

# GGWhisPer

企業溝通無國界翻譯系統  
建立多語言的即時語音翻譯

Team C2 : 哭給我汽油庫存兔兔去  
Member : 余振揚、莊上緣、江宇浩、林志芸



# Outline

- ▶ **User Scenario & Story**
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ Insights & QA



# User Story

作為一位台灣工程師，  
我希望會議中我的技術講解能被即時翻譯，  
以便全球同事能立即理解我的想法。

作為一位會議參與者，  
我希望系統能儲存翻譯後的會議記錄，  
以便之後能隨時查閱重要討論內容。



## 即時翻譯模組

支援多語言即時語音識別及翻譯，專注於技術對話場景

## 術語智能處理

建立專業術語知識庫，確保翻譯的準確性

## 會議助手功能

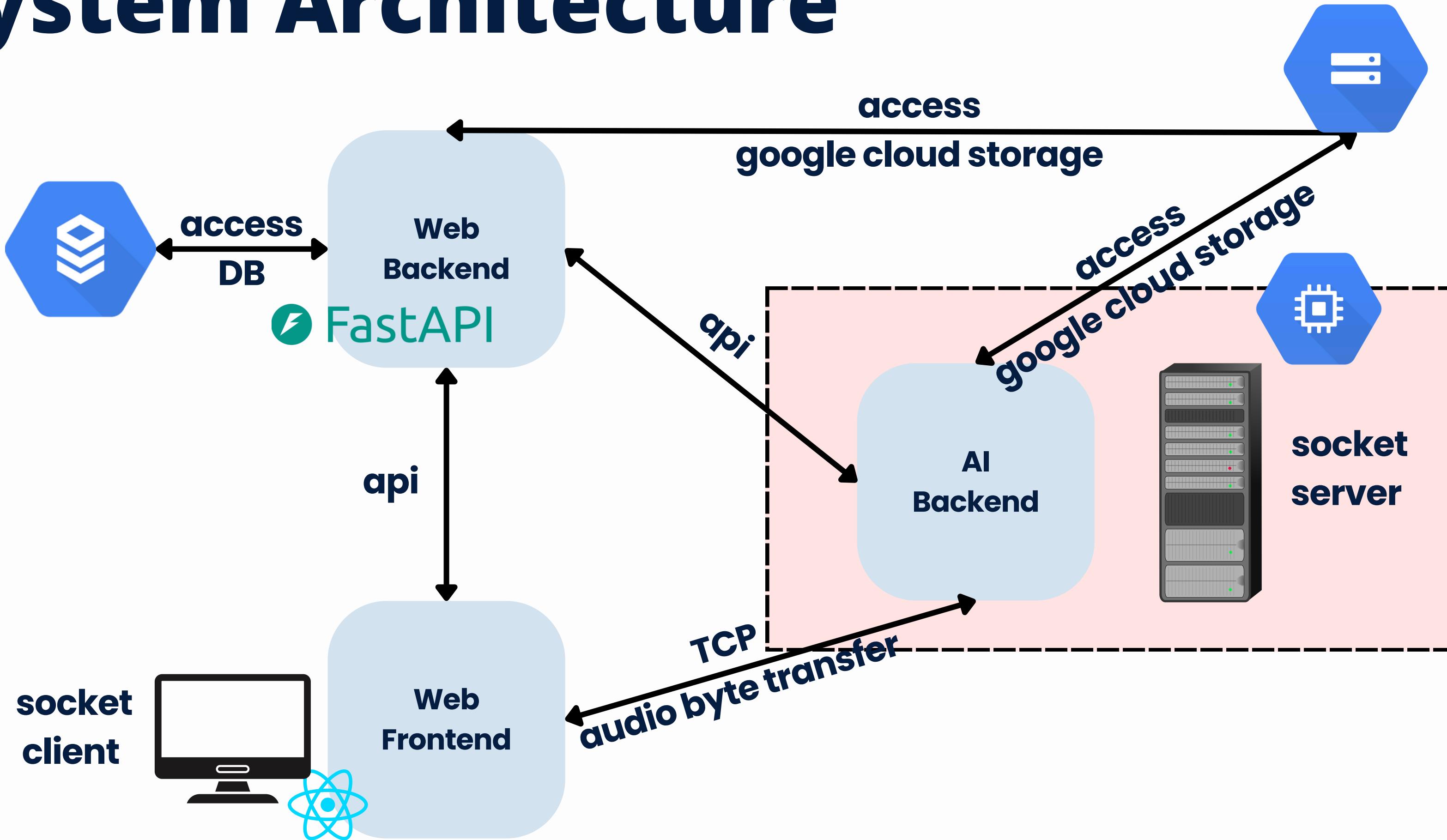
生成多語言會議紀錄與摘要，追蹤重要決策點

# Outline

- ▶ User Scenario & Story
- ▶ **System Architecture**
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ Insights & QA

