

CHARALAMBOS ROTSIDES

Forward Deployed AI Engineer

Production AI systems · AI agents & RAG · Document intelligence · Private LLM deployment · Full-stack

Limassol, Cyprus — open to remote | charalambosrotsides@gmail.com | LinkedIn | GitHub | charalambosrotsides.com

PROFILE

AI and software engineer with 3+ years of professional experience building, deploying and operating production software and AI systems. Assistant Scientist at the Cyprus Marine and Maritime Institute (CMMI), where I own systems end to end: architecture, backend, databases, frontend, deployment and infrastructure.

I build AI agents, voice agents, RAG and document-processing pipelines and self-hosted LLM infrastructure, and take AI prototypes to reliable production systems. Most useful on complex or loosely defined problems that need someone to understand the domain quickly, choose the architecture and build the full solution.

PROFESSIONAL EXPERIENCE

Cyprus Marine and Maritime Institute (CMMI)

2023 – Present · Cyprus

Assistant Scientist / Software & AI Engineer — Maritime Digitalisation Centre

- Design, build and deploy software and AI systems for maritime digitalisation, machine learning and EU-funded R&I projects. Own systems end to end, lead architecture decisions and supervise a junior developer.
- Real-time maritime platforms: AIS vessel data from multiple sources, PostgreSQL/PostGIS geospatial queries, Redis caching and WebSocket delivery to React frontends (vessel history, geofencing, areas of interest).
- OCR, information-extraction and LLM-assisted analysis pipelines for technical and research documents.
- CI/CD and production infrastructure on Linux, Docker, Nginx and Cloudflare; turning experimental research outputs into reliable software.

Self-Hosted LLM Infrastructure

CMMI · AI Infrastructure / Model Serving

Designed and optimised a self-hosted **~27B-parameter LLM inference platform** supporting multi-user workloads.

- Benchmarked quantisation and precision configurations, trading off model quality, GPU memory, throughput, concurrency and latency.
- Request queueing, APIs and monitoring of active and waiting users; validated with concurrent agent load tests.

Internal Application Platform

CMMI · Platform Engineering

Designed and deployed an internal containerised application platform supporting **multiple production research applications**.

- Isolated Docker workloads, private networks, persistent volumes and restricted secrets; an Nginx configuration watcher validates routing changes before applying them, so a bad deploy cannot break production.

INDEPENDENT PRODUCTS & AI PROJECTS

MealVerve

Founder / AI Product Engineer · mealverve.com

- Designed, built and operate a live AI meal-planning product: weekly plans, recipe import, automatic calorie and macro tracking, grocery lists.
- Own the full stack and operations: Next.js, Prisma, Auth0, Stripe, Sentry, installable PWA; features evolve with user feedback.

Kai — Bilingual AI Voice Agent

Voice AI / Tool Calling

- English/Greek voice agent for outbound and inbound calls that checks real calendar availability, books the meeting and texts a confirmation with an add-to-calendar link.
- Vapi pipeline (GPT-4o, ElevenLabs, Deepgram) over Twilio/Telnyx with a Flask tool backend; ran **5 concurrent outbound calls**; detects voicemail and IVR systems and ends the call.

JARVIS — Voice-Driven Coding-Agent Orchestrator

Multi-Agent / Realtime Voice

- A phone app streams speech to a realtime voice model; a Qwen router sends each task to Claude Code, Codex or a Qwen coding agent in an isolated git worktree, reports back by voice and merges on command.
- Resumable agent sessions per project; a rate-limited executor goes on cooldown and the router fails over to another.

agent-loop — Autonomous Plan / Code / Review Loop

Multi-Agent / MCP

- Planner and reviewer agents direct coding agents through a task checklist; file-disjoint subtasks run in parallel (up to 4) in separate git worktrees and merge back.
- The reviewer checks UI changes in the running app through Playwright MCP and database state through a read-only database MCP. Used it to build a full-stack compliance web app.

Historic Handwriting Recognition

Document AI / Zindi Competition

- 5-fold evaluation harness and data audit, then fine-tuned Kraken/McCATMuS on legacy Kepler GPUs: weighted **CER 35.4 → 9.4** and **WER 11.4 → 4.9** on a held-out fold of 816 lines.

TECHNICAL SKILLS

AI / LLM: OpenAI and Anthropic APIs, open-source models, vLLM, RAG, embeddings and vector search, tool calling, MCP, AI agents and multi-agent systems, structured outputs, model routing, LLM evaluation, OCR and document processing

Machine learning: PyTorch, TensorFlow/Keras, LSTMs, reinforcement learning, GNNs, computer vision, quantisation, model serving

Software & data: Python, TypeScript/JavaScript, Node.js, Express, FastAPI, React, REST APIs, WebSockets, queues, PostgreSQL/PostGIS, MongoDB, Redis

Infrastructure: Docker, Linux, Nginx, PM2, Cloudflare, CI/CD, GPU inference servers, monitoring, load testing

EDUCATION & ADDITIONAL

B.Eng. Computer Engineering — University of Cyprus

2019 – 2023

Thesis: artificial pancreas. Lead programmer in a five-person team; LSTM glucose forecasting (**RMSE ~10 mg/dL ±3**, 23 simulated patients) and RL insulin dosing (**>70%** time in range), deployed on Android with TensorFlow Lite.

Awards: 1st Place Team Award — University of Cyprus Student Innovators Competition, 2023 · European Innovation Academy, A grade, 2023

Certificates: Coursera — Improving Deep Neural Networks · Structuring ML Projects | **Languages:** Greek (native), English (fluent)