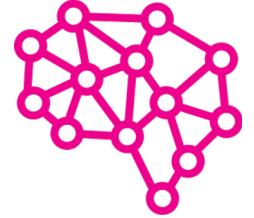




Real-time Multilingual Custom Keyword Spotting

Nasim Alamdari, and Christos Maganas
April 2023



Motivation

Problem

Models like “Alexa” work great, but are not customizable or open-sourced.



Solution

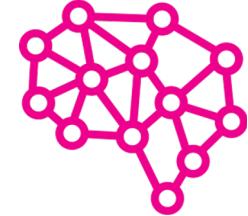
What if there was a way to create a custom keyword for my business/brand like “Hey FourthBrain” which was personalized to your voice **in only a few minutes?**



Applications

- **Customizing activation of voice assistance**
- Keyword spotting in Mental Health Monitoring (e.g Depression)
- Detect keywords/phrases in phone calls or audio recordings

Success Metrics (Features)



Real-time Audio
Processing



Low Processing time
35 milliseconds



High Performance
Based on False Acceptance
& Rejection Rates

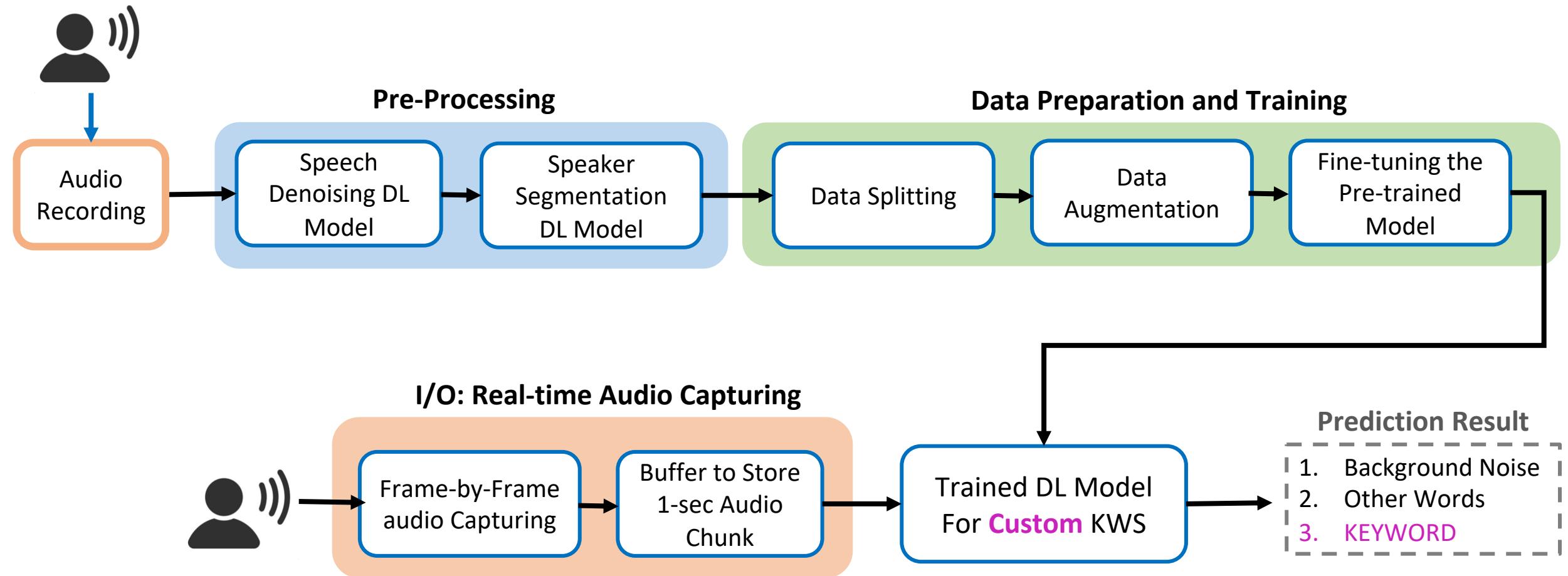
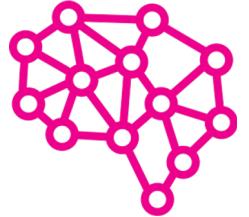


Works on 50 Languages

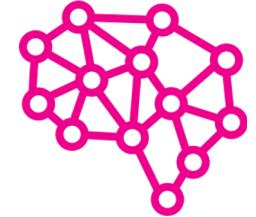


Works in Presence of
Background Noise

ML Model Pipeline



Customization via Few-Shot Learning



Data Augmentation:

