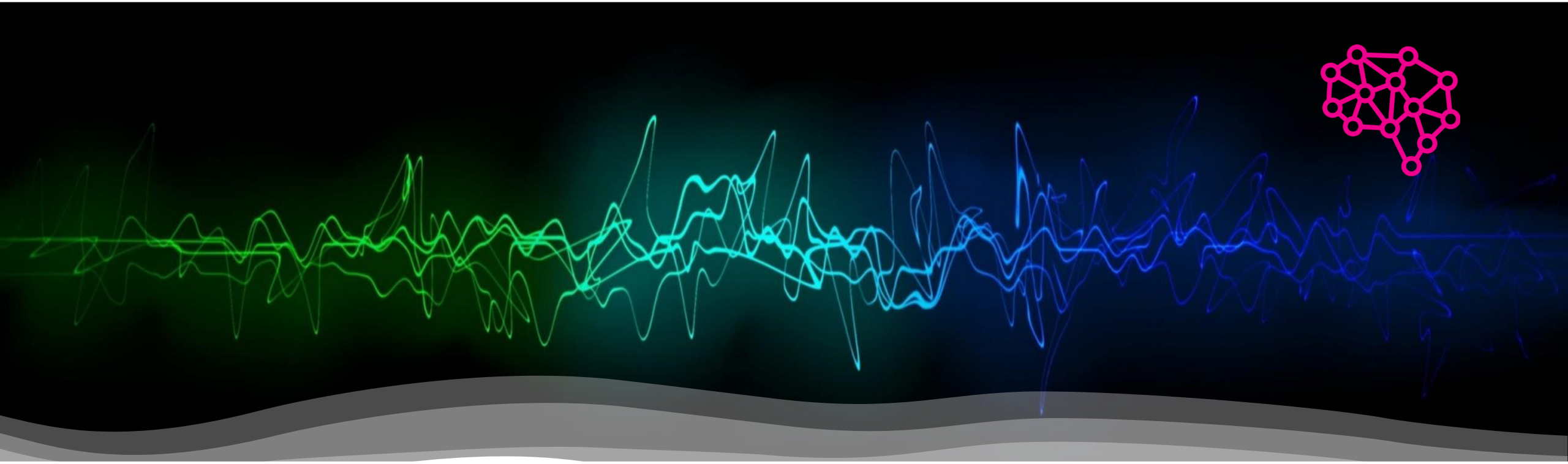


Real-time Multilingual Custom Keyword Spotting

Nasim Alamdari, and Christos Magganas

April 2023

Motivation

Problem

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

Solution

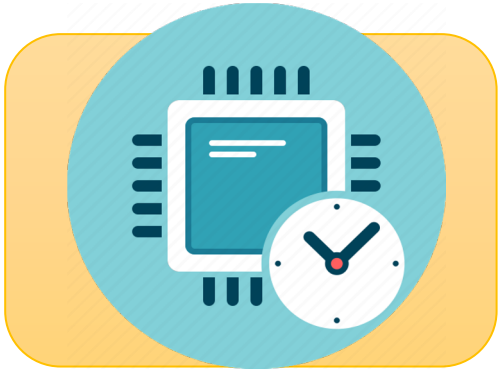
What if there was a way to create a custom keyword for my business/brand like “Hey FourthBrain” which was personalized to your voice in only a few minutes?

