Chang-Yu Tsai

MA in Linguistics, NCCU
BA in Chinese Literature, NTPU

+886 932872107
entropia.changyu@gmail.com
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

Sales Assistant

Fastener Fair Stuttgart 2025
- Introduced company products and supported booth operations.
- Communicated with international visitors in English to deliver product overviews.

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

Customised GPT "Sin-khu Liah Jit-le"   Oct. 2024 - Feb. 2025
- Used prompt engineering and in-context learning to identify the target entities.
- Expected to develop the AI healthcare services, combining speech recognition.

Browser game "MisInfoDex: Brave Expedition"   Nov. 2022 - Mar. 2023
- Created an immersive learning experience by integrating storytelling and gameplay to learn language patterns of misinformation.

LINE bot "MisInfoDex"
Aug. 2022 - Nov. 2022

- 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

- Extracted body part names from Taiwan Southern Min healthcare texts.
- Analysed the semantics of body part names.
- Designed the task from preprocessing to evaluation.

Popular Comment Detection on Social Media: Taking Dcard for Example

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

Entropy Approach of Mandarin-Speaking Children's Language Diversity

- Measured language diversity in children at different ages using entropy.
- Suggested entropy as an effective alternative to the type-token ratio for assessing vocabulary richness.

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.

GCP & Linux CLI

- Deployed Doccano using Compute Engine to support remote collaborative annotation.
- Configured virtual machines and set up isolated Python virtual environments.

Praat

- Analysed acoustic phonetic data to investigate sound patterns and variations.
- Interpreted spectrograms to examine segmental features across age groups, identifying devoicing patterns in younger speakers.

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

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

- 運用 Prompt Engineering 與 In-Context Learning 技術，辨識目標實體。
- 預期整合語音辨識技術，開發 AI 健康照護服務。

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。

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

- 從台語健康照護語料中進行命名實體辨識，擷取身體部位名稱。
- 探討身體部位詞在語意上的表現與使用脈絡。
- 獨立設計任務流程，涵蓋資料前處理、語意標註、模型訓練與預測評估。

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

- 以訊息熵分析兒童語言多樣性隨年齡的變化。
- 提出訊息熵可有效取代 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