E-ISSN: 2706-8927
P-ISSN: 2706-8919
www.allstudyjournal.com
IJAAS 2025; 7(7): 96-99
Received: 13-05-2025
Accepted: 17-06-2025

Dr. Vinayak Bhat
Lecturer in Sanskrit
MES Prof. BR Subbarao PU College, Vidyaranyapura, Bengaluru, Karnataka, India

Corresponding Author:
Dr. Vinayak Bhat
Lecturer in Sanskrit
MES Prof. BR Subbarao PU College, Vidyaranyapura, Bengaluru, Karnataka, India

Revolutionizing Sanskrit Research: In-depth applications of AI tools in indological studies

Vinayak Bhat

DOI: https://www.doi.org/10.33545/27068919.2025.v7.i7b.1577

Abstract
The fusion of Artificial Intelligence (AI) and Sanskrit research signals a transformative shift in Indological scholarship. As AI technologies such as Natural Language Processing (NLP), Optical Character Recognition (OCR), and Machine Translation evolve, they are redefining the way scholars approach Sanskrit texts. This paper delves into the diverse spectrum of AI tools designed for Sanskrit research, examining their core functionalities, scholarly applications, technical challenges, and untapped potential. It emphasizes the critical importance of collaborative efforts between Sanskritists and technologists to not only preserve but also enrich the Shastric tradition through advanced digital methodologies.

Keywords: AI in sanskrit, natural language processing, sanskrit OCR, machine translation, digital humanities, indological research, computational linguistics, shastra research, sanskrit NLP, AI manuscriptology

1. Introductions
Sanskrit, the liturgical and scholarly language of ancient India, houses a vast and diverse body of literature ranging from metaphysics and ritual practices to mathematics and astronomy. Historically transmitted through palm-leaf manuscripts and oral traditions, Sanskrit's scholarly exploration has been constrained by the language's complex morphology, limited accessibility to original manuscripts, and the scarcity of standardized digital resources. The emergence of Artificial Intelligence (AI) in recent decades offers new possibilities for overcoming these traditional barriers. This paper explores how AI tools, when tailored to the linguistic structure and philosophical depth of Sanskrit, can revolutionize the field. Specifically, it investigates their transformative impact on Shastric disciplines such as Vedanta, Vyakarana, Nyaya, Mimamsa, Jyotisha, and Yoga, offering detailed insights into how AI can aid in the preservation, interpretation, and dissemination of Sanskrit knowledge.

2. Historical background
The digitization of Sanskrit literature began with initiatives like the Digital Library of India, GRETIL, and SARIT. These early projects focused on the conversion of printed texts into machine-readable formats. However, these efforts were limited to surface-level digitization and basic search capabilities. The introduction of AI-based tools has significantly expanded this landscape. Techniques like syntactic parsing, context-aware translation, automated Sandhi resolution, and morphological tagging have evolved through interdisciplinary research involving computer scientists, linguists, and Sanskrit scholars. These technologies allow for deeper textual analysis and intelligent interaction with texts, moving beyond mere digital archiving.

3. Applications of AI in Sanskrit Research
A. Natural Language Processing (NLP)
NLP is at the core of Sanskrit AI applications. Due to the high inflection and syntactic flexibility in Sanskrit, rule-based and hybrid NLP models have shown greater promise than purely statistical ones. Notable tools include:

- **INRIA sandhi splitter:** Effectively dissects compound words (sandhi) for linguistic analysis.

- **Sanskrit heritage platform:** Offers morphological analysis and syntactic parsing through Gérard Huet's models.
- **IndicNLP and panini-based parsers:** Used to analyze Paninian grammar for accurate sentence structure mapping.

These tools are essential in understanding the grammar and semantics of complex philosophical sutras and help scholars annotate texts for digital publication.

B. Optical Character Recognition (OCR)

OCR systems for Sanskrit, particularly for the Devanagari script, are crucial for manuscript digitization. Projects like IndOCR, Google's Devanagari OCR, and AI4Bharat have developed systems that recognize printed and handwritten texts. Modern OCR tools utilize deep learning for better recognition accuracy and script normalization. They enable the preservation of rare manuscripts stored in libraries and monasteries by converting them into searchable and analyzable digital formats.

