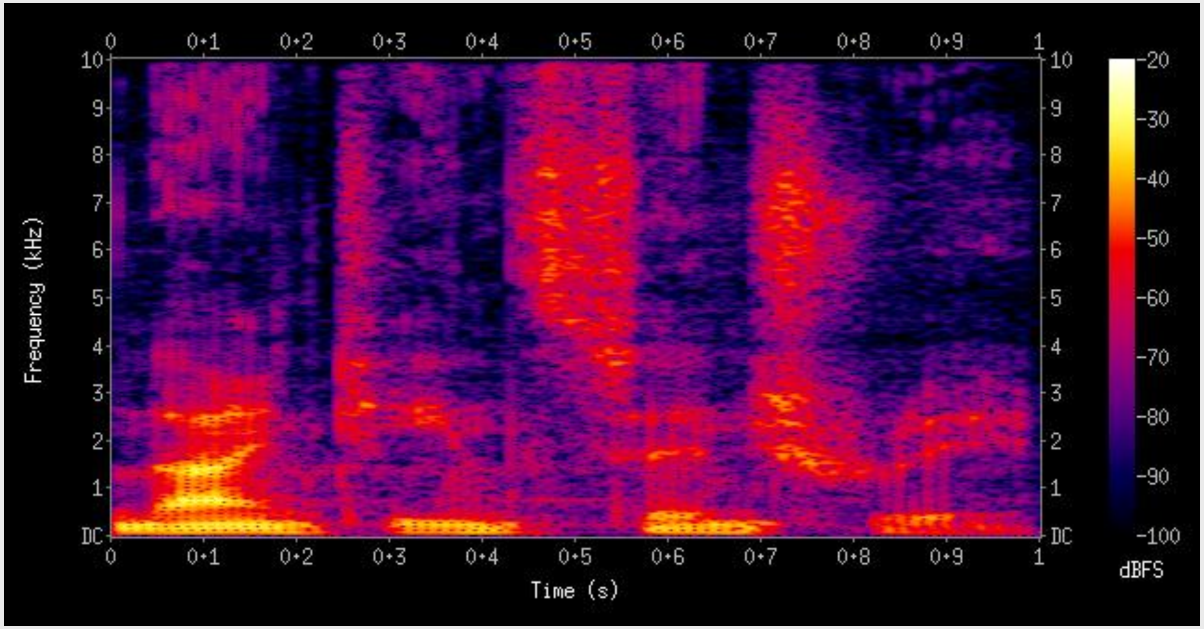
VOICE CLONING

AI Voice Cloning

- AI can easily mimic someone’s speaking style in text with transformers
- This is possible because the AI is able to represent text with a good embedding vector
- If the AI wanted to represent the sound of someone’s voice, what would the embedding vector represent?

Spectrograms: How Computers Represent Voices

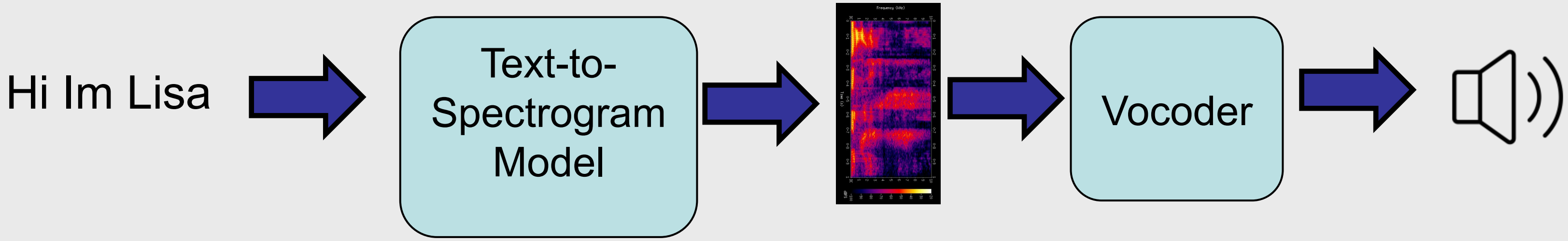
- Spectrogram – represents the frequencies in an audio signal as it varies in time
  - Like an audio fingerprint



Text-to-Speech (TTS) Models

- Text to speech (TTS) models are neural networks that convert text into spectrograms

- The spectrograms are then turned into voice audio using neural networks called vocoders

