

Обо мне

AI-разработчик с коммерческим опытом работы **6 лет** в командах от 5 до 20 человек. Интегрировал платформенные ИИ-решения в действующие и проектируемые информационные системы. Выстроил **полный цикл разработки**: от сбора требований до реализации системы с автоматизированной системой релизов с высоким покрытием тестами и использованием передовых практик CI/CD и DevOps.

В дополнение к основным знаниям по Python/AI имею опыт работы с Kubernetes и микросервисной архитектурой. Интересуюсь развитием экосистемы PyTorch, постоянно обновляю свои знания по этим темам. Мое хобби — развитие личного бренда через доклады на профильных конференциях, публикацию научных статей и вклад в Open Source.

Навыки

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
Architecture	DDD, TDD, BDD, Hexagonal Architecture, Microservices

Опыт работы6 лет

VK	6 месяцев
Ведущий AI Инженер	Июль 2025 - н/в
<ul style="list-style-type: none">Разработал ML-based систему балансировки партиций в in-house решении потоковой обработки данныхРеализовал RAG-систему для анализа инцидентов и автоматического обновления документации SRE. Это позволило сократить время закрытия инцидентов и ускорить подготовку новых SRE инженеров Stack: Python, RAG, LangChain, Java, Spring Boot, CatBoost	

T1 Innotech	1 год
Ведущий AI Инженер	Июль 2024 - Июль 2025
<ul style="list-style-type: none">Интегрировал в систему RAG fusion подход, что позволило повысить качество ответов на 11% и удовлетворенность пользователейРеализовал подсистему AI ассистента для ответа на письма в in-house почтовом клиенте с учетом истории сообщений и использования персонализированных мастер промтов пользователейПостроил многоагентную MCP-based систему навигации по документации разных бизнес-доменов. Она позволило облегчить процесс написание документации и ускорить онбордингИнтегрировал в приложение систему сбора метрик и трассировок, что позволило повысить observability, улучшить прогнозируемость сбоев и ускорить локализацию инцидентов Stack: Python, FastAPI, LangChain, Lang Graph, RAG, NLP, PGvector, Apache Kafka	

Bell Integrator	1 год 8 месяцев
AI Инженер	Нояб. 2022 - Июль 2024
<ul style="list-style-type: none">Добавил в RAG-систему логику, выявляющую запросы с недостаточными правами доступа и предлагающую повышение привилегий, что устранило ложные ответы об отсутствии данных и повысило прозрачность системыРелизовал систему уведомлений о нахождении персональных данных в базе знаний, что повысило уровень безопасности исходного источника данныхРеализовал систему семантического кеширования. Это улучшение позволило понизить latency на похожих запросах и повысить удовлетворенность пользователей Stack: Python, FastAPI, LangChain, RAG, NLP, PGvector, Apache Kafka	

Inline Group	1 год 8 месяцев
ML-ops/Backend Разработчик	Март 2021 - Нояб. 2022
<ul style="list-style-type: none">Разработал сервис просмотра аналитики результатов действий работника, повысив скорость анализа workflow на 30%Разработал Computer Vision сервис для сокращения времени анализа дефектов на 20.7% Stack: Python, FastAPI, Java, Spring Boot, OpenCV, Docker, PostgreSQL	

Стройсервис	9 месяцев
Full-stack Разработчик	Июнь 2020 - Март 2021
<ul style="list-style-type: none">Backend/Frontend разработка корпоративной логистической системы Stack: Java, Spring Boot, Hibernate, PostgreSQL	

Образование

«Прикладная математика и информатика»
Воронежский государственный университет, Воронеж, Россия

Сертификаты

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

Публикации

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

Знание языков

Русский Родной
Английский Professional working proficiency