International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0320
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 2045 - 2049

Designing an AI Agent for Translating Scientific Plant Names to Indian Vernaculars Using Large Language Models – A Review

Radha N Chauhan[1], Mitali B Hora[2]
[1,2]Assistant Professor, The Maharaja Sayajirao University of Baroda, Vadodara, India.
Emails: *radha.c-compapp@msubaroda.ac.in[1], mitali.hora-compapp@msubaroda.ac.in[2]*

Abstract
This paper explores creating an AI agent that uses free versions of Large Language models (LLMs) such as Gemini to convert English and Latin biological plant names into Indian languages, focusing on the Gujarati language. The paper also investigates what kind of influence such an application might have on farmers, ayurvedic practitioners, and rural people in India, with specific attention towards socio-economic usage, advantages for agricultural specialists, and its broader implications pertinent to agriculture, education, medicine, and its role in the preservation of traditional knowledge.
Keywords: *AI Agent; Agricultural Technology; Ayurvedic Knowledge; Empowering Farmers; Gujarati Language; Indian Languages; Large Language Models; Plant Name Translation; Life in Rural India; Socio-Economic Impact; Traditional Wisdom.*

1. Introduction
The access to research information regarding agriculture and botany is limited by language. In countries with diversity in regional languages, especially countries such as India, communication and understanding of information are mostly accomplished in regional languages. While school-aged children learn some English and science and gravitate around Latin and English, most farmers and rural people can only understand communication from agriculture or informal knowledge in their local or regional language. Language as an agent of communication to understand farmers' knowledge of crop varieties, crop pests, crop disease, and sustainable agricultural practices represents a significant constraint on other research and researchers disseminate similar kinds of important information. This paper intends to develop an AI agent using LLMs, such as Gemini from Google, as a research-critical application to provide credibility in terms of translation of biological plant names in English and Latin into Indic language, noting possible insights for Gujarati-speaking communities and the wider picture for rural India,

2. Literature Review
Traditional ways of translating the names of plants have often been based on either bilingual dictionaries or expertise [4], but many bilingual dictionaries do not include all plant species and adequate explanations of local dialects are difficult to obtain. The emergence of Natural Language Processing (NLP) and machine translation (MT) has offered new possibilities for automated translation. In recent developments, LLMs have exhibited an ability to comprehend and produce human-like text in different languages [6]. LLM models like Gemini have shown early promise, especially where the technology has trained on extensively multilingual datasets. The model at Gemini's scale is well-resourced to address lexicon, including translation of specialized terms like biological names, in a quantitatively similar manner across all languages to share knowledge on it [3] Studies in biodiversity informatics have also highlighted linguistic diversity during documentation of traditional ecological knowledge [1]

3. Proposed Approach
The AI agent shall be built using the following elements.

- **Input Interface:** The agent will take input from the user usually a farmer or agricultural professional through a simple interface that may involve text input for the common name (English or Latin, depending on the plant) and in more sophisticated versions, also take voice input based on the user's language.
- **Underlying Large Language Model (LLM):** The agent would be driven by a large language

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0320
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 2045 - 2049

model (LLM) such as Gemini Pro from the google-generativeai Python library, or an equivalent API [2].

- **Augmenting Knowledge:** For enhanced accuracy, the agent would preferably be derived with community curated databases of local plant names, knowledge repositories of available agricultural knowledge, and possibly community sourced knowledge.
- **Output in Indian Languages:** The agent would return translation in the target Indian languages (e.g. Gujarati, Hindi, Marathi, and Tamil). The output could also include the plants' common names and perhaps a small written description or application of the plant.
- **Accessibility Consideration:** The agent will have to consider rural accessibility factors in creating its interface - such as low literacy and low access to high-speed internet. Some of these considerations might involve offline mode or SMS interaction.

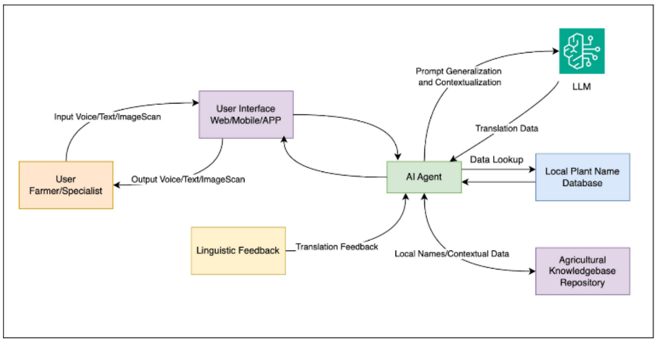

Figure 1 Architecture of proposed AI Agent

The above diagram Figure 1 shows the architecture of an AI agent that translates English/Latin plant names to Gujarati. It shows the information flow between various components of the system [5].

