

Generative AI Enabled Preservation and Global Promotion of Sanskrit Language: Opportunities, Challenges and Computational Framework

Dr.Arzoo

Assistant Professor, Department of Computer Science
Government College Jatauli Haily Mandi, Gurugram
Haryana, India

ABSTRACT

Sanskrit is among the world's oldest and most systematically structured languages, preserving extensive knowledge related to philosophy, medicine, astronomy, mathematics, linguistics, and literature. Despite its historical significance, Sanskrit faces challenges such as declining usage, manuscript deterioration, limited digital accessibility, and inadequate computational resources. Earlier studies emphasized Sanskrit's grammatical precision and its relevance in artificial intelligence and natural language processing. Recent advancements in Generative Artificial Intelligence (GAI), Natural Language Processing (NLP), machine learning, and digital humanities offer new opportunities for preserving and promoting Sanskrit globally. This paper proposes an AI-enabled framework for manuscript digitization, automated translation, semantic search, intelligent tutoring systems, and knowledge dissemination. The study also examines challenges related to data scarcity, ethical concerns, OCR limitations, and linguistic complexity. The paper concludes that collaborative efforts between Sanskrit scholars and AI researchers can ensure sustainable preservation of Sanskrit heritage.

Keywords:- Generative AI, Sanskrit, NLP, Digital Preservation, Machine Translation, Computational Linguistics

I.

INTRODUCTION

Sanskrit has contributed significantly to India's intellectual traditions through texts such as the Rigveda, Mahabharata, Ramayana, Charaka Samhita and Arthashastra.[2,7]

Although Sanskrit preserves ancient scientific and philosophical knowledge, it faces major problems:

- degradation of manuscripts
- shortage of Sanskrit experts
- limited computational datasets

- declining global usage
- accessibility barriers

Traditional preservation methods are no longer sufficient.

Generative AI tools such as OpenAI ChatGPT, Google Gemini, and transformer-based language systems can create new opportunities for preserving Sanskrit content at scale.

II. COMPUTATIONAL FOUNDATION OF SANSKRIT IN AI

Earlier research by Dr. Arzoo highlighted that Sanskrit possesses a highly formal grammatical structure that makes it highly suitable for computational processing.[1]

- **Role of Pāṇini's Ashtadhyayi**

Pāṇini's Ashtadhyayi is widely regarded as one of the earliest examples of a highly formalized rule-based linguistic framework. Its systematic arrangement of grammatical rules demonstrates principles that closely align with modern computational linguistics and artificial intelligence. The structured methodology used by Pāṇini reflects concepts similar to contemporary parsing algorithms, formal grammar models, computational rule engines, syntactic analysis, and language generation techniques. The precision and logical sequencing embedded in this ancient grammatical system continue to inspire modern Natural Language Processing (NLP) research, particularly in areas involving machine translation, semantic analysis, and intelligent language processing systems.[1]

- **Morphological Analysis**

Sanskrit's highly inflectional and morphologically rich structure makes it particularly valuable for computational linguistic research. The language's complex system of word formation, sandhi transformations, declensions, and derivational patterns provides a strong foundation for advanced Natural Language Processing applications. Computational tools developed for Sanskrit increasingly focus on tasks such as morphological analysis, root identification, lexical normalization, grammatical annotation, and syntactic

interpretation. These capabilities are essential for building efficient AI-driven systems for machine translation, automated text processing, semantic analysis, and intelligent retrieval of Sanskrit knowledge resources.[3]

- **Semantic Networks**

As discussed in previous work, A semantic network is a graph-based representation that captures the relationships between various elements in a language. In the case of Sanskrit, semantic networks can map out the intricate connections between words, concepts, and their semantic meanings, unraveling the language's rich tapestry. With the help of AI-driven natural language processing (NLP) algorithms, computers can analyze vast Sanskrit texts and build comprehensive semantic networks. These networks not only reveal the meaning of individual words but also shed light on the contextual associations and underlying structures within the language. By understanding the semantic relationships between words, AI systems can accurately interpret and translate Sanskrit texts, making them more accessible to a global audience. Semantic relationships in Sanskrit can support AI-based knowledge graphs and contextual understanding.

III. GENERATIVE AI APPLICATIONS IN SANSKRIT PRESERVATION

Generative AI offers innovative solutions for preserving Sanskrit literature and making ancient knowledge more accessible in the digital age. It supports various applications ranging from manuscript preservation to intelligent learning systems, helping bridge

traditional Sanskrit knowledge with modern technological advancements.

- **AI-Assisted Preservation of Ancient Manuscripts**

A large portion of Sanskrit literature exists in fragile manuscripts that are vulnerable to damage and loss. Generative AI helps preserve these resources through text extraction, image enhancement, and reconstruction of damaged content, converting them into sustainable digital formats.

- **Cross-Lingual Knowledge Transformation**

The complexity of Sanskrit often limits accessibility for modern readers. Generative AI can translate Sanskrit texts into English, Hindi, and regional languages, making ancient knowledge available to a wider audience.

- **Knowledge Compression and Interpretation**

Many Sanskrit texts are lengthy and difficult to understand. AI systems can generate summaries, simplified explanations, and interpretations to help learners and researchers understand key concepts more efficiently.

