

Real-time Voice Cloning in the Age of GenAI

Dr Jackie Tan
Unemployed Bum



Dr Jackie Tan

Ex-academic, 3x serial entrepreneur

Enjoys solving problems

And making ~~money~~ impact

Broad areas of expertise = AI + web3

Broad domains = fintech + DS + edu

Not All Financial Advice Is Created Equal

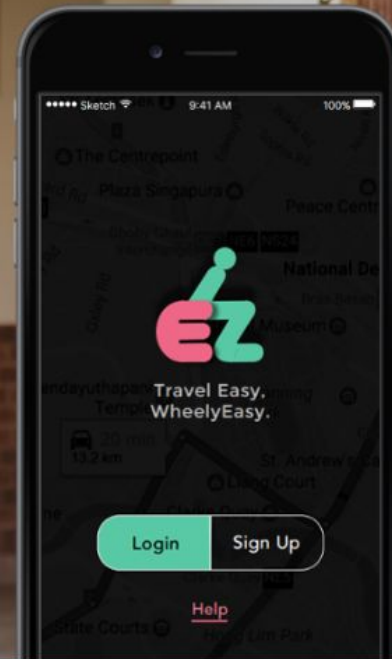
fundMyLife connects your financial planning questions to
the right advisers using our intelligent matching system

[ASK A QUESTION](#)

WheelyEasy

Mobility - made Wheely Easy

Get Notified When It's Released



Discover Wheely-Friendly Routes and Places of Interest



Woofer AI

[BLOG](#)

[LOG IN](#)

[SIGN UP FREE](#)

Turn your audio into blog articles easily

Expand your content writing strategy by going beyond audio transcripts

[SIGN UP FREE](#)

FLAUNT WHAT YOU'VE ACHIEVED

Consolidate all of your certs and NFT badges onto a single coherent builder's profile.

GET YOUR PROFILE





Join our beta

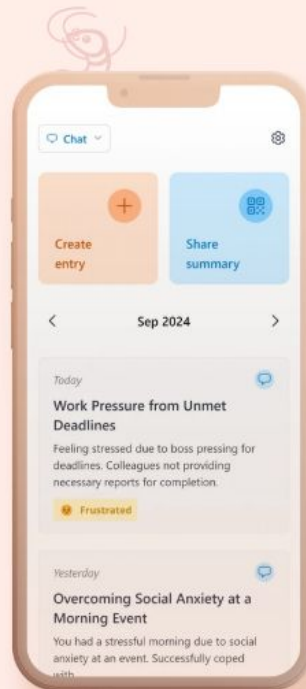
Tell us your story, we'll handle the rest

We're making it easier than ever to share your mental health journey, with seamless journaling and instant summaries that capture your progress effortlessly.

Keep me updated

Read our [Privacy Policy](#) and [Terms of Service](#).

In collaboration with



Let's talk about GenAI...

World / Asia

Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'



By Heather Chen and [Kathleen Magramo](#), CNN

🕒 2 minute read · Published 2:31 AM EST, Sun February 4, 2024

Finance director nearly loses \$670k to scammers using deepfakes to pose as senior execs



 Connect

🚩 SCAMMER ALERT: Hey friends, it has come to my attention that there is someone out there impersonating me on WhatsApp, using my voice (probably AI) claiming I changed my number and asking for some money to "test my paynow account".

THAT IS NOT ME. And if you do know me, you know I'm a giver and not a taker.

The only thing I will reach out for is a good cup of coffee, sometimes lunch, great founders introduction and to support [Bareborn](#).

What is deepfake?

ARTIFICIAL INTELLIGENCE VS MACHINE LEARNING VS DEEP LEARNING

1 Artificial Intelligence

Development of smart systems and machines that can carry out tasks that typically require human intelligence

2 Machine Learning

Creates algorithms that can learn from data and make decisions based on patterns observed

Requires human intervention when it doesn't know what to do

3 Deep Learning

Uses an artificial neural network to reach accurate conclusions without human intervention

Deepfake technology is a form of artificial intelligence that employs machine learning algorithms to

generate realistic media content,

including synthetic audio, images, and text.



VOICE CLONING HISTORY

1968



IBMs Shoebox machine is introduced in San Jose, CA.

