Asian Journal of Information Science and Technology
ISSN: 2231-6108 (P); 3049-2661 (E)
Vol.15 No.1, 2025, pp.25-29

www.crijournals.org
DOI: https://doi.org/10.70112/ajist-2025.15.1.4320

A Perspective on the Use of Sanskrit Language and Literature in Developing AI and GenAI Systems

Sushant K. Singh
Director, The Center for Artificial Intelligence and Environmental Sustainability (CAIES) Foundation,
Anisabad, Patna, Bihar, India
E-mail: sushantorama@gmail.com
(Received 31 December 2024; Revised 30 January 2025; Accepted 18 February 2025; Available online 28 February 2025)

Abstract - **Artificial intelligence (AI) and generative AI (GenAI) have transformed various industries by enabling machines to understand, interpret, and generate human languages with remarkable precision, a capability popularly known as natural language processing (NLP). While dominant languages such as English and Mandarin have traditionally played a significant role in AI and GenAI model training, there is growing interest in exploring the potential of Sanskrit for AI and GenAI system development. Using its highly structured grammar and rich semantic framework, this article explores how Sanskrit offers unique advantages to AI and GenAI systems. Sanskrit's deterministic and precise grammatical rules, as codified in Maharishi Panini's Ashtadhyayi, present an opportunity to reduce ambiguity and enhance the computational efficiency of NLP models. The language's inflected morphology allows for more compact and flexible expressions, which may improve AI's ability to handle complex word relationships. Sanskrit's cultural and philosophical significance also enables AI to engage with ancient wisdom and interdisciplinary research. However, challenges such as the limited availability of modern Sanskrit corpora, the lack of native speakers, and the computational complexity of its grammatical rules must be addressed to realize its full potential. Despite these challenges, incorporating Sanskrit into AI and GenAI systems could lead to innovations in linguistic research, philosophical AI, and computational logic.**
Keywords: **Sanskrit, Generative AI (GenAI), Natural Language Processing (NLP), Panini's Ashtadhyayi, Computational linguistics.**

I. INTRODUCTION

Artificial intelligence (AI) and generative AI (GenAI) drive fundamental transformations across almost every industry. The automation, efficiency, and creativity enabled by these technologies push boundaries and create previously unimaginable opportunities. For example, AI has been leveraged in healthcare to improve diagnostics, treatment plans, and drug discovery (Ueda, Kakinuma, *et al.,* 2024). AI systems can help doctors quickly examine large amounts of medical data, including patient records, medical images, and research studies. This allows them to diagnose diseases more accurately and recommend better treatment plans (Topol, 2019; Ueda, Kakinuma, *et al.,* 2024). In genomics, AI models help identify genetic markers for diseases and accelerate the development of personalized medicine.

Similarly, in finance, AI has redefined the sector by optimizing trading algorithms, fraud detection, and personalized banking services (Jain, 2023). AI models analyze market data to predict trends, manage portfolios, and streamline risk assessment. Robo-advisors powered by AI provide personalized investment strategies and democratize access to financial services. In manufacturing and supply chains, AI technologies enable predictive maintenance, robotics, and supply chain optimization, resulting in reduced downtime and increased efficiency. AI-driven robotics can handle complex manufacturing tasks, whereas natural language processing (NLP) helps optimize logistics through demand forecasting and real-time monitoring (Hughes, Dwivedi, *et al.,*2022).

In addition, GenAI systems have revolutionized content creation, including news writing, video game design, and film production. GenAI-generated music, art, and stories are becoming increasingly popular, with AI co-authors and composers contributing to the creative industry. For instance, GPT-4 can write stories, generate plot ideas, and draft articles that mimic human creativity (Brown, 2020). GenAI models also help generate content tailored to individual tastes, such as personalized playlists or customized product suggestions. AI-driven tools transform education by enabling personalized learning. Adaptive learning platforms use AI to assess students' strengths and weaknesses and offer customized lessons and feedback (Luckin & Holmes, 2016). GenAI models generate learning material summaries and answer student queries in real time.

Furthermore, AI and GenAI systems have helped address critical global challenges. AI technologies are applied to model climate patterns, optimize renewable energy sources, and reduce waste. GenAI systems can simulate environmental impacts and propose solutions for mitigating the effects of climate change (Rolnick, Donti, *et al.,* 2022). During the COVID-19 pandemic, AI systems were used to model the spread of the virus, optimize resource allocation, and even aid vaccine development by analyzing viral genomes. AI-driven solutions improved contact tracing and helped governments make informed decisions to contain the virus (Vaishya, Javaid, *et al.,* 2020). AI applications in agriculture help farmers optimize crop yields through precision farming techniques and reduce food scarcity.

