

LINGOBRIDGE AI: A MULTILINGUAL INTELLIGENT SYSTEM FOR TRANSLATION AND SUMMARIZATION OF INDIAN LANGUAGES

¹Dr.Pramod Kumar Naik, ²Divyank Jain Singhvi, ²Ekta Ghosh, ²Saswat.S Rout, ²Savhar Verma

¹Assistant Professor, ²UG Students, Dept. of Computer Science and Engineering, School of Engineering, Dayananda Sagar University Devarakaggalahali, Karnataka

Corresponding author: ektaghosh2002@gmail.com

Received 28 April 2025 Received in Revised form 01 May 2025 Accepted 02 May 2025
Available Online 06 May 2025

ABSTRACT

This project aims to develop an efficient multilingual translation and summarization tool tailored for Indian languages, addressing the digital communication challenges faced by speakers of low-resource languages. Existing tools often lack contextual understanding, resulting in inaccurate translations. LingoBridge AI introduces a transformer-based LLM that bridges this gap by accurately translating and summarizing documents, speech, and text. It employs language clustering and a virtual intermediary called the "Lamp Language" to reduce errors in inter-cluster translation. The model ensures real-time performance and accessibility through web and mobile platforms. Users receive results in both text and audio formats in their preferred language—Hindi or English. This system promotes digital inclusivity and empowers citizens by improving access to essential information. Ultimately, the initiative supports India's linguistic diversity while fostering an equitable and digitally connected society.

Keywords: Multilingual LLM, Translation, Indian Languages, Lamp Language, Real-time Summarization, Digital Inclusivity.

I. INTRODUCTION

The rise of multilingual large language models (LLMs) has transformed natural language processing (NLP) by giving machines the ability to understand and create text in many different languages. These advanced models can handle important language tasks, like translation, summarization across multiple languages. This technology opens up big possibilities for communication between languages and makes digital information accessible to more people.

In a country as linguistically diverse as India, with hundreds of languages, multilingual LLMs have great potential to break down language barriers. By supporting communication in regional languages, these models can help more people access technology and information, promoting digital inclusivity. This project focuses on creating a multilingual LLM system to support several regional Indian languages, including Hindi, English, Tamil, Bengali, and others, so that users can interact with technology in their preferred language across different applications.

Our project focuses on developing a multilingual transformer-based model tailored specifically for Indian languages[1]. It is designed to support Hindi, English, and a variety of regional languages to foster inclusive communication across the country's linguistically diverse population. By bridging linguistic gaps between communities, the model enhances digital participation and ensures that no

one is left behind in the digital ecosystem. One of the standout features of this system is its real-time translation and summarization capabilities, which provide users with instant access to information in their preferred language. This functionality greatly improves accessibility, particularly for individuals residing in remote and underserved areas. Furthermore, the system contributes to better governance by simplifying the delivery of government content and services. Businesses can also benefit from this technology through efficient multilingual communication with clients and stakeholders[2]. By enabling seamless interaction in native languages, the platform empowers users at every level. Ultimately, the project supports the creation of a more connected, informed, and equitable digital India.

II. LITERATURE SURVEY

The Jiten Parmar, Naveen Saini, and Dhananjay Dey proposed the ILrLSUMM model, which employs a differential evolutionary algorithm for extractive summarization of Indian regional languages, including Hindi and Gujarati. Their approach achieved notable performance improvements with a focus on optimization. This research underscores the significance of leveraging evolutionary algorithms for low-resource Indian languages and optimizing summarization processes. [3]