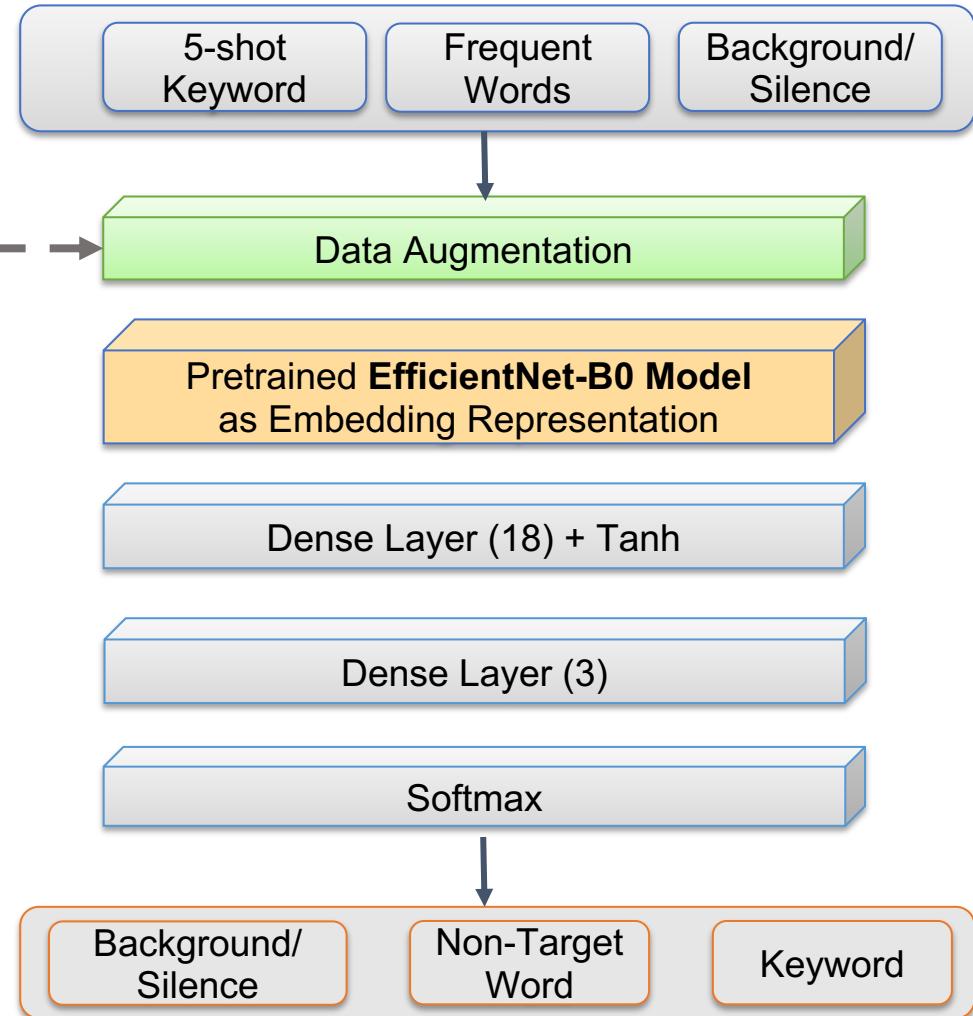
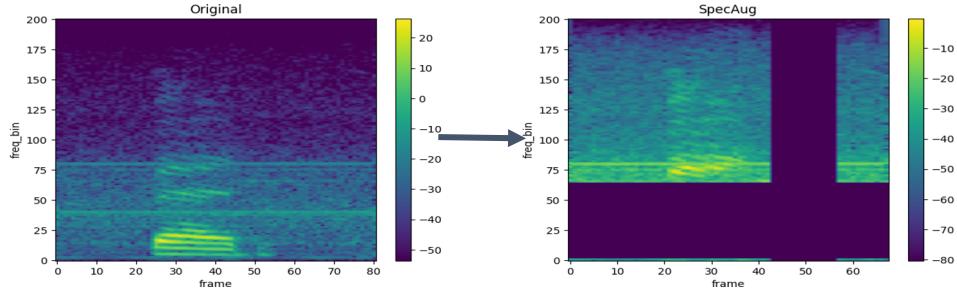
To Increase Train Data From **5-shot** to **1000 Data Samples**

