

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

2025 斯圖加特緊固件展業務助理

Mar. 2025 - Mar. 2025

金祐昇實業有限公司

- 介紹公司產品並協助攤位營運。
- 以英文與國際訪客溝通，說明產品內容。

NLP 與語言學課程教學助理

Sep. 2021 - Jul. 2024

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

客製化 GPT：身軀掠一咧

Oct. 2024 - Feb. 2025

- 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。

網頁遊戲：討厭詭圖鑑 勇者遠征

Nov. 2022 - Mar. 2023

- 結合故事敘述與遊戲設計，打造沉浸式學習體驗，引導使用者理解錯假訊息的語言模式。

- 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。

以資訊熵衡量中文母語兒童的語言多樣性

- 以資訊熵分析兒童語言多樣性隨年齡的變化。
- 提出資訊熵可有效取代 type-token ratio 評估詞彙豐富度

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。

GCP & Linux CLI

- 透過 Compute Engine 部署 Doccano，提供多人遠端標註工作。
- 設定虛擬機並建立獨立的 Python 虛擬環境。

Praat

- 分析語音聲學資料，探討語音模式與變異現象
- 解讀聲紋圖，比較不同年齡層的語音片段特徵，發現年輕說話者中常見的清音化傾向