

# ZHONGYAN (HUGO) LUO

☎ 858-241-4996 ✉ [zh1203@ucsd.edu](mailto:zh1203@ucsd.edu) [in LinkedIn Profile](#) [Github Profile](#)

## Education

### University of California, San Diego (UCSD)

Bachelor of Science, Data Science

Sep 2024 - Jun 2026

San Diego, CA

### University of California, Santa Barbara (UCSB)

Bachelor of Science, Data Science

Sep 2022 - Jun 2024

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 **15%**. ([www.wuya-ai.com](http://www.wuya-ai.com))
- Conducted a comparative evaluation of leading LLMs (GPT-4, LLaMA-2, BERT etc.) on 20+ **Text2SQL** 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 **18%** increase in accuracy.
- Established a data pipeline based on **schema-linking** to create 10k+ new training and testing examples to corpus.
- 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 **10% 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.

Mar 2023 – Jun 2023

Data Engineer

- Developed **end-to-end ETL pipelines** with **Spark**, **AWS Glue**, **Redshift** and third-party APIs, processing 100GB+ of data daily to empower the BI team with real-time analytics for data-driven decision-making.
- Migrated legacy ETL processes to AWS, leveraging serverless architecture to reduce operational costs by **30%**.
- Deployed data pipelines to production via **AWS CDK**, implementing Infrastructure as Code (IaC) for scalability.

## 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.

### 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:** Transformers, RAG, PEFT, LoRA, p-tuning, DDP, FSDP, RNN, CNN, LSTM, Hadoop, Spark, HTML/CSS, JavaScript, AWS S3, AWS Glue, AWS Redshift, Google Cloud Platform (GCP), MySQL