

Chow Siew Yeng

LANGUAGE ENGINEER · COMPUTATIONAL LINGUIST

✉ [github-username]dot[first-letter-of-surname] at gmail.com | 🏠 siewyeng.github.io | 🌐 siewyeng | 📄 Chow Siew Yeng

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.

Nanyang Technological University

B.A. IN LINGUISTICS AND MULTILINGUAL STUDIES

Singapore

Aug. 2015 - Jun. 2019

- Graduated with Honours (Highest Distinction).
- Minor: Mathematics.

Scholarships

- | | |
|-----------|--|
| 2015-2019 | College Scholarship , awarded to outstanding freshmen pursuing full-time undergraduate programmes |
| 2020-2021 | NTU Research Scholarship , awarded to outstanding graduate students for research |
| July 2021 | DAAD Scholarship , German summer course for foreign students |

Experience

Temasek Lab@NTU

RESEARCH ASSOCIATE

Singapore

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.
- Guided the technical direction for a multi-million-dollar funded research initiative.
- Used LangChain to prompt tuned local and cloud-based LLMs to automate text annotation tasks.
- Designed text pre-processing algorithms for data analysis and downstream tasks.
- Participated in qualitative content analysis, gaining practical experience in human annotation workflows.

AI Singapore

AI APPRENTICE

Singapore

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 pytestes for CI/CD with Github Actions.

Transperfect

LANGUAGE ENGINEER

Singapore

Mar. 2022 - Aug. 2023

- Collaborated with global teams 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.
- 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.

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.

Worldwide Opportunities on Organic Farms

Galway, Ireland

VOLUNTEER

Aug 2019

- Learnt about sustainable agriculture and living.
- Handled daily farm operations including harvesting and planting.

Lions Befrienders

Singapore

RESEARCH INTERN

May. 2018 - Jul 2018

- Co-developed and executed a health-centric project, promoting active lifestyles among 20 socially isolated elderly individuals.
- Designed, translated, and carried out Mandarin/English questionnaires tracking activity and motivation of the elderly.
- Conducted literature review and research on befriending mechanisms to improve volunteer training.

Skills

Machine Learning	vLLM, LangChain, Transformers, PyTorch, Pandas, NumPy
NLP	NLTK, Regex
Deployment / DevOps	FastAPI, Docker, MLflow, GCP CLI
Programming	Python, React Native, SQL, LaTeX, Svelte
Languages	English, Cantonese, Mandarin, German (B2), Japanese (basic)
Music	ABRSM Music Theory (Grade 8), ABRSM Piano (Grade 8)

Publications

Chow, S. Y., Shin, C. U., & Bond, F. (2024, May). This word mean what: Constructing a Singlish dictionary with ChatGPT. In Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI)@ LREC-COLING 2024 (pp. 41-50).

Chow, S. Y., & Bond, F. (2022, June). Singlish where got rules one? constructing a computational grammar for Singlish. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 5243-5250).

Extracurricular Activity

NTU Welfare Services Club

Singapore

MATHS TUTOR

Sep. 2018 - Oct. 2021

- Mentored ITE students in mathematics weekly.

NTU GEM Club

Singapore

CLUB AMBASSADOR

Aug. 2016 - May. 2019

- Welcomed and guided exchange students around Singapore.
- Buddy to a group of 12 exchange students and arranged group outings to foster bonding and help them feel more comfortable away from home.

NTU Hall 1

MAIN COMMITTEE MEMBER AND INTER-HALL REPRESENTATIVE

- Represented hall in Inter-Hall Cheer Competition and Inter-Hall Recreational Games.
- Planned and executed an annual 3-day camp for over 50 underprivileged children over 4 months of preparation.

Singapore

Aug. 2015 - Jan. 2017

Additional Courses

2023	German B2 , Goethe Institute
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	German Summer Course by Technische Universität Dresden , scholarship awarded by DAAD
2021	11th Lisbon Machine Learning School , Instituto Superior Técnico

Projects

Eurali (LREC-COLING) 2024

Torino, Italy

THIS WORD MEAN WHAT: CONSTRUCTING A SINGLISH DICTIONARY WITH CHATGPT (IN PUBLICATIONS SECTION)

May 2024

- 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 2022

Marseille, France

SINGLISH WHERE GOT RULES ONE (IN PUBLICATIONS SECTION)

May 2022

- 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.

TiMi Studio and Tencent

International (online)

GREEN GAME JAM

Jul. 2021 - Aug 2021

- Top 16 out of 200 teams globally.
- Created an original game to the theme of environmental conservation.

Defence Science and Technology Agency (DSTA)

Singapore

BRAINHACK

Jun 2021

- Top 8 of over 40 participating university teams.
- Implemented neural network and used filtered data, optimising model for Computer Vision challenge.