Applications

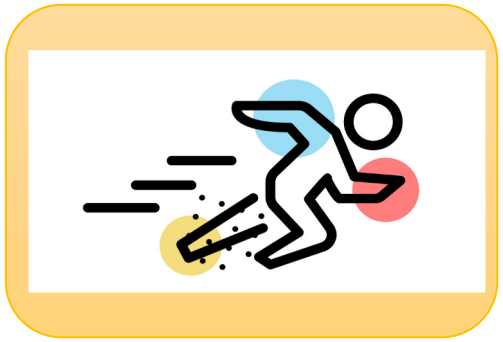
- **Customizing activation of voice assistance**
- Keyword spotting in Mental Health Monitoring (e.g Depression)
- Detect keywords/phrases in phone calls or audio recordings

Success Metrics (Features)

Real-time Audio Processing



Low Processing time
35 milliseconds

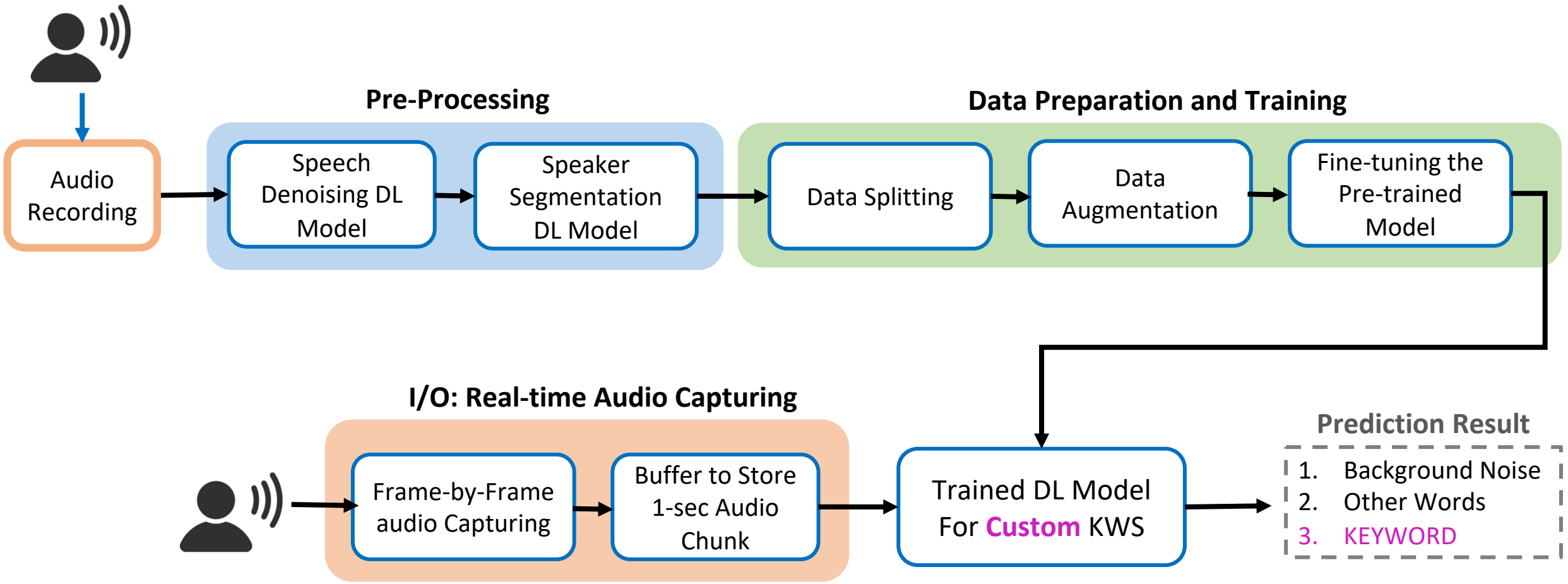


High Performance
Based on False Acceptance & Rejection Rates

Works on 50 Languages

Works in Presence of Background Noise

ML Model Pipeline



Customization via Few-Shot Learning

Data Augmentation:

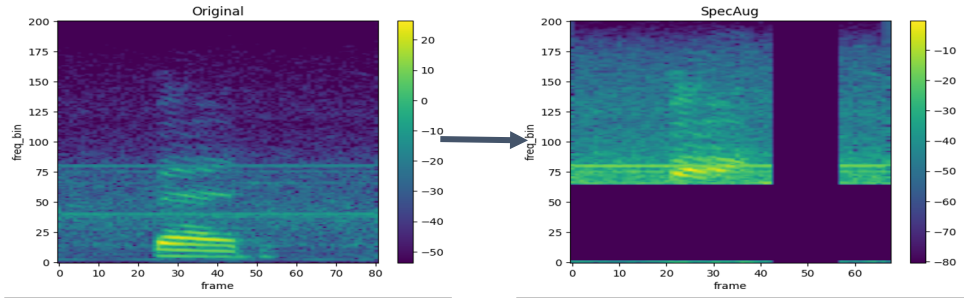
To Increase Train Data From **5-shot to <u>1000 Data Samples</u>**

Adding Background Noise

Time Shifting, from -100 ms to 100 ms

Spectrogram Features (size 49×40)

SpecAugment: Time and Frequency Masking



5-shot Keyword | Frequent Words | Background/ Silence

Data Augmentation

Pretrained **EfficientNet-B0 Model** as Embedding Representation

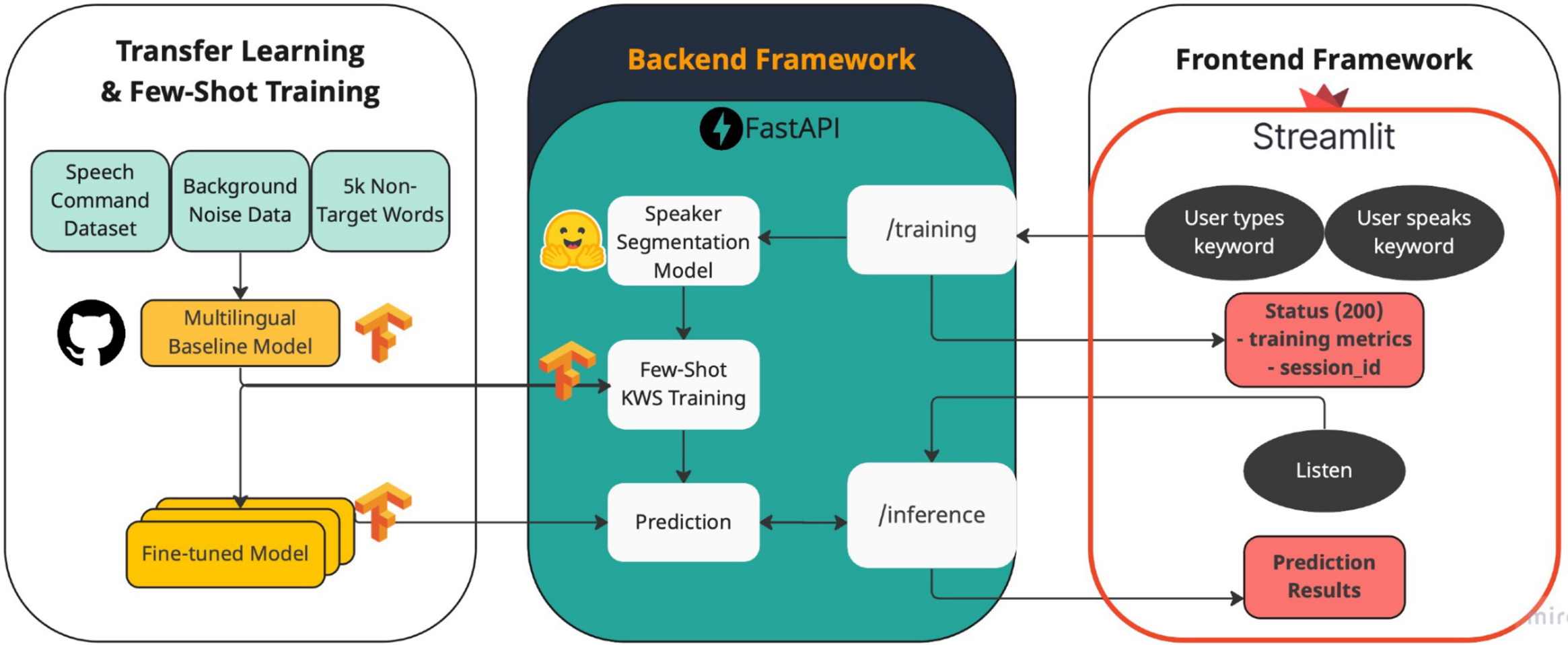
Dense Layer (18) + Tanh

Dense Layer (3)

