International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 04 | April – 2026 DOI: 10.55041/ISJEM06511
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Enhancing Accessibility of Government Notices Through LLM-Based Multilingual Translation

Dr.M.U.Karande[1], Mr. Rohit. V. Talele[2], Miss.Shruti.S.Ujjainkar[3], Miss.Shiva.D.Hinge[4], Mr.Saurabh.R.Patil[5],

[1] Professor, Department of Computer Science Engg,
[2] Students, Department of Computer Science Engg, Dr. V. B. Kolte College of Engineering, Malkapur, India
[3] Students, Department of Computer Science Engg, Dr. V. B. Kolte College of Engineering, Malkapur, India
[4] Students, Department of Computer Science Engg, Dr. V. B. Kolte College of Engineering, Malkapur, India
[5]Students, Department of Computer Science Engg, Dr. V. B. Kolte College of Engineering, Malkapur, India

---

Abstract –

In the field of computational linguistics, addressing machine translation (MT) challenges for low-resource languages remains crucial, as these languages often lack extensive data compared to high resource languages. General large language models (LLMs), such as GPT-4 and Llama, primarily trained on monolingual corpora, face significant challenges in translating low-resource languages, often resulting in subpar translation quality. This study introduces Language-Specific Fine-Tuning with Low-rank adaptation (LSFTL), a method that enhances translation for low-resource languages by optimizing the multi-head attention and feed-forward networks of Transformer layers through low-rank matrix adaptation. LSFTL preserves the majority of the model parameters while selectively fine-tuning key components, thereby maintaining stability and enhancing translation quality. Experiments on non-English centered low-resource Asian languages demonstrated that LSFTL improved COMET scores by 1-3 points compared to specialized multilingual machine translation models. Additionally, LSFTL's parameter-efficient approach allows smaller models to achieve performance comparable to their larger counterparts, highlighting its significance in making machine translation systems more accessible and effective for low-resource languages.

Key Words**:** *Machine translation, low-resource languages, large language models, parameter-efficient fine-tuning, LoRA.*

1. INTRODUCTION

The Indian government places significant reliance on the distribution of notices as a primary means of conveying vital and pertinent information to the masses. This method has proven to be a robust and dependable way to establish a direct connection with a large and diverse population, ensuring that every citizen can remain informed about the government's ongoing activities and developments.

However, a notable challenge arises in this process–the government often employs the language that is predominantly spoken within a particular state. While this choice effectively reaches the majority of the population, it leaves a significant portion of the minority population struggling to grasp the content of these notices. The linguistic diversity of India is one of its defining characteristics, with numerous regional languages spoken across the nation. Consequently, a substantial portion of the populace may not be fluent or familiar with the language chosen for official notices in their respective regions. This language barrier can be a substantial impediment to the government's objective of ensuring that all citizens, regardless of their linguistic background, can access critical information. To address this challenge, a systematic solution is imperative. The government should develop and implement a robust system capable of seamlessly translating and converting local languages into various Indian languages. By doing so, these notices can be made accessible to the minority population, transcending linguistic boundaries and enabling everyone to comprehend and engage with the government's communications effectively. Such a system would not only promote inclusivity and equal access to information but also uphold the principles of democracy by ensuring that every citizen can actively participate in the governance process. It would reflect the government's commitment to fostering unity in

diversity and strengthen its relationship with all segments of society, reinforcing the bonds of a vibrant and pluralistic nation. In essence, the development of a comprehensive language conversion system is a vital step towards creating a more connected, informed, and equitable society in India.