GenAI models analyze environmental conditions and recommend strategies for sustainable farming, providing solutions to global hunger (Casten Carlberg & Jerhamre, 2021).

In addition to helping various industries, AI and GenAI have enhanced everyday human life. AI-powered virtual assistants, such as Alexa, Siri, and Google Assistant, have become vital tools for managing daily tasks, setting reminders, and providing quick information. These assistants use NLP to understand and respond to voice commands, improving human-computer interactions. AI algorithms behind platforms such as Netflix, Amazon, and Spotify analyze user preferences to offer personalized content recommendations, significantly enhancing the user experience (Gomez-Uribe & Hunt, 2015). Therefore, it is critical to understand the importance of NLP in developing AI and GenAI systems.

II. ROLE OF NATURAL LANGUAGE IN DEVELOPING AI AND GenAI SYSTEMS

NLP has played a central role in modern advancements in AI and GenAI systems. By enabling machines to interpret, understand, and generate human language, NLP significantly expands AI systems' capabilities across various applications, from chatbots and virtual assistants to automated content creation and sentiment analysis (Cambria & White, 2014; Chowdhary & Chowdhary, 2020). This transformation reshapes how machines communicate and interact with humans.

NLP is a subfield of AI focused on enabling machines to comprehend and process human language. It combines computational linguistics, computer science, and cognitive psychology to enable machines to recognize, interpret, and generate language data in meaningful ways (Cambria & White, 2014; Chowdhary & Chowdhary, 2020). Through tokenization, parsing, sentiment analysis, and entity recognition, NLP systems can process unstructured data from text or speech, transforming it into a format understandable to machines.

NLP enhances traditional AI systems by allowing them to interact effectively with human language. Early AI systems relied heavily on structured data, but NLP introduced the ability to handle the ambiguity and variability inherent in natural languages (Cambria & White, 2014; Chowdhary & Chowdhary, 2020). This enables machines to understand user queries better, perform accurate text-based searches, and engage in limited conversational exchanges. For instance, NLP-powered systems such as IBM's Watson demonstrated that AI could outperform humans in tasks like answering trivia questions by accurately parsing and understanding natural language (Ferrucci, Brown, *et al.,* 2010).

GenAI refers to systems that learn from existing data to create new content-text, images, music, or videos. Large language models (LLMs) such as GPT-3, GPT-4, and BERT exemplify how NLP has transformed GenAI by enabling machines to generate human-like text with unprecedented fluency and coherence (Brown, 2020). The core of these models relies on NLP techniques such as language modeling and machine translation to generate text that can mimic human dialogue or produce narratives, articles, and essays from prompts. GenAI systems use NLP to understand and maintain the context of input, making the generated output highly relevant and meaningful.

The ability to produce human-like, cohesive, and creative text stems from advancements in NLP, which allow GenAI systems to simulate human cognitive patterns of language use. In a recent study, the authors reported that LLMs outperformed several GenAI models trained in Indian languages, including Sanskrit, Hindi, and Bangla (KJ, Jain, *et al.,* 2024).

III. THE INFLUENCE OF LANGUAGE CHOICE ON AI AND GenAI SYSTEMS

A. Language Representation and Its Impact

One of the most critical factors in developing AI and GenAI systems is the choice of the language used during model training. Language diversity, structure, and even specific linguistic features of a given language can significantly influence the robustness and adaptability of an AI system across different regions and user bases (Alexandris, 2024).

B. Data Availability and Language Representation

GenAI models such as GPT-4 are typically trained on vast datasets predominantly composed of English-language data. This leads to superior performance in English while potentially compromising effectiveness in other languages. AI and GenAI models trained in languages with extensive online presence, such as Mandarin, Spanish, and French, benefit from rich datasets, enhancing their robustness and capability in these languages (Bender, Gebru, *et al*., 2021). However, languages with limited digital representation may result in weaker performance, highlighting the issue of linguistic inequity in AI and GenAI.

C. Multilingual Models and Cross-Language Transfer

Recent advancements in multilingual NLP models, such as mBERT and XLM-R, have shown how AI systems benefit from multilingual training. These models demonstrate that training in multiple languages can help AI systems generalize better across different linguistic contexts, reducing reliance on a single dominant language and improving robustness in low-resource languages (Conneau, 2019). Cross-lingual transfer learning allows AI to leverage shared linguistic features, thereby improving the performance of less-represented languages based on knowledge learned from others.

