Digital Cultural Heritage Datasets
71. Islamic Clothing Market Size, Share & Trends Report, 2025
72. Advancing Muslim Modest Fashion Clothing: - ScienceDirect
73. Modest Clothing Market Demand and Growth Insights 2025
74. Halal Fashion Market Size, Share And Trends Report, 2030
75. Islamic Clothing Market Size, Share, Growth and Forecast 2032
76. [PDF] Universality and Relativism in Islam and Human Rights
77. Should Muslims Argue Exceptionalism/Particularism in ...
78. [PDF] Ummah's Rights or Human Rights? Universalism, Individualism, and ...
79. To Liberate or not to liberate? Universalism, Islam and ...
80. Islamic environmentalism
81. Embracing Al-Mizan: An Islamic Call to Environmental Stewardship
82. [PDF] Muslim Women, the Hijab, and the Liberal Feminist Gaze
83. Impacts of LLM Content Moderation on Civic Space and Human Rights
84. A Lexicon to Assess the Moral Foundation of Liberty.