A Comparative Study: Exploring Methods for Domain-Specific AI Chatbot Development

Dr Dikshan N. Shah

Assistant Professor, Vanita Vishram Women's University, Surat, Gujarat

Abstract

In today's technological age, a chatbot is an intelligent system that can behave like a virtual person and effectively communicate with any human being using natural language. Major work on chatbots has already been done in English, however Indian languages have a rich morphology but limited resources. This requires further work. It is worth noting that chatbots serve as a bridge between Indian users and expert systems. This document describes the work done in Chatbot for Indian languages, as well as the problem gaps in its development. Because Indian languages have their own rules, developing a chatbot for all languages would be impossible. This study also demonstrated the difficulties in developing chatbots for Indian languages because each language has its own syntax and semantics. An expert system is based on machine learning and Artificial Intelligence (AI) concepts, and because of their dynamic nature, these chatbots include built-in AI, NLP, programming, and conversion services, which can make design and development more difficult. Although major work has not been done for Indian language chatbots, numerous scholars and organizations have begun developing chatbot systems in this subject.

Keywords : Artificial Intelligence, Machine Learning, Chatbots, Indian Languages, Natural Language Processing

1. Introduction

India is on the verge of rapid growth, with digital innovations altering even traditional government functions. Domain-specific, multilingual, knowledge-rich, intelligent chatbots are an important step forward in this direction. A chatbot is a program designed to imitate intelligent communication via text or voice input. Human engagement in their regional tongue is feasible thanks to Chatbot. AI-powered chatbots might be domain or language-specific. The chatbot classifies user input and, using pattern matching, retrieves information to provide a predetermined acknowledgment based on the user's demand. When the input query is entered into the database, the user is provided a response based on a specified pattern.

Natural Language Processing (NLP) has the potential to increase internet access for a larger proportion of India's population. Only ten percent of Indians speak English; the other ninety percent speak their regional languages, which include Hindi, Marathi, Gujarati, Bengali, Kannada, Telugu, and Tamil [30]. NLP, a subset of AI, is crucial for understanding and managing human language. Knowing a language entails understanding words, phrases, definitions, grammatical structures, and how to connect those concepts in an understandable way. NLP's key features include POS tagging, lemmatization, phrase extraction, text categorization, entity extraction, and topic extraction.

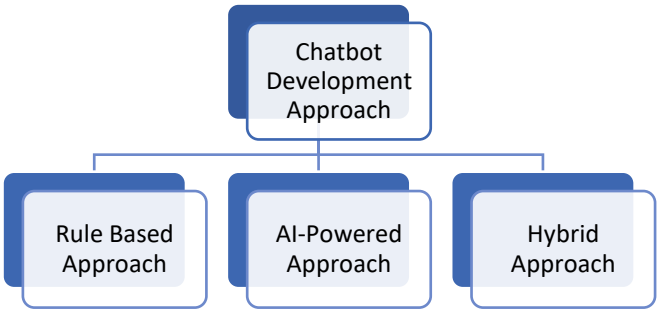
2. Applications of Chatbot

AI-based chatbots have emerged as the most adaptable tools in a variety of businesses as artificial intelligence (AI) and natural language processing (NLP) advances. And Indian language chatbots play an important function as a communication medium. AI-enabled chatbot development can be done in any domain, including:

1. Customer support, including 24/7 assistance, automated responses, and e-commerce.
2. Healthcare features include a symptom checker, appointment scheduling, and mental health support, Telemedicine
3. Education in Personalized Learning, Virtual Teaching Assistant, and Language Learning
4. Banking and Finance Services: Account Management, Loan Assistance, and Fraud Detection
5. E-commerce and retail services, including product recommendations, order tracking, and customer feedback.
6. Travel and Hospitality includes travel booking, itinerary management, and customer service.
7. Entertainment and Media for Content Recommendations: Suggesting movies, music, games, and books based on user preferences, Interactive Stories, and Event Assistance.
8. Human Resources (HR) tasks include recruitment, employee support, and onboarding.
9. Smart Cities and Public Services: Transportation Information, Complaint Management, and Utility Management.
10. Marketing and sales efforts include lead generation, campaigns, and feedback collection.
11. In Gaming, Game Guides, Player Engagement, and In-game Support
12. Legal Assistance with Document Preparation, Legal Advice, and Case Management
13. In Agriculture, Crop Advisory, give weather updates and market trends.
14. Real Estate: Property Search, Virtual Tours, and Query Handling
15. Non-profit and social services such as awareness campaigns, donor management, Volunteer Coordination

3. Existing Approach for Chatbot Development

Expert System or Human Interaction System or chatbot development approach basically classified into mainly three :



[Figure – 1 Chatbot Development Approaches]

Figure 1 depicts the two basic ways to constructing chatbots: rule-based and AI-based. Chatbot performance is improved by a combination of rule-based and AI-based approaches.

