Experience with Podcast LM

December 30, 2024

1 Generation with Gemini

I used a vietnamese law document as an input for Gemini to generate a podcast script with a host and a guest. The input document is about duties and missions of a department in Lao Cai. See result in this github file. In the overall, I think Gemini did a good job in summary, create a script with mix of tones, verbal fillers as in a normal discussion.

2 Generation with Llama3.3

I used the same document and prompt as used for Gemini, but Llama3.3 failed to create a script. It instead give us just a summary of the document (result). But if we take a look at the summary and the result of Gemini, we can see that they both captured what the document is about. I think we can break down the system prompt into summary and script generation as 2 separeted tasks, and run through Llama3.3 2 times.

3 Voice Generation

The primary challenge with voice generation was achieving a natural and expressive tone. I found 2 open-source models to serve out needs.

One is facebook/mms-tts-vie, I generated an audio file from Gemini's script. There are 2 problems with this model, the voice is not intersting and there is only 1 male voice.

Another model is capleaf/viXTTStext. This model is extended from coqui/XTTS-v2, it is very good with voice tones, and we can generate voice with our own voice sample. But a huge problem is that Coqui is shutting down since January 2024. The model is still in an unstable state and I can't run it on local, even on Google colab, it is not stable, usually crash kernel.