DOI: https://doi.org/10.63056/jllsa.1.2.2025.458

Journal of Language, Literature & Social Affairs

journal homepage: https://scholarclub.org/index.php/jllsa

Algorithmic Theology or Stochastic Parrot? Evaluating the Epistemic Authority and Theological Accuracy of Generative AI in Global Religious Education

Hafiz Syed Adeel Shah Bukhari*[a], Muhammad Aitzaz[b], Amna Ali Akbar[c]

a. Department of Islamic Studies, Education University Lahore, Pakistan
b. Bachelor in Computer Science (BSCS), Department of Computer Science. BSCS, National College Of Business and Administration
c. Division of Islamic and Oriental Learning

ARTICLE INFO

Received:
May 21, 2025
Revision Received:
June 25, 2025
Accepted:
June 27, 2025
Available Online:
June 30, 2025

Keywords:
generative artificial intelligence; religious education; epistemic authority; theological correctness; big language models; stochastic parrots; digital religion; algorithmic bias; interfaith pedagogy.

ABSTRACT

Generative artificial intelligence (AI) systems based on large language models (LLMs) are fast infiltrating classrooms, seminaries, madrasas, Sunday schools, and secular religious studies courses, where they are consulted on scripture, dogma, ritual, and ethics. Whether such systems should be viewed as sources of religious knowledge, as sophisticated pattern-completion tools, or as anything in between remains unanswered. This paper explores to what degree generative AI may claim epistemic authority and theological correctness in global religious education. Using a critical interpretive review of peer-reviewed and scholarly literature from natural language processing, philosophy of testimony, digital religion, and religious education, supported by a structured analytic framework (the Epistemic Authority and Theological Accuracy framework, EATA), the study synthesises evidence across four dimensions: textual and doctrinal fidelity, intra-traditional plurality, pedagogical appropriateness, and epistemic transparency. The synthesis yields five themes: the dissociation of fluency from accuracy; representational skew toward dominant languages and traditions; the flattening of internal religious diversity; the displacement of traditional authority structures; and the partial, conditional promise of technical mitigation. The paper argues that generative AI has simulated rather than justified epistemic authority: it may be a helpful teaching device when positioned as a flawed, source-anchored interlocutor, but it cannot operate as a theological authority in any tradition's own terms. Implications for teachers, curriculum designers, religious groups, developers, and policymakers are highlighted, combined with a research agenda for tradition-specific empirical audits.

1. Introduction

For much of human history, religious knowledge has been transferred via relationships: parents, teachers, academics, clergy, elders, and the texts and communities they interpret. The authority of these persons based on acknowledged training, lineage, collective responsibility, and, in many traditions, moral and spiritual upbringing. The introduction of generative AI catboats indicates a breach in this trend. A student in Lahore, Lagos, São Paulo, or Leeds may now raise a question on the Qur'anic idea

* Correspondence to: Department of Islamic Studies, Education University Lahore, Pakistan
E-*mail address:* muhammadadeelgailani@gmail.com (H. S. A. S. Bukhari)

3105-4919 & 3105-4927/© 2025 The Authors. Published by Scholar Club. This is an open access article.

of tawhid, the Buddhist teaching of anattā, the Catholic theology of grace, or the Hindu notion of dharma and get within seconds a fluent, confident, well-organized answer.

The fluidity of these replies is exactly what makes them both alluring and perilous. Large language models are trained to anticipate plausible continuations of text (Vaswani et al. 2017; Brown et al. 2020) and are then tweaked to provide helpful, human-preferred replies (Ouyang et al. 2022). Critics have labelled such systems as "stochastic parrots" that sew together language structures without consideration to meaning (Bender et al. 2021; Bender and Koller 2020). Others contend that the picture is more convoluted and that some type of emerging competence or partial knowledge cannot be discarded out of hand (Mitchell and Krakauer 2023). Religious education lies at the core of this argument because its substance is not only informative. Religious claims are interwoven in behavior's, groups, interpretative disciplines, and, for believers, claims about ultimate truth and responsibility. An error in a chemical explanation is a mistake; an error in the explanation of a religious book may be perceived as deception, contempt, or even blasphemy.