- **User (Ayurvedic Practitioners /Farmer/Specialist):** Placed on the left, this is the system's end-user, either a farmer or an agricultural specialist. They use the system by giving input and getting desired output from the system.
- **Input Voice/Text/Image Scan:** An arrow from the "User" to the "User Interface" indicates the different modes through which the user can enter the plant name. These include:
  1. **Voice:** Verbal description of the plant name.
  2. **Text:** Keyboard typing of the plant name.
  3. **Image Scan:** Possibly scanning an image of the plant in the future
- **User Interface (Web/Mobile/APP):** This is the layer that is referred to as the user interface, in which the user engages directly. The goals of the user interface are How the user is providing input in terms of voice or text(e.g English/Latin Name of the plant). How the system generates output for the user that could possibly still be voice, or text (e.g. the plant with the Gujarati name).
- **AI Agent:** This is a software module representing the artificial intelligence core intelligence, intermediation between the user interface, LLM and other sources of knowledge.

The AI agent has to do the following

1. Accept the user input related to the plant name from the "User Interface"
2. Data Retrieval: Access the "Local Plant Name Database" and "Agricultural Knowledgebase Repository" to find related data.
3. Data Lookup: Searching the "Locally Named Plant Database" and "Agricultural Knowledgebase Repository" for applicable information.
4. Getting the translation information from the "LLM". Formatting and sending the translated information back to the "User Interface".
5. Managing feedback from users and language experts.

- **LLM:** This is the Large Language Model (like Gemini Pro) that performs the translation. The AI Agent sends a prompt to the LLM and processes it and returns the Translated Data (the Gujarati or Indian language name of the plant name) back to the AI Agent.
- **Local Plant Name Database:** This takes the form of a database with a structure mapping various English and/or Latin biological plant names to the corresponding Gujarati or other

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0320
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 2045 - 2049

Indic language names. The AI Agent can utilize this for quick reference and be able to translate common plant names relevant to the user in appropriate language.

- **Agricultural & Ayurvedic Knowledge base Repository:** This represents information around plants, such as uses, cultivation, common names, etc.
- **Linguistic & Translation Feedback:** This shows a possible route for linguistic experts to give feedback on how accurate and good the translations are created by the system. This feedback could be used to enhance the LLM's performance (could be ultimately improved through fine-tuning in a more comprehensive situation) and the data in the knowledge bases. This shows that users (ayurveda practitioners/farmers/specialists) can provide feedback on the translations produced. This practical feedback is important to identify errors, uncertainties, and ways to improve the system.

4. Impact on Ayurveda Practioners, Farmers & Rural India

Creation and implementation of such an AI agent could have a revolutionary effect on farmers, rural societies, and Ayurveda practitioners in India:

- **Improved Access to Information:** Farmers would have access to critical information on crop management, pest and disease diagnosis, and agro-practices in their language—overcoming the hurdle presented by English or Latin scientific names. For practitioners of Ayurveda, it also translates into simpler identification and authentication of medicinal plants, since the agent would be able to correlate botanical names with classical Ayurvedic nomenclature (Dravyas) and regional names.
- **Enhanced Communication:** Improve the possible connections a farmer and another farmer, a farmer and an expert, or a farmer and traditional medicine practitioner. For example, if the farmer was to provide a local name to the Ayurveda practitioner, the pratitioner would obtain the vernacular name and traditional knowledge of that plant, thus increasing the interactions between the modern and indigenous systems of knowledge.
- **Preservation and Recovery of Traditional Knowledge:** The agent is the mediator between Ayurvedic literature (Nighantu Granthas) and modern database.It correlates between scientific names to Ayurvedic vernacular names like Haritaki (Terminalia chebula) or Guduchi (Tinospora cordifolia). This not only assists in the preservation of India's ethnobotanical wealth but also facilitates the validation and sharing of centuries-old health practices based on Ayurveda.
- **Support for Holistic and Sustainable Agriculture:** Ayurveda gives importance of harmony with nature by using prakriti-based techniques and rasa-panchaka principles (five-fold plant properties: rasa, guna, virya, vipaka, and prabhava). AI agents provide access to such information in local languages, farmers and Ayurveda professionals alike can encourage and practice more ecologically sustainable and health-focused cultivation like cultivating herbs following dosha-balancing properties or seasonal patterns (ritucharya).
- **Rural Ayurveda Practitioner Empowerment:** Ayurveda is often applied in the villages informally with locally prescribed herbs. The AI agent may provide assistance to these practitioners by confirming the local names (i.e. local plants) with recognized medicinal uses, dosages and contraindications as a means to improve safety and authenticity of traditional therapy [7-8].

