Tianyao (Alan) Su

Los Angeles, California 90007 | (213) 301-2764 | tianyaos@usc.edu | linkedin.com/in/tianyao-su-293254325/

Basic Skills

- Programing: C++, Python, Pytorch, Tensorflow, Keras, HTML, CSS, Git
- AI Knowledge: Machine Learning, Deep Learning, Computer Vision, Reinforcement Learning, NLP, LLMs
- Database: MySQL, MongoDB, Elasticsearch, Neo4j

EDUCATION

University of Southern California August 2024-June 2026
Master of Science in Computer Science (AI)

Shanghai University September 2020-July 2024
Bachelor of Engineering in Computer Science (AI), ***Shanghai University 2024 Outstanding Graduate***

- Cumulative GPA of 3.70/4.0, departmental GPA of 3.86/4.0. Ranked in top 3% during undergraduate studies, third in CS major

WORK EXPERIENCE

Large Language Model Development Intern January 2024-April 2024
Faqianyan Inc., Shanghai, China

- Worked in a team responsible for legal assistance Large Language Model development based on ChatGPT 4.0 API Key
- Append RRF algorithm to a basic BGE_reranker. Trimmed input tokens with prompt mapping table (***Prompt Engineering***) for meaning completion. Finally improved LLM RAG (***Retrieval-Augmented Generation***) reranking top 10 recall rate from 68% to 100%
- Connected AI Lawyer LLM retrieval module to Alibaba Elasticsearch pool (***Cloud Platform***), increasing data retrieval speed by 130%
- Designed data cleaning (***Python***) and database importation (***MySQL***) algorithms for 130 million legal html documents (***Data Cleaning***)

PROJECTS

Multi-Model Integration for Marine Meteorological Forecasts November 2023-June 2024

- Completed a sea fog forecasting system (***Time Series Analysis***) for next 48 hours within a one-square-kilometer area by hydro-meteorological factors and visibility data for Pacific region obtained from China Oceanic Administration (***Meteorological Analysis***).
- Combined Informer, Temporal Fusion Transformer, DeepVAR, and N-BEATS models (***Deep learning***) for time series forecasting output of meteorological factors by a weight matrix (***Ensemble Learning***). Decreased prediction loss by 40%, compared to single model predictions
- Pioneeringly used Graph WaveNet (***Graph Neural Network***) to extract spatio-temporal dependencies between locations, fitting residuals of time series forecasting outputs (***Spatio-temporal Analysis***), further reduced meteorological factor outputs loss by 20%
- Integrated XGBoost and LightGBM (***Machine Learning***) to predict whether sea fog will take place within 48 hours, reached an accuracy of 85% and an ROC area of 0.93 on test set, compared to an accuracy of 75% in previous study (***Pattern Recognition***)

Research on Multimodal Product Recommendation System June 2023-August 2023

- Designed a multimodal fusion product recommendation framework based on Item2Vec, LightGCN, and CLIP using relevant product data from 1data Technology Shanghai, including views, product types, product photos, etc.
- Implemented LightGCN to extract product relationships from the original user-item collaborative filtering matrix, analyzing the most relevant connections between products (***Data Mining***)
- Employed CLIP to transform product images into latent vector representations (***Computer Vision***). Combined CLIP ouputs with other information and processed through a Variational Autoencoder to obtain a comprehensive distribution of product information, achieving data abstraction and dimensionality reduction
- Used Item2Vec to build a product recommendation system, utilizing negative sampling for continuous generation of product recommendations, achieved an accuracy of 70% and a recall rate of 77% on the test set (***Recommendation System***)

Offshore Areas Ulva Prolifera Identification with Remote Sensing Images November 2022-January 2023

- Constructed an Ulva Prolifera identification system, including optical chlorophyll detection, disaster assessment, and time series analysis
- Obtain ocean remote sensing images from Modis L1B 500m satellite data. Extracted Ulva Prolifera distribution with NDVI and RVI optical detection method. Normalized gray values to analyze coverage boundaries and Ulva Prolifera severity (***Remote Sensing Data Processing***)
- Analyzed Ulva Prolifera growth trend by Informer neural network with chlorophyll and temperature data obtained from NASA. Achieved over 75% accuracy of Ulva Prolifera occurring within 10 days

Shanghai University Course Selection Mock System June 2022-August 2022

- Promoted development of a student course selection mock system for Shanghai University (***Collaboration***)
- Develop course query APIs to send requests to SHU academic office, returning course credits, professor information, timetable, etc. from 5000 on campus courses (***Back End***) before semester starts