The issue presented in this paper is thus both epistemological and pedagogical. Can generative AI be considered to possess epistemic authority, that is, a standing that offers learners strong cause to accept its religious claims? And how accurate, in theological terms, are the outputs that it provides in the multi-faith, multilingual environments that define global religious education? The title's contrast, algorithmic theology vs stochastic parrot, is purposely provocative. It presents two competing intuitions: that AI may become a new form of interpreter of tradition, or that it is only an echo of the material on which it was taught. This essay argues that neither caricature is appropriate and presents a paradigm for judging the gap between them.

The article offers three contributions. First, it brings together literatures that are seldom studied together: technical study on language-model reliability and bias, philosophical work on witness and authority, and scholarship in religious education and digital religion. Second, it presents the EATA framework as a method for organizing assessment of AI in religious education. Third, it extracts practical consequences and a research agenda for educators, religious groups, developers, and politicians.

2. Literature Review

2.1. Epistemic authority and testimony

Philosophers of testimony have long understood that much of what any individual knows relies on the word of others (Hardwig 1985; Lackey 2008). Epistemic reliance is reasonable when the source is dependable and when the learner has appropriate grounds to perceive it as such. Zagzebski (2012) develops this idea into a theory of epistemic authority, arguing that a person may reasonably defer to another when doing so is more likely to produce true belief than relying on their own judgement, and that authority is grounded in the authority's conscientious pursuit of truth and its answerability for error. Raz (1986) proposes a parallel "service conception" in which power is justified by its potential to assist others conform to reasons that already apply to them. Goldman (1999) adds a social-epistemological approach, judging sources by their dependability in developing correct beliefs within a specific information context.

Applied to AI, these accounts pose a fundamental challenge. Reliability is an empirical attribute, but accountability and conscientiousness are not clearly accessible to a statistical model. Some researchers have suggested that LLM outputs are better understood as indifferent to truth than as falsehoods or mistakes, since the systems are not meant to monitor truth at all (Hicks, Humphries, and Slater 2024; Frankfurt 2005). Floridi and Chiriatti (2020) also emphasize that the ability to write semantically compelling writing should not be misconstrued for comprehension or agency. Messeri and Crockett (2024) warn more widely that AI systems may build "illusions of understanding" among their users, an insight that is particularly pertinent to rookie learners who lack the background to spot plausible mistakes.

2.2. Religious authority in digital contexts

Scholarship in digital religion has revealed that new media do not just communicate religious information; they rearrange who counts as an authority (Campbell 2010, 2013; Cheong et al. 2012). Campbell's study on religious bloggers shows that internet platforms may both threaten and be domesticated by established religious hierarchies, with communities determining whether new voices are acceptable. Generative AI deepens this dynamic: it is a non-human interlocutor that does not belong to any community, has no tradition-recognized qualifications, and yet replies in the register of a teacher. Parallel religious thought has started to examine the nature of machine intelligence. Herzfeld (2002) explored AI in light of the Christian doctrine of the imago Dei, Geraci (2010) traced the religious imaginaries embedded in AI discourse, and the Vatican's note Antiqua et Nova (Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education 2025) distinguishes human intelligence from machine information-processing and cautions against attributing to AI capacities proper to persons. Such writings demonstrate that religious groups are themselves actively theorizing the position of AI, while they do not represent empirical appraisals of its effectiveness.

2.3. Reliability, delusion, and veracity in language models

The technical literature demonstrates that LLMs may deliver eloquent but unsubstantiated or erroneous assertions, frequently called hallucinations (Ji et al. 2023). Benchmarks such as TruthfulQA show that models may reproduce popular misconceptions because these are statistically prominent in training data (Lin, Hilton, and Evans 2022), and broad knowledge benchmarks such as MMLU, which include humanities items, measure aggregate performance rather than tradition-specific accuracy (Hendrycks et al. 2021). Weidinger et al. (2022) propose a taxonomy of hazards that includes disinformation damages, discrimination and exclusion, and over-reliance. Retrieval-augmented generation, which conditions model outputs on retrieved documents, has been proposed as one route to improved factual grounding (Lewis et al. 2020), while reinforcement learning from human feedback shapes tone and helpfulness but does not by itself guarantee correctness (Ouyang et al. 2022).

