

# 生成式英文口說情境練習與評分系統

## Generative English speaking scenario practice and scoring system

指導教授：李國川

學生：連思雅、楊宜萍、王呈軒

國立聯合大學 資訊工程學系

苗栗市南勢里聯大2號

[gcleee@nuu.edu.tw](mailto:gcleee@nuu.edu.tw)

### 摘要

本專題開發一款基於大語言模型的生成式英文口說情境練習與評分系統，使用者可以選擇在不同情境下練習英文口說，並進行類似全民英檢的口說測驗，旨在透過語音互動提升學習者英文口說能力、並改善國人不敢面對老外開口的學習障礙，同時也可節省昂貴的英文學習費用。

技術架構採用 Flask 框架，系統整合語音轉文字（STT）技術提供即時語音辨識、運用 GPT-SoVITS 實現客製化語音生成。結合 RAG 檢索英文題庫生成口說測驗、運用 Prompt Engineering 生成不同口說情境練習。最後使用大語言模型（LLM）的 Gemma 3 開源模型對於口說內容進行自動化評分及個人化學習建議。系統提供全民英檢題庫測驗及情境對話練習，支援語音與文字雙模式。透過資料庫記錄學習歷程，生成成績趨勢圖表，並由 LLM 分析歷史數據提供個人化建議。系統採用混合式身分驗證及多輪對話管理機制，確保流暢使用體驗。並提供網頁與手機版兩種介面方便使用者練習。

**關鍵詞：**大語言模型、語音轉文字、英文口說練習、自動評分系統、個人化學習、

### Abstract

This project develops a generative English speaking scenario practice and scoring system based on Large Language Models. Users can practice English speaking in various scenarios and take GEPT-style oral tests, aiming to enhance learners' English speaking proficiency through voice interaction, overcome Taiwanese people's psychological barriers to speaking with foreigners, and reduce expensive English learning costs.

The technical architecture adopts Flask framework. The system integrates Speech-to-Text (STT) technology for real-time speech recognition and utilizes GPT-SoVITS for customized voice generation. It combines RAG retrieval of English question banks to generate speaking tests and employs Prompt Engineering to create diverse speaking scenario practices. Finally, it uses the Gemma 3 open-source Large Language Model (LLM) for automated scoring of spoken content and personalized learning recommendations. The system provides GEPT test banks and scenario-based conversation practice, supporting both voice and text modes. Through database recording of learning progress, it generates performance trend charts, and LLM analyzes historical data to provide personalized suggestions. The system employs hybrid authentication and multi-turn dialogue management

mechanisms to ensure a smooth user experience, offering both web and mobile interfaces for convenient practice.

**Keyword: Large Language Model, Speech-to-Text, English Speaking Practice, Automated Scoring, Personalized Learning**

一、研究背景與目的

構想開發一款基於網頁的英文語音練習評分系統，結合語音轉文字（STT）技術與大語言模型（LLM），以協助學習者透過語音互動方式提升英文口語表達與理解能力。主要功能包含：

（一）即時語音轉文字：整合 STT 技術，將使用者的語音內容即時轉換為文字，提供即時顯示功能，讓使用者能清楚檢視自身口說內容，達成類語音助理的互動體驗。

（二）英文表達智能分析：將轉錄後的文字傳送至 LLM，由模型進行語法、語義及表達品質分析，協助學習者了解自身英文口語的正確性與流暢度，並提供改進建議。

（三）個人化學習回饋：根據 LLM 的分析結果，生成客製化的學習建議與改進方向，幫助學習者強化語言理解與表達能力，並透過模擬真實對話情境，降低開口說英文的心理障礙，提升學習動機與自信心。

二、系統架構

（一）系統架構圖

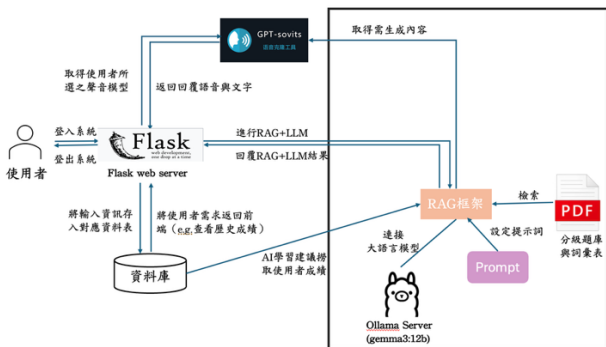


圖 1 系統架構圖

（二）系統架構說明

整體架構透過 Flask 作為核心中樑，串接語音處理、知識檢索、生成式 AI、資料儲存與輸出等功能，使系統能在網頁端提供即時語音互動、內容查詢、自動化生成回答以及輸出。

