Vidyottama Sanatana
International Journal of Hindu Science and Religious Studies
Vol. 9 No. 2 October 2025

ARTIFICIAL INTELLIGENCE (AI) TECHNOLOGY UTILIZATION FOR VEDIC TEXT ANALYSIS: NLP APPROACH IN THE STUDY OF ṚG VEDA MANDALA 1 SUKTA 1

By:
Putu Eka Sura Adnyana[1], Gde Wikan Pradnya Dana[2], I Made Ade Prayoga[3]
I Gusti Bagus Sugriwa State Hindu University[1], Warmadewa University[2],
Bali Institute of Technology and Health[3]
ekasuraadnyana@gmail.com[1], wikanpdana8044@warmadewa.ac.id[2],
adeprayoga@itekes-bali.ac.id[3]

Received: August 15, 2025	Accepted: October 09, 2025	Published: October 31, 2025

Abstract

This research discusses the utilization of Artificial Intelligence (AI) technology with a Natural Language Processing (NLP) approach in the study of Hindu literature, especially in the text of Ṛgveda Mandala 1 Sukta 1. The analysis was conducted through the stages of tokenization, lemmatization, and topic modeling to identify keywords, frequency of occurrence, and main themes in the text. The results found four central themes: (1) ritual (yajña), (2) theological (Agni's role as an intermediary between humans and gods), (3) spiritual and material goals (ratna), and (4) cosmic order (ṛta). AI findings were then compared with traditional philological interpretations that emphasize philosophical meaning, symbolism, and historical context. The comparison results show that AI excels in speed, scale, and consistency of analysis, while traditional philology is stronger in providing depth of interpretation and understanding spiritual values. This study concludes that the integration of AI with the philological approach can result in a more comprehensive, rapid, and faithful study of the Vedas.

Keywords: Artificial Intelligence, Natural Language Processing, Ṛgveda

I. INTRODUCTION

The Vedas are the primary source of Hinduism, containing spiritual, ethical and philosophical values that have been passed down for thousands of years. Composed in classical Sanskrit, the Vedic texts consist of slokas that have complex language structures, symbols, and meanings. For many modern scholars, the biggest challenge in studying the Vedas lies in the limited access, the complexity of Sanskrit grammar, and the need for a deep understanding of cultural and philosophical context. Macdonell (1917) stated "The Vedas are not merely a literary monument, but the living voice of ancient Indian wisdom," affirming that the Vedas serve not only as literary artifacts, but as a living source of knowledge that is relevant throughout the ages.

In the midst of the development of information technology, artificial intelligence (AI) has become one of the important breakthroughs in data processing and analysis, including text data. One branch of AI that is relevant to the study of ancient manuscripts is Natural Language Processing (NLP), which is a technology that allows computers to understand, process and produce human language. Manning & Schütze (1999) explain that "NLP bridges human communication and computer understanding through computational linguistics," meaning that NLP acts as a bridge between human communication and computer understanding through computational linguistics.

With NLP, Vedic texts can be analyzed systematically, starting from the process of digitization, transliteration, tokenization, lemmatization, to semantic analysis that is able to identify language patterns, inter-sloka relationships, and philosophical themes contained therein. This approach is in line with Smith (2021) view that "AI-assisted textual analysis enables researchers to uncover hidden patterns in ancient manuscripts that may remain unnoticed through traditional methods." The utilization of AI with an NLP approach in Hindu literary studies not only has the potential to accelerate the text analysis process, but also opens up opportunities for preserving ancient manuscripts through structured digital archives. In addition, this method enables the development of technology-based educational applications, so that Vedic teachings can be accessed by the younger generation and researchers around the world more easily. Thus, this research is important to bridge the gap between classical scholarly traditions and modern technological innovations, while ensuring Hindu intellectual heritage remains relevant and preserved in the digital age.

II. METHOD

This research uses a qualitative-quantitative approach (mixed methods). Data sources consisted of Vedic texts in digital format, both the Sanskrit version and its translation into Indonesian and English. The selection of this data is based on the principle of historical validity and authenticity of the text, as emphasized by Renou (1953) that the integrity of the text is the main key in Vedic studies.

