

Ветров Алексей Алексеевич

ML / RAG Developer | Python Backend

Москва · Telegram: @aleksey_vet · E-mail: alex_v685@mail.ru · github.com/Alexeyokey

Английский C1 · Готов рассматривать работу в пределах Москвы

ПРОФИЛЬ

Студент 2 курса направления «Системы искусственного интеллекта» МГТУ им. Н. Э. Баумана. Разрабатываю прикладные ML-сервисы: от подготовки документов и retrieval до инференса LLM, оценки качества и контейнерного развёртывания. Быстро осваиваю новые технологии.

ОПЫТ

ML / RAG-разработчик — пилотный проект для компании

Летний срочный контракт, 1 месяц + последующая поддержка проекта · 2026

- Разрабатывал и участвовал в проектировании сервисов в рамках инфраструктуры ИИ-агента для Ведомственного ситуационного центра Минобрнауки России по мониторингу сферы образования и науки.
- Спроектировал RAG-чат-бота для работы по правовым документам: ответы с источниками для ускорения поиска информации и снижения нагрузки на поддержку.
- Реализовал обработку PDF/DOCX/ODT/XLSX/CSV/Markdown через Docling, структурное чанкование, embeddings и безопасную переиндексацию Qdrant со сменой alias.
- Настроил hybrid retrieval: dense search + два BM25-канала, weighted RRF и CrossEncoder; добавил оценку RAGAS/DeepEval и диагностические отчёты.
- Разделил систему на отдельные контейнеры для многосервисной архитектуры; подключил MongoDB, Docker Compose, health-checks и логи. Запустил Qwen3.6-35B-A3B-AWQ через vLLM на 2×RTX 4090 и настроил память, параллелизм и prefix caching.

ПРОЕКТЫ

Поиск объявлений для Авито — система кандидатогенерации: BM25, плотный поиск с дообученной Octen, географические каналы и табличный ранкер. Отбирает до 50 объявлений из корпуса в 189 тыс.; локальный Recall@50 — 0,8438 (5 157 отложенных групп).

Hybrid RAG — опубликованная с разрешения компании ранняя версия рабочего проекта: Python-пайплайн индексации и поиска, Qdrant, BM25, CrossEncoder, vLLM, RAGAS/DeepEval, Docker и тесты.

ChessAI — desktop-приложение на PyQt5 с шахматным движком и CNN на PyTorch для оценки позиций; представлен на конференции «Инженеры будущего».

English Learning Bot — Telegram-бот на aiogram: словарь, диктанты, статистика и AI-собеседник.

ТЕХНОЛОГИИ

ML / LLM: PyTorch, Transformers, SentenceTransformers, embeddings, CrossEncoder, RAG, RAGAS, DeepEval, vLLM

Search / Data: Qdrant, dense retrieval, BM25, weighted RRF, Docling, MongoDB, Pandas, NumPy

Backend / Infra: Python, FastAPI, Pydantic, Docker Compose, Linux, Git, pytest

AI-инструменты: поиск ошибок, code review, тесты и документация в рамках требований проекта

Применял ранее: PyQt5, Keras, aiogram, Flask, SQLAlchemy, Peewee, Pygame, JavaScript, C++

ОБРАЗОВАНИЕ И ДОСТИЖЕНИЯ

МГТУ им. Н. Э. Баумана — бакалавриат, «Системы искусственного интеллекта», 2 курс; последняя сессия — на отлично.

Яндекс Лицей — программа по разработке на Python, окончил с отличием.

Призёр предпрофессиональной олимпиады (проект SmartFridge); участник программы «Шаг в будущее».

Языки: русский — родной; английский — C1; китайский — изучаю (текущий уровень A2).