Adding Background Noise

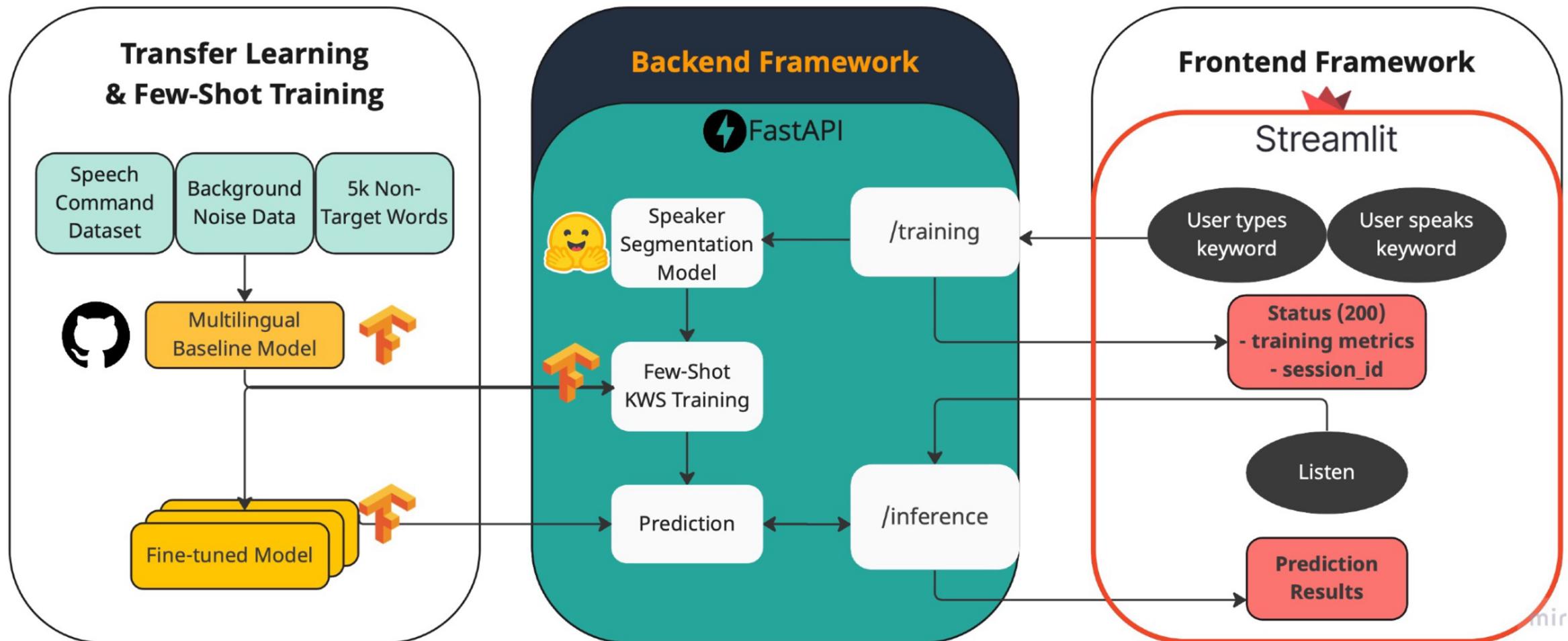
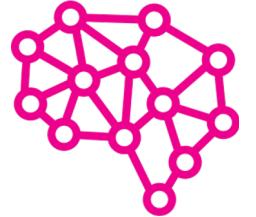
Time Shifting, from -100 ms to 100 ms

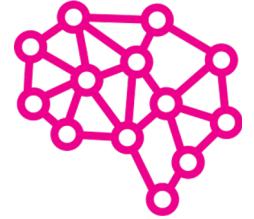
Spectrogram Features
(size 49×40)

SpecAugment:
Time and Frequency
Masking



Solution Architecture





Demo 1: AWS EC2 Deployment

Model Serving via Streamlit app (front-end) and FastAPI (backend)

Multilingual Custom Keyword Spotting App

About this App

Enter Custom Keyword

Step 1: Please Enter Your Custom Keyword with No Spcae Between Words (maximum 2 words)

e.g. heyGPT

The current Custom Keyword is:

Train

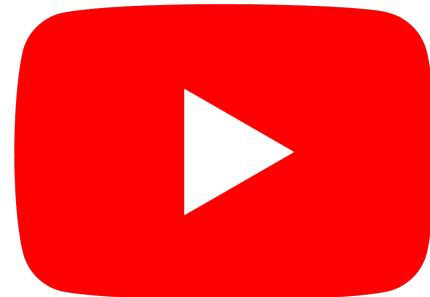
Step 2: Uploading an audio file for training

Upload your (.wav) audio file that has at least 10 times keyword (only) repeated

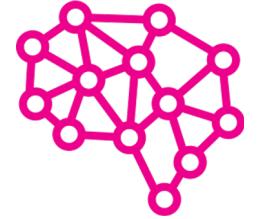
Drag and drop file here
Limit 200MB per file • WAV

Browse files

file is not uploaded yet.



Demo Link: <https://youtu.be/quaSahxFPLQ>



Demo 2: Real-time Streamlit App

Real-Time Multilingual Custom Keyword Spotting App

About this App

Enter Custom Keyword

Please Enter Your Custom Keyword with No Spcae Between Words (maximum 2 words)

e.g. heyGPT

Submit Keyword

Record Audio

Select recording duration (in seconds)

20

15 60

Press "Record" and Repeat Your Custom Keyword for 20 Seconds.

Record 🎧

Running Locally
Demo Link: <https://youtu.be/didUvy0FCV8>

Train

Press "Train" to Fine-Tune the Model

Training may take 2-3 minutes... 🕒

Train ✘

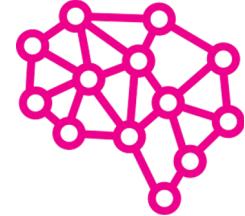
Real-Time Evaluation

Press "Start Inference" to Test the Fine-Tuned Custom Keyword Spotting Model in Real-Time

Stop ⏹

Start Inference ➡

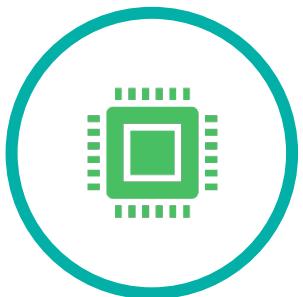
Delivered Solution



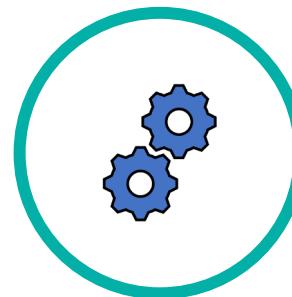
1. End-to-End *Real-time* Multilingual Custom Keyword Spotting



2. Quantitative and qualitative performance of custom keyword detectors.

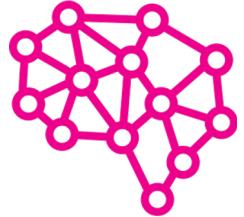


3. Deployed the ML model on AWS EC2.
Model serving through Streamlit (frontend) and FastAPI (backend)



4. Technical report on System setup and performance.

Responsible AI



Protecting User Data

Ensuring the Secure Handling of Recorded and Uploaded Audio during Training and Evaluation



Fair and Unbiased Keyword Spotting

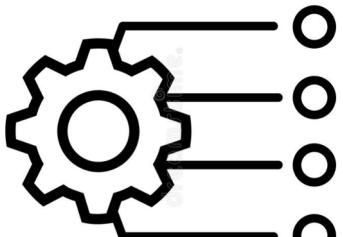
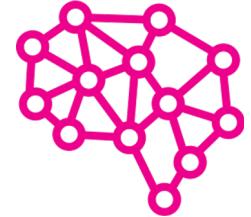
Eliminating Gender, Ethnic, and Dialect Biases in Speaker Recognition



Strategies for Improper Customization

Future Work

Future Work

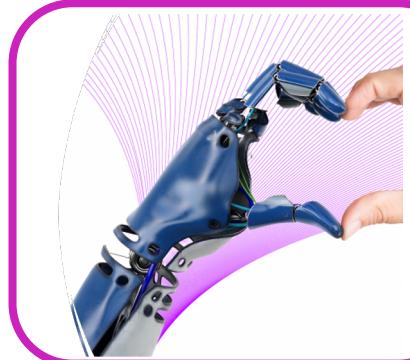


I/O:
Multi-threading
Multi-processing

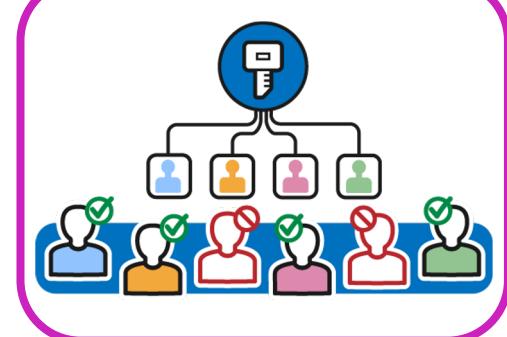


Improve Model to
Various Speech
Distortion Scenarios:

1. Echo Removal
2. Double Talker Detection
3. Speaker Diarization



Responsible AI



Handling
Simultaneous Usage
of App on the Cloud
by Multiple Users



Embedding the
Model with a
Voice Assistance
on Edge Devices

Shout out

Thank you to **all of our
instructors**



A big thank you to **FourthBrain**



Finally, we want to express our
gratitude to **MLE11 participants**





**Thank you!
Questions?**