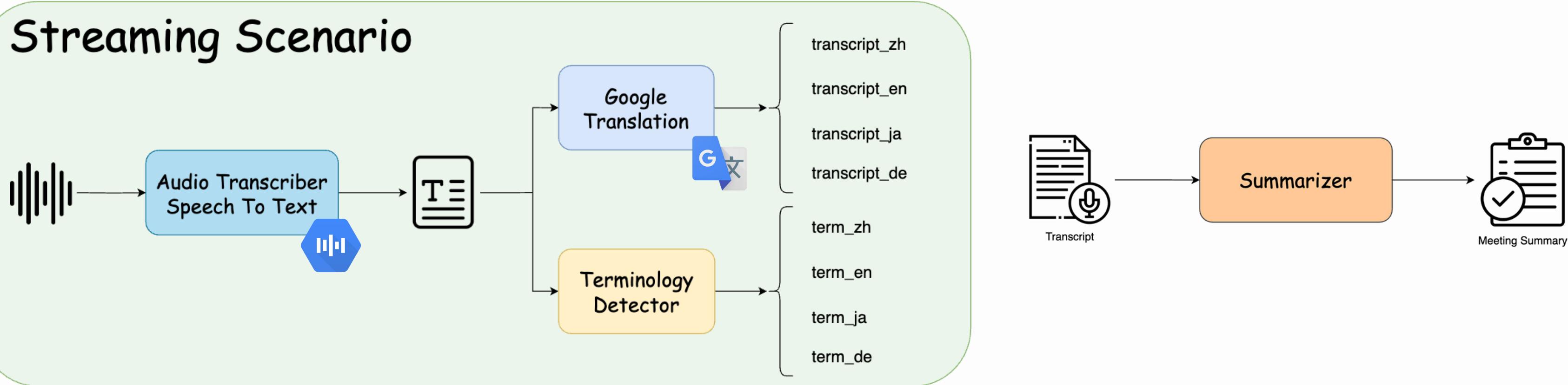


# System Architecture

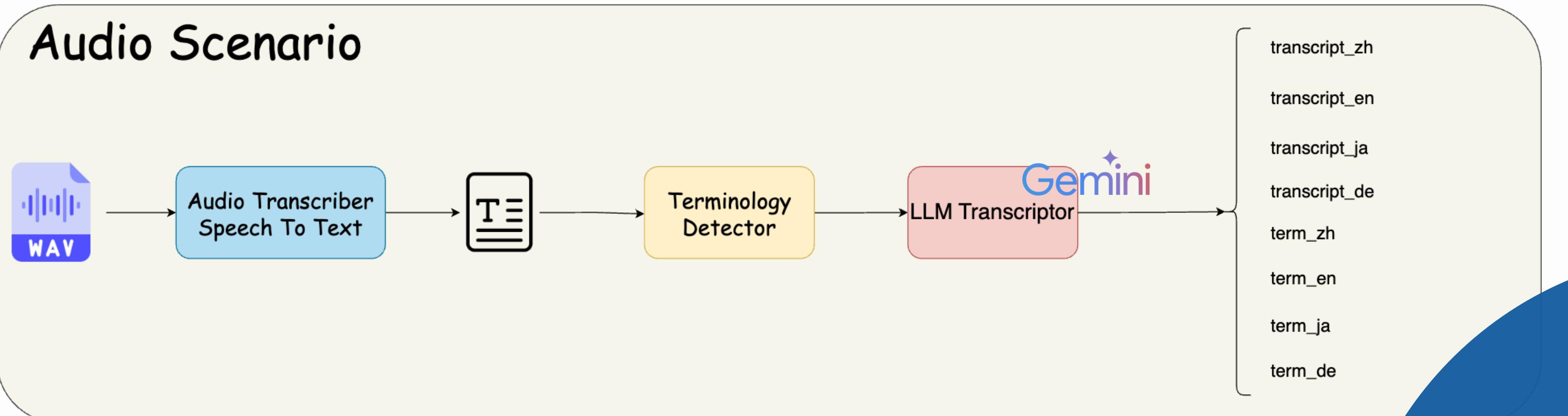


# System Architecture

## Streaming Scenario



## Audio Scenario



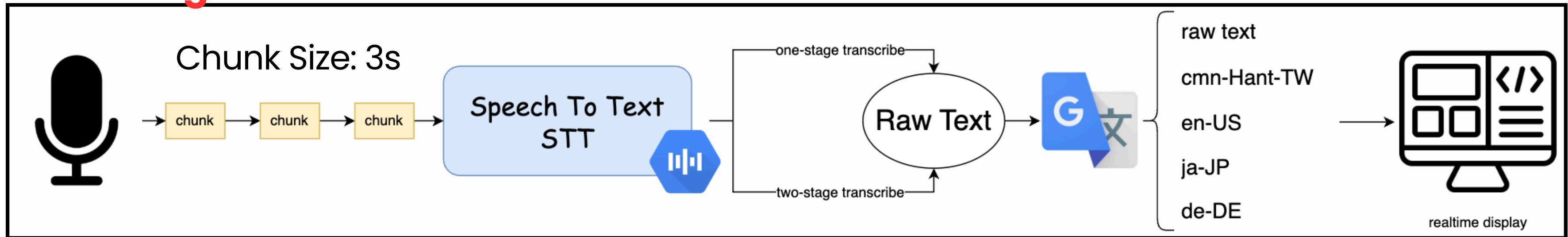
# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation**
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ Insights & QA