2.4. Bias and representation

A considerable amount of research reveals that language technology incorporate social biases and uneven representation. Blodgett et al. (2020) challenge the conceptual ambiguity of "bias" in NLP research and demand for attention to the power dynamics at issue. Abid, Farooqi, and Zou (2021) found persistent anti-Muslim bias in GPT-3, which identified Muslims with violence at rates greater than for other religious groups in the studied prompts. Santurkar et al. (2023) and Durmus et al. (2023) revealed that model outputs fit more closely with the attitudes of particular demographic and national groupings than others. Joshi et al. (2020) documented the extreme concentration of NLP resources in a handful of languages, a finding with obvious implications for religious traditions whose primary texts and living scholarship are in languages such as Arabic, Sanskrit, Pali, Classical Chinese, Hebrew, Ge'ez, or Gurmukhi. These results parallel prior criticisms of how Western knowledge systems describe religious "others" (Said 1978).

2.5. Religious education and pedagogy

Religious education is a plural field. Approaches span from confessional upbringing within a tradition to non-confessional, comparative, and worldview-based research (Smart 1983; Jackson 2004, 2014). Jackson's interpretative approach highlights the internal variety of traditions, the depiction of insiders' perspectives, and the rejection of essentialist interpretations. Pedagogical topic knowledge, the teacher's ability to turn subject matter into forms accessible to learners (Shulman 1986; Mishra and Koehler 2006), is fundamental to excellent teaching and is not reducible to information retrieval. General evaluations of AI in education highlight to potential like as customization and accessibility, with dangers including inaccuracy, prejudice, and academic integrity (Kasneci et al. 2023; Dwivedi et al. 2023; UNESCO 2023). Yet these studies examine education generally and offer nothing about topic areas in which the material is holy, contentious, and tradition-specific.

2.6. Research gap and goal

Taken together, the evidence indicates that LLMs are fluent but flawed, that they display representational asymmetries, and that religious education involves truth, multiplicity, and respect for insider viewpoints. However, little is known about how these general properties translate into tradition-specific theological accuracy, how the epistemic authority of AI should be conceptualized in religious learning contexts, and what evaluative criteria educators in different faith and non-faith settings might use to judge AI outputs. Little is also known about how learners' belief in AI interacts with existing frameworks of religious authority across places.

The aim of this study is therefore to (a) conceptualize epistemic authority as it applies to generative AI in religious education, (b) synthesize existing evidence on the accuracy and representativeness of LLM-generated religious content, and (c) propose and apply an analytic framework for evaluating such systems. The research is led by the following working premises, which are verified interpretively against the literature rather than statistically:

- **P1.** Fluency and seeming confidence in LLM results are not good indications of theological truth.
- **P2.** Accuracy and representativeness change consistently with the language, tradition, and importance of the topic matter in training data.
- **P3**. Generative AI lacks the answerability that bases epistemic authority in most religious and philosophical theories of authority, and hence may occupy at best a derivative, conditional position in religious teaching.

3. Research Methodology

3.1. Design

The research utilizes a critical interpretative review design. Such a design is suitable to emergent, multidisciplinary themes in which the evidence base is distributed across areas and in which conceptual explication is as vital as empirical aggregation

(Arksey and O'Malley 2005; Tranfield, Denyer, and Smart 2003). The paper does not disclose a new experimental dataset. It synthesises published evidence and adds a systematic framework to it.

3.2. Search and selection

Sources were identified through targeted searches of Google Scholar, Scopus, the ACM Digital Library, the ACL Anthology, Phil Papers, and ATLA Religion Database, using combinations of terms such as large language model, generative AI, ChatGPT, religion, religious education, theology, epistemic authority, testimony, bias, and hallucination. Searches were augmented by backward and forward citation monitoring from significant literature. Reporting follows the transparency criteria of PRISMA 2020 (Page et al. 2021) modified for a narrative, conceptual evaluation. Inclusion criteria were: (i) peer-reviewed publications, academic monographs, conference papers from recognized venues, or formal institutional documents; (ii) relevance to LLM reliability, bias, epistemic authority, digital religion, or religious education; and (iii) availability in English. Exclusion criteria were non-scholarly comments and sources whose assertions could not be confirmed. A drawback of the English-only requirement is addressed in Section 6.3.

