

Chow Siew Yeng

LANGUAGE ENGINEER · MACHINE LEARNING ENGINEER

✉ siewyeng001@e.ntu.edu.sg | 🏠 siewyeng.github.io | 🌐 siewyeng | 📱 Chow Siew Yeng

Experience

Temasek Lab@NTU

Singapore

RESEARCH ASSOCIATE

Sep. 2024 - present

- Designed and trained machine learning models to detect deception based on facial expression data using facial action units.
- Built a scalable, multiprocessing data pipeline for efficient video preprocessing and feature extraction.
- Directed the technical approach for a multi-million-dollar funded research initiative.
- Fine-tuned local LLMs and applied advanced prompt engineering techniques to automate text annotation tasks, replicating human-coded definitions of extremism indicators.
- Participated in qualitative content analysis, gaining practical experience in human annotation workflows.

AI Singapore

Singapore

AI APPRENTICE

Aug. 2023 - May 2024

- Completed intensive course involving topics such as time-series, Neural Networks, NLP and MLOps.
- Developed advanced LLM agents capable of conversing through Time Weighted Memory Retrieval, showcasing innovation in AI conversation systems.
- Successfully deployed these agents on Streamlit, enabling real-time user interactions and enhancing user experience.
- Utilized prompt engineering techniques to steer LLMs towards generating targeted outputs, ensuring alignment with project goals.
- Completed end-to-end Inverse Text Normalization model trained with Pytorch lightning using a Transformer within an Agile environment, deploying inference with FastAPI.
- Made use of Kedro framework in creation of the data pipeline.
- Communicated with the external principal investigator through presentations on project updates every sprint to build an AI solution for the product sponsor.
- Worked on Linux-based code server, making use of MLFlow for experimentation logging.
- Experienced in using docker registry, cloud storage and artifact registry in google cloud platform with cli.
- Wrote detailed technical documentation of code in read the docs format, contributing to knowledge sharing.
- Designed pytest for CI/CD with Github Actions.

Transperfect

Singapore

LANGUAGE ENGINEER

Mar. 2022 - Aug. 2023

- Collaborated with global teams, doing code reviews to improve digital assistant's proficiency in Singapore English and localised user experience.
- Acted as a subject matter expert, creating guidelines that drive locale-specific phonetic transcription for ASR, reducing speech recognition error rates in the region.
- Analysed speech model data to identify utterance misrecognitions and engineer data-driven augmentations to the model, raising end-to-end performance and safety.
- Leveraged Docker for continuous testing, ensuring high quality of training dataset used for new model development.
- Assigned work to and guided a team of 3 data curators, fostering their skills in accurate transcription and data output.
- Utilised Python to increase efficiency of completing day-to-day tasks.

A*STAR (Institute of Higher Performance Computing and Institute of Infocomm Research)

Singapore

INTERN

Jan. 2021 - Mar. 2021

- Trained CNN and MLP models for computer vision (CV) task of mapping facial landmarks to 7 distinct emotions, enhancing online learning tools in the COVID context.
- Researched on Discourse Representation Structure and evaluation benchmarks in Natural Language Inference (NLI), offering linguistic expertise to a query answering project.

Nanyang Technological University

Singapore

TEACHING ASSISTANT

Aug. 2020 - Dec. 2020

- Delivered comprehensive conceptual support in semantics and pragmatics to over 50 undergraduates, enhancing their academic understanding.
- Provided consultations and delivered weekly tutorials on concepts such as logical grammar, set theory and non-classical logics.
- Strategised lesson flow, guided discussion, graded group assignments and promoted understanding of linguistic concepts.

Nanyang Technological University

Singapore

RESEARCH ASSISTANT

Dec. 2017 - Mar. 2021

- Conducted in-depth analysis of over 3000 sentences, primarily focusing on grammatically incorrect sentences, to refine treebanking software for text parsing.
- Supervised adjudication between various pairs of annotators to facilitate inter-annotator agreement and quality of gold data.

Education

Nanyang Technological University

M.A. (RESEARCH) IN COMPUTATIONAL LINGUISTICS

Singapore
Jan. 2020 - Dec. 2021

- GPA: 5.00/5.00.
- Built a rule-based syntactic parser for Singlish.
- Recipient of Research Scholarship.

Nanyang Technological University

B.A. IN LINGUISTICS AND MULTILINGUAL STUDIES

Singapore
Aug. 2015 - Jun. 2019

- Graduated with Honours (Highest Distinction).
- Minor: Mathematics.
- Recipient of College Scholarship.

Skills

Machine Learning	Neural Network, LLM, Pytorch
NLP	NLTK, Regex, Inverse Text Normalisation, Transformers
Programming	Python, React Native, SQL, LaTeX, Bash
Languages	English, Cantonese, Mandarin, German (B2), Japanese (basic)

Publications

Eurali (LREC-COLING)

Torino, Italy
May 2024

THIS WORD MEAN WHAT: CONSTRUCTING A SINGLISH DICTIONARY WITH CHATGPT

- Created new open-source resource for Singlish using ChatGPT through continuous API calls.
- Annotated and analysed findings according to results of annotation.
- Collaborated on dictionary front-end using typescript and svelte.

LREC

Marseille, France
May 2022

SINGLISH WHERE GOT RULES ONE

- Developed a formal grammar of Singlish that can generate syntactic trees and DMRS from input sentences.
- Increased accuracy in parsing Singlish sentences from 46.7% to 70.0%.
- Presented findings and international DELPH-IN summits and LREC 2022.

Courses

2022	Mathematics for Machine Learning Specialization , Coursera
2022	Natural Language Processing with Probabilistic Models , DeepLearning.AI
2022	Natural Language Processing with Classification and Vector Spaces , DeepLearning.AI
2021	11th Lisbon Machine Learning School , Instituto Superior Técnico