

Nikolas Bakalis

+1 646-250-0176 | bakalis.nikolas@gmail.com | Manhattan, NY | Dual Citizenship US/EU
linkedin.com/in/nikolas-bakalis | github.com/NikosBakalis

Senior Software Engineer | Production AI & Foundation Models | Applied AI

Senior engineer with 5+ years shipping production systems end-to-end — full-stack Python, foundation model integration (Claude MCP, OpenAI, Bedrock, Llama), retrieval-augmented generation, and ML pipelines with measurable business impact. Comfortable owning the full loop: problem framing, evals, deployment, and monitoring. Looking to build agentic AI systems at higher stakes.

Professional Experience

Channel Factory

New York City, NY

Senior Software Engineer, Applied AI & Backend

11/2024 – Present

- Named **Product/Tech Star of the Year 2025–2026** after designing and shipping real-time **AdTech MVPs** from scratch, ingesting live campaign data to surface insights for strategic accounts (Apple, Hyundai, Mazda); each engagement projected at **\$25K–\$50K** in monthly revenue per account.
- Delivered **\$0.46–\$0.51M** in annual EBITDA with a 12-month payback by owning end-to-end automation of CSM proof-of-execution across **YouTube, Meta, TikTok, and DV360**: scoped with sales/ops leadership, designed and shipped a Salesforce-integrated Playwright-based RPA system that replaced an 8-step manual workflow, automated **80–90%** of screenshot generation, cut cycle time from hours to minutes, and impacted **70+** global AM/CSM roles.
- Architected a **data lakehouse** on **Apache Iceberg** and **AWS Athena** using **Medallion Architecture** to unify ingestion and analytics across millions of advertising channels, defining service contracts for downstream microservices and replacing a fragmented SQLite-based pipeline.
- Built a **GenAI Q&A tool** over proprietary internal data using **RAG** with a custom n-gram retrieval engine and the **OpenAI** streaming API exposed via **MCP**; selected from a company-wide hackathon as the foundation for production rollout.
- Mentored junior engineers on integrating **MCP** servers into agentic workflows and raising code to production standards, covering testing, observability, error handling, and code review practices.

ACERTUS

New York City, NY

Software Engineer, ML Engineering & Full Stack

05/2022 – 11/2024

- Built and operated ML algorithms to **forecast ETAs, delays, and fraud**, improving route efficiency by **23%** and cutting fuel costs by **20%**; owned model performance in production including drift monitoring and retraining cadence.
- Automated large-scale deployments with **Python, AWS, and Docker**, reducing costs and processing time while ensuring cloud scalability.
- Led client-facing projects for Ford, Volkswagen, and Holman, turning product ideas into scalable cloud tools that improved customer experience.

Vivpro Corp.

New York City, NY

Full Stack Engineer, ML Engineering

06/2021 – 05/2022

- Architected the **pre-ChatGPT** conversational interface for Vivpro's flagship product RIA (R&D Intelligence Assistant) using **LSTM**-based sequence modeling, enabling natural language querying over regulatory and clinical trial data before LLMs became commercially available.
- Earned recognition from Pfizer, AstraZeneca, and Novo Nordisk for ML and full-stack solutions that improved data accuracy and research workflows.

Education

Anglia Ruskin University

Cambridge, UK

M.S. in Artificial Intelligence

09/2022 – 09/2023

Thesis: Adversarial Attacks on Intel and MobileODT Cervical Cancer Screening — 9% classification accuracy improvement post adversarial attack and retraining.

University of Patras

Patras, Greece

Integrated B.Eng. & M.Eng. in Computer Engineering

10/2016 – 06/2022

Thesis: Format Conversion & Data Transfer between Relational Databases.

Skills

Languages & Backend: Python (Django, DRF, Flask, FastAPI, SQLAlchemy, Alembic, Pytest), Elasticsearch, Polars, Pandas, NumPy.

AI / ML: Foundation Models & APIs (OpenAI GPT-4, AWS Bedrock, SageMaker, Llama, Claude); Production AI Patterns (RAG, prompt engineering, structured output, MCP, vector search); Classical ML (regression, time-series/ARIMA, LSTM/sequence models, statistical modeling).

Cloud & Infrastructure: AWS (EC2, S3, DynamoDB, Bedrock, SageMaker, OpenSearch), Apache Iceberg, AWS Athena, PostgreSQL, Docker.

Methodologies: Agile/Scrum, SOLID, DRY, YAGNI, MVC, end-to-end ownership, cross-functional collaboration with sales/ops.