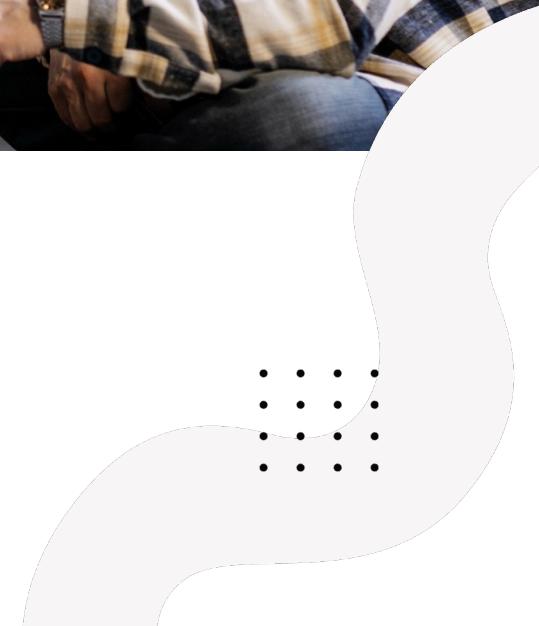




ALDI TALK

ADAPTIVE LANGUAGE/DIALECT INTERPRETER FOR ELDERLY

Improving Communication in Care Facilities

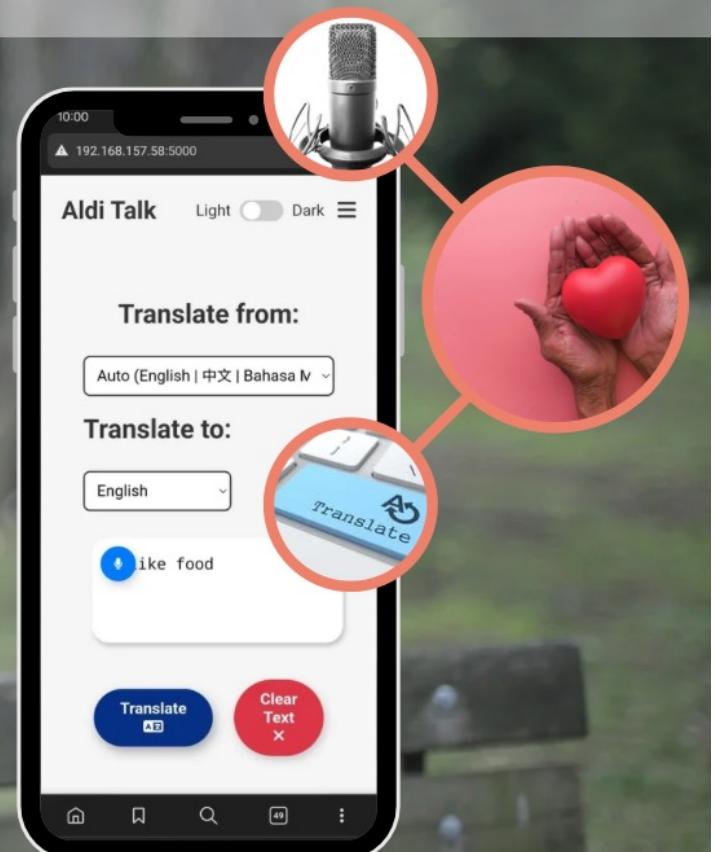


ALDI TALK



ALEX, HABIB, JINBIN

ALDI TALK



ALEX, HABIB, JINBIN

ALDI TALK

KEY FEATURES



One-click Translation

Detects spoken dialects and translates them automatically



Real-Time Translation

Instantly converts speech or text between dialects and English



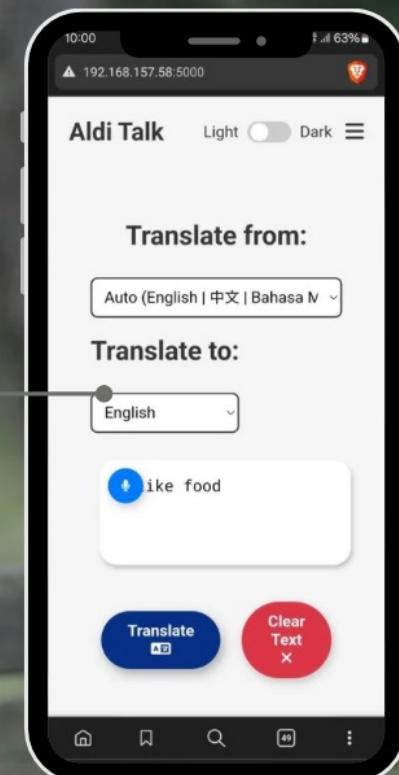
Specialised Vocabulary

Focuses on care-specific and medical terms for better accuracy

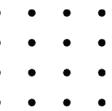


Simple Interface

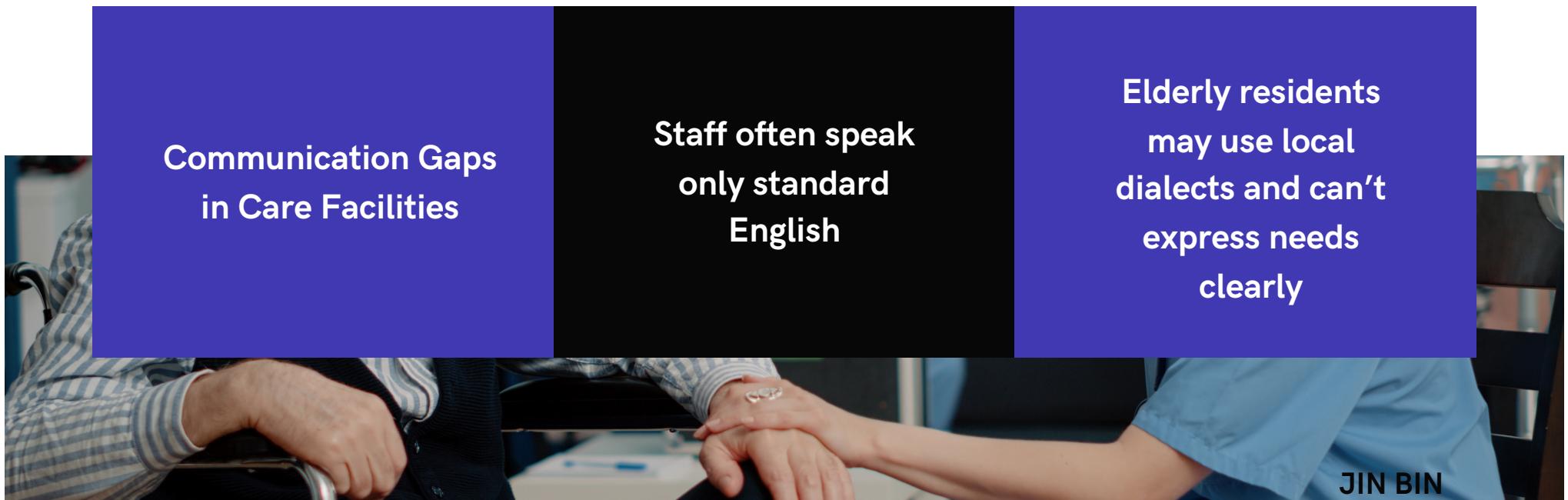
Large buttons, clear fonts, and easy navigation for elderly users.

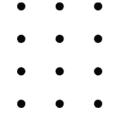
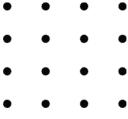


ALEX, HABIB, JINBIN



WHAT'S WRONG?





WHAT IS ALDI TALK?

Solution

A Singapore-focused dialect translator app

Goal

Quick, accurate, and user-friendly translation

Impact

Reduced stress, improved care, and better resident-staff relationships

JIN BIN

PROJECT SOLUTION APPROACH



Supported languages

English
Mandarin
Malay
Tamil

Cantonese
Hokkien

Factors to look out for:

- Local dialects
- Languages with limited resources
- Elderly user friendliness



DATA COLLECTION

1 **Data source:** Singapore Hokkien Dramas on YouTube

2 **Frame sampling:** segmented each video into frames and extract suitable frames



3 **Optical Character Recognition:** extracted the subtitles from each frame

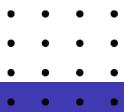
What's the scale?