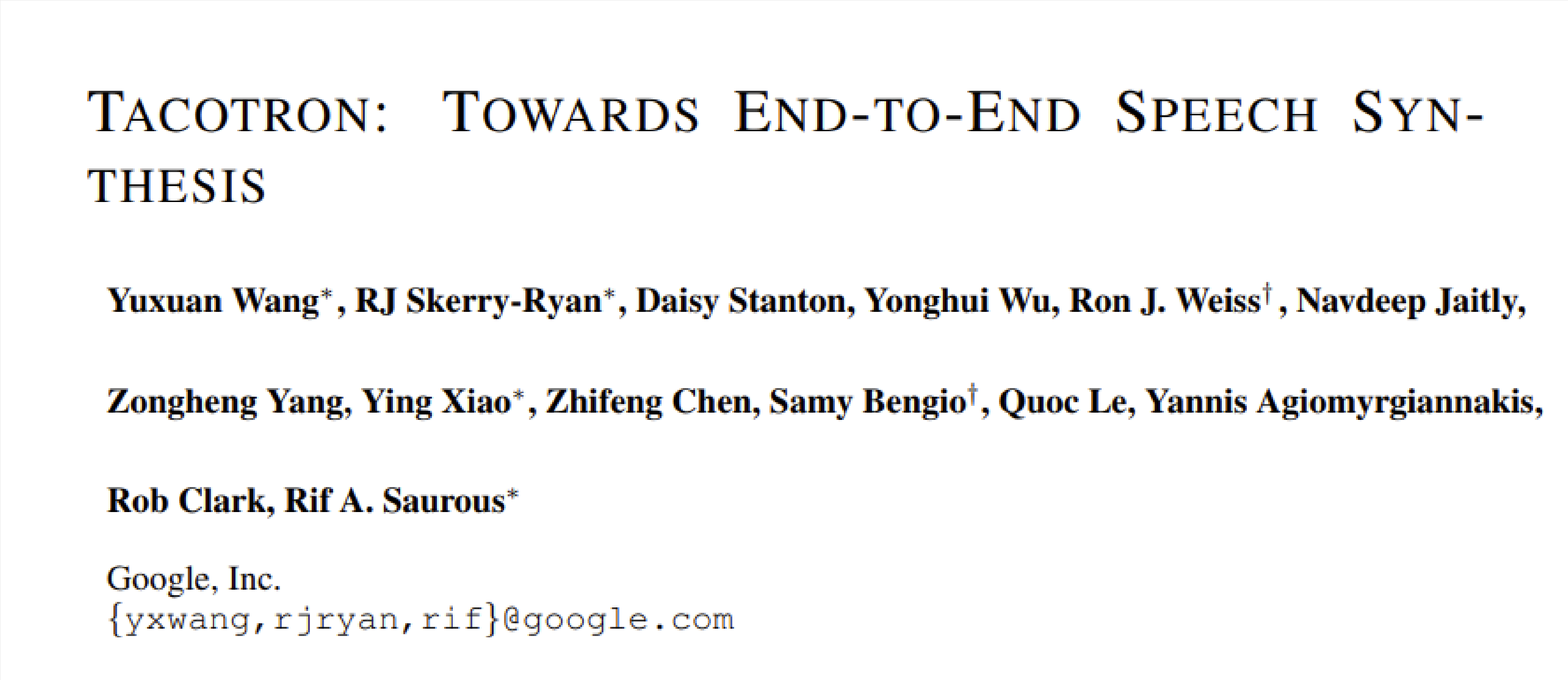


Hi Im Lisa

Text-to-Spectrogram Model

Vocoder

First Good TTS Model: Tacotron

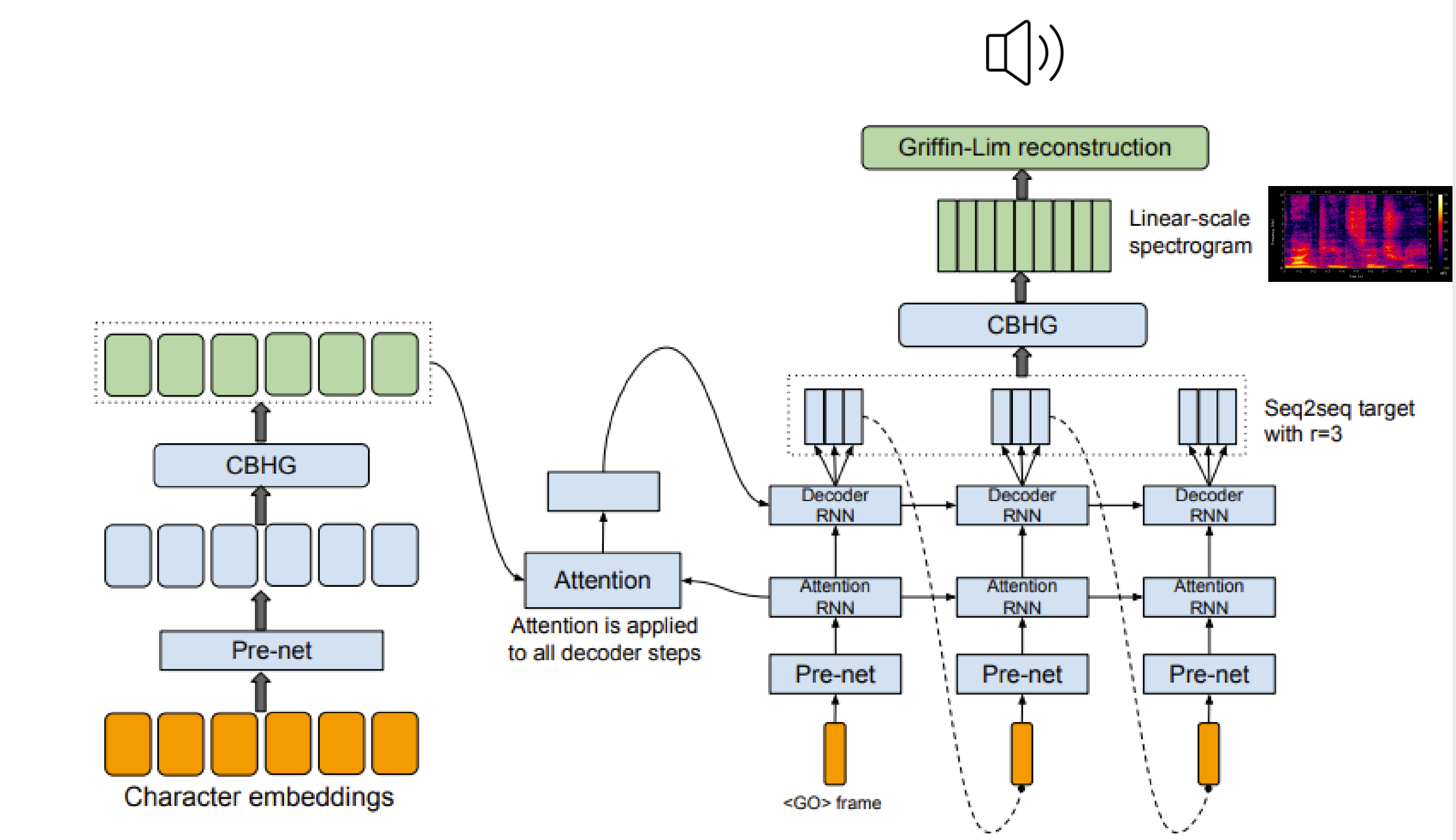


Tacotron: Towards End-to-End Speech Synthesis

Yuxuan Wang[*], RJ Skerry-Ryan[*], Daisy Stanton, Yonghui Wu, Ron J. Weiss[†], Navdeep Jaitly, Zongheng Yang, Ying Xiao[*], Zhifeng Chen, Samy Bengio[†], Quoc Le, Yannis Agiomyrgiannakis, Rob Clark, Rif A. Saurous[*]