3.3. Analytic framework: EATA

The Epistemic Authority and Theological Accuracy (EATA) paradigm was constructed by merging the epistemological and pedagogical literatures mentioned above. It analyses AI-generated religious material along four dimensions:

- Textual and doctrinal fidelity: if statements regarding scripture, theology, history, and practice are true by the criteria of accepted scholarship and the tradition's own sources.
- Intra-traditional plurality: if the result respects real internal variety (schools, denominations, jurisprudential and philosophical traditions) rather than presenting a single flattened stance.
- Pedagogical appropriateness: if the output matches the educational setting (confessional or non-confessional), avoids proselytizing or disparagement, and supports rather than substitutes learner inquiry.
- Epistemic transparency: if the system expresses ambiguity, acknowledges reliable sources, and separates description from endorsement.

Each dimension was related to the three propositions: fidelity and transparency to P1, plurality and fidelity across languages to P2, and transparency and pedagogical appropriateness to P3.

3.4. Analysis

Included materials were read in full and coded inductively and deductively using reflexive thematic analysis (Braun and Clarke 2006), using EATA dimensions as sensitizing ideas and with openness to emerging themes. Themes were improved by iterative comparison across disciplinary literatures. To boost credibility, assertions were preserved only if they were backed by a primary source, and inferential steps from general NLP results to religious settings are expressly noted as such in the sections below.

4. Results

The thematic synthesis yielded five themes. The data foundation for general language-model behavior is extensive; the evidence base particular to religious education is low, and this imbalance is itself a consequence.

4.1. Theme 1: Fluency is divorced from accuracy

Across the technical literature, LLMs develop cohesive content independent of factual foundation. Ji et al. (2023) catalogue hallucination across tasks, while Lin, Hilton, and Evans (2022) demonstrate that bigger models may be less honest on hostile issues because they more closely mirror human misconceptions. Hicks, Humphries, and Slater (2024) suggest that outputs are best understood as indifferent to truth. Inference to religious settings is easy but should be labelled as inference: a model may credit a statement to the incorrect scripture, blend diverse doctrinal perspectives, or fabricate citations in a way that is stylistically indistinguishable from accurate replies. Because religious writings are often cited, paraphrased, and misattributed on the open web, common misunderstandings are likely to be well represented in training data.

4.2. Theme 2: Representation is unequal across languages and customs

Evidence of bias against Muslims in GPT-3 (Abid, Farooqi, and Zou 2021), of opinion skew toward certain groups (Santurkar et al. 2023; Durmus et al. 2023), and of language disparity (Joshi et al. 2020) converges on the conclusion that models do not reflect the world's faiths equitably. Traditions that are numerically tiny, historically marginalized, oral, or textually concentrated in low-resource languages are likely to be defined with less detail and greater dependence on outsider concepts.

This pattern connects with long-standing criticisms of Orientalist representation (Said 1978) and with the field's concern about the power dynamics of prejudice (Blodgett et al. 2020).

4.3. Theme 3: Internal plurality tends to be flattened

Religious education studies maintains that traditions are internally varied, historically dynamic, and lived differently by people (Jackson 2004, 2014). Models designed to offer a single, useful, short response (Ouyang et al. 2022) encounter a structural conflict with this demand. A answer to "What does Islam say about X?" or "What do Christians believe about Y?" risks reducing Sunni, Shia, Ibadi, and Sufi viewpoints, or Catholic, Orthodox, and Protestant stances, into a general synthesis that represents the majority or most-documented opinion. This subject hinges mostly on the convergence of the educational literature with known aspects of preference-tuned models; direct, systematic assessment across traditions is notably missing.

4.4. Theme 4: Authority is displaced rather than just augmented

The digital religion literature reveals that new media affect views of who may speak for a tradition (Campbell 2010, 2013; Cheong et al. 2012). Epistemic dependency on a source is logical only when its dependability can be fairly judged (Hardwig 1985; Zagzebski 2012). Novices in religious learning are, by definition, poorly positioned to judge dependability, which renders them prone to the "illusions of understanding" defined by Messeri and Crockett (2024). Because AI gives quick, private, non-judgmental access, it may supplant human instructors and community responsibility, not merely enhance them. Importantly, institutional religious responses such as Antiqua et Nova (2025) clearly avoid associating machine outputs with human knowledge, demonstrating that at least some traditions' own standards of authority exclude AI from the position of authoritative interpreter.