- **Smart Learning Environments for Sanskrit Education**

Generative AI can support personalized Sanskrit learning through virtual tutors that assist with grammar, pronunciation, vocabulary, and textual interpretation, making learning more interactive.

- **Intelligent Knowledge Discovery**

AI-powered semantic search systems can help researchers efficiently retrieve relevant information from large digital repositories of Sanskrit texts by identifying contextual relationships between concepts and documents.

IV. DIGITAL ECOSYSTEM FOR SANSKRIT PROMOTION

Several digital platforms have significantly contributed to the computational development and preservation of Sanskrit resources by providing structured datasets, digitized manuscripts, and searchable textual repositories.[4-6]

- **Digital Corpus of Sanskrit**

The Digital Corpus of Sanskrit provides linguistically processed Sanskrit texts that include morphological annotations, lexical information, and grammatical tagging. These structured datasets assist researchers in developing Natural Language Processing models for tasks such as parsing, word segmentation, and machine learning-based language analysis.

- **GRETIL**

GRETIL (Göttingen Register of Electronic Texts in Indian Languages) offers a large collection of digitized Sanskrit texts covering literature, philosophy, religious scriptures, and classical works. It supports researchers by providing machine-readable texts that can be used for corpus development, textual mining, and generative AI training.

- **SARIT**

SARIT provides digitally encoded Sanskrit texts in structured formats that enable efficient search, retrieval, and computational processing. It helps researchers perform textual analysis and supports AI systems in understanding contextual relationships within large Sanskrit datasets.

- **Muktabodha Digital Library**

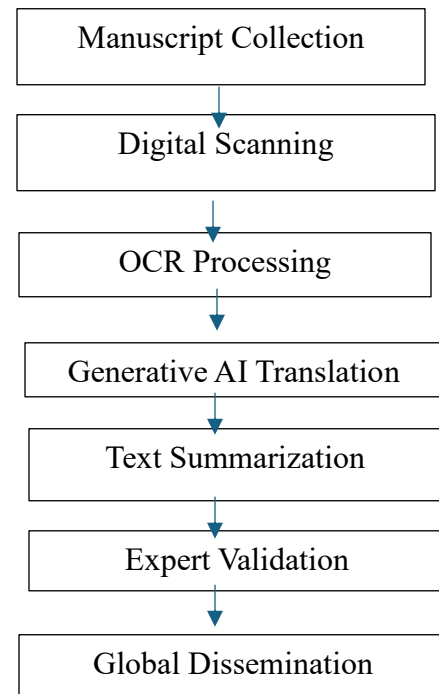
Muktabodha Digital Library focuses on preserving rare Sanskrit manuscripts, particularly traditional and ancient texts that are difficult to access physically. It offers valuable source material for manuscript digitization, semantic interpretation, and knowledge preservation initiatives using artificial intelligence.

Together, these platforms create a strong digital ecosystem that supports generative AI applications such as translation, text generation, semantic analysis, and preservation of Sanskrit knowledge systems.

V. PROPOSED GENERATIVE AI FRAMEWORK

The proposed framework presents a systematic approach for preserving and promoting Sanskrit literature through the integration of Generative AI technologies. It begins with manuscript collection and digitization, followed by OCR-based text extraction, AI-driven translation and summarization, and concludes with expert validation to ensure authenticity before making the content globally accessible through digital platforms. This framework ensures both technological efficiency and

scholarly reliability in preserving ancient Sanskrit knowledge.



VI. CHALLENGES

Despite the transformative potential of generative AI in Sanskrit preservation, its implementation faces several significant challenges. One of the primary issues is the limited availability of high-quality digital datasets, as Sanskrit lacks large annotated corpora and digitized textual resources required for training efficient AI models. Additionally, the complex grammatical structure of Sanskrit, including extensive morphological variations and compound formations, makes computational processing difficult. Ethical concerns also arise because many Sanskrit texts hold religious and cultural significance, requiring careful interpretation to preserve their original meaning. Furthermore, copyright restrictions and limited access to digitized manuscripts

may create barriers for large-scale research and data sharing. Another major concern is the possibility of AI-generated inaccuracies, where incorrect translations, summaries, or interpretations may distort the authenticity of ancient Sanskrit knowledge.

VIII. CONCLUSION

This paper extends earlier research on Sanskrit's computational relevance by integrating modern generative AI applications. While earlier work established Sanskrit's compatibility with AI systems, this study demonstrates how generative AI can actively preserve and globally promote Sanskrit in the digital era. Human supervision remains essential for authenticity and ethical preservation.

REFERENCES

- [1] Dr. Arzoo, "Significance of Sanskrit in Artificial Intelligence," *International Journal of Computer Science Trends and Technology (IJCTST)*, vol. 11, no. 4, pp. 98–101, Jul.–Aug. 2023.
- [2] R. Briggs, "Knowledge Representation in Sanskrit and Artificial Intelligence," *AI Magazine*, vol. 6, no. 1, 1985.
- [3] R. Sinha, "Sanskrit in the Digital Era: Challenges and Opportunities Ahead," SSRN, 2025.
- [4] A. Vasudevan et al., "SansGPT: Advancing Generative Pre-Training in Sanskrit," 2024.
- [5] ByT5-Sanskrit Research Paper, 2024.
- [6] SanskritShala NLP Toolkit, 2023.

- [7] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*.