

Alexander Fal

AI Agent Engineer · Tech Lead

Moscow · Telegram @faleksander · me@afal.dev · linkedin.com/in/alex-fal · afal.dev/cv · Updated: August 2026

SUMMARY

I build agentic systems hands-on and lead their development — in any proportion, from player-coach to 100% coding. 9+ years of engineering: 3 years of commercial Python (my entire current stack), 6 years of Java before that; ~5 years of management — a backend team plus coordination of a 12-person product team (–80% release defects, time-to-market from a week to a day). Since December 2025 I lead the GenAI track at Kamaz Digital: a logistics AI assistant — pilot in operation since June 2026 (tool calling, agentic RAG, a custom MCP server, evals, human escalation); agent-driven analytics — 9 → 65 metrics within a $\leq 3\%$ tolerance vs the reference report; an AI development process for teams. Coding agents have been my primary way of writing code since August 2025 (Claude Code, Codex): my own infrastructure of skills, MCP servers, subagents and a deterministic guard for destructive actions.

KEY RESULTS

- **pilot · June 2026** — logistics AI assistant in operation — architecture, tool calling, agentic RAG, MCP server, evals
- **9 → 65** — agent-driven analytics metrics in one month; verified against the reference report, $\leq 3\%$ tolerance
- **1 sprint** — trading module for logistics marketplaces — idea to production (Python, backend + frontend)
- **–80%** — release defects after changes to planning and code review
- **week → 1 day** — time-to-market after moving to a release-free delivery model
- **3 days → 30 min** — analytics turnaround on ClickHouse / DataLens

EXPERIENCE

Kamaz Digital

July 2023 — present · Moscow · GenAI Track Lead · System Architect

Career path: Data Platform Lead → System Architect (position) → GenAI track lead (role, since December 2025)

- Leading the GenAI track since December 2025: a logistics AI assistant from business requirements to operation — scenarios, tool calling, agentic RAG, knowledge layer, evals, human escalation; pilot in operation since June 2026 with live users; team of two.
- Built agent-driven analytics over the corporate DWH: a semantic layer of 9 → 65 verified metrics in one month; reconciled with the reference management report within a $\leq 3\%$ tolerance — a

closed improvement loop driven by a coding agent.

- Trading module for logistics marketplaces — idea to production in one sprint: requirements, architecture, backend (Python) and frontend; ~5 marketplace integrations, partly hands-on, partly supervising a developer and an automation QA.
- A custom MCP server for the assistant (coding-agent access to the product); an A2A PoC of four cooperating agents; browser automation for marketplaces without APIs — by agreement with the platforms.
- Established an AI development process for teams: specs, role skills (analyst → system analyst → developer), an AI-ready repository checklist; piloted on three projects; a series of internal AI meetups.
- Crisis management: when the neighbouring team's backend developers left at once, took over their module — fixed production, interviewed and onboarded new engineers; newcomers on new features, legacy triage on me.
- Cut analytics turnaround from 3 days to 30 minutes: a data platform on ClickHouse, 10+ sources, DataLens reporting.

Globaltruck IT

July 2020 — July 2023 · Moscow · Principal Developer / Backend Lead

- Led a product team of 12 (developers, analysts, QA) and a backend team of 3 developers.
- Reduced release defects by 80% through changes to planning, code review and retrospectives.
- Cut time-to-market from a week to 1 day by initiating the move to a release-free delivery model.
- Brought average resolution time for critical incidents under an hour: with DevOps, introduced CI/CD, monitoring, tracing (Prometheus, Grafana, Tempo) and structured logging (Logstash, Loki).
- Designed and built backend services for the logistics platform and integrations with external systems.

Center of Financial Technologies (CFT) / Diasoft

October 2017 — July 2020 · Novosibirsk · Software Engineer

- Built backend modules for an online banking system and integrations with UFEBS, SWIFT, Oracle and PostgreSQL.
- Set up CI/CD in Jenkins and aligned integration requirements with adjacent teams.

Parallels–NSU Laboratory

September 2014 — August 2015 · Novosibirsk · Research Software Assistant

- Built a Linux kernel tool for analyzing web server parameters (graduation project).

STACK

LLM systems and agent architectures

LLM orchestration: LangChain / LangGraph, production LLM model selection and operation, MCP and tool calling, agentic RAG: query expansion, agentic retrieval, evals: datasets and quality metrics, run-cost control, Claude Code / Codex / Cursor

Backend

Python, FastAPI, Java, Spring Boot, PostgreSQL, Apache Kafka, ClickHouse, event-driven integrations

Leadership & process

teams up to 12, hiring & onboarding, SDD-style AI development process (specs, role skills, AI-ready checklist), ADR / RFC, internal AI meetups

Architecture

DDD, Clean Architecture, system design, API contracts, ADR / RFC

Operations

Docker, Kubernetes, CI/CD, Prometheus / Grafana, Loki / Tempo

EDUCATION

Novosibirsk State University (NSU), Faculty of Information Technologies — B.Sc. in Computer Science and Engineering, 2018

Novosibirsk State University (NSU), Faculty of Mechanics and Mathematics — B.Sc. in Mathematics and Computer Science, 2015

Courses: IT Solution Design Workshop · TOGAF 10 & ArchiMate 3.2 · DDD & Clean Architecture · Business Oriented System Design · Microservice Architecture · Team Management Toolkit

LANGUAGES

Russian — native · English — C1