Softmax

Background/ Silence | Non-Target Word | Keyword

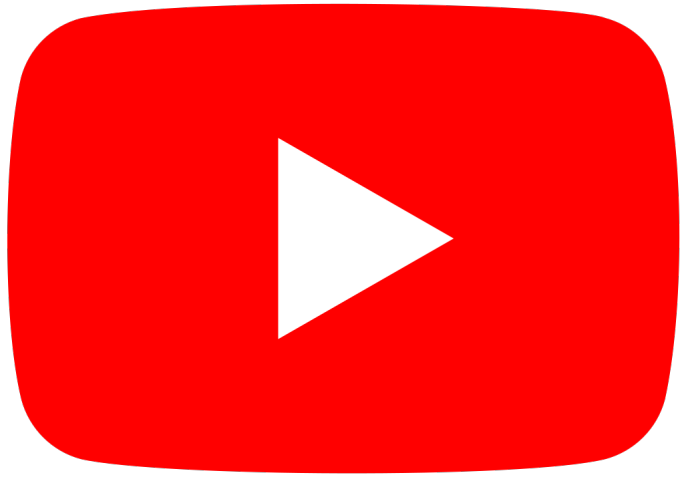
Solution Architecture



Demo 1: AWS EC2 Deployment

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

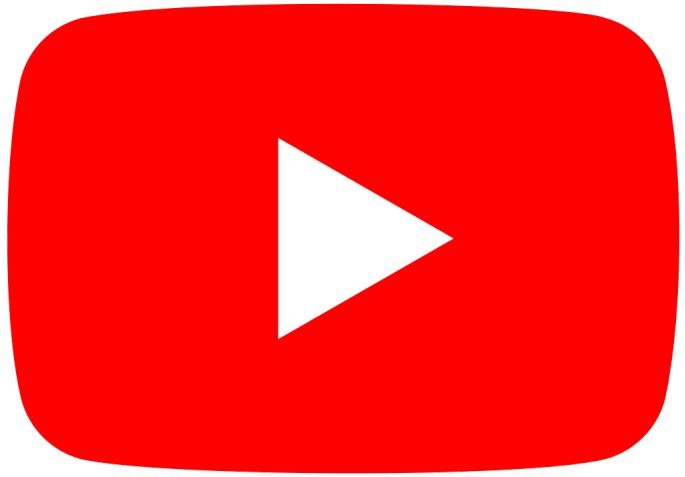
Demo Link



Demo 2: Real-time Streamlit App

Running Locally

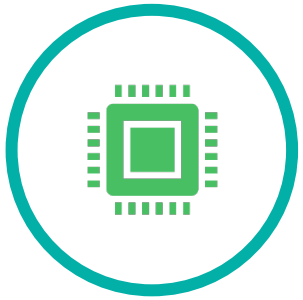
Demo Link



Delivered Solution

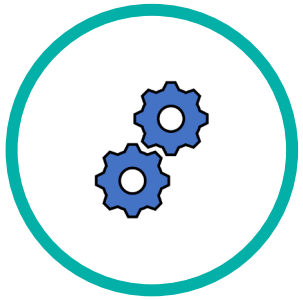
1. End-to-End Real-time Multilingual Custom Keyword Spotting

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.