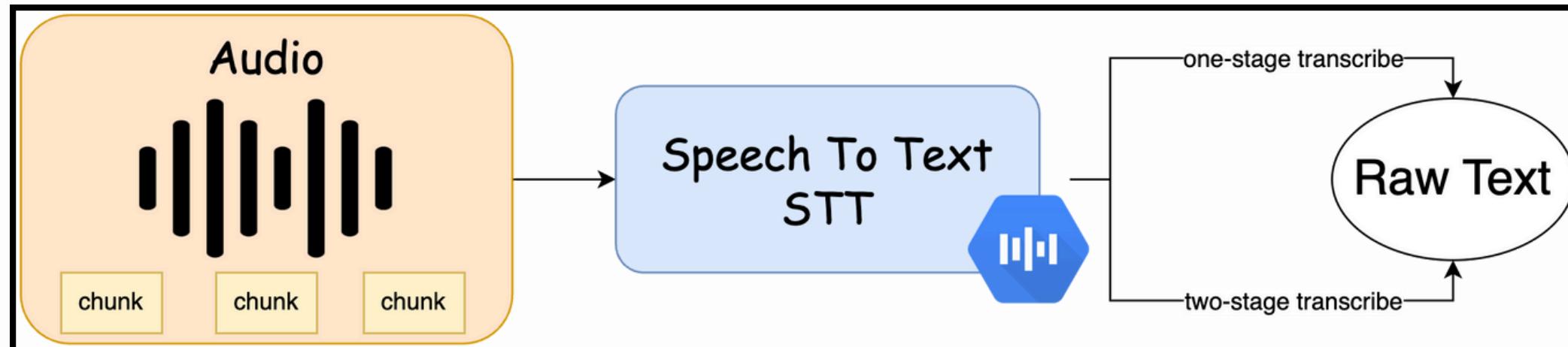


# Task 1: Real-Time Speech To Text Translation

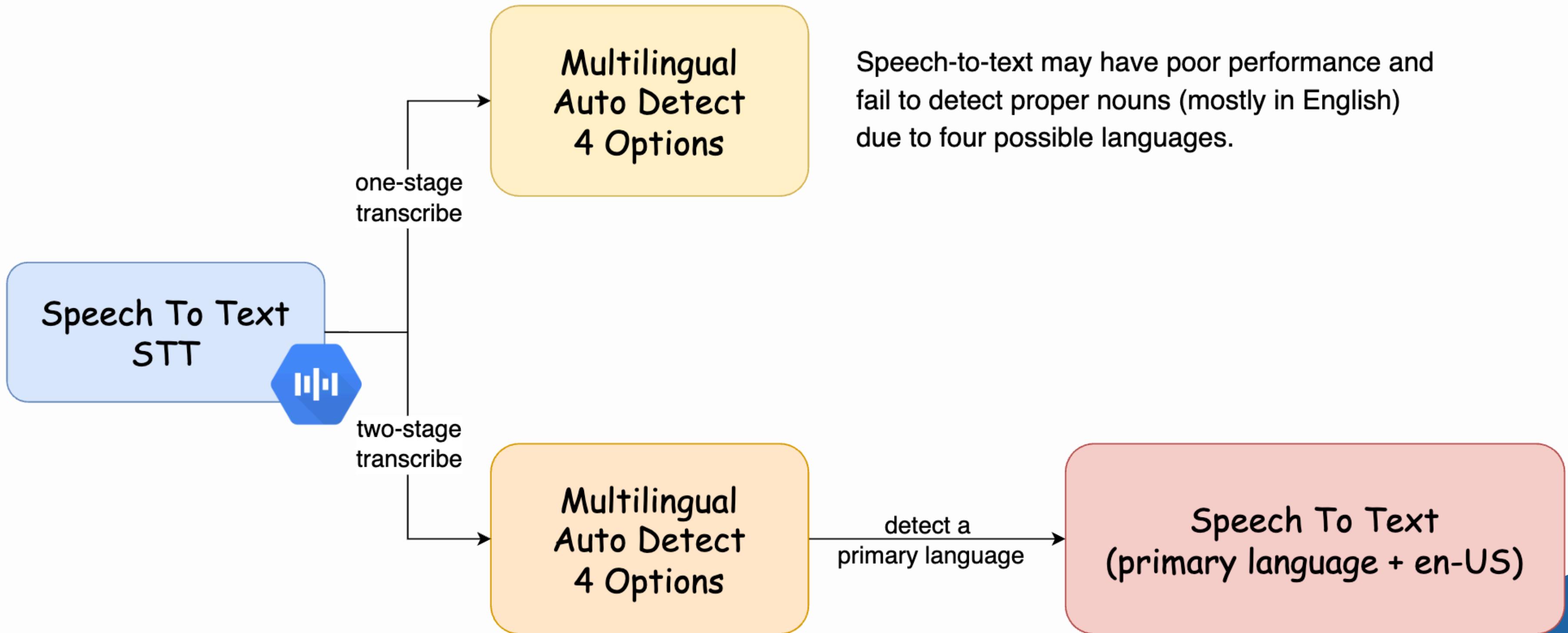
## Streaming Mode



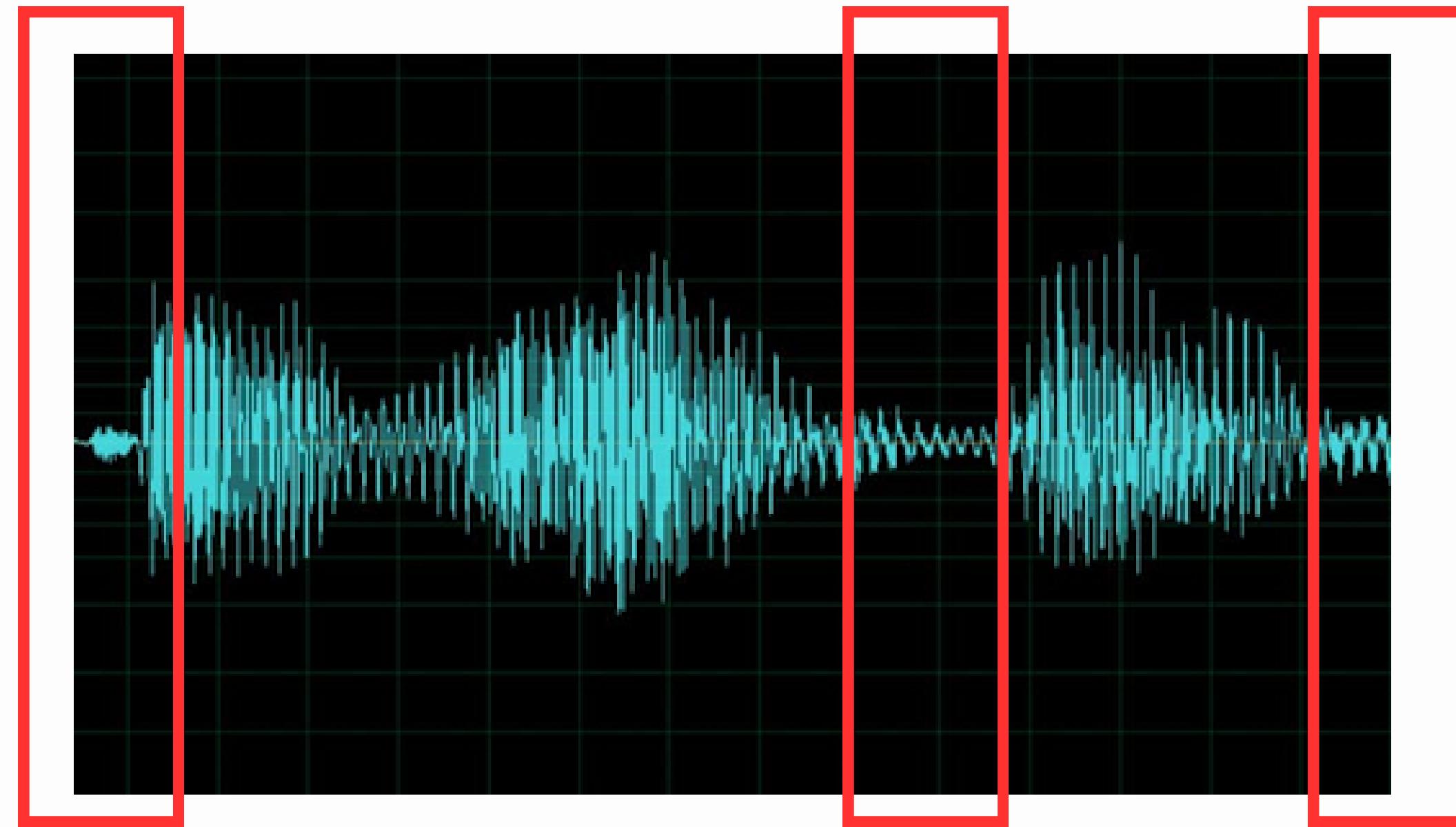
## Audio Mode



# Transcribe: one-stage v.s. two-stage



# Full Audio Mode: Split by Silence



Unlike the fixed trunk size in streaming mode, in full audio mode, we use **silent segments** as **sentence separators** to capture full sentence trunks for speech-to-text.

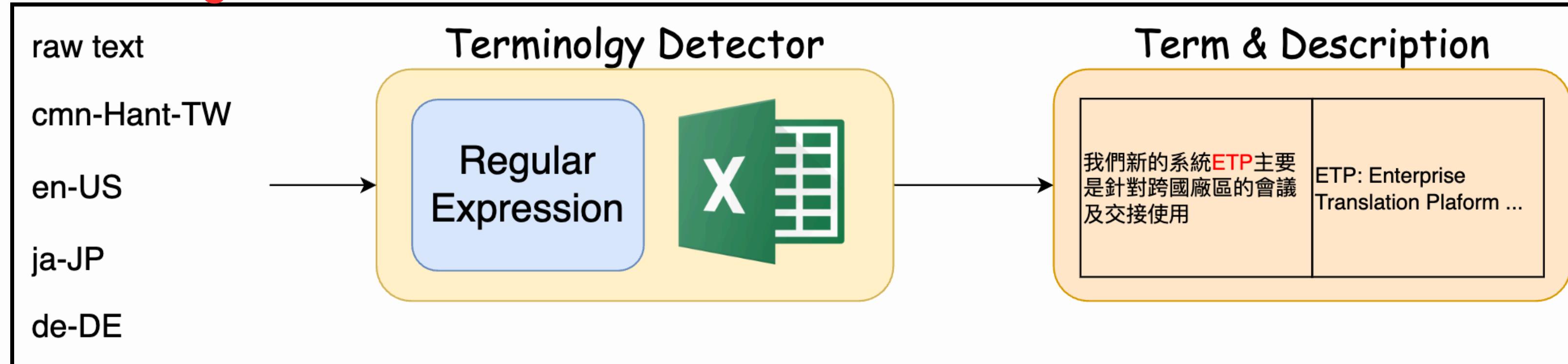
# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ **Task 2: Company-Specific Terminology Detector**
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ Insights & QA