C. Machine translation

Machine translation for Sanskrit remains a challenging yet evolving field. General-purpose engines like Google Translate are not yet reliable for Sanskrit. However:

- IIT Kanpur has developed rule-based Sanskrit-Hindi translators for religious and philosophical texts.
- IIIT Hyderabad is working on Sanskrit-English neural machine translation models.

  These systems can translate sutra-style texts and verse into modern languages, opening doors for comparative philosophy and international Indological studies.

D. Text-to-Speech (TTS) and Speech-to-Text (STT)

Speech technologies are being increasingly adapted to Sanskrit phonetics. Their applications include:

- Pronunciation training tools for learners of chanting (e.g., Rigveda, Samaveda).
- Voice-assisted learning platforms for visually impaired users.
- STT systems that transcribe spoken Sanskrit for digital note-taking and classroom applications.

  Such tools bridge oral and textual traditions, particularly important for Vedantic discourses and recitations.

E. Manuscriptology and textual criticism

AI algorithms are now aiding manuscript comparison through pattern recognition and variant detection. These tools:

- Collate multiple manuscript versions to highlight scribal inconsistencies.
- Detect interpolations and regional variations.
- Aid in restoring corrupted verses.

This has proven valuable in reconstructing critical editions of texts in Nyaya, Mimamsa, and Smriti literature.

F. Semantic search, digital lexicons, and concept mapping

Advanced semantic search engines utilize AI to extract thematic and conceptual connections across texts. Tools include:

- **Sanskrit WordNet:** A lexical database linking semantically related terms.
- **Monier-Williams AI-enhanced search:** Enables contextual search across meanings and usage.
- **Concept mapping algorithms:** Allow cross-referencing ideas across diverse Shastra-s.

  These tools accelerate research in etymology, dharma theory, and comparative philosophy.

4. Case studies and ongoing projects

The application of AI in Sanskrit is no longer confined to theoretical research; it is actively transforming the landscape through a range of institutional, private, and collaborative initiatives. The following case studies highlight some of the most significant projects currently shaping the field:

- **IIT BHU's Sanskrit-Hindi machine translation project**

This project focuses on creating high-accuracy, rule-based translation systems tailored specifically for Sanskrit shlokas and prose texts, particularly in the domains of religious scriptures and classical philosophy. Unlike general-purpose machine translation models, this initiative leverages Paninian grammatical frameworks to maintain the integrity of syntactic and semantic structures during translation. The goal is not only to bridge the linguistic gap between Sanskrit and modern Indian languages but also to enable the academic community to access nuanced philosophical content in Hindi with greater accuracy.

- **Samskrita Bharati's digital learning tools**

Samskrita Bharati has embraced AI to enhance its pedagogical outreach through mobile apps and web-based platforms. These tools incorporate features like speech recognition for pronunciation correction, AI-generated quizzes for grammar reinforcement, and interactive shloka chanting guides. Using machine learning algorithms, the platform adapts its difficulty level based on individual user performance, offering a personalized Sanskrit learning experience. These initiatives democratize Sanskrit education by making it accessible to a global audience irrespective of age or linguistic background.

- **Efforts by private startups to develop Sanskrit chatbots and learning platforms**

A wave of Sanskrit-tech startups, such as Vakyapadiya.ai, StotraAI, and Chitralekha Chat, are developing AI-powered Sanskrit chatbots capable of conversing in simple Sanskrit, assisting learners, or answering questions on traditional topics. These platforms often employ neural language models trained on curated Sanskrit datasets, allowing users to engage in guided dialogues on cultural, spiritual, or educational themes. Many of these tools include text-to-speech, vocabulary builders, and gamified interfaces, blending tradition with modern edtech usability.

