

Education

Case Western Reserve University , Cleveland, OH	Aug. 2024 – Present
Master of Science in Computer Science. Current GPA: 4.0/4.0; Thesis direction: Embodied AI, VLA, VLN. Coursework including: Machine Learning, Data Mining, Computer Vision, Machine Learning on Graph, Deep Generative Model.	
University of Washington , Seattle, WA	Sep. 2015 – Mar. 2020
Bachelor of Science in Applied Physics	

Working Experience

VULab CWRU, Cleveland	May 2025 – Present
Research Assistant	
<ul style="list-style-type: none">As the most junior level member on the team, presented research findings on Vision-Language-Action (VLA) models at 10+ internal research meetings, leading to new insights and identified 4+ potential areas for future research.	
ZHIPU.AI , Beijing	May 2021 – Aug. 2024
Algorithm engineer, Part-time	
<ul style="list-style-type: none">Developed models that enable metric-based document discovery and semantics-enhanced retrieval methods. These models were validated on millions of data entries and developed using Python, NumPy, the Neo4j Graph database, and MongoDB.Implemented plugins and Retrieval-Augmented Generation (RAG) pipelines with the company's Large Language Model (GLM) for 3+ B2B solutions, including LLM-enhanced retrieval, text-to-video, and image-to-video generation.Built large-scale knowledge graphs from journal articles and patents, comprising over 90 million nodes with an average degree of 12, to support metric computation and analysis.Nominated for 5+ patents focused on big data classification and identification, NLP, and GNN-related methods.Built and refactored a web application backend with optimized methods for executing asynchronous calculation tasks.Managed the terabyte-scale databases, including MongoDB to Elasticsearch migration and database migration to Alibaba Cloud. Deployed and maintained local cluster CPU/GPU/containers monitoring using Grafana.Implemented in-memory relational storage using Redis bitmaps and data stream processing using RabbitMQ. Designed and proposed a microservices architecture, boosting online system performance by 7x.Led and mentored a team of 4 interns, driving consistent performance and fostering strong cross-functional collaboration with the data department.	
Founder Securities Co., Ltd. , Beijing	Sep. 2020 – Dec. 2020
Quantitative analyst intern	
<ul style="list-style-type: none">Developed quantitative trading strategies in Python, including a multi-factor model based on research reports and a statistical model targeting on northbound Hong Kong capital flows affecting the mainland A-share market.	

Personal Project Experience

An Agentic Navigation Framework Utilizing Vision-Language Models , Cleveland	Jan. 2025 – May 2025
<ul style="list-style-type: none">Designed and implemented a Vision-Language-based navigation agent leveraging Qwen 2.5-VL (7B) as a cognitive “main brain” with structured memory and reflective reasoning. Engineered custom system + user prompting strategies to enable contextual planning and semantic understanding for embodied tasks. Integrated and evaluated the framework within Habitat-Lab and Room-to-Room (R2R) environments, demonstrating improved instruction following and environment grounding. https://arxiv.org/abs/2506.10172	
Enhancing Video Retravel Using VLM , Cleveland	Oct. 2024 – Dec. 2024
<ul style="list-style-type: none">Designed and developed a scalable backend and database layer for an application focused on retrieving relevant videos based on video or image input. Responsibilities include engineering on VLM model Qwen 2-VL (7b), retrieval method, and Vector DB integration. Used Transformer, Neo4j, and Pinecone, etc. for implementation. https://arxiv.org/abs/2503.17415	
Low-Rank Adaptation Defense with Robustness , Cleveland	Oct. 2024 – Dec. 2024
<ul style="list-style-type: none">Developed and evaluated a LoRA-based defense pipeline for a ResNet-18 model to counter Feature Importance Attacks (FIA). Achieved a best validation adversarial accuracy of 98.60% within only 2 epochs, demonstrating rapid model robustness improvement. Used Pytorch for implementation.	

Skills

Programming Languages: Python, Linux Bash. **ML Frameworks:** Transformer, Pytorch, Numpy, Scikit-learn. **Database Management Systems:** MongoDB, Elasticsearch, Kibana, Neo4j, Pinecone, PostgreSQL, Redis. **Message Queues:** RabbitMQ, Kafka. **Container:** Docker, Grafana.

Accomplishments

Patents	May 2021 – Present
Document fining methods in large corpus: CN 114510584 B CN 114969251 A CN 115471483 A ; Topic crusting: CN 116644338 B CN 116561605 B ; Information Retrieval: CN 117216417 B ; Generating training patterns: CN 118277794 A ;	