

Vitaliy Kuzmin

Lead AI Architect & Senior Engineer | Remote / B2B Contract

hello@hi-ai.app | linkedin.com/in/vitaly-k-ai-engineer | vitalyk.xyz

Career Profile

Lead AI Architect and Senior Engineer with 9+ years of complex software system development and a deep specialization in multi-agent orchestration, local LLM infrastructure, and enterprise AI platforms. Proven track record of architecting secure, on-premise AI ecosystems, transitioning legacy enterprise workflows to deterministic LLM pipelines, and leading cross-functional engineering teams. Focuses on building context-driven AI solutions that bridge the gap between isolated models and real business impact.

Career Highlights

- **Enterprise AI Integration:** Scaled AI-assisted analytics across internal data silos, decreasing manual enterprise documentation navigation by up to 50%.
- **Workflow Automation:** Automated 60% of routine formatting and editorial preparation tasks from single-source drafts, significantly enhancing cross-platform delivery speed.
- **Technical Leadership:** Steered a cross-functional 10+ person engineering division (backend, frontend, QA, and analytics), driving the technical roadmap and CI/CD release cycles.

Skills

- **AI & Orchestration:** Agentic AI Development, LangGraph, Retrieval-Augmented Generation (RAG), SQL Agents, Model Context Protocol (MCP), Context-Driven Development, Prompt Engineering.
- **Infrastructure & Evaluation:** Local LLMs, Ollama, vLLM, LLM Evaluation, LLM-as-a-Judge, Langfuse, Vector Databases (LanceDB), Docker, GCP, CI/CD, DevOps.
- **Backend Engineering:** Software Solution Architecture, Python, FastAPI, C++, PostgreSQL, REST APIs.
- **Leadership:** Technical Leadership, Team Leadership, Product Development, Architecture Documentation.

Professional Experience

InterProCom | *Lead AI Architect* | Mar 2025 - Present

InterProCom specializes in Enterprise Asset Management (EAM) platforms for complex industrial sectors.

- Engineered the core architecture for the Axioma enterprise AI assistant, tailoring it specifically for complex asset management in the energy domain.
- Developed advanced multi-agent workflows utilizing LangGraph to handle SQL querying, norm-control verification, and sophisticated RAG over proprietary corporate documentation.
- Implemented rigorous LLM-as-a-judge evaluation pipelines using Langfuse to ensure high reliability, quality control, and observability of agent outputs.
- Orchestrated the seamless integration of private AI infrastructure with existing internal APIs, structured databases, and deployment pipelines via Docker.
- Established context-driven development practices, managing global prompts, reusable agent skills, and architecture documentation alongside the codebase.

BL Group R&D Center | *Lead AI Engineer* | Jul 2022 - Mar 2025

Spearheaded technical execution within an internal R&D division dedicated to building production-ready AI automation workflows.

- Architected "Smart Publish", a centralized AI-driven platform designed to automate content adaptation, cross-channel scheduling, and editorial review processes.
- Engineered deterministic LLM pipelines utilizing OpenAI API for automated text drafting, structural formatting, and channel-specific metadata extraction prior to human-in-the-loop validation.
- Developed robust backend microservices with FastAPI to orchestrate data flow between Telegram infrastructure, proprietary CMS platforms, and n8n automation layers.
- Guided early-stage R&D initiatives, leveraging Google Cloud Platform (GCP) to prototype automated cloud workflows prior to the flagship product launch.

BL Group | *Team Lead / Lead Software Engineer* | Apr 2017 - Jun 2022

Directed the engineering lifecycle for industrial software and geospatial machine learning platforms.

- Steered a cross-functional 10+ person engineering division, driving the technical roadmap and release cycles for "Light in Night", an enterprise-grade lighting calculation suite.
- Architected "Urban Lighting", a comprehensive ML geospatial platform utilizing LightGBM for illuminance forecasting, integrating mobile laboratory metrics with field inventory data.
- Streamlined the entire product delivery pipeline, translating complex business requirements into technical specifications, managing backlog prioritization, and enforcing CI/CD best practices.

Selected Projects

- **Axioma AI (Enterprise AI):** An enterprise-grade AI assistant platform engineered securely for the energy sector. Core Architecture includes LangGraph multi-agent workflows, domain-specific RAG, SQL Agents, Langfuse observability, and on-premise local LLM infrastructure.
- **hAi (Life-Context AI Assistant):** An independent R&D project developing a "Life-Context AI"—a personal layer integrated directly into the OS. Utilizes Python/FastAPI, React/Electron, MCP, and local LLMs to provide highly personalized, zero-prompt assistance while keeping user data strictly private.
- **Smart Publish (Workflow Automation):** A centralized AI platform built to streamline the editorial lifecycle. Leveraged FastAPI, OpenAI API, and custom CMS integrations to automate 60% of routine formatting tasks.
- **Urban Lighting (ML Platform):** A geospatial machine learning platform developed for city infrastructure planning. Utilized LightGBM, Python, C++, and GIS datasets to accelerate field data processing speed by 60%.

Education

Moscow Power Engineering Institute (MPEI) Master's Degree, Electronics and Nanoelectronics | 2018