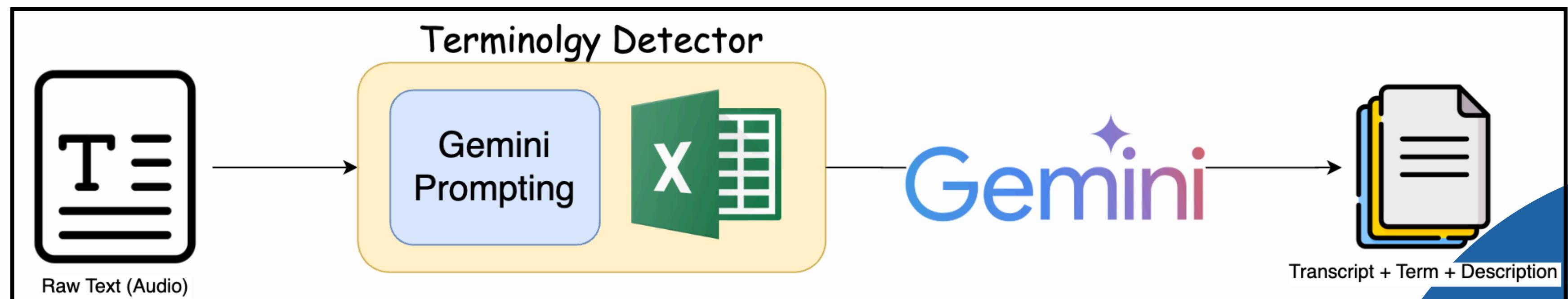


# Task 2: Company-Specific Terminology Detector

## Streaming Mode



## Audio Mode

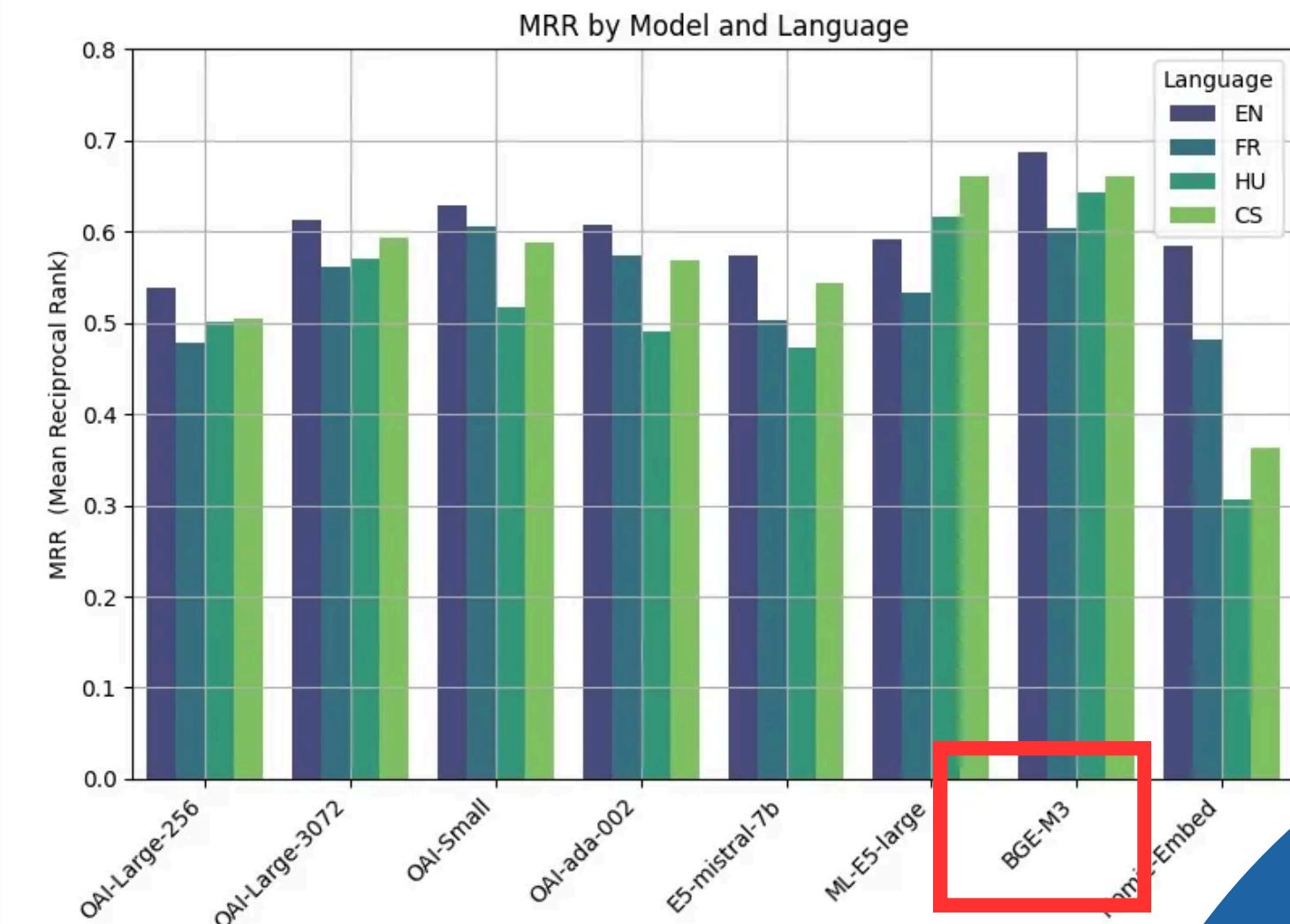
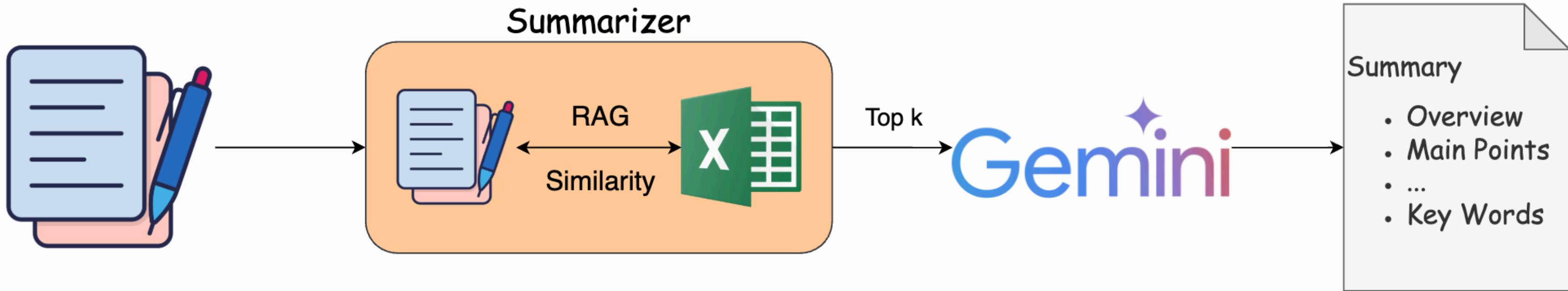


# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings**
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ Insights & QA

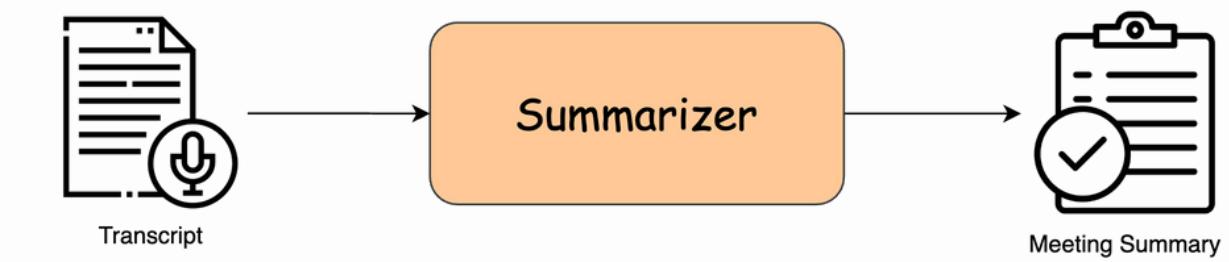
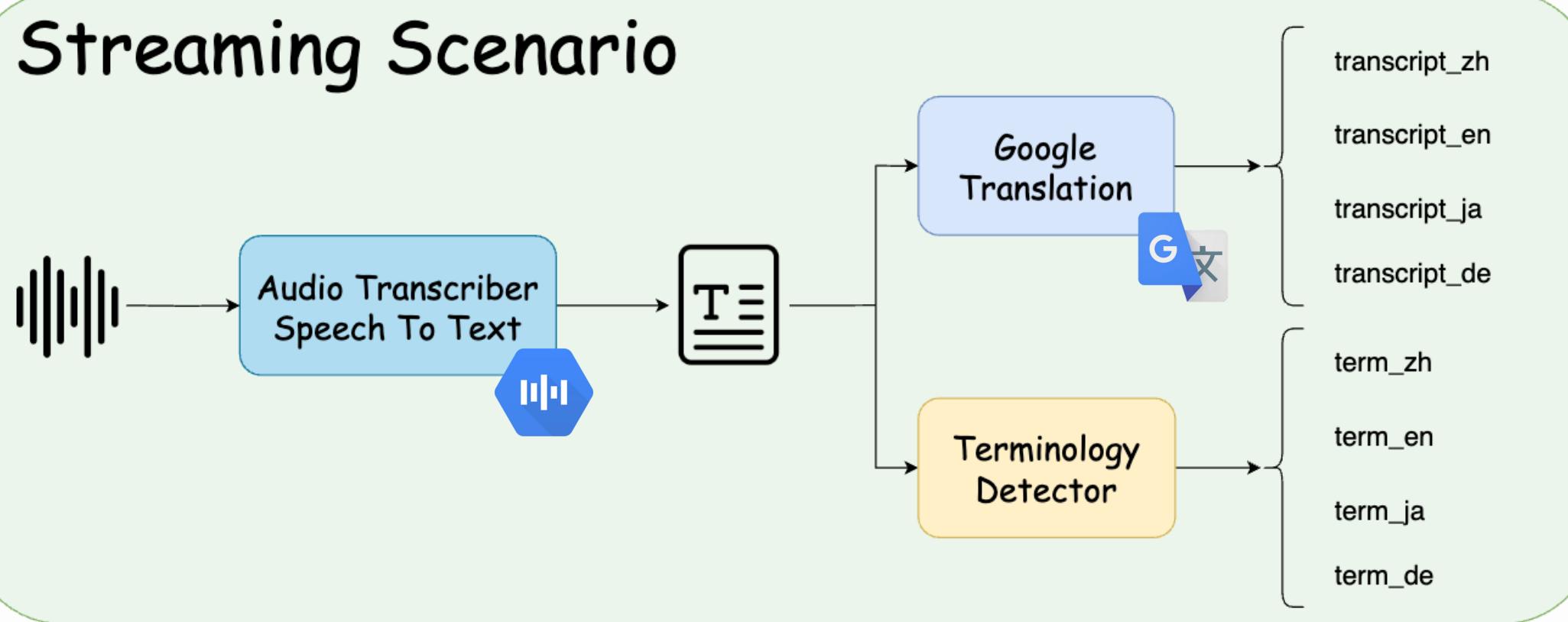


