

Konstantinos Chervatidis

Senior AI/Software Engineer | Node.js, Python, React, LLMs

✉ konstantinos.chervatidis@gmail.com

📍 Zurich, Switzerland

🌐 linkedin.com/in/konstantinos-chervatidis/

🐙 github.com/KostasCherv



My Contribution

Senior AI/Software Engineer with 8+ years of experience designing and shipping production systems across AI, backend infrastructure, and high-throughput distributed platforms. I specialize in building agentic AI and RAG systems (LangGraph, LangChain, Pinecone, FastAPI) with the same production discipline I bring to backend architecture: async processing, observability, failure-aware design, and clean API boundaries.

My strongest contribution is architectural ownership end-to-end, from system design and data modeling through deployment and monitoring. I have built AI orchestration platforms, real-time data pipelines, high-throughput transaction infrastructure, and API-driven products, consistently improving reliability, latency, and delivery speed. As a senior engineer, I combine strong hands-on coding with the judgment to choose the right architecture for a given constraint, whether technical or business-driven, and communicate that reasoning clearly to both engineering and business stakeholders.

Work Experience

Independent AI Engineer

Jan 2026 – Present | Switzerland

- Architected Cortex, a multi-workflow AI orchestration platform using LangGraph and LangChain, supporting structured report generation, itinerary planning, and a grounded conversational interface with citations to source documents and prior context
- Designed a stateful LangGraph architecture instead of a simpler sequential chain, enabling multi-turn follow-up conversations without repeating retrieval steps, with every response grounded in per-run source chunks
- Built a durable RAG ingestion pipeline using a transactional outbox pattern and Inngest async workers for idempotent, retry-safe document processing, validated with test documents up to 50 pages
- Used Pinecone as the retrieval layer for its ability to scale as a managed service in production RAG systems, rather than self-hosting a vector store
- Implemented auth-scoped data isolation for session-scoped follow-up chat and end-to-end LangSmith observability for tracing agent behavior in production

Software Engineer, Maple Finance

Sep 2025 – Nov 2025 | Remote

- Delivered backend services for financial analytics and data reconciliation in production environments
- Improved the primary historical data aggregation endpoint with Redis caching and Subgraph DataLoaders, reducing query latency by approximately 70%
- Built real-time indexing and wallet position tracking with accurate cross-source reconciliation

Senior Backend Engineer, Gelato Network

Oct 2024 – May 2025 | Switzerland

- Owned backend services for distributed transaction infrastructure across 10+ networks.
- Built APIs and middleware for high-throughput transaction flows, focusing on reliability, latency, and operational stability.
- Automated deployment workflows and reduced new network rollout time from days to minutes, and improved transaction infrastructure through logging, monitoring, alerting, and failure-aware backend design.
- Worked with TypeScript, Node.js, Redis, API services, distributed systems, and production infrastructure.

Clearstar Labs AG, Senior Software Engineer

Sep 2022 – Sep 2024 | Switzerland

- Led a team of 4 developers, translating requirements from analysts into structured tasks and technical guidelines for delivery
- Designed event-driven data pipelines aggregating data from multiple independent market sources with consistent cross-source accuracy

- Choose a serverless architecture (AWS Lambda and Step Functions) over dedicated machines to scale quickly and reuse functionality across backend components without managing dedicated infrastructure state
- Built and deployed CI/CD pipelines on AWS, reducing release time from production deploys from 4 hours to 15 minutes
- Collaborated closely with product and business stakeholders to translate financial data requirements into reliable, maintainable software

Backend Developer, *Quintessential SFT* [🔗](#)

Apr 2021 – Aug 2022 | Greece

- Led backend development for a banking transactions mobile app, designing scalable APIs and handling high-volume financial data flows

Fullstack Developer, *Atypon* [🔗](#)

Jan 2020 – Mar 2021 | Greece

- Built a full-stack analytics platform for real-time data visualization, contributing across backend services and React frontend

JavaScript Developer, *NMN*

Jun 2018 – Dec 2019 | Greece

- Developed a browser-based rendering engine powering graphics for 20+ games, focusing on performance and real-time execution

Backend Developer, *Mita Travel*

Sep 2017 – May 2018 | Greece

- Implemented high-throughput backend APIs for a booking platform using microservices architecture and distributed systems technologies

Education

MSc in Artificial Intelligence, *Aristotle University of Thessaloniki*

2020 – Sep 2022 | Greece

Thesis: Reinforcement Learning for Forex Trading using Actor-Critic and PPO algorithms

BSc in Computer Science, *Hellenic Open University*

2016 – 2020 | Greece

Certificates

- AI Model Development & Deployment — Coursera — June 2026 [🔗](#)
- IBM RAG and Agentic AI — Coursera — May 2026 [🔗](#)
- Machine Learning in Production — DeepLearning.AI — May 2026 [🔗](#)
- Model Context Protocol: Advanced Topics — Anthropic Academy — March 2026 [🔗](#)
- Building with the Claude API — Anthropic Academy — March 2026 [🔗](#)

Skills

AI and LLM Engineering: LLM agents, RAG systems, LangGraph, LangChain, LangSmith, vector search (Pinecone), MCP

Languages and Frameworks: Python, TypeScript, FastAPI, Node.js, NestJS, React, Next.js

APIs and Data: REST APIs, GraphQL, PostgreSQL, Redis, Supabase, async processing, MongoDB

Cloud and Infrastructure: AWS (Lambda, Step Functions, ECS Fargate, CDK), Docker, Kubernetes, CI/CD, deployment automation, Azure

Systems and Reliability: Event-driven workflows, transactional outbox pattern, idempotent job processing, Inngest, monitoring and alerting (Datadog, PagerDuty), observability, Grafana, Prometheus

Languages

English — Fluent, **German** — Intermediate, **Greek** — Native