**Kannada AI ChatBot**

**Indian Lanaguage Translation:** Initially , I was looking for kannada language translator (from kannada language to English). In that process, I found this IndicTran2 model from AI4Bharat.

Here is the github link: <https://github.com/AI4Bharat/IndicTrans2/>

Research paper link: <https://arxiv.org/pdf/2305.16307>

And I found this website really helpful for us to get the dataset and models especially on Indian language data: [here](https://ai4bharat.iitm.ac.in/)

This model is available in 3 different versions.

1. English to any 22 indian language - it has 1B parameterized large model which can runs on GPU and CPU (~4gb) , And also it has smaller version, it is available in distilled version (size ~1.5gb)

Large model: [ai4bharat/indictrans2-indic-en-1B](https://huggingface.co/ai4bharat/indictrans2-en-indic-1B)

Smaller Model: [ai4bharat/indictrans2-indic-en-dist-200M](https://huggingface.co/ai4bharat/indictrans2-indic-en-dist-200M)

1. Any Indian language to English : - Same it also available in both larger and in smaller versions

Large Model: [ai4bharat/indictrans2-indic-en-1B](https://huggingface.co/ai4bharat/indictrans2-indic-en-1B)

Smaller Model: [ai4bharat/indictrans2-indic-en-dist-200M](https://huggingface.co/ai4bharat/indictrans2-indic-en-dist-200M)

1. Any Indian Language to any other Indian Language : - Same this is also available in larger and smaller versions

Large model: [ai4bharat/indictrans2-indic-indic-1B](https://huggingface.co/ai4bharat/indictrans2-indic-indic-1B)

Smaller Model: [ai4bharat/indictrans2-indic-indic-dist-320M](https://huggingface.co/ai4bharat/indictrans2-en-indic-dist-200M)

Available Indian languages are: [languages](https://github.com/AI4Bharat/IndicTrans2/blob/main/utils.map_token_lang.tsv)

Check the colab code once: [code here](https://colab.research.google.com/github/AI4Bharat/IndicTrans2/blob/main/huggingface_interface/colab_inference.ipynb#scrollTo=_7TxTTCoKjti)

My Thoughts on this model: These models are working good. I just tested with kannada and English language only. It gives me the correct high accurate responses. Needs to test more with different inputs.

**Speech Recognition:** Now , when it comes to speech recognition (live transcription), from my past experiences, I have tried with different open source models.

1. As per my experience, the best model I can say is Openai’s [whisper large v3](https://huggingface.co/openai/whisper-large-v3) model. Its an open source and able to recognize the words correctly even in the noise environment , and able to add symobls (like ?, . ! etc. ) at respective places and also giving the time lines for each words , sentences and it can help for [speaker diarization](https://huggingface.co/pyannote/speaker-diarization) task as well. But It has some latency issues when it is running on cpu, and also it’s a large model (~3.9gb). It is available in smaller distilled version, like [distil-large-v3](https://huggingface.co/distil-whisper/distil-large-v3). And not suitable for live transcription.
2. Next, I can say Google’s Speech recognition library. Its simple and powerful. It can generate the live transcriptions and with high accurate. It takes the speeches into chunks (with time period let’s say 5sec) and transcribe it into text. After every time period , it takes some time to transcribe (let’s say 2-3 sec) , in that time , it was unable to capture the speech and transcribing it. Also it has some latency issues when displaying these transcriptions on frontend with javascript (using web sockets.)
3. I have also tried and experimented with some speech to text live transcription models such at Vits (not suitable in noise environment) and faster whisper model (takes more time to load the model and generating the transcriptions but it can give high accurate transcripts)
4. I have tried with some paid api models such as deepgram and assembly ai, in this case , deepgram works very well when compare to assembly ai
5. Apart from these models, I have used javascript web kit speech recognition library which can truly perform better and display the live transcriptions faster and support for indian languages like kannada, tamil, Telugu …etc. But it is also have one drawback, even it gives the transcriptions faster , some times the transcriptions are not much accurate , and it might face challenges in noise environment to generate transcriptions.

Here, in this application I am using this javascript web kit speech recognition for transcribing user message ,If it is a English , then set the language as ‘en-IN’ and if it is a kannada language , then set the language as ‘kn-IN’.

And also tested this [addy88/wav2vec2-kannada-stt](https://huggingface.co/addy88/wav2vec2-kannada-stt) model , but it takes the audio file as input. Not supported for live transcription

**Text to Speech:** Here, I simply use google’s text to speech library (gtts library from python), it is simple and effective. But the speech is looks more robotic and not able to pronouns some words accurately, some times it face some challenges with numbers and special symbols.

2. Also tested facebook’s massively multilingual speech (mms) model , works better that gtts and takes time to process the audio. Available with different gender tones.

Needs to test more libraries and models for better accurate speech with different tones and understanding the emotion in words.

**LLM model:** Here, I have used llama3 with 8b parameters model using groq api inference engine. It can give me the better response when compare to other llm models like mixstral, llam2.