# Task 3: Clerk Of Meetings

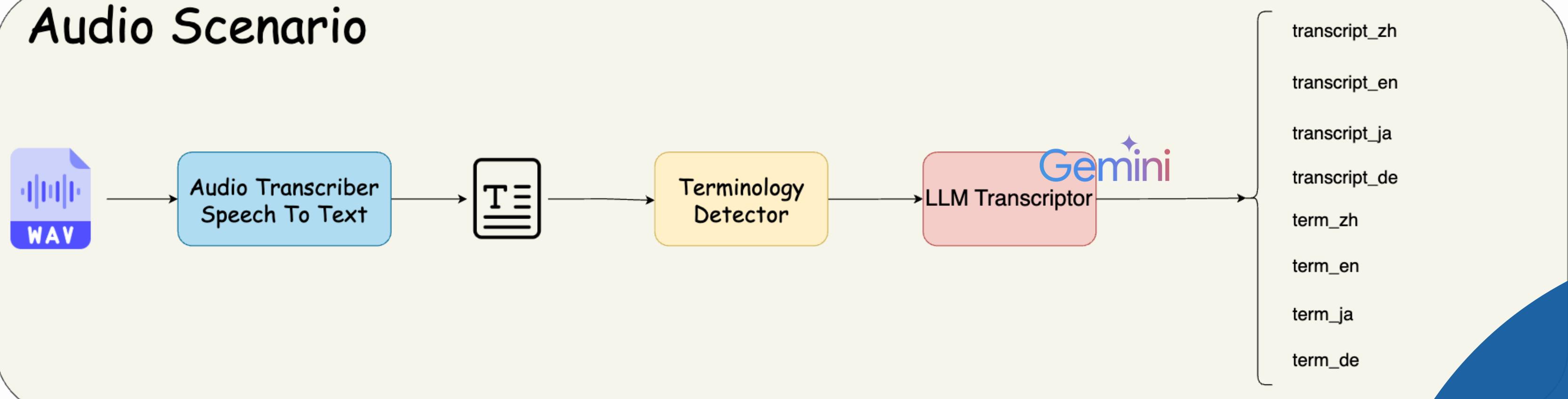


# System Architecture Recap & Results

## Streaming Scenario



## Audio Scenario



# Testing Dataset Results

## Transcript

大家好，今天的會議主要是請Ivan來介紹他們的新系統。為了加快生產速度並減少錯誤，我們請Ivan開發了Genai平台來協助各位的日常工作。謝謝的意思。

今天很高興可以和大家介紹的新系統。

### SGP

該系統包含最愛的服務以及那個服務。

不可以安全地在公司使用就那個服務。

各位影片、像技術文件、回應內容的文件，都可以上傳到系統。

根據SGP來整理文件內容以及問答。

工廠裡的機器人經常出現問題。

主要原因是當港口還有貨物時，機器手臂會先將貨物移到邊緣。

這個系統可以解決這個問題。

可以。

除了以攝影機偵測外，也可以加入最愛的功能，判斷何時該讓下一批貨物進來。

新工廠裡有很多新人，STB之後應該做什麼？還不太清楚如何修理示波器。這個系統可以幫助我們嗎？

沒問題。

## Term

SGP

SGP

STB

SGP

GKE

Cloud Run

ALP

## Summary

### 會議概述

Ivan介绍了新開發的Secure GenAI Platform (SGP)系統，旨在提高生產效率並減少錯誤。該系統能處理技術文件和流程說明，並通過AI解答員工問題，尤其能解決工廠機器人因港口貨物堆疊導致的問題，並協助新員工熟悉操作流程。系統利用GKE叢集處理文件，並通過Cloud Run部署供使用者查詢。

### 主要討論點

- SGP系統的功能介紹及應用場景。
- SGP系統如何解決工廠機器人及新員工培訓問題。
- SGP系統的資料處理流程及安全措施。
- ALP問題的處理。

### 決策事項

- 將技術文件和流程說明上傳到SGP系統，由AI處理並解答員工問題。
- Martin需注意ALP問題。

### 待辦事項

項目	負責人	期限
將技術文件及流程說明整理好並上傳到SGP系統	所有相關人員	盡快
處理ALP問題	Ivan及相關團隊	盡快

### 關鍵詞彙

SGP (Secure GenAI Platform), DP (Data Platform), Cloud Storage, ETP (Enterprise Translation Platform), STK (Stocker), ALP (Active Load Port), Cloud Run, GKE (Google Kubernetes Engine)

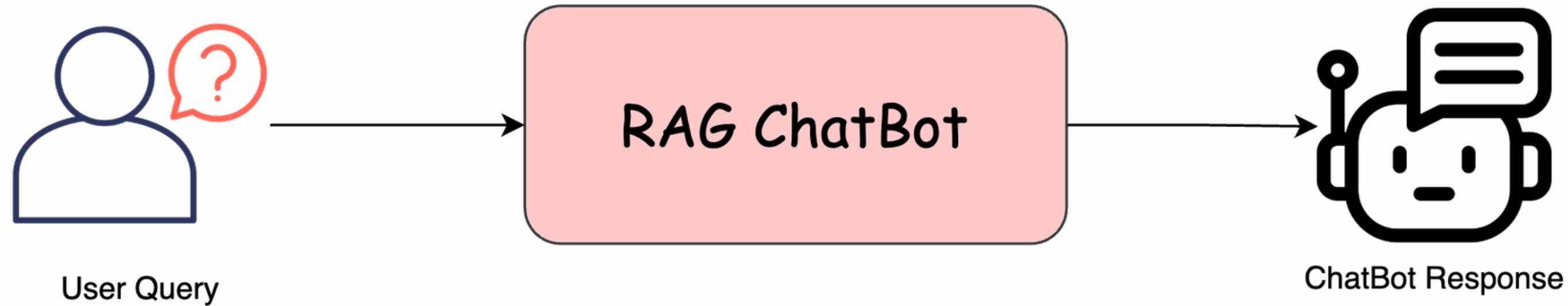
# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records**
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ Insights & QA

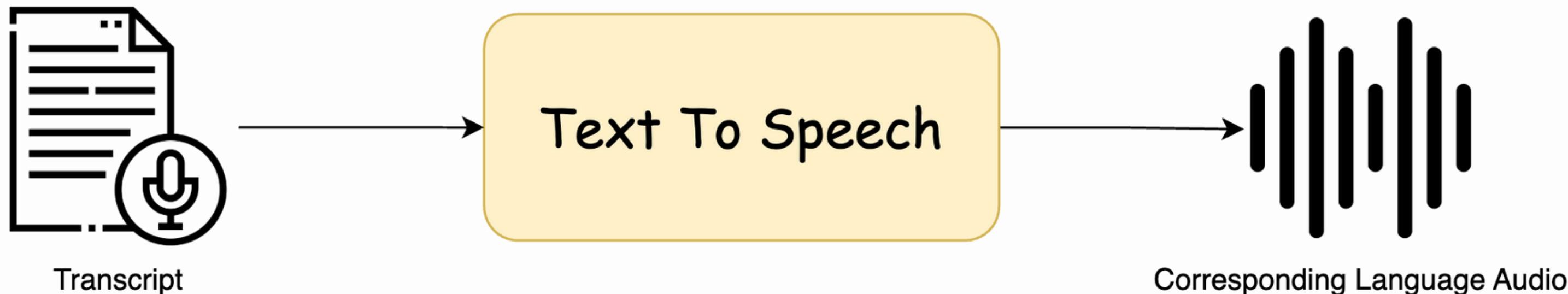


# Bonus Feature: Rag ChatBot, Text-To-Speech

same technique as RAG Summarizer



invoke Google Text To Speech API



# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)**
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ Insights & QA



# Metric Evaluation – Task 1

## How to evaluate speech-to-text transcription quality?

- **WER** (Word Error Rate)
- **CER** (Character Error Rate)

We use WER for **English** and **German**, and CER for **Chinese** and **Japanese**, and calculate a **weighted error rate** based on each language's **proportion in the text**.

Character Error Rate (CER):

$$CER = \frac{\text{Number of incorrect characters}}{\text{Total number of characters in the reference text}} \times 100\%$$

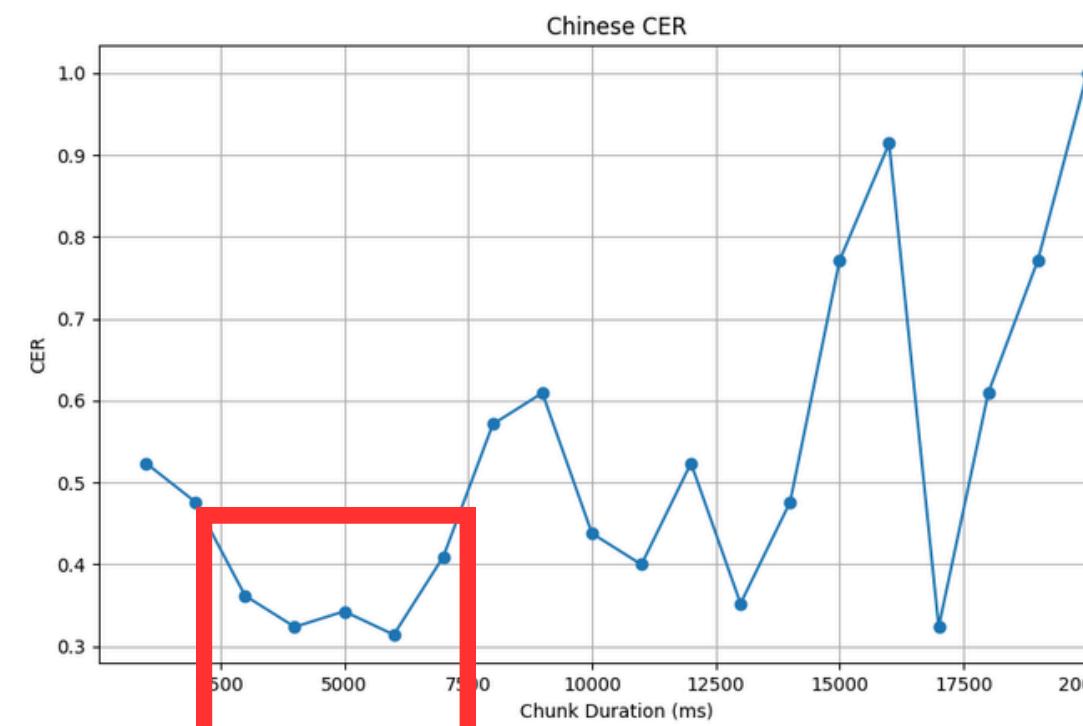
Word Error Rate (WER):

$$WER = \frac{\text{Number of incorrect words}}{\text{Total number of words in the reference text}} \times 100\%$$

