

Manos Adamopoulos

Software/AI Engineer | LLM Data & ML Systems

man.adamopoulos@gmail.com | github.com/manKraut | +30 694 752 6197

Summary

Senior Software Engineer with 3+ years designing, shipping, and operating Python backend services and ML/LLM systems on Google Cloud (Cloud Run, Kubernetes), owning the full lifecycle from requirements through CI/CD, deployment, and monitoring. Two MSc degrees in STEM, both with honors. 6+ years working directly with clients and startups, translating real business problems into high-impact technical solutions. Currently building agent skills and RAG systems at the core of GWI's AI transformation, and a member of the AI Enablement team defining agentic engineering standards across the organisation.

Professional Experience

AI Cloud Engineer | GWI | Oct 2024 – Present

Lynx - Synthetic Survey Generation Agent (RAG)

- Architected and shipped a **production AI orchestration backend** (Python/Django, DRF) for an LLM-driven synthetic survey platform: FSM-based job lifecycle with row-level concurrency control and idempotent retries; parallelized the respondent batch-generation pipeline with a Redis-backed async worker tier and graceful long-job shutdowns, cutting end-to-end runtime by ~94% (77+ PRs, ~400 commits)
- Deployed and operated **cloud-native infrastructure on GKE** via custom **Helm** charts and **Terraform/Terragrunt**: Workload Identity, KMS-encrypted secrets, Cloud SQL, multi-env (staging/prod) with HPA, PDBs and Prometheus alerting; automated object-storage → BigQuery ingestion feeding the company's data platform.
- Built an **LLM synthetic-respondent engine**: semantic few-shot retrieval over a 483-question, 1,536-dim embedding index, Big5-persona-conditioned generation, largest-remainder demographic quota sampling, and parallel batch execution; created an evaluation harness benchmarking synthetic vs. real respondents across 5 metric families.

Segmentation Integration Automation (SIA) - ML pipeline to predict custom audience segments on GWI's Core dataset

- Engineered and productionized an event-driven ML post-processing microservice on GCP (Cloud Run, Pub/Sub, Docker, Terraform) that rebalances weighted audience distributions for tens of thousands of respondents per run, guaranteeing convergence within 0.5% of market targets via a deterministic, parallelized iterative optimizer with early stopping.
- Designed a novel **LLM-agent optimizer** (**GPT-4o** function-calling loop, **3 custom domain tools**) that autonomously measures distribution divergence and reweights segment assignments; extended the engine to multilabel segmentation and achieved 100% ground-truth respondent retention, fixing a production-critical reassignment defect.

Platform and Company wide events

- Led adoption of engineering standards for the department as **DevOps Champion**, driving **AI Agentic engineering harness** adoption via the AI Enablement team and migrating department infrastructure into GWI's central **Terraform** repo with resource labelling that enabled per-product cost attribution
- **Drive AI literacy** across the organisation: speak at public discussions, run an **AI News Forum** that keeps teams informed on developments and encourages adoption

Data Engineer | GWI | Oct 2023 – Oct 2024

- Integrated the ML pipeline with the product team's Segmentation's API over authenticated REST calls, decoupling GCS file I/O from the workflow engine.
- Replaced legacy SPSS processing with a BigQuery ETL pipeline, consolidating 3 data sources into one version-controlled transformation layer. The data pipeline reduced its time from 12 hours to 1 hour (~90% reduction).

Solutions Engineer | Linked Business | Mar 2021 - Oct 2023

Market and credit intelligence (KYC) platform covering **800K+** Greek companies; clients include systemic banks and credit-servicing firms.

- Owned technical relationships with systemic banks and credit servicers, leading integration scoping, security reviews, and go-live of KYC onboarding and credit-check workflows; turned client pain points into prioritised roadmap fixes.
- Cut authentication tickets **40%** in one quarter via root-cause analysis of recurring login failures; shipped fix with Playwright regression tests in **CI/CD** so the failure class cannot silently return.
- Drove **50% API usage** growth in 6 months: turned client interviews into a prioritised roadmap item, then scoped and delivered a bulk Excel export endpoint.

Technical Project Manager | Sociality co-op | Dec 2019- Mar 2021

Led a team of 6 developers to deliver web platforms, including a unified CMS, member portal and ticketing system (TypeScript, React, Node.js, PostgreSQL), driving a **40%** increase in ticket sales within one quarter

Education

- **MSc** in Software Development & AI - University of Piraeus 9.8/10
- **MSc** in Civil Engineering (5-year degree) - Democritus University of Thrace 7.04/10

Projects

[Job Matcher](#) - full-stack AI platform for job hunters.

A platform that helps job seekers describe roles in natural language, search listings from extracted preferences, and get concise, preference-aware explanations of why results fit.

- Sequential **multi-agent backend** (preference extractions and match advisor) with evidence-backed prompts from vector retrieval over API docs (**text-embedding-3-small**, **ChromaDB**).
- **Failure-tolerant RAG integration** (lazy service init, continue without context on errors).
- **React/TypeScript UI**, **FastAPI** services, **Docker** images and **Kubernetes** deployment.
- **GitHub Actions** CI builds backend/frontend Docker images on every push/PR to main.

Stack: Python (FastAPI) + React/TypeScript (Vite, Tailwind) + OpenAI + ChromaDB RAG, containerized with Docker and deployable on Kubernetes, with GitHub Actions CI

[SkyScribe-CLI](#) - multi-agent weather app

Python CLI that turns live forecast data into short, actionable weather briefs (temperature ranges, rain windows, wind, practical tips) for travelers and daily planning, using Open-Meteo APIs.

- **Open model + PEFT:** Fine-tuned **TinyLlama with LoRA** adapters (**Transformers** + **PEFT**)—train and ship small adapters instead of full weights.
- **Training/inference stack:** Supervised pairs with prompt-token label masking; optional 4-bit QLoRA for GPU-efficient training; adapter hot-load at inference via env config.

Stack: Python, Typer, httpx, **Hugging Face Transformers**, PEFT.

Skills

Languages: Python, SQL, TypeScript, JavaScript, HTML/CSS

Backend & APIs: FastAPI, Django REST Framework, REST API design, event-driven microservices (Pub/Sub), PostgreSQL, Redis

Cloud & DevOps: GCP (Cloud Run, GKE, Cloud Build, Artifact Registry, Secret Manager, GCS, BigQuery), Docker, Kubernetes/Helm, Argo Workflows, Terraform, GitHub Actions

Reliability & security: Secret Manager/SOPS, signed URLs, Sentry, structured logging, pytest, Playwright

ML/LLM: LightGBM, scikit-learn, SHAP, RAG, embeddings, ChromaDB, Hugging Face/PEFT

Frontend: React, Tailwind