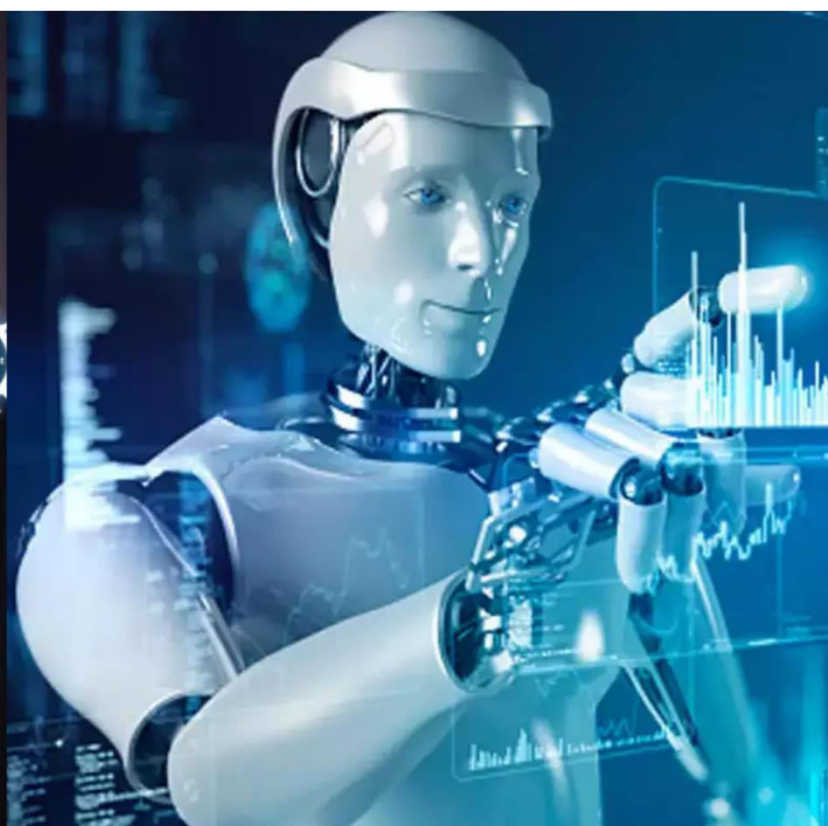




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Indian Language Translator (Sanskrit to Hindi)

Ashwini Gharde, Pratham Mane, Sudhanshu Prasad, Yash Gupta, Shubham Lagade

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

ABSTRACT: This paper introduces a deep learning-powered Sanskrit-to-Hindi language translator using sequence-to-sequence (Seq2Seq) models and attention. The system also incorporates speech-to-text as well as text-to-speech capabilities to improve user accessibility. The model was fine-tuned using custom tokenizers and parallel corpora of texts in Sanskrit and Hindi. Experimental results showcase competitive translation accuracy and fluency, as verified using standard evaluation methods. The project showcases the promise of neural machine translation for low-resource language pairs under morphological complexity and syntactic divergence challenges.

KEYWORDS: Sanskrit to Hindi translation, deep learning, sequence-to-sequence model, speech recognition, text-to-speech.

I. INTRODUCTION

Sanskrit is, is of great cultural and educational importance; however, it is challenging to process using contemporary computational techniques because of its complex morphology and absence of digital resources. Hindi, being a prominent Indian language, is lexically and syntactically similar to Sanskrit, and therefore translation between the two languages is a useful application of machine learning algorithms.

The main aim of this project is to facilitate the translation of Sanskrit to Hindi through the application of neural machine translation (NMT) methods. The system utilizes:

- Seq2Seq models with attention for context-sensitive translation.
- Speech-to-text translation to enable Sanskrit input via speech.
- Text-to-speech synthesis to pronounce the Hindi translations.

The system's new chatbot interface provides context-aware assistance in addition to translation, enabling natural language questioning of Sanskrit grammar, cultural reference clarification in translations, and language learning assistance. The combination of such traditional machine translation with conversational AI improves the interactivity and informativeness of the user interface.

