

TAMIL AND ARTIFICIAL INTELLIGENCE

Dr. Chandrika J

Assistant Professor, Department of Tamil, Shri Krishnaswamy College for Women, Anna Nagar, Chennai – 600040

ABSTRACT

Artificial Intelligence (AI) has emerged as a transformative technology with applications across diverse fields, including language, education, translation, and cultural preservation. This paper explores the evolving relationship between the Tamil language and Artificial Intelligence, tracing Tamil's transition from its classical heritage to its growing presence in the digital and AI-driven era. It examines how AI technologies can support the learning, preservation, enrichment, and global dissemination of Tamil.

The study highlights the importance of comprehensive and reliable Tamil language data for developing accurate Large Language Models (LLMs), particularly because Tamil is an agglutinative language with distinctive linguistic structures and cultural nuances. AI-based applications offer significant opportunities in grammar and spell correction, prosody analysis, translation, speech-to-text and text-to-speech conversion, and the digitization and preservation of inscriptions, palm-leaf manuscripts, temple records, folk songs, and other cultural resources. AI-powered learning tools can also support Tamil education among younger generations and Tamil communities living across the world.

At the same time, the paper draws attention to challenges associated with inadequate language data, translation accuracy, incorrect digital inputs, and excessive dependence on AI-generated creative content, which may affect human creativity and the natural evolution of literary expression. The paper therefore emphasizes the need for authentic linguistic resources, responsible data curation, and collaboration among Tamil scholars, linguists, educators, technologists, and institutions.

The study concludes by identifying future possibilities for AI applications in Tamil literature, grammar, inscriptions, folklore, tribal languages, dictionaries, manuscripts, and cultural documentation. Harnessing AI while preserving human creativity and linguistic authenticity can contribute significantly to strengthening Tamil in the contemporary digital world and ensuring its transmission to future generations.

Keywords: Tamil Language, Artificial Intelligence, Natural Language Processing, Digital Preservation, Language Technology

I. INTRODUCTION

Intelligence is an inherent mental attribute encompassing cognitive abilities such as planning, problem-solving, abstract thinking, comprehension of concepts and language, and learning. Psychologists distinguish intelligence from other human traits such as creativity, personality, intellectual acumen, and individual characteristics. In contrast to natural intelligence, artificial intelligence is currently on the rise; it represents the pinnacle of advancements in computer science. The applications of artificial intelligence are steadily expanding across diverse fields, including medicine, agriculture, engineering, linguistics, education, intelligence services, and the military. In this context, this article chronicles the evolutionary trajectory of the Tamil language—from its origins as a classical language to its growth in the computer age and its current dimension within the realm of artificial intelligence.

Learning Ability and Data Input

Continuous learning paves the way for intelligence; as the poet Avvaiyar aptly stated, "Intelligence is commensurate with the extent of one's learning." For humans, both formal education and experiential knowledge serve as key factors in developing intelligence. Conversely, artificial intelligence is instilled in machines through the input of data. Thiruvalluvar emphasizes the significance of learning in chapters dedicated to Education, Listening (to the learned), Wisdom, and the Ignorance of the Unlearned. To acquire true wisdom, one must learn thoroughly and without error. Furthermore, one must steadfastly adhere to the principles prescribed in the texts one has mastered. While a human may or may not always follow the path they have learned, there is no doubt that a machine will operate strictly according to the principles dictated by the data fed into it. Valluvar emphasizes the indispensability of the intellect that deeply analyzes what has been learned, stating: "The beauty of one lacking subtle, profound insight is akin to that of a finely crafted clay doll." This human capacity for nuanced analysis—encompassing data compilation, analysis, classification, reflective action, imaginative expression based on data, and problem-solving—is now being imparted to machines through Artificial Intelligence (AI). There are various avenues for creating Tamil works using AI tools, software, and web platforms. Tech-savvy Tamil enthusiasts and creative minds are doing their utmost to expand Tamil's presence in the technological landscape. However, regarding accuracy, the Tamil language requires more data compared to English; the accuracy of an LLM (Large Language Model) is directly proportional to the volume of data it possesses. Since Tamil is an agglutinative language, its linguistic nuances differ from those of other languages; consequently, a generic LLM is not suitable for Tamil. Accuracy improves only with a substantial volume of Tamil data; without sufficient data, high accuracy cannot be expected. Currently, Tamil literary and grammatical works are being digitized chronologically, ensuring these texts will be available to future generations as historical records. If incorrect data is entered, the truth may be obscured, resulting in a system that projects a false sense of reliability. Therefore, those inputting data must ensure that the information provided to the machine is accurate and trustworthy.

