

Speech-to-Text Transcription Report

Differences between prompted and unprompted transcriptions

The guided transcription is clearly closer to the original source than the unguided one, preserving wording, structure, and meaning more faithfully. While both outputs are readable, the unguided version introduces a critical factual error by hallucinating “Maine” instead of “Earth” and tends to smooth or interpret speech, which reduces authenticity. The guided transcript maintains more accurate lexical choices, better reflects the original sentence flow, and preserves hesitations, fillers, and informal speech patterns. Although neither version captures speaker turns or conversational features like laughter and overlaps, the guided approach consistently delivers higher semantic accuracy and avoids misleading additions. Overall, prompted transcription stayed closer to the original text and preserved meaning more reliably.

Benefits of chunking for long audio

Chunking long audio into smaller segments improves transcription stability and accuracy. Shorter chunks reduce decoding complexity, improve timestamp precision, and limit error propagation. Chunking also makes it easier to isolate, correct, or reprocess errors and enables scalable processing for long recordings. In practice, it helped reduce semantic drift in longer transcriptions.

Challenges faced

Several practical challenges were encountered. Audio recording on macOS caused PortAudio and microphone permission issues that required manual configuration. Environment setup problems, such as missing .env files or incorrect paths, initially prevented API key loading. Additionally, committing to GitHub failed because the .env file was included in an early commit before being added to .gitignore, requiring history cleanup.

Recommendations for improving accuracy

- Use guided prompts for domain-specific or technical content
- Apply chunking for longer recordings, ideally with overlap or sentence-aware splitting
- Prefer verbose output with timestamps for review and post-processing
- Manually validate critical transcripts to detect subtle hallucinations
- Ensure good audio quality through consistent sample rates and minimal background noise

Combining prompted transcription with chunked processing provides the most reliable results for long-form speech-to-text tasks.