Yaoning(Kevin) Yu

yyn030600@gmail.com | +1 2064842304 | LinkedIn

EDUCATIONS

University of Chicago | Applied Data Science *Sep 2023 – Dec 2024*

- Core Coursework: NLP(Natural Language Processing), CV(Computer Vision) and Deep Learning, Data Mining, Big Data, Time Series, Advanced ML(Machine Learning) and AI, DS(Data Science) in Healthcare

University of Washington | Mathematics *Sep 2019 – Aug 2022*

PUBLICATIONS/PREPRINTS

[1] Yaoning Yu, Kai wei, Ye, Yu, Haohan Wang, Haojing Luo(2025). **Data-Augmented Prompt Optimization for LLMs.** Under Review for *ACL 2025*.

EXPERIENCES

***Research Assistant* | University of Illinois Urbana-Champaign | Champaign, IL** *April 2024 – Present*

In the area of AI Automation under the supervision of Dr. Wang | Due to client confidentiality, our code is privately protected on GitHub.

- Led the development of an advanced application using **python**(PyQt, pdf2image, pdfplumber, docx, ReportLab, OCR, etc) and **LLMs(Large Language Models)** for underwriting file classification and interpretation, employing **prompt tuning and customized privacy-preserving methods** by LLM(Large Language Models) **APIs Gemini and OpenAI.**
- Spearheaded the automation of final underwriting reports, including generating detailed recommendations, plan comparisons, and interest rate calculations using **LLM-agents**.
- Designed a **self-prompt learning loop as a data augmentation method** with LLM(Large Language Model) multi-agent systems**, achieving 96% accuracy in underwriting classification tasks despite limited data** (20 clients).

***Industry Data Analyst Intern* | Guotai Junan Securities | Shanghai, China** *Jan 2023 – April 2023*

- Developed **financial valuation models** for company financial analysis using **Excel**. Models incorporated total revenue, product-specific revenue, gross profit margin, product sales volume, pricing, and market share data to assess company performance
- Created a Power Conversion System (PCS) **database in Excel**, sourcing information from company annual reports, prospectuses, and industry reports. This database served as a valuable resource for tracking and analyzing PCS-related data
- Authored **industry research reports** covering microinverters, Power Conversion Systems (PCS), and PV ribbon, in factors across different areas of PCS such as functionality and classification.
- Produced detailed analyses of various microinverter companies, evaluating and comparing based on their historical performance, product offerings, etc.

***Data Science Intern* | Quant Investment | Shanghai, China** *Sep 2022 – Jan 2023*

- Analyzed financial datasets, including trading data, position data, and daily stock market data, using **Python(pandas and numpy)** to calculate daily and monthly profits for each account based on their invested stocks
- Developed highly efficient and automated Python programs for daily web scraping of financial websites using **Python (mainly Selenium, Requests, and BeautifulSoup)**
- Created data visualizations and analysis reports using various types of plots to effectively convey insights by **Python (mainly Seaborn and Matplotlib)**
- Researched the applications of **futures tick data** and its relevance to market behavior, contributing to a better understanding of financial markets

***Research Assistant* | University of Washington | Seattle, WA** *April 2022 – Oct 2022*

In topics of Ocean Dynamic under the supervision of Dr. Manucharyan | Research Website | Github

- Transformed extensive datasets from 1978 to 2021, including ice concentration, sea ice velocity, ocean velocity, and atmospheric wind velocity, into the Ease Grid format using **Python (Pandas and Numpy)**
- Created visual representations of these datasets in Ease Grid format to verify data alignment with the original datasets by Python
- Utilized Python for data interpolation and ML(Machine Learning) modeling, employing techniques such as **Linear Regression and MLP regression**, to estimate deep ocean activity (velocity) with interpolated datasets

RELEVANT PROJECTS

NYC taxi demand prediction for different time *Autumn 2023*

- Using NYC Yellow Taxi data spanning from 2009 to 2023, which encompasses 13.6 billion records, we built and **fine-tuned ML(Machine Learning) regression models** achieving up to 85% accuracy. The objectives were twofold: 1) to estimate taxi demand across different locations and times, and 2) to predict the duration of each trip, employing PySpark for these analyses.

Natural Language Prcessing and Cognitive Computing (Individual) *Winter 2024*

- Applied Transformer, topic modelling (**LDA model with ktrain and BERTopic**), Zero-shot (NLI) modelling, sentimental analysis and classification (**CNN/Pooled Bi-Directional GRU**), entity identification (**SpaCy/NLTK**) in order to provide actionable recommendations on what can be done with AI to automate the jobs and improve employee productivity based on collection of ~200k news articles.

TradingHero AI Application *Spring 2024*

- Building a comprehensive **Conversational AI Streamlit application** designed to empower traders and investors with advanced tools for stock market analysis through vertex AI and GCP. The application integrates a wide range of functionalities, including market status updates, historical data visualization, stock recommendations, fintuned **FinBERT** for news sentiment analysis, advanced nueral network with Facebook Prophet model for technical trend analysis, and AI-driven insights. With a focus on the US stock market, Trading Hero aims to provide users with actionable insights to inform their trading decisions (by python implementing **streamlit, streamlit_plotly_events, yfinance, finnhub-python, pandas, numpy, google.generativeai, prophet, scikit-learn, Pytorch, transformers, chardet**).

SKILLS & LANGUAGES

Technical Skills: Python (Scikit-learn, Numpy, Pandas, Scipy, Matplotlib, Seaborn, Pytorch, Cartopy, TensorFlow, Beautifusoup, Request), MATLAB, Java, R, SQL, Spark (ML Rgression, ML Classification), LLMs(Large Language Models), LLM-Agents

Others: Latex, Markdown; MS Office; AWS, Google Cloud Platform; Cloud Computing, Tableau; SOLIDWORKS, Google VertexAI

Interests: Music, Reading, Films, Photographs