Felix Desyatirikov

AI Engineer, **6 year** experience
GMT+3

LinkedIn.com/felixdesyatirikov · Telegram: @felix_des · E-mail: felixdesyatirikov@gmail.com · Github.com/FelixDes

About

AI engeenier with commercial experience of **6 year** working in teams of 5 to 20 people. Integrated platform-based AI solutions into existing and planned information systems. Built the **full development cycle**: from requirements gathering to implementing a system with an automated release pipeline, high test coverage, and using **cutting-edge CI/CD and DevOps practices**.

In addition to core Python/AI expertise, I have great experience working with Kubernetes and microservice architectures. I am interested in the development of the PyTorch ecosystem and continuously update my knowledge in these areas. My hobby is building a personal brand through presentations at professional conferences, publishing scientific papers, and contributing to Open Source.

Skills

AI/ML	Python, FastAPI, LangChain, LangGraph, RAG, NLP, PyTorch, FAISS, Qdrant, PGvector, Bert, MLops, DevOps, TTS, STT, Hugging Face, asyncio, CatBoost, XGboost, LGBM, LoRA, QLoRA, oLLaMA, llama.cpp, Java, Kotlin, mlfow, PostgreSQL
LLM models	Llama, Mistral, OpenAi ChatGPT, DeepSeek, Anthropic, Grok
Big Data	Apache Kafka, Hadoop, Spark, Hive, Impala, Airflow
DevOps	Kubernetes, OpenShift, Microservices, Docker, Vagrant, Docker Compose, CI/CD, Ansible, Linux, Prometheus, Grafana, MinIO, Google Cloud Platform
Architecture	DDD, TDD, BDD, Hexagonal Architecture, Microservices

Experience   6 year

VK
Senior AI Software Engineer

6 month
Jul 2025 - present

- Developed **ML-based partition balancing system** for an in-house stream data processing solution
- Implemented a **RAG system** for incident analysis and **automatic SRE documentation updates**. This **reduced incident resolution time** and accelerated the onboarding of new SRE engineers
- Integrated a metrics and tracing collection system into the **MPC-based system**, which increased observability, improved failure predictability, and accelerated incident localization

Stack: Python, RAG, LLM, FastAPI, LangChain, Java, Spring Boot, CatBoost, YTsaurus

T1 Innotech
Senior Backend Developer

1 year
Jul 2024 - Jul 2025

- Built a **multi-agent MCP-based system** for documentation navigation across different business domains. It helped simplify the documentation writing process and accelerate onboarding
- Implemented an **AI assistant** subsystem for replying to emails in an in-house mail client, taking into account message history and the use of users' personalized master prompts
- Completed the full cycle of **design, implementation, and deployment** of a **client chatbot validation service** based on **Mistral 3 8B** and **LoRA**. The deployment of the service made it possible to ensure the safety of the system's responses

Stack: Python, FastAPI, LLM, Mistral, LoRA, LangChain, Lang Graph, RAG, MCP, PGvector, Apache Kafka

Bell Integrator
Backend Developer

1 year 8 month
Nov 2022 - Jul 2024

- Implemented a scenario for the **RAG system validator module** that detects requests with insufficient access rights and suggests privilege escalation to the user, which eliminated false responses about missing data and **increased system transparency**
- Developed a **semantic query caching** system in the **RAG** system. This improvement made it possible to **reduce latency by 13.5%** for semantically similar queries and increase **user satisfaction**
- Built a notification system for detecting personal data in the knowledge base, which increased the security level of the source data

Stack: Python, FastAPI, LangChain, RAG, NLP, PGvector, Apache Kafka

Inline Group
Backend Developer

1 year 8 month
Mar 2021 - Nov 2022

- Developed a service for displaying analytics on employee action results, increasing workflow analysis speed by **30%**
- Developed a **Computer Vision** service to reduce defect analysis time by **20.7%**
- Increased the number of unit tests from **2,058** to **3,891** to reduce support requests

Stack: Python, FastAPI, OpenCV, YOLO, Docker, PostgreSQL, CI/CD, Java, Spring Boot

Stroyservice
Full-stack Developer

9 month
Jun 2020 - Mar 2021

Stack: Python, FastAPI, SQLAlchemy, PostgreSQL, React, Java

Education

«Applied Mathematics and Computer Science» - Voronezh State University, Voronezh, Russia
Master's degree

«Applied Mathematics and Computer Science» - Voronezh State University, Voronezh, Russia
Bachelor's degree

Certificates

Architecting with Google Kubernetes Engine Specialization   Mar 2024

Papers

Scopus   Web of Science   IEEE*Xplore*

Languages

Russian   Native proficiency

English   Professional working proficiency