

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:** MangoDB, FAISS, MySQL
- Language:** Mandarin, English

Education

Institute of System Science, National University of Singapore (Singapore) <i>Master of Technology in Knowledge of Engineering</i> <ul style="list-style-type: none">GPA 4.19 / 5.0, Top 10%The Best FYP project prize <i>Graduate Diploma of System Analysis</i> <ul style="list-style-type: none">GPA 4.41 / 5.0, Top 3	Jan 2014 - Jun 2015
Computer Science and Technology, Harbin Institute of Technology (China) <i>Bachelor of Computer Science and Technology</i> <ul style="list-style-type: none">GPA 82.4 / 100.0, Top 10%	Sep 2008 - Jul 2012

Work Experience

NLP Group, ALI, I2R, A*Star (Singapore) <i>Lead Research Engineer II</i> <ul style="list-style-type: none">60% Project / 20% Research / 20% AdminResponsible for establishing our LLM framework and platform to integrated our data pipeline, models and resources cross projectsCoordinated with the project management team and assist the resources planning and budget management ensure the project successfully completed <i>Senior Research Engineering II</i> <ul style="list-style-type: none">60% Project / 20% Research / 20% AdminResponsible for establishing data pipeline and data quality control for integrated the resource cross projectsConducted the experiments of model training, finetune and integration to achieve the customer requirementAssisted with project budget and allocation management to ensure the project completed within the time and budget <i>Senior Research Engineering I</i> <ul style="list-style-type: none">70% Project / 30% ResearchResponsible for standardizing the data annotation SOP and overseeing the data annotation processDeveloped tools for boosting the data annotation and data quality improvement for the projectsConducted the experiments of model training, finetune and integration to achieve the customer requirement <i>Research Engineering</i> <ul style="list-style-type: none">90% Project / 10% ResearchAssisted the scientists to collect, prepare, clear and provide high-quality annotated data for the projectsConducted the error analysis of our model and optimized the performance to achieve the customer requirement	Jul 2023 - Present
SDE, National University of Singapore (Singapore) <i>Research Assistant (Internship)</i> <ul style="list-style-type: none">Developed an IFC parser to convert the BIM model into a 3D scene with a multi-agent simulator, achieved best FYP prize	Apr 2019 - Jun 2023
IT Department, Keppel Land (Singapore) <i>Software Engineer (Internship)</i> <ul style="list-style-type: none">Developed and launched 3 internal Android mobile applications enhanced the team connection	Apr 2018 - Mar 2019
	Aug 2015 - Mar 2018
	Aug 2014 - Apr 2015
	Feb 2013 - Jul 2013

Project Experience

LLM Integrated Pipeline Framework (Core Member & Proposer) <ul style="list-style-type: none">Designed and implemented the LLM framework APIs centralizing our data pipeline, prompt design and model finetune and evaluation, which improve our knowledge sharing.Developed a web interface allow end-users easily to access and manage resources with UI and easy to deployed on local GPU as a on-premise LLM platform	2023 - Present
Document-level Event Extraction on Multi Domain with LLM (Core Member & Co-PI) <ul style="list-style-type: none">Developed and refined the model that increase customer about 4.6 times productivity boostManaged the data annotation that provide high-quality data with 90% data agreementPropose and build the annotation platform which boost the 2-3 times speed and enhanced the annotation quality	2023 - Present
Multi-turn Dialogue Emotion Analysis (Team Member) <ul style="list-style-type: none">Enhanced the annotation tool for dialogue emotion on utterance-level annotation, reduce about 3 annotation timeSupported the team to refine the guideline and verify the data quality, achieved 86% internal data agreement on 5K data	2022 - 2023
Customer Satisfaction Analysis of China UNION Hotline Manuscript (Core Member & Co-PI) <ul style="list-style-type: none">Coordinated with partners to obtain the clean business data from the target customer for model trainingEnhanced the annotation tool functionality with some automated validation and reduce time for data quality controlResponsible for overseeing the data annotation and clear process, data achieved 90% internal data agreement on 5K dataTrained and optimized the model which achieved the 86% F1-score on satisfied and 84% F1-score on unsatisfied conversationDelivered and integrated into the customer's platform, boost 4.6 times productivity and able to pre-screen 300K conversation per day	2021 - 2023
Customer Satisfaction Analysis of Helpdesk Conversation on Weibo (Core Member & Co-PI) <ul style="list-style-type: none">Created the annotation tool for dialogue data boosting the annotation speed 3 timesResponsible for the data collection, clear and manage the data annotation process. The data achieved 91% internal data agreement on 10K dataTrained and optimized the model which achieved the 86% F1-score on 4 category conversations, passed the customer 1st phase review	2019 - 2020
Machine Reading Comprehension on Yes/No Question Answering (Team Member) <ul style="list-style-type: none">Developed and refined the model which achieved the top 3 performance on the Yes/No leaderboardPropose and build an annotation tool for complex text data, which boost about 2 times annotation speed	2018 - 2019
Aspect Sentiment Analysis of New Title (Core Member) <ul style="list-style-type: none">Managed the data annotation that provide high-quality data with 90% data agreement on 30K titlesCreated the annotation tool for aspect sentiment annotation, which boost about 3 times annotation speedRefined the model which achieved 2-3% F1-score improvementDelivered the model to the customer as an enhanced module on 1 products	2017 - 2018
Sentiment Analysis of Movie Comments (Core Member) <ul style="list-style-type: none">Prepared the guideline and conduct the data annotation, achieved 90%+ average data internal agreement on 90K movie commentsRefined the sentiment dictionary and developed a conjunction rules module on the baseline model, which get around 7-8% F1-score improvementDeployed the model on the customer's platform which satisfied 8 domains over 3 products	2015 - 2017

Certifications

Deep Learning Specialization by DeepLearning.AI Coursera. No.BDQ6PV84R4VU	May 2019
Google Cloud Big Data and Machine Learning Fundamentals Coursera. No.3C5P9W8A53NW	Jan 2022
Introduction to TensorFlow Coursera. No.T6K3GLP5XUCQ	Mar 2022