

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, LLM, OpenAi, Claude, LLama, RAG, NLP, PyTorch, PGvector, FAISS, Bert, MLops, DevOps, TTS, Hugging Face, asyncio, CatBoost, XGboost, LGBM, Java, Kotlin, mlflow, PostgreSQL
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

Experience

VK	6 year
Senior AI Software Engineer	5 month Jul 2025 - present
<ul style="list-style-type: none">Developed ML-based partition balancing system for an in-house stream data processing solutionImplemented a RAG system for incident analysis and automatic SRE documentation updates. This reduced incident resolution time and accelerated the onboarding of new SRE engineers Stack: Python, FastAPI, RAG, Lang Graph, LangChain, Java, Spring Boot, CatBoost, YTsaurus	
T1 Innotech	1 year
Senior Backend Developer	Jul 2024 - Jul 2025
<ul style="list-style-type: none">Integrated a RAG fusion approach into the system, which made it possible to improve answer quality by 11% and increase user satisfactionImplemented 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 promptsBuilt a multi-agent MCP-based documentation navigation system across different business domains. It helped simplify the documentation writing process and accelerate onboardingIntegrated a metrics and tracing collection system into the application, which increased observability, improved failure predictability, and accelerated incident localization Stack: Python, FastAPI, LangChain, Lang Graph, RAG, NLP, PGvector, MCP, Apache Kafka	
Bell Integrator	1 year 8 month
Backend Developer	Nov 2022 - Jul 2024
<ul style="list-style-type: none">Added logic to the RAG system to detect queries with insufficient access rights and suggest privilege escalation, which eliminated false 'no data' responses and improved system transparencyImplemented a notification system for detecting personal data in the knowledge base, which increased the security of the original data sourceDeveloped a semantic caching system. This improvement reduced latency for similar queries and increased user satisfaction Stack: Python, FastAPI, LangChain, RAG, NLP, PGvector, Apache Kafka	
Inline Group	1 year 8 month
Backend Developer	Mar 2021 - Nov 2022
<ul style="list-style-type: none">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, Java, Spring Boot, OpenCV, Docker, PostgreSQL	
Stroyservice	9 month
Full-stack Developer	Jun 2020 - Mar 2021
<ul style="list-style-type: none">Backend/Frontend development of a corporate logistical system Stack: Java, Spring Boot, Hibernate, PostgreSQL	

Education

«Applied Mathematics and Computer Science»
Voronezh State University, Voronezh, Russia

Certificates

Architecting with Google Kubernetes Engine Specialization	Mar 2024
---	----------

Papers

[Scopus](#) [Web of Science](#) [IEEEXplore](#)

Languages

Russian	Native proficiency
English	Professional working proficiency