SHEN, Yuyang

Shenzhen, Guangdong Prov., P.R. China
224020343@link.cuhk.edu.cn / yuyangshen889@gmail.com
(+86)18674805326
Personal Page | Github | ORCID

Education

The Chinese University of Hong Kong, Shenzhen 2024.08 – 2026.08
Shenzhen Finance Institute
M.Sc. in Information Management and Business Analytics (BA Track)

- Courses: Machine Learning for Business, Fintech and Applications, Programming for Business Intelligence, AI For Decision-Making, GenAI and Business Applications, Blockchain Technology and Applications, Economic Model Analysis

Southwestern University of Finance and Economics (SWUFE) 2020.09 – 2024.06
School of Finance
Bachelor of Finance

- GPA: 3.65
- Courses: Investment, Corporate Finance, Monetary Finance, Probability Theory, Statistics, Econometrics, Python Programming, Data Mining, Data Visualization, Algorithm Analysis, Database

Publications

A Theoretical Framework About Deep Reinforcement Learning In Trading - Agent As A Predictor
Working Paper (2025)

- **Abstract:** The application of Deep Reinforcement Learning (DRL) in quantitative trading faces key challenges. The Efficient Market Hypothesis (EMH) questions the validity of technical analysis, which DRL in trading relies on. There's also a lack of standardized frameworks for DRL model definition. Finally, most studies rely on backtesting without rigorous evaluation. This study addresses these issues by rethinking EMH, framing DRL as a prediction task, and introducing new evaluation methods, and completes the theoretical development, experimental design and execution, and assumption validation.
  - Key Words: Portfolio Management, Deep Reinforcement Learning, Quantitative Trading, EMH, DQN, PPO, DGPO, DDPG, RLHF, Actor-Critic, Hypothesis Testing
  - Innovations: Introduces theoretical frameworks long recognized in traditional financial theory, making slight adjustments to its underlying assumptions, and reexamines the theoretical foundations of DRL in finance. Explores the essence of how automated trading agents generate returns, leading to adjustments in the model definition.
  - Area: AI In Decision Science

Research Experience

Peking University HSBC Business School 2024.12 – Present
Research Assistant

- **Project I: Web-Scraping and Econometrics**
  - **Web Scraping:** Combined entity names from a specific industry dataset with market participation "activity" cue words to form targeted keywords. Automatically scraped relevant Google News data via *Selenium* and *requests*, thereby constructing an "entity activity" variable to indicate whether market participants were active.
  - **Econometric Analysis:** Following reference literature, employed a panel regression framework to investigate whether the activity variable had a statistically significant impact on the chosen dependent variable.
- **Project II: NLP (Contextual Understanding and Emotion Analysis)**
  - **Data Collection:** Focused on data from an overseas e-commerce platform spanning multiple industries, enabling richer insights into diverse textual feedback and user interactions.
  - **Model Development:** Leveraged LoRA technology to fine-tune a *BERT* model for multi-label emotion classification using *BCEWithLogitsLoss*. Deployed the model for real-time emotion prediction, thereby

enhancing performance in processing and analyzing textual data for deeper contextual and emotional insights.

CBIT Chain Intelligence Technology Center, CUHK-Shenzhen 2024.09 – 2024.11
Research Assistant

- **Finance Research LLM Project:**
  - **Objective:** Developed an automated large language model (LLM) system to precisely evaluate finance research papers, focusing on theoretical vs. empirical contributions, innovation, and other quality indicators.
  - **Divide-and-Conquer Framework:** Deconstructed the concept of a "high-quality" finance paper, designing a structured question-flow framework to guide the LLM's paper assessment.
  - **Model Workflow Design:** Implemented an LLM-As-A-Judge approach by fine-tuning multiple LLMs in parallel. Combined their outputs via majority voting to improve the overall reliability and accuracy of evaluation results.

Projects

Fintechathon International Fintech Contest - AI Track 2024.10 – 2024.12
Top 10 in AI Track

- **Federal Learning:** Developed an industrial equipment failure prediction and financial risk control system using the FATE federated learning framework, covering design, tuning, and analysis, with key components like pipeline, reader, and heterxgboost.
- **System Development:** Built and integrated a federated learning system for press machine quality monitoring using HTML, CSS, JavaScript, Python, and SQL, achieving a top 10 ranking.
- **Project Links:** Github | External Access | China Access

Multiple Automated Web-Scraping Projects 2023.11-2024.01

- **Automated Literature Retrieval:** Engineered Python-based scraping scripts to continuously discover and aggregate relevant research papers in real time, integrating keyword matching and deduplication logic for data quality.
- **WeChat Auto-Responder Bot:** Designed a conversational agent that employs *requests* and *Selenium* for real-time monitoring of incoming messages and automated replies, dramatically improving user engagement.
- **Institutional Admissions Data Mining:** Deployed headless browser techniques to systematically capture and parse admissions statistics from multiple academic institutions, enabling large-scale data-driven insights on enrollment trends.

Happy Companion AI - ByteDance Mars Code Competition 2024.12-2024.01
Frontend Development, Workflow Design

- **AI Platform and Workflow Design:** Developed a voice interaction platform with dynamic workflows for service classification and switching, enhancing user experience.
- **Frontend Development:** Created a multi-device accessible, user-friendly interface.
- **Project Links:** Github | Access

Fintech Modeling Contest 2022.11 – 2022.12
Third Prize (Provincial Level)

- **Machine Learning:** Developed a risk control model using de-identified bank credit data with default rate as the target. After preprocessing and feature engineering, a stacking method combining XGBoost and CatBoost achieved an AUC of 0.84.

Work Experience

Ping An Technology 2024.11– 2025.01
Complex Information Game Theory Model Platform Group - AI Development and
Data Analyst Intern

- **LLM Development:** Participated in a government-funded initiative dedicated to designing and implementing advanced large language models for financial and macroeconomic analysis. This involved developing robust data pipelines, refining training methodologies, and conducting rigorous evaluations to ensure domain-specific performance and accuracy.
- **Knowledge Base:** Analyze the impact pathways and interrelationships between economic indicators in financial

systems and macroeconomics. This involves creating a knowledge base to train large language models, enhancing their ability to predict economic trends, market behavior, and policy effects.

Kuaishou Technology (Beijing) 2024.02 – 2024.05
KSIB - Data Analyst Intern

- **Data Mining:** Used KwaiBI and SQL for data extraction, Python for machine learning and visualization. Developed video scoring metrics, analyzed sales factors with logistic regression and XGBoost, and created charts for video duration vs. playback time to inform KPIs like CVR. Segmented buyer behavior into touchpoints and analyzed engagement and conversion metrics with funnel charts.

Skills and Awards

- **Awards:** Merit Student (2023.09); Second-class Academic Scholarship (SWUFE Top 15%) (2021.03)
- **English:** IELTS: 6.5 (6); CET-6: 552
- **Programming:** Advanced Python (Data Analysis, Machine Learning, Deep Learning, Web Scraping, Natural Language Processing), Front-end Development (HTML, JavaScript, CSS)