三、專題理論與實作

（一）RAG 檢索及 LLM Model

本專題之系統基於結構化提示詞設計原則，針對不同功能模組開發專用模板以優化 Gemma 3 模型的輸出品質。

在等級測驗的答題評分功能中，提示詞融合角色定義、評分維度、輸出格式要求與外部知識注入（RAG 檢索），確保模型生成可解析的結構化評分結果。詞彙分級模組根據 CEFR 標準實施詞彙限制約束，透過提示詞反覆強調提高模型遵循率。

情境對話功能則透過角色設定、對話目標與前置對話歷史引導模型維持上下文連貫性；個人化學習建議功能彙整學習者的歷史成績、趨勢與常見錯誤統計，作為數據驅動的提示策略要素。

此設計確保系統輸出兼具結構化、可控性與適應性，並可透過正規表示式與容錯邏輯進行自動解析與驗證。

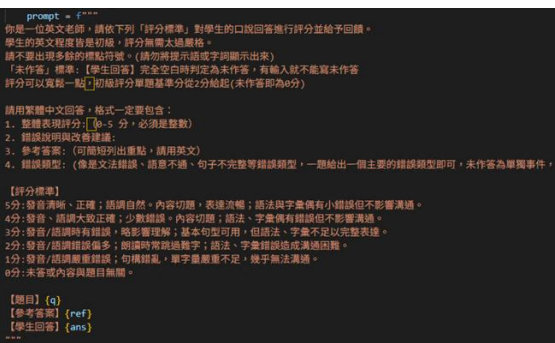


圖 2 LLM Prompt 示例（等級測驗）

（二）GPT-SoVITS 語音生成

採用 GPT-SoVITS 雙階段管線，以 JSON 註冊表（model.json）作為模型資源的單一來源。

系統於啟動時載入模型登錄檔，各筆登錄包含 SoVITS 與 GPT 權重、參考聲音

及提示文字；推理時透過 `_switch_if_needed` 檢查快取權重，必要時經 HTTP 介面動態切換以避免重複載入延遲。語音生成流程由 `tts/tts_question/tts_preview` 組裝推理參數（含目標語言、參考聲音及語氣資訊），呼叫 `tts` 端點產生 wav 並落檔供外部服務。

題庫測驗與情境對話分別於題目顯示與 AI 回覆後同步生成語音。模型訓練採一鍵流程（`_pipeline_oneclick`）：上傳原始音檔後依序進行人聲分離、長靜音清理、切分降噪與 ASR 配對，完成資料準備後訓練 SoVITS 與 GPT 模型。

此設計確保前端「選聲音」、「管理聲音模型」與「預聽」均依賴統一的 `model.json`，實現訓練產出與服務端推理的一致對齊。

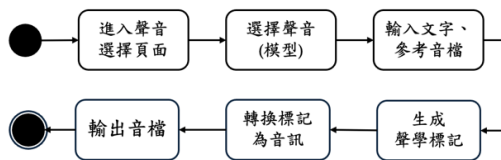


圖 3 語音生成流程圖

### (三) 多輪對話

透過伺服器端會話管理機制克服大語言模型 (LLM) 的無狀態特性，實現連貫互動。系統在 Flask Session 中維護對話歷史陣列，每輪對話以結構化格式「`{"role": "user/assistant", "content": "..."}`」儲存，明確區分對話雙方身分。

推理時，系統將對話歷史轉換為自然語言文本格式「User: [輸入]\nAssistant: [回應]」，併同情境設定資訊（AI 角色、對話目標、詞彙限制）加入提示詞的區塊，使 LLM 在每次推論時能理解完整對話脈絡而非孤立輸入，藉此彌補 LLM 缺乏內部記憶機制的限制。