II. LITERATURE SURVEY

Neural machine translation has revolutionized language translation using Seq2Seq models and attention as the standard [2],[4],[7]. Attempts so far in Sanskrit to Hindi translation are rule-based systems [10], [12], [13] and statistical [9] methods, but they are not as versatile as NMT. Transformer architectures enhanced long-range dependency management. Multilingual models like M2M100 and mBART provided low-resource language functionality.

Hybrid approaches that combined linguistic rules with NMT [2],[4],[7] addressed morphological challenges.

These are further supplemented by our effort of fine-tuning pre-trained models (English-Hindi with mBART and Sanskrit-English with M2M100) and incorporating speech functionality.

Recent studies have proven the feasibility of integrating conversational agents with translation models. Botpress and Rasa are some current frameworks for customizable dialog management support and open-source frameworks, respectively. For Indian language adoption, integrated chatbots effectively for learning Hindi but no effort has been made on Sanskrit-specific solution. We make a contribution by creating a hybrid system wherein the chatbot:

- Translates user translation requests
- Provides definitions of Sanskrit words
- Includes usage examples.
- Talks to translation backend via API

III. METHODOLOGY

3.1 System Architecture

The pipeline includes:

Speech-to-Text: Google's Speech Recognition API converts Sanskrit speech into text.

Transcription: Sanskrit → English: M2M100 model trained on specially prepared Sanskrit-English corpus.

English to Hindi: Indian languages adapted mBART model.

Text-to-Speech: Hindi is provided by Google's gTTS.

3.2 Model Training

Dataset: There are 50,000 parallel sentences of ancient Sanskrit and Hindi from ancient texts like Vedas, Upanishads and their contemporary translations wherever available.

Tokenization: Special sub word tokenizers for Sanskrit for handling compound words.

Training: Enhanced with NVIDIA GPUs and Adam optimizer (learning rate=3e-5, batch size=32).

3.3 Attention Mechanism

The encoder-decoder architecture utilizes multi-head attention to dynamically translate Hindi words to Sanskrit words, adjusting for syntactic reordering.