2. SYSTEM ARCHITECTURE

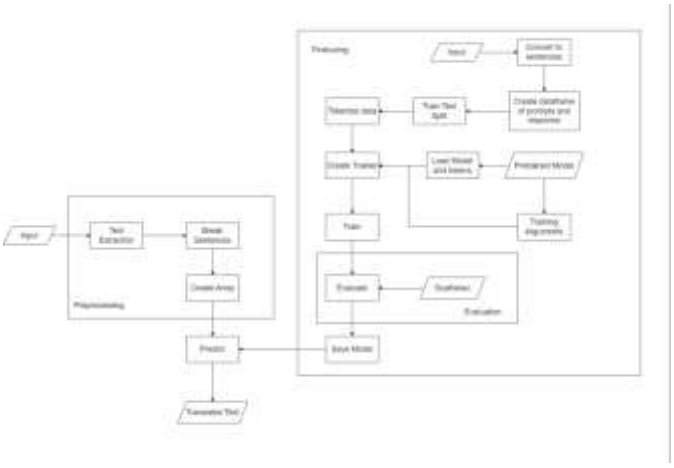


Fig -1: System Architecture

Input The input to the system is the source text to be translated. This text can be in any language, but it is typically preprocessed to clean it up and normalize it.
Text Extraction The text extraction module extracts the text to be translated from the input. This may involve breakingup the input into sentences, identifying the source language, and detecting any special characters or formatting.
Preprocessing The preprocessing module performs a variety of tasks on the extracted text, such as tokenizing the text, removing stop words, and normalizing the case.
Pretrained Model The pretrained model is a neural network that has been trained on a large corpus of parallel text. The model is used to generate translations of the preprocessed text.
Finetuning The finetuning module is used to fine-tune the pretrained model on a specific domain or task. This can be done by providing the model with additional training data or by adjusting the model's hyperparameters.
Evaluation The evaluation module is used to evaluate the performance of the MT system on a held- out test set. This helps to ensure that the system is producing high-quality translations.
Output The output of the MT system is the translated text in the target language. This text can be used in a variety of applications, such as translation websites, machine translation APIs, and multilingual software.

3. IMPLEMENTATION AND RESULT

Users upload images or PDFs to the website, where the text is extracted and processed. Extracted text is structured into sentences, then segmented into word arrays. These arrays are fed into a language model (LLM) for translation, with the model leveraging training data. Predicted translations are reconstituted into sentences and displayed to users. This process enhances accessibility and comprehension of Marathi government documents, fostering citizen engagement and alleviating complexities. It combines Optical Character Recognition (OCR)for text extraction, LLM for translation, and user-friendly web interfaces for effective information delivery.

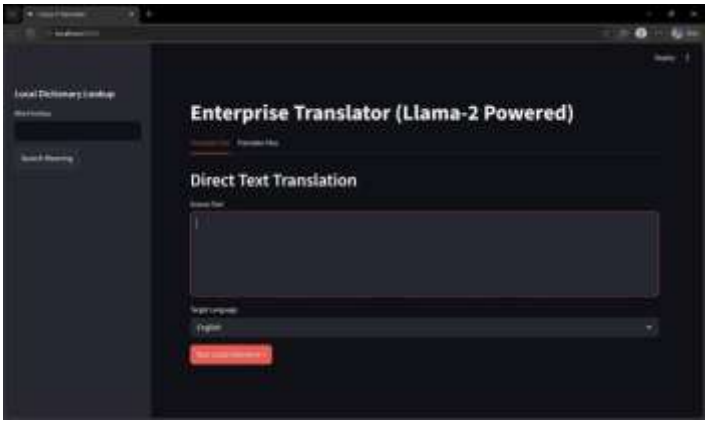


Fig -2: **Home Page**

3.1 Direct Text Translation Module:

The **Direct Text Translation Module** is a core component of the system designed to translate user-provided text from a source language into a selected target language using a locally deployed language model (such as LLaMA-2). This module enables real-time translation without relying on external APIs, ensuring data privacy and faster processing.

The module provides a user-friendly interface where users can input text in any supported language and choose their desired output language. Upon clicking the "Run Local Inference" button, the system processes the input text using the local inference engine and generates the translated output.

Key Functionalities:

- Accepts multilingual input text from the user.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 04 | April – 2026 DOI: 10.55041/ISJEM06511
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

- Allows selection of the target language (e.g., English).
- Performs translation using a locally hosted AI model.
- Displays translated output clearly in the result section.
- Provides status updates such as "Inference Complete."