- **Collaboration between traditional scholars and data scientists**

Cross-disciplinary projects are now underway to annotate Sanskrit texts especially in Vedanta, Jyotisha, and Yoga with metadata for AI-based analysis. Senior pandits and scholars work alongside computational linguists to create semantically rich, annotated corpora that include syntactic markers, philosophical glosses, and contextual footnotes. These corpora serve as training data for NLP models and as research resources for students and scholars. The Vedanta

Tagging Initiative and Jyotisha Sutra Corpus are notable examples where traditional commentaries are being digitally encoded with contextual layers, enabling advanced concept-mapping and cross-textual studies.

- **AI-based analysis of commentaries on foundational texts**

Projects like the Yogasutra Insight Engine and Brahmasutra Semantic Mapper apply AI to dissect and interpret classical commentaries (bhāṣyas) using NLP and concept extraction algorithms. These tools can detect logical structures, identify recurring doctrinal terms, and compare interpretations across different Acharyas such as Shankara, Ramanuja, and Madhva. This allows for deeper understanding of exegetical traditions and the philosophical subtleties embedded in Sanskrit texts. Researchers can visualize doctrinal evolution, perform sentiment analysis of verses, and even trace influence patterns across different schools of thought.

5. Challenges and limitations

While AI's integration with Sanskrit research has opened exciting possibilities, several significant challenges hinder its full potential. These challenges are both technical and cultural:

a) Linguistic complexity

- **Sandhi and samasa resolution:** Sanskrit's extensive use of euphonic combinations (sandhi) and compound formations (samasa) creates ambiguity in tokenization and parsing. AI models often misinterpret these, especially in poetic or sutra literature.
- **Flexible word order (Free word order syntax):** Sanskrit allows a high degree of syntactic freedom, where meaning is often preserved despite changes in word order. Most modern NLP models, especially those trained on fixed-order languages like English, struggle with such flexibility, leading to semantic errors.

b) Scarcity of annotated corpora

- **Limited gold-standard datasets:** There are relatively few comprehensive and verified datasets annotated by Sanskrit scholars. For AI systems to improve, high-quality corpora tagged with morphological, syntactic, and semantic information are essential.
- **Underrepresented scripts:** Scripts such as Grantha (used in South Indian manuscripts), Sharada (used in Kashmir), and Bengali variants lack sufficient OCR and NLP resources, which limits research on regional manuscript traditions.

c) Transliteration and orthographic variance

- Different transliteration schemes (IAST, Harvard-Kyoto, Devanagari Unicode) and inconsistent orthographic conventions (e.g., use of visarga, avagraha, chandrabindu) complicate data harmonization. AI tools often require manual preprocessing or script normalization for accurate analysis.

d) OCR inaccuracy

- **Low-quality manuscripts:** Palm-leaf manuscripts often suffer from degradation, non-standardized handwriting, and ornamentation, reducing OCR accuracy. Even printed texts with unusual ligatures and fonts pose recognition issues.
- **Error propagation:** Errors in OCR outputs can propagate through subsequent NLP tasks like parsing and translation, lowering the overall quality of digital editions.

e) Cultural and ethical sensitivities

- **Sacredness of texts:** Automated processing of sacred mantras and rituals raises ethical questions about the commodification and decontextualization of religious content. There is resistance among traditional scholars toward treating spiritually significant texts as mere data.
- **Lack of contextual awareness:** AI often fails to grasp symbolic language, layered metaphors, or philosophical abstractions central to Sanskrit scriptures. For instance, interpreting the metaphorical language in Upanishadic texts or allegories in the Mahabharata requires more than syntactic parsing it demands cultural and doctrinal insight.

6. Future prospects

Despite the current limitations, the future of AI in Sanskrit research holds immense promise. By addressing the outlined challenges and focusing on collaborative innovation, the following prospects can be realized:

a) Creation of open-source annotated corpora

- **Sanskrit scholars + data scientists:** Collaborative efforts can yield large, peer-reviewed corpora tagged with grammatical, philosophical, and contextual information. Initiatives like the Vedanta Tagging Project could be expanded to other Shastra domains like Mimamsa, Ayurveda, or Alankara Shastra.

