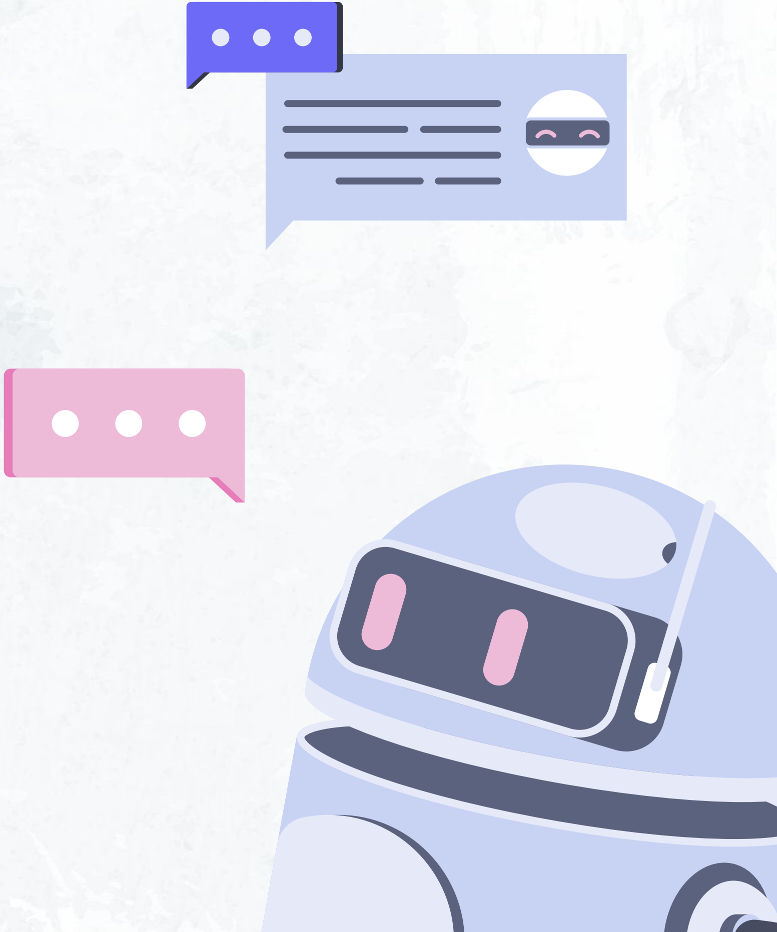
Multi Speaker Transcription



Group2 KJYIHUAYYDS

Table of contents

01 → Purpose

02 → Our solution-Back end

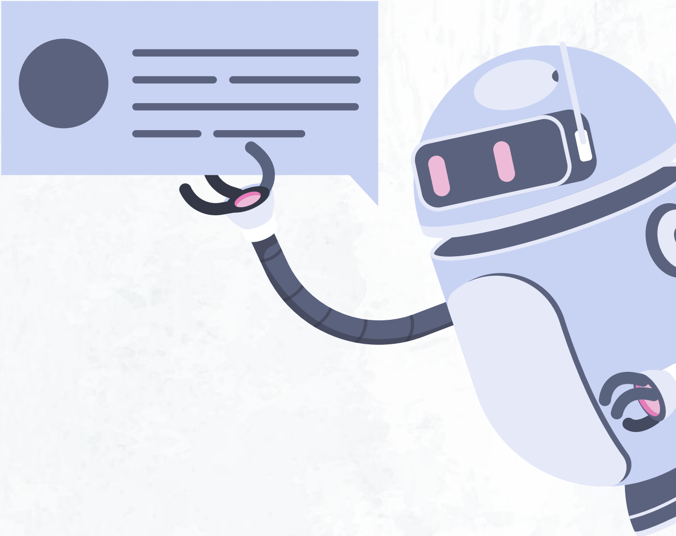
03 → Our solution-Front end

04 → Translation

05 → Demo

01 →

Purpose



Purpose

Target Users →

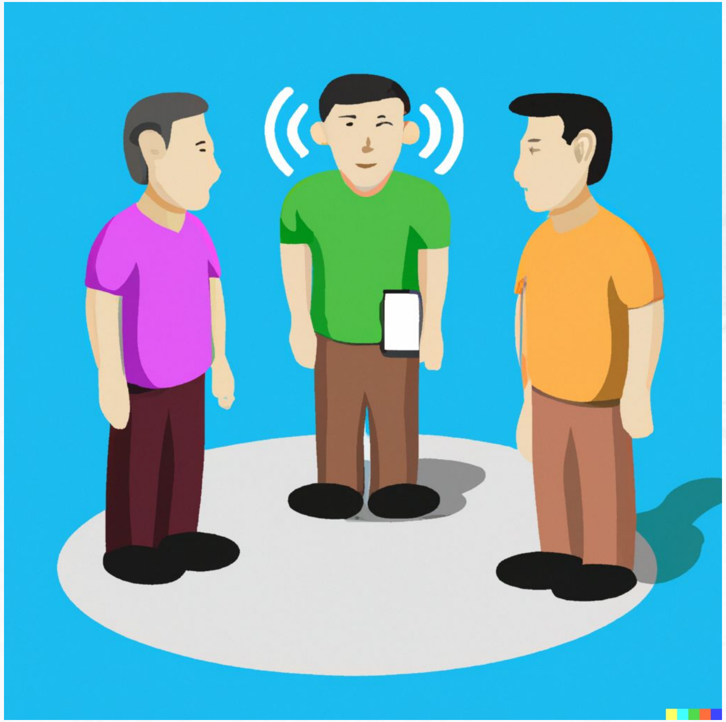
- Low hearing people

Motivation →