Google, Inc.
`{yxwang,rjryan,rif}@google.com`

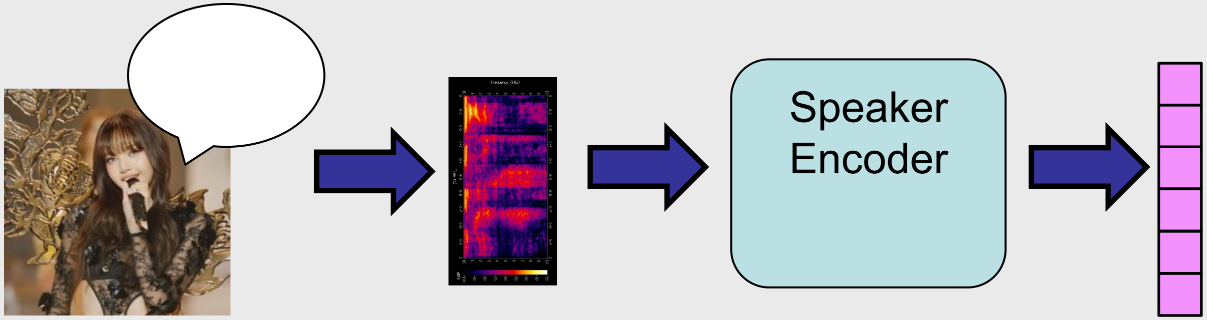
Tacotron Architecture



Hi Im Lisa

Representing a Speaker

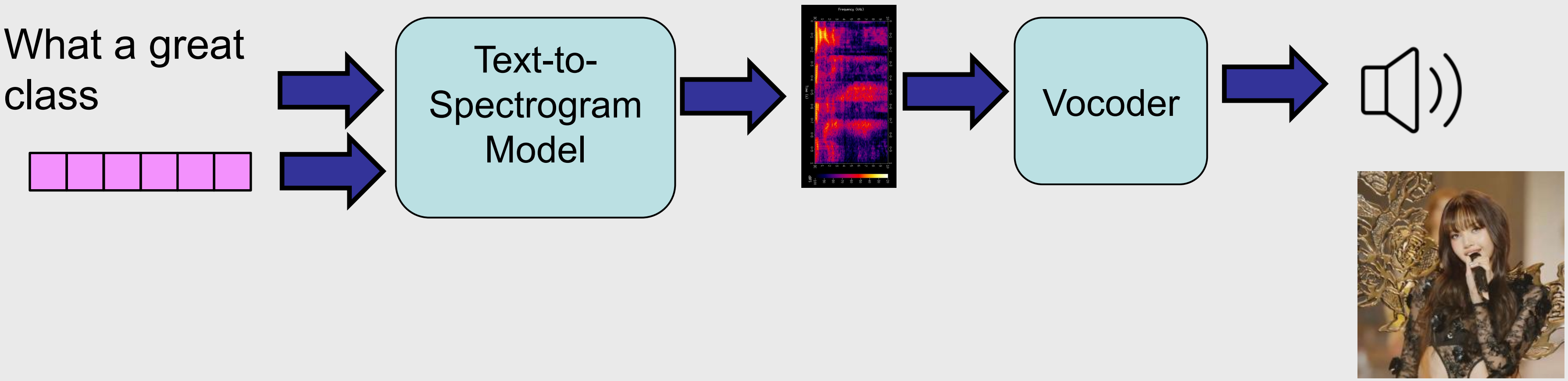
- Spectrograms are like voice fingerprints
- Voice cloning neural networks called “speaker encoders” take spectrograms and turn them into vectors
- Trained using contrastive learning