5. Challenges & Future Scope

Designing an effective LLM-based AI agent for plant name translation into Indic languages has a number of challenges:

- **Accuracy and Nuances:** Most importantly we need accuracy for each translation. The fact that there are often regional variations to a single local name, and that there can be many local names for a single scientific name is going to complicate this issue.
- **Dialects:** Indian languages frequently have

e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 2045 - 2049

quite different dialects which further complicate the design of our agent and we need to consider how these dialects can be effectively translated so that the agent is useful across the variety of regions.

- **Access for Low-Resourced Farmers:** We can consider access for farmers who are low resourced and have limited internet access and have lower digital literacy by adding additional features like accessibility of content offline or accumulatively a voice-based model.
- **Scarcity of Data:** Although we can train and fine-tune LLM's using specific botanical terminology in any Indian language, we will continue to run into issues around lacking labeled data to continue to update and support the ongoing development of the original agent.
- **Creating Image Based Application:** In future this study further leads to image-based identification of ayurvedic plant in vernacular languages of India and creation of dataset for the same [9-15].

Conclusion

The construction of a LLM-based AI agent to translate English and Latin plant names into one or more Indian languages, specifically Gujarati has excited potential for enabling farmers and rural communities in India. This technology can improve access to critical information blocking language as a barrier to communication, keeping cultural knowledge alive, and ultimately benefiting socio-economic development in rural areas. Challenges remain in ensuring the translations are accurate and that there are no barriers to access, but the pace of progress in LLMs and the prospect of education and training the sector will look very promising in providing viable solutions for farmers and rural communities and ayurveda practitioners.

References

[1]. Hirschfeld, L. A. (1994). Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies. BRENT BERLIN. American Ethnologist, 21(2), 430431. https://doi.org/10.1525/ae.1994.21.2.02a00220

[2]. Gadgil, M., Berkes, F., & Folke, C. (1996). The importance of local knowledge in biodiversity conservation. Ambio, 25(3), 166–173.

[3]. Google AI. (n.d.). Gemini API Documentation. Retrieved from https://ai.google.dev/

[4]. Jain, S. K. (1991). Dictionary of Indian folk medicine and ethnobotany. Deep Publications.

[5]. World Bank. (2019). The Future of Food and Agriculture: Trends and Challenges. World Bank Publications.

[6]. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[7]. D. Chetia, S. K. Kalita, P. P. Baruah, D. Dutta and T. Akhter, “Identification of Traditional Medicinal Plant Leaves Using an Effective Deep Learning Model and Self-Curated Dataset,” arXiv preprint, arXiv:2501.09363, Jan. 2025. [Online]. Available: https://arxiv.org/abs/2501.09363

[8]. A. Kumar and R. Singh, “Ayurvedic Plant Identification Using Image Processing and Artificial Intelligence,” ResearchGate, 2023. [Online]. Available: https://www.researchgate.net/publication/357594831

[9]. P. Sharma and S. Verma, “Ayur-PlantNet: An Unbiased Lightweight Deep Convolutional Neural Network for Ayurvedic Plant Classification,” Pattern Recognition Letters, vol. 170, pp. 10–17, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S2214786123000037

[10]. M. Patel and K. Desai, “Identification of Ayurvedic Medicinal Plant Using Deep Learning,” International Journal of

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0320
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 2045 - 2049

Innovative Science and Advanced Engineering (IJISAE), vol. 3, no. 4, pp. 67–72, 2023. [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/5016

[11]. Jangalni Jadibutti- (Part-1 to 6, Illustrated)- Gujarati by Shamaldas Sevakram Gor

[12]. Kansara, Meera & Parikh, Ajay. (2020). 3RD INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING AND COMMUNICATION Indian ayurvedic plant identification using multi-organ image analytics: creation of image dataset of Indian medicinal plant organs. 10.2139/ssrn.3563074.

[13]. Singh, Harish. (2008). Importance of local names of some useful plants in ethnobotanical study. Indian Journal of Traditional Knowledge 7(2): 365-370.. 7. 365-370

[14]. Lankasena, Bentotage Nalaka Samantha & Wisumperuma, Dhanesh & Seneviratne, Bathiya & Perera, H & Nugara, R & Perera, Pathirage. (2022). Misidentifications in Ayurvedic Medicinal Plants: Convolutional Neural Network (CNN) to Overcome Identification Confusions.

[15]. Sensarma, P. (1992). Plant names - sanskrit and latin. Ancient science of life. 12. 201-20.