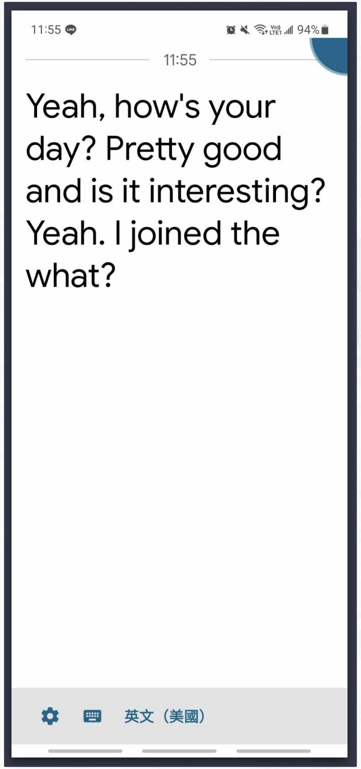
- Current live transcript application perform well on English real-time transcript

What if it’s a conversation with many people?

Setting



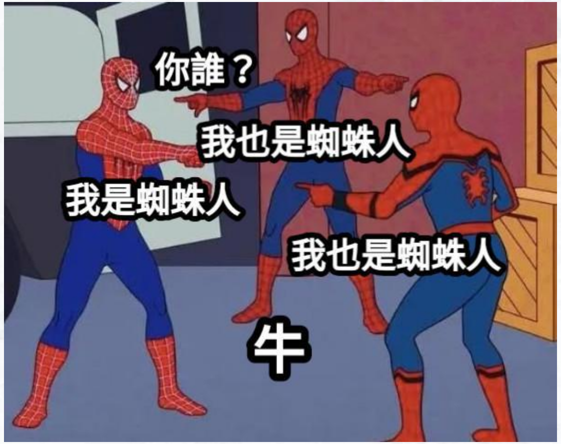
Example - 5-people conversation


11:55
94%

11:55

Yeah, how's your day? Pretty good and is it interesting? Yeah. I joined the what?