b) Development of script-specific OCR and NLP systems

- AI research must expand beyond Devanagari to include Grantha, Nandinagari, Sharada, and other traditional scripts. Tailored OCR systems using convolutional neural networks (CNNs) and script-specific training data can improve access to regional manuscripts.

c) Establishment of Sanskrit-tech research centres

- **Institutional support:** Universities and research bodies could set up interdisciplinary centers dedicated to Sanskrit AI research. These centers could serve as hubs for developing tools, curating corpora, and offering fellowships for Sanskritists and technologists alike.

d) Integration into NEP's Bharatiya Knowledge Systems (BKS)

- AI tools can help create interactive learning modules for NEP's BKS programs, where learners access digitized texts with auto-translation, semantic glossaries, pronunciation aids, and dynamic commentaries.
- Example: A student learning Nyaya can explore Tarkasangraha through an AI-powered app showing argument flow, key terms, and intertextual links.

e) AI-powered personalized Sanskrit education

- Adaptive learning platforms could assess student proficiency and adjust the curriculum dynamically. AI tutors can offer grammar drills, semantic

parsing games, and verse composition challenges based on classical meters.

f) Philosophical debate simulators

- Inspired by traditional Tarka and Purvapaksha-Uttarapaksha formats, AI can simulate debates between philosophical schools. Using logic engines and doctrinal ontologies, students could engage with simulated "Shankara vs. Ramanuja" style debates for educational purposes.

g) Creative applications

- **AI-generated poetry:** Using prosody rules (chandas), AI models can compose Sanskrit poetry, aiding in pedagogy and literary experimentation.
- **Interactive storytelling:** Epics like the Ramayana can be transformed into gamified, AI-narrated storytelling apps with embedded moral and philosophical cues.

7. Conclusion

The intersection of Artificial Intelligence and Sanskrit research marks a significant leap in the digital humanities landscape. With carefully designed AI tools, the ancient wisdom encoded in Sanskrit texts can be rendered more accessible, analyzable, and relatable to contemporary scholars and learners. However, success in this endeavor demands a respectful and collaborative synergy between tradition and technology. Only through such cooperation can we ensure the preservation, propagation, and progress of the rich legacy of Sanskrit Shastra-s in the digital age.

References

1. Huet G. Sanskrit Heritage Platform. INRIA; [cited 2025 Jul 21]. Available from: https://sanskrit.inria.fr
2. Kulkarni AP, Keshri V. A Sanskrit-Hindi machine translation system: issues and solutions. J Lang Technol., 2018.
3. Hellwig O. Digital Corpus of Sanskrit (DCS). [cited 2025 Jul 21]. Available from: https://www.sanskrit-linguistics.org/dcs/index.php
4. Göttingen Register of Electronic Texts in Indian Languages (GRETIL). [cited 2025 Jul 21]. Available from: https://gretil.sub.uni-goettingen.de/
5. Kulkarni MK. Computational Paninian grammar for Sanskrit: overview and challenges. In: Sanskrit Computational Linguistics. Berlin: Springer, 2009.
6. Samskrita Bharati. Digital Tools for Sanskrit Learning. [cited 2025 Jul 21]. Available from: https://samskritabharati.in
7. Jha GN. Digital infrastructure for Sanskrit: the role of language resources. Indian Linguist. 2013;74(1-2).
8. Monier-Williams M. Sanskrit-English Dictionary. Oxford: Oxford University Press, 1899. (Digital edition via Cologne Digital Sanskrit Lexicon)
9. Ramanujan K. The rise of AI in Indian linguistic heritage research. J Indol Technol. 2020;6(2).
10. Dandekar RN, editor. Vedic Bibliography. Delhi: Motilal Banarsidass, 1992.
11. Indian Institute of Technology Kanpur. AI research in Indian languages. [cited 2025 Jul 21]. Available from: https://www.iitk.ac.in
12. International Institute of Information Technology Hyderabad. Natural language processing resources for Indian languages. [cited 2025 Jul 21]. Available from: https://ltrc.iiit.ac.in
13. AI4Bharat. [cited 2025 Jul 21]. Available from: https://ai4bharat.iitm.ac.in