1) **Rule based Approach**
   In this approach, user defined rules are framed for responses. It is required to have strong linguistic knowledge to develop the rules. This approach can be useful in case of simple Use cases.
2) **AI-Powered Chatbots**: The implementation of Machine Learning (ML) and Natural Language Processing (NLP) to make more dynamic, responsive, efficient and context-aware interactions.
3) **Hybrid Approach**: This is a Combination of rule-based approach with Artificial Intelligence for flexibility and robustness.

4. Evaluation Metrics

Evaluation metrics are critical for measuring a chatbot's effectiveness and user happiness. These metrics can be classified according to their purposes, such as interpreting user inputs, creating suitable replies, and delivering a consistent user experience.

[Figure – 2 Objective based Evaluation metrics for Chatbot]

Figure 2 shows seven potential category metrics for measuring ChatGPT performance. Accuracy is assessed using the purpose, objective, user reaction, and feedback criteria.

5. Existing Work

Usneek Singh et al. [2023] proposed a multilingual chatbot for Indian languages that focuses on fixed-response questions and employs natural language processing methods. The authors fine-tuned the MuRIL BERT model for the SQuAD dataset and reached 76.34% accuracy, while XLM-RoBERTa achieved 49.02%, Indic 48.72%, and Monsoon NLP at 16.09%. [1]

Urmil Bharti et al. [2020] described 'Aapka Chikitsak', an AI-based chatbot for telehealth created following COVID-19 to increase healthcare access, particularly in rural parts of India

during the pandemic. It includes home cures, location-based diets, recommendations, age- and gender-specific health information, check-up guidance, and emergency helpline numbers, as well as the ability to connect to a real-time messaging service such as WhatsApp. [2]

Lasha Labadze et al. [2023] examined the benefits of AI chatbots in education for students and teachers. They also explored the problems and issues associated with adopting AI-powered chatbots for education, as well as opportunities. [3]

Sanjay Chakraborty et al. [2022] discussed how chatbots can treat infectious diseases in the medical field. They presented an AI-powered medical chatbot employing deep feedforward multilayer perceptrons. Their datasets cover a variety of symptoms and disorders, including Covid-19, Deep Fever, and chronic ailments. This information also covers their treatments, methodologies, RT PCR testing, side effects, previous infections, and more trustworthy solutions. It also has doctor's numbers, neighboring hospitals, and all necessary phone numbers. Their approach resulted in 94.32% accuracy. [4]

Athira Susan George et al. [2022] developed several chatbots that interact in Indian languages. They evaluated a variety of chatbots built for domain-specific and language-specific goals. They explored the challenges of creating chatbots in Indian languages such as Hindi, Marathi, Tamil, Malayalam, Odia, and Bengali. They also discussed a number of multilingual AI-enabled chatbots. [5]

Sagar Badlani et al. [2021] created a multilingual health-care chatbot for India's rural population, who communicate in regional languages on a daily basis. The algorithm suggests diagnoses depending on the user's health-related questions and symptoms. They assessed the system using the following classification algorithms: Random Forest, Decision Tree, SVM, MNB, and KNN, and obtained accuracy values of 98.43%, 97.12%, 96.22%, 95.39%, and 97.88%, respectively. [6]

Sachin Kuberkar and Tarun Kumar Singhal [2020] propose a chatbot for public transportation in a smart city when there is adequate infrastructure, societal acceptance, ease of use, and performance. The chatbot would be extremely useful for commuters, as a big portion of India's working population relies on an accurate schedule and high-quality public transportation systems. [10]

Varad Bhagwat et al. [2019] illustrated chatbots created exclusively for the Hindi language. They proposed query-based chatbots that employ OCR (Optical Character Recognition) to extract information from PDFs, pictures, or text-based interactions. They discussed the difficulties of constructing a chatbot in Hindi due to the intricacies of NLP. [13]