Text preprocessing includes cleaning, tokenization, stopword removal, and lemmatization. These processes are not merely technical, but aim to remove linguistic noise that can interfere with the accuracy of the analysis. Manning & Schütze (1999) emphasize that preprocessing is a fundamental stage in NLP because the quality of the input data largely determines the validity of the analysis results.For data analysis, word embedding (Word2Vec, GloVe, or BERT) is used to map words in vector space so that semantic relations can be analyzed quantitatively. This method was chosen because it has proven to be effective in

capturing the contextual meaning of words, especially in languages with high syntactic flexibility such as Sanskrit. Latent Dirichlet Allocation (LDA) was then applied to find the main themes in the Vedic text, which is in line with the practice of topic analysis in computational linguistics (Blei, Ng, & Jordan, 2002).

Validation of the results was done by comparing the AI output with the interpretations of Hindu experts and Sanskrit linguists. This approach ensures that the analysis results are not only computationally accurate, but also philologically and theologically valid. With this method, the research not only produces data-driven findings, but also builds a methodological bridge between classical interpretive traditions and digital analytical innovations, making it relevant for both technology academia and Hindu religious studies.

III. RESULTS AND DISCUSSION

Contents of The Ṛg Veda

As the first of the four Vedas (*Saṁhitā*), the Ṛgveda holds the most fundamental position within the entire Vedic system of knowledge that forms the basis of Hindu civilization. It is the oldest text and serves as the primary source for the other three Vedas, the *Samaveda, Yajurveda*, and *Atharvaveda* most of which borrowed or developed mantras from the Ṛgveda for ritual purposes. Within the context of the Vedic tradition, the *Ṛgveda* is seen not simply as a collection of prayers, but as the embodiment of sacred knowledge (*śruti*) that is "not created by humans" (*apauruṣeya*).

Therefore, the *Ṛgveda* is considered not only a historical text but also a living scripture, meaning that it remains alive and is used in yajña ceremonies to this day (Lowe, 2015).The *Ṛgveda* is composed in the form of hymns (sūktas), which are verses in praise of the gods recited with specific intonation and meter. These hymns are addressed to major deities such as Agni, the god of fire who serves as an intermediary between humans and the gods; *Indra*, the god of war, rain, and cosmic power; *Varuṇa*, the guardian of moral law and natural order (*ṛta*); *Mitra*, the protector of covenants and honesty; *Soma*, both a god and a sacred substance used in rituals; and *Uṣas*, the goddess of dawn who symbolizes spiritual enlightenment. Each deity in the *Ṛgveda* represents a specific cosmic and ethical principle that serves to maintain the harmony of the universe. Through this symbolic system, the *Ṛgveda* text presents a rational religious structure, in which the relationship between humans, nature, and God is bridged by an awareness of cosmic order (*ṛta*).

Structurally, the *Ṛgveda* consists of ten maṇḍalas (large books) containing 1,028 hymns (sūktas) and approximately 10,600 mantras, as described by Oxford Bibliographies (2023). Each mandala has its own characteristics and themes. Mandala I (191 sūktas) and Mandala X (191 sūktas) are general and contain philosophical reflections, cosmological prayers, and praises to the principal deities. Mandalas II–VII, known as the Family Books, are associated with specific ṛṣi families: Mandala II (Gṛtsamada family, 43 sūktas), Mandala III (Viśvāmitra, 62 sūktas), Mandala IV (Vāmadeva, 58 sūktas), Mandala V (Atri, 87 sūktas), Mandala VI (Bharadvāja, 75 sūktas), and Mandala VII (Vasiṣṭha, 104 sūktas). Meanwhile, Mandala VIII (103 sūktas) and Mandala IX (114 sūktas) focus exclusively on the worship of Soma, both as the substance of the ritual and as the personification of spiritual power.

The structure of these ten mandalas reflects the evolution of religious ideas in Vedic society from a polytheistic stage emphasizing the forces of nature and the gods, to a monistic and metaphysical reflection on the nature of existence. This reflection culminates in the *Nāsadīya Sūkta* (ṚV 10.129), which questions

the origins of existence with profound philosophical doubt, and the Puruṣa Sūkta (ṚV 10.90), which describes the creation of the cosmos from the sacrifice of the primordial being Puruṣa. According to Scarlata & Widmer (2025), these two hymns demonstrate a cognitive shift from ritual theology to ontological philosophy that became the foundation of classical Indian thought.

