ЭДУАРД ХАН

AI Product Engineer | Backend & Agentic Systems

fanatsira@gmail.com | github.com/fanat-sira | fanat-sira.art | t.me/fanat_sira | Удаленно

ПРОФИЛЬ

Разрабатываю AI-продукты и backend-сервисы от идеи до production: проектирую решения, интегрирую AI с внутренними данными и бизнес-процессами и довожу системы до стабильной эксплуатации. Использую AI-агентов как основной инструмент разработки, при этом самостоятельно принимаю архитектурные и технические решения, контролирую реализацию и отвечаю за итоговый результат. Работаю с LLM, RAG, эмбеддингами, реранжированием, обработкой речи и инференсом моделей. Фокусируюсь на качестве, стоимости, автоматизации и производительности процессов.

ОПЫТ

Элитный Сочи | Агентство недвижимости **Июнь 2025 - настоящее время**

Backend & Product Engineer

Система проверки звонков

- Самостоятельно спроектировал, разработал и запустил production-систему автоматической проверки звонков.
- Сервисом пользуются сотни сотрудников, ежемесячно он обрабатывает десятки тысяч проверок, ручное вмешательство требуется только в исключительных случаях.
- Реализовал backend, веб-интерфейс, автоматизацию обработки и инфраструктуру для масштабирования.
- При росте числа проверок примерно в 4 раза снизил стоимость одной проверки в десятки раз за счет автоматизации и оптимизации системы.
- После релиза отвечаю за развитие, стабильность и поддержку системы.

AI-ассистент по сделкам

- Разработал внутреннего AI-ассистента, объединяющего в одном интерфейсе данные сделки, звонки, транскрипты, результаты проверок, учебные материалы и внутреннюю базу знаний.
- Ассистент анализирует контекст сделки, помогает менеджеру определить следующие действия и генерирует черновики ответов клиентам.
- Довел продукт от идеи до рабочего внутреннего пилота.

Сервис семантической маршрутизации для голосового бота

- Предложил и разработал сервис классификации ответов для существующего голосового бота.
- Снизил среднