



AI in Tamil Literary Translation: Bridging Technology and Cultural Heritage

Dr. John Britto Parisutham, Dean, JOBA Academy, Australia

pjohnbritto@gmail.com

+61424698383

Dr. P. Rajendran, Madurai

drrajendran555@gmail.com

+919486302381

APA Citation:

Parisutham, John B. & Rajendran. P. (2025). AI in Tamil Literary Translation: Bridging Technology and Cultural Heritage, *Journal of Indian Languages and Indian literature in English*, 03(02), 258- 266; 2025

Submission Date:04.03.2025

Acceptance Date:24.03.2025

Abstract

This paper explores AI's potential and challenges in translating Tamil literature, a complex and culturally rich language, leveraging NMT and NLP advancements. While AI promises scalability and global access, it faces hurdles like linguistic complexity, data scarcity, and the risk of losing literary subtleties, as evidenced by case studies. Addressing ethical concerns like bias, the study advocates for hybrid models combining AI and human expertise. Future efforts should enhance datasets, develop context-aware algorithms, and establish ethical frameworks. Ultimately, successful Tamil literary translation with AI requires collaboration between technologists and linguists to preserve the language's legacy while ensuring culturally resonant translations.

Keywords: Tamil Translation, Neural Machine Translation, Natural Language Processing, Scarcity of Digitised Tamil Text, Synergy between AI Techs and Tamil Linguists

1. Introduction

Tamil, one of the world's oldest classical languages, boasts a literary tradition spanning over 2,000 years, from Sangam poetry to modern novels, tradition spanning millennia, encompassing profound philosophical treatises, epic poetry, and intricate lyrical works. To assess AI's capability in translating Tamil literature, AI chatbots like Gemini, DeepSeek, Grok and ChatGpt, were tested using a verse from the Thirukural¹. The outcomes of this experiment are provided below.

Kural from original Poem:

எப்பொருள் யார்யார்வாய்க் கேட்பினும் அப்பொருள்

மெய்ப்பொருள் காண்ப தறிவு (423)

Translation by a person called Puliur Kesihan:

எந்தப் பொருளைப் பற்றியும், எவரெவரிடமிருந்து கேட்பதானாலும், அப்பொருளினது மெய்யான தன்மைகளைக் காண்பதுதான் அறிவு ஆகும். புலியூர் கேசிகன் (423)

Translations by AI:

Gemini AI: Whatever you hear from whomever, discerning the truth in it is wisdom.

DeepSeek AI: Whatever is heard from anyone's lips, to perceive the true essence of that is wisdom.

Chat GPT AI: Whatever the subject may be, regardless of whom it is heard from, true wisdom lies in discerning its real nature.

Grok AI: Whatever subject is heard from whomever, the ability to discern the true nature of that subject is wisdom.

All the AI Chatbots gave, other than direct translation, a breakdown of every word or phrase of the original poem, except Chat GPT and an explanation of its own.

The intricate grammar of Tamil language, layered metaphors, and cultural depth pose unique challenges for translation. In recent years, Artificial intelligence (AI), with its burgeoning capabilities in natural language processing (NLP) and neural machine translation (NMT), offers a potential solution to bridge the linguistic and cultural divide, democratising access to Tamil literature on a global scale. However, the application of AI to Tamil literary translation remains underexplored, fraught with technical and ethical complexities. This article delves into the transformative potential of AI in translating Tamil literary works, while critically examining the inherent challenges and ethical considerations that accompany this endeavour.

1.1. Literature review

Charles-Kenechi, S. (2024)² in the research paper titled, ‘Artificial intelligence in translation studies: Benefits and challenges’ notes that translation studies has experienced significant growth in recent decades, fuelled by technological advancements and the forces of globalisation. These developments have made it easier to address challenges posed by language barriers, cultural differences, and contextual nuances.

Accepting the above theory, Aravinthan, A., & Eugene, C. (2024)³ in their article, ‘Exploring Recent NLP Advances for Tamil: Word Vectors and Hybrid Deep Learning Architectures’ note that the potential of advanced Natural Language Processing (NLP) techniques, such as deep learning, for Tamil, and evaluate the effectiveness of word embeddings (word2Vec, FastText) and hybrid models combining Convolutional Neural Networks (CNN) and Bidirectional Gated Recurrent Units (Bi-GRU) for Tamil text classification.

Zaki, M. Z. (2024)⁴ in the research paper titled, ‘Transforming worlds: The intersection of translation technology and transformers’ note that the role of transformers in translation technology, focusing on advancements in machine translation using Natural Language Processing (NLP) is important in democratising access to language and literature transnationally. Using a comparative and interpretative approach, the study examines the intersection of translation technology and transformers, drawing insights from language engineers and translation experts.