1990s



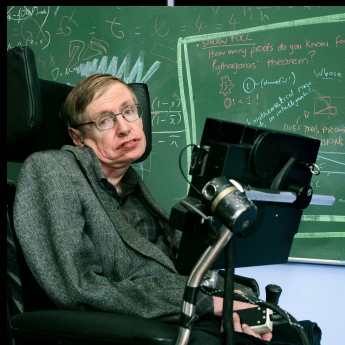
The intro of neural networks and machine learning leads to improvements in synthesized speech.

2017



Tacotron an AI-powered speech synthesis system converts text to speech.

Voice Cloning Historical Timeline



Commercial speech synthesis products released.
iTalk
MacinTalk



2016

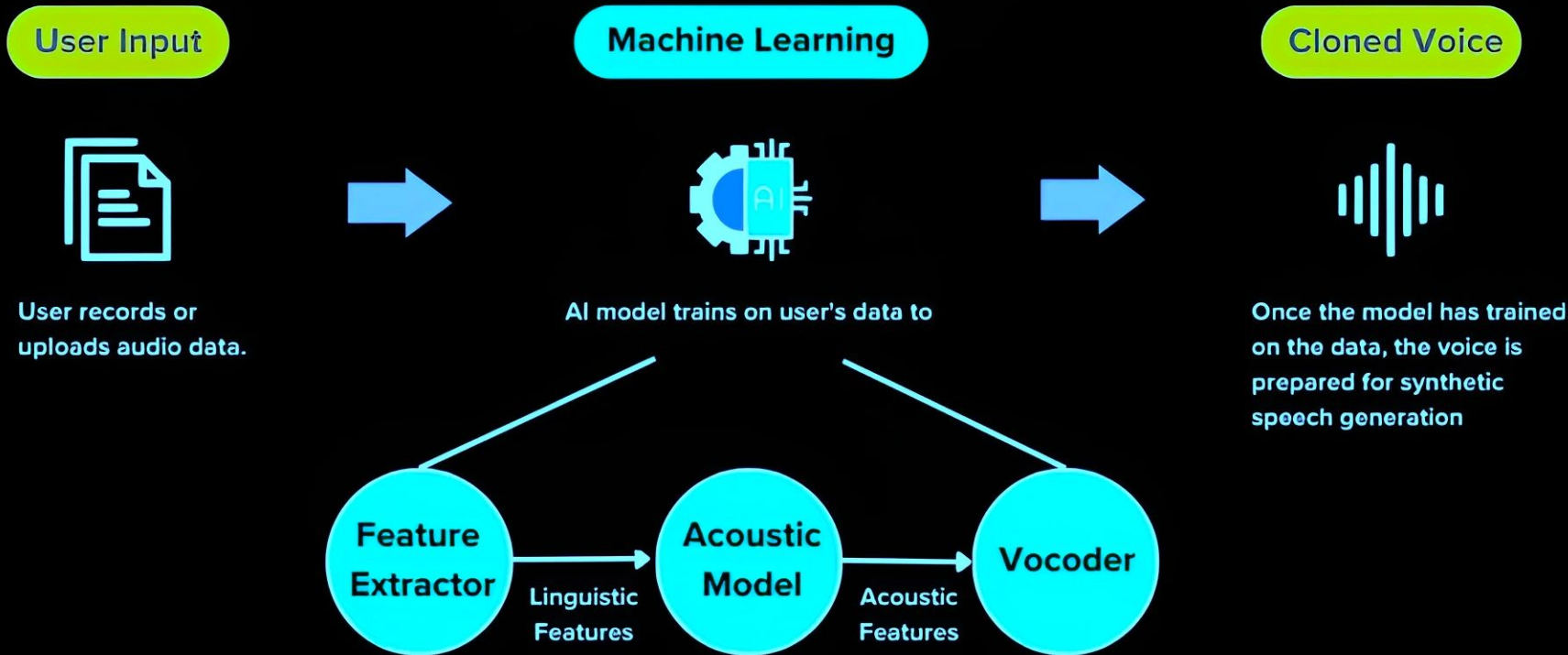
WaveNet, a deep neural network is introduced for generating raw audio.