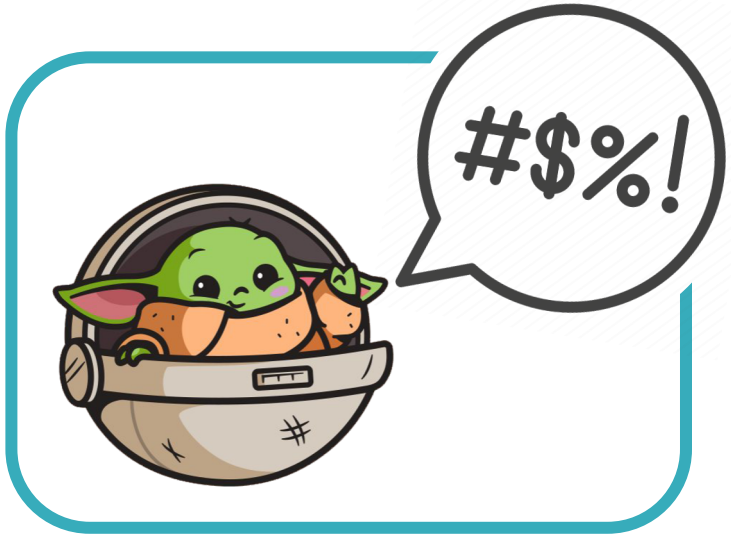
Responsible AI

Protecting User Data

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

Fair and Unbiased Keyword Spotting

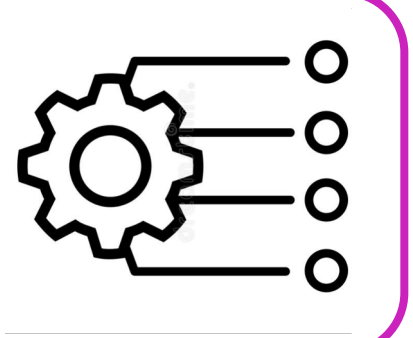
Eliminating Gender, Ethnic, and Dialect Biases in Speaker Recognition



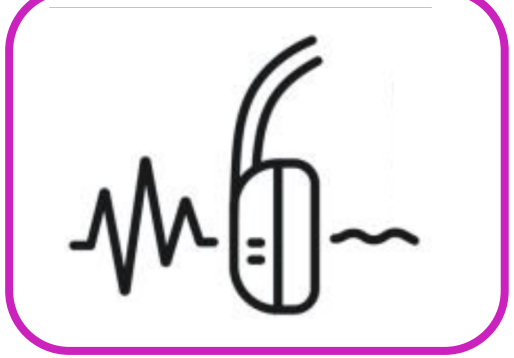
Strategies for Improper Customization

Future Work

Future Work

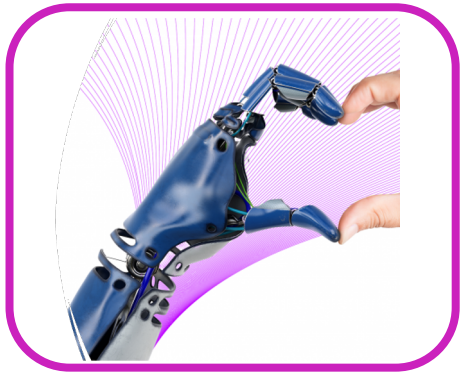


I/O:
Multi-threading
Multi-processing

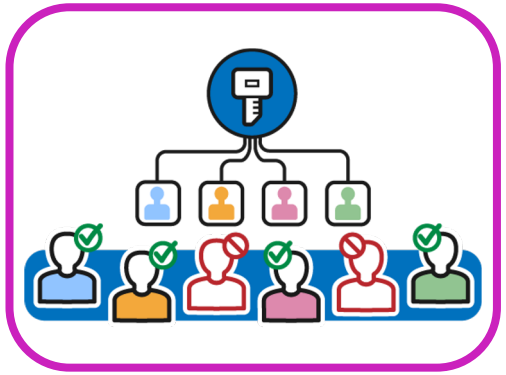


Improve Model to Various Speech Distortion Scenarios:

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?