YU CHUAN OH

AI Engineer / ML Engineer | LLM Agents | Applied ML

Berlin, Germany | +49 176 77628402 | ohyuchuan@gmail.com
linkedin.com/in/yu-chuan-oh-169a3419b | github.com/yc82628

SUMMARY

Builds AI that knows when it's wrong: agents with guardrails and fallbacks that never fail silently, models that defer uncertain cases to humans, and evaluation that catches data leakage. Shipped 4 AI systems in 2026, from a live LLM shopping agent to medical imaging, including a Top 10 hackathon finish.

SKILLS

ML Reliability: model calibration, leakage prevention, uncertainty-based abstention, evaluation design
LLM & Agent Systems: tool-calling agents, multi-provider fallback, prompt-injection defence, LLM-as-judge
Debugging & Testing: root-cause analysis, regression rollback, audit logging, pytest
Core Stack: Python, TypeScript, PyTorch, FastAPI, Next.js, SQL
Spoken Languages: English (C1), Mandarin (B1), German (A2)

PROJECTS

SigPath: LLM Shopping Agent with Scam Checks | Solana x Superteam Germany Hackathon, 2026

TypeScript, Next.js, LLM tool calling, Groq, Gemini, Ollama, Solana | sigpath.vercel.app

- Built Ai-chan, a tool-calling LLM agent that turns plain-language requests into searches over 50 live eBay listings, with guardrails that add any warning the model omits and defences against prompt injection.
- Provider-agnostic (Groq, Gemini, Claude, local Ollama) with automatic fallback; measured Gemini at 19-60 s per answer vs 1-2 s on gpt-oss-120b and moved the live site to Groq. 629 automated tests.

Diabetic Retinopathy Screening System | Solo project, 2026

Python, PyTorch, OpenCV, FastAPI, SQLite, pytest | github.com/yc82628/Medical_Diagnosis_w_DL

- Cleaned messy public retinal datasets (broken files, duplicates, no patient IDs) and caught a split that leaked 142 of 400 patients (35%) into both train and test; fixed it with patient-grouped splits.
- Built for cases the model has never seen: blurred or borderline scans go to doctors, every verdict is logged, and model-doctor agreement is tracked to detect drift. Measured both calibration methods and kept isotonic, cutting calibration error by 70%.

DayTonator: Supply-Chain Malware Detector | Top 10, Daytona AI Builder Hackathon, Jun 2026

Python, FastAPI, Daytona, strace, Anthropic API | github.com/yc82628/Daytonator

- Built the sandbox-forking layer: forked one clean snapshot into a control plus 4 baited copies (SSH keys, AWS credentials, prod host, clock jump), exposing evasive malware a normal scan misses.
- Cross-fork strace evidence feeds an LLM judge (Qwen2.5-72B, Anthropic API) that returns a scored verdict, with a deterministic fallback so a scan never fails when a model is unavailable.

Midknight Watcher: Autonomous Coding Agent | 42 Berlin x Needle Hackathon, May 2026

Python, Cerebras, Groq, Ollama, Git | github.com/yc82628/Ai-Agent-

- Owned the unattended overnight run of a self-repairing coding agent (plan, code, test, fix) with 3-tier cross-family LLM fallback, in a team of two.
- Traced 0/150 overnight scores to a test-harness timestamp bug rather than agent code, then rolled back an infinite-loop regression to recover the 92/150 build (30/100 on hidden tests).

EDUCATION

BSc (Hons) Computer Science | Berlin School of Business and Innovation (BSBI) | 2023 - 2026