Speaker Encoder

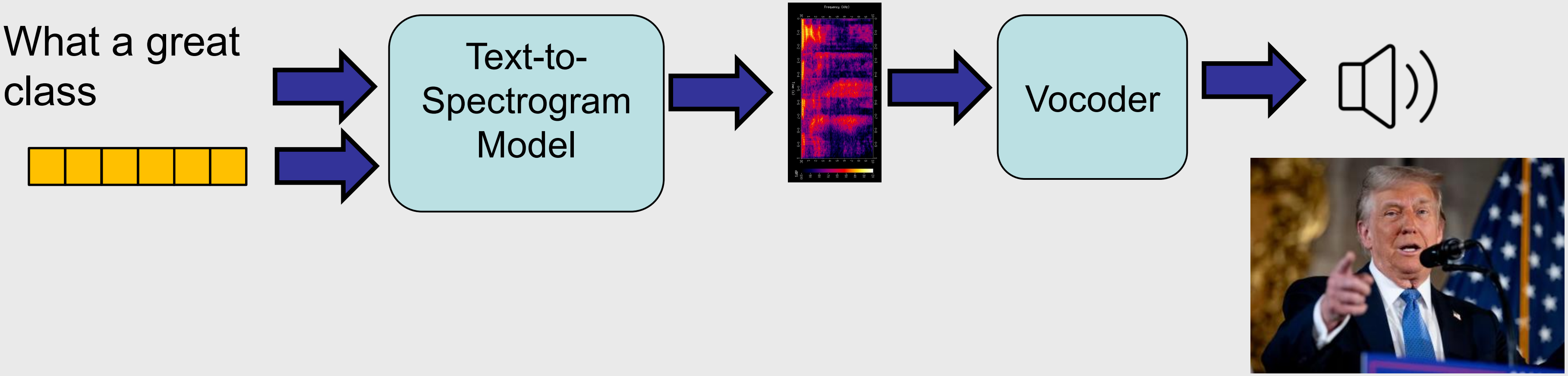
Voice Cloning

- Once we have the speaker vector, we can just add it to the TTS model



Voice Cloning

- **Once we have the speaker vector, we can just add it to the TTS model**



What a great class

Text-to-Spectrogram Model

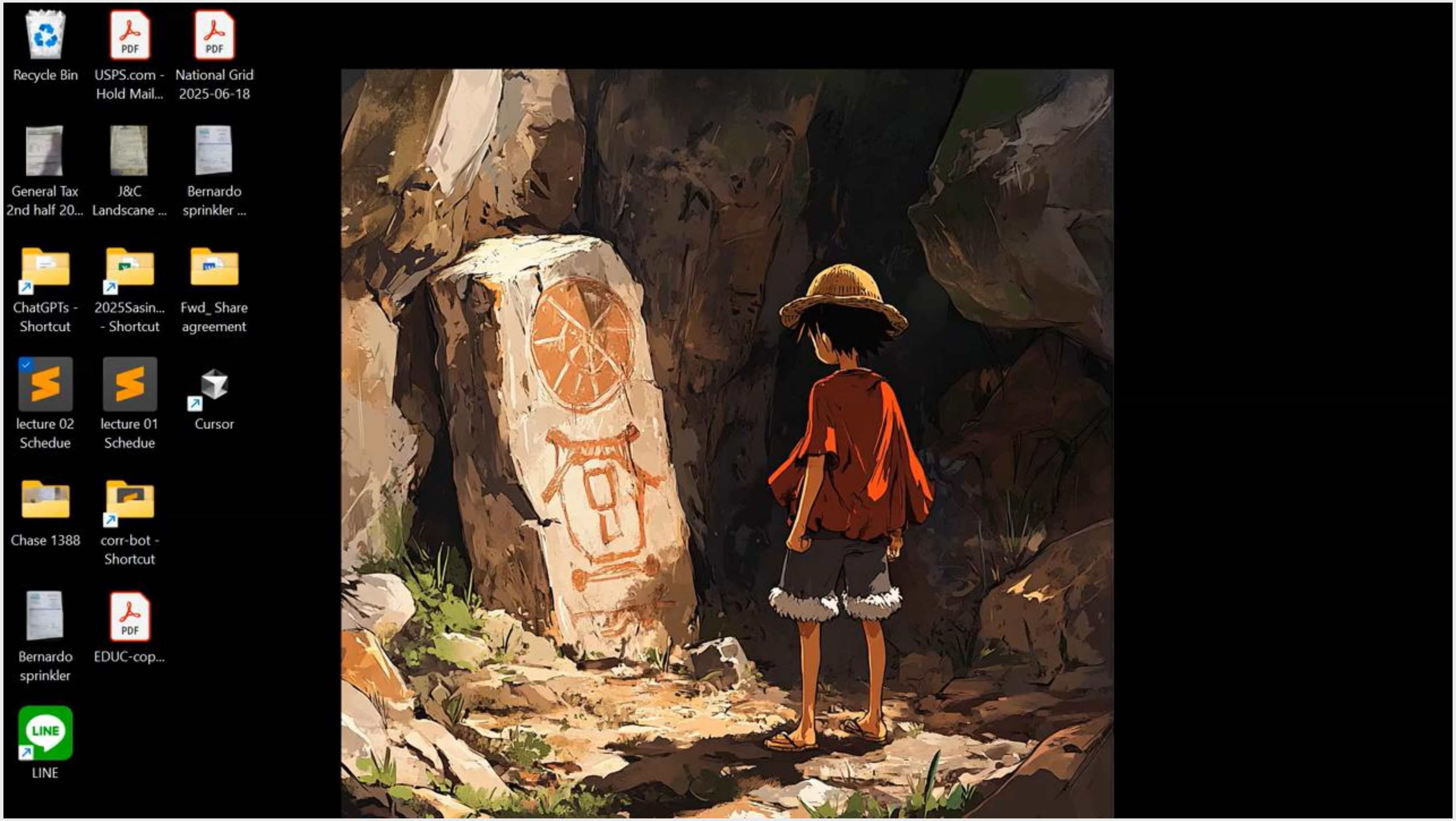
Vocoder

ElevenLabs

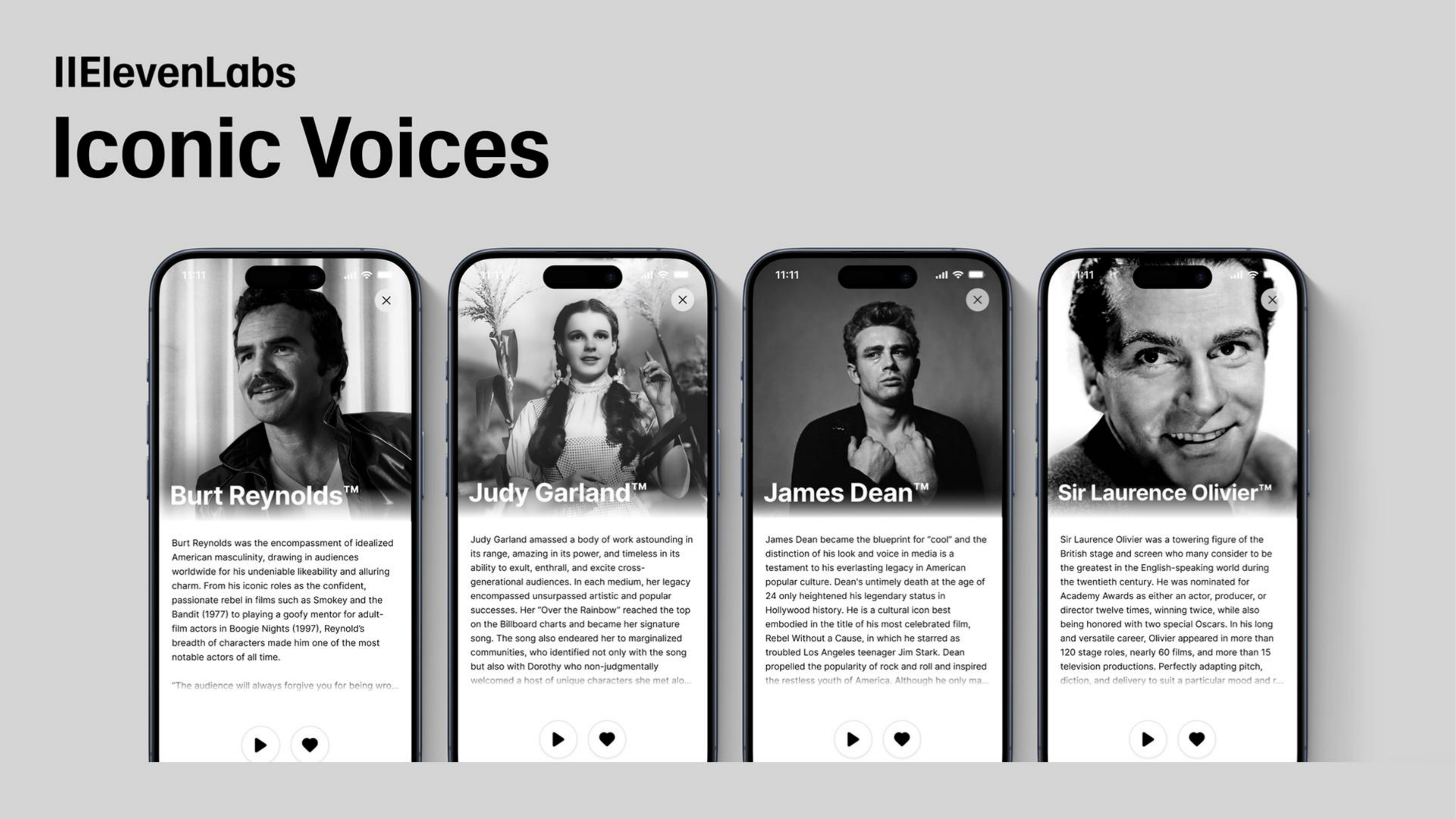
- ElevenLabs is one of the most advanced voice cloning tools: https://elevenlabs.io/

- Many voice cloning features
  - Voice library (use voices others cloned)
  - Instant voice cloning (10 seconds of audio)
  - Professional voice cloning (30 minutes of audio)
  - Voice design (create voice by text prompt)
  - Voice changer (change the voice in an audio file)

Voice Changer



Permission Granted to Voice Clone



IIElevenLabs

Iconic Voices

Burt Reynolds™

Burt Reynolds was the encompassment of idealized American masculinity, drawing in audiences worldwide for his undeniable likeability and alluring charm. From his iconic roles as the confident, passionate rebel in films such as Smokey and the Bandit (1977) to playing a goofy mentor for adult-film actors in Boogie Nights (1997), Reynold's breadth of characters made him one of the most notable actors of all time.

"The audience will always forgive you for being wro...

▶ ♥

Judy Garland™

Judy Garland amassed a body of work astounding in its range, amazing in its power, and timeless in its ability to exult, enthrall, and excite cross-generational audiences. In each medium, her legacy encompassed unsurpassed artistic and popular successes. Her "Over the Rainbow" reached the top on the Billboard charts and became her signature song. The song also endeared her to marginalized communities, who identified not only with the song but also with Dorothy who non-judgmentally welcomed a host of unique characters she met alo...

▶ ♥

James Dean™

James Dean became the blueprint for "cool" and the distinction of his look and voice in media is a testament to his everlasting legacy in American popular culture. Dean's untimely death at the age of 24 only heightened his legendary status in Hollywood history. He is a cultural icon best embodied in the title of his most celebrated film, Rebel Without a Cause, in which he starred as troubled Los Angeles teenager Jim Stark. Dean propelled the popularity of rock and roll and inspired the restless youth of America. Although he only ma...

