Viachaslau Spetsyian

GitHub: github.com/saequus spetsyian.com
LinkedIn: www.linkedin.com/in/spetsyian/
Warsaw, Poland

ABOUT

Senior Software & AI Engineer with 9+ years of experience building production systems in Python, with a track record of deploying ML models at scale and a focus on LLM applications: agentic workflows, retrieval-augmented generation (RAG), evaluation-driven development and LLMOps. Combines strong backend and cloud architecture with a business and finance background. Fluent in Polish, English, Russian.

KEY ACHIEVEMENTS

• Led and mentored a team of 5 middle and junior engineers, improving code quality by 30% and cutting onboarding time by 80% • Led backend development of key features that reduced average response time from 4.2s to 0.9s • Architected microservices on Google Cloud, improving scalability and reducing infrastructure costs by 25% • Deployed production ML models that cut latency by 40%, lifted accuracy by 25% and boosted adoption by 35%

EXPERIENCE

Razor Labs – SENIOR SOFTWARE ENGINEER Jan 2024 - Mar 2026
MLflow, TorchServe, Grafana, Terraform, GCP, FastAPI, PostgreSQL, MongoDB: Led a team of 5 engineers delivering ML-powered product features. Owned model serving, monitoring and deployment pipelines for production-grade ML models. Bridged research algorithms and real-world products, reducing processing time by 40%, improving prediction accuracy by 25% and increasing user adoption by 35%.

Zageno Inc. – SENIOR SOFTWARE ENGINEER Nov 2021 - Jan 2024
Python, GCP, AWS, Kubernetes, FastAPI, PostgreSQL, Django: Designed integration solutions consolidating multiple services into a unified platform. Built high-impact product features and improved pricing algorithms.

Emerging Travel Group – BACKEND DEVELOPER Oct 2019 - Oct 2021
Python, Go, Django, TypeScript, Express.js, K8s, PostgreSQL, Redis: Built REST API integrations in Go and Python, developed and operated microservices, improved CI/CD flows and built a backend-for-frontend layer in TypeScript.

AI PROJECTS

Serpentaria Capital – AI ENGINEER Mar 2026 - Present
LangGraph, MCP, pgvector, Langfuse, FastAPI: Building LLM agents and RAG pipelines for finance and research use cases, with tool calling, structured outputs, automated evals and tracing.

TECHNICAL SKILLS

LLM & Agents: OpenAI / Anthropic / Gemini APIs, LangGraph, LangChain, LlamaIndex, Pydantic AI, Model Context Protocol (MCP), tool calling, structured outputs
RAG & Vector Search: pgvector, Qdrant, embeddings, hybrid search, reranking
LLMOps & Evals: Langfuse, LangSmith, Ragas, DeepEval, prompt versioning, guardrails
ML & Serving: MLflow, TorchServe, PyTorch, Hugging Face, vLLM, Ollama
AI Dev Tools: Cursor, Claude Code, GitHub Copilot
Cloud Platforms: AWS, Google Cloud Platform (Vertex AI), Microsoft Azure
Languages: Python, Go, Rust, TypeScript / JavaScript
Databases: PostgreSQL, MongoDB, Redis
Frameworks: FastAPI, Django, React, Vue, Node, Express, pandas, numpy, Matplotlib
DevOps: Kubernetes, Docker, Terraform, Grafana, Nginx, CI/CD

EDUCATION

University of Finance and Management in Warsaw, CGPA: 4.39/5.00
BSc in Business Administration and Service Management,
Management Department. Warsaw, Poland. September, 2016

RELEVANT COURSES

• Strategic Leadership: Impact, Change, and Decision-Making Specialization • Team Leadership • Statistics & Calculus

PROFESSIONAL INTERESTS

• Agentic AI Systems • Distributed Systems • System Architecture Design • Life optimization using AI