$25\text{fps} * 60\text{s} * 40\text{min}$
= 60,000 frames per ep

60,000 frames * 42eps
= **3,120,000 frames**

MIN

DATA COLLECTION



4 Matching the English and Chinese transcripts with Frames

frame_name	transcript
0 frame_10020.jpg	你不要过来 我现在就跳下去
1 frame_10020.jpg	Don't come near me! I will jump down if you do!
2 frame_10050.jpg	你不要过来 我现在就跳下去
3 frame_10050.jpg	Don't come near me! I will jump down if you do!
4 frame_10080.jpg	我求你

5 Split into matching audio segments

frame_name	transcript_en	start	end	transcript_zh
0 frame_1950.jpg	There are days where I would wonder why I was ...	74.6	79.0	我连一粒米的贡献都没有
1 frame_1980.jpg	I really was a rice-waster	78.2	80.2	我真的是浪费米
2 frame_2040.jpg	always seemed to fail at everything ...	80.6	82.6	做什么都失败
3 frame_2130.jpg	brother ...	84.2	86.2	哥
4 frame_2160.jpg	Wake up!	85.4	87.4	喂 醒醒

Start: 2.6s | End: 5.8s
English: Let's get you cleaned up
Chinese: 来帮你擦干净

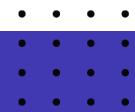
Start: 9.8s | End: 13.0s
English: Says who?
Chinese: 谁说的

Start: 12.2s | End: 15.4s
English: Don't you know
Chinese: 你没听说过吗

MIN

DATA COLLECTION

What's really the scale???



Ho Seh Bo/Ho Seh Bo 2

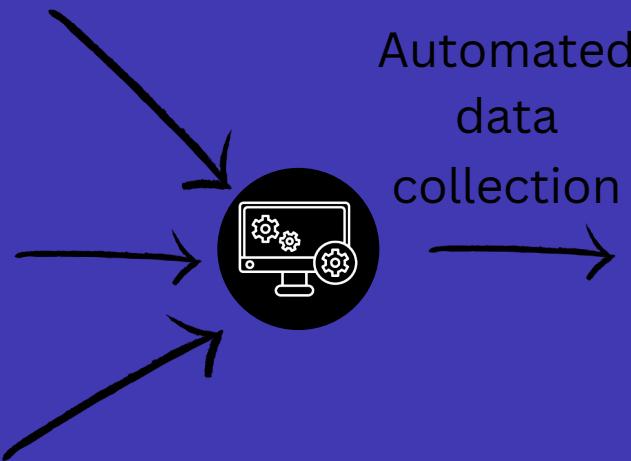
- 16 Episodes each
- ~ 736 minutes each

I Want To Be A Tokay

- 20 Episodes
- ~ 880 minutes

吃饱没 (S1-S4)

- 41 Episodes
- ~ 820 minutes



21 hours of training data

MIN

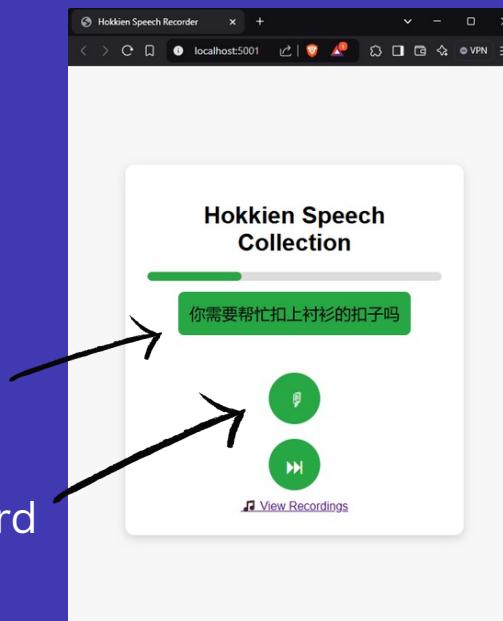
DATA COLLECTION 2

1 Data source:

Speech data from friends,
family and lecturers

2 Phrase to
read

3 Click to record



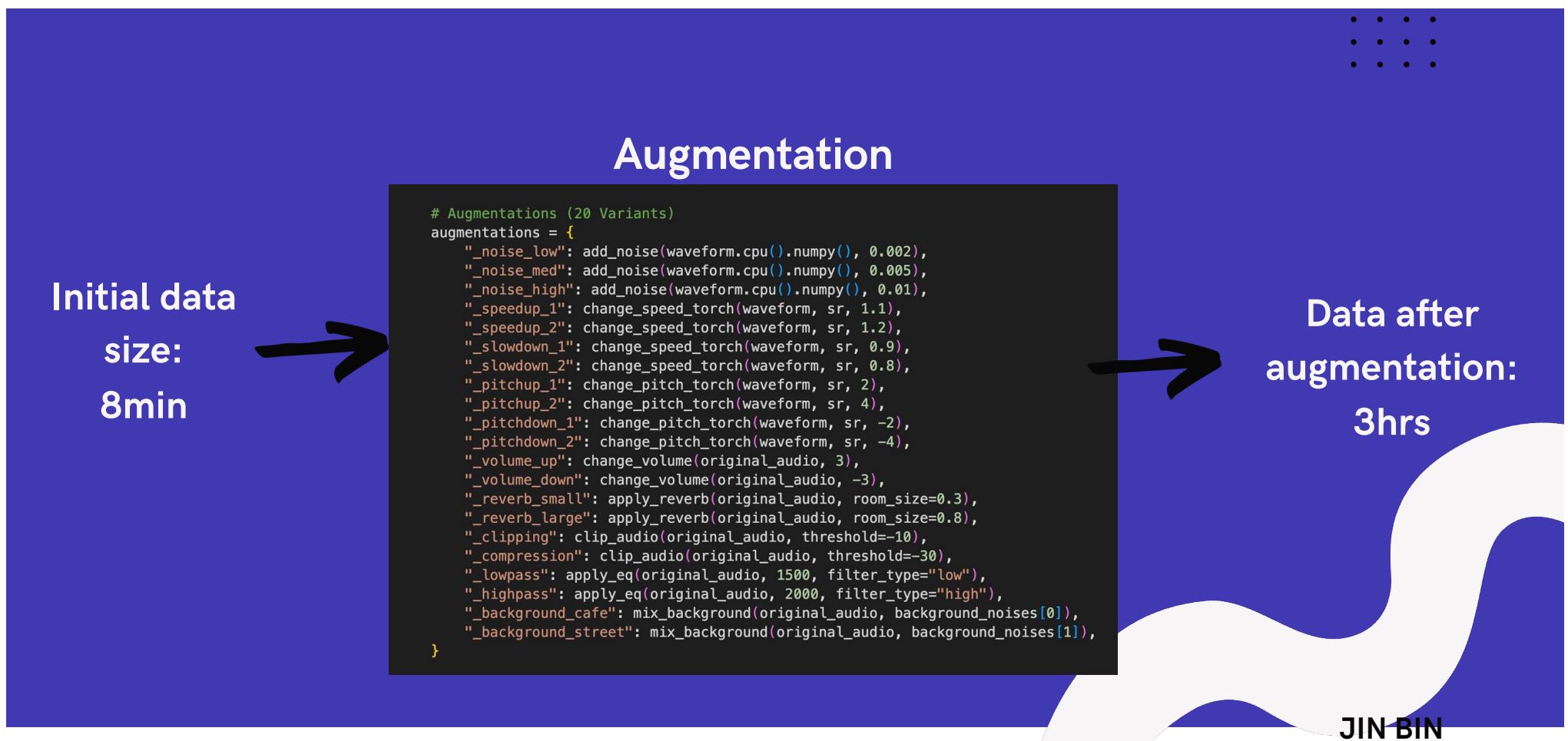
The screenshot shows a window titled 'Recorded Phrases'. It displays a table of recorded phrases with columns for Timestamp, Phrase, Audio, and Action. Each row includes a play button, a back button, and a delete button. The table data is as follows:

