

YAO XIANG

✉ xiangyao888@gmail.com | 🌐 yxiang-828.github.io | 🔗 linkedin.com/in/yao-xiang-733b06329 | 📄 github.com/Yxiang-828

SUMMARY

Computer Engineering undergraduate at NUS, specializing in Advanced Electronics & IoT with a Minor in AI, building across **AI infrastructure, model inference, agent runtimes, and hardware/software systems**. Experience spans TensorRT GPU inference optimization, autonomous agent infrastructure, CI/CD, robotics, and independent research into persistent autonomous agents.

EDUCATION

National University of Singapore – B.Eng. Computer Engineering, specializing in Advanced Electronics & IoT · Minor in AI Aug 2024 – Aug 2028 (Projected)

- Coursework: Data Structures & Algorithms, Software Engineering, Machine Learning, Digital Design, ARM Assembly, Embedded Systems, and Information Security.

EXPERIENCE

Robotics Software Intern – KABAM Robotics | Python, C++, ROS 2, Docker, Jetson Orin, TensorRT, SAM3, May 2026 – Dec 2026 YOLOv8, FFmpeg, CAN bus, LiDAR, SSH, Apps Script

Matrix-4 autonomous delivery robot

- Optimized real-time SAM3 vision-language inference on Jetson Orin AGX using TensorRT, integrating FFmpeg pipelines and YOLOv8 overlays for low-latency deployment.
- Built an agentic debugging toolchain automating SSH execution, ROS 2 / CAN telemetry extraction, log parsing, and test workflows across deployed robots.
- Led outdoor field testing and calibration of Matrix 4 robot, optimizing LiDAR SLAM, Nav2 MPPI controller parameters (`vx\_max`), docking, and MWAN3 routing.

Teaching Assistant, CS2113 (Software Engineering & OOP) – NUS | Java, GitHub, CI/CD Jan 2026 – May 2026

- Teach software engineering practices including OOP design, JUnit testing, Git workflows, code review, and CI/CD to undergraduate development teams.

HACKATHONS & RECOGNITION

DSTA BrainHack CODE_EXP – Open Finalist (2026) [certificate] · Tencent "Age Well" Social Good Challenge Singapore (2026) · NUS × Synapse × IMDA AI Innovation Challenge (2026) · VibeWithSG (2026) · AI Engineer Hackathon (2026) · TinyFish SG Hackathon (2026) · Hack4Good (2026) · HacX – CS7 (2025)

SELECTED PROJECTS

Aiko – Persistent OS Agent & Multi-Model Runtime | TypeScript, Node.js, SQLite, React Native, Tailscale, Apr 2026 – Present Ollama / OpenRouter / Kimi, Qwen3 TTS

Persistent autonomous-agent runtime across Linux, mobile, Telegram, and VRM interfaces · [details](#) · [aiko-core](#) · [aiko-avatar](#) · [aiko-app](#)

- Designed a 5-tier memory architecture using Hindsight BM25 + cosine RRF over SQLite vectors, episodic memory, and a self-map; redesigned dynamic context assembly to cut token payload from 109KB to 34KB (-69%).
- Built a multi-model runtime with automatic failover across local Ollama, OpenRouter, and Kimi, stall watchdogs, and local Qwen3-TTS GPU voice inference (RTF ~0.19).
- Engineered self-backup, self-diagnosis, multi-surface IPC, and calendar-driven autonomous scheduling across the open-source aiko-core, aiko-avatar, and aiko-app stack.

Replay – Singapore Family Safety & Asymmetric AI Engine | React, Vite, Node.js, Vercel, Render, OpenAI, Aug 2026 – Present SQLite

Asymmetric multiplayer scenario game for Singapore family safety against online harms · [details](#) [Live Demo](#)

- Engineered asymmetric role-locked scenario engine resolving multiplayer evidence views via a deterministic authoritative room state engine with zero cross-role data leaks.
- Implemented source-locked AI narrative generator for dynamic scam/safety simulation and structured real-time AI post-game debriefing.

AnAn – Autonomous Senior Care AI Companion & Care Kernel | Python, FastAPI, React, WASM, MediaPipe, SEA- Aug 2026 – Present LION, Qwen3-TTS, MapLibre GL, OneMap SG, SQLite

Autonomous senior care companion engine treating time, silence, and state as core inputs ({time, silence, state}) · [details](#)

[Live Demo](#) [Video Demo](#) [Tencent Age Well Challenge](#)

- Architected 7-phase decision kernel (WAKE → THINK → REVALIDATE → GATE → ACT → RECEIPT → COMMIT) with 3-tier permission gating and stale-decision revalidation before action dispatch.
- Built silence FSM escalation with walking family contact tree, on-device browser CV health screening (rPPG heart rate, pose mobility, FAST stroke facial symmetry), local Qwen3-TTS voice cloning, and deterministic code safety floors.

Agora – Real-Time Agent-Native Development Environment | TypeScript, VS Code Fork, Mattermost, Node.js, Jun 2026 – Present WebSockets, Docker, Bash

Agent-native development environment integrating Mattermost and VS Code forks · [details](#) [Video Demo](#)

- Built custom orchestrator service resolving parallel AI agent git merge conflicts in real time, featuring 3D spatial presence rooms over WebSockets and single-script container deployment (`up.sh`).

Kampung Kaki – Civic Emergency Coordination & Role-Aware Map | Python, FastAPI, MQTT 5, Redis Streams, May 2026 – Jun 2026 MapLibre GL, OneMap SG, React, Docker

Distributed emergency-coordination system · [details](#) [Video Demo](#) [DSTA Brainhack Open Finalist](#)

- Delivered unified operational picture for citizens, responders, and ops teams via single canonical state over retained MQTT 5 QoS 1 messaging, Redis streams, FastAPI, and OneMap SG APIs.

TECHNICAL SKILLS

Languages:	Python, JavaScript/TypeScript, Java, C/C++, SQL, Verilog, ARM Assembly
AI / ML:	PyTorch, DINOv2, SAM3 VLM, OpenAI / OpenRouter, Gemini CLI, Ollama, Whisper, MediaPipe WASM, SEA-LION, MERaLION, Qwen3-TTS
Robotics:	ROS 2 Jazzy, Nav2, MQTT, Jetson Orin AGX, CAN bus, SLAM, Frigate
Frameworks:	React, React Native (Expo), Vite, Electron, FastAPI, Next.js, Mermaid C4, Tailwind, OpenCV
Infrastructure:	Docker, Vercel Edge, Render, Tailscale (WireGuard), Git/GitHub Actions, SQLite (WAL + FTS5), Linux/Unix
Hardware:	Arduino, Verilog / Vivado (Basys 3 FPGA), Raspberry Pi