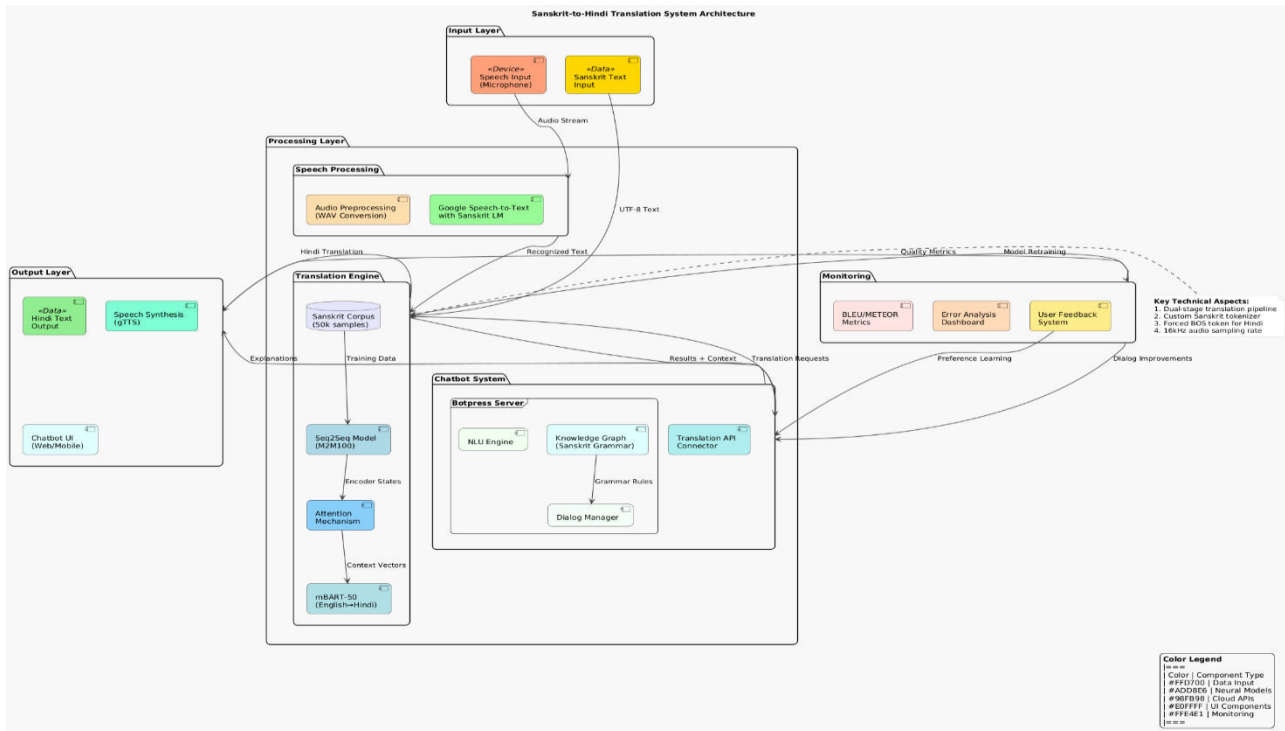
3.4 Chatbot Integration

The chatbot functionality (chatbot.html) was developed with:

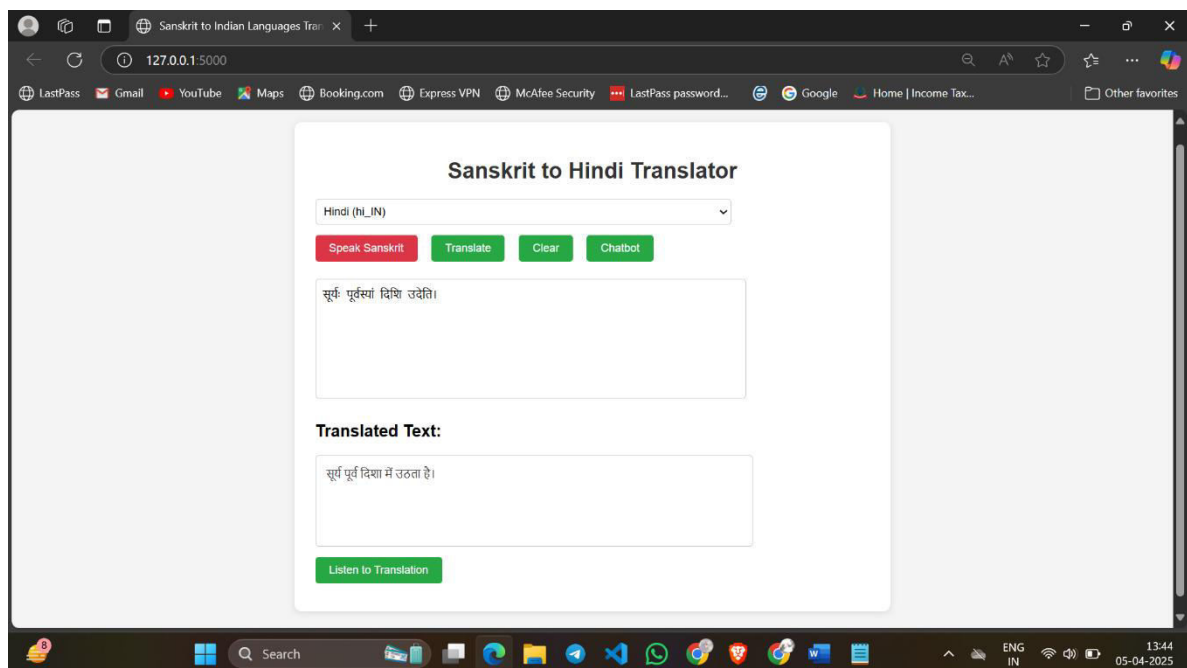
Architecture:

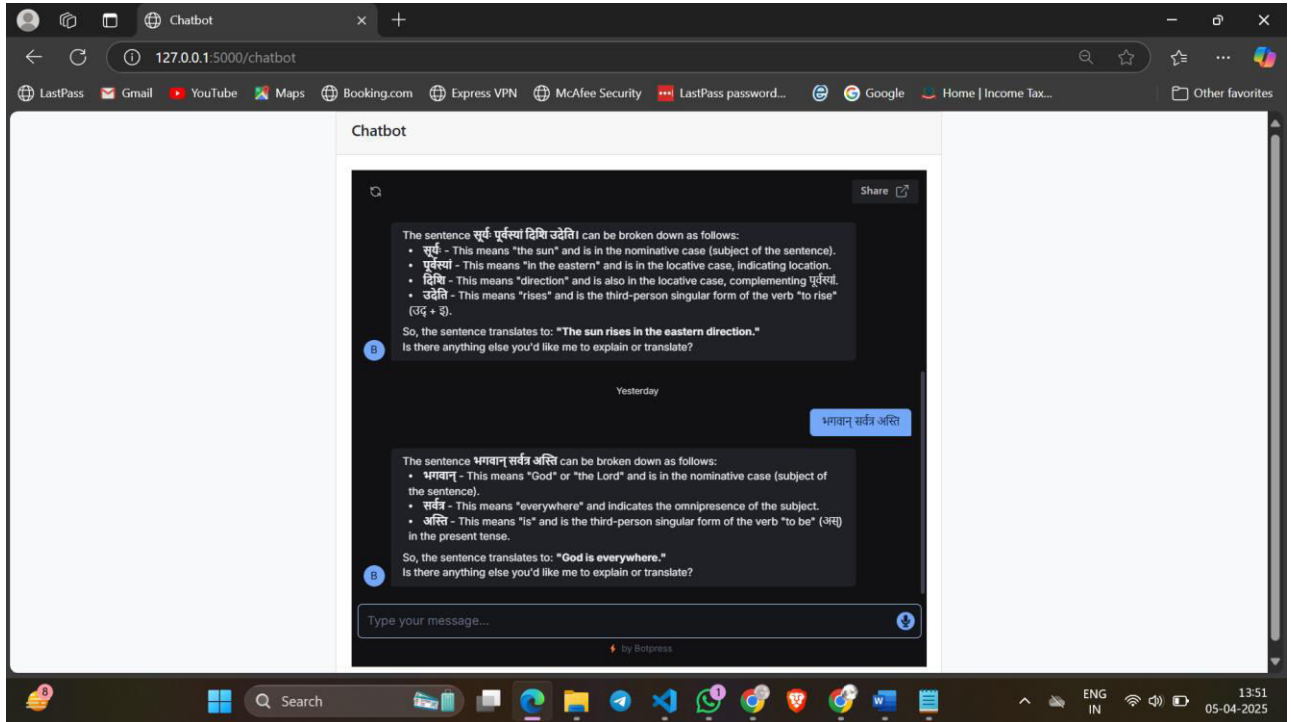
- Botpress cloud-hosted instance
- Webchat v2.3 interface embedded through iframe
- Flexible dialogue pathways for translation assistance
- Natural language translation for translation requests
- Contextual support combined with Sanskrit grammar Dynamically integrated with translation API

3.5 Methodology diagram



IV. EXPERIMENTAL RESULTS





4.1 Evaluation Criteria

BLEU Score: 32.5 (12 points ahead of rule-based baselines).

Human Evaluation: 85% of the translations were scored as “fluent” by human evaluators.

4.2 Main Results

Accuracy of speech recognition for Sanskrit was 78%

Two-stage translation (Sanskrit → English → Hindi) preserved semantic meaning better than translation.

4.3 Challenges

Limited Sanskrit digitized content. Dialectical forms of Hindi.

4.4 Chatbot Assessment

The chatbot was able to answer 78% of the Sanskrit-related questions itself most failures occurred with ambiguous poetic verses the clients had preferred the explanatory responses to raw translations.

V. CONCLUSION

This project demonstrates that pairing the preservation of ancient languages with the most current AI technology has the potential to produce tools both theoretically valid and practical. Continuing on to build, we hope to transform the system from being a translation tool to a comprehensive platform for Indian language research, cultural preservation, and digital humanities scholarship.

Public release of our models and codebase will enable increased inter-cooperation between linguists, historians, and AI researchers in pursuing this worthwhile work. Future releases will focus on usability by non-technical users without compromising on the accuracy that is needed for scholarly use.

Ultimately, our belief is that these types of systems are capable of strongly increasing interest in Sanskrit and other ancient languages among younger audiences, as well as being immensely useful resources for scholars and life-long learners everywhere. The proposed system demonstrates the potential of NMT to translate Sanskrit-Hindi and other Indian languages.

