

Yanzhi Yao

Burnaby, BC, Canada | Renoyao9@gmail.com | [LinkedIn](#) | [Personal Website](#)

SUMMARY

AI engineer and technical founder with hands-on experience building production-grade LLM, RAG, and real-time voice AI systems. Skilled in Python, Flask, AWS, Azure, and enterprise AI deployment, with a record of architecting scalable AI products that improve efficiency, reliability, and response quality.

EXPERIENCE

We Link Technology LLC — San Jose, California, USA

Sep 2023 – Jan 2025

Co-Founder | AI Engineer

Built and deployed production-grade generative AI systems for a Fortune 500 elevator company, spanning LLM assistants, real-time voice AI, and domain-grounded enterprise workflows.

AI Sales & Service Assistant (JuTong)

- Architected and deployed a GPT-4o-powered sales and service assistant using Python/Flask, Docker, and AWS EC2, enabling domain-specific recommendations, automated diagnostics, and customer support workflows.
- Designed and implemented a RAG pipeline over 1Gb of internal sales documentation, improving response grounding, reducing hallucinations, and increasing query accuracy for enterprise use cases.
- Improved answer quality through prompt refinement, feedback-driven optimization, and production iteration, increasing sales team efficiency by ~30% and contributing to \$1M in cost savings.

AI Elevator Emergency Call Center Platform

- Built a real-time voice AI emergency response platform integrating Twilio, STT/TTS pipelines, and LLM-based conversation handling for elevator entrapment scenarios.
- Engineered low-latency, fault-tolerant voice workflows using Azure AI Services and GPT-4o fallback mechanisms, improving response reliability in live emergency interactions.
- Reduced emergency intervention time by 30%, improved grounded response accuracy from 80% to 90%, and cut support escalations by 50% through SOP-grounded optimization and post-incident feedback loops.

JTANG Consulting Inc — Vancouver, BC, Canada

Jan 2025 – Present

ERP System Software Engineer & Functional Consultant

- Delivered Microsoft Dynamics 365 Business Central customizations for 10+ suppliers, streamlining warehouse and retail workflows, improving operational efficiency by 30%, and reducing inventory errors by 50%.
- Led end-to-end migration of a medical supplier's legacy ERP system to an Azure cloud-based platform, including MySQL data migration and system upgrade; ensured 99.9% uptime post-launch.

UC San Diego, SOC Lab — San Diego, California, USA

Jun 2022 – Sep 2022

Research Assistant

- Developed Python-based traffic control models and simulation workflows for autonomous vehicle research, improving simulation efficiency by 20% and validating a 15% improvement in traffic flow stabilization.
- Designed and deployed a Flask web application to demonstrate mixed-traffic highway control strategies.

EDUCATION

Simon Fraser University — Burnaby, BC, Canada

Sep 2023 – Dec 2024

M.Sc. in Computer Science, GPA: 3.7/4.0

University of California, San Diego — San Diego, California, USA

Sep 2021 – Mar 2023

M.S. in Machine Learning & Data Science, GPA: 3.6/4.0

Cardiff University — Cardiff, UK

Sep 2019 – Jun 2021

B.S. in Computer Science (First Class Honors), GPA: 4.0/4.0

SKILLS

Languages: Python, Java, C++, JavaScript, SQL, HTML/CSS, MATLAB, AL

AI / GenAI: LLM Applications, RAG, Prompt Engineering, Model Fine-Tuning, STT/TTS, Real-Time Voice AI

Frameworks / Cloud / Data: Flask, Django, React, AWS, Azure, Docker, Redis, MySQL, MongoDB, Spark, REST APIs, Postman, JMeter

AWARDS & HONORS

Silver Award for 6th China College Students' Internet Plus Innovation Competition.

2020

National Project Approval for Development and Promotion of Lizi Visitor Platform Based on SaaS Mode.

2019

Excellent Award and Most Popular Award for 2019 Microsoft Imagine Cup Competition.

2019