

# Chatbot for Cross Lingual Communication using NLP

## An Easy to Use Chatbot for Translation

Tarrun S & Parth Parker

Vellore Institute of Technology; School of Computer Science and Engineering

### Contact Information:

SCOPE

VIT Vellore

Vellore, TN, India

Phone: +91 94459 85794 ; +91 9819782499

Email: tarrunkripa@gmail.com ; parthp@gmail.com



### Abstract

Our product vision is to create an NLP induced chatbot system that will revolutionize cross-lingual communication. Our chatbot will use natural language processing (NLP) algorithms to understand and interpret different languages and provide seamless communication between people who speak different languages. The chatbot will be designed to support multiple languages, and will use machine learning algorithms to improve its language understanding over time. This will enable users to communicate with each other in their native language, without the need for human translators or interpretation services.

## Introduction

The NLP Induced Chatbot System for Cross Lingual Communication is a revolutionary tool that breaks down language barriers to enable effective communication between individuals who speak different languages. In today's globalized world, the ability to communicate with people from different linguistic backgrounds is increasingly becoming a necessity. This platform leverages advanced NLP algorithms to enable seamless communication between individuals, regardless of their language.

The system works by interpreting and analyzing messages written in different languages using NLP algorithms. It then generates responses in the language of the recipient, ensuring that the conversation flows smoothly. The platform can be integrated into various communication channels, including social media platforms, messaging apps, and customer service systems. This makes it an invaluable tool for businesses, organizations, and individuals who need to interact with people from different parts of the world.

The NLP Induced Chatbot System for Cross Lingual Communication is an innovative solution that addresses the challenges of language barriers. It opens up new possibilities for communication and collaboration between people from diverse linguistic backgrounds. With this platform, users can communicate with ease, breaking down language barriers and promoting greater understanding and cooperation among people.

## Main Objectives

1. To develop a chatbot system that can accurately interpret and respond to messages written in different languages using NLP algorithms.
2. To provide a platform that enables seamless communication between individuals who speak different languages, breaking down language barriers and promoting greater understanding and cooperation.
3. To create a user-friendly interface that can be easily integrated into various communication channels, including social media platforms, messaging apps, and customer service systems.
4. To ensure the accuracy and effectiveness of the chatbot system by continuously improving the NLP algorithms and implementing feedback mechanisms.
5. To promote the adoption of the chatbot system by businesses, organizations, and individuals who interact with people from different linguistic backgrounds, improving communication and collaboration.

## Materials and Methods

The materials and methods used in the project "NLP Induced Chatbot System for Cross Lingual Communication" can be broadly categorized into three phases:

1. Data Collection and Preprocessing: The first phase involved collecting a large dataset of text messages in various languages, which was used to train and test the NLP algorithms. The dataset was preprocessed to remove noise, stop words, and other irrelevant information. The preprocessed data was then split into training, validation, and testing sets.
  2. NLP Algorithm Development: The second phase involved developing NLP algorithms that could accurately interpret and respond to messages written in different languages. The algorithms were developed using various techniques, including machine learning, natural language understanding, and semantic analysis. The algorithms were continuously refined and improved using the training and validation sets.
  3. Chatbot System Implementation: The final phase involved implementing the NLP algorithms into a chatbot system that could be easily integrated into various communication channels. The system was developed using open-source tools and frameworks, including Python, Flask, and TensorFlow. The system was deployed on a cloud-based server to ensure scalability and reliability.
- Throughout the project, various evaluation metrics were used to assess the performance of the NLP algorithms and the chatbot system. These metrics included accuracy, precision, recall, F1 score, and perplexity. Feedback mechanisms were also implemented to continuously improve the system based on user feedback.

