Zihan Wang

Tel: (484) 597-9438 | Email: zihan_wang3@brown.edu | GitHub: https://github.com/ziw224

EDUCATION

BROWN UNIVERSITY   **Providence, RI**
Master of Computer Science (Sc.M.), AI/ML Pathway   **Expected May 2026**
- Selected Courses: Software Engineering, Deep Learning, Computational Linguistic, Computer Vision

LEHIGH UNIVERSITY   **Bethlehem, PA**
Bachelor of Science (BS), Computer Science and Business Honors Program, Minor in Data Science (GPA: 3.79)   **2024**
- Selected Courses: Machine Learning, Big Data Analytics, Statistical Computing, Linear Methods, Statistical Methods, Design & Analysis of Algorithm, Data Structure, Computer Organization Architecture

WORK EXPERIENCES

EcoForge   **Boston, MA**
Founding | AI & Full Stack Engineer   **Dec 2024 - Present**
- **Architected full-stack AI-powered** cement production platform serving enterprise manufacturing clients, evolving from ML optimization API to **intelligent agent system** with natural language interface, securing **¥2M seed funding** and trials with **China's largest cement manufacturer**
- Built **production-grade AI agent** (LangChain, GPT-4o) orchestrating multi-tool workflows across two core capabilities: (1) RAG-based Q&A system querying 50+ plant datasets for historical production insights and (2) intelligent parameter recommendations, achieving **<2s query latency** using vector embeddings and hybrid search
- **Engineered a high-performance backend** using FastAPI, Redis, and WebSockets to handle real-time data streaming; developed a responsive Next.js/TypeScript frontend to visualize **100+ daily production KPIs.**
- **Established robust CI/CD pipelines** with automated testing for stochastic agentic tool-calling, ensuring 24/7 system availability and reliable model versioning/rollbacks.

LUNAR Lab, Brown University   **Boston, MA**
Research Assistant, *advised by Ellie Pavlick*   **Dec 2024 - Present**
- **Researched emergent properties** of Video Large Language Models (Video LLMs) to improve ML pipeline explainability and performance optimization
- **Developed and executed control experiments on EgoSchema** (e.g., shuffling frames and shuffling text infills) to benchmark and compare Image LLMs and Video LLMs performance on long-form video question-answering tasks.
- Implemented **AnyRes** technique in multi-image scenarios, enabling LLaVA-NeXT-Image to process videos with fewer tokens, and utilized **visual Chain-of-Thought** to generate automated text infills between frame pairs for enhanced data processing efficiency
- Co-authored research paper **"Video Finetuning Improves Reasoning Between Frames"** accepted to **CogInterp workshop at NeurIPS 2025** [paper]

PROJECT EXPERIENCES

Oracle Central Bank Digital Currency (Oracle Sponsored Open-Source Project, Fall '23):
- Forked an open-source CBDC platform from MIT and Boston's Federal Reserve's Project Hamilton known as "OpenCBDC" and extensively researched their 2-Phase Commit architecture, especially their use of localized data storage and transaction flow.
- Developed and executed a technical plan in coordination with Oracle's Vice President of Blockchain Technologies to port OpenCBDC's local storage of transaction information to a manually sharded **Oracle Autonomous Database**
- Engineered a **React-based frontend interface** to visualize UTXO transaction flows and developed an administrative dashboard for benchmarking and evaluating system throughput performance.
- **Published an Oracle Blog Post** titled "Extending CBDC Scalability and Resilience with Oracle Database Blockchain Tables." [GitHub, Blog Post]

SKILLS AND TECHNOLOGIES

- **AI/ML & LLMs:** LangChain, LlamaIndex, RAG Pipelines, Pinecone/Chroma (Vector DBs), PyTorch, TensorFlow, HuggingFace Transformers, OpenAI API
- **Languages:** Python (Expert), TypeScript, Javascript, SQL, Rust, C++, Java
- **Backend & Cloud:** FastAPI, Flask, Spring Boot, PostgreSQL, Redis, Docker, AWS (Lambda, EC2, S3), CI/CD (GitHub Actions)
- **Frontend:** React, Next.js, Playwright, Jest, Tailwind CSS