Timestamp	Phrase	Audio	Action
05/02/2025, 21:09:31	让我帮你起床	▶ 0:00 ←	>Delete
05/02/2025, 21:09:17	如果你听到声音请告诉我	▶ 0:00 ←	>Delete
05/02/2025, 21:08:56	你需要帮忙洗澡吗	▶ 0:00 ←	>Delete
05/02/2025, 21:08:42	请小心坐下	▶ 0:00 ←	>Delete
05/02/2025, 21:08:23	请深呼吸	▶ 0:00 ←	>Delete

4 Recordings

JIN BIN

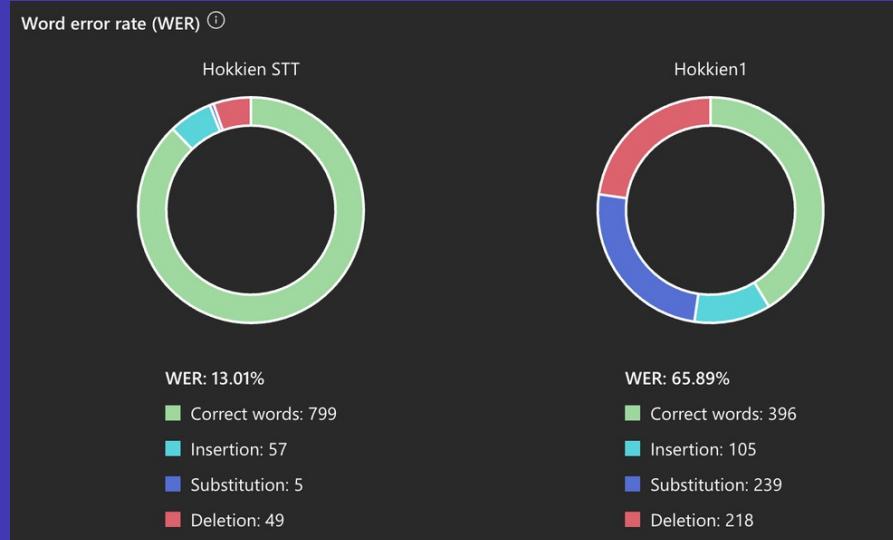
DATA COLLECTION 2



HOKKIEN STT

Azure Custom
Speech Model

Training
6 hours total
\$10/hr



JIN BIN

API INTEGRATION

Azure Translation Service

Provides language translation for text input.

