Liangtao LIN

+65 83159103 | e1127456@u.nus.edu

EDUCATION

National University of Singapore **Singapore**
Master of Science of Data Science and Machine Learning Aug 2023 - Jan 2025

- **Mark:** 4.17/5.0

Sun Yat-sen University **Shenzhen, China**
Bachelor of Engineering in Intelligent Science and Technology Sep 2019 - Jun 2023

- **Mark:** 3.8/4.0
- **Honours:** Second Scholarship (university level, 2019, 2020 and 2021)
- **Courses:** Machine Learning 94(1/138) Deep Learning 89(11/138) Natural Language Processing 91(15/122) Text Data Mining 94(1/98) Computer Vision 89(16/130) Operating Systems 87(20/133) Computer Network 85(23/102)

WORK EXPERIENCE

Tencent Holdings Limited **Shenzhen, China**
Technology Research Intern, IEG-User Platform Department Jul 2022 - Oct 2022

- Expanded the Q&A system based on the NLP pre-trained language models and established Q2Q (query-to-query) and A2Q (answer-to-query) generative model, to solve the cold-start problem of the Q&A system
- Supplemented the FAQ corpus, including generating QA pairs from unstructured documents and expanding similar queries; managed to facilitate the construction of the FAQ system and enlarge the base of FAQs
- Participated in the discussion and exploration of several projects, such as the Game for Peace, multi-round dialogue robots, and the intelligent NPC of games

RESEARCH EXPERIENCE

Research on Large Language Model Alignment **Singapore**
Cooperation Sep 2023 - June 2024

- Found that it is possible to break model defenses simply by appending a space to the end of a model's input. A study of eight open source models demonstrated that this acts as a strong enough attack to cause the majority of models to generate harmful outputs with very high success rates. Underscored the fragile state of current model alignment and encouraged the importance of developing more robust alignment methods
- Proposed a defense against adversarial attacks on LLMs utilizing self-evaluation. Our method requires no model fine-tuning, using pre-trained models to evaluate the inputs and outputs of a generator model, significantly reduces the cost of implementation in comparison to other, tuning based methods, and more resilient to various attacks than existing methods
- Both works have been written into papers and submitted to AAAI 2025

Research on Question-Answer Generation Algorithm Based on FAQ Database Construction **Shenzhen, China**
Individual, Undergraduate Dissertation Feb 2023 - May 2023

- Proposed a QAG pipeline for constructing the FAQ database; a comprehensive method consisting of NER and event extraction is proposed in the answer extraction module, which enhanced the diversity of the generated QA pairs; a novel method that combined MRC and voting mechanism is proposed in the QA pairs ranking module to improve the ranking effectiveness and the overall quality of the generated QA pairs
- The Answer Selection model, Question Generation model and MRC model in the pipeline are all fine-tuned based on the pre-trained language models such as BART, using the Chinese QA dataset collected and constructed by myself
- Experimental results on multiple Chinese QA datasets demonstrated that the generated QA pairs achieved the highest quality and the method performed well in data augmentation for Chinese QA
- Received the award of "Excellent Undergraduate Dissertation" at the faculty level

Research on Answer Selection Based on BERT **Shenzhen, China**
Leader Mar 2022 - Jun 2022

- Researched the topic of Answer Selection, completed the literature review and contribute to Chinese journal
- Ran the open-source code of existing research; adopted the BERT-based answer selection model as a baseline to optimise the method and helped members to process the coding problems
- Delved into the clustering-based classification learning methods, including BERT word vectors, SimCSE, and SCCL clustering models and KMeans and OPTICS clustering algorithms, to solve data problems

Intelligent E-commerce Customer Service Solution Based on Multi-modal QA System **Shenzhen, China**
Leader Feb 2021 - Dec 2021

- Searched relevant materials further to enhance the comprehension of QA systems and multi-round dialogues; organised group meetings and established a data-sharing platform
- Organised data set, conducted the reproduction of generative and retrieval models, and integrated the coding of the two models to generate the optimised QA system, which was highly recognised by the team
- Added voice and image to the system and constructed a new system with a function of understanding the context based on multiple rounds of historical dialogue, generating various replies, and selecting the most suitable reply
- Won the excellent project at “Innovation and Entrepreneurship Training Program for College Students”

Image Caption: A Bilingual Model based on Encoder-Decoder with Attention Mechanism **Shenzhen, China**
Leader May 2021 - Jul 2021

- Reproduced the image caption model code of the Encoder-Decoder with the Attention mechanism in Kelvin Xu's paper; adopted the Beam Search method, modified the classification model in the encoder to VGG, and applied the model to the COCO English dataset and AI Challenger 2017 Chinese dataset respectively to realise the bilingual model
- Improved the model; used the Flask framework to complete the server back-end deployment and developed a front-end WeChat applet to show the effect of the model

Visual Tracking Task and Its Implementation of Traffic Flow Detection Application **Shenzhen, China**
Leader Dec 2021 - Jan 2022

- Undertook in-depth research on visual tracking; analysed the codes of the Yolov3+Sort model and Yolov5+DeepSort model, optimised the Yolov5 model and wrote the actual traffic flow detection program

PUBLICATIONS

Single Character Perturbations Break LLM Alignment **AAAI-AIA 2025**
Accepted

- Co-first author with Hannah Brown, supervised by Michael Shieh, from NUS School of Computing

Self-Evaluation as a Defense Against Adversarial Attacks on LLMs **IJCAI 2025**
Submitted

- Co-first author with Hannah Brown, supervised by Michael Shieh, from NUS School of Computing

EXTRACURRICULAR ACTIVITY

2021 Mathematical Contest in Modeling / MCM **Shenzhen, China**
Leader Feb 2021

- Established a machine learning model to analyse the existing data of 4000 reports concerning the migration of Asian giant hornets; interpreted the possible migration traces of Asian giant hornets and predicted the possible distribution locations
- Accomplished pre-processing, analysis and visualisations of the data, compiled the code of the VGG-based image classification model and the SVM-based location classification model and established the Fuzzy Synthetic Assessment model; won the Finalist Prize

2022 Mathematical Contest in Modeling / MCM **Shenzhen, China**
Leader Feb 2022

- Developed a price prediction, trading decision and risk assessment model that used only the past stream of daily prices to help the trader make daily bitcoin trading decisions
- Accomplished pre-processing, analysis and visualisations of the data, compiled the code of the price prediction model based on the Grey model and LSTM model and the risk assessment model based on the clustering GMM model

SKILLS & INTERESTS

Computer Skills: Python, C, C++, C#, Go, MATLAB, JavaScript, SQL, Pytorch, Unity, AutoCAD, MS Office
Languages: Mandarin (Native), English (Fluent)
Interests: Drum, Photography, Basketball, Martial Art