Artificial Intelligence in Sacred Text Analysis: A Survey of Methodologies and Applications

Keyur Sameer Mahajan, Ali Abbas Gazge, Tanmayi Prakash Chaure, Neel Patel Mitulkumar, Prof. Rachana Karnavat,

Dept. of Information Technology, PICT, India

DOI: https://doi.org/10.51584/IJRIAS.2026.11060141

Received: 18 April 2026; Accepted: 23 April 2026; Published: 02 July 2026

ABSTRACT

The past decade has witnessed a remarkable convergence of artificial intelligence (AI) and the scholarly study of sacred texts. This paper offers a critical survey of AI-based sacred text analysis, drawing on studies involving the Bhagavad Gita, the Quran, and the Bible. It reviews approaches ranging from lexicon-based sentiment analysis and topic modeling to transformer-based models such as BERT, and it examines where these methods succeed and where they fall short. A central argument of the paper is that sacred texts cannot be treated as ordinary sentiment datasets because they are culturally embedded, symbolically rich, and theologically nuanced. Rather than claiming empirical validation beyond what is reported in the literature reviewed, this paper synthesizes prior work and proposes a conceptual framework built on interpretive partnership, methodological transparency, and domain-informed training. The study contributes a more careful and responsible perspective on the use of AI in spiritually sensitive domains.

Keywords: Artificial Intelligence, Natural Language Processing, Sacred Text Analysis, Sentiment Analysis, Digital Humanities, BERT, Hermeneutics

INTRODUCTION

The past ten years have seen a quiet but important shift in how sacred texts are approached, interpreted, and shared. What was once the preserve of theologians, linguists, and historians working through manual exegesis is now increasingly shaped by computational methods, especially those emerging from artificial intelligence (AI). From natural language processing models that analyze ancient Sanskrit verses to machine learning systems that group thematic patterns in the Quran or the Bible, the analytical landscape has grown substantially in both scope and speed. At the same time, this growth raises a deeper question: how should meaning, context, and interpretation be handled when algorithmic systems begin to mediate texts that are layered, symbolic, and culturally embedded?

Ideally, the study of sacred texts requires a balance between philological accuracy, historical awareness, and interpretive sensitivity. Scholars try to preserve contextual subtlety while still uncovering patterns that may not be immediately visible in close reading. Within this setting, AI can serve as an augmentative tool, helping to identify patterns, compare across languages, and widen access to large corpora. Even so, many current systems still favor computational efficiency over interpretive richness, reducing theological narratives to tokenized data points or statistical correlations. They perform reasonably well on surface-level tasks such as text classification or sentiment labeling, but they often struggle with metaphor, allegory, and doctrinal distinction.

This gap is not entirely unrecognized. A growing body of literature has attempted to connect computational techniques with religious text analysis. Computational linguistics has been used to identify stylistic differences in biblical corpora (Burrows, 2002), while topic modeling methods have been applied to investigate thematic patterns in Islamic texts (Alhawarat et al., 2015). More recently, deep learning models, including transformer-based architectures, have been used to conduct semantic analysis of multilingual scriptures (Devlin et al., 2019). These are valuable contributions, but they tend to operate within narrow methodological frames, focusing on

tasks such as authorship attribution, translation alignment, or keyword extraction without engaging the broader interpretive challenge.

The consequences of this limitation are both direct and indirect. At a practical level, misinterpretations generated by AI systems can spread across online platforms and shape how people perceive religious texts in ways that are difficult to track or correct. More subtly, there is a risk of encouraging reductive readings of complex traditions, particularly when computational outputs are treated as objective or authoritative. This concern is especially important in pluralistic societies where sacred texts intersect with cultural identity, legal frameworks, and ethical discourse.

What remains underexplored is not the use of AI on sacred texts itself, but the methodological and conceptual frameworks that should guide such applications. A clear gap exists for integrative studies that critically examine the assumptions embedded in computational models when they are applied to spiritually significant literature. Current research rarely asks how meaning, context, and interpretation are preserved or distorted in algorithmic pipelines, nor does it fully address the ethical implications of using AI in domains where interpretive authority has historically been contested.

This paper addresses these gaps by offering a structured survey of methodologies and applications in AI-based sacred text analysis, alongside a conceptual framework that foregrounds interpretive sensitivity. It builds on existing research in computational linguistics and digital humanities while focusing on the interaction between technical design and hermeneutic principles. Rather than treating sacred texts as mere datasets, it views them as dynamic artifacts embedded in historical, cultural, and theological contexts.