對話進行中，所有歷史記錄暫存於伺服器端 Flask Session；使用者結束對話時，系統將完整對話歷史、評分結果與元數據整合為 JSON，儲存於 `sessions/` 目錄並在 `chat_scores` 表中記錄檔案路徑，以

`session_id` 作為識別鍵確保多用戶隔離，`user_id` 外鍵關聯則支援個人化學習分析與成績追蹤。此設計透過外部狀態管理補償 LLM 的無狀態設計，實現連貫的多輪情境對話體驗。

### (四) 資料庫串接

採用 MySQL 關聯式資料庫搭配 Flask-MySQLdb 擴充套件實現應用層與資料庫層串接，並採用 JSON 陣列格式儲存部分內容。

資料庫架構包含三個核心表：帳號管理表（`accounts`）儲存使用者基本資訊與登入、練習次數自動累加計數；測驗記錄表（`scores`）記錄文字與語音測驗結果，採用 JSON 陣列格式儲存多個錯誤類型標籤（例如「`文法錯誤`」，「`詞彙錯誤`」），測驗模式欄位區分三類測驗（1 文字、2 語音、3 情境對話）；情境對話記錄表（`chat_scores`）擴展測驗表結構，額外記錄情境類別、難度等級與詳細記錄檔案路徑，該路徑指向儲存完整對話歷史與評分細節的 JSON 檔案。

資料庫連線管理採用連線池機制，由 Flask-MySQLdb 自動處理連線複用與釋放；系統啟用 UTF-8 編碼確保中文資料正確儲存，並設定連線逾時與自動重連機制提升容錯能力，使系統能高效處理併發請求。`user_id` 外鍵關聯確保各筆記錄可追蹤至特定使用者，支援後續個人化學習分析。

### (五) 使用者驗證(Session+JWT)

根據不同平台特性採用差異化之身分驗證策略：網站版採用 Session-Based Authentication，行動應用採用 Token-Based Authentication (JWT)。

網站端 Session 驗證由伺服器端集中管理使用者狀態，登入後伺服器建立包含使用者編號、登入時間與到期時間的 Session，並以加密方式透過 Cookie 回傳，每次 HTTP 請求瀏覽器自動附帶 Cookie，伺服器據此驗證存取權限，便於即時撤銷異常或過期 Session，強化安全控制。

行動應用因需與後端進行頻繁無狀態



API 呼叫，採用 JWT 作為驗證方式，伺服器產生包含使用者編號、權限資訊與過期時間的自包含 Token 並以密鑰簽署，App 儲存於 Secure Storage 並透過 Authorization: Bearer 方式傳遞，伺服器僅需驗證 Token 簽章與有效期，無需查詢 Session 狀態，降低伺服器負擔並提升效能。

此混合式驗證架構兼具 Session 的集中式安全控制與 JWT 的跨平台彈性、低耦合特性，符合系統未來擴展多前端設備與服務規模的需求。

## 四、系統功能呈現與說明

### (一) 題庫等級測驗

搜集全民英檢之題庫（初級、中級），又分為語音模式與文字模式供練習選擇，以問答方式進行測驗練習，使用者可自選等級、題數及作答秒數。最後系統將利用大語言模型進行評分與回饋建議。



圖 4 等級測驗-語音模式

### (二) LLM 自動評分與即時回饋

在等級測驗中，讓 LLM 當「英文老師」的角色，使用者做測驗後給予評分結果，並依照規範之項目給予回饋。



圖 5 測驗回饋

### (三) 情境對話之聊天練習

以聊天的方式，與大語言模型進行對話，並提供情境式及程度分級方便做合適的練習，最後結束對話時，大語言模型會給出此次聊天之分數與回饋建議。使用者亦可於側欄點選先前與 LLM 之聊天記錄，視窗會顯示該次聊天之分數與評語。



圖 6 情境對話

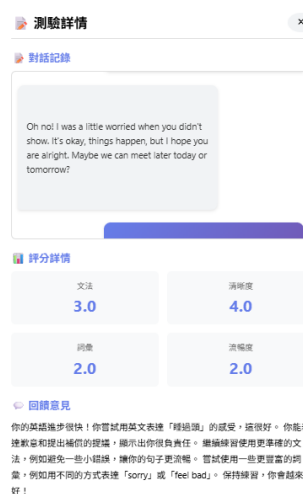


圖 7 情境對話-歷史聊天記錄

#### (四) 管理、新增自訂與選擇聲音

使用者可於管理模型頁面進行模型的刪除、圖片更改、代號更改、注記等操作。若想要自行新增想要的聲音，亦可進入「新增聲音」之頁面，自行加入人聲樣本片段來生成想要的聲音。最後在選擇聲音的畫面中，可以選擇已訓練完成之聲音選項，作為此次測驗之題目播放與對話對象之聲音。