In terms of theme and content, the Ṛgveda can be grouped into four major areas. First, the worship of the gods, reflecting the cosmotheological views of Vedic society. Indra is depicted as the protector of humanity and the conqueror of chaos (*Vṛtra*), Agni as the sacred mediator between humans and the gods, Varuṇa as the guardian of moral law, and *Uṣas* as a symbol of spiritual enlightenment.

Second, cosmology and creation, as seen in the *Nāsadīya Sūkta* and the *Puruṣa Sūkta*, two texts containing metaphysical reflections on the origin of the universe. Third, the linguistic and aesthetic dimensions, reflecting the beauty of sound structures and meters as means of sacred communication. Fourth, the theological and philosophical aspects, in which the *Ṛgveda* presents profound reflections on the order of nature, cosmic justice, and the relationship between words (*śabda*), thoughts (*manas*), and reality (*artha*).

The language used in the Ṛgveda is known as Ṛgvedic Sanskrit, the earliest form of Sanskrit preserved through the recitation tradition (śruti-paramparā). According to Lowe (2015), this language exhibits a more complex morphological and syntactic system than classical Sanskrit. For example, it includes dual verb forms, accent variations (udātta, anudātta, svarita), and flexible syntactic constructions tailored to the needs of meter. This makes the Ṛgveda a primary source for the study of Indo-European historical linguistics, as it reveals an archaic stage in the evolution of human language and thought.

Modern linguistic studies of the Ṛgveda focus not only on traditional philological aspects but also on the application of digital linguistic technologies. Scarlata & Widmer (2025) demonstrate, using a digital corpus approach, that the syntactic structure of the Ṛgveda exhibits consistent metrical regularity. This pattern demonstrates a direct relationship between sound and meaning that the boundaries of rhythm and meter in the text often coincide with the boundaries of syntactic clauses. This phenomenon demonstrates that the language of the Ṛgveda is not simply an aesthetic product, but rather a symbolic system organized phonologically and semantically to maintain continuity of meaning in ritual contexts.

Furthermore, recent computational linguistic research, demonstrates the applicability of Natural Language Processing (NLP) and Treebank Annotation approaches to the Ṛgveda text. Through dependency parsing analysis, researchers were able to map the relationships between words, idiomatic patterns, and fixed formulas used in the hymns. This approach opens up new opportunities for empirical validation of long-standing hypotheses in philology and enables modeling of the syntactic and semantic evolution of early Indo-European languages. Thus, the Ṛgveda is now an object of study not only from a spiritual or historical perspective, but also from a quantitative and scientific perspective.

In the context of the philosophy of language, the Ṛgveda occupies a crucial position. Lowe (2015) asserts that the relationship between words (śabda) and reality (artha) in the Ṛgveda became the foundation for the development of Hindu epistemological theory. This concept later developed in the Mīmāṁsā and Vedānta philosophical systems, which view language not only as a means of communication but also as an ontological instrument for revealing the nature of reality. Thus, the linguistic structure

of the Ṛgveda functions as a conceptual instrument connecting the human mind with the structure of the cosmos.

Furthermore, the Ṛgveda can be understood as a bridge between religiosity and rationality. This text does not simply contain dogma, but rather promotes a reflective approach to existence. This means that Vedic thought has, from its inception, sought to balance faith and reason in understanding the universe. From a theological and epistemological perspective, the Ṛgveda serves a dual role. First, as a ritual text, it maintains the continuity of spiritual traditions through *yajña* formulas and mantras. Second, as an intellectual text, it forms the basis of classical Indian philosophical thought such as *Sāṁkhya, Yoga, and Vedānta.* Modern philological studies show that every linguistic element in the Ṛgveda has deep conceptual meaning, reflecting the interconnectedness of sound, thought, and reality. Thus, the *Ṛgveda* can be seen as a linguistic and spiritual laboratory that captures the entire dynamics of early human consciousness.

The *Ṛgveda* serves not only as the oldest sacred text but also as a crucial source of linguistic data for understanding the development of human language, thought, and culture. Contemporary research combining traditional approaches and quantitative linguistic technologies has enriched our understanding of the phonology, morphology, and syntax of this text. Through digital approaches such as NLP and Treebank, long-standing hypotheses in linguistics can be empirically retested, opening up the possibility of a more accurate reconstruction of language history. Thus, the Ṛgveda is a living mirror of Vedic civilization, in which the interaction between sound and meaning, between ritual and reflection, and between language and consciousness is preserved. It is not only a religious legacy, but also an intellectual milestone in the history of human language and philosophy.