Objectives of the Study

The main objectives of this study are threefold. First, the paper reviews current AI approaches to the analysis of sacred texts, including natural language processing, machine learning, and deep learning methods. Second, it critically evaluates the limitations of these methods in terms of interpretive specificity, contextual richness, and theological nuance. Third, it proposes a conceptual model capable of combining computational efficiency with hermeneutic sensitivity and outlines the reporting requirements needed for a transparent systematic study.

More specifically, the paper examines whether existing AI models can be adapted to better accommodate multi-layered meanings in sacred texts and considers how interdisciplinary expertise from religious studies and linguistics can guide such adaptation. The paper is organized following the CARS (Create A Research Space) model: it first establishes the territory by stating the growing relevance of AI to sacred text analysis, then identifies the niche by pointing to the limitations of existing methodologies, and finally fills the niche by providing a critical survey and a conceptual framework that integrates computational and hermeneutic perspectives.

LITERATURE REVIEW

The application of artificial intelligence methods to spiritual and religious texts has been growing steadily, providing new opportunities to identify patterns and sentiments within these complex corpora. Theoretically, AI and in particular natural language processing could uncover thematic patterns or emotional tones within vast quantities of religious text that would not otherwise be discernible through traditional reading. A large-scale analysis of fourteen different sacred texts found that, even despite archaic language, many sacred writings share similar themes, historical contexts, and emotional tones while also exhibiting culture-specific vocabulary and polarity differences (Felipe-Ruiz, 2024). These studies highlight the potential of AI to reduce linguistic and cultural barriers in interpreting faith-based literature.

However, this potential has only been partially realized. Present AI systems tend to simplify rich scripture to sentiment ratings or topic labels and treat sacred writings as though they were conventional social media data. This raises genuine concerns: when algorithmic analyses are interpreted too literally, they may reduce highly nuanced traditions to oversimplified categories. For example, one recent study applied a transformer model based on BERT to compare five English translations of the Sermon on the Mount and found varying sentiment scores

across verses, with subtle indications of humor or empathy (Vora et al., 2024). Yet by concentrating on statistical sentiment alone, the theological background and doctrinal significance of these passages risk being overlooked.

Review of Important Literature

Several recent papers represent the current state of AI-based religious sentiment analysis, and examining them in detail reveals both the progress made and the persistent gaps. Some researchers apply relatively straightforward tools over broad datasets. For instance, topic modeling and the lexicon-based VADER analyzer have been used on translations of the Bhagavad Gita, the Quran, and the Bible, with findings suggesting that each of the three texts emphasizes themes such as generosity and devotion and that traditionally distinct religious traditions share common moral values (Goel and Arsiwala, 2024). Similarly, Felipe-Ruiz (2024) conducted lexical, sentiment, and statistical analyses across fourteen sacred texts, concluding that numerous scriptures exhibit widespread emotional colorings, while also noting polarity variations across cultures, suggesting that translation choices and historical context can alter the perceived tone of a text.

These pieces of work contribute to the field by demonstrating that AI can identify broad similarities and differences across traditions. However, they also reveal significant limitations. VADER, for instance, was designed for contemporary social media sentiment and can struggle to classify spiritually neutral or metaphorical language correctly. Neither study discusses how cultural idioms or theological context might confound the analysis, and both rely primarily on off-the-shelf lexicons or topic models without engaging with deeper interpretive questions.

Other initiatives are more technically focused. Using a BERT-based deep model, Chandra and Kulkarni (2022) attempted to align the English translations of the Bhagavad Gita with their original Sanskrit meaning. After fine-tuning a sentiment dataset by hand, they found that, though translation styles vary considerably, sentiment analysis and semantic similarity measures suggest that the underlying emotional message is largely preserved across versions. This implies that deep learning can, to some degree, abstract the underlying emotional meaning of scripture. Vora et al. (2024) similarly applied a large language model to the Sermon on the Mount and demonstrated how different Bible translations differ in vocabulary, use of humor, and overall tone. These studies advance the application of state-of-the-art NLP to sacred texts, yet they share a critical assumption: that high-level sentiment classification is a sufficient or complete measure of scriptural meaning.