2021

BERT improves performance on a range of natural language processing tasks

AI VOICE CLONING PROCESS



Live Demo

OpenVoice: Versatile Instant Voice Cloning

Zengyi Qin *
MIT & MyShell.ai

Wenliang Zhao
Tsinghua University

Xumin Yu
Tsinghua University

Xin Sun
MyShell.ai

Abstract

We introduce OpenVoice, a versatile instant voice cloning approach that requires only a short audio clip from the reference speaker to replicate their voice and generate speech in multiple languages. OpenVoice represents a significant advancement in addressing the following open challenges in the field: **1) Flexible Voice Style Control.** OpenVoice enables granular control over voice styles, including emotion, accent, rhythm, pauses, and intonation, in addition to replicating the tone color of the reference speaker. The voice styles are not directly copied from and constrained by the style of the reference speaker. Previous approaches lacked the ability to flexibly manipulate voice styles after cloning. **2) Zero-Shot Cross-Lingual Voice Cloning.** OpenVoice achieves zero-shot cross-lingual voice cloning for languages not included in the massive-speaker training set. Unlike previous approaches, which typically require extensive massive-speaker multi-lingual (MSML) dataset² for all languages, OpenVoice can clone voices into a new language without any massive-speaker training data for that language. OpenVoice is also **computationally efficient**, costing tens of times less than commercially available APIs that offer even inferior performance. To foster further research in the field, we have made the source code³ and trained model publicly accessible. We also provide qualitative results in our demo website⁴. OpenVoice has served more than 2M users worldwide as the voice engine of MyShell.ai.

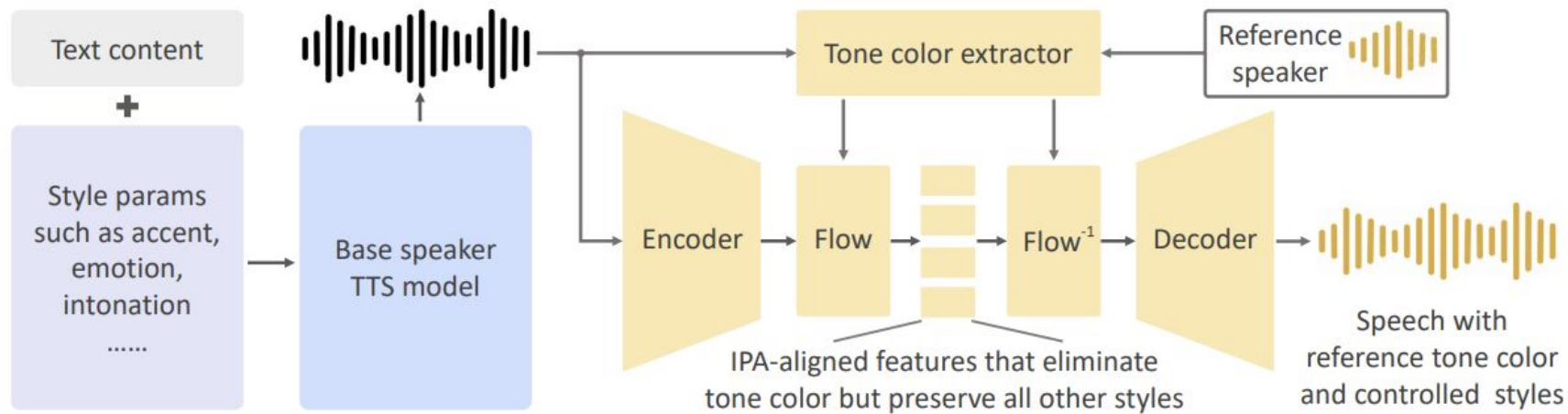


Figure 1: Illustration of the OpenVoice framework. We use a base speaker model to control the styles and languages, and a converter to embody the tone color of the reference speaker into the speech.

Live Demo

(for real)

So what?

Create the most realistic speech with our AI audio platform

Pioneering research in Text to Speech, AI Voice Generator, and more

GET STARTED FREE

CONTACT SALES



TEXT TO SPEECH



SPEECH TO TEXT



CONVERSATIONAL AI



DUBBING



VOICE CLONING



ELEVENREADER

In the ancient land of Eldoria, where the skies were painted with shades of mystic hues and the forests whispered secrets of old, there existed a dragon named Zephyros. Unlike the fearsome tales of dragons that plagued human hearts with terror, Zephyros was a creature of wonder and wisdom, revered by all who knew of his existence.