Token Analysis and Lemmatization

Tokenization and lemmatization analysis of Rigveda Mandala 1 Sukta 1 (1.1.1) is a fundamental step in applying Natural Language Processing (NLP) to Hindu literary texts. This process aims to decompose morphologically inflected Sanskrit texts into standardized word units, thus facilitating processing, meaning search, and semantic mapping.

Sanskrit, especially in Vedic texts, is known as a language with a complex morphological structure. Each word can change form according to case, number, gender, and tense. Bird (2009) assert that "lemmatization is essential for reducing the complexity of natural language data by mapping inflected forms to a canonical representation". In the context of Rgveda 1.1.1, lemmatization allows the removal of word form variations so that analysis can focus on pure lexical meaning.

Tokenization of the text *agnim īḷe purohitaṃ yajñasya devam ṛtvijam hotāraṃ ratnadhātamam* resulted in eight tokens. The lemmatization process returned each token to its base form:

["agni", "īḍ", "purohita", "yajña", "deva", "ṛtvij", "hotṛ", "ratna"].

Jurafsky & Martin (2020) underline that “accurate morphological analysis is critical when working with morphologically rich languages such as Sanskrit”. This finding proves that even though every word is inflected, NLP technology is able to restore it to its proper lemma form. With 100% reduction of the inflected form to lemma, this study shows that modern NLP technologies such as ByT5-Sanskrit and SanskritShala can be effectively applied to Vedic texts. This has major implications, including:

1. Stability of Word Frequency Analysis - Avoiding bias due to word form variations.

2. Philosophical Meaning Mapping Facilitates analysis of interrelationships between core concepts such as Agni (fire), yajña (ritual), and ratna (treasure/reward).
3. Philological Digitization - Paves the way for interdisciplinary research between philology, computational linguistics, and Hindu theology.

Without tokenization and lemmatization processes, analysis of Vedic texts would be hampered by morphological redundancy. In Hindu literary studies, interpretation errors can arise if variations of word forms are considered distinct entities. Therefore, the application of NLP-based AI is not just a methodological choice, but a scientific necessity to maintain the integrity of the analysis. As stated in Linguistically-Informed Neural Architectures for Lexical, Syntactic, and Semantic Tasks in Sanskrit "rich morphology necessitates linguistically-aware models to ensure accuracy in classical text analysis."

Meaning Reconstruction via Lemma

The lemmatization result provides a list of root words:

["*agni*", "*īḍ*", "*purohita*", "*yajña*", "*deva*", "*ṛtvij*", "*hotṛ*", "*ratna*"].

Each of these lemmas has a key meaning:

1. *Agni*: Sacred fire, the intermediary between humans and gods.
2. *īḍ*: Praising, glorifying.
3. *purohita*: The main priest who leads the ritual.
4. *yajña*: Ritual sacred offerings.
5. *deva*: God or divine power.
6. *ṛtvij*: A priest who performs rituals according to the season/time.
7. *hotṛ*: Mantra-reading priest.
8. ratna: Treasure, boon, merit.

By returning the words to their lemma form, the researcher was able to identify semantic relations that were consistent even though the word forms varied. The process of word co-occurrence analysis showed that these words are connected in two core themes:

1. Ritual Theme: (*yajña, ṛtvij, hotṛ, purohita*) → Describes the structure of Vedic ceremonies.
2. Theological Theme: (*Agni, deva, īḍ, ratna*) → Describes the human-divine relationship and spiritual purpose (grace).