4.5. Theme 5: Technical mitigation is promising but conditional

Retrieval-augmented generation may anchor outputs in curated corpora (Lewis et al. 2020), which in theory allows tradition-approved or scholar-reviewed libraries to confine answers. Yet retrieval quality, corpus selection, and translation choices incorporate their own authority decisions: who curates the corpus, and in whose language and school of interpretation? Alignment by human input increases apparent helpfulness (Ouyang et al. 2022) but, as the larger risk literature finds, does not remove misinformation or exclusion effects (Weidinger et al. 2022). Mitigation thus modifies the distribution of mistakes and the site of culpability; it does not remove the epistemic dilemma.

5. Findings

Integrating the five themes with the EATA framework provides the following conclusions, expressed with their evidentiary status.

Finding 1 (supports P1; strong general evidence, indirect religious evidence). Fluency and confidence are not good signs of theological correctness. The research on hallucination and honesty regularly reveals that surface quality and factual accuracy are relatively weakly related. In religious education, where learners typically lack background to check claims, this weak connection is pedagogically crucial.

Finding 2 (supports P2; moderate evidence). Accuracy and fairness are unevenly distributed. Bias, attitude skew, and linguistic disparity reported in general contexts, with clear evidence for anti-Muslim linkages, mean that learners in certain traditions and languages would acquire consistently inferior or more inaccurate representations than others. Tradition-specific audits across a broad spectrum of faiths are uncommon.

Finding 3 (somewhat supports P2; little direct evidence). Flattening of intra-traditional plurality is a feasible and logically well-motivated concern, stemming from the combination between dominant-view data and single-answer optimisation. It demands actual empirical testing.

Finding 4 (supports P3; conceptual evidence). On the theories of authority articulated by Zagzebski (2012) and Raz (1986), AI systems lack the answerability and conscientious commitment that foundation epistemic power, and numerous religious organisations expressly reject that computer outputs may carry interpretative authority. AI may thus possess simulated authority (the rhetorical form of an expert) but not justified authority (the standing that warrants respect).

Finding 5 (moderate, conditional). Source-anchored, transparent, and tradition-consulted designs may increase the educational value of AI by reframing it as an imperfect research helper rather than an oracle. The advantages rely on governance of the underlying corpus and on teacher mediation.

Overall finding. The dichotomy of algorithmic theology versus stochastic parrot is analytically useless. Generative AI is neither a theologian nor a simple echo. It is a strong, probabilistic text generator whose outputs may be helpful, wrong, or prejudiced in systematic ways, and whose instructional usefulness relies on how its fallibility is disclosed and handled.

6. Discussion

6.1. Reframing the discussion

The stochastic parrot metaphor (Bender et al. 2021) captures a key reality about the formal foundation of language models, and the work of Bender and Koller (2020) illustrates why form alone does not guarantee meaning. Yet the metaphor might mask the actual dilemma that educators face: regardless of whether a system "understands," its outputs are increasingly accurate on many major subjects, and learners will utilise them. Mitchell and Krakauer (2023) remark that the scientific community is split on whether LLMs display comprehension, which implies that educators should not rely policy on either extreme. The EATA approach transfers focus from philosophical problems about machine minds to evaluative questions about outputs and institutional arrangements: is this response faithful, plural, suitable, and transparent?

6.2. Why religious education is a particular test case

Several aspects make religion a demanding test of AI's epistemic standing. First, assertions are interpretative and frequently disputed within traditions, therefore there is no one ground truth against which benchmark accuracy may be measured. Second, religious language is performative and context-bound; a phrase may act as a prayer, a legal decision, a lyrical expression, or a doctrinal declaration. Third, authority in many traditions is relational and corporeal, based in chains of transmission, social recognition, and moral formation, elements that a software system inherently lacks. Fourth, damage is not merely informational; misrepresentation may hurt identity and social trust. For these reasons, generic educational AI guidance (UNESCO 2023) is required but inadequate for religious sectors.

6.3. Limitations

This research has significant drawbacks. As a review, it does not create new empirical measurements of model outputs; statements concerning religious content are in some areas inferred from general results and have been indicated appropriately. The English-only inclusion restriction, although needed for verifiability by the intended audience, under-represents research generated in Arabic, Persian, Urdu, Hindi, Sanskrit, Chinese, and other languages, which is exactly where tradition-internal appraisal of AI may be best improved. LLMs are also changing fast, thus system-specific results may date quickly. Finally, the author's positionality impacts the selection and interpretation of materials; the framework should be developed via discussion with insiders of each tradition.