Pinhanez, et. al., (2024)⁵ explain in their paper, ‘Harnessing the Power of Artificial Intelligence to Vitalize Endangered Indigenous Languages: Technologies and Experiences’ that they have been investigating AI and NLP's potential to support endangered Indigenous languages since 2022 and by adopting a community-driven AI development approach, they address the unique ethical challenges, showcasing the effective creation of machine translators with minimal data and the development of practical writing tools with Brazilian communities. They also propose ILMs for scalable language support and envision a future where AI transforms language documentation into interactive, living resources for preserving indigenous languages through digitisation.

Sarveswaran (2024)⁶ in his research paper titled, ‘Tamil Language Computing: the Present and the Future’ explains that language computing integrates linguistics, computer science, and cognitive psychology to enable tasks like speech recognition, machine translation, and text summarization, fostering meaningful human-computer interactions and recent advances in deep learning have improved accessibility and enabled computers to learn and adapt independently, while foundational efforts like Tamil's transition from ASCII to Unicode have enhanced digital communication.

To summarise, Artificial Intelligence, especially Neural Machine Translation, is leading this transformative shift, advancing in understanding context, subtleties, and idiomatic expressions significantly boosting translation accuracy. It highlights the improved effectiveness of machine translation through the combined efforts of human expertise and AI. Prospects include multimodal translation, integrating image and voice recognition, which could enable more inclusive communication. The research also stresses the need to address linguistic diversity by developing adaptive translation systems attuned to contextual nuances (Mohamed, Y. A., et. al., 2024).⁷

1.2. Research questions

- What are the benefits of AI in Tamil literary translation?
- What are the challenges of AI in Tamil literary translation?

2. Discussions

Benefits of AI in Tamil Literary Translation

Rapid scalability

AI-powered tools offer unparalleled scalability, enabling the rapid processing and translation of vast volumes of text. This is particularly crucial for preserving endangered Tamil manuscripts and making them accessible to a wider audience. Artificial Intelligence (AI) has emerged as a central force in this transformation, enabling more efficient and accurate translation processes. AI-powered tools, such as neural machine translation (NMT) and natural language processing (NLP), have revolutionised how translations are performed, offering faster and more context-aware solutions.

This has not only improved the quality of translations but also expanded access to global communication, bridging gaps between diverse languages and cultures. As a result, translation studies have evolved from a primarily human-driven discipline to one that increasingly integrates technology, making it more dynamic and inclusive in addressing the complexities of multilingual and multicultural interactions. The results confirm that deep learning approaches, particularly when combined with joint embeddings, are highly effective for Tamil NLP tasks, bridging the gap for low-resource languages in the digital era.

Democratising access to Tamil literature globally

AI can facilitate the democratisation of Tamil literature, breaking down geographical and linguistic barriers. Through readily available translation platforms and applications, readers worldwide can engage with the beauty and wisdom of Tamil literary works. It addresses challenges in transformer-based systems, such as data efficiency, rare word handling, context sensitivity, bias, and societal impacts. The research also explores innovative tools like wearable translation devices and emotion recognition systems, emphasizing their role in overcoming language barriers and fostering global collaboration.

It is further emphasised that translation is a fundamental necessity, not merely a linguistic exercise, but a vital tool in our interconnected world, essential for cultural exchange as cultures share their stories, ideas, and values through translated works, and also for promoting global understanding by making information accessible, helping people comprehend different perspectives; translation effectively bridges linguistic divides by converting languages, enabling communication despite barriers, and also bridges cultural divides by interpreting nuances, preventing misunderstandings and fostering respect; by facilitating communication across diverse communities, translation allows dialogues and interactions between groups that would otherwise be isolated, fostering empathy by exposing people to different experiences and encouraging compassion, thus strengthening global comprehension and contributing to a more informed and interconnected world, the ‘effectively’ emphasizing the power of well-executed translation, which is not just about changing words, but conveying meaning; in essence, translation acts as a critical link in the chain of human connection, allowing us to build bridges of understanding across the globe (Zaki, M. Z., 2024).⁸

Preserving endangered texts through digitisation

AI can contribute to the standardisation and preservation of the Tamil language itself. By digitising and analysing large corpora of Tamil texts, AI can identify patterns, document variations, and contribute to the development of comprehensive linguistic resources. Near-instant translations of Thirukkural examples prove AI's rapid processing, free accessibility, and potential for language standardization. Tamil palm leaf manuscripts have been a cornerstone of South Indian intellectual and cultural traditions for centuries. These manuscripts, typically inscribed on dried palm leaves, encompass a wide array of content, including religious texts like the ‘Tirukkural’ and ‘Agamas’, medical treatises in Siddha medicine, and historical chronicles.

