

Zhiheng Wang

Houston, Texas | zhihengw02@outlook.com | +1 608-707-7364 | [linkedin.com/in/zhihengwang02](https://www.linkedin.com/in/zhihengwang02)

Education

Rice University <i>Master of Computer Science</i>	Houston, TX, US Aug 2025 - Dec 2026 (Expected)
University of Wisconsin–Madison <i>Bachelor of Science in Computer Science</i>	Madison, WI, US Jan 2023 - Dec 2024
Beijing University of Posts and Telecommunications <i>Undergraduate in Telecommunication Engineering</i> (transferred)	Beijing, China Sep 2020 - Dec 2022

Experience

Research Intern (Data & Agent Engineering), Beijing Academy of Artificial Intelligence (BAAI)
Feb 2025 – Aug 2025

- **Chain-of-Thought (CoT) Data Pipeline Engineering**
 - Engineered a data generation pipeline for long-text reasoning by iteratively refining prompt templates for text chunking, CoT extraction, and logic summarization.
 - Generated a core batch of high-fidelity reasoning datasets that directly contributed to a 15% performance improvement in the team’s SFT base model on Math-500 and GSM8K benchmarks.
- **Agentic Workflow Proof-of-Concept (PoC) Validation**
 - Integrated 6 core ML operational scripts (data preprocessing, labeling, loss calculation, evaluation, training, inference) into standard tool schemas for Agentic automated workflows.
 - Validated the end-to-end toolchain execution via the Model Context Protocol (MCP) utilizing local Qwen-2.5 models and Cherry Studio, successfully executing automated data processing loops on sample datasets.

Projects

Empath-AI: Local-First Stateful Persona Agent Dec 2025 – Present

- Architected a dual-model conversational agent utilizing a lightweight Extractor for real-time state tracking and a high-capacity Generator for empathetic synthesis.
- Integrated **Mem0** and **ChromaDB** for localized NER and fact embedding, engineering a dynamic Python logic layer to retrieve and inject context, drastically reducing multi-turn hallucinations.

Multimodal Knowledge Graph (MMKG) Construction & Evaluation Jan 2026 – Present

- Engineered a data extraction pipeline leveraging **CLIP** to align multimodal features, constructing a domain-specific graph with explicit relational and implicit k-NN visual edges.
- Implemented **GCN/GAT** models via **PyTorch Geometric** for downstream link prediction, quantitatively proving the representational superiority of multi-sensory fusion over text-only baselines.

Interactive Developer Portfolio & Blog Feb 2026 – Present

- Developed a responsive full-stack web hub utilizing **Next.js**, **React**, and **Tailwind CSS** to host technical publications and interactive API-driven engineering demos.
- Established automated CI/CD deployment pipelines via **Vercel** to ensure seamless iteration and production-grade system availability.

Honors

- **5th Place** (of 128), ICPC North Central NA Regional Contest 2023
- **Silver Medal** (49th of 329), CCPC (China Collegiate Programming Contest) Weihai 2021
- **Top 1.5%** (of 7,000+), CCF-CSP Certification 2021
- **1st Prize** (Top 100 of 20,000+), NOIP (National Olympiad in Informatics in Provinces) 2019

Skills

- **Languages:** Python, C/C++, Java, JavaScript/TypeScript, SQL, Bash/Shell, R, Matlab
- **AI & ML:** PyTorch, HuggingFace, vLLM, CLIP, FlagScale, Megatron, Ollama
- **Agent & Data:** MCP, Mem0, ChromaDB, Pandas/NumPy, Neo4j, Bigtable
- **Web & Backend :** Next.js, React, Tailwind CSS, SpringBoot, Vue.js, Flask, FastAPI, Redis
- **DevOps & Deployment:** Linux/WSL 2, Docker, Git, Vercel, Kubernetes, GCP