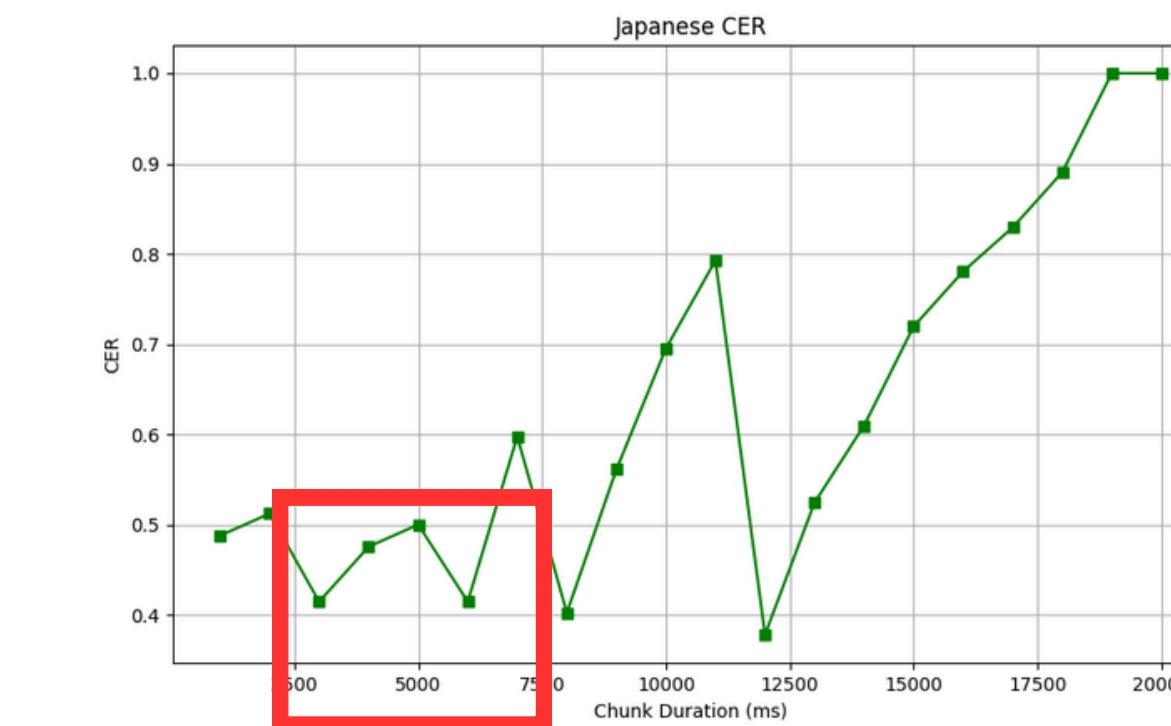
# Task 1 - Chunk Size Experiment

Why is the chunk size in streaming mode set to 3 seconds?

