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

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

Personlised 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.

REALBAR: 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 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.