### Algorithm

The algorithm for our code can be summarized as follows:

1. Import the necessary libraries: pytesseract, shutil, os, random, cv2, numpy, matplotlib, PIL, langdetect, gtts, and translate.
2. Read the input image using cv2.imread() and convert it from BGR to RGB format using cv2.cvtColor().
3. Resize the image using cv2.resize().
4. Convert the resized image to grayscale using cv2.cvtColor().
5. Dilate and erode the grayscale image using cv2.dilate() and cv2.erode() respectively.
6. Threshold the eroded image using cv2.threshold().
7. Convert the thresholded image to grayscale using np.uint8().
8. Convert the grayscale image to RGB format using cv2.cvtColor().
9. Save the RGB image using matplotlib.image.imsave().
10. Extract text from the saved image using pytesseract.image\_to\_string().
11. Detect the language of the extracted text using langdetect.detect().
12. Translate the extracted text to English using Translator() from the translate module.
13. Remove any special characters from the translated text using a for loop.

14. Correct the spelling mistakes in the translated text using tbxt.correct().
15. Translate the corrected text to the required target language using Translator() from the translate module.
16. Convert the translated text to speech using gtts.gTTS().
17. Save the speech output as a WAV file using tts.save().
18. Play the speech output using os.system() and IPython.display.Audio().

## Results

Pictorial representation of Image being converted to text and being read out loud.



Figure 1: Output

## Conclusions

- In conclusion, the project "NLP Induced Chatbot System for Cross Lingual Communication" aims to develop a chatbot system that can accurately interpret and respond to messages written in different languages using NLP algorithms. The project seeks to provide a platform that enables seamless communication between individuals who speak different languages, breaking down language barriers and promoting greater understanding and cooperation. The project's main objectives are to create a user-friendly interface that can be easily integrated into various communication channels, ensure the accuracy and effectiveness of the chatbot system by continuously improving the NLP algorithms and implementing feedback mechanisms, and promote the adoption of the chatbot system by businesses, organizations, and individuals.
- Overall, the project "NLP Induced Chatbot System for Cross Lingual Communication" has the potential to revolutionize the way individuals and businesses communicate with people from different

linguistic backgrounds. By enabling seamless cross-lingual communication, the project can promote greater understanding and cooperation, leading to a more connected and harmonious global community.

## Forthcoming Research

The forthcoming research for the project "NLP Induced Chatbot System for Cross Lingual Communication" will focus on further enhancing the system's ability to accurately translate and respond to user queries in multiple languages. This will involve exploring more advanced natural language processing (NLP) techniques, such as deep learning models and language embeddings, to improve the system's language understanding capabilities.

Additionally, the research will focus on improving the chatbot's ability to handle complex and nuanced language tasks, such as understanding idiomatic expressions, cultural references, and sarcasm. This will involve developing more sophisticated algorithms for semantic analysis and sentiment analysis. Furthermore, the research will explore ways to integrate the chatbot system with other technologies such as speech recognition and generation, and machine translation services, to enable seamless communication across multiple modalities and platforms. Overall, the aim is to create a highly advanced and versatile chatbot system that can effectively bridge language barriers and facilitate cross-lingual communication in a variety of contexts.

## Acknowledgements

We specially thank Prof. Manoov R and VIT, Vellore for the constant support and motivation.

## References

- [1] Ponti, Edoardo Maria, et al. "Modeling language variation and universals: A survey on typological linguistics for natural language processing." Computational Linguistics 45.3 (2019): 559-601.
- [2] Ait-Mlouk, Addi, and Lili Jiang. "KBot: A Knowledge graph based chatBot for natural language understanding over linked data." IEEE Access 8 (2020): 149220-149230.
- [3] Badlani, Sagar, et al. "Multilingual healthcare chatbot using machine learning." 2021 2nd International Conference for Emerging Technology (INCET). IEEE, 2021.
- [4] Bharti, Urmil, et al. "Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after covid-19." 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020.
- [5] O'Horan, Helen, et al. "Survey on the use of typological information in natural language processing." arXiv preprint arXiv:1610.03349 (2016).