Imaginative Power and Linguistic Richness

New technologies will continue to emerge over time. It is a fact that embracing these technologies—rather than fearing or resenting them—opens the door to new opportunities. Humans, who once roamed the forests, began by recounting their hunting exploits as a pastime. Storytelling—the wellspring of imagination—eventually evolved into literary works that enriched the Tamil language. In this technological age, it is doubtful whether a machine can possess imaginative capabilities equal to those of a human; however, Artificial Intelligence (AI) is adept at synthesizing provided data to generate creative works. While a search engine like Google merely retrieves information relevant to a user's query, AI-driven tools like ChatGPT and Gemini function more like humans. Just as a person—when asked a question—selects the most appropriate response based on their knowledge rather than simply reciting everything they know, AI applications provide tailored answers, thereby saving the user time. Yet, there is a potential downside. In the past, people could mentally calculate even large sums involving income and expenses; today, however, everyone—from children to adults—relies on calculators for even the simplest arithmetic. This raises a troubling question: are we destined to lose our innate creativity and become too lazy to write songs, poems, stories, or plays born of our own imagination, choosing instead to rely on machines? Musicians and lyricists are already inputting melodies into AI platforms and using the lyrics generated in response. The reality is that we are paving the way for a future where Tamil is deprived of the natural, spontaneous flow of words that once defined its poets.

Learning Basic Grammar

Today's generation takes pride in speaking and writing in English but shows little interest in learning Tamil properly. Consequently, the essays, sentences, poems, and paragraphs they write often contain errors. Artificial Intelligence (AI) applications prove helpful in identifying and correcting these mistakes. Tools such as spell checkers, word correctors, sandhi (euphonic junction) correctors, Tamil Lexicon, Vani, and Siruthiruthi clearly highlight and rectify various Tamil grammatical errors—including the incorrect placement of *otru* (consonant doubling), confusion regarding *mayangoli* (confusable sounds) and singular-plural forms, spelling mistakes, and issues with phonetic markers. Furthermore, AI applications like Sulagu, Avalokitham, Sangacholai, Sangamedia, and Paadal assist in identifying elements of prosody (such as *asai*, *todai*, and poetic forms) and in converting poetic structures.

The Growth of Tamil Abroad

Tamils living overseas are playing a significant role in fostering the growth of the Tamil language in the digital realm. They are keen on teaching Tamil to children in their communities and are increasingly utilizing AI-powered tutors for this purpose. Through this process, learners acquire not only the language itself—including pronunciation and writing systems—but also insights into traditional knowledge and cultural values. As the Tamil language adapts to keep pace with cutting-edge technological advancements, it has become much easier to use widely.

AI platforms provide immense assistance to individuals living thousands of miles away, enabling them to easily read and study grammatical texts, literary manuscripts, or ancient inscriptions located in remote corners of Tamil Nadu—all from the comfort of their own homes.

Applications in Translation

It is essential for linguists to possess multilingual proficiency. It was precisely because of his multilingual expertise that Bharathiyar was able to read Japanese poetry and English and French literature, and subsequently enrich Tamil with similar works. Today, even without multilingual skills, we can translate and read English articles in Tamil with the help of various applications such as Google Translate, Google Lens, Sanskrit Dictionary, and Cloud Vision. Google Translate alone offers the capability to translate across 93 languages. Translation applications play a significant role in enabling us to explore foreign literature and understand other cultures, while also allowing others to appreciate the richness of our own culture. Artificial Intelligence (AI) applications are making the traditionally challenging field of translation easier and are fostering an increase in job opportunities.

Audio-Visual-Text Converters

Artificial intelligence applications contribute across various domains, enabling the conversion of speech to text, text to speech, images to text, and text to images. This capability allows for the transcription and recording of speech, songs, and music—not only from the general population but also from rare tribal communities. Folk songs such as lullabies and dirges, sung by unlettered village women, can be recorded, transcribed into text, and easily utilized. At a time when many modern women are familiar only with film songs, preserving these folk songs in both audio and text formats enables them to sing such traditional lullabies to their own children.