Debanjali Roy and Tanmoy Putatunda [2023] proposed an AI-based chatbot for the pedagogical process of undergraduate English literary instruction in the Indian context. Their research has also demonstrated interaction, efficacy, and collaboration among learners through a variety of activities, including case studies and practitioner interviews. [14]

Nitirajsingh Sandu and Ergun Gide [2020] investigate how chatbot technology might enhance students' learning and engagement experiences in higher education in India. They also differentiate between teacher-centered/traditional learning and AI bot-based learning. The sample survey had 24 male and 23 female respondents. [15]

Netral Pal Singh and Devender Singh [2019] proposed that different types of banks, namely private and public sector banks in India, use chatbots and virtual assistants. They also gave a brief overview of India's banking business, include ng its history, characteristics, and architecture for chatbots and virtual assistants. They discussed the essential features, ways to connect, services given, accuracy, and technology providers for chatbots and virtual assistants adopted by several Indian banks. They have included data for the chatbot from the bank's existing website regarding clients' general inquiries. [16] Table 1 demonstrates numerous chatbots created for Indian languages using different domain-specific datasets and approaches.

[Table – 1 Chatbot development for various Indian Languages]

Sr No	Year	Language	Developed Tool	Dataset Used	Approach used
1	2022	Bengali [5]	Golpo	Manual Bangla Corpus	Retrieval-based closed domain
2	2022	Bengali [17]	Doly	Educational	Machine Learning (ML) algorithms and Bengali Natural Language Processing (BNLP)
3	2022	Bengali [18]	Bengali Informative Chatbot (BIC)	Manual Dataset	Trie and Dictionary Based Search by Removing Affix (DBSRA), as well as Edit Distance
4	2022	Bengali [19]	Disha: Bangla Healthcare Chatbot	Healthcare	SVM
5	2022	Urdu [7]	Nubot	Roman Urdu	Semantic-based natural language paradigm
6	2022	Urdu [5]	Umair	Customer Service	Word Order Wizard
7	2022	Urdu [5]	Neural Encoder-Decoder Based Urdu Chatbot	History of Pakistan	Seq2Seq, Bi-directional encoder
8	2022	Odia [12]	Chatbot that implements Odia POS Tagging in IoT Devices	Natural language human generated text	POS Tagging in IoT Devices
9	2022	Marathi [20]	Query-Based Chatbot System in Marathi Language		Text detection, localization and tracking modules
10	2022	Marathi [5]	AMBER	Paraphrasing Detection Library	phonetic and stripping algorithms
11	2022	Hindi [21]	Query-Based Chatbot in Hindi	PDF and Images	Query-based System
12	2022	Hindi [11]	AskDISHA	IRCTC and CoRover Passenger	artificial intelligence (AI), machine learning (ML), and natural language processing (NLP)
13	2022	Malayalam [8]	An Interactive AI chatbot in Malayalam	Kerala govt policies	NLP and AI
14	2022	Malayalam [23]	Malayalam Chatbot based on Chatterbot Python Library	Malayalam Tourism corpus	Machine Learning

15	2022	Tamil [5]	Poongkuzhali - - An Intelligent Tamil Chatbot	Manually Created	Artificial Intelligence
16	2022	Tamil [22]	Tamil Chatbot using Intent Recognition	Educational	Natural Language Processing (NLP) along with the Dialog Flow API
17	2019	English [13]	A Graph Based Chatbot for cancer Patients	Forums Data	Knowledge Based Chatbot

Table 2 describes a number of existing chatbots designed for Indian languages that handle multiple languages and use various domain-specific datasets and techniques.

[Table – 2 Multi-Lingual supported Chatbot]

Sr No	Year	Language	Developed Tool	Dataset Used
1	2019	Multi- Lingual [24]	Niki	AI-Powered
2	2019	Multi- Lingual [25]	Vernacular.ai	Business Owner
3	2020	Multi- Lingual [26]	Aham	Covid-19
4	2020	Multi- Lingual [27]	COVID-19 Self Help Chatbot	Guidelines of WHO and Indian Government
5	2020	Multi- Lingual [28]	Haptik	Customer Ineractive
6	2020	Multi- Lingual [29]	Flow.ai	Manually Created
7	2021	Multi- Lingual [22]	Recharge Bot by Payjo	Customer Services for Mobile
8	2021	Multi- Lingual [23]	Dhee.ai	Business

6. Challenges

Precise AI-powered chatbot systems are increasingly accessible in European languages, particularly English. The issues faced by South and Southeast Asian languages continue to demand significant attention. The problem definition demonstrates the many obstacles present with Indian language chatbots. They are as follows.