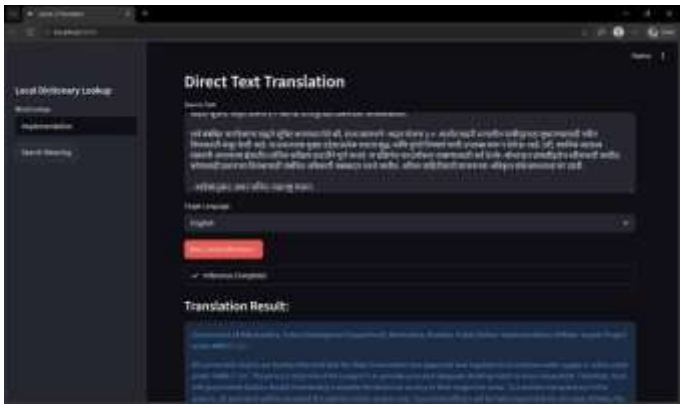


Fig -3: **Direct Text Translation**

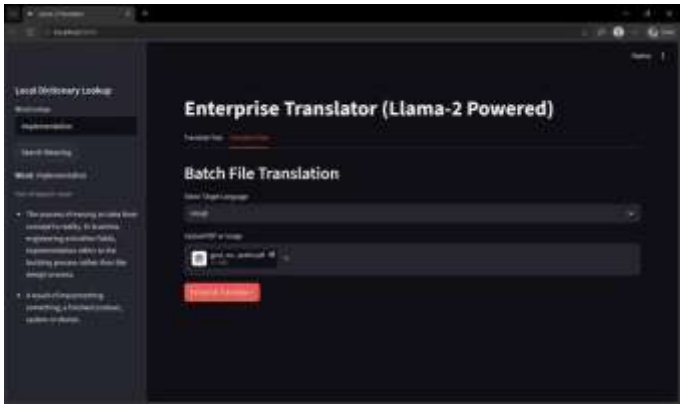
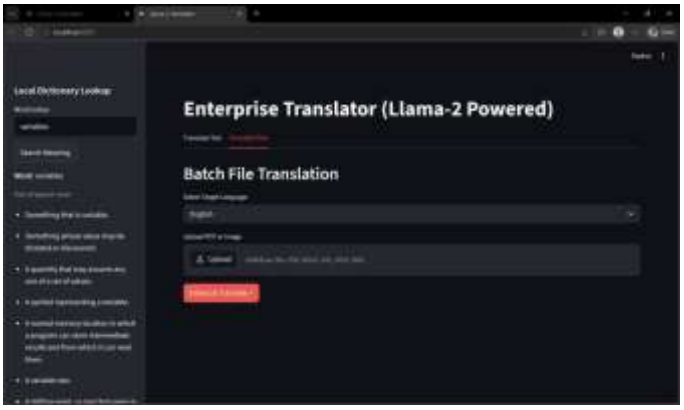


Fig -4: **Translate File**

6. LLM Based Translation System

The proposed system demonstrates a high level of accuracy and efficiency in the translation of government notices across multiple languages. Experimental evaluation indicates that the system effectively preserves the semantic meaning and contextual relevance of the original text, thereby ensuring reliable communication of critical information. The implementation of a locally deployed Large Language Model (LLM) significantly enhances data privacy and security, as sensitive information is processed without reliance on external servers. Additionally, the system reduces dependency on continuous internet connectivity, making it suitable for deployment in resource-constrained or rural environments.

The results further highlight the system's capability to bridge linguistic barriers by enabling multilingual access to government communications, thereby promoting inclusivity and wider dissemination of information. Performance analysis shows consistent response times and satisfactory. translation quality across commonly used regional languages.



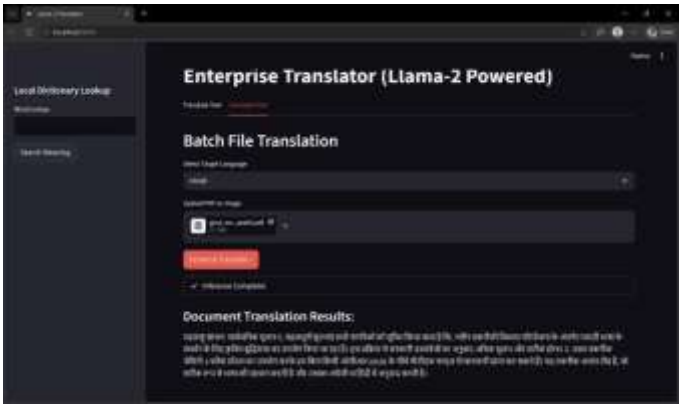
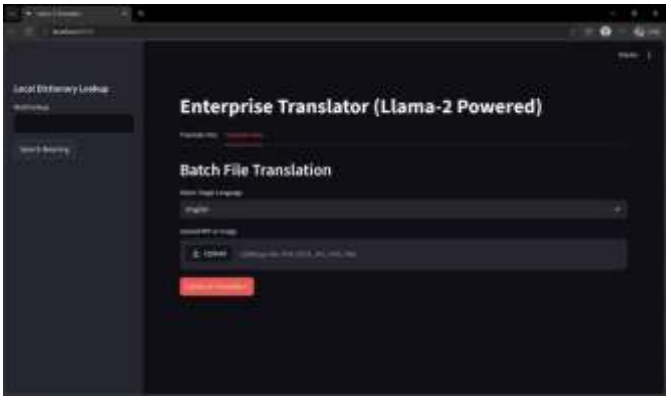