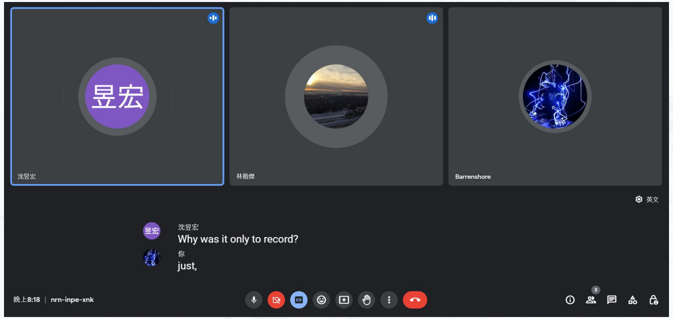
英文（美國）


你誰？
我也是蜘蛛人
我是蜘蛛人
我也是蜘蛛人
牛

Possible solution - voice identification

Hardware - google meet →

- not a good solution in our setting


昱宏
沈昱宏
林楷傑
Barrenshore
英文
昱宏
晚上8:18 | nrn-inpe-xnk
3

沈昱宏
Why was it only to record?

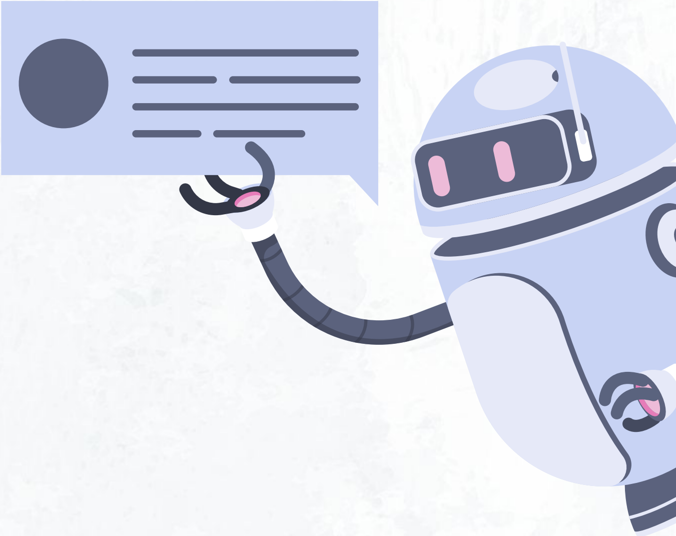
你
just,

Software - speaker diarization(說話者分辨) →

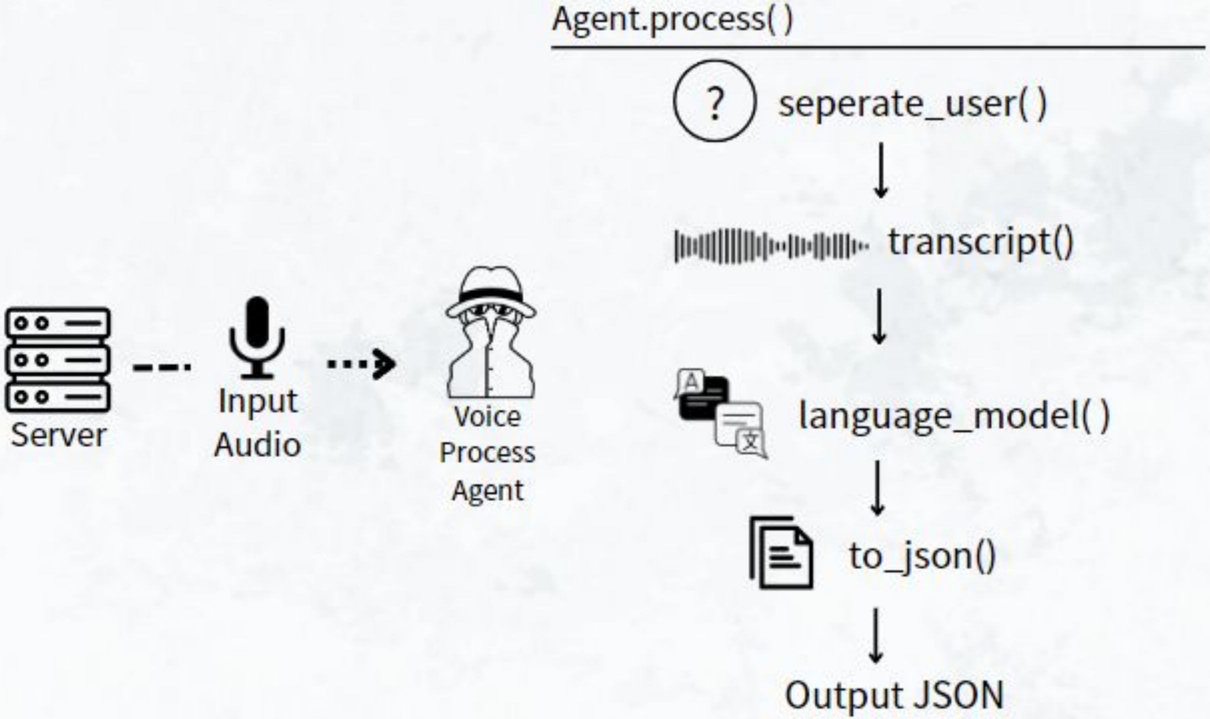
- Google cloud platform
- Amazon model
- often used for off-line data

