Imperial College
London

Yuyao Jiang

☎ +86 19921331454 ✉ yuyao.jiang22@gmail.com

Education Experience

Imperial College London **Applied Computational Science and Engineering** Oct 2022 – Nov 2023

- Master's degree, Merit
- Core courses: Machine Learning, Deep Learning, Python, C++, Parallel Programming, Mathematical Modeling and Numerical Methods, Convex Optimization

Nanjing University of Science and Technology , University of Reading **Mathematics and Applied Mathematics** Sep 2018 – June 2022

- Bachelor's Degree Dual Degree, Top 10% in Major, First Class Degree
- Core courses: probability theory and probability statistics, mathematical analysis, advanced algebra, abstract algebra, numerical analysis, fluid mechanics

Internship

- **Zhongdian Hongxin Information Technology Co.** **Natural Language Processing Intern** June 2023 – Sep 2023
  - Main job: Responsible for the research and development of NLP algorithms, including data processing, model training and optimization, and model evaluation, etc., to identify and label entity types or extract relationships between different entities from various types of text data.
  - Technical Points: Project I uses two methods (1.training BiLSTM+CRF model 2.fine-tuning the address structured element parsing model of the Moda platform) to identify and label the entity types in address information. Project 2 adopts the Universal Information Extraction model to extract entities and relationships of arbitrary text information, and the difficulty lies in processing data, adjusting labels and extracting features from the business perspective. The model is encapsulated and can be used to realize the inference function by calling the API through post requests.
  - Output : **Address Entity Recognition Model** and **Appointment Information Extraction Model** were adopted by the company's Business Acceptance System and Human Resource Management System respectively, and the accuracy rate of marking address entities and identifying the extraction fields of appointment documents reached more than 96%.
- Finvolution **Large Language Models Intern** Dec 2023 – Feb 2024
  - Main job: 1.Based on the company' s business data, used multiple frameworks (TRL, LLaMA-Factory, Firefly, FastChat) performed SFT finetuning and DPO training to achieve multi-round dialogue reasoning and deploy large customer service models 2. Use large models for precise dialogue management, and use RAG to realize speech selection and speech generation for intelligent customer service in the preset speech library.
  - Technical Points: Fine-tuning, Quantization, RerankerModel, RLHF, API and web interfaces for deploying large model inference.
  - Output: 1. Conduct a large number of fine-tuning experiments and adjust learning rate, lora rank and other hyper-parameters, and launch the best customer service multi-round dialogue model in production environment 2. Based on the RAG idea, independently write the code for dialogue management, and using BCEmbedding model to compare user input and Q&A pairs in vocabulary library for similarity, recall the top-k Q&A pairs, and call the API of the large language models to select or generate answers.

Project experience

- **"Probability and statistics" Project** **University of Oxford** June 2021 – Nov 2021
  - Main work: try different machine learning algorithms on selected datasets, and finally select the random forest and cross-validation algorithms to extract the key influencing factors with larger weights, to further

  realize and improve the prediction accuracy of the multiple linear regression algorithm.
  - Difficulty: selecting machine learning algorithms to reduce the number of features (32) based on different test metrics (residuals, standard deviation, R-squared, p-value, etc.)
  - Outputs: 84.18/100 for project completion, resulting in two papers, ”Random Forest Regression for Predicting Student Performance” and ”Statistical Analysis of Several Factors in Predicting Student Performance”

- **Filtering, Projection and Slicing of Images using C++** **Imperial College London** March 2023 – May 2023
  - Main tasks: implementation of various filters such as color correction, image blurring, edge detection, average/maximum/minimum intensity projection and slicing of 3D images.
  - Difficulty: Optimize algorithms to improve processing efficiency, such as the use of separated Gaussian filters instead of Gaussian filters, decompose a two-dimensional Gaussian kernel function into two one-dimensional kernel functions in the horizontal direction and vertical direction, reduce the computational cost while ensuring the filtering effect.
  - Output: Packaged executable that renders the input image according to the user's multiple needs. 34% reduction in processing time before and after algorithm optimization.

Publication of papers

- **A study of convolutional neural network algorithms** Published in the Journal of Arts and Sciences Navigation in April 2018 on CNN basic framework, implementation principles and application scenarios.
- Statistical Analysis of Several Factors in Predicting Student Performance Published as CSAMCS 2021 conference paper in November 2021 in SPIE journal, retrieved by EI, Scopus.

Skills

- Mastery of C++, Python, proficiency in deep learning frameworks like PyTorch and PaddlePaddle, LLMs training and inference frameworks like TRL, LLaMA-Factory.
- Familiarity with traditional NLP tasks like entity recognition and information extraction, with practical experience in the application of LLMs fine-tuning technology, customer service models, multi-round dialogue management and other downstream tasks.

Awards

- 2020 National College Students' English Competition Third Prize, IELTS 6.5, PTE 71, College Students' English Level 4 and Level 6 Passed
- School-level third-rate student and school-level first-rate scholarship for several times
- School Class of 2022 Outstanding Graduates