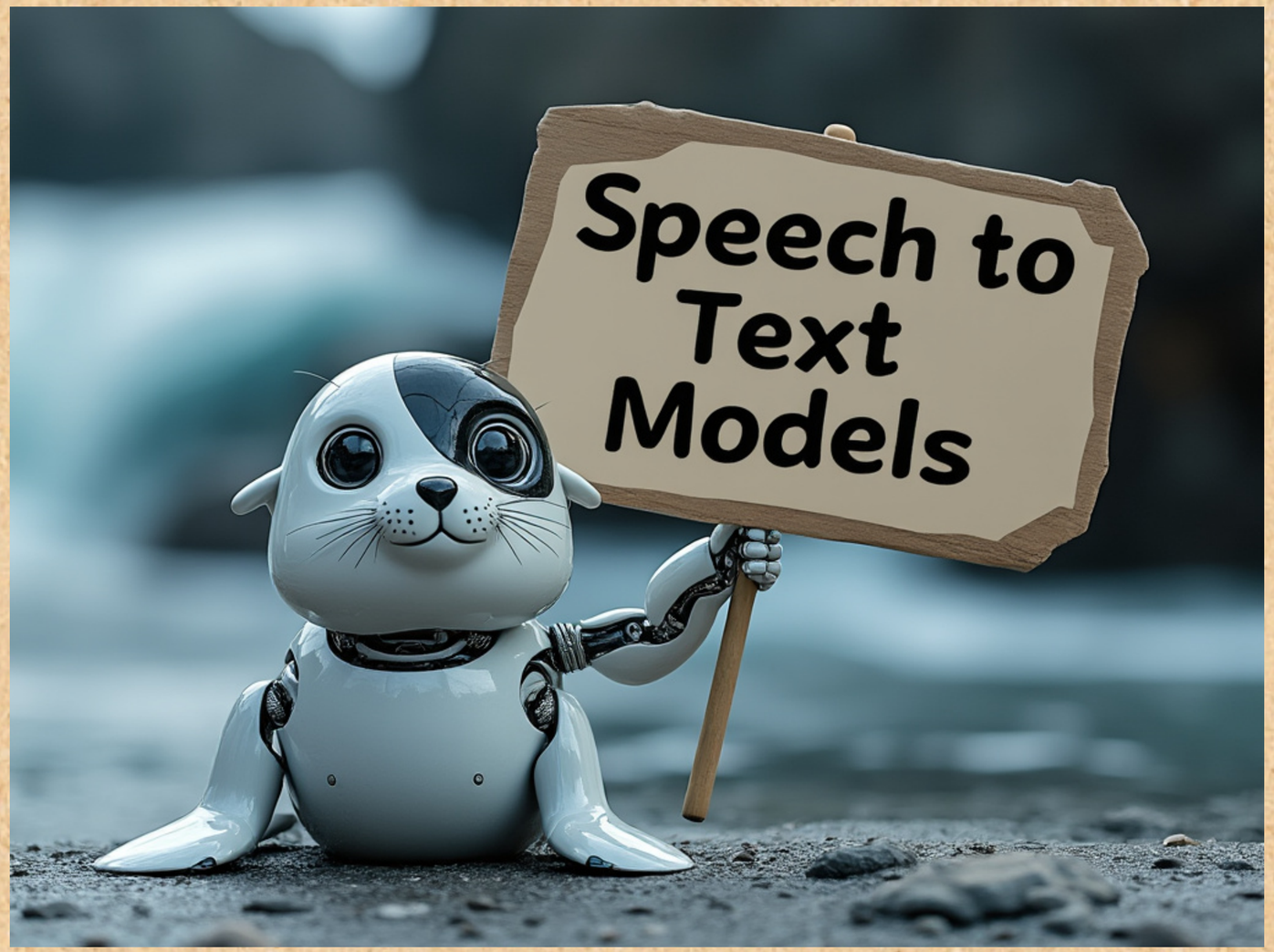
Module 3 - Topic 4

3.4.3 Speech to Text Models


Speech to
Text
Models

What are Speech to Text Models?

Speech to text models, also known as Automatic Speech Recognition (ASR) systems, convert spoken language into written text.

- Process audio input through neural networks trained on vast amounts of speech data
- Convert sound waves into text while handling different accents and languages
- Leverage transformer architecture for improved accuracy

Applications of ASR

Speech-to-text technology has revolutionized how we interact with devices and process audio content.

- Real-time captioning for videos and live events
- Voice assistants and automated customer service
- Meeting transcription and documentation
- Accessibility tools for hearing-impaired individuals

Popular Tools and Models

Some of the leading solutions in STT space

- OpenAI Whisper
- Amazon Transcribe
- Google Speech-to-Text

Using STT Models using API

ft. Whisper