- Translating idioms and slang remains challenging, despite advancements in the field. Indian languages contain ambiguous terms.
- NLP algorithms should address ambiguities and complexities.
- Indian languages have smaller data sets than Western languages.
- Not all Indic languages have NLP features like intent recognition, entity identification, and grammatical phrasing.
- There are limited resources accessible for Indian language literature and grammar.
- Some chat platforms have inconsistent features.
- English sentences follow subject-verb-object word order, but most Indic languages use subject-object-verb.

7. Conclusion

This research mainly focuses on the design, development, and performance of numerous Indian-language chatbots. The rise of chatbots has simplified the lives of both its providers and users, as both profit from it. However, the bots are still in their early stages of learning. The accuracy will improve as more data is sent to them, hence increasing the system's efficiency. Chatbots can be developed using a variety of approaches and technologies, depending on their intended application and expected results. The study explores several available methodologies and platforms for creating chatbots in Indic languages. Chatbot installation has resulted in cost

savings, more sales, higher customer happiness, and massive growth for organizations. This tremendous increase in multilingual chatbots should be viewed positively, and many Indian businesses should invest in this technology. Despite the fact that most Indic languages have limited resources, the field of Indic language NLP will see greater activity in the future.

References

[1] Singh, U., Vora, N., Lohia, P., Sharma, Y., Bhatia, A., & Tiwari, K. (2023, July). Multilingual chatbot for Indian languages. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

[2] Roy, D., & Putatunda, T. (2023). From textbooks to chatbots: Integrating AI in English literature classrooms of India. Journal of e-Learning and Knowledge Society, 19(3), 65-73.

[3] Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education, 20(1), 56.

[4] Chakraborty, S., Paul, H., Ghatak, S., Pandey, S. K., Kumar, A., Singh, K. U., & Shah, M. A. (2022). An AI-based medical chatbot model for infectious disease prediction. Ieee Access, 10, 128469-128483.

[5] George, A. S., Muralikrishnan, G., Ninan, L. R., Varrier, P. S., & Dhanya, L. K. (2021, June). Survey on the design and development of Indian language chatbots. In 2021 International Conference on Communication, Control and Information Sciences (ICCISc) (Vol. 1, pp. 1-6). IEEE.

[6] Badlani, S., Aditya, T., Dave, M., & Chaudhari, S. (2021, May). Multilingual healthcare chatbot using machine learning. In 2021 2nd International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.

[7] Shabbir, J., Arshad, M. U., & Shahzad, W. (2021). Nubot: Embedded knowledge graph with rasa framework for generating semantic intents responses in roman urdu. arXiv preprint arXiv:2102.10410.

[8] George, A. S., Muralikrishnan, G., Ninan, L. R., Varrier, P. S., & Dhanya, L. K. (2021, June). Survey on the design and development of Indian language chatbots. In 2021 International Conference on Communication, Control and Information Sciences (ICCISc) (Vol. 1, pp. 1-6). IEEE.

[9] Bharti, U., Bajaj, D., Batra, H., Lalit, S., Lalit, S., & Gangwani, A. (2020, June). Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after covid-19. In 2020 5th international conference on communication and electronics systems (ICCES) (pp. 870-875). IEEE.

[10] Kuberkar, S., & Singhal, T. K. (2020). Factors influencing adoption intention of AI powered chatbot for public transport services within a smart city. International Journal of Emerging Technologies in Learning, 11(3), 948-958.

[11] Dey, S., & Shukla, D. (2020, March). Analytical study on use of AI techniques in tourism sector for smarter customer experience management. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-5). IEEE.

[12] Sahoo, R., Das, B. R., & Mishra, B. K. (2020, March). Analysis and Implementation of Odia Part of Speech Tagger in recent IoT based devices through Chatbot: A review. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-4). IEEE.