D. Cultural and Semantic Differences

Languages carry cultural and semantic nuances that vary widely among different linguistic groups. For example, how politeness, formality, and emotions are conveyed differs between languages, impacting the performance of AI systems trained predominantly in one language but used in another. For AI systems to perform optimally across languages, NLP models must consider cultural and contextual differences (Bender, Gebru, *et al*., 2021).

E. Bias in Language Models

The choice of training languages and data sources introduces potential biases in AI systems. Models trained on English-heavy datasets may implicitly learn biases rooted in Western cultures, thus failing to account for the linguistic diversity and cultural contexts of non-English-speaking populations. This can compromise the fairness and inclusivity of the models, affecting their deployment in global applications (Mehrabi, Morstatter, *et al*., 2021).

F. Performance in Low-Resource Languages

Many languages worldwide lack the extensive digital corpora required to train robust AI systems. This leads to weaker NLP performance in low-resource languages, where models may struggle to understand, generate coherent sentences, or correctly interpret user intent. Ensuring that AI systems perform robustly across various languages is a significant challenge for NLP researchers, who must develop data augmentation and transfer learning techniques to improve language inclusivity (Roundtree, 2023).

G. Understanding Nuance and Ambiguity

Natural language includes subtleties such as idioms, sarcasm, and cultural references. GenAI models still struggle to fully capture these nuances, leading to misinterpretations or inappropriate responses (Bender, Gebru, *et al.,* 2021).

H. Data Privacy

As NLP models require vast amounts of text data, concerns have arisen regarding the collection and storage of personal information. Balancing the need for large datasets with the privacy rights of individuals is a significant challenge for NLP and AI researchers (Roundtree, 2023).

IV. SANSKRIT AND ITS POTENTIAL IN AI AND GenAI

Sanskrit, an ancient language of the Indian subcontinent, has drawn the interest of AI researchers for several reasons. Its rich grammatical structure, logical syntax, and clarity make it an exciting candidate for developing AI and GenAI systems. Some linguists consider Sanskrit one of the most "scientifically structured" languages because of its precise and logical grammatical system, as described in Maharishi Panini's Ashtadhyayi. This ancient treatise laid down an exhaustive framework of rules for generating valid Sanskrit sentences (Bharati, Chaitanya, *et al*., 1995; van Wyk, 2015). More than three decades ago, Briggs (1985) had already provided evidence that Sanskrit could be a better language for training AI systems; however, his recommendation has not yet materialized.

Sanskrit, an ancient Indian language developed before the Rigveda, the world's earliest known text, was compiled between 6500 B.C. and 1500 B.C. (Vasu, 1897; CSU, 2024). The language used in the Vedas, called Vedic Sanskrit, was slightly different from modern Sanskrit and was explained through grammatical texts known as Pratishakhyas (Vasu, 1897; CSU, 2024). Over time, various grammar schools contributed to a vast body of Vedic literature, including the Vedas, Brahmanas, Aranyakas, and Upanishads (Vasu, 1897; CSU, 2024). Around 500 B.C., Panini's Ashtadhyayi became a landmark in Sanskrit's development, consolidating grammar of the time. His system shaped both literary and spoken Sanskrit and remains the standard for evaluating the correctness of the language today (CSU, 2024).

Sanskrit was the original language in which the Vedas, Ramayana, Mahabharata, and Gita were written, and it is known that these events occurred as early as 12209 B.C.E. (Oak, 2024). The world's first residential university, Nalanda University, had millions of books in Sanskrit, covering science, mathematics, astronomy, physics, Ayurveda, yoga, material science, and literature. However, a Muslim tyrant, Khilji, killed most scholars and burned the entire library, which was said to have burned for months (Khichar, 2024). During that period, several Chinese and other scholars who either studied at Nalanda or visited succeeded in carrying scriptures to China and other parts of the world. India went through several invasions and was ruled by foreigners. As a result, the education system and Sanskrit practitioners, primarily in Hindu temples, were systematically destroyed. Over time, many foreign scholars translated the Vedas, Ramayana, Mahabharata, and Gita into English and other languages.

In this regard, the father of the atomic bomb, J. Robert Oppenheimer, was influenced by the Bhagavad Gita and the Mahabharata; he studied Sanskrit and read the Gita in its original form (Hijiya, 2000). Likewise, Erwin Schrödinger (Nobel Laureate, 1933; Bitbol & Darrigol, 1992), Werner Heisenberg, Niels Bohr (Nobel Peace Prize; Prothero, 2011), Carl Sagan, Nikola Tesla, and many more referred to Indian ancient scriptures in Sanskrit, which had a significant impact on their personal and professional lives. This further suggests that Sanskrit was not restricted to a single country or continent but was meaningful to those spiritually connected to Indian consciousness.

