

# Chang-Yu Tsai

MA in Linguistics, NCCU  
BA in Chinese Literature, NTPU

+886 932872107  
entropia.changyu@gmail.com  
[linkedin.com/in/changyutsai](https://www.linkedin.com/in/changyutsai)

## CURRENT POSITION

---

### NLP Research and Teaching Assistant

*National Chengchi University*

- Developed NLP pipelines using PyTorch, Tensorflow, pandas, numpy, scikit learn, seaborn, matplotlib, re, and so on.
- Conducted NLP tasks such as classification, sequence labelling, seq2seq, and topic modelling.
- Assessed inter-annotator agreement using Cohen's Kappa.
- Evaluated model prediction using McNemar's test.

## WORK EXPERIENCES

---

### NLP and Linguistics Course Teaching Assistant

*National Chengchi University*

- Led a team of 8 teaching assistants, coordinating tasks and bridging communication with three instructors.
- Designed and conducted lab sessions on Python for NLP tasks.
- Led discussion sessions on the linguistic properties of Taiwan Southern Min and the development of Taiwanese culture.

## PROJECTS

---

### Personalised LINE Chatbot

Apr. 2023 - Present

- Developed a personalised chatbot using Python and integrated it with the LINE Messaging API.
- Implemented a Retrieval-Augmented Generation (RAG) system using FAISS for semantic search and the Gemini API for natural language generation.
- Containerised the application with Docker and deployed it on Google Cloud Run for serverless and scalable operations.
- Handled real-world issues including API key management, memory overuse errors, and Git history rewriting for sensitive file removal.

### Customised GPT "Sin-khu Liah Jit-le"

Oct. 2024 - Present

- Provided two linguistic features for identifying the body part names.
- Used Chain of Thought and In-Context Learning in the instruction.

- Designed and developed a learning-oriented bot on LINE, a major social messaging app widely used in Taiwan.
- Demonstrated the project results and promoted the LINE bot at the 2022 Taiwan IT Month, gaining over 100 users during the event.

## RESEARCH WORKS

---

### Entity Extraction and Semantics Exploration of Body Part Names In Taiwan Southern Min Healthcare Texts

- Built an mBERT-CRF model to extract body part names from Taiwanese texts.
- Analysed radicals and POS tags of the characters within the body part names.
- Improved the prediction with F1 increasing from 87% to 89%.

### Popular Comment Detection on Social Media: Taking Dcard for Example

- Analysed popular comments on Dcard.
- Discovered sarcasm as a significant linguistic trait.

### Pronunciation Variations of Voiced Stops [b] and [g] among Younger Taiwanese Speakers

- Investigated pronunciation variations in Taiwanese using Praat for acoustic analysis.
- Found that devoicing is a common phonetic feature among younger speakers.

## PROGRAMMING & TOOLS

---

### Python & R

- Conducted data preprocessing, model training, and evaluation for NLP tasks.
- Performed statistical analyses in quantitative linguistics contexts.
- Visualised statistical results to support analysis and interpretation.

### Google Cloud Platform

- Deployed Doccano using Compute Engine to support remote collaborative annotation.
- Deployed LINE chatbot using Cloud Run to provide a scalable backend service.

### Docker

- Containerised a personalised chatbot application and built custom Docker images.
- Managed dependencies and optimised memory allocation via container configuration.

### Git & Version Control

- Maintained clean version history using Git for both local and remote repositories.
- Managed GitHub-based collaboration and deployments in personal projects.

# 蔡長祐

國立政治大學語言學研究所碩士  
國立臺北大學中國文學系學士

+886 932872107  
entropia.changyu@gmail.com  
[linkedin.com/in/changyutsai](https://www.linkedin.com/in/changyutsai)

## 現職

### NLP 研究及教學助理

國立政治大學

- 使用 PyTorch、Tensorflow、pandas、numpy、scikit-learn、seaborn、matplotlib、re 等工具建構 NLP 任務流程。
- 執行各類自然語言處理任務，如分類、序列標註、序列轉換 (seq2seq) 與主題建模。
- 採用 Cohen's Kappa 評估人工標註者間的一致性。
- 使用 McNemar's 檢定分析模型預測結果之顯著性差異。

## 工作經驗

### NLP 與語言學課程教學助理

國立政治大學

- 帶領 8 人教學助理團隊，統籌任務分配，並作為三位授課教師與助教團隊之間的溝通橋樑。
- 規劃並執行以 Python 為基礎的 NLP 任務之實作課程。
- 主持討論課，探討台語語言特性及其文化發展脈絡。

## 專案經驗

### 個人化聊天機器人整合專案

Apr. 2023 - Present

- 使用 Python 串接 LINE Messaging API，建構個人專屬聊天機器人。
- 採用 RAG 架構 (Retrieval-Augmented Generation)，結合 FAISS 向量檢索與 Gemini 模型生成回應。
- 將應用容器化 (Docker)，並部署至 GCP Cloud Run，完成無伺服器自動擴展部署流程。
- 處理 API 金鑰管理、記憶體超限錯誤與 Git 歷史清理，具備實務問題排解經驗。

### 客製化 GPT：身軀掠一咧

Oct. 2024 - Present

- 使用部首及詞性標註，用於辨識台語文本中的身體部位。
- 透過 Chain of Thought 及 In-Context Learning 設計 GPT 的後台設置。

### 真的訊息事務所：LINE 機器人討厭詭圖鑑

Aug. 2022 - Nov. 2022

- 設計及開發學習導向之 LINE 機器人，協助用戶識別錯假訊息中的語言表達形式。
- 於 2022 台灣資訊月展示專案結果，展覽期間獲得民眾熱烈迴響，展期三天內即累計超過百名使用者新增為好友。

## 研究成果

---

### 臺灣閩南語健康照護文本的身體部位名稱：實體提取與語意探索

- 搭建 mBERT-CRF 模型，擷取台語文本中的身體部位名稱。
- 分析身體部位名稱中部首及詞性標註的語意訊息。
- 有效提升預測能力，F1 由 87% 提升至 89%。

### 社交媒體熱門留言偵測:以Dcard 為例

- 探討 Dcard 熱門留言的語言使用特性。
- 發現諷刺語氣為吸引互動的重要語言策略。

### 年輕台語使用者的發音變異：以塞音 [b] 及 [g] 為例

- 透過 Praat 分析發音人之聲學數值，觀察年輕世代的發音狀況。
- 結論指出清音化現象（devoicing）在年輕族群中較為普遍。

## 相關技能

---

### Python & R

- 操作自然語言處理任務中的資料前處理、模型訓練與成效評估。
- 進行量化語言學相關的統計分析。
- 視覺化統計結果以輔助資料詮釋與研究解讀。

### Google Cloud Platform

- 使用 Compute Engine 部署 Doccano，支援遠端協作式標註。
- 使用 Cloud Run 部署 LINE 聊天機器人，提供可擴展且無伺服器的後端服務。

### Docker

- 將個人化聊天機器人專案容器化，建立自訂的 Docker 映像檔。
- 透過容器設定管理依賴套件與記憶體資源分配。

### Git & Version Control

- 使用 Git 管理本地與遠端專案版本歷史，維持乾淨的開發紀錄。
- 透過互動式 rebase 與 filter-repo 移除敏感資訊、解決合併衝突。
- 在個人專案中操作 GitHub 進行協作與部署。