Other scholars have examined religious discourse on social media or mixed-corpus data. Nath, Das, and Ghosh (2024) introduced a wide-ranging religious sentiment detector based on recurrent neural networks and long short-term memory architectures, trained on data related to twelve major religions, and reported remarkably high classification accuracy. However, reported accuracy metrics alone raise questions: What was the distribution of sentiment classes in the training data? How was ambiguity managed? Without this information, the practical utility of such a model for nuanced scriptural interpretation is difficult to assess. Listiyono et al. (2024) used a Naive Bayes classifier on religious posts in Indonesian language forums and achieved reasonable precision and accuracy by adapting the algorithm to religious vocabulary. While this demonstrates that simpler models can capture emotional signals in faith contexts, it also illustrates a broader trend: many studies prioritize algorithmic performance metrics over interpretive validity.

Identified Gaps and the Contribution of This Paper

Looking across this body of work, several consistent patterns emerge. Numerous studies, whether examining scriptures or social media religious content, confirm that AI can classify or rank texts by sentiment and identify common motifs such as charity and forgiveness across traditions. However, significant inconsistency exists in terms of interpretive depth. Some analyses detect subtle emotional tones within scripture, while others suggest that deep models essentially abstract stylistic differences into coarse sentiment categories. These discrepancies likely reflect methodological differences between rule-based lexicons, deep neural networks, and simpler classifiers.

Most strikingly, no existing study adequately resolves the central question of how well computationally derived sentiments correspond to human or religious community interpretations. This remains the most significant

knowledge gap in the field. Sacred texts are frequently ambiguous, metaphorical, or contextually dependent in ways that standard sentiment models are not equipped to handle. A verse rated as emotionally neutral by an NLP system may carry profound spiritual significance within its tradition. Furthermore, the majority of studies presuppose English or contemporary-language inputs and rarely examine how sentiment shifts through translation beyond noting stylistic differences.

This paper addresses these gaps in alignment with its core objectives. By reviewing studies that explicitly differentiate surface-level tasks from deeper semantic inference, by comparing AI outputs against expert human interpretations where such comparisons are reported in the literature, and by proposing a conceptual framework that integrates hermeneutic sensitivity with computational analysis, it charts a more complete path forward for the field.

METHODOLOGY

The research design of this paper is a structured literature-survey approach. It synthesizes existing work on AI-based sacred text analysis and organizes the literature by method, text tradition, and interpretive limitation. The present manuscript should therefore be read as a critical survey and conceptual paper rather than as an original experimental study.

Because this is a review paper and does not report new human-subject data, a separate ethics approval statement is not required.

Similarly, no participant recruitment was carried out for this manuscript. Any discussion of expert judgment in the paper refers to the studies reviewed in the literature, not to a new annotation exercise conducted by the authors.

The primary materials discussed in this survey consist of sacred writings drawn from multiple traditions, including the Bhagavad Gita, the Bible, and the Quran in both original and translated editions, as reported in the literature reviewed. Where the paper refers to passage-level comparison, those examples should be understood as illustrative summaries of published studies rather than as original data produced for this manuscript.

The review process was organized thematically. First, the literature on AI-based sacred text analysis was grouped by methodological family, including lexicon-based methods, topic models, machine learning classifiers, and transformer-based approaches. Second, each study was examined for its treatment of context, metaphor, translation, and doctrinal nuance. Third, the review synthesized recurring strengths and limitations in order to build the proposed conceptual framework.

Because the paper is a survey rather than an experiment, the outcome measures are descriptive. The discussion compares how studies report accuracy, precision, thematic consistency, translation sensitivity, and interpretive validity. This makes it possible to identify patterns across the literature without claiming new empirical results that were not actually collected in the present manuscript.

RESULTS AND DISCUSSION

The findings of this survey reveal a field that is technically productive but interpretively uneven. Across the studies reviewed, AI methods have demonstrated genuine capability in detecting broad thematic patterns, identifying shared values across religious traditions, and classifying sentiment at a gross level. Topic modeling approaches consistently identified common motifs such as devotion, generosity, compassion, and moral instruction across Hindu, Christian, and Islamic texts, offering empirical support for the intuition that major world religions share overlapping ethical frameworks.

However, the literature also shows that these outputs are often too coarse for theological interpretation. Verses or passages that scholars treat as spiritually significant are frequently reduced to neutral or negative sentiment by baseline models, especially when the language is archaic, poetic, or liturgical. BERT-style models perform

better on semantic similarity, but they still struggle with paradox, allegory, and meanings that depend on context rather than surface wording.

The literature reviewed does not support a single pooled reliability estimate, largely because the studies use different datasets, labels, and evaluation settings. Even so, the overall pattern is clear: agreement is usually stronger for straightforward narrative passages and weaker for poetic, parabolic, or doctrinally dense material. This supports the central argument of the paper: the interpretive dimensions of sacred texts are not reducible to the statistical patterns current NLP models are designed to capture.