A. Advantages of Sanskrit in AI and GenAI

1. Highly Structured Grammar

Sanskrit is often considered one of the most grammatically precise languages worldwide. Panini's grammar, which consists of approximately 4,000 rules, provides a deterministic way of generating valid sentences. This strict rule-based structure can be leveraged to develop highly efficient and error-resistant algorithms for NLP tasks (Bharati, Chaitanya, *et al*., 1995). Precision in Sanskrit's syntax can reduce ambiguity, making it easier for AI to parse and understand.

2. Rich Semantic Framework

Sanskrit has an extensive vocabulary of words that convey specific meanings without heavily relying on context, which is often necessary in other languages. Sanskrit's grammatical structure includes eight cases (vibhaktis) that clarify relationships between words in a sentence, significantly reducing ambiguity (Bathulapalli, Desai, *et al*., 2016). This semantic clarity may enhance GenAI's ability to generate meaningful content with reduced risk of misinterpretation.

3. Inflection and Morphology

Sanskrit uses a highly inflected morphological system, modifying words based on roles in a sentence (nouns, verbs, adjectives, etc.). This feature allows compact and flexible expression of ideas. AI models trained on morphologically rich languages like Sanskrit may develop better capacities to handle complex word relationships and improve understanding of other morphologically complex languages (Bathulapalli, Desai, *et al*., 2016).

4. Cultural and Philosophical Significance

Sanskrit is closely linked to ancient Indian texts such as the *Vedas* and *Upanishads*, which hold vast knowledge of philosophy, physics, mathematics, medicine, and astronomy. Incorporating Sanskrit into AI and GenAI systems could help preserve and analyze these texts digitally, opening pathways for AI models to engage with ancient wisdom and interdisciplinary research (Bathulapalli, Desai, *et al*., 2016).

5. The Uniqueness of Sanskrit

The systematic grammatical framework, especially as articulated in Panini's Ashtadhyayi (with approximately 3,959 syntax and semantics rules), enables precise semantic representation-critical for AI applications that require accurate language understanding and generation (Bathulapalli, Desai, *et al*., 2016).

6. Rich Literature and Cultural Heritage

Sanskrit has a vast literary corpus, including philosophical texts, epics, and scientific treatises. This repository can serve as valuable training data, enabling AI to produce culturally and contextually relevant content (Singh, Kumar, *et al*., 2020). Translating texts like the Bhagavad Gita can also enhance cross-cultural understanding.

7. Potential for Knowledge Representation

Sanskrit's formal grammar supports effective knowledge representation, a crucial AI capability. Its parallels with modern programming languages suggest Sanskrit could inspire efficient AI algorithms. Semantic nets in AI can align closely with Sanskrit grammar, improving information processing (Briggs, 1985).

B. Challenges in Training AI and GenAI on Sanskrit

1. Limited Resources

Despite its strengths, developing AI trained in Sanskrit faces challenges due to limited linguistic resources and computational tools. While efforts exist to build morphological analyzers and parsers, these tools lag behind those for widely spoken languages (Mishra & Mishra, 2012).

2. Complexity of Morphological Analysis

Sanskrit's rich morphology makes parsing complex. The intricacies of *vibhakti* and *karaka* relations demand sophisticated algorithms that can manage these complexities (Mishra & Mishra, 2012). Effective systems may require hybrid rule-based and statistical methods.

3. Lack of Native Speakers

Although Sanskrit has a rich literary heritage, it is not widely spoken today. NLP and GenAI models often benefit from conversational data provided by native speakers. The limited speaking population may constrain real-world applications like chatbots (Huet, Kulkarni, *et al*., 2009). However, increasing interest in learning Sanskrit could gradually meet this need.

4. Complexity in Training AI Models

While Panini's rule-based grammar brings clarity, its complexity (~4,000 rules) can challenge computational efficiency in large-scale AI systems (Bathulapalli, Desai, *et al*., 2016). Designing architectures to balance precision with efficiency is essential.

5. Cultural and Philosophical Depth

Sanskrit words often carry layered meanings, posing difficulties for AI and GenAI models in generating nuanced interpretations (Deshpande, 2019). Capturing these depths requires sophisticated models and extensive datasets.

V. CONCLUSION

Training AI and GenAI systems in Sanskrit presents a unique opportunity to leverage the language's unambiguous structure and rich literary heritage. Although challenges such as limited resources and the complexity of morphological analysis exist, the potential benefits of reducing ambiguity and enhancing knowledge representation are significant. As research in this area continues to evolve, Sanskrit could play a pivotal role in the future of AI and GenAI by bridging cultural gaps and enriching the technological landscape.