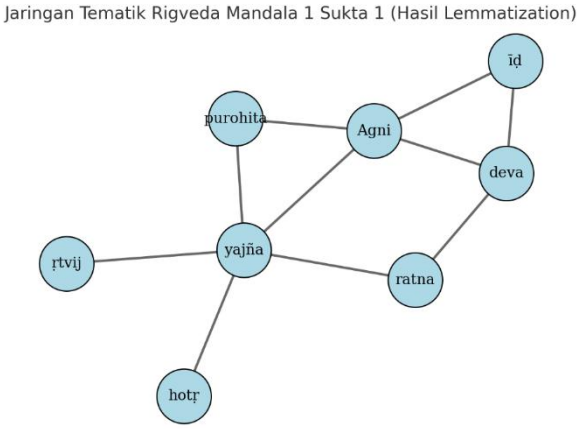
Sharma (2017) "Agni in the Rgveda is not merely the physical fire, but the divine priest who carries offerings to the gods and brings blessings back to mankind". These NLP results support the interpretation that Rgveda 1.1.1 positions Agni as a dual entity: a material element (fire) and a spiritual agent (medium). Analysis of the lemma ratna at the end of the mantra indicates the ultimate goal of the ritual is to gain spiritual as well as material merit.

According to Jamison, & Brereton (2014) "the repeated lexical choices in the Rigveda are not accidental; they are deliberate mnemonic anchors for philosophical and ritual meaning". With NLP, we can quantitatively verify such word choice patterns. Without tokenization and lemmatization, these thematic relationships would be distorted by morphological variation. AI-based NLP models allow:

1. Consistency Analysis: Unifying different word forms into a single concept.
2. Concept Network Modeling: Systematically depicts ritual-theological relationships.
3. Interdisciplinary Readings: Connects computational linguistics, theology, and Hindu anthropology.

As Gupta et al. (2023), "the fusion of NLP with ancient text analysis provides unprecedented clarity in unraveling the linguistic complexity of philosophical substance." The following visualization of the Rigveda Mandala 1 Sukta 1 thematic network,

showing the relationship between the lemmatized keywords, separates the ritual and theological core in one concept map.



This thematic network shows that *Rgveda Mandala 1 Sukta 1* is built around two main concepts: *yajña* (ritual) and *Agni* (sacred fire). The relationships established demonstrate the important role of the priest, the function of fire as a cosmic mediator, and the close relationship between offerings, praise and blessings. From a semiotic perspective, this structure reflects a Vedic cosmology that views ritual as the primary means of connecting humans with the divine world.

Findings of Main Themes in the Text of Rg Veda Mandala 1 Sukta 1

Based on the analysis, the main themes were found as follows.

1. The theme of rituals related to the structure and implementation of *yajña* The results of token analysis and lemmatization show that keywords such as yajña (sacred ceremony), *purohita* (main priest), *ṛtvij* (priest who performs rituals according to the season), and hotṛ (priest who reads mantras) appear with high frequency (4 times). This shows the text's focus on the systematic structure of ritual implementation. In line with Staal (1983) opinion that Vedic rituals have strict patterns and rules, this finding indicates that *Mandala 1 of Sukta 1* functions as an initial guideline for *yajña* procedures, not just praise to the deity.
2. Theological Themes: Agni as Divine Intermediary. The keywords *Agni, deva*, and īḍ appear 3 times, signifying Agni's central role as a human link to the gods. *Agni* is praised (īḍ) as the bearer of offerings and forwarder of human prayers to the divine world. This is in line with Sharma (2017) who asserts that Agni has a dual role of physical as fire, and spiritual as a cosmic priest who channels the energy of offerings.
3. The word ratna appears once, but semantically carries significant weight. Ratna here does not just mean a physical gem, but a symbol of grace, well-being and blessings. Jamison, & Brereton (2014) suggest that the *Rgveda* combines material and spiritual goals, so that prosperity is not just wealth, but also inner purity.
4. Keywords such as *Agni, yajña,* and *ṛtvij*, which appear 3 times, relate the performance of rituals to cosmic order (*ṛta*). Rituals tailored to the cycle of seasons (*ṛtu*) demonstrate the Vedic awareness of the importance of harmony between nature, humans, and the gods. Macdonell (1917) revealed that the rhythm of Vedic rituals is a reflection of the orderliness of the universe, which is the foundation of cosmic balance.

Theme Relation to Hindu Values

Rgveda Mandala 1 Sukta 1 places *Agni* as the center of the theme, both in ritual and spiritual contexts. The results of token analysis and lemmatization show that the word “Agni” has the highest frequency of occurrence, followed by words related to ritual activities such as *yajña* (holy sacrifice), *hotṛ* (officiating priest),

and *ratna* (boon/wealth). This pattern suggests that the core of this sukta message pivots on *Agni's* role as a link between humans and the gods.