The reviewed studies also point to an important qualitative issue: sentiment in sacred contexts does not behave like sentiment in everyday discourse. A verse expressing divine wrath may be experienced by practitioners as consoling in its affirmation of justice, while a passage of blessing may carry deep communal meaning that no automated system can detect without context. These observations confirm that interpretive validity, rather than classification accuracy alone, should be central in future evaluation of AI systems for sacred literature.

Comparison across studies reveals a methodological divergence with direct implications for interpretation. Rule-based lexicon tools tend to produce surface-level emotional labels, whereas deep-learning approaches offer more nuanced outputs but remain highly sensitive to the quality and assumptions of the training data. Very few studies integrate domain expertise from religious scholarship in a systematic way, and this remains the most important practical limitation identified by the survey.

CHALLENGES IN AI-BASED SACRED TEXT ANALYSIS

Despite the progress documented in this survey, several structural challenges continue to limit the reliability and depth of AI-based sacred text analysis. Understanding these challenges is essential not only for improving technical systems but also for ensuring that their deployment in culturally sensitive domains is responsible and ethically sound.

Interpretive Complexity and Metaphorical Language

Sacred texts are not ordinary prose. They are composed in registers that rely heavily on metaphor, allegory, symbolism, and intertextual reference, features that standard NLP models are not well equipped to handle. A verse that describes God as a consuming fire, for instance, carries theological meaning that is neither literal nor easily mapped onto conventional sentiment categories. Current models that assign such passages a negative or alarming emotional score misrepresent the intended meaning in ways that can mislead users who rely on automated analysis.

Translation and Multilingual Challenges

Most AI studies reviewed in this paper work primarily with English translations of sacred texts, and relatively few examine how meaning and sentiment shift across translations or between a text and its original language. This is a significant limitation, since translation decisions are themselves interpretive acts that introduce the assumptions and cultural contexts of translators. The Bhagavad Gita in English, for example, reflects not only the original Sanskrit but also the interpretive frameworks of its translators, which vary considerably across scholarly, devotional, and colonial traditions.

Data Quality and Bias

Machine learning models are highly sensitive to the quality and composition of their training data. When models trained on general-purpose sentiment datasets are applied to sacred texts, they import the assumptions embedded in those datasets, including biases toward contemporary social media language, Western cultural norms, and secular frameworks for understanding emotion. These biases can produce systematic misclassifications that are not immediately visible to users and are difficult to detect without expert validation.

Ethical and Epistemological Concerns

The application of AI to sacred texts raises profound ethical questions about interpretive authority and the nature of religious knowledge. When computational outputs are presented without appropriate caveats, they risk being perceived as objective or authoritative by non-expert users, potentially displacing the interpretive traditions of religious communities. In pluralistic societies where sacred texts intersect with cultural identity and community practice, the stakes of misrepresentation are not merely academic. Responsible AI deployment in this domain requires transparency about model limitations, engagement with religious communities, and clear communication about the difference between computational analysis and theological interpretation.

PROPOSED CONCEPTUAL FRAMEWORK

Based on the findings of this survey, this paper proposes a conceptual framework for the responsible integration of AI methods into sacred text analysis. The framework rests on four interconnected principles: interpretive partnership, methodological transparency, domain-informed training, and iterative validation. These principles are intended as design and reporting guidance for future work, not as a claim that the present manuscript includes a fully executed empirical system.

Interpretive partnership refers to the deliberate integration of religious studies scholars and community representatives into the design, evaluation, and deployment of AI systems used on sacred texts. Rather than treating expert knowledge as a post-hoc validation step, this framework positions it as foundational to the analytical process. This means engaging theologians and linguists not only in annotating training data but also in defining what counts as meaningful output and what interpretive questions the system is designed to address.

Methodological transparency requires that AI systems used in religious contexts clearly communicate their assumptions, limitations, and the nature of their training data. This includes flagging passages where model confidence is low, where cultural or linguistic context significantly affects interpretation, and where automated classification diverges from expert consensus. Such transparency is not merely a technical feature but an ethical commitment to honest communication with users.

Domain-informed training involves adapting or fine-tuning AI models using datasets that have been curated and annotated by domain experts, rather than relying on general-purpose sentiment or language models. This requires sustained investment in the creation of high-quality, expert-labeled corpora for sacred texts across multiple traditions and languages, a task that is labor-intensive but essential for producing systems that are genuinely fit for purpose.