Fig -5: Document Translation Result in Hindi



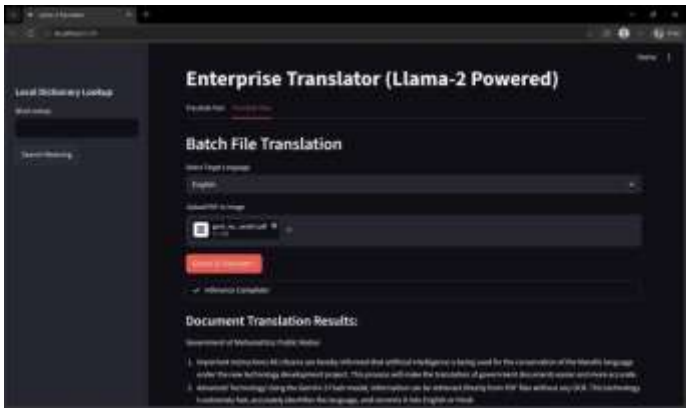


Fig -6: **Translate File Result in English**

4. CONCLUSIONS

In conclusion, The project *"Translation of Government Notices in Indian Languages using LLM"* provides an effective solution to overcome the linguistic barriers that citizens often face in accessing government information. By combining **OCR-based text extraction**, **fine-tuned Large Language Models**, and a **user-friendly web interface**, the system ensures that official notices are translated into clear, understandable, and inclusive formats. This approach not only enhances **citizen–government communication** but also empowers minority communities who may not be proficient in the regional language, thereby fostering **transparency, inclusivity, and democratic participation**. While the current implementation focuses on Marathi, the framework can be scaled to support multiple Indian languages, making it a powerful tool for nationwide deployment. In essence, the project demonstrates the practical potential of **AI-driven language technologies** in solving real- world socio-technical challenges, ultimately contributing to a more connected, informed, and equitable society.

5. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who supported and guided **Dr. M. U. Karande** mam us throughout the development of this project on the **Enhancing Accessibility of Government Notices through LLM-based Multilingual Translation**. This work would not have been possible without their encouragement, guidance, and valuable contributions. First and foremost, I would like to thank my project guide for their constant support, patience, and insightful suggestions. Their guidance helped me understand the project requirements clearly and motivated me to complete the work successfully. I am truly grateful for their time and effort in helping me at every stage of this project.

I would also like to thank all the faculty members of my department for providing me with the knowledge and resources required to carry out this project. Their teachings have played a significant role in shaping my understanding and skills.

I extend my heartfelt thanks to my friends and classmates who supported me with ideas, discussions, and encouragement whenever I faced difficulties. Their cooperation made this journey much easier and enjoyable.

Finally, I would like to express my deepest gratitude to my family for their unconditional support, understanding, and motivation. Their encouragement gave me the strength and confidence to complete this project successfully.

This project has been a great learning experience, and I am thankful to **Padm. Dr. V. B. Kolte College Of Engineering Malkapur** and everyone who directly or indirectly contributed to its completion.

6. REFERENCES

[1] Touvron, Hugo, et al. "Llama 2: Open Foundation and Fine-Tuned Chat Models." arXiv.Org, 19 July 2023, arxiv.org/abs/2307.09288.

[2] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi and S. Jain,"Machine translation using deep learning: An overview," 2017 International Conference on Computer, Communications and Electronics (Comptelix), Jaipur, India, 2017, pp. 162-167, doi: 10.1109/COMPTELIX.2017.8003957.

[3] P. Salunkhe, A. D. Kadam, S. Joshi, S. Patil, D. Thakore and S. Jadhav, "Hybrid machine translation for English to Marathi: A research evaluation in Machine Translation: (Hybrid translator)," 2016 International Confer ence on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 2016, pp. 924-931, doi: 10.1109/ICEEOT.2016.7754822.

[4] Daron Acemoglu and Pascual Restrepo. Artificial intelligence, automa tion, and work. In The economics of artificial intelligence: An agenda, pages 197–236. University of Chicago Press, 2018.

[5] Joshua Ainslie, James Lee-Thorp, Michiel de Jong, Yury Zemlyanskiy, Federico Lebr´ on, and Sumit Sanghai. Gqa: Training generalized multi query transformer models from multi-head checkpoints, 2023.

[6] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessan dro Cappelli, Ruxandra Cojocaru, Merouane Debbah, Etienne Goffinet, Daniel Heslow, Julien Launay, Quentin Malartic, Badreddine Noune, Bap tiste Pannier, and Guilherme Penedo. Falcon-40B: an open large language model with state-of-the-art performance. 2023.

[7] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. Program synthesis with large language models, 2021.

[8] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforce ment learning from human feedback. arXiv preprint arXiv:2204.05862, 2022a.

[9] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmar garet Shmitchell. On the dangers of stochastic parrots: Can language mod els be too big? In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pages 610–623, 2021b.

[10] Shaily Bhatt, Sunipa Dev, Partha Talukdar, Shachi Dave, and Vinodkumar Prabhakaran. Re-contextualizing fairness in nlp: The case of India, 2022