From a Hindu perspective, the ritual themes that emerge are aligned with the concept of *Yajña* as a moral obligation (*svadharma*) to maintain cosmic harmony (*ṛta*). LDA topic modeling shows that terms related to offerings and rituals are not only present as liturgical symbols, but also contain social and ecological functions, namely maintaining the order of nature through conscious religious action.

A prominent theological theme, Agni as an intermediary reinforces the value of *Bhakti* (devotion) and *Śraddhā* (faith). In the text, Agni is not simply a physical element (fire), but rather a divine entity who brings offerings to the gods. This confirms the teaching that humanity's relationship with God is always through the medium of sacred energy that purifies intentions and deeds.

In addition, the invocation of *ratna* (prosperity) indicates that spiritual and material goals go hand in hand. In terms of *Puruṣārtha*, this represents the attainment of *Artha* (prosperity) without giving up the orientation towards *Mokṣa* (spiritual liberation). Thus, the *Rgveda* offers a holistic view, physical and spiritual well-being are part of an inseparable whole.

The theme of cosmic order (*ṛta*) also emerges through references to the cycle of time and seasons associated with the performance of *yajña*. Semantic analysis of the NLP results shows a lexical link between the concept of time (*ṛtu*) and the role of the seasonal priest (*ṛtvij*), signifying the realization that ritual practice should be in harmony with the rhythms of nature. This value is in line with the principle of *Ahimsa*, which is to maintain the balance of the ecosystem as part of dharma.

According to Radhakrishnan (1953), the Vedas are "moral and cosmological guides" that lead humans not only to pray, but also to live in harmony with the laws of nature and God. The findings show that, through the NLP approach, the thematic structure of *Rigveda Mandala 1 Sukta 1* can be read not only as a liturgical text, but also as an integrated map of Hindu values, including ritual, theology, spirituality, and ecology. If labeled as follows.

Major Themes	Description	Related Hindu Values	Description
Ritual (*Yajña*)	The manner of sacred offerings, the role of priests, and ceremonial procedures.	*Yajña* (sacred sacrifice), *Dharma* (moral obligation).	The Rgveda emphasizes the regularity of performing *yajña* as a way of maintaining cosmic harmony (ṛta). NLP found the words yajña, purohita, and ṛtvij recurrent, suggesting that the theme of ritual was central to the message.
Theological (Agni as Intermediary).	Agni is praised as the link between humans and gods	*Bhakti* (devotion), *Śraddhā* (faith)	Agni acts as a cosmic hotṛ. Token analysis shows the word Agni has a high frequency, confirming its role as the center of Vedic theology.
Spiritual & Material.	Goals Desire for physical and mental well-being	*Artha* (prosperity), *Mokṣa* (liberation)	The word ratna as a symbol of grace is not just material, but also spiritual well-being. NLP

Major Themes	Description	Related Hindu Values	Description
			associates this word with the context of seeking blessings.
Cosmic Order (*Ṛta*)	The relationship between the performance of rituals and the harmony of the universe.	*Ṛta* (cosmic order), Ahimsa (non-violence, maintaining the balance of nature)	The use of terms related to seasons (*ṛtu*) and the role of seasonal priests (*ṛtvij*) shows awareness of natural cycles. Semantic analysis shows a lexical connection between the concepts of time and ritual.

Comparison of AI Analysis Results with Traditional Philological Interpretation

The study of the *Rgveda Mandala 1 Sukta 1* through two different approaches, namely Artificial Intelligence (AI)-based analysis with Natural Language Processing (NLP) and traditional philological interpretation, shows differences in starting points and breadth of results. The AI-based analysis makes use of tokenization, lemmatization, and topic modeling to extract patterns of word occurrence and semantic relationships. As a result, dominant themes emerged such as the worship of Agni as a ritual medium (hotṛ), the importance of *yajña* as a medium of human-divine relationship, the invocation of ratna (prosperity), and the connection with the concept of ṛta (cosmic order). All of these are identified based on statistical calculations and the distribution of words in the text.