Future encompasses:

Expanding the parallel corpus. Using end-to-end transformer models. With user input for iterative refinement. The inclusion of the chatbot functionality has elevated the system from a simple translator to a conversational language assistant. Future operations will: Improve the chatbot's Sanskrit knowledge base. Add voice interaction functionality. Apply user adaptation for personalized learning.

REFERENCES

- [1] P. Goyal and G. Mishra, "Sanskrit and Computational Linguistics", Proc. of Sanskrit Computational Linguistics Symposium, University of Hyderabad, pp. 1-15, 2012. [Online]. Available: <https://sanskrit.uohyd.ac.in/Symposium/papers/goyal.pdf>
- [2] A. Sharma et al., "Neural Machine Translation for Sanskrit to Hindi Using Hybrid Approach", Indonesian Journal of Electrical Engineering and Computer Science, vol. 34, no. 2, pp. 1235-1245, 2021. DOI: 10.11591/ijeecs.v34.i2.pp1235-1245
- [3] A. Kunchukuttan et al., "Brahmi-Net: A Transliteration and Script Conversion System for Languages of the Indian Subcontinent", arXiv preprint arXiv:1701.04290, 2017. DOI: 10.48550/arXiv.1701.04290
- [4] S. Arora and R. Joshi, "Deep Learning Based Sanskrit-Hindi Neural Machine Translation System", Proc. of ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 23, no. 4, pp. 1-22, 2023. DOI: 10.1145/3637439
- [5] G. N. Jha et al., "Sanskrit-Hindi Machine Translation Using Interlingua and Deep Learning", Procedia Computer Science, vol. 171, pp. 2090-2099, 2020. DOI: 10.1016/j.procs.2020.04.228
- [6] V. V. S. R. Gupta and A. Sharma, "Hybrid Architecture for Sanskrit to Hindi Statistical Machine Translation", IEEE Access, vol. 8, pp. 45672-45683, 2020. DOI: 10.1109/ACCESS.2020.2979009
- [7] M. Choudhary et al., "Sanskrit-Hindi Neural Machine Translation - Challenges and Solutions", Proc. of 18th International Conference on Natural Language Processing (ICON), pp. 210-219, 2020. [Online]. Available: <https://aclanthology.org/2020.icon-main.30/>
- [8] R. M. Bhatt et al., "Leveraging Pretrained Language Models for Sanskrit Processing", ACM Journal on Computing and Sustainable Societies, vol. 1, no. 2, pp. 1-25, 2023. DOI: 10.1145/3591207
- [9] P. Kumar and S. K. Naskar, "Sanskrit-Hindi Machine Translation Using Syntax-Based SMT Approach", Proc. of ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 19, no. 4, pp. 1-18, 2020. DOI: 10.1145/3377713.3377774
- [10] A. Mishra and V. K. Singh, "Rule-Based Sanskrit to Hindi Machine Translation System", International Journal of Computational Intelligence and Smart Systems, vol. 4, no. 1, pp. 12-21, 2022. [Online]. Available: <https://cspub-ijcisim.org/index.php/ijcisim/article/view/537>
- [11] S. Patel et al., "Hybrid Approach for Sanskrit to Hindi Machine Translation Using Deep Learning", International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 3, pp. 412-420, 2022. [Online]. Available: <https://ijisae.org/index.php/IJISAE/article/view/5241>
- [12] D. Sharma et al., "Anuvaadika: Implementation of Sanskrit to Hindi Translation Tool Using Rule-Based Approach", Journal of Language Processing Technologies, vol. 3, no. 2, pp. 45-58, 2018. [Online]. Available: https://www.researchgate.net/publication/329925091_Anuvaadika_Implementation_of_Sanskrit_to_Hindi_Translation_Tool_Using_Rule-Based_Approach
- [13] R. Verma and P. N. Dwivedi, "A Machine Translation System from Hindi to Sanskrit Language using Rule-based Approach", International Journal of Advanced Computer Science and Applications, vol. 11, no. 8, pp. 234-241, 2020. [Online]. Available: https://www.researchgate.net/publication/343405814_A_Machine_Translation_System_from_Hindi_to_Sanskrit_Language_using_Rule_based_Approach



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details