

THiNK STT/TTS PROJECT.

The THiNK STT/TTS project presented several challenges and lessons that informed both its implementation and areas for improvement.

Challenges Faced

One of the primary challenges was ensuring **transcription accuracy**. The initial STT model struggled with accurately transcribing spoken input, especially in scenarios involving diverse accents or noisy environments.

To address this, i implemented enhanced audio pre-processing techniques, including normalization and resampling of the audio data, which significantly improved the model's performance.

Another issue arose with **response generation**. The language model occasionally produced irrelevant or nonsensical outputs, particularly when handling less common queries.

To resolve this, i introduced predefined responses for frequently asked questions and refined the prompting strategy to better guide the model toward generating contextually relevant responses.

Additionally, **the language model's thought process was sometimes visible in its outputs**, resulting in verbose and less clear responses.

By simplifying and refining the prompts, i ensured that the final outputs were concise and direct, effectively hiding the model's internal processing steps.

The lightweight nature of the distilgpt2 model also posed challenges, as it struggled to handle a wide range of queries or generate contextually appropriate responses.

Recognizing the inherent limitations of the model, I compensated by further enhancing predefined responses and carefully controlling the prompts to improve the overall relevance of its outputs.

Lessons Learned

Through this project, I learned the critical importance of proper audio pre-processing. Ensuring that audio input is normalized and resampled correctly is essential for achieving accurate transcription results from the STT model.

I also recognized the significant value of predefined responses. By incorporating them into the system, I improved the relevance and accuracy of the assistant's outputs, providing a more reliable and user-friendly experience.

Crafting effective prompts for the language model proved to be another essential skill. Thoughtful and precise prompting greatly influenced the quality and contextual relevance of the model's responses.

Lastly, the project highlighted the limitations of lightweight models. While these models offer efficiency, they may fall short in handling complex or diverse queries. This underscored the need for predefined responses and refined prompting strategies to mitigate these limitations.

Improvements Needed

Moving forward, there are several areas for improvement. Upgrading to more advanced language models, such as GPT-3.5-turbo, would enhance the system's ability to generate accurate and contextually relevant responses. Additionally, implementing more sophisticated audio pre-processing techniques could further improve the STT model's performance, particularly in handling diverse accents and noisy environments.

Expanding the set of predefined responses to cover a broader range of common questions and scenarios would make the assistant more robust and versatile. Integrating a feedback loop to learn continuously from user interactions would also be beneficial, enabling the system to refine both its transcription and response generation capabilities over time.

In summary, the THiNK STT/TTS project offered valuable insights into the challenges and opportunities of building a lightweight voice assistant. By addressing these lessons and implementing the suggested improvements, future iterations of the system can achieve greater accuracy, relevance, and user satisfaction.