02 →

Our solution - Back end



Back end - concept



Server

Input Audio

Voice Process Agent

Agent.process( )

seperate_user( )

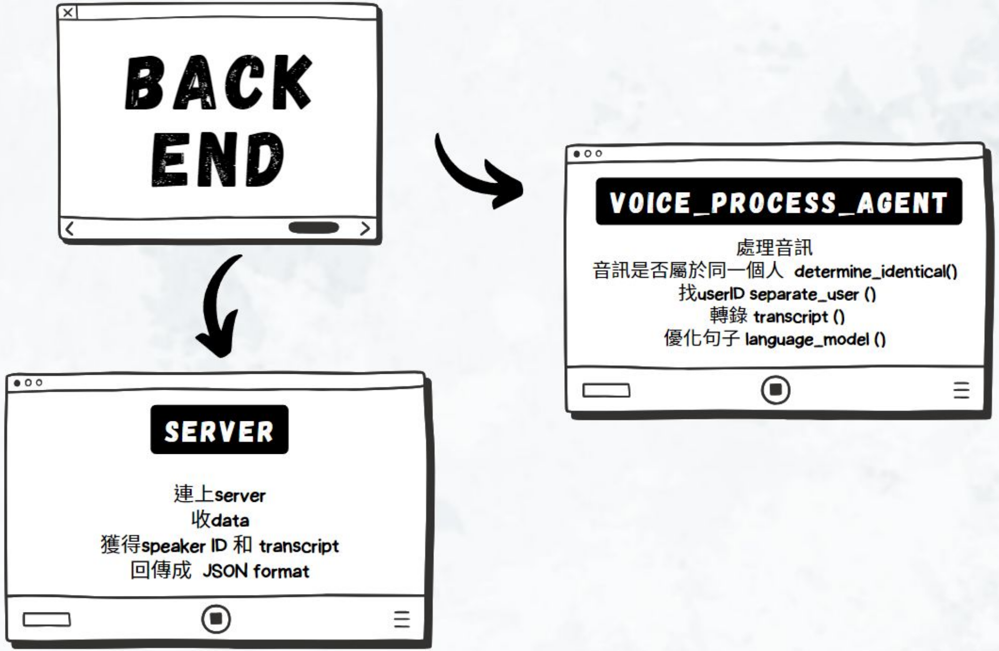
transcript()

language_model( )

to_json()

Output JSON

Back end - function



BACK END

VOICE_PROCESS_AGENT

處理音訊
音訊是否屬於同一個人 determine_identical()
找userID separate_user ()
轉錄 transcript ()
優化句子 language_model ()