Custom Azure OpenAI Integration

Enables advanced text generation for specialized tasks like Hokkien translation

Azure Speech Service

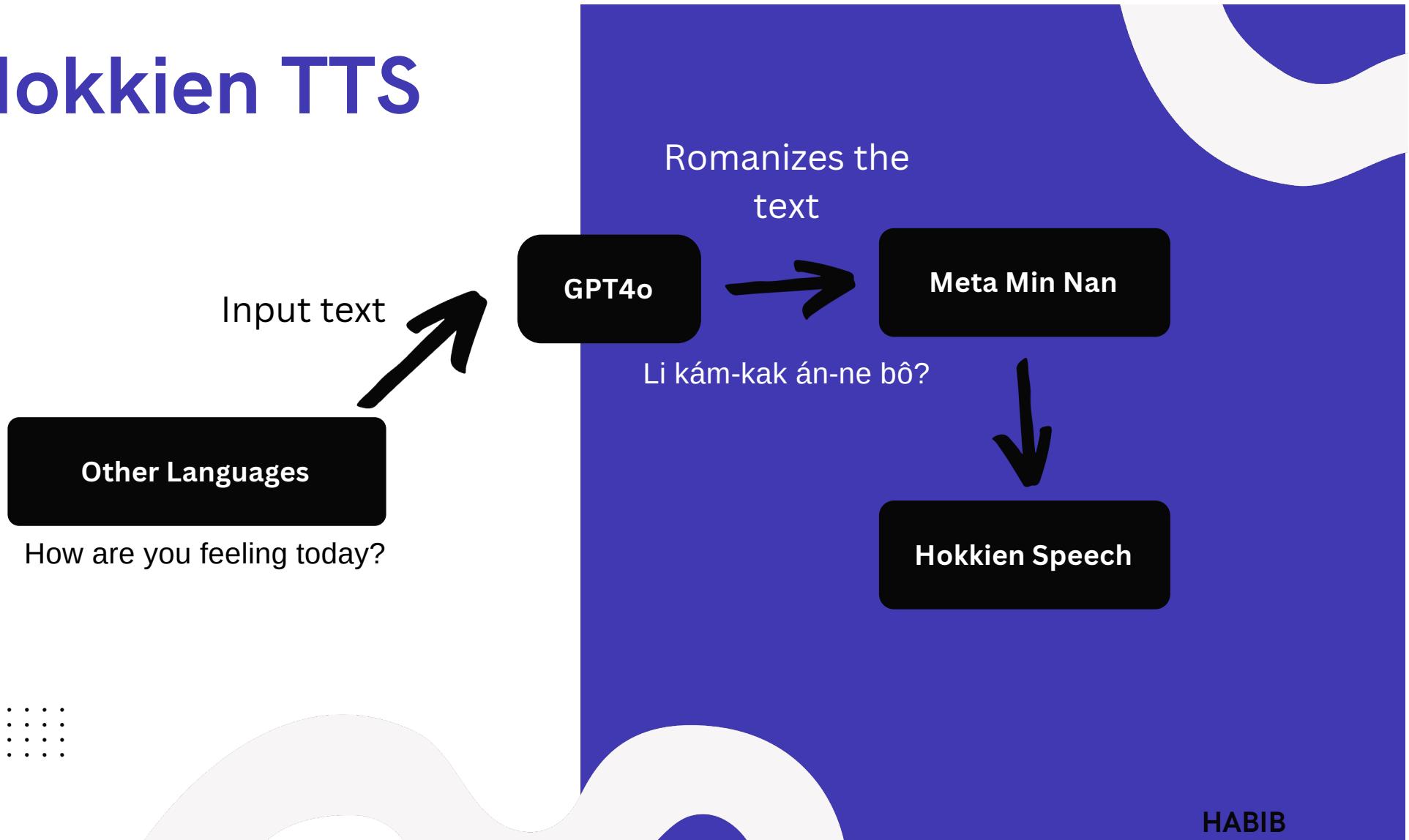
Provides both Speech-to-Text (STT) and Text-to-Speech (TTS) functionalities.

Meta API Integration

generates speech audio from the Pēh-ōe-jī (POJ) transliteration obtained from the Azure OpenAI output

HABIB

Hokkien TTS



APPLICATION DEPLOYMENT

Flask Deployment

The app is deployed using Flask, which serves HTML templates and static assets from designated folders.

Backend Integration

Backend integration is achieved by exposing RESTful endpoints in app.py that invoke functions from AldiTalk/alditalk.py to process requests and return JSON responses.

Config and Security

The application securely loads configuration and API credentials from a .env file to integrate with external services like Azure Cognitive Services.

ALEX

CHALLANGES AND LIMITATIONS

No Hokkien

Azure does not include Hokkien as one of the available languages

Data

Lack of readily available data for Hokkien speech/text

Lack of money and features

OCR and training costs too much, and student subscription has limited capabilities

Lack of Sufficient Outreach

Little research or model development available to the public often requiring more innovation

ALEX

Aldi Talk is ETHICAL & IMPACTFUL



MIN

POTENTIAL IMPROVEMENTS

Support more languages

Make the model more diverse and adaptable to lesser known dialects and languages. Model may also be trained to take responses from people with thick accents