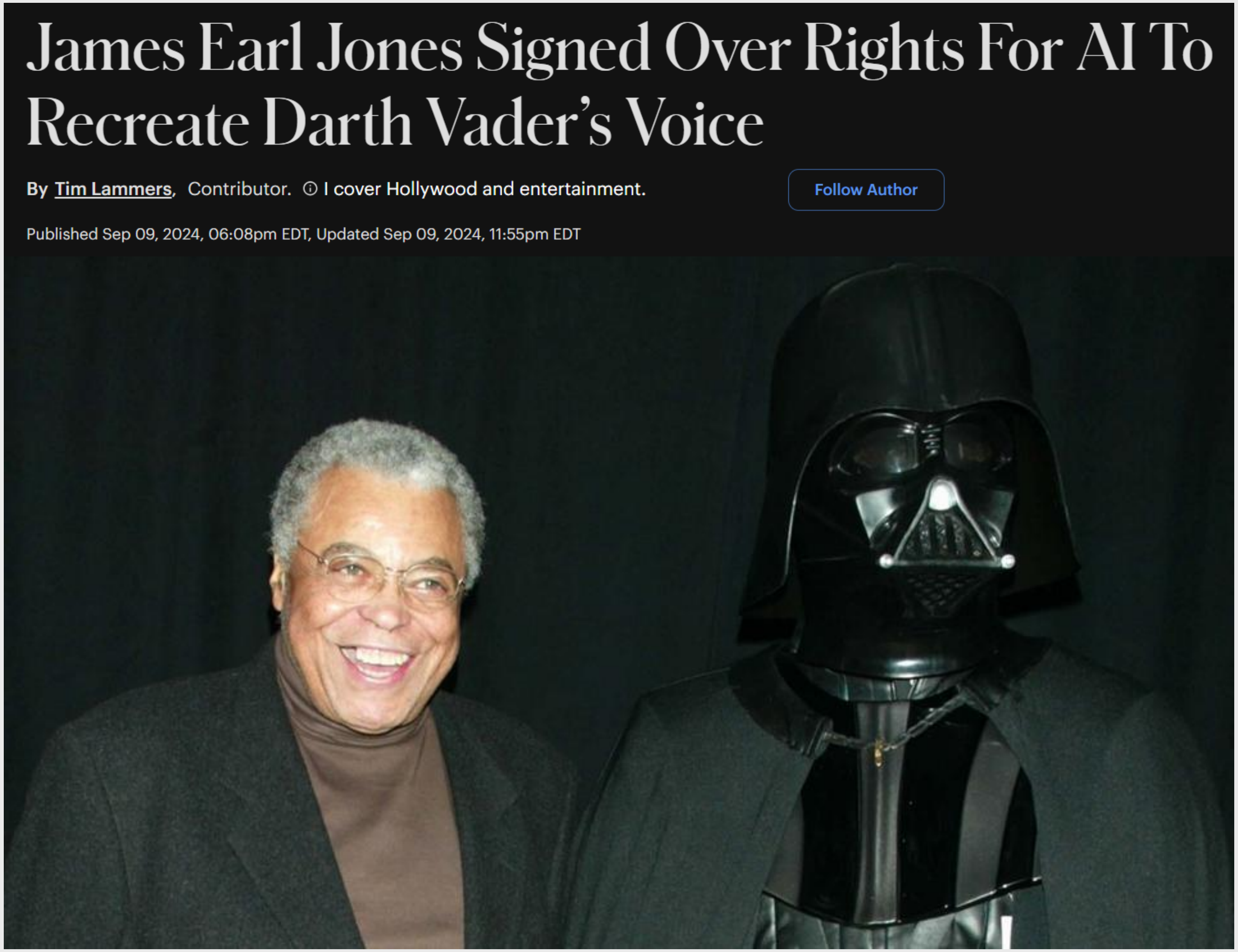
▶ ♥

Sir Laurence Olivier™

Sir Laurence Olivier was a towering figure of the British stage and screen who many consider to be the greatest in the English-speaking world during the twentieth century. He was nominated for Academy Awards as either an actor, producer, or director twelve times, winning twice, while also being honored with two special Oscars. In his long and versatile career, Olivier appeared in more than 120 stage roles, nearly 60 films, and more than 15 television productions. Perfectly adapting pitch, diction, and delivery to suit a particular mood and r...

