1. Course overview

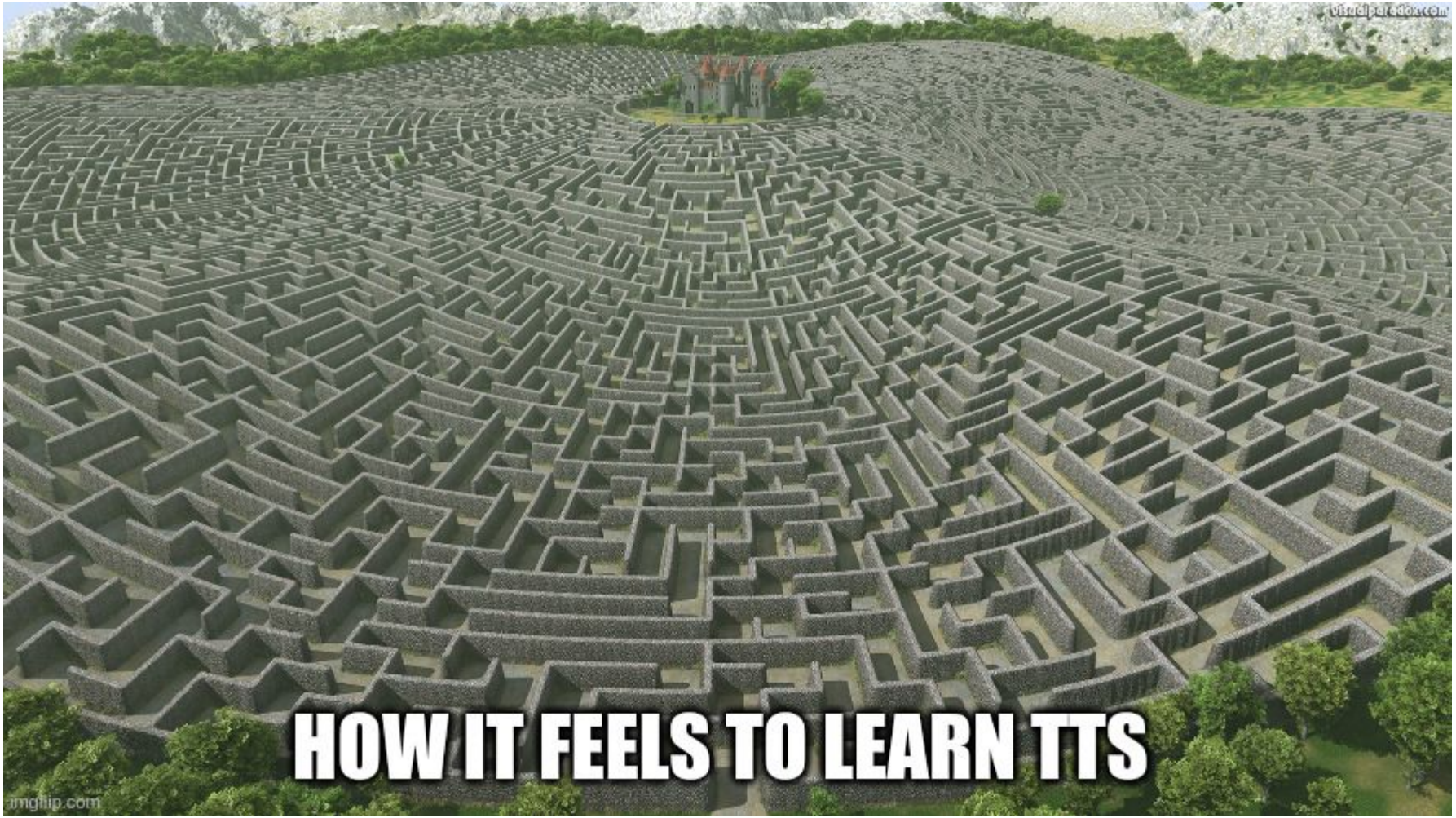
The Monster Text to Speech & Voice Cloning Course

