

Xiangyu Zheng

Graduate AI Engineer | Agentic AI Systems | Applied ML

Sydney, Australia · xiangyuzheng0802@gmail.com · [LinkedIn](#) · [Portfolio](#)

Professional Summary

Early-career AI Engineer (Master of AI & Data Science) who builds and evaluates AI agents and machine learning workflows over messy, real-world data, then ships clean, maintainable code. Experienced designing agentic LLM systems with schema-constrained generation, tool allowlists, and confidence gating; engineering pipelines from structured and unstructured inputs; and running experiments, validating outputs, and iterating. Grounded in machine learning fundamentals, statistics, and data analysis, with a Python data stack (NumPy, Pandas, scikit-learn, PyTorch) and hands-on delivery across Qwen, OpenAI, and full-stack systems.

Skills

- **Agentic AI & LLMs:** AI agents · agentic AI workflows · multi-step LLM orchestration · schema-constrained (structured) generation · tool allowlists · human-in-the-loop approval · confidence gating · Qwen Cloud (qwen-plus) · OpenAI API · Google Generative AI · prompt engineering
- **Machine Learning & Data:** machine learning fundamentals · statistics and data analysis · classical ML models · structured and unstructured datasets · engineering features from incomplete and noisy real-world data · run experiments, validate outputs, and iterate
- **Python Stack:** Python · NumPy · Pandas · scikit-learn · PyTorch (ML foundations, coursework & bootcamp)
- **AI Concepts (exposure):** Retrieval-Augmented Generation (RAG) · multi-agent collaboration (MAC) · domain-specific fine-tuning · vector/pgvector-compatible schemas · embeddings
- **Data & Delivery:** data visualisation and analysis pipelines · PostgreSQL · PolarDB · MongoDB · SHA-256 deduplication · clean, maintainable code · shared tooling and codebases · Vitest · contract/domain/end-to-end testing · GitHub Actions · CI/CD · Docker · Agile/Scrum · Claude Code · Codex
- **Cloud & Full Stack:** Alibaba Cloud Function Compute · Cloudflare Workers/R2 · AWS S3 · Vercel · Next.js · React · TypeScript · Node.js · Express · REST APIs

Selected AI Projects

[EmailAgent — Job Opportunity Autopilot](#)

Two-Person Hackathon Team | Jun 2026 – Jul 2026

- Designed a Qwen-powered **AI agent workflow** that engineers **unstructured, noisy real-world** recruitment emails into an event-sourced opportunity board, using structured generation, deterministic matching, an **eight-tool Zod allowlist**, and confidence gates of **≥0.85 automatic**, **0.60–0.84 confirmation**, and **<0.60 review** — a validate-and-iterate loop over model outputs.
- Built event-driven ingestion with Cloudflare Email Routing, raw MIME parsing, PostgreSQL persistence, `FOR UPDATE SKIP LOCKED`, five-minute leases, exponential retry, and SHA-256 deduplication; verified the application with **60 passing tests across 12 test files**.
- Designed the opportunity schema, event normalization, deterministic stage projection, preference matching, immutable event timeline, human approval state machine, and valuable-deal filtering across **11 API routes** — modular AI components integrated into an automation workflow.
- Automated container delivery through GitHub Actions, running tests and production builds, building an **AMD64 Docker image**, pushing it to Alibaba Cloud ACR, and deploying to Function Compute; used PolarDB for persistence. Authored **42 of 77 commits** in the two-person repository snapshot.

UNSW AI Academic Advisor

Full-Stack AI Project | Feb 2025 – Jun 2025

- Designed a multi-step advising **agent workflow** using OpenAI API, Google Generative AI, contextual memory, and **rule-grounded constraints** across **four guidance tasks**: course planning, prerequisite validation, degree requirements, and policy interpretation.

- Developed prompt workflows combining LLM reasoning with user context and deterministic academic constraints, running experiments and iterating on prerequisite validation, degree-progress checks, course-path planning, and policy questions.
- Built the React, Node.js, Express, and MongoDB application with JWT authentication and dedicated dashboards for **three user roles**: students, administrators, and program managers.
- Optimized the Vite-powered responsive interface for cross-device academic planning workflows.

Professional Experience

[VVS Education Pty Ltd](#) — Junior Full-Stack Developer

Hybrid, Australia | Jun 2026 – Present

- Architected an **AI-powered mathematics question bank**, converting analysis of **210 pages across 8 units and 55 exercise sets** into approximately **95 reusable problem archetypes** and a **300–500-question MVP pipeline** with programmatic grading, prerequisite-aware diagnostics, and LLM-generated student feedback.
- Gathered requirements, developed domain understanding, and shipped four ChatGPA planning tools with Next.js, TypeScript, and MongoDB for GPA prediction, target-grade planning, admission matching, and application guidance; protected datasets behind rate-limited server APIs.
- Engineered admission-rule snapshots, matcher contracts, server-side **data projections**, and domain tests to keep published requirements aligned with output, maintaining **5,813 passing tests across 161 test files**.
- Established a PR- and push-triggered **CI/CD pipeline** that installed dependencies, enforced ESLint, ran **5,813 Vitest tests**, and validated production builds before Vercel deployments.
- Built the [VVS Academy / Pluto Academy](#) production platform across **55 page routes and 43 API routes** with transactional bookings, auditable points-ledger workflows, direct-to-R2 uploads, and role-specific dashboards; owned the repository end to end, authoring **693 of 693 commits** across **912 tracked files**.

Earlier Software Engineering Experience

Jobpin AI — Software Engineering Intern

Sydney, Australia | Jul 2025 – Sep 2025

- Designed an OpenAI prompt-chaining and contextual-reasoning workflow for resume analysis, candidate profiling, and personalized job matching during a **three-month internship**.
- Built responsive React/Next.js experiences and Node.js/Express REST APIs; integrated AWS S3 for secure resume upload and retrieval into the analysis workflow, collaborating via GitHub, Jira, and Agile iteration.

IFA Translator — Software Engineering Intern

Seven-Person Team | Jan 2025 – May 2025

- Built a React, Next.js, Node.js, and MongoDB translation platform integrating Azure AI Translator across **four input modes**—speech, text, image, and documents—in a **seven-person team**; designed backend services for sessions, translation history, and persistent data with end-to-end testing in an Agile environment.

Education

University of New South Wales (UNSW)

Master of Artificial Intelligence and Data Science | 2023 – 2025 Sydney, Australia Relevant Coursework: Machine Learning · Natural Language Processing · Deep Learning · Big Data Analytics

Southwest Jiaotong University (SWJTU)

Bachelor of Materials Science and Engineering | 2019 – 2023 — science/engineering discipline with applied AI focus Chengdu, China

JR Academy

AI Engineer Full-Stack Project Bootcamp | 2024 – 2025 — Australia

- Project-based training across OpenAI, RAG, React.js, Node.js, Spring Boot, and AWS; full SDLC across **two individual and one commercial project**.