Iterative validation requires that AI outputs be regularly compared against human expert interpretations, with discrepancies treated as learning opportunities rather than errors to be minimized. This cyclical process of comparison, reflection, and refinement is consistent with the hermeneutic tradition in which meaning is understood as contextual and dynamic rather than fixed. It also ensures that AI systems remain accountable to the communities whose texts they analyze.

CONCLUSION

This paper has presented a critical survey of methodologies and applications in AI-based sacred text analysis, drawing on recent studies across Hindu, Christian, and Islamic traditions while situating these within the broader context of computational linguistics and digital humanities. The literature shows that AI methods, including sentiment analysis, topic modeling, and transformer-based deep learning, offer genuine promise for identifying broad thematic patterns and enabling large-scale comparative study of religious corpora. At the same time, the review reveals a persistent gap between what these systems can compute and what sacred texts mean within their living traditions.

The central argument of this paper is that interpretive validity, rather than classification accuracy, should become a primary evaluative standard for AI systems applied to sacred literature. This requires moving beyond task-based optimization toward an interdisciplinary approach in which computational analysis is guided by, and

accountable to, the hermeneutic traditions of religious scholarship. The conceptual framework proposed here, built on interpretive partnership, methodological transparency, domain-informed training, and iterative validation, offers a practical path in this direction.

FUTURE RESEARCH DIRECTIONS

Future work should move from broad sentiment scoring toward domain-specific evaluation. That means building expert-annotated corpora in original languages, comparing multiple translations side by side, and developing metrics that capture interpretive validity rather than surface polarity alone. It would also be useful to test whether transformer-based models can be adapted with religious-studies guidance so that they better handle metaphor, ambiguity, and doctrinal context. Just as important, future studies should document their search strategy, screening process, and selection criteria more transparently so that survey claims can be checked and reproduced more easily.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Prof. Rachana Karnavat for her guidance and support throughout this research. The authors also thank the Department of Information Technology at PICT for providing the resources and academic environment that made this work possible.

REFERENCES

1. Alhawarat, M., Hegazi, M., & Hilal, I. (2015). Revisiting k-means and topic modeling, a comparison study to cluster Arabic documents. IEEE Access, 3, 2041–2054.
2. Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 5185–5198). https://doi.org/10.18653/v1/2020.acl-main.463
3. Burrows, J. (2002). Delta: A measure of stylistic difference and a guide to likely authorship. Literary and Linguistic Computing, 17(3), 267–287. https://doi.org/10.1093/llc/17.3.267
4. Chandra, R., & Kulkarni, V. (2022). Semantic and sentiment analysis of selected Bhagavad Gita translations using BERT-based language framework. IEEE Access, 10, 21291–21315. https://doi.org/10.1109/ACCESS.2022.3154258
5. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019 (pp. 4171–4186). https://doi.org/10.18653/v1/N19-1423
6. Felipe-Ruiz, A. (2024). Lexical, sentiment and correlation analysis of sacred writings: A tale of cultural influxes and different ways to interpret reality. Natural Language Processing Journal, 9, Article 100121. https://doi.org/10.1016/j.nlp.2024.100121
7. Listiyono, S., Budiarso, Z., Susilowati, S., & Windarto, A. P. (2024). Comprehensive sentiment analysis of religious content: Naive Bayes algorithm model. Jurnal Media Informatika Budidarma, 8(1), 602–611. https://doi.org/10.30865/mib
8. Nandan, A. D. M., Godbole, I., Kapparad, P., & Bhattacharjee, S. (2025). Comparative analysis of religious texts: NLP approaches to the Bible, Quran, and Bhagavad Gita. Coling-Rel Proceedings. Association for Computational Linguistics.
9. Nath, S., Das, U., & Ghosh, D. (2024). A religious sentiment detector based on machine learning to provide meaningful analysis of religious texts. In K. Dasgupta et al. (Eds.), Computational Intelligence in Communications and Business Analytics. Springer. https://doi.org/10.1007/978-3-031-48876-4_13
10. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
11. Smith, J. (2021). Reflective AI in the humanities: Toward a framework for responsible computational text analysis. Digital Humanities Quarterly, 15(2).
12. Vora, M., Blau, T., Kachhwal, V., Solo, A. M. G., & Chandra, R. (2024). Large language model for Bible sentiment analysis: Sermon on the Mount. arXiv:2401.00689. https://doi.org/10.48550/arXiv.2401.00689