LI Wenfeng

Email: *liwenfeng20112012@gmail.com*   **HP:** *+65 8342 2901*   **GitHub:** *Gzgylwf*

Skills

Programming & ML Related: Python, Java, Pytorch, Scikit-learn, deepspeed, C/C++
Web Development: FastAPI, Vue
Database: FAISS, MySQL, MangoDB
Language: Mandarin, English

Education

Institute of System Science, National University of Singapore (Singapore)

Master of Technology in Knowledge of Engineering   Jan 2014 - Jun 2015
- GPA 4.19 / 5.0, Top **10%**
- The **Best** FYP project prize

Graduate Diploma of System Analysis   Jul 2012 - Aug 2013
- GPA 4.41 / 5.0, Top **3**

Computer Science and Technology, Harbin Institute of Technology (China)

Bachelor of Computer Science and Technology   Sep 2008 - Jul 2012
- GPA 82.4 / 100.0, Top **10%**

Work Experience

NLP Group, ALI, I2R, A*Star (Singapore)

Lead Research Engineer II (reorganized)   Jul 2023 - Present
- 60% Project / 20% Research / 20% Admin
- Responsible for establishing our local framework and platform for data pipeline, model finetune and evaluation
- Coordinated with the project development team and resources planning and budget management

Senior Research Engineering II (promoted)   Apr 2019 - Jun 2023
- 60% Project / 20% Research / 20% Admin
- Responsible for establishing local data pipeline and data quality control for the projects
- Performed the model training and optimization to achieve the customer required performance
- Assist with project budget and allocation management to ensure the project completed within the time and budget

Senior Research Engineering I (promoted)   Apr 2018 - Mar 2019
- 70% Project / 30% Research
- Responsible for the build of data annotation SOP and created the annotation guideline
- Developed tools for boosting the data annotation and data quality improvement for the projects
- Conducted the experiments for project model performance improvement and interagtion

Research Engineering   Aug 2015 - Mar 2018
- 90% Project / 10% Research
- Assist the scientist to prepare, clear, and annotate the high-quality dataset for model training
- Conducted the error analysis of the varioud model and search for the optimal results

SDE, National University of Singapore (Singapore)

Research Assiastant (Internship)   Aug 2014 - Apr 2015
- Developed an IFC parser to convert the BIM model into a 3D scene with a multi-agent simulator, achieved **best** FYP prize

IT Department, Keppel Land (Singapore)

Software Engineer (Internship)   Feb 2013 - Jul 2013
- Developed and launched 3 internal Android mobile applications enhanced the team connection

Project Experience

LLM Interagrated Pipeline Framework (Core Member)   2023 - Present
- Designed and built the LLM pipeline framework that support the data annotation and model evaluation

Document-level Event Extraction on Multi Domain with LLM (Core Member & Co-PI)   2023 - Present
- Developed and refined the model that increase customer about **4.6** times productivity boost
- Managed the data annotation that provide high-quality data with **90%** data agreement
- Propose and build the annotation platform which boost the **2-3** times speed and enhaced the annotation quality

Multi-turn Dialogue Emotion Analysis (Team Member)   2022 - 2023
- Enhanced the annotation tool for dialogue emotion on utterance-level annotation, reduce about **3** annotation time
- Supported the team to refined the guideline and verify the data quality, achieved **86%** interal data agreement on **5K** data

Customer Satisfaction Analysis of China UNION Hotline Manscription (Core Member & Co-PI)   2021 - 2023
- Coordinated with partners to obtain the **clean** business data from the target customer for model training
- Enhanced the annotation tool functionality with some **automated** validation and **reduce time** for data quality control
- Responsible for overseeing the data annotation and clear process, data achieved **90%** interal data agreement on **5K** data
- Trained and optimized the model which achieved the **86%** F1-score on satisfied and **84%** F1-score on unsatisfied conversation
- Delivered and Interagrated into the customer's platform, boost **4.6** times productivity and able to pre-screen **300K** conversation per day

Customer Satisfaction Analysis of Helpdesk Conversation on Weibo (Core Member & Co-PI)   2019 - 2020
- Created the annotation tool for dialogue data boosting the annotation speed **3** times
- Responsible for the data collection, clear and manage the data annotation process. The data achieved **91%** interal data agreement on **10K** data
- Trained and optimized the model which achieved the **86%** F1-score on 4 category conversations, passed the customer 1st phase review

Machine Reading Comprehension on Yes/No Question Answering (Team Member)   2018 - 2019
- Developed and refined the model which achieved the top **3** performance on the Yes/No leaderboard
- Propose and build an annotation tool for complex text data, which boost abour **2** times annotation speed

Aspect Sentiment Analysis of New Title (Core Member)   2017 - 2018
- Managed the data annotation that provide high-quality data with **90%** data agreement on **30K** titles
- Created the annotation tool for aspect sentiment annotation, which boost about **3** times annotation speed
- Refined the model which achieved **2-3%** F1-score improvement
- Delivered the model to the customer as an enhanced module on **1** products

Sentiment Analysis of Movie Comments (Core Member)   2015 - 2017
- Prepared the guideline and conduct the data annotation, achieved **90%+** average data internal agreement on **90K** movie comments
- Refined the sentiment dictionary and developed a conjection word rules on the baseline model, which get around **7-8%** F1-score improvement
- Deployed the model on the customer's platform which satisfied **8** domains over **3** products

Certifications

Deep Learning Specialization by DeepLearning.AI Coursera. No.BDQ6P9W84R4VU   May 2019
Google Cloud Big Data and Machine Learning Fundamentals Coursera. No.3C5P9W8A53NW   Jan 2022
Introduction to TensorFlow Coursera. No.T6K3GLP5XUCQ   Mar 2022