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

- Built an mBERT-CRF model to extract body part names from Taiwanese 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.