1. Differences between prompted and unprompted transcriptions

When we compared the two versions, we noticed some interesting differences. The **unprompted transcription** was more literal—it included all the small noises, hesitations, and even little mistakes in the speech, basically a word-for-word capture. On the other hand, the **prompted transcription** was cleaner and more structured. It removed some of the filler words, noise, and unnecessary repetitions, which made it easier to read.

However, we ran into a **big issue**: in the prompted version, the third chunk of audio wasn't transcribed correctly at first. Instead of giving the transcription, the model sometimes just repeated the text of the prompt itself. I've seen this happen in other Whisper-based transcription services too. It took several attempts to adjust the prompt text so that the third chunk would finally get transcribed properly and so that the prompt would actually improve the transcription overall.

This experience shows that prompts can both **improve and worsen transcription quality**, so they need to be carefully tested and adjusted.

---

2. Benefits of chunking for long audio

Chunking the audio into smaller pieces (in our case, 60-second chunks) was really helpful. First, it **prevented API errors**, because Whisper has a limit on the maximum file size it can handle. Without chunking, longer files could crash or fail to process.

Second, chunking allowed us to **focus on each part individually**, which made it easier to apply different prompt strategies or settings if needed. And third, it helped with **timing and timestamps**, so we could attach precise start and end times to every segment in the TXT, JSON, and SRT outputs.

Overall, chunking makes working with long audio files much more reliable and manageable.

---

3. Challenges you faced

We faced several challenges while doing this lab:

- **Prompt behavior:** The biggest one was that the prompt sometimes made the transcription worse, instead of better. As mentioned, the third chunk would sometimes output the prompt text itself. We had to test different prompt versions to make it work correctly.
- **ChatGPT code limitations:** While writing the code, ChatGPT didn't always handle all the tasks in one go. For example, initially it forgot to include timestamps or didn't export all three formats (TXT, JSON, SRT) for both prompted and unprompted versions. We had to remind it and adjust the code several times.

- **Audio quality issues:** Some parts of the audio were quiet or unclear, which required careful chunking and prompt wording to ensure the model could still capture most of the speech.

---

4. Recommendations for improving accuracy

Based on our experience, here are some tips:

- **Test multiple prompt versions:** A prompt can improve transcription, but it can also make it worse. Always try a few variations to see which one works best for your audio.
- **Chunk long audio:** Break audio into smaller segments (like 60 seconds) to avoid API size limits and make it easier to manage transcription and timestamps.
- **Include timestamps:** Always add start and end times for each chunk, especially if you need SRT or JSON outputs.
- **Expect some manual cleanup:** Even with a good prompt and chunking, some unclear or noisy sections might need manual review.
- **Keep prompts short and clear:** Avoid overly long instructions or assumptions about audio clarity; Whisper works better when the prompt tells it to **transcribe everything as accurately as possible**, even if the speech is unclear.