Brian Narrate a story



Alice Tell a joke



Bill Record an ad



Callum Speak different languages



Charlotte Hear a video game character



Chris Lead a meditation



George Direct a dramatic movie scene

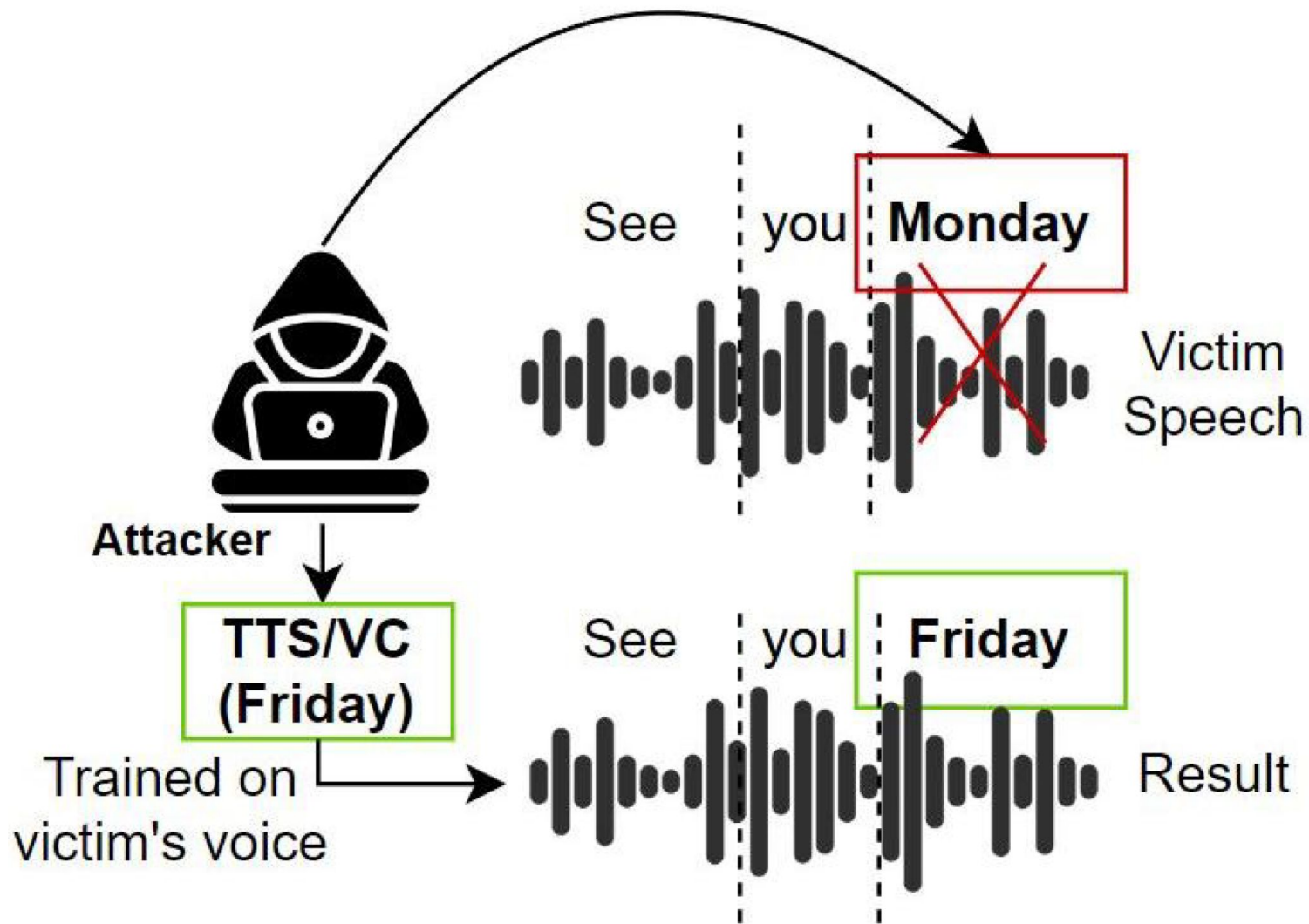
That's true



Live Demo Part II

What's stopping massive fraud at scale for now?

- Real time STT transcription
- Static TTS generation
- Luckily, we don't have real time STTS technology
- ...yet



Are we doomed to be scammed?

- Not really
- Make sure you verify that the other person is legit
- There's still a lot of problems with perfect voice cloning

Thanks

@jackietanyen