SERVER

連上server
收data
獲得speaker ID 和 transcript
回傳成 JSON format

Models

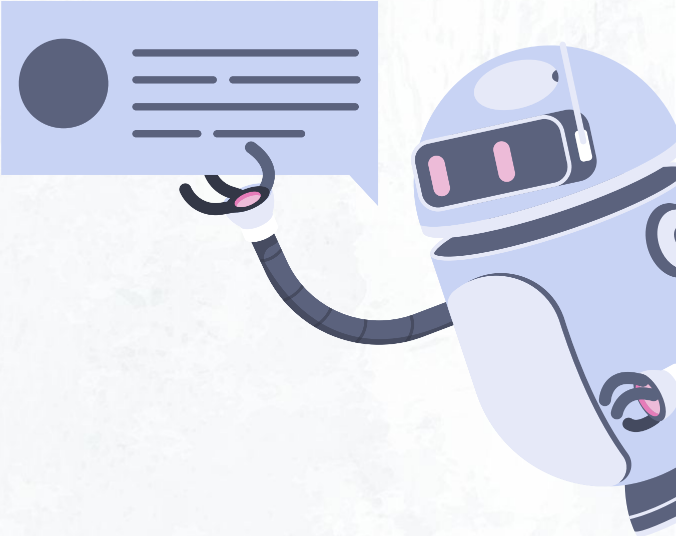
Separate user → speech brain

Transcript → speech_recognition library (in PYPI) (google)

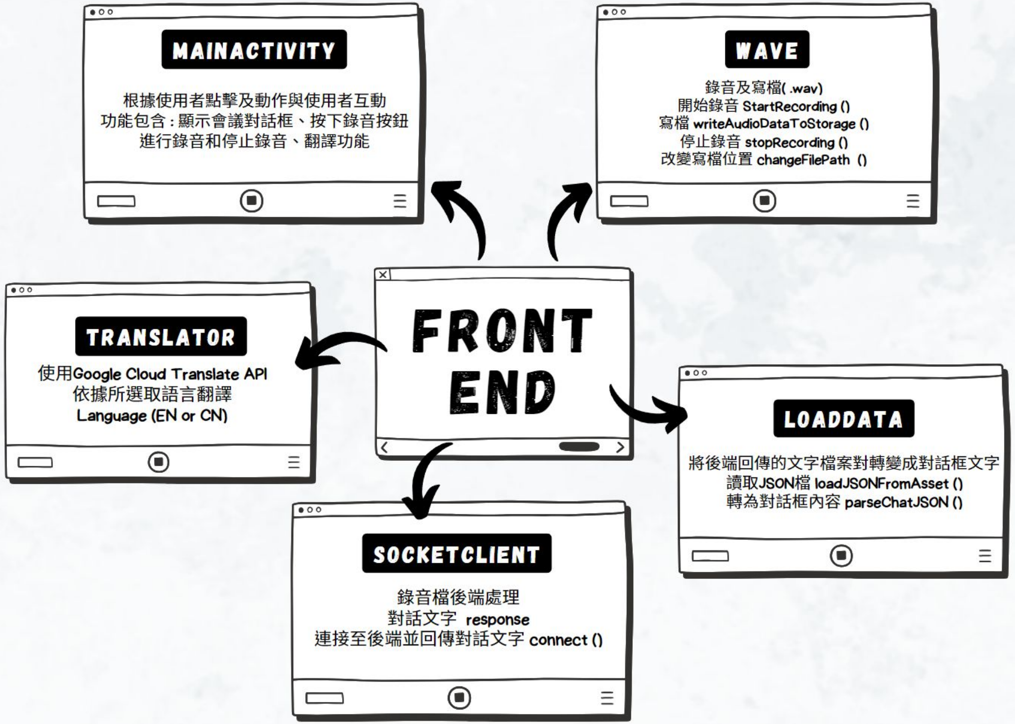
Language model → PALM

03 →

Our solution - Front End



Front end - function



FRONT END

MAINACTIVITY

根據使用者點擊及動作與使用者互動
功能包含：顯示會議對話框、按下錄音按鈕
進行錄音和停止錄音、翻譯功能

WAVE

錄音及寫檔(.wav)
開始錄音 StartRecording ()
寫檔 writeAudioDataToStorage ()
停止錄音 stopRecording ()
改變寫檔位置 changeFilePath ()

TRANSLATOR

使用Google Cloud Translate API
依據所選取語言翻譯
Language (EN or CN)

