fine-tuning the summarization model

Created	@August 1, 2025 7:19 PM
Tags	
Type	Research

training a GPT or sumarization transformer is technically infeasible

- do not have 100s of GPUs and many training takes many months on end
- dont have large volumes of data

the next most challenging thing: fine tune on the AMI meeting dataset

some things to consider

- meeting transcripts will be chunked- for better processing
- the longest meeting we can transcribe while mantaing performance constraints is 15-20 minutes- that too with chunking.

two ways of doing this:

1. train an extractive summarizer first and then train an abstractive summarizer- like a t5 on the output from the extractive summarizer
2. use a pre-trained extractive summarizer and then train an abstractive summarizer on the output from the extractive summarizer
3. we could just use an abstractive summarizer as well, but alone it can be slower for large transcripts

first is a bit slow but more techically challenging

second is slightly less technically challenging compared to first, but app performace is improved due to analready robust extractive summarizer

extractive summary

you pick key sentences from the text, join them.

abstractive summary

you look at keywords and important sentences from the text and then paraphrase it as well. it handles the “um.. uh...”s of a meeting and filler words/sentences better.