Improve Hokkien model

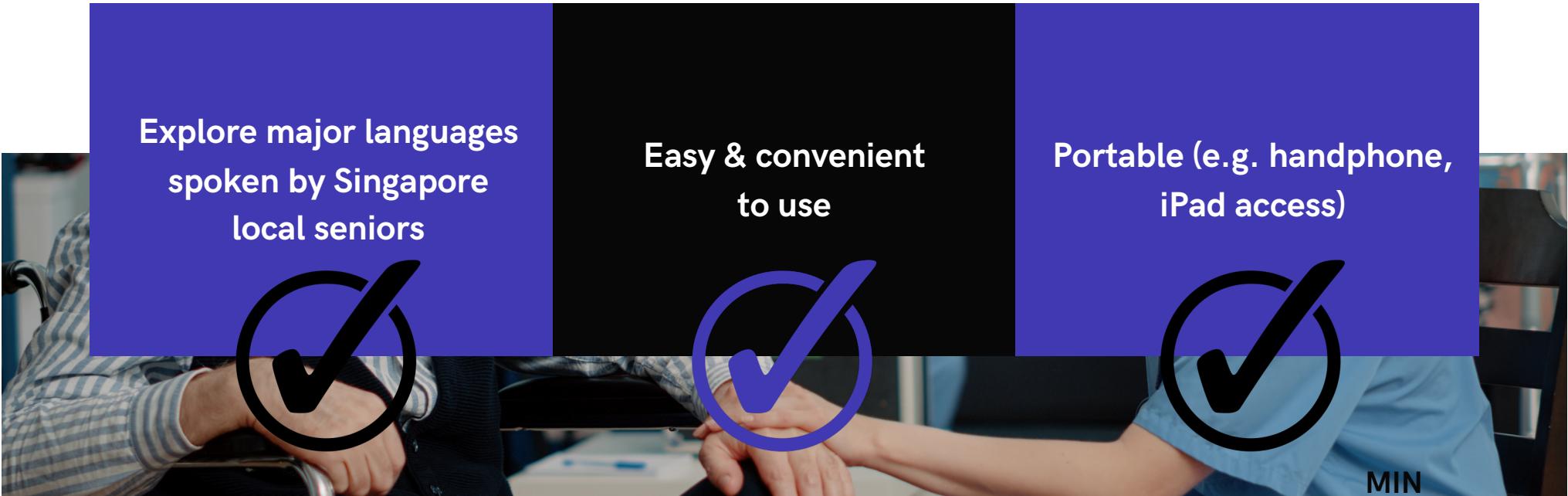
Seek out more diverse data from different sources like Hokkien based organisations to improve the model

More context awareness

Fine-tune every common language with caregiving scenarios

MIN

OBJECTIVES



Explore major languages
spoken by Singapore
local seniors



Easy & convenient
to use



Portable (e.g. handphone,
iPad access)



MIN



THANK YOU!
Q&A TIME!

MODEL ARCHITECTURE

Azure Supported Languages

English, Mandarin, Tamil, Malay, Hindi, Cantonese, ...

- Azure Speech - STT
- Azure Speech - Translation
- Azure Speech - TTS

Azure Unsupported Languages

Hokkien

- Data Collection
 - Azure Vision OCR
- Hokkien STT
 - Azure Custom Speech
- Hokkien Pèh-ōe-jí Translation
 - Azure OpenAI - GPT4o
- Hokkien TTS
 - Meta [Hugging Face] - mms-tts-nan