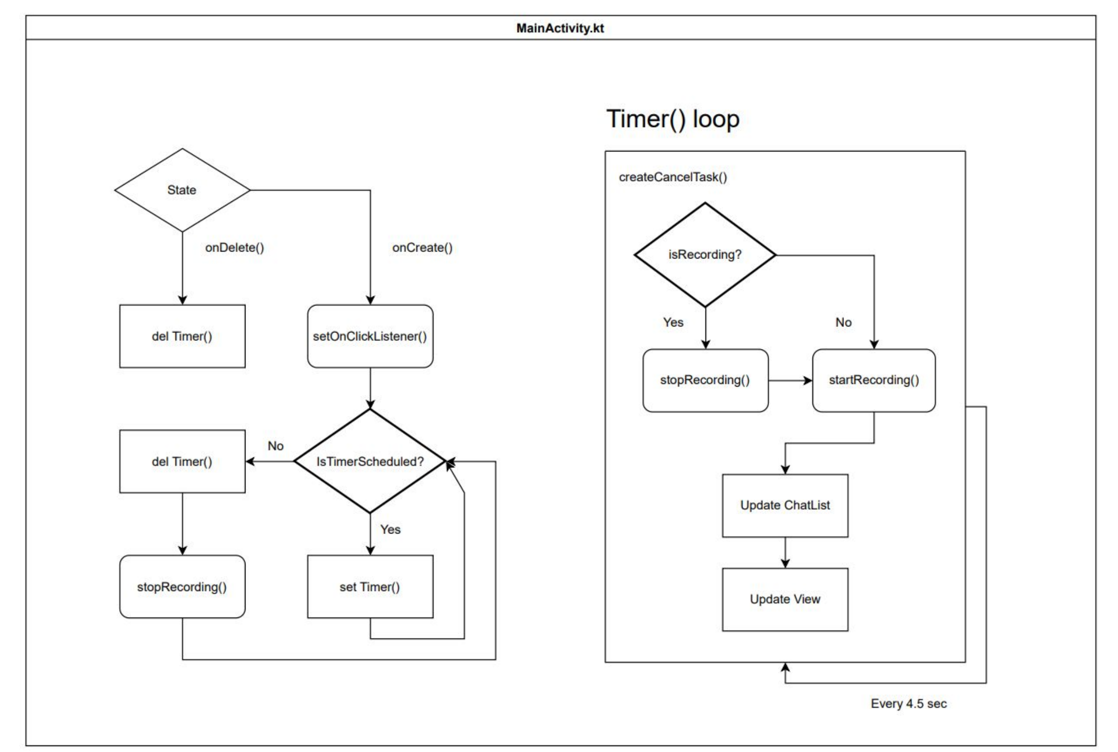
LOADDATA

將後端回傳的文字檔案對轉變成對話框文字
讀取JSON檔 loadJSONFromAsset ()
轉為對話框內容 parseChatJSON ()

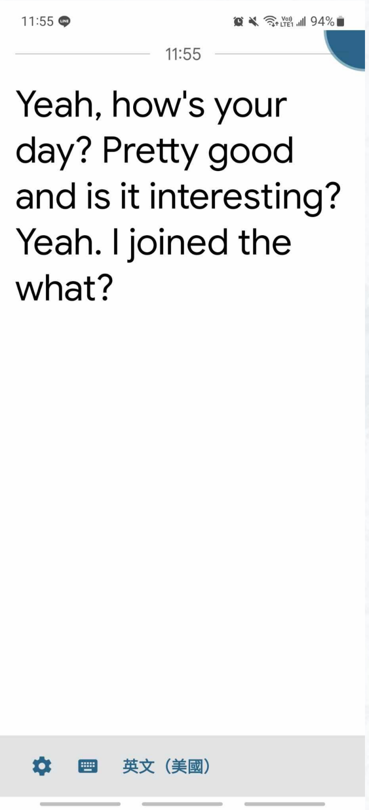
SOCKETCLIENT

錄音檔後端處理
對話文字 response
連接至後端並回傳對話文字 connect ()