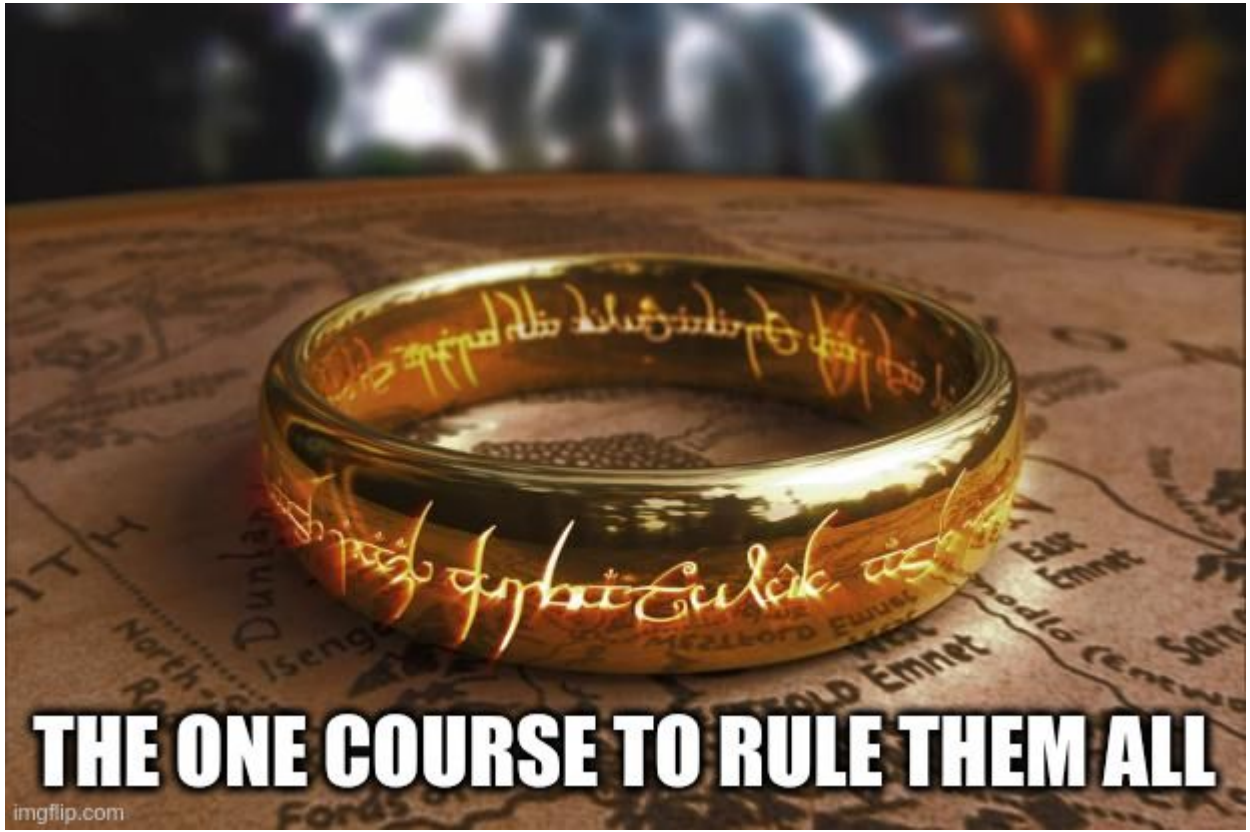
THE SOUND OF AI

Understand how machines learn to speak, from phonemes and text to realistic voices and expressive AI speech

Why learn TTS / Voice Cloning?

- Voice AI is exploding
- Voice = main interface between humans and AI
- Lots of jobs for voice ML specialists


HOW IT FEELS TO LEARN TTS


THE ONE COURSE TO RULE THEM ALL
imgflip.com

Who’s this course for?