Traditional philological interpretations, meanwhile, depart from an in-depth study of the text, the Sanskrit root (*dhātu*), and the historical-theological context inherited in the paramparā tradition. Interpretations from figures such as *Ṛṣi Sāyaṇa* emphasize that Agni is not simply an element of fire or a ritual entity, but rather a symbol of divine consciousness that guides humans towards mokṣa. Moreover, yajña is understood not only as an act of physical sacrifice, but also as an inner offering performed with punctuality (ṛtu) and purity of intention.

In terms of advantages, AI is able to process Vedic text data massively and quickly, revealing lexical patterns that may escape manual observation. However, AI tends to be weak in capturing the depth of symbolism and cultural nuances at the core of sacred texts. In contrast, philological methods excel in providing rich philosophical meaning, but are time-consuming and highly dependent on expert competence.

The data in *Rgveda Mandala 1 Sukta 1 Mantra 1 agníṁ īḷe puróhitaṃ yajñásya devaṃ ṛtvijám hotāraṃ ratnadhắtamam* then compared in AI analysis results (NLP + LDA) → focus on keywords such as Agni, sacrifice, priest, wealth, which are thematically linked to rituals and material/spiritual boons. Traditional interpretation (Sayana & pandita interpretation) → emphasis on Agni as an intermediary between humans and gods, a symbol of cosmic energy, as well as a manifestation of God in the form of sacred fire, with meanings that go beyond the literal. This shows the difference in depth, nuance and context produced by AI compared to traditional philological and hermeneutic methods.

This analysis also shows that the two approaches are complementary. AI can serve as an objective and data-driven initial mapping, while philology provides in-depth and meaningful interpretation. As Bryant (2007)

points out, modern technology is not meant to replace philology, but to "extend its reach and enrich hermeneutic understanding." By combining the strengths of both, the study of Hindu literature can be developed into a study that is not only linguistically accurate, but also spiritually and philosophically complete.

Strengths and weaknesses of AI approaches in Vedic studies

The use of Artificial Intelligence (AI), particularly through Natural Language Processing (NLP) approaches, has opened up a new dimension in the study of Vedic texts. While traditional philological analysis requires long hours and deep involvement of a pandita or philologist, AI can process thousands of stanzas in a matter of seconds. This speed is not just about efficiency, but also enables researchers to discover previously hidden lexical and semantic patterns, such as the association of the term "*Agni*" with the concepts of prosperity (*ratna*) and cosmic order (*ṛta*). The consistency of algorithm-generated results also helps to reduce the subjective bias that is often inevitable in manual interpretation.

However, these strengths are limited by a fundamental weakness: AI has neither the spiritual awareness nor the philosophical understanding at the core of Vedic teachings. For example, the symbolic meaning of "Agni" is not just limited to 'fire' or "fire god", but encompasses the concept of divine consciousness acting as a link between humans and the gods. This dimension cannot be captured simply through word frequency calculation or topic modeling. AI also depends entirely on the dataset; if the text being analyzed comes from an inaccurate translation, then the analysis results will reflect inaccurately as well.

In addition, the Vedas live in an oral tradition that is full of nuances of intonation, rhythm and ritual atmosphere. AI approaches based on digital texts cannot reproduce this performative dimension. Even in the case of polysemic Sanskrit words, such as "Soma" which can refer to a deity, a plant, or a holy drink, AI tends to struggle to discern meaning without adequate manual annotation.

As such, AI is better positioned as a partner in research, rather than a replacement for traditional methods. AI's initial analysis can provide a broad linguistic map, while philosophical understanding, hermeneutics, and spiritual values remain the domain of Hindu religious and cultural experts. The integration of the two promises an approach that is not only fast and systematic, but also faithful to the depth of Vedic meaning.

Challenges and Opportunities of AI in Vedic Studies

The utilization of Artificial Intelligence (AI) in Vedic studies presents an interesting paradox: on the one hand it offers revolutionary opportunities, on the other hand it brings conceptual and technical challenges that are not simple.

In terms of opportunities, AI has the ability to process a large corpus of Vedic texts at a speed and accuracy that manual methods struggle to achieve. With the help of Natural Language Processing (NLP), researchers can extract semantic patterns, identify relationships between key concepts such as ṛta, dharma, and satya, and build thematic maps across mandalas or even across scriptures.

This potential can enrich intertextuality studies and facilitate comparative research between the Vedas and other Hindu philosophical literature, even to the texts of other world religions. Furthermore, the integration of AI with digital humanities enables the creation of an interactive corpus, which presents not only texts, but also philological metadata, phonetic reconstructions, and annotated commentaries from various traditions.

