



ERROR - 404

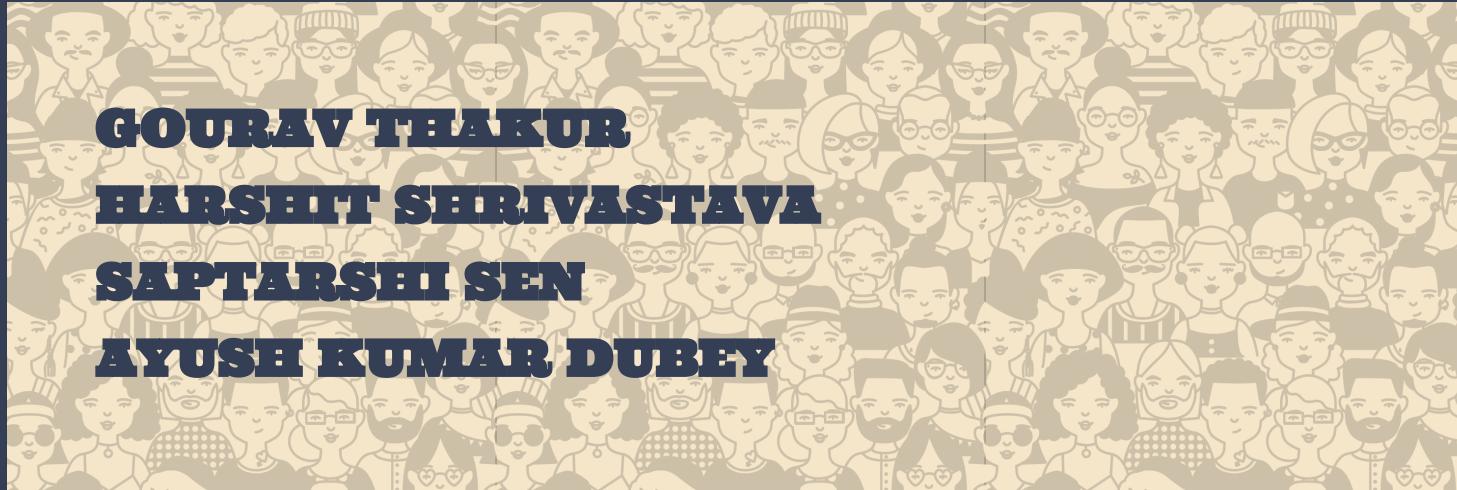
TRANSLINGO-GPT





ERROR - 404

TEAM MEMBERS :





ERROR - 404



PROBLEM STATEMENT

Our world is interconnected, yet language differences hinder communication. We're building a web app for direct translation between English, Sanskrit, Hindi and Hinglish.

This user-friendly tool breaks language barriers by translating between any of these languages, fostering collaboration across cultures.



ERROR - 404



PROBLEM APPROACH

Our approach tackles the challenge of building a website that translates between English, Sanskrit, and Hindi. It emphasizes overcoming limitations of current translation tools by proposing a character-level language model with a state-of-the-art Transformer architecture. We will train this model to understand complex grammar and translate text or speech input into the desired language (text or speech output).



ERROR - 404

Overview

We are building a language translator by training a transformer on large language data and deploying the model on a website.

Our project stands out from other translators because we have added a new feature to affect the tone of translation. Generally we do not get the option to manage our tone in a translator app or website, but we managed to add 2 tones to the translation of english to hinglish(Interrogative and assistant type). We can add multiple tones to multiple translations if the dataset is available of that kind.



ERROR - 404

SOLUTION APPROACH

Our approach tackles the challenge of building a website that translates between English, Sanskrit, Hindi and Hinglish. It emphasizes overcoming limitations of current translation tools by proposing a character-level language model with a state-of-the-art Transformer architecture. We will train this model to understand complex grammar and translate text or speech input into the desired language (text or speech output). We propose the development of a language translator web application equipped with advanced features to facilitate cross-language communication. The application will be built using the NodeJS web framework for seamless deployment of machine learning models.



