

Fedor Kabachenko

AI Engineer

me@fedor.page linkedin.com/in/fedor-kabachenko github.com/padobrik

I design and ship end-to-end **machine learning and LLM systems**, from quick prototypes to high-load production. My background is in **distributed systems**, and lately most of my work is **LLMOps** and **agentic workflows**. I care about architectures that stay maintainable and about building what the product actually needs.

Renesas Electronics

Senior Staff AI Analyst

Nov. 2026 – Present

Tokyo, Japan

- Part of the **AI Strategy, Governance & Enablement** team in the Information Systems Division of a global semiconductor company.
- Design, build and deploy **AI-powered tools** that bring **LLM APIs** and **ML models** into products and internal workflows.
- Work with business teams to turn their needs into **AI-driven features**, from building the solution to supporting it in use.

Education

St. Petersburg Polytechnic University

St. Petersburg, Russia

M.Sc. in Biomedical Computing 2021 – 2023

B.Sc. in Medical Biophysics 2017 – 2021

GrainFox

AI Systems Engineer

Feb. 2023 – Oct. 2026

Winnipeg, Canada (Contract)

- Architected and deployed a production-grade **Agentic Ecosystem** of **state-driven, multimodal agents** running diverse business workflows end-to-end with an **85% autonomous task completion rate**.
- Engineered a **Generative Analytics engine (RAG)** over a **continuously updated knowledge base (LlamaIndex, hybrid vector search, optimized chunking)**, cutting manual research time by **90%**.
- Built **LLMOps infrastructure** on **GCP Cloud Run** with a custom gateway for prompt management and smart model routing: full provider independence and a **30% cut in inference costs**.
- Shipped **human-in-the-loop tooling (React, TypeScript, MCP)** that lets operations teams supervise and correct agent decisions without engineering involvement.

Publications

Development of CNN for Defining a Renal Pathology using CT Images

Springer (2022)

Evaluation of Haptoglobin and Its Proteoforms as Glioblastoma Markers

IJMS (2021)

Yandex

Machine Learning Engineer, Search Antifraud

Feb. 2024 – Jan. 2025

Moscow, Russia

- Engineered high-throughput **SERP antifraud models (CatBoost + BERT embeddings)** served via low-latency Python APIs (**<20ms**) across **millions of requests daily**.
- Mined **petabyte-scale clickstream logs (MapReduce, SQL/YQL)** for adversarial behavioral patterns — lifted fraud detection **precision by 6%** without sacrificing recall.
- Owined reliability of distributed model serving end-to-end, cutting production **incident response time by 40%** for search clusters.

Technical Skills

GenAI & LLMOps

Agent Frameworks LangChain, LangGraph, AutoGen

RAG Pipelines LlamaIndex, Haystack, GraphRAG

Evaluation Ragas, DSPy, LangSmith

Vector Infrastructure Vertex AI, Pinecone, Weaviate

Cloud & MLOps

Cloud Platforms GCP, AWS

Infrastructure as Code Terraform

Model Serving vLLM, Triton

Containerization Kubernetes, Docker

Observability Prometheus, Grafana

Data & ML Engineering

Distributed Computing Spark, MapReduce, SQL/YQL

Workflow Orchestration Airflow, MLflow

Deep Learning PyTorch, TensorFlow

Classical ML CatBoost

Software Architecture

Backend Python, FastAPI, Django

Frontend Vue.js, React, TypeScript

Data Stores PostgreSQL, Redis, MongoDB

API Standards REST, OpenAPI, gRPC

Inkast

Machine Learning Engineer

Feb. 2022 – Jul. 2023

St. Petersburg, Russia

- Developed a hybrid **GNN-based Recommender System** behind a **Django API** — **12% uplift in user retention** from better suggestion relevance.
- Contributed across the stack: **PostgreSQL** schema design and **Vue.js** visualization components delivering model insights directly to end users.