Document Preservation

Artificial Intelligence (AI) applications—such as *Aavanam*, *Keezhadi*, and *Varnam*—play a vital role in preserving ancient Tamil inscriptions, palm-leaf manuscripts, temple chronicles (*Koyil Ozhugu*), copper plates, and coins. Such digital archiving is immensely helpful in safeguarding our heritage for future generations and facilitating research.

The Role of AI in Teaching

Today's students are born and raised in a digital world. Having grown up interacting with mobile phones, they may not always find traditional blackboard-based instruction engaging. Consequently, teachers are showcasing their pedagogical skills through AI-driven tools and platforms—utilizing PowerPoint presentations, videos, audio clips, and services like Google Meet, Zoom, and Google Classroom (for collecting Tamil assignments). As AI-powered robotic teachers emerge as potential competitors, human teachers must master and adopt comparable technologies to remain effective. In an era where machine learning rivals human

learning, educators must leverage technology to demonstrate the unique value and superiority of human teaching. To achieve this, professors need to equip themselves with proficiency in both language and technology.

II. THE TIDAL WAVE OF ARTIFICIAL INTELLIGENCE IN THE TAMIL DIGITAL LANDSCAPE

Artificial Intelligence, the pinnacle of modern technology, has surged like a massive tidal wave across the Tamil digital landscape. Although the Tamil language initially faced challenges due to a multitude of fonts and software systems—unlike the more standardized English environment—it has since transformed these very elements into its own unique identity and is expanding rapidly across diverse sectors. Indeed, the proliferation of the internet and digital devices has led to a colossal accumulation of Tamil language data essential for training computer systems. Furthermore, advancements in high-speed computing chips—such as Graphics Processing Units (GPUs)—have made it possible to train complex Artificial Intelligence models. With new techniques like deep learning and advanced algorithms exponentially enhancing the self-learning capabilities of machines, the Tamil language is also integrating into these systems and adapting itself accordingly.

III. AREAS FOR ARTIFICIAL INTELLIGENCE (AI) APPLICATION IN TAMIL

- Applications should be designed to instantly identify the poetic genre of a verse upon its input, while also providing details on literary devices (*anis*), metaphors (*uvamaigal*), and explanations of the poem's meaning.
- Integrating programming logic for scripts found in inscriptions and copper plates—such as Tamil, Grantha, and Vatteluttu—into applications would greatly assist future researchers.
- To ensure 100% accuracy in translation tools (moving beyond the current 80%), the input data must encompass the language's morphemes, allomorphs, and phonemes, as well as the cultural context of the people who speak the language.
- Given the growing need for technical terminology (*kalaichollakkam*), efforts must be made to create new terms and bring them into widespread use.
- Resources such as encyclopedias, dictionaries, archives of old periodicals, and unpublished manuscripts should be digitized and uploaded to digital libraries.
- Artificial Intelligence should be utilized to document the 30 undocumented tribal languages.
- Future applications should be developed to cover a wide range of areas, including: folklore, minor literary genres (*sitirilakkiyam*), temple chronicles (*thalapuramams*), medieval literature, plays, films, biographies and quotes of Tamil scholars, notable lines from Tamil literature, general knowledge guides, spiritual texts, Tamil dictionaries,

astrological notes, recipes (including traditional cuisine), devotional literature, information on various temples, moral tales, arts, and cultural records. Developing such applications ensures we provide invaluable documentation for future generations. It is somewhat regrettable that developing a mobile application requires a minimum of three months and costs between ₹25,000 and ₹2 lakh, as it is not feasible for the government to provide financial assistance for every single app. However, business enterprises that generate substantial revenue through the use of such applications should step forward to provide upfront funding and encourage their development.

IV. CONCLUSION

Tamil's journey, powered by Artificial Intelligence, will continue across various domains: preserving the richness of its literature and grammar, passing this heritage on to a modern-minded future generation, addressing the need for linguistic input in machine learning, leveraging the role of language in deep learning, and expanding its global reach. By adapting to the times, Tamil will triumph.