Yeshuai Cui

Email: cuiyeshuai@gmail.com Mobile: 07927 697698

CAREER

Citi   **London, Unitied Kingdom**
Machine Learning Engineer   08/2024-

- **Short-text Thematic analysis solution**: Designed, implemented and deployed a pipeline that utilizing a hybrid feature extraction algo that combines embeddings, TF-IDF. Utilized K-Means++ for real time response.

- **Multi-Agent solution to identify operational risks:** Designed, implemented and deployed a pipeline that includes: 1. Chat-based HITL plan generator 2. Multi-Agent Execution pipeline that has data-loading, python, semantic, and report agent.

- **Multiple Prompt-based Solution end to end**: Designed, implemented and deployed multiple prompt based solutions include multi-turn generation with Human-In-The-Loop.

Full Stack Software Engineer   09/2023-08/2024

- **BE Engineer**: Implemented **RestAPI** Endpoints using **Java SpringBoot** and **JDBC**, implemented **Oracle SQL** Consumption layer **tables** & **Materialized Views** for data query. Primarily developed the **distributed caching layer** using **Couchbase**, including custom solutions for data refreshing, eviction, and concurrency control to efficiently handle parallel database read/write operations. This improves response time by 10x. Introduced Key Design Decisions to ensure **time consistency** across FE & BE.

- **FE Engineer**: Implemented webpage components using **React**, handling **data fetching** / data mocking, **state management** of highly correlated states. Optimized development and testing efficiency by **encapsulating** common utility code, reducing redundancy by 5x.

EDUCATION

Imperial College London   **London, United Kingdom**
MSc Advanced Computing, Distinction   09/2022 – 09/2023

King's College London   **London, United Kingdom**
BSc Computer Science, First-Class Honours   09/2019 – 06/2022

Pennon Education   **Shandong, China**
A-level   09/2016 – 06/2019
Courses: Maths: A*, Further Maths: A*, Physics: A*. Each single unit passed with at least 90%.

EXTRACURRICULAR EXPERIENCE

Microsoft   **Suzhou, China**
Software Engineer Intern, Bing   06/2021 – 09/2021

- Conducted **data processing/feature analysis** using Aether pipeline, wrote Scope(**SQL**-like) and **C#** code to extract time information from data source in order to calculate document ages then calculate the distribution of document ages in time buckets.

- Simulated the scorecard offline and created modules that calculates CTR and similar metrics of fresh documents.

- Evaluated MEB model and revealed it was not fresh-fair, the model biased against more recent documents and documents came from fresh tier.

- **Trained the model** to give even scores for documents from different time and discovered possible reason for the tier gap: feature coverage difference across tiers.

Google
Google Summer of Code, Cloud Native Computing Foundation   06/2021 – 09/2021

- Joined and contributed to **KubeVela**, a modern application platform that makes it easier and faster to deliver and manage applications across hybrid, **multi-cloud** environments.

- Contribution mainly revolved around **Rollout Controller**, which was used in Rollout Plan and Rollout Traits (canary update).

PROJECTS

Load-Aware L4 Load Balancing for Microservices and Cloud Native Systems:
C, Golang, eBPF, Linux, Networking, Kubernetes, bash
Utilized Cilium CNI/eBPF and Kubernetes Service abstraction to develop a centralized L4 Load Balancer, preventing IO bottlenecks. Idea: https://marioskogias.github.io/docs/crab.pdf

Processing & Classification of Provenance On Spark(link):
Python, Data Processing, Machine Learning
Built a **PySpark** pipeline to generate provenance types from provenance graphs and aggregate them into **feature vectors**. Trained **ML** models to classify these graphs and identified the most influential provenance types for improved **model interpretation**.

Deep Learning Projects: Python, Pytorch, LLM, Text Generation, Computer Vision
Developed DCGAN, VAE generative models using Pytorch. (link)
Performed age regression from Brain MRI using segmentation, volume feature, and regression / CNN. (link)
Developed discriminative language classifiers using BoW, DeBERTa-Base, and BERT. Improved performance through keyword pre-processing and ensemble methods (Bagging, Voting). (link)
Implemented DP, MC, TD agents for a maze environment, and DQN/DDQN agents for pole balancing in reinforcement learning. (link)

A.I.D. Application for Intervening Depression: Mobile App, Flutter/Dart, MongoDB
Delivered a mobile app developed for patients, an admin webpage for data collection and training item set modification, and utilized MongoDB for data analysis and collection.

SKILLS

Languages: English (fluent), Chinese mandarin (native)
Programming Languages: Python, Java, C++, C, Go, C#, Ruby, Haskell, Scala, SQL, JavaScript, Lean4
Skills: Machine Learning, GenAI, Software Development, Distributed Systems, Database Systems

Books: Probabilistic Machine Learning (Intro) (100%), Principles of Mathematical Analysis (100%), Measure, Integration and Real Analysis (50%), Stochastic Calculus for Finance II(50%), Options, Futures and Other Derivatives(50%), CSAPP(100%), DDIA(100%), xv6(100%),

Courses (MSc):
- Computational Finance: Brownian motion, Option pricing, Black and Scholes model, Futures.
- Scalable System and Data: BigTable, Dynamo, Spanner, Spark, Memory indexing, Zookeeper.
- Reinforcement Learning: Markov Process, Bellman Optimality, TD, Q-Learning, DQN.
- Scheduling and Resource Alloc: Moore-Hodgson, Muntz-Coffman, Potential Games, Auctions.
- Deep Learning: GoogLeNet, ResNet, VAE, GAN, RNN, Attention, Diffusion
- Machine Learning for Imaging: Registration, YOLO, Atlas, Federated learning, Interpretability
- Natural Language Processing: Encoding, RNN, LSTM, AutoEncoder, Transformer, Bert, GPT
- Cryptographic Engineering: Perfect Secrecy, HASH, MAC, Commitment. 0 knowledge proof

Courses (BSc Selective):
- Optimization Methods: LP, Shortest Path, Convex Optimization, (Projected) Gradient Descent
- Programming practice and Application: Java, OOP, Design Patterns
- Database Systems: Relational Algebra, Buffer Pool, Query Execution, Lock/Latches.
- Internet Systems: HTTP, HTML, TCP/IP, TLS, JavaScript, REST.
- Operating Systems and Concurrency: Locks, Semaphore, Virtual Memory, Paging, Concurrency.
- Machine Learning: Decision Tree, K-means, OLS / ridge / lasso, SVM, Evo Algos, Neural Nets..