Preserving this knowledge is vital not only for understanding ancient Tamil civilization but also for broader studies of Indian and South Asian cultural history. Previous research has emphasized the unique role of these manuscripts in transmitting Tamil culture, particularly through oral traditions, where they served as the primary medium for codifying and preserving texts. Despite their significance, Tamil palm leaf manuscripts face considerable preservation challenges, such as degradation from climatic conditions, fungal growth, and physical wear due to handling. These issues have led to the loss or inaccessibility of countless manuscripts over time (Pradeep, et. al., 2024).⁹

Challenges in Tamil Literary Translation

Scarcity of Digitized Texts

A significant obstacle to the development of effective AI translation tools for Tamil is the scarcity of digitised literary texts. However, it is pointed out that the scarcity of digitized texts in Tamil literature and the need for computational resources, linguistic annotation, and practical applications, calling for research collaboration, digitization of historical texts, and increased digital usage to advance language processing and global communication (Sarveswaran, 2024)¹⁰.

Moreover, the lack of comprehensive and well-annotated datasets limits the ability of AI models to learn the nuances of the language and its literary traditions. Furthermore, the available datasets may be biased towards certain genres or periods, leading to translations that reflect these biases. Addressing this challenge requires a concerted effort to digitise and curate a diverse range of Tamil literary texts, ensuring that AI models are trained on representative and unbiased data.

Linguistic Complexity - Risk of oversimplifying metaphors and aesthetic subtleties

Ramesh, et. al., (2021)¹¹ in their research article titled, 'Comparing statistical and neural machine translation performance on hindi-to-tamil and english-to-tamil' claim that translation to or from under-resourced languages has been historically seen as a challenging task and that despite producing state-of-the-art results in many translation tasks, NMT still poses many problems such as performing poorly for many low-resource language pairs mainly because of its learning task's data-demanding nature. The authors investigated the performance of PB-SMT and NMT on two rarely tested under-resourced language pairs, English-To-Tamil and Hindi-To-Tamil, taking a specialised data domain into consideration and found out that Tamil's agglutinative structure allows words to combine morphemes (word units) to convey complex meanings. For example, the word "செல்லப்பூங்காவில்" (sellapūṅkāvil) translates to "in the flower garden of love," blending emotion, location, and metaphor. AI models like Google Translate often struggle with such constructs, producing literal or disjointed outputs.

Furthermore, AI translation models, driven by algorithms and statistical patterns, may struggle to capture the subtle nuances of literary language, such as metaphors, symbolism, and emotional depth. This can lead to translations that are accurate in terms of literal meaning but fail to convey the aesthetic beauty and emotional impact of the original work. The challenge lies in developing AI models that can understand and appreciate the artistic and cultural context of Tamil literature.

Cultural and Aesthetic Nuances - Emotional depth intrinsic to its literature.

Tamil literature relies heavily on ‘rasa’ (emotional essence) and ‘dhvani’ (suggestive meaning), concepts challenging to encode algorithmically. For instance, Subramania Bharati’s nationalist poetry intertwines cultural pride with metaphor, risking oversimplification by AI. Translation acts as a conduit for cultural knowledge. In the modern context, its role has evolved, becoming more prominent and direct. Operating within a polyphonic landscape marked by subversion, translation now navigates cultural crises prevalent globally, particularly within the Tamil region (Indra, C. T., & Rajagopalan, R. (2017)).¹²

Kumar (2021)¹³ confirms in his dissertation titled, ‘Multilingual NMT for Indian Languages’ that Neural Machine Translation relies on first mapping each word into the vector space, and traditionally we have a word vector corresponding to each word in a fixed vocabulary. This concludes that vocabulary plays an important role in NMT. The author also articulates that Tamil’s agglutinative grammar, where words are formed by adding multiple suffixes to a root, presents a unique challenge for AI translation. The intricate interplay of suffixes creates a vast array of grammatical possibilities, requiring sophisticated algorithms to accurately parse and translate sentences. Furthermore, Tamil literature is replete with idiomatic expressions, metaphors, and cultural nuances that are deeply embedded in the language’s historical and social context. AI systems must be trained to understand and interpret these complexities to produce accurate and culturally sensitive translations.

3. Conclusions

In conclusion, the integration of AI in Tamil literary translation offers significant benefits, including rapid scalability, democratizing access to Tamil literature globally, and preserving endangered texts through digitization. These advancements hold the potential to make Tamil’s rich literary heritage more accessible and enduring. However, challenges remain, such as the scarcity of digitized texts, the linguistic complexity of Tamil, and the risk of oversimplifying metaphors, aesthetic subtleties, and emotional depth intrinsic to its literature. Addressing these challenges requires a balanced approach that combines technological innovation with cultural sensitivity, ensuring that the essence of Tamil literature is preserved while leveraging AI’s transformative potential.