- ML engineers / researchers
- Audio programmers
- Developers
- Engineering managers
- Product managers with technical understanding

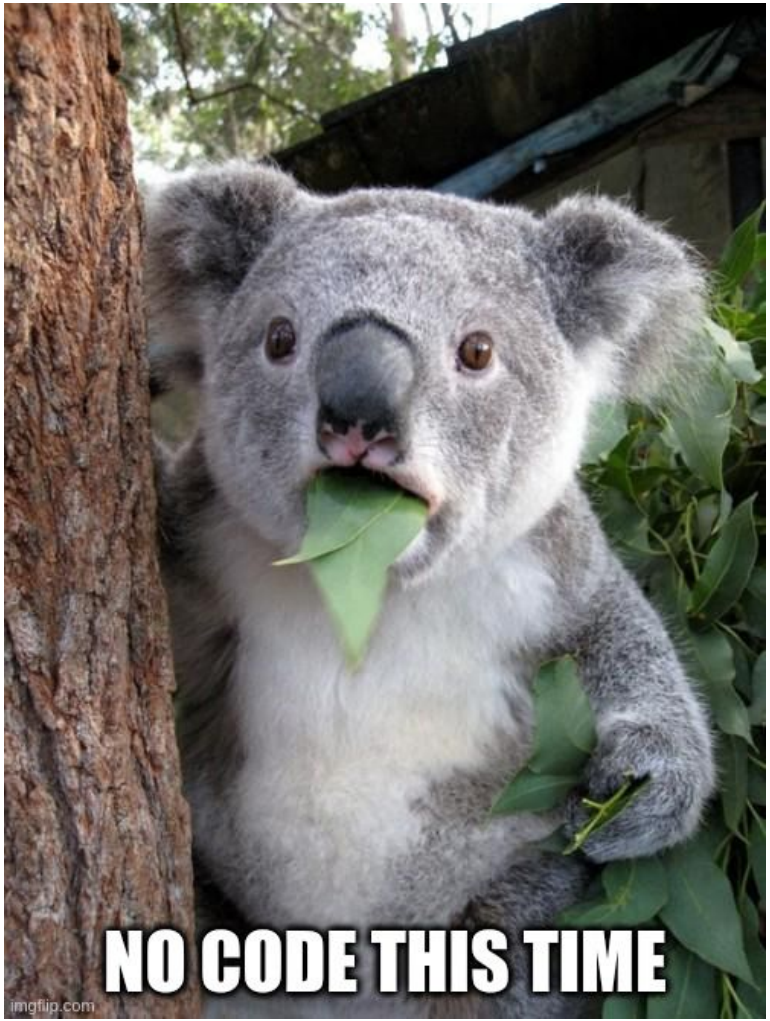
Pre-requisites

- Basic ML concepts (e.g., embeddings, training)
- DL architectures (e.g., transformers, GANs)
- Basic DSP (e.g., power spectrum, spectrogram, waveform)
- No coding experience is OK
- Willing to learn… a ton


INTUITION
THEORY
imgflip.com

Teaching style

- Big-picture conceptual videos
- Zoom-ins on specific systems
- Pointers to papers, models, and demos


NO CODE THIS TIME
imgflip.com

What you'll learn

What you'll learn

Foundations

- How humans speak
- How machines process text and phonemes

What you'll learn

Foundations

- How humans speak
- How machines process text and phonemes

Core technologies

- From traditional TTS to neural vocoders (e.g., HiFi-GAN)
- Neural codecs and discrete audio tokens (e.g., EnCodec)
- Self-supervised speech representation (e.g., WavLM)
- Codecs-based generation and voice cloning (e.g., VALL-E)

What you'll learn

Foundations

- How humans speak
- How machines process text and phonemes

Core technologies

- From traditional TTS to neural vocoders (e.g., HiFi-GAN)
- Neural codecs and discrete audio tokens (e.g., EnCodec)
- Self-supervised speech representation (e.g., WavLM)
- Codecs-based generation and voice cloning (e.g., VALL-E)

Frontiers

- Emotion, prosody, and accent modeling
- Voice singing and speech-to-speech generation
- Conversational and voice agents

Where do I get the learning material?

Where can I discuss course topics?

- The Sound of AI Slack community -> #tts-course
- YouTube comments

How do I get the most out of it?

- Watch actively: pause, think, sketch the flow
- Study (some of) the papers I cover
- Play around with the models
- Follow all the videos in order


PUT THE TIME IN!
imgflip.com

Course pace

- 1-2 videos per week
- 3-5 months to publish all videos
- 3-5 hours of study / week


LET'S GET STARTED
imgflip.com