[13] Bhagwat, V., Nagarkar, M., Paramane, P., Jindam, S., & Kharate, N. G. (2019). Review of Chatbot system in Hindi language. International Research Journal of Engineering and Technology (IRJET), 6(11), 2395-0072.

[14] Sandu, N., & Gide, E. (2019, September). Adoption of AI-Chatbots to enhance student learning experience in higher education in India. In 2019 18th international conference on information technology based higher education and training (ITHET) (pp. 1-5). IEEE.

[15] Sandu, N., & Gide, E. (2019, September). Adoption of AI-Chatbots to enhance student learning experience in higher education in India. In 2019 18th international conference on information technology based higher education and training (ITHET) (pp. 1-5). IEEE.

[16] Singh, N. P., & Singh, D. (2019). Chatbots and virtual assistant in Indian banks. Industrija, 47(4).

[17] Kowsher, M., Tithi, F. S., Alam, M. A., Huda, M. N., Moheuddin, M. M., & Rosul, M. G. (2019, May). Doly: Bengali chatbot for bengali education. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (pp. 1-6). IEEE.

[18] Kowsher, M., Alam, M. A., Uddin, M. J., Islam, M. R., Pias, N., & Saifullah, A. R. M. (2019, July). Bengali Informative Chatbot. In 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) (pp. 1-7). IEEE.

[19] Rahman, M. M., Amin, R., Liton, M. N. K., & Hossain, N. (2019, December). Disha: an implementation of machine learning based Bangla healthcare chatbot. In 2019 22nd International Conference on Computer and Information Technology (ICCIT) (pp. 1-6). IEEE.
[20] Navalakha, D., Pittule, M., Mane, R., Rathod, A., & Kharate, N. G. (2019). Review of Chatbot system in Marathi language. International Research Journal of Engineering and Technology (IRJET), 6(11).
[21] Bhagwat, V., Nagarkar, M., Paramane, P., Jindam, S., & Kharate, N. G. (2019). Review of Chatbot system in Hindi language. International Research Journal of Engineering and Technology (IRJET), 6(11), 2395-0072.
[22] Udhayakumar, S., & Rajeswari, P. (2019). Human-Computer text conversation through NLP in Tamil using Intent Recognitio. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-5).
[23] Sandhini, S., Sreeja, P., & Sreeja, P. (2018). A proposal of chatbot for Malayalam. International Journal of Computer Sciences and Engineering, 6(6), 21-25.
[24] "AI-based chatbot Niki to add seven more Indian languages," Sap.io, 14-Oct-2019. [Online]. Available: https://sap.io/ai-based-chatbot-niki-toadd-seven-more-indian-languages/.
[25] S. Nair, "Multilingual chatbots making conversational AI vernacular," Haptik.ai, 24-Apr-2019. [Online]. Available: https://www.haptik.ai/blog/building-multilingual-chatbots/.
[26] "Meet aham! Your new virtual assistant providing information on COVID-19 in 7 Indian languages," Outlookindia.com, 21-Mar-2020. [Online].
Available: https://www.outlookindia.com/newsscroll/meetaham-your-new-virtual-assistant-providing-information-on-covid19-in-7-indian-languages/1785923.
[27] H. T. Correspondent, "Covid-19 app, chatbot launched for people in Bihar," 02-Apr-2020. [Online]. Available: https://tech.hindustantimes.com/tech/news/covid-19-app-chatbotlaunched-for-people-in-bihar-story-291GpnvSHza27YcqqCuw5K.html.
Available:https://analyticsindiamag.com/7-chatbot-building-platforms-fromindia-that-are-revolutionising-businesses/.
[28] H. T. Correspondent, "Reliance Jio Haptik-powered coronaviruschatbot arrives on WhatsApp," 22-Mar-2020. [Online].
Available: https://tech.hindustantimes.com/tech/news/reliance-jiohaptik-powered-coronavirus-chatbot-arrives-on-whatsapp-story-KvS0RDxGNQRYungitkycSO.html.
[29] "Multi-language design," Flow.ai. [Online]. Available:https://flow.ai/docs/learn/multi-language-design.
[30] Shah, D. N., & Bhadka, H. B.. Named entity recognition from Gujarati text using rule-based approach. In *Intelligent Systems Design and Applications: 17th International Conference on Intelligent Systems Design and Applications (ISDA 2017) held in Delhi, India, December 14-16, 2017* (pp. 797-805). Springer International Publishing.