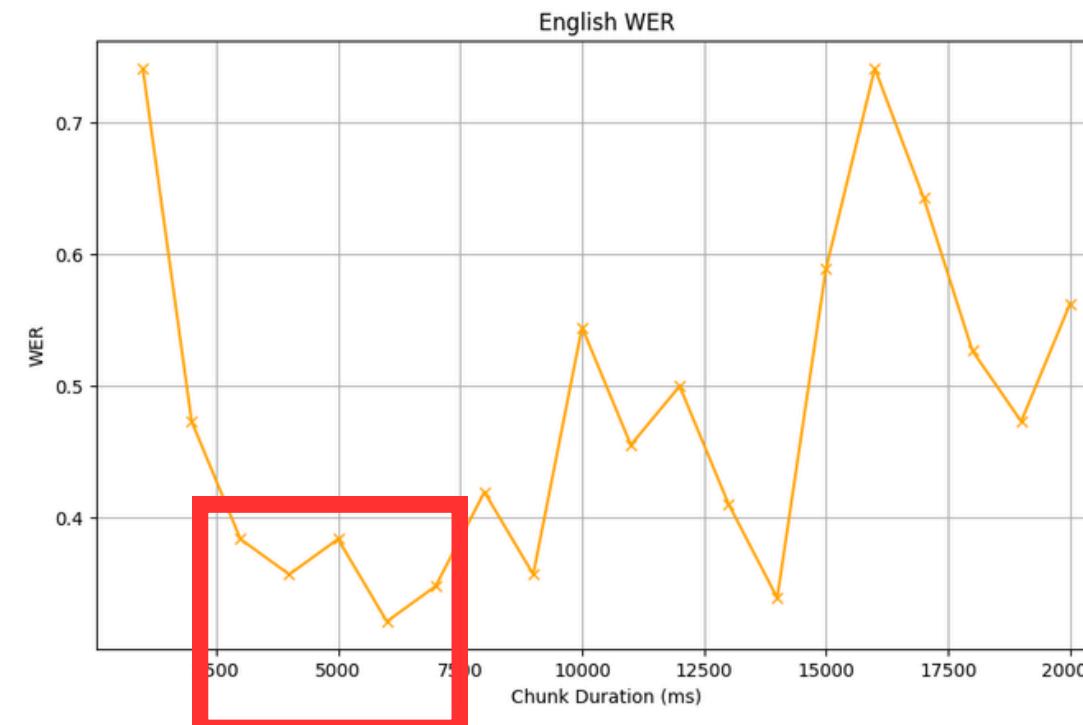
Chinese CER



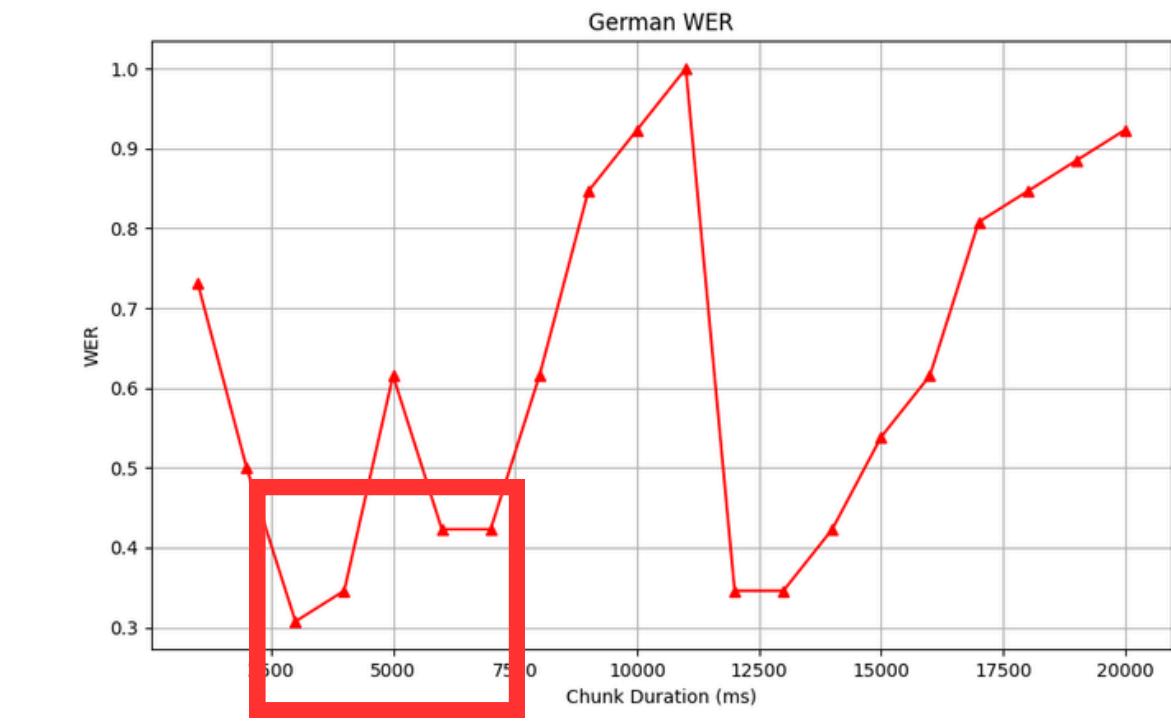
Japanese CER



English WER



German WER



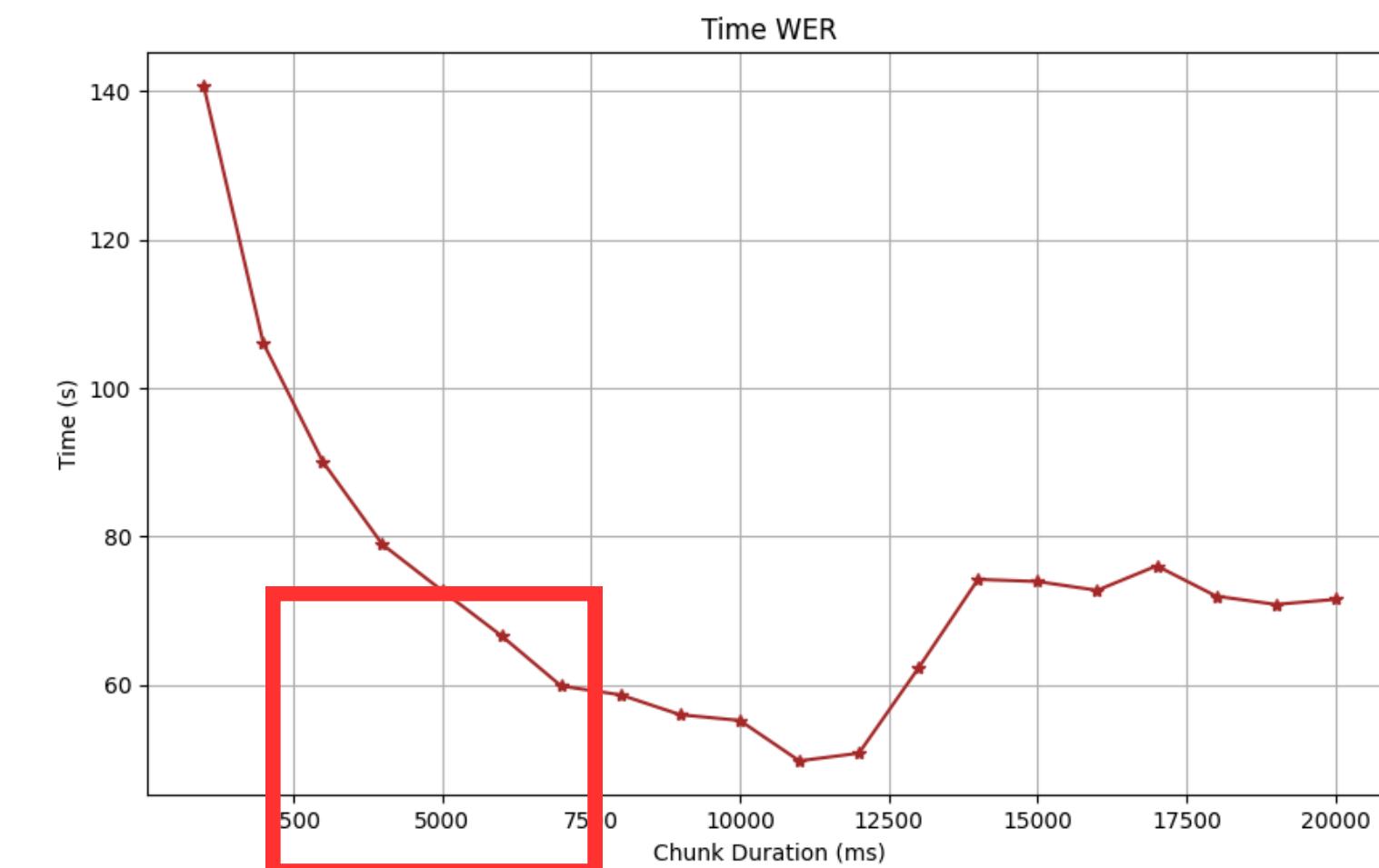
# Task 1 - Chunk Size Experiment

**Why is the chunk size in streaming mode set to 3 seconds?**

**Weighted Error Rate**



**Time WER**



# Ablation Study

## Training Dataset Chinese Transcript Results

	CER (compared to Training wav answer)
<b>one-stage transcribe + RegExp PN detection</b>	0.634
<b>two-stage transcribe + RegExp PN detection</b>	0.601
<b>two-stage transcribe + LLM PN detection</b>	0.581
<b>Ground Truth</b>	0.0

# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ **DEMO**
- ▶ Frontend Workflow
- ▶ Insights & QA



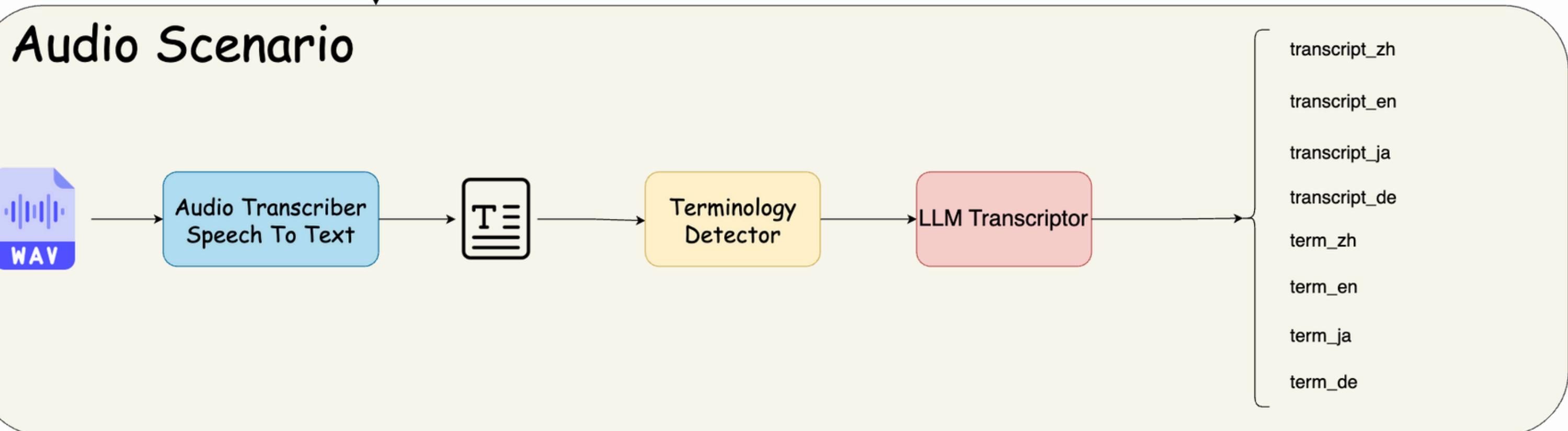
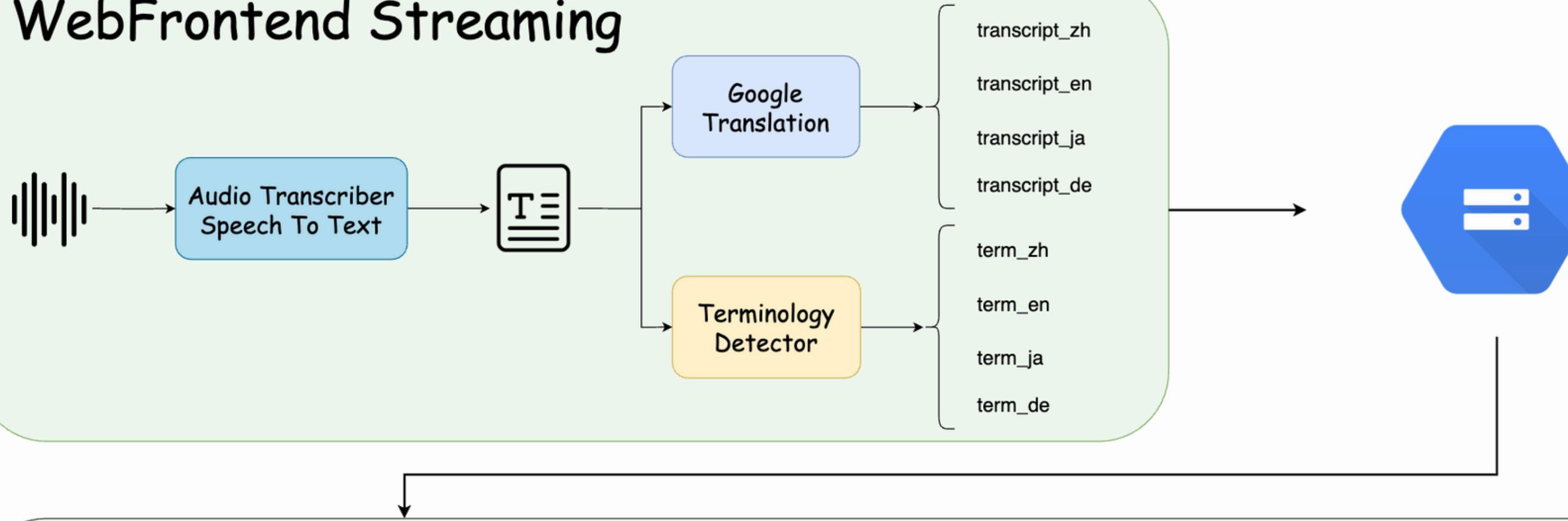
# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ **Frontend Workflow**
- ▶ Insights & QA