ERROR - 404

DATA ACQUISITION AND PREPROCESSING:

- Gather English to Sanskrit and English to Hindi translation datasets.

- Preprocess the datasets to clean and normalize the text.
Sources: Hugging Face, Kaggle



ERROR - 404

MODEL TRAINING:

We will train a character-level language model based on the Transformer architecture to understand the grammatical intricacies of each language. The Transformer will have multiple encoder and decoder blocks. The decoder will use the output from the encoder along with the text generated till now and predict the next character in the required language. The self attention mechanism will help in translation longer texts and still give accurate results.

Library Used:

- PyTorch
- PyTorch.nn

- Numpy
- Pandas



ERROR - 404

NODEJS WEB APPLICATION DEVELOPMENT:

- Develop the front-end of the web application using HTML, CSS, and JavaScript for user interaction and interface design.
- Develop the front-end of the web application using HTML, CSS, and JavaScript for user interaction and interface design.
- Develop the front-end of the web application using HTML, CSS, and JavaScript for user interaction and interface design.



ERROR - 404

INTEGRATION OF MACHINE LEARNING MODELS:

- Integrate the trained translation models into the Node JS application.
- Implement translation functionality that takes input text in input language and outputs translations in Sanskrit, Hindi, Hinglish or English.
- The GPT model will be integrated into a user-friendly website using NodeJS in the backend and HTML in the client-side. The model will be converted to “.pt” file.



ERROR - 404

TEXT-TO-SPEECH (TTS) INTEGRATION:

- ▶ Integrate a text-to-speech system to convert translated text into spoken words.

- ▶ Utilize libraries like Google Text-to-Speech or Mozilla TTS for high-quality audio output.



ERROR - 404

TESTING AND DEPLOYMENT:

- ▶ Thoroughly test the application to ensure accurate translations and proper functioning of text-to-speech and speech-to-text features.

- ▶ Deploy the Flask application to a web server, ensuring scalability and accessibility to users worldwide.

OBJECTIVE



ERROR 404

By implementing this solution, we aim to bridge the language gap and provide users with a convenient and efficient tool for cross-language communication, enhancing global connectivity and collaboration. This solution leverages Flask for its simplicity and flexibility in deploying machine learning models, ensuring a smooth user experience and efficient translation services.





ERROR 404

EXPECTED OUTCOMES

The project is expected to perform well for medium and short length sentences.

The English and Hinglish model is trained on a dataset that contains mostly conversations of a human and a online assistant. So, the model is expected to perform extremely well on those kinds of sentences. The translation of Sanskrit, in particular, is expected to be grammatically accurate because the grammar of Sanskrit is very strict and rule-based. Since the model is trained using shlokas and religious texts, the output might sound very formal.

This may or may not be ideal for everyday conversation, but it ensures high accuracy in conveying the core meaning, especially for similar contexts. We plan to explore incorporating more diverse training data in the future to improve the model's ability to handle casual language for Sanskrit and Hinglish model.



ERROR 404

ACTUAL OUTCOMES

Tone - Assistant

Model name :- **eng_to_hing_st.pt**

Input sentence	Target sentence	Model's output
What is the time?	Time kya ho raha hai?	Time kya hai?
start the timer.	Timer start karo.	Timer ko start karo.
I want to listen to a song.	Mai gaana sunna chahata hu.	Mai ek song sunna chahta hu.
Turn on the fan.	Fan on karo.	Ek fan ko on karo.
Cancel all meetings of next week.	Agle hafte ke saare meetings cancel karo.	Next week ke liye meetings cancel kare.



ERROR 404

ACTUAL OUTCOMES

Tone - Interrogative

Model name :- english_hinglish_double_train.pt.pt

Input sentence	Target sentence	Model's output
What is the time?	Time kya ho raha hai?	Yeh time kya hai?
start the timer.	Timer start karo.	Start the timer.
Can i play with your friend?	Mai gaana sunna chahata hu.	kya main khel sakte hain?
What are you doing there?	Aap waha kya kar rahe hain?	Kya aap kar rahe hain?
Can you do it?	Kya aap kar sakte hai?	Kya aap kar sakte hai?