Languages

Data Enginnering

- Python

Backend

- PythonFlask

- JavaScript

- System Prompts

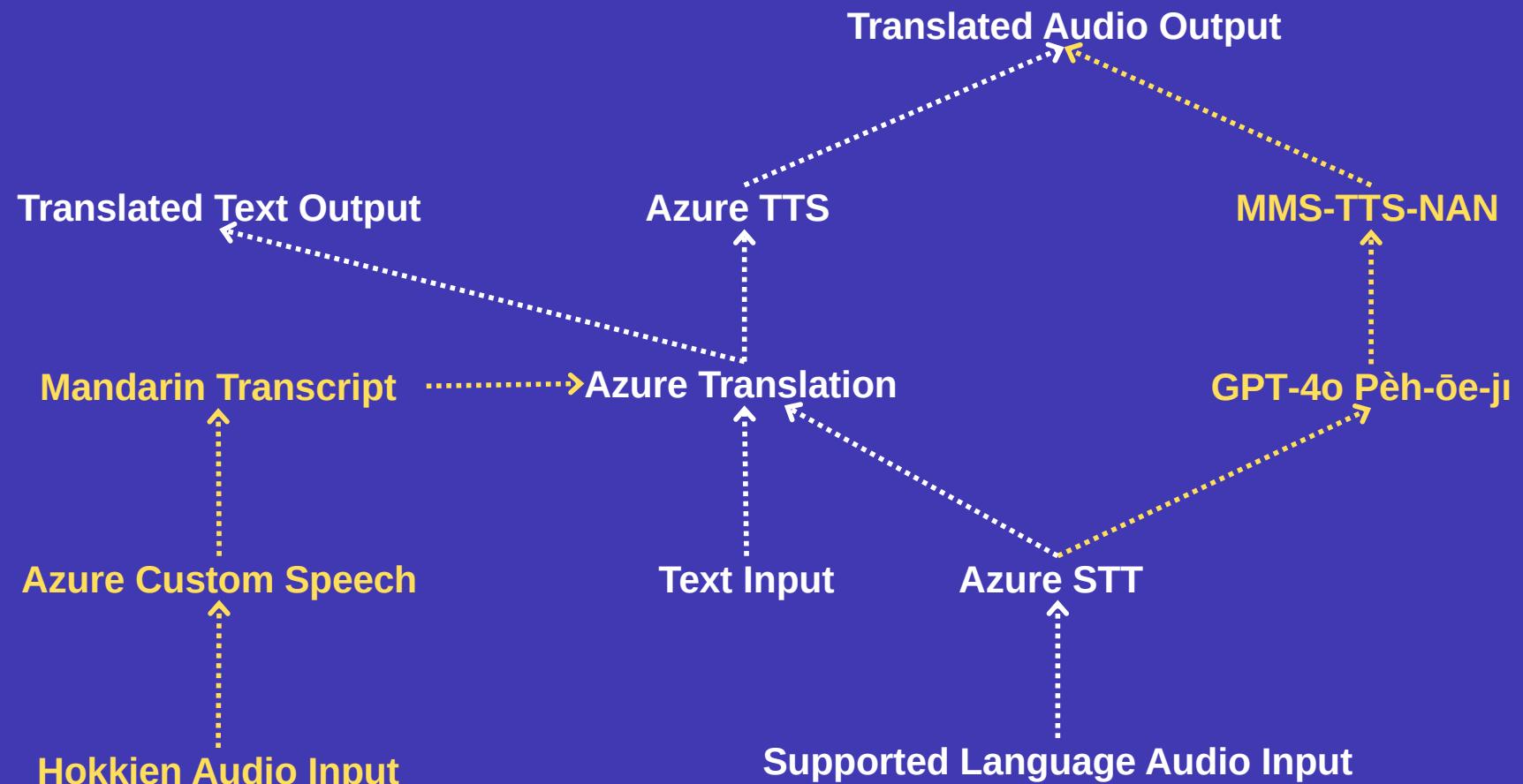
Frontend

- HTML

- CSS

- JavaScript

MODEL ARCHITECTURE



CONTRIBUTIONS

Names ordered in terms of contributions.

Algorithm for YouTube
Hokkien Audio Collection \longleftrightarrow Min, [Alex, Jin Bin, Habib]

JS for in-person Hokkien
Audio collection \longleftrightarrow Min, Alex, Habib, Jin Bin

Sourcing of Hokkien Data \longleftrightarrow Min, Alex, Habib, Jin Bin

HTML & CSS \longleftrightarrow Jin Bin, Alex

UI \longleftrightarrow Jin Bin, Alex, Habib,

Flas \longleftrightarrow Alex, Habib, Min, Min, Jin

API Integration^k \longleftrightarrow Habib, Min, Bin

Custom Model Training \longleftrightarrow Mi

Slides & Report \longleftrightarrow nAlex, Lim Jin Bin, Habib, Min