

Tien Cheng Oh

Singapore | tiencheng.oh@gmail.com | [linkedin.com/in/ohtiencheng](https://www.linkedin.com/in/ohtiencheng) | github.com/Tien-Cheng

Education

National University of Singapore, B.Comp. in Computer Science – Singapore Expected 2029

- NUS Global Merit Scholarship.
- GPA 4.79/5.0.

Singapore Polytechnic, Diploma in Applied AI and Analytics – Singapore Mar 2023

- GPA 4.0/4.0; Singapore Computer Society Silver Medal.

Experience

Backend Software Engineer Intern, TikTok – Singapore July 2026 – present

- Fixed a production video moderation agent that often flagged entire videos instead of the violating segment, raising precise timestamp output from 72% to 91% by enforcing structured JSON output and fixing a validation bug.
- Built a 3,900-case benchmark to test whether retrieving past moderation decisions (RAG) improves an LLM moderation agent, and prototyped an autoresearch agent that lets ops teams run their own experiments.
- Added token, latency, and cache metrics and a Grafana dashboard to a Go captioning service to track LLM cost.

Co-founder, Conform Labs – Singapore Aug 2025 – Dec 2025

- Built a voice-agent evaluation framework to compare agent quality across simulated conversations using automated scoring and Langfuse trace analysis.
- Developed form analytics, document-to-form generation, and real-time voice-agent pipelines with TypeScript, FastAPI, and Pipecat.

Instructor and Software Developer, National Service, Digital and Intelligence Service – Singapore May 2024 – July 2025

- Built an LLM-powered exam-marking platform with custom RAG over exam materials that reduced instructor grading time by over 50% in an air-gapped environment.
- Improved grader accuracy and robustness by benchmarking prompts and automating prompt optimization with DSPy and open-source data.
- Integrated Langfuse telemetry to monitor traces, latency, and scoring quality, reducing time spent debugging the pipeline.

Co-founder, Flora – Singapore Feb 2023 – Oct 2023

- Built a neural search API for unstructured enterprise documents with Elasticsearch, ONNX Runtime, Jina, FastAPI, and MongoDB, supporting public-sector and enterprise pilots.
- Improved document-layout detection throughput by 400% by distilling and quantizing an RTMDet model for production inference.
- Deployed pilot systems on GCP and refactored services for self-hosted environments to enable on-premise adoption with SkillsFuture SG, NSCC, and Coherent, Inc.

Software Engineer Intern, Defence Science and Technology Agency – Singapore Sept 2022 – Feb 2023

- Built an internal MLOps model hub with Vue.js, FastAPI, and MongoDB that let teams deploy ML models as Kubernetes services.
- Automated deployment to air-gapped Kubernetes environments with Skaffold, Helm, and service-scaling APIs for 2 departments.
- Established a GitHub Actions CI pipeline for unit testing and linting to improve code quality and release confidence.

Awards & Open Source

- 1st Place (University Category), HPC Innovation Competition 2023, for OwlShield, an LLM safety firewall.
- Contributed a [LiteLLM integration](#) to NVIDIA's Garak, an open-source LLM security scanner (8k+ stars).

Skills

Languages: Python, Go, Java, TypeScript, SQL

ML/LLM: PyTorch, Hugging Face Transformers, DSPy, RAG, ONNX Runtime, Scikit-learn

Backend/Serving: FastAPI, Streamlit, Celery

Data/Vector/Search: PostgreSQL, PGVector, MongoDB, Elasticsearch

Infrastructure/Observability: Docker, Kubernetes, GCP, Langfuse, OpenTelemetry, Grafana