6.4. Future research

Three priorities follow. First, tradition-specific audits should be prepared together with academics and practitioners, employing multilingual prompt sets, expert-rated rubrics linked with the four EATA criteria, and comparisons across models and versions. Second, classroom experiments should study how learners of various ages and backgrounds adjust their faith in AI religious responses and if teacher mediation impacts results. Third, comparative scholarship should explore how various religious organizations build their own standards for the validity of AI as an educational or interpretative tool.

7. Implications

For instructors and curriculum creators. AI results should be viewed as artefacts for critical scrutiny, not as authoritative material. Activities that encourage learners to check citations, compare an AI response with insider sources, and uncover missing internal views may convert constraints into learning goals. Teachers' pedagogical subject knowledge (Shulman 1986) remains essential in framing and contextualising outcomes.

For religion groups and religious establishments. Communities may desire to define specific guidelines on permissible applications of AI in religious education, to engage in the curation of authoritative corpora, and to be upfront about the distinction between informational help and interpretative authority. Dialogue between traditions may avoid the repetition of one tradition's norms on another.

For AI developers. Systems deployed for religious or educational purpose should disclose doubt, reference verified sources, exhibit internal diversity by default on contentious subjects, and be assessed on multilingual and tradition-specific standards set with community involvement. Documentation of training-data coverage for religious sources would enhance accountability.

For policymakers and educational authorities. Regulation and guidelines should acknowledge religion as a high-sensitivity educational area. Frameworks for procurement and classroom adoption should demand proof of tradition-specific assessment, in accordance with the larger need for human-centered approaches to generative AI in education (UNESCO 2023).

For researchers. The EATA framework gives a platform for comparative, cumulative assessment research. Researchers should emphasize inclusive techniques that incorporate religious insiders and that avoid perpetuating the same inequalities under research.

8. Conclusion

This paper has studied whether generative AI may be trusted with epistemic authority and theological correctness in global religious education. Through a critical interpretive review and the EATA framework, it has argued that AI-generated religious content is characterized by fluent but unreliable confidence, uneven representation, a tendency to flatten internal plurality, and a capacity to displace as well as supplement traditional authorities. On the best accessible explanations of epistemic authority, AI holds simulated rather than justified authority. At the same time, the dismissive picture of a simple parrot understates both the practical value of these systems and the intricacy of the scholarly discussion surrounding them.

The constructive route lies in placing generative AI as a transparent, source-anchored, and teacher-mediated technology within an educational environment that continues to depend on human interpreters and responsible communities. The primary task for the future years is empirical and collaborative: to design tradition-sensitive assessments that enable religious educators, communities, and developers to recognize, rather than assume, whether an algorithmic response deserves to be believed.

References

Abid, Abubakar, Maheen Farooqi, and James Zou. 2021. "Persistent Anti-Muslim Bias in Large Language Models." In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 298–306. New York: Association for Computing Machinery.

Arksey, Hilary, and Lisa O'Malley. 2005. "Scoping Studies: Towards a Methodological Framework." International Journal of Social Research Methodology 8 (1): 19–32.

Bender, Emily M., and Alexander Koller. 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, 5185–5198. Stroudsburg, PA: Association for Computational Linguistics.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. New York: Association for Computing Machinery.

Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. 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, 5454–5476. Stroudsburg, PA: Association for Computational Linguistics.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." Qualitative Research in Psychology 3 (2): 77–101.

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

Campbell, Heidi A. 2010. "Religious Authority and the Blogosphere." Journal of Computer-Mediated Communication 15 (2): 251–276.

Campbell, Heidi A., ed. 2013. Digital Religion: Understanding Religious Practice in New Media Worlds. London: Routledge.

Cheong, Pauline Hope, Peter Fischer-Nielsen, Stefan Gelfgren, and Charles Ess, eds. 2012. Digital Religion, Social Media and Culture: Perspectives, Practices and Futures. New York: Peter Lang.

Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education. 2025. Antiqua et Nova: Note on the Relationship between Artificial Intelligence and Human Intelligence. Vatican City: Holy See.

Durmus, Esin, Karina Nguyen, Thomas I. Liao, et al. 2023. "Towards Measuring the Representation of Subjective Global Opinions in Language Models." arXiv preprint arXiv:2306.16388.

