

Опыт работы | 3 года

Сбер | MLOps инженер

Фев. 2022 - Настоящее время

- Вывод моделей в промышленную среду, с настройкой ML и сетевой архитектуры. Успешно развернуто 4 модели под различные домейны данных
- Аналитика введенных моделей и имплементация архитектур из открытых статей. Использование более эффективного механизма внимания в архитектуре позволило сократить время работы на 30%
- Автоматизация мониторинга моделей с использованием MLFlow, а также пайплайнов данных с помощью Hadoop, PySpark и AirFlow, что упростило общий цикл работы с моделями и данными для нескольких команд

Навыки

ЯП	Python • Golang
DL ML	PyTorch Lightning • XGBoost CatBoost LightGBM • Ray • MLFlow • Ollama • BentoML
Data	GreenPlum • AirFlow • Redis • Weaviate • RabbitMQ • Hadoop PySpark
DevOps	k3s Kubernetes (k8s) • Containerd Docker • GitFlow • NetBird • Rancher
Инструменты	Triton • WandB • Gradio • GitLab
Языки	Русский • Английский (B2)

Проекты

PersIK | LangGraph | Ollama

2025

- Персональный помощник на базе агента на LangGraph, с распознаванием и генерацией голоса в реальном времени

LLMCity | Ollama | k3s | Docker | Ray

2024

- Исследовательский проект симуляции социальной сети на базе кластера предоставленного ВУЗ-ом

VKIS | Svelte | transformers.js | Golang | onnxruntime

2024

- Сервис, предоставляющий возможность выполнять мультимодальный поиск на странице, работающий локально в браузере

NROS | Keras | Django | PostgreSQL | Vue

2020

- Выпускной проект с курсов МШП. CRM система, использующая генетический алгоритм с LSTM, для предсказания спроса и предложения на точках продаж. Команда заняла второе место в конкурсе выпускных проектов

Образование

Финансовый университет при Правительстве РФ, ИТиАБД, Бакалавр, ПМИ

Сен. 2021 - Настоящее время | Москва