

Ref Image

Ref Audio

Text Prompt



A man stands at the podium in OpenAI's luxurious conference room, behind him a massive electronic screen displays the company's glowing profit data. He grips the microphone firmly, gazes across the audience below, and announces in a steady tone: <S>The board wants to sell OpenAI to Zuckerberg, which is unacceptable.<E>

Ref Image

Ref Audio

Text Prompt



A woman stands before the Trevi Fountain at dawn, holding two coins in her palm. She says to the empty piazza: <S>May your wishes find their way to water, and your heart always know the way back.<E>

Baseline



+ Face&Timbre embeddings



+ Contrastive learning losses