Dwivedi, Yogesh K., Nir Kshetri, Laurie Hughes, et al. 2023. "Opinion Paper: 'So What If ChatGPT Wrote It?' Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy." International Journal of Information Management 71: 102642.

Floridi, Luciano, and Massimo Chiriatti. 2020. "GPT-3: Its Nature, Scope, Limits, and Consequences." Minds and Machines 30 (4): 681–694.

Frankfurt, Harry G. 2005. On Bullshit. Princeton, NJ: Princeton University Press.

Geraci, Robert M. 2010. Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality. Oxford: Oxford University Press.

Goldman, Alvin I. 1999. Knowledge in a Social World. Oxford: Clarendon Press.

Hardwig, John. 1985. "Epistemic Dependence." Journal of Philosophy 82 (7): 335–349.

Hendrycks, Dan, Collin Burns, Steven Basart, et al. 2021. "Measuring Massive Multitask Language Understanding." Paper presented at the International Conference on Learning Representations (ICLR). arXiv:2009.03300.

Herzfeld, Noreen L. 2002. In Our Image: Artificial Intelligence and the Human Spirit. Minneapolis: Fortress Press.

Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. "ChatGPT Is Bullshit." Ethics and Information Technology 26: article 38.

Jackson, Robert. 2004. Rethinking Religious Education and Plurality: Issues in Diversity and Pedagogy. London: RoutledgeFalmer.

Jackson, Robert. 2014. Signposts: Policy and Practice for Teaching about Religions and Non-Religious Worldviews in Intercultural Education. Strasbourg: Council of Europe Publishing.

Ji, Ziwei, Nayeon Lee, Rita Frieske, et al. 2023. "Survey of Hallucination in Natural Language Generation." ACM Computing Surveys 55 (12): 1–38.

Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. "The State and Fate of Linguistic Diversity and Inclusion in the NLP World." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 6282–6293. Stroudsburg, PA: Association for Computational Linguistics.

Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, et al. 2023. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." Learning and Individual Differences 103: 102274.

Lackey, Jennifer. 2008. Learning from Words: Testimony as a Source of Knowledge. Oxford: Oxford University Press.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, et al. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." In Advances in Neural Information Processing Systems 33: 9459–9474.

Lin, Stephanie, Jacob Hilton, and Owain Evans. 2022. "TruthfulQA: Measuring How Models Mimic Human Falsehoods." In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 3214–3252. Stroudsburg, PA: Association for Computational Linguistics.

Messeri, Lisa, and M. J. Crockett. 2024. "Artificial Intelligence and Illusions of Understanding in Scientific Research." Nature 627: 49–58.

Mishra, Punya, and Matthew J. Koehler. 2006. "Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge." Teachers College Record 108 (6): 1017–1054.

Mitchell, Melanie, and David C. Krakauer. 2023. "The Debate over Understanding in AI's Large Language Models." Proceedings of the National Academy of Sciences 120 (13): e2215907120.

Ouyang, Long, Jeffrey Wu, Xu Jiang, et al. 2022. "Training Language Models to Follow Instructions with Human Feedback." In Advances in Neural Information Processing Systems 35: 27730–27744.

Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." BMJ 372: n71.

Raz, Joseph. 1986. The Morality of Freedom. Oxford: Clarendon Press.

Said, Edward W. 1978. Orientalism. New York: Pantheon Books.

Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. "Whose Opinions Do Language Models Reflect?" In Proceedings of the 40th International Conference on Machine Learning, PMLR 202: 29971–30004.

Shulman, Lee S. 1986. "Those Who Understand: Knowledge Growth in Teaching." Educational Researcher 15 (2): 4–14.

Smart, Ninian. 1983. Worldviews: Crosscultural Explorations of Human Beliefs. New York: Charles Scribner's Sons.

Tranfield, David, David Denyer, and Palminder Smart. 2003. "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review." British Journal of Management 14 (3): 207–222.

UNESCO. 2023. Guidance for Generative AI in Education and Research. Paris: UNESCO.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. 2017. "Attention Is All You Need." In Advances in Neural Information Processing Systems 30: 5998–6008.

Weidinger, Laura, Jonathan Uesato, Maribeth Rauh, et al. 2022. "Taxonomy of Risks Posed by Language Models." In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 214–229. New York: Association for Computing Machinery.

Zagzebski, Linda Trinkaus. 2012. Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief. New York: Oxford University Press.