▶ ♥

Permission Granted to Voice Clone

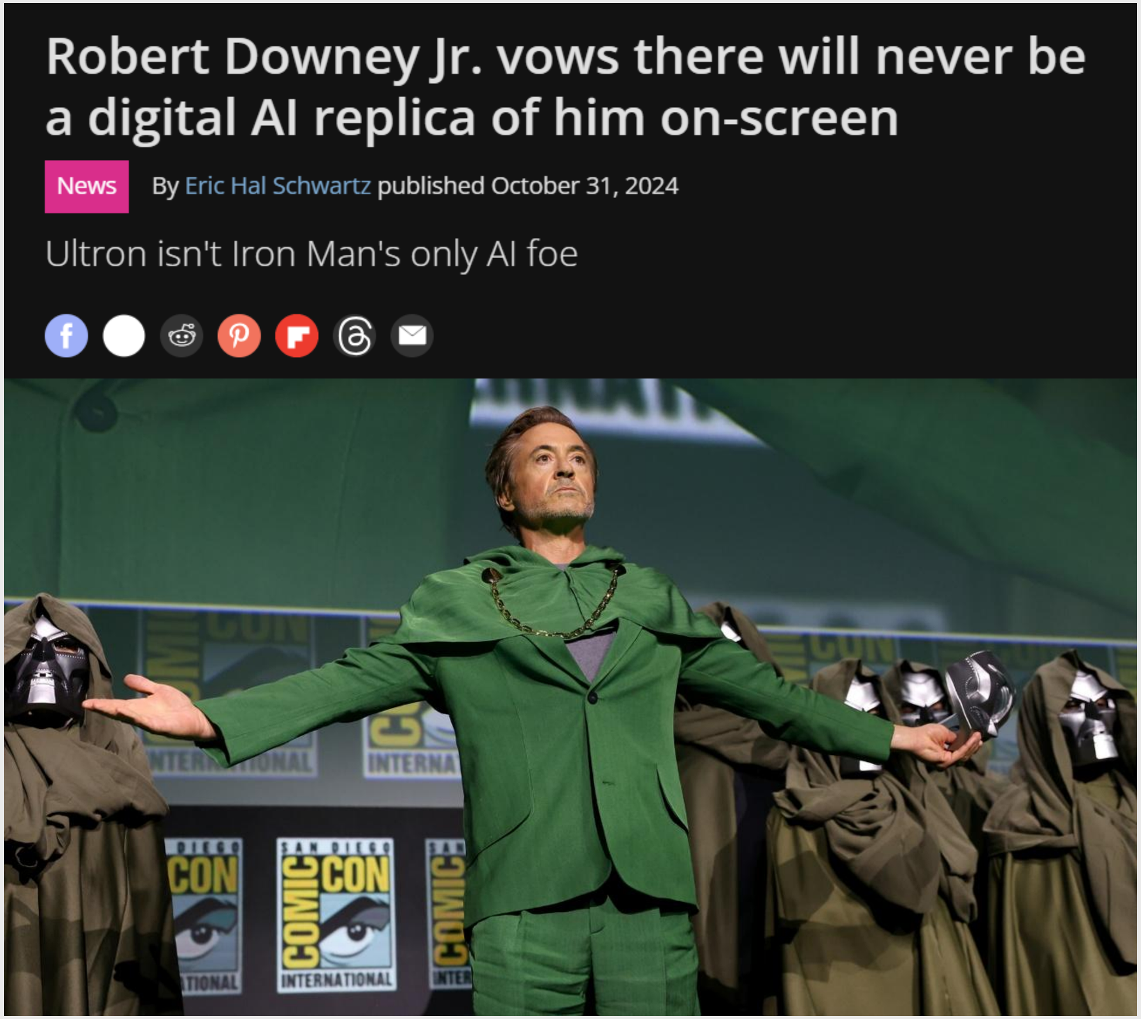


James Earl Jones Signed Over Rights For AI To Recreate Darth Vader's Voice

By Tim Lammers, Contributor. ⓘ I cover Hollywood and entertainment.

Published Sep 09, 2024, 06:08pm EDT, Updated Sep 09, 2024, 11:55pm EDT

Permission Not Granted to Voice Clone



Robert Downey Jr. vows there will never be a digital AI replica of him on-screen

News By Eric Hal Schwartz published October 31, 2024

Ultron isn't Iron Man's only AI foe

Voice Cloning Ethics

- If using a voice clone, be transparent
- Don’t use anyone’s voice without their permission
- Don’t make the voice clone say offensive things
- Don’t create deepfake audio that will cause panic…

How to Clone a Voice

- ElevenLabs allows us to clone voices with only a few seconds of audio

- Can obtain audio files from YouTube
  - – Choose a YouTube video of the subject speaking alone or mostly alone
  - – Download audio of YouTube video (https://tuberipper.net/73/)
  - – Clip the audio to 10 to 30 second clips (https://audiotrimmer.com/)

Video Narrations

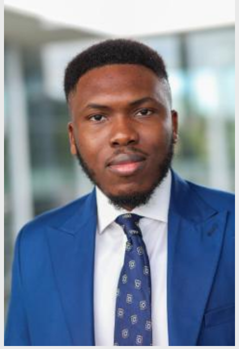
- If we give the frames of a video to AI plus some instructions, it can give us a narration for the video

- We already have methods to gives multiple images and context text to AI and get a text description

- We just need to extract the images from the video
  - We will usually sample around 10-20 images per video
  - This is near the limit of what OpenAI allows

Old HW Problem

- **Make a narration of this video**
- **Bonus – clone a voice and have it narrate the video**



Richard Ogu
Yale MAM
Class of 2024

Coding Session

- We will clone a voice

- We will clone a GitHub repository with some useful code to get us started

- We will build an app that narrates a video using the cloned voice

- AI tools used
  - OpenAI
  - ElevenLabs