Yuxi Zheng

Odysseusailoon@gmail.com | 193-8331-1208

Introduction

At 25, I'm a senior software engineer with a physics background, fueled by curiosity and a passion for AI. With four years of experience in distributed systems and database kernel development, I specialize in designing large-scale, high-availability infrastructure. My blend of analytical rigor and technical expertise drives innovation in AI and embodied agents, where I focus on creating robust safety frameworks and scalable pipelines to enhance their spatial-temporal understanding, planning, and decision-making in the physical world.

Education

Nanjing University 2016/08 — 2020/06

Bachelor of Science in Physics China

- Courses: PDE, Statistical Physics, Operating Systems, Optics, Quantum Mechanics, Database Systems, Computer Networks, Computational Physics

Work Experience

USC Viterbi School & World Engine Startup 2025/01 — Now

Research Engineer & Funding Engineer Los Angeles & Santa Clara, CA

- **Robot Data Agent System (0-1 Development)**: Built comprehensive system using S3, message queues, multi-agent architecture, and GPU framework (Ray) enabling users to create and view various simulated robotics data through natural language prompts
- **Large-Scale Dataset Platform (0-1 Development)**: Developed low-latency platform allowing users to select data ranges and types through histogram interface, edit metadata, and download from GCS with robust authentication and seamless latency
- **Research Contribution**: Co-authored and submitted paper to CoRL (top robotics learning conference) using state-of-the-art ML models to convert images into robotics data with motion and video
- Working with Prof. Yue Wang's research group while simultaneously serving as funding engineer for his robotics data collection company

Grab 2022/06 — 2025/01

Senior Software Engineer Singapore

- **AI Intelligent Search Platform Assistant**: Led a team of 5 engineers to build an AI-powered assistant for infrastructure platform
- Tech Stack: Python, LlamaIndex, ETL tools, Graph DB, OpenAI API, VectorDB, S3, Redis
- Designed and implemented multi-agent system supporting semantic caching, context pooling, adaptive RAG, GraphRAG and function calling
- Built real-time ETL pipeline for generating and updating Knowledge Graph from structured data
- Integrated with Slack API for seamless user experience and enhanced company efficiency
- **Redis Expertise**: Recognized as the go-to Redis expert within Grab, specializing in source code level understanding
- Developed and maintained widely utilized Redis common library and Hystrix lib, serving all Grab customers and driver partners
- Created resilient Redis chaos testing framework using Chaos Mesh, enhancing service reliability
- **Automation Platform**: Contributed to microservices system automating infrastructure requests
- Built metadata storage for all services and resources, connection management API, and infrastructure resources management system
- Used Python, Steampipe, and Event Bridge to build HA system as single source of truth for Grab infrastructure-related metadata

... continues on the next page ...

- **Leadership**: Founded and chaired Grab's Mental Health Community, organizing wellness initiatives and support programs

Shopee 2022/03 — 2022/05
Software Engineer Singapore
- Joined data management system development team
- Gained experience with Java tech stacks and demonstrated rapid learning abilities

Kingsoft Cloud 2020/06 — 2022/01
Redis Kernel Developer Beijing, China
- Investigated Redis offerings across cloud providers (Redis, AWS, Alibaba) and drafted implementation plans for new features
- Provided technical support for customers including top-tier companies: Xiaomi, Bilibili, and Bytedance
- Led development as core Redis developer, implementing key features: Hybrid Storage, HA Control Thread, and Redis Proxy
- Developed deep understanding of Redis source code for effective troubleshooting of production issues

Research Experience

Robot Learning from Any Images (RoLA) 2025/04 — 2025/05
Data Pipeline Lead - Submitted to CoRL (under review)
- Led development of scalable data pipeline transforming single images into interactive, physics-enabled robotic environments
- Built framework enabling massive visuomotor robotic demonstrations generation from camera captures, robotic datasets, and Internet images
- Designed novel visual blending strategy for photorealistic data collection combining real images with virtual assets

Vision-Language Planning Benchmark 2025/04 — Now
Co-First Author - Collaboration with UPenn, submitting to AAAI
- Co-leading comprehensive benchmark evaluation of VLM zero-shot planning capabilities across multiple models (GPT-4.1, InternVL3, QwenVL2.5, Gemma3)
- Developed evaluation pipeline for PDDL generation from visual inputs using scene graph and captioning approaches
- Analyzed solver pass rates, plan success rates, and action success rates across Blocksworld and other robotic domains

Adaptive scalable multi-agent system design 2024/08 — 2024/10
Researcher & Engineer, Camel AI Community
- Joined Camel AI community as consulting collaborator focusing on LLM and AI infrastructure
- Applied expertise in large-scale distributed systems to research and engineering of adaptive multi-agent systems
- Contributed to ongoing research projects and previous paper-related implementations

Task Scheduling for Edge Computing: Delay-Sensitive Application 2020/01 — 2020/06
Student, mentored by Prof. Zhuzhong Qian Nanjing University
- Designed and implemented algorithm using Lyapunov optimization for job dispatching and data redistribution on edge cloud
- Used Markov approximation to solve optimization problem for delay-sensitive applications

Skills

Programming:
- PyTorch, CUDA, Golang, C, C++

Software Engineering Tools & Technologies:
- Redis (expert level), Linux kernel, Network, RDS, TiDB, VectorDB, S3, Git, Docker, Kubernetes, ETL, CI/CD pipeline, Datadog, Chaos Mesh

... continues on the next page ...

- Distributed Systems Architecture, Database Kernel Development, High-Availability Infrastructure Design, Scalable System Optimization, Data Processing Pipeline Development, Cloud Computing Platforms (AWS)
- AI-Powered Development: Proficient use of Cursor, Claude Code, Supabase and Vercel for fast prototyping and deployment

AI & Robotics Domain Knowledge:
- Model and build training pipelines for LLM/VLM/Diffusion models
- Robotics policies and models: Diffusion policy, ACT, Octo, OpenVLA, PiO
- Simulation: IsaacSim, IsaacGym
- PEFT methods (LoRA), MoE, Distributed training, Distributed inference

Other Skills:
- Problem-Solving, Fast Learning and Adaptability, Marketing

Languages:
- English (proficient), Chinese (native)

Last Updated on July 20, 2025