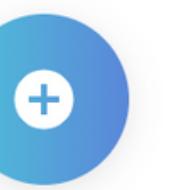
# Frontend Workflow

## WebFrontend Streaming





上傳音檔



錄製音檔

### 會議紀錄搜尋

開始日期 :

年 / 月 / 日



結束日期 :

年 / 月 / 日



標題 :

搜尋標題

描述 :

搜尋描述

逐字稿內文:

搜尋逐字稿

專有名詞:

搜尋專有名詞

搜尋

重置

2025 / 2 / 14 15:11

2025 / 2 / 14 16:44

2025 / 2 / 14 16:47

2025 / 2 / 14 20:43

2025 / 2 / 14 21:07

2025 / 2 / 14 21:15

2025 / 2 / 14 21:19

test

test

test

odflaibnea;ofgnae;ogjv

**標題**

請輸入標題

選擇檔案 未選擇任何檔案

**描述**

請輸入描述

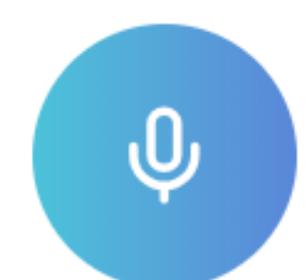
**標題**

請輸入標題

**描述**

請輸入描述

上傳



0:00

點擊即開始錄音

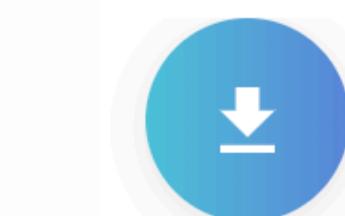
開始錄製

# 測試測試

大老現身店家又驚又喜，但這可不是張忠謀頭一回被民眾野生捕獲。張忠謀現身美式餐廳大口吃漢堡、薯條，身旁不見任何隨扈，儘管講究養生但私底下的張忠謀似乎對速食情有獨鍾。台積電董事長張忠謀：「昨天我的確是在大直美麗華跟太太一起看電影，看完電影在Fridays吃飯。」身穿紫色毛衣，張忠謀又被拍到出現在百貨公司速食店裡跟老婆甜蜜約會，桌上吃的同樣是漢堡、薯條，夫妻倆輕鬆用餐就像對平凡夫妻，儘管高齡80幾歲卻看不出大老闆架勢，也讓民眾感受私下張忠謀親民的一面。

## 原始音檔

▶ 0:00 / 0:50



檔案下載

中文逐字稿

英文逐字稿

日文逐字稿

德文逐字稿

中文專有名詞

英文專有名詞

日文專有名詞

德文專有名詞

中文總結

## 中文音檔

▶ 0:00 / 0:47

## 英文音檔

▶ 0:00 / 0:38

## 日文音檔

▶ 0:00 / 0:52

## 德文音檔

▶ 0:00 / 0:45

升高的原因不明。

大家好，今天我們要討論關於 DDR 比例的問題。

這週 DP 的比例相當高。

Martin 知道原因嗎？

我昨天值大夜班，並已將事情交接給 Lisa。

誰能解釋原因？

我在這裡，沒有 data radio.com。

與主控端相比，溫度等數值的使用量相當大。

為什麼這麼容易被更改？

這些數值不應該與主控端副本一致嗎？

我可以教系統紀錄檔識別是誰做的更改嗎？

可以。

## 專有名詞

### DDR Ratio

Defective Dies Ratio,為wafer defect的一個指標，為一個基準定義此批產品是否會報廢，而此defect影鄉高品質yield的

原文

中文

English

日本語

Deutsch

皆さん、今日はDDR Ratioに関する問題について話  
し合います。

今週のDPの比率がかなり高いです。

Martinはその理由を知っていますか？

私は昨日夜勤で、Lisaに業務を引き継ぎました。

誰がその理由を説明できますか？

ここにいますが、data radio.comのデータがありません。

マスターパートと比較したところ、温度などの数値  
がかなり使われていました。

なぜ簡単に変更できたのでしょうか？

値はマスターコピーと一致するべきではないでしょ  
うか？

誰が変更を加えたかを特定するために、システムロ  
グを参照でさしますか？

こゆうめいし

### DDR Ratio

不良ダイ比率: ウェーハの不良  
の指標であり、この基準で製  
品の廃棄を判断します。これ  
らの不良は製品の歩留まりに  
影響を与えます。

## 專有名詞小字典

你好！本系統專注於提供即時翻譯與專有名詞優化，我們將盡力提供最佳回答！

下午1:01:14

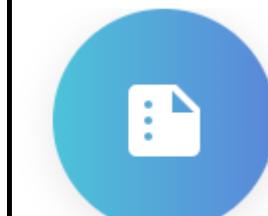
ETP 指的是「企業翻譯平台」(Enterprise Translation Platform)。它是一個提供安全且即時翻譯服務的平台，包含會議紀錄和查詢等功能，方便企業內部或跨企業之間的翻譯需求。

請輸入訊息...

下午1:01:21

ETP是什麼

下午1:01:20



生成總結

## 會議概述

本次會議主要討論DP資料平台中DDR比例（疑似指Defective Dies Ratio）異常升高問題。會議探討了異常升高的原因，懷疑與主控端資料不一致以及資料被輕易更改有關，並討論了如何追蹤更改記錄的可能性。

## 主要討論點

- DP資料平台DDR比例異常升高。
- DDR比例升高的原因不明。
- 數據與主控端數據不一致。
- 數據容易被修改。
- 追蹤數據修改記錄的方法。

## 決策事項

- 可以教系統記錄檔識別更改數據的使用者。

## 待辦事項

- 找出DP資料平台DDR比例異常升高的原因 | Martin/Lisa | 盡快
- 教系統記錄檔識別更改數據的使用者 | 待定 | 盡快

## 關鍵詞彙

DDR Ratio (Defective Dies Ratio), DP (Data Platform), EC (機台設定值)

# Outline

- ▶ User Scenario & Story
- ▶ System Architecture
- ▶ Task 1: Real-Time Speech To Text Translation
- ▶ Task 2: Company-Specific Terminology Detector
- ▶ Task 3: Clerk Of Meetings
- ▶ Bonus Feature: Rag ChatBot, Text-To-Speech, Meeting Records
- ▶ Metric Evaluation (Task 1)
- ▶ DEMO
- ▶ Frontend Workflow
- ▶ **Insights & QA**



# Insights & QA

**Is it worth finetuning a NER model for the terminology detector?  
What are the pros and cons?**

! Starting on September 15, 2024, you can only customize classification, entity extraction, and sentiment analysis objectives by moving to Vertex AI Gemini prompts and tuning. Training or updating models for Vertex AI AutoML for Text classification, entity extraction, and sentiment analysis objectives will no longer be available. You can

**Some words are hard to form... for NER to detect**

**ET..P      DDR Racial      GK1**

**Accent is annoying**

**Cloud C⃝X⃝(SQL)      Big Curry**

# Contributors



余振揚

NCKU AI Robotics 碩二

莊上緣

NCKU CSIE 碩二

江宇浩

NCKU CSIE 碩一

林志芸

NCKU CSIE 大四

- Team Leader
- AI Backend
- Summarizer
- ChatBot

- Audio Transcriber
- Text to Speech
- Ablation Study

- Term Detector
- RAG Summarizer
- RAG ChatBot

- Web Frontend
- Web Backend

# THANK YOU!