Front end - structure

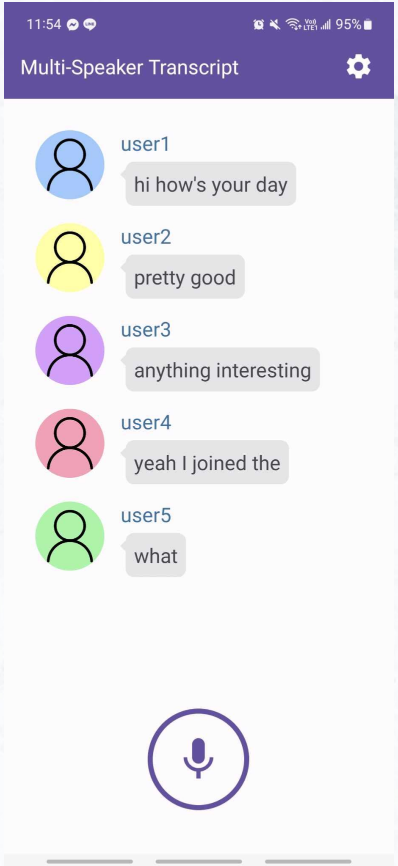

MainActivity.kt
State
onDelete()
onCreate()
del Timer()
setOnClickListener()
No
del Timer()
IsTimerScheduled?
Yes
stopRecording()
set Timer()
Timer() loop
createCancelTask()
isRecording?
Yes
No
stopRecording()
startRecording()
Update ChatList
Update View
Every 4.5 sec

Result


11:55
94%
11:55
英文（美國）

Yeah, how's your day? Pretty good and is it interesting? Yeah. I joined the what?