The convergence of artificial intelligence and Tamil literary translation presents a landscape of extraordinary potential, yet one that necessitates meticulous and thoughtful guidance. While AI offers the transformative capacity to democratize access to Tamil’s rich literary heritage on a global scale, it is crucial to recognize that the essence of literature extends beyond mere linguistic

conversion. True translation, particularly of culturally profound works, requires a deep understanding of the subtle nuances, emotional depth, and historical context embedded within the original text.

Therefore, the path forward lies in a collaborative synergy between AI technologists and Tamil linguists. By forging partnerships that leverage the computational power of AI alongside the nuanced expertise of human scholars, we can ensure that translations not only achieve linguistic accuracy but also faithfully capture the literary soul of Tamil works. This deliberate and balanced approach will enable us to harness the benefits of AI to bridge cultural divides and make Tamil literature accessible to a wider audience, all while safeguarding the integrity and enduring legacy of this ancient and revered language. Ultimately, the successful fusion of AI and Tamil literature will depend on our collective commitment to honoring the past as we embrace the possibilities of the future, ensuring that translations resonate with both intellectual understanding and emotional connection.

Initiatives like the Tamil Nadu Digital Library which aim to digitize 50,000 texts by 2030 (தமிழிணையம். மின்னுகலம். தமிழ் இணையக் கல்விக்கழத்தின் ஒரு பிரிவு, 2021)¹⁴ has to be supported by technocrats and linguists. The Tamil Digital Library (<https://www.tamildigitallibrary.in/>) is a valuable online resource dedicated to preserving and providing access to a vast collection of digitized Tamil books and manuscripts. It serves as a crucial repository for researchers, scholars, and anyone interested in Tamil literature, history, and culture. It offers a wide range of digitized materials, including ancient texts, literary works, historical documents, and other valuable resources. The library aims to make these resources freely accessible to a global audience, promoting the preservation and dissemination of Tamil knowledge. It plays a vital role in preserving fragile and aging texts by digitizing them, ensuring their long-term availability. The website contributes significantly to the preservation and promotion of Tamil cultural heritage. It is a vital resource for scholars and researchers studying Tamil language and literature.

4. Reference

1. அறிவுடைமை. குறள். <https://www.thirukkural.net/ta/kural/adhigaram-043.html>
2. Charles-Kenechi, S. (2024). Artificial intelligence in translation studies: Benefits and challenges. *Cascades, Journal of the Department of French & International Studies*, 2(1), 5-15.

3. Aravinthan, A., & Eugene, C. (2024). Exploring Recent NLP Advances for Tamil: Word Vectors and Hybrid Deep Learning Architectures. *The International Journal on Advances in ICT for Emerging Regions*, 17(2).
4. Zaki, M. Z. (2024). Transforming worlds: the intersection of translation technology and transformers. *Computer Science&engineering: An International Journal (CSEIJ)*, Vol 14, No 3.
5. Pinhanez, C., Cavalin, P., Storto, L., Finbow, T., Cobbinah, A., Nogima, J., ... & Gonçalves, I. (2024). Harnessing the Power of Artificial Intelligence to Vitalize Endangered Indigenous Languages: Technologies and Experiences. *arXiv preprint arXiv:2407.12620*
6. Sarveswaran, K. (2024). Tamil Language Computing: The Present and the Future. *arXiv preprint arXiv:2407.08618*.
7. Mohamed, Y. A., Khanan, A., Bashir, M., Mohamed, A. H. H., Adiel, M. A., & Elsadig, M. A. (2024). The impact of artificial intelligence on language translation: a review. *Ieee Access*, 12, 25553-25579.
8. Zaki, M. Z. (2024). Transforming worlds: the intersection of translation technology and transformers. *Computer Science&engineering: An International Journal (CSEIJ)*, Vol 14, No 3.
9. Pradeep, N., Subramanian, D., & Ganapathy, M. K. (2024). Digitizing India's Ancient Texts: AI for Tamil Palm Leaf Manuscript Preservation and Accessibility.
10. Sarveswaran, K. (2024). Tamil Language Computing: The Present and the Future. *arXiv preprint arXiv:2407.08618*.
11. Ramesh, A., Parthasarathy, V. B., Haque, R., & Way, A. (2021). Comparing statistical and neural machine translation performance on hindi-to-tamil and english-to-tamil. *Digital*, 1(2), 86-102.
12. Indra, C. T., & Rajagopalan, R. (2017). Mapping the nuances of language. In *Language, Culture and Power* (pp. 187-205). Routledge India.
13. Kumar, S. (2021). Multilingual NMT for Indian Languages (Doctoral dissertation, International Institute of Information Technology Hyderabad).
14. தமிழிணையம். மின்னுகலம். (தமிழ் இணையக் கல்விக்கழத்தின் ஒரு பிரிவு). <https://www.tamildigitallibrary.in/>