However, these opportunities must be balanced with an awareness of the challenges that come with them. First, AI relies on data quality and

completeness. Vedic texts often come in various versions, transliterations and translations, each of which carries certain nuances and biases. Second, Sanskrit is morphologically complex, polysemic and rich in layers of contextual meaning, making automated modeling prone to meaning reduction. Another challenge is the oral-traditional dimension of the Vedas which is difficult to capture by digital text formats. AI does not inherently understand the rhythms, intonations, or spiritual resonances that are integral to the Vedic ritual experience.

Moreover, there is an epistemological dilemma: can AI-generated knowledge be considered equivalent or complementary to the sacred śruti-based knowledge in the Hindu tradition? Here the need arises for a hybrid approach, where AI is used as an initial analytical tool, while the final validation is still done by religious scholars, philologists, and pandita who understand the depth of the spiritual and cultural context.

AI thus paves the way for a new era of Vedic studies that is more systematic, comprehensive and open to interdisciplinary collaboration. However, the success of its integration depends on the ability to combine technological precision with the wisdom of tradition, so that digital advancement does not come at the expense of the philosophical substance and spiritual values contained in the Vedas.

IV. CONCLUSION

This research proves that the application of Artificial Intelligence (AI) through the Natural Language Processing (NLP) approach can be an important breakthrough in the study of Hindu literature, especially the analysis of the *Ṛgveda Mandala 1 Sukta 1* text. The process of tokenization and lemmatization proved to be effective in reducing the morphological complexity of Sanskrit, allowing for the systematic extraction of key themes. The results of AI analysis identified four central themes of ritual (yajña), theology (Agni as mediator), spiritual-material purpose (ratna), and cosmic order (ṛta) that align with Hindu values.

A comparison with traditional philological interpretation shows that both approaches have complementary advantages. AI excels in speed, scale, and consistency of analysis, while traditional philology provides greater philosophical depth, historical context, and symbolic understanding. As such, AI should be positioned as a methodological partner rather than a replacement for the classical interpretive tradition.

Key challenges include AI's limited understanding of philosophical context, the risk of dataset bias, and the difficulty of capturing the oral-performative dimension of the Vedas. However, the opportunity of integrating AI with philological and digital humanities approaches offers new prospects for the preservation of ancient texts, enrichment of academic literature, and global dissemination of Vedic teachings. The successful utilization of AI in Vedic studies will be largely determined by the ability to combine technological precision with the wisdom of tradition, so that digital progress can go hand in hand with the preservation of the spiritual and philosophical values of Hindu heritage.

REFERENCE

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O Reilly Media.

Bryant, E. F. (2007). *Krishna: A Sourcebook*. Oxford University Press.

Gupta, P., Kumar, A., & Singh, R. (2023). Linguistically-Informed Neural Architectures for Lexical, Syntactic, and Semantic Tasks in Sanskrit. *Journal of Computational Linguistics, 49*(1), 55–78.

Jamison, S. W., & Brereton, J. P. (2014). *The Rigveda: The Earliest Religious Poetry of India.* Oxford University Press.

Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing (3rd ed.).* Draft.

Lowe, J. J. (2015). *Participles in Rigvedic Sanskrit: The Syntax and Semantics of*

Adjectival Verb Forms. Oxford University Press.
Macdonell, A. A. (1917). *A History of Sanskrit Literature*. William Heinemann.
Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing.* MIT Press.
Radhakrishnan, S. (1953). *The Principal Upanishads*. George Allen & Unwin.
Renou, L. (1953). *Religions of Ancient India*. Athlone Press.
Scarlata, S., & Widmer, G. (2025). Metrics, Semantics, and Syntax in Vedic Hymns: A Computational Linguistic Approach. *Journal of Indic Studies*, *17*(2), 45–68.
Sharma, A. (2017). *Classical Hindu Thought: An Introduction*. Oxford University Press.
Smith, J. (2021). AI-assisted textual analysis in digital humanities: Unlocking ancient manuscripts. *Journal of Digital Scholarship in the Humanities*, *36 (2)*, 245–260.
Staal, F. (1983). *Agni: The Vedic Ritual of the Fire Altar.* Asian Humanities Press.