ERROR 404

ACTUAL OUTCOMES

Hinglish to English model

Input sentence	Target sentence	Model's output
Aaj kitne alarms hai?	How many alarms today?	How many alarms today?

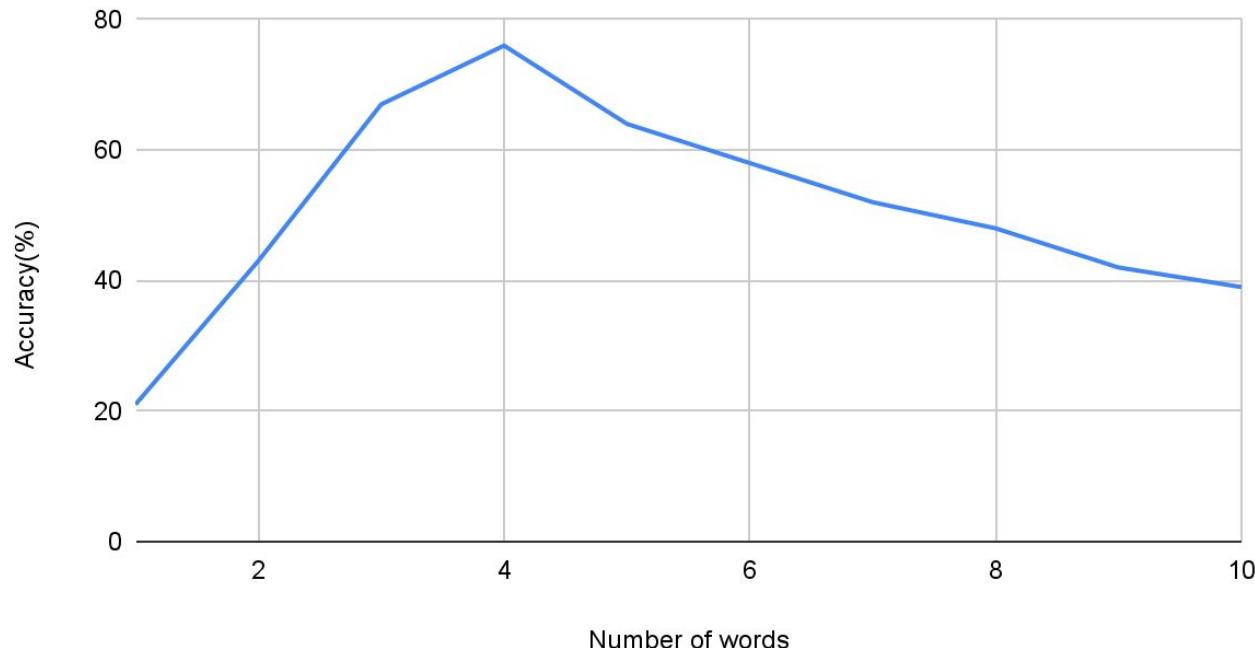
Outputs similar to the English to Hinglish model.

Same goes for the interrogative tone model of Hinglish to English



ERROR 404

Accuracy(%) vs. Number of words



**ERROR - 404**

CONCLUSION

This problem statement was more over solved by us as our target was to build a character level language model using transformer that translates between several languages. We expected to get accurate results on longer sentences which we could achieve in some of the models we trained. The english to hinglish model could translate longer sentences very easily if input was of similar type as the dataset we feeded (user and assistant communication). The multitone feature also worked fine because the interrogative tone model could translate questions, better than the assistant model.

We expected the sanskrit translation to be grammatically accurate, which the model was, Because sanskrit has very mathematical grammar which the model could easily understand but it could not figure out the meaning of the translation because of the lack of sufficient data set.



ERROR 404

THANK YOU



ありがとう

شُكْرًا

謝謝

ধন্যবাদ