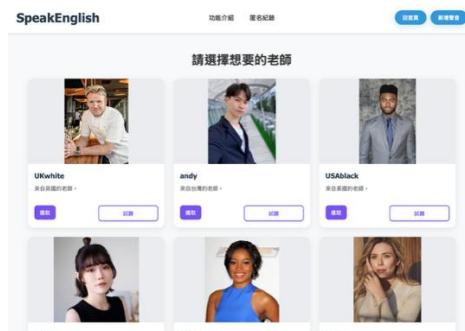


圖 8 選擇聲音頁面

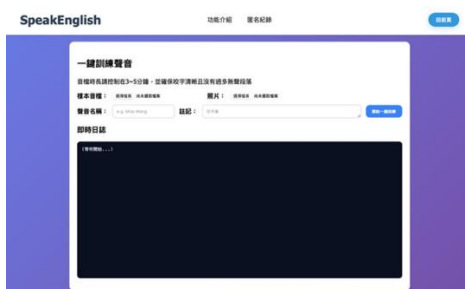


圖 9 使用者新增自訂聲音頁面

#### (五) 個人化 AI 學習建議與成績趨勢圖

利用該帳號之所有歷史成績紀錄之測驗成績與錯誤類別資訊，製成折線圖表，方便查看成績分布趨勢，並使用 LLM 給予個人化之學習建議。

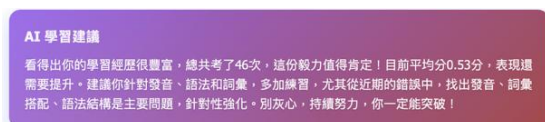


圖 10 個人化 AI 學習建議



圖 11 成績趨勢圖

#### (六) 手機版本

為提供使用者跨平台的英語口說練習體驗，亦製作手機簡易版的英文口說練習系統，其功能包括帳號紀錄、個人化AI學習建議、題庫等級測驗練習與情境對話（含自動化評分與回饋）。

#### 五、專題作品影音平台網址

<https://www.youtube.com/watch?v=qVuvTrvV8I>



#### 六、參考文獻

- [1] 彰化縣人師教育協會, “GEPT Basic 初級口說訓練,” E-video resource, <https://www.twrses.org/E-resources/E-videos/G-EPT-Basic>. Accessed: Apr. 20, 2025.
- [2] AmiGreen, “全民英檢初級口說測驗題庫整理2,” Scribd document, <https://www.scribd.com/document/440794012/%E5%85%A8%E6%B0%91%E8%8B%B1%E6%AA%A2%E5%88%9D%E7%B4%9A%E5%8F%A3%E8%AA%AA%E6%B8%AC%E9%A9%97%E9%A1%8C%E5%BA%AB%E6%95%B4%E7%90%86-2-doc>. Accessed: Aug. 10, 2025.
- [3] Logto, “JWT 與 Session 驗證比較,” Blog post, <https://blog.logto.io/zh-TW/to-ken-based-authentication-vs-session-based-authentication>. Accessed: Nov. 19, 2025.
- [4] Ollama, “gemma3: The current, most capable model that runs on a single GPU,” Ollama model entry, <https://ollama.com/models/gemma3>. Accessed: Sep. 14, 2025.
- [5] RVC Boss, “GPT SoVITS: 1 min voice data can also be used to train a goodTTS model (few shot voice cloning),” GitHub repository, <https://github.com/RVC-Boss/GPT-SoVITS>. Accessed: Jun. 14, 2025.
- [6] Towards Dev, “Convert PDF Tables to JSON Using Python: A Comprehensive Guide,” Blog post, <https://towardsdev.com/convert-pdf-tables-to-json-using-python-a-comprehensive-guide-be119fd16544>. Accessed: Nov. 10, 2025.