ZHONGYAN (HUGO) LUO

📞 858-241-4996 ✉ zhl203@ucsd.edu in LinkedIn Profile

Education

University of California, San Diego (UCSD)   **Sep. 2024 – Jun. 2026**
Bachelor of Science, Data Science   *San Diego, CA*

University of California, Santa Barbara (UCSB)   **Sep. 2022 – Jun. 2024**
Bachelor of Science, Data Science   *Santa Barbara, CA*

- **GPA: 3.96/4.00**
- **Coursework:** Machine Learning | Deep Learning | NLP | Statistical Models | Data Structure | Database | Big Data | Distributed system | Cloud Computing | Stochastic Process | Algorithm Design | Computer Graphics

Experience

Transwarp   **Jun. 2024 – Sep. 2024**
AI Research Intern

- Implemented a LoRA-based fine-tuning pipeline in **PyTorch**; fine-tuned multiple LLMs on **A100 GPU cluster** using **FSDP**, improving the accuracy of downstream tasks e.g. financial consultation by **20%**.
- Conducted an comparative evaluation of leading LLMs (GPT-4, LLaMA-2, BERT etc.) on 20+ **NL2SQL** benchmark datasets using **SQLite3** & **Pandas**, measuring execution accuracy and robustness across complex SQL queries.
- Built a **Retrieval-Augmented Generation (RAG)** data augmentation pipeline by leveraging **OpenAI text embedding** models and **FAISS** vector indexing to retrieve domain knowledge, achieving a **25%** increase in SQL generation accuracy by reducing hallucinations and improving entity linking.
- Authored patent "*A method and system for SQL generation based on background knowledge augment*".

ZTE Corporation   **Jun. 2023 – Sep. 2023**
AI Research Intern

- Developed a scalable **synthetic data generation pipeline** based on **Evol-Instruct**, expanding a 20K dataset to **over 100K** while achieving a **15% improvement** in general task performance.
- Designed a multi-layered Explainable AI **(XAI)** architecture, integrating solutions including **SHAP**, **LIME**, **counterfactual explanation** etc. for deep learning models; enhanced interpretability and model transparency in AI-driven decision making, which boosts the user retention rate by **25%**.
- Contributed to the design, development, and verification of a sensitive word monitoring feature in AI applications, delivering **5+ key features** and deploying via **CI/CD** pipelines using **GitHub Actions**.
- Authored paper "*Exploration and practice of XAI human-machine trust mechanism*", published on *Big Data Research*.

Zhongfu Information Inc.   **Sep. 2022 – Mar. 2023**
Data Engineer

- Migrated legacy ETL processes to **S3, AWS Glue** and **Redshift**, leveraging serverless architecture to reduce operational costs by **20%** while scaling for larger datasets.
- Developed **ETL pipelines** using **AWS, PySpark** and **SQL** to identify top-performing regional real estate agents from transaction records, facilitating customer-agent connections while improving transaction rate.

Projects

N-gram Statistical Language Model | *Python, NLP, Web Scrapping*   **Feb. 2025**

- Fetched and tokenized large text corpora using custom Python scripts with **requests & regex**.
- Built multiple statistical language models (Uniform, Unigram, N-gram) to handle large text (e.g. Homer's epic) in under 20 seconds, significantly improving context awareness and coherence of generated text compared to baseline models and demonstrating higher sequence prediction accuracy.

Heart Disease Prediction Model | *Python, Sklearn, XGBoost*   **May 2024**

- Engineered multiple heart disease classification models based on health metrics, including **Logistic Regression**, **SVM** and **XGBoost**, achieving a prediction accuracy of 92%.

C++ Path Tracing Render | *C++, Computer Graphics, Offline Rendering*   **Apr. 2024**

- Engineered a path tracing renderer in C++; Implemented sophisticated rendering techniques including Multiple Importance Sampling (MIS), Monte Carlo Algorithm, and microfacet BRDF to simulate realistic light interactions.

Technical Skills

Languages: Python (PyTorch, Scikit-learn, request, PySpark, Pandas, NumPy), C++, R, Java, Swift, Matlab
Technologies/Frameworks: SQL, Transformers, RAG, LoRA, GROP, FSDP, RNN, CNN, LSTM, AWS, Hadoop, Spark