

Tan Jing Kai

Singapore · jingkai_tan@mymail.sutd.edu.sg · +65 9726 4570 · linkedin.com/in/jingkait27 · tanjingkaivercel.app · github.com/jingkai27

PROFESSIONAL SUMMARY

Computer Science and Technology Entrepreneurship student at SUTD who builds AI-powered applications, not just explores them. At Helios (New York), designed and shipped production LLM workflows -- AI document generation and explain-with-audio features -- then refactored the prompt-handling architecture to cut AI generation errors by over 60%. At Panasonic R&D, deployed computer vision AI (YOLOv8) to live operational environments for fall detection and fire safety surveillance, end-to-end from enterprise use case to working system. Strong in Python, LLM API integration, prompt engineering, and translating customer requirements into deployed AI solutions. Portfolio and project walkthroughs: tanjingkaivercel.app

CORE COMPETENCIES

LLM API Integration Prompt Engineering & Iteration AI-Powered Workflow Design Python & Data Structures
Computer Vision (YOLOv8 / Edge AI) Technical Prototyping & Demos API & Backend Integration Conversational AI Development

WORK EXPERIENCE

Helios May 2025 – Jul 2025

Full Stack + LLM Infra Engineer · New York, USA

- Built and shipped production LLM-powered workflows for Proxi -- AI document generation and explain-with-audio features -- integrating LLM APIs directly into customer-facing product features used by live users.
- Refactored prompt-handling architecture to separate formatting responsibilities, **cutting AI generation errors by over 60%** and significantly improving long-term system reliability.
- Architected prompt-based document workflows using Liveblocks-powered real-time collaboration, reducing content setup time by 30-40% across multi-user editing environments.
- Unified template architecture across Scribe and Library views into a single standardized entry point, increasing developer velocity and ensuring system-wide consistency.

Panasonic Research and Development Aug 2024 – Dec 2024

AI / Computer Vision Engineer · Singapore

- Deployed computer vision AI (YOLOv8) to live operational environments for fall detection and fire safety surveillance -- translated enterprise use cases into working AI systems, end-to-end from model selection to production deployment.
- Built end-to-end ML pipeline scripts fully in Python covering data processing, model inference, and deployment, working directly with real enterprise datasets and safety-critical use cases.

PROJECTS

KubeVision – Kubernetes Observability Stack Mar 2026 – Present

ML Infrastructure · Self-Initiated

- Engineered a production Kubernetes observability stack for a chaos-injected FastAPI ML inference service -- full three-pillar coverage (metrics, logs, traces) via Prometheus, Grafana, Loki, and Jaeger via Helm.
- Instrumented p95 latency, error rate, and distributed traces via OpenTelemetry SDK, wiring the full pipeline from pod to dashboard for end-to-end AI system accuracy and latency monitoring.
- Diagnosed two silent production failures through systematic layer-by-layer debugging: a ServiceMonitor label mismatch and cross-namespace DNS failures severing the tracing pipeline.

Stack: Kubernetes, Prometheus, Grafana, Loki, Jaeger, OpenTelemetry, FastAPI, Python, Helm

MCPOS – Multi-Currency Payment Orchestration Jan 2026 – Feb 2026

Backend + API Integration · Self-Initiated

- Engineered a high-concurrency B2B payment engine using PostgreSQL pessimistic locking -- 100% data integrity under concurrent load with serializable transaction isolation.
- Architected multi-provider failover with Adapter Pattern and Circuit Breakers across Stripe and PayPal APIs; integrated HMAC-verified async webhooks and a real-time React dashboard with optimistic updates.

Stack: PostgreSQL, Node.js, TypeScript, React, React Query, Stripe API, PayPal API, HMAC

WhoCodes Better – Real-Time Competitive Coding Platform Nov 2025 – Jan 2026

Real-Time Systems · Self-Initiated

- Engineered a bidirectional event-driven architecture using Socket.IO and Redis with high-speed matchmaking queues and live state management for real-time head-to-head competitive interactions.
- Architected a scalable TurboRepo monorepo with a unified TypeScript type system; orchestrated Docker Compose deployment integrating Node.js, PostgreSQL, Redis, and a sandboxed code execution engine.

Stack: Node.js, TypeScript, Socket.IO, Redis, PostgreSQL, Docker Compose, TurboRepo, React

EDUCATION

Singapore University of Technology and Design (SUTD) Expected 2027

B.Sc. Computer Science and Design | MSc Technology Entrepreneurship (concurrent)

GPA: 4.58 / 5.0 · First Class Honours

SKILLS

Languages: Python, JavaScript / TypeScript, SQL

AI / ML: LLM APIs (OpenAI, Anthropic), Prompt Engineering, YOLOv8 / Computer Vision, OpenTelemetry, LangChain (in progress), RAG pipelines (in progress)

Backend: Node.js, FastAPI, PostgreSQL, Redis, REST APIs, HMAC / webhook security

Infrastructure: Kubernetes, Docker, Prometheus, Grafana, Loki, Jaeger, Helm, OrbStack

Frontend / Tools: React, Socket.IO, TurboRepo, Streamlit, Plotly, Liveblocks