Declaration of Conflicting Interests
The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation
The authors confirm that no AI-assisted technologies were used in the preparation or writing of the manuscript, and no images were altered using AI.

REFERENCES

[1] Alexandris, C. (2024). *GenAI and socially responsible AI in natural language processing applications: A linguistic perspective*. Proceedings of the AAAI Symposium Series.

[2] Bathulapalli, C., Desai, D., & Kanhere, M. (2016). Use of Sanskrit for natural language processing. *International Journal of Sanskrit Research, 2*(6), 78–81.

[3] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

[4] Bharati, A., Chaitanya, V., Sangal, R., & Ramakrishnamacharyulu, K. (1995). *Natural language processing: A Paninian perspective*.

[5] Bitbol, M., & Darrigol, O. (1992). Erwin Schrödinger. *Philosophy and the Birth of Quantum Mechanics*.

[6] Briggs, R. (1985). Knowledge representation in Sanskrit and artificial intelligence. *AI Magazine, 6*(1), 32.

[7] Brown, T. B. (2020). Language models are few-shot learners. *arXiv preprint*, arXiv:2005.14165.

[8] Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine, 9*(2), 48–57.

[9] Casten Carlberg, C. J., & Jerhamre, E. (2021). Artificial intelligence in agriculture: Opportunities and challenges.

[10] Chowdhary, K., & Chowdhary, K. (2020). Natural language processing. In *Fundamentals of Artificial Intelligence* (pp. 603–649).

[11] Conneau, A. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint*, arXiv:1911.02116.

[12] CSU. (2024). About Sanskrit. Retrieved October 20, 2024, from https://www.sanskrit.nic.in/about_sanskrit.php

[13] Deshpande, M. M. (2019). Scope of early Sanskrit usage: A wider approach.

[14] Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J. W., Nyberg, E., & Prager, J. (2010). Building Watson: An overview of the DeepQA project. *AI Magazine, 31*(3), 59–79.

[15] Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems, 6*(4), 1–19.

[16] Hijiya, J. A. (2000). The "Gita" of J. Robert Oppenheimer. *Proceedings of the American Philosophical Society, 144*(2), 123–167.

[17] Huet, G., Kulkarni, A., & Scharf, P. (2009). *Sanskrit computational linguistics*. Springer.

[18] Hughes, L., Dwivedi, Y. K., Rana, N. P., Williams, M. D., & Raghavan, V. (2022). Perspectives on the future of manufacturing within the Industry 4.0 era. *Production Planning & Control, 33* (2–3), 138–158.

[19] Jain, R. (2023). Role of artificial intelligence in banking and finance. *Journal of Management and Science, 13*(3), 1–4.

[20] Khichar, V. (2024). Nalanda University: A dream Asian university of the 21st century. *CHETANA International Journal of Education, 9*(1), 77–79.

[21] KJ, S., Jain, V., Bhaduri, S., Roy, T., & Chadha, A. (2024). Decoding the diversity: A review of the Indic AI research landscape. *arXiv preprint*, arXiv:2406.09559.

[22] Luckin, R., & Holmes, W. (2016). Intelligence unleashed: An argument for AI in education.

[23] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35.

[24] Mishra, V., & Mishra, R. (2012). English to Sanskrit machine translation system: A rule-based approach. *International Journal of Advanced Intelligence Paradigms, 4*(2), 168–184.

[25] Oak, N. N. (2024). Epic literature as source of Bhartiya cultural history. *Pratnakīrti, 5*(1), 1–20.

[26] Prothero, S. (2011). *God is not one: The eight rival religions that run the world, and why their differences matter*. Black Inc.

[27] Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., & Waldman-Brown, A. (2022). Tackling climate change with machine learning. *ACM Computing Surveys, 55*(2), 1–96.

[28] Roundtree, A. K. (2023). AI explainability, interpretability, fairness, and privacy: An integrative review of reviews. In *International Conference on Human–Computer Interaction*. Springer.

[29] Singh, M., Kumar, R., & Chana, I. (2020). Corpus-based machine translation system with deep neural network for Sanskrit to Hindi translation. *Procedia Computer Science, 167*, 2534–2544.

[30] Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Hachette UK.

[31] Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., Matsui, Y., Nozaki, T., Nakaura, T., & Fujima, N. (2024). Fairness of artificial intelligence in healthcare: Review and recommendations. *Japanese Journal of Radiology, 42*(1), 3–15.

[32] Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14*(4), 337–339.

[33] van Wyk, K. (2015). Evidence of Hindu religion on the theory of Chomsky's transformational grammar.

[34] Vasu, S. C. (1897). *The Ashtadhyayi of Panini*. Satyajnan Chaterji.