→


11:54
95%
Multi-Speaker Transcript

user1
hi how's your day

user2
pretty good

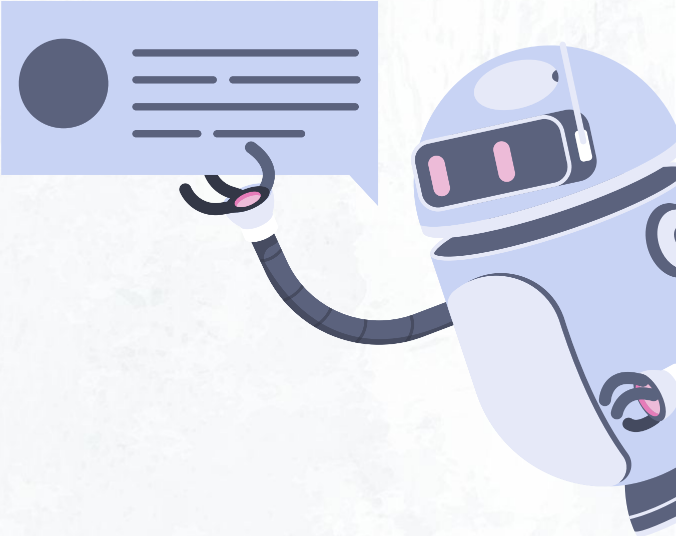
user3
anything interesting

user4
yeah I joined the

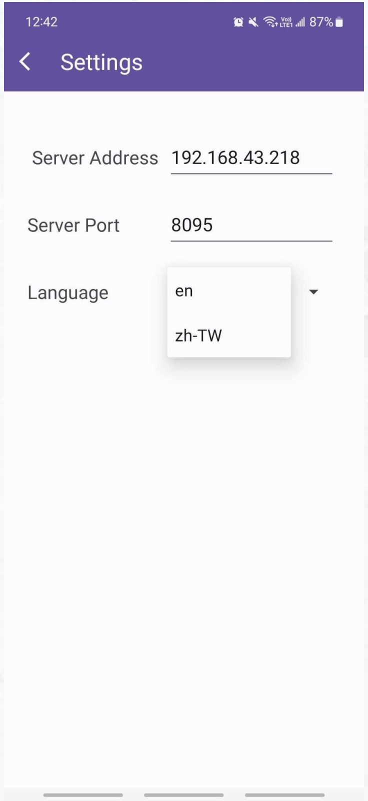
user5
what

04 →

Translation - for more users



Setting Layout



12:42
87%
Settings

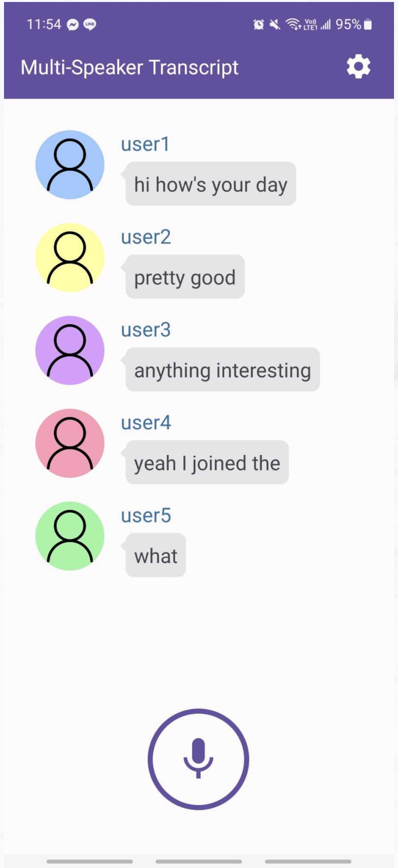
Server Address 192.168.43.218

Server Port 8095

Language en

zh-TW

Result


11:54 95%

Multi-Speaker Transcript

user1
hi how's your day

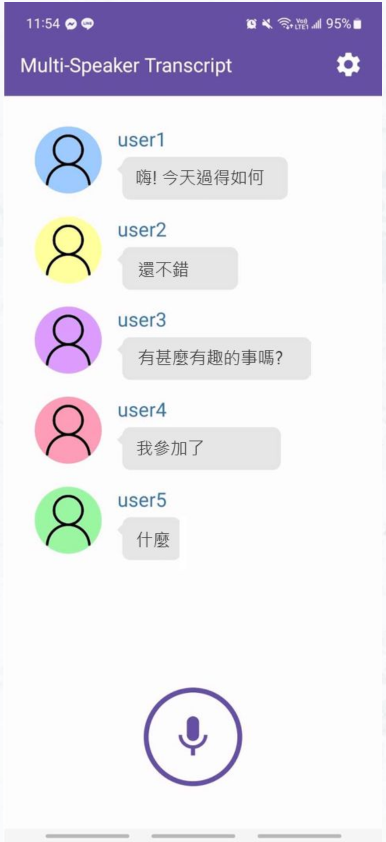
user2
pretty good

user3
anything interesting

user4
yeah I joined the

user5
what

→


11:54 95%

Multi-Speaker Transcript

user1
嗨! 今天過得如何

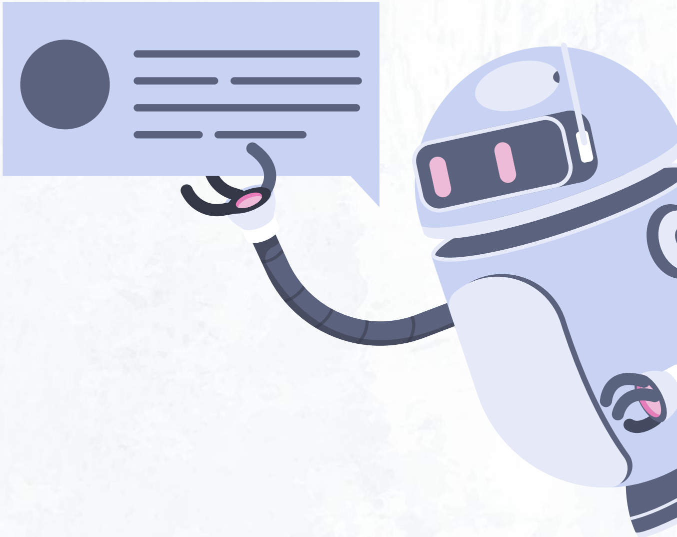
user2
還不錯

user3
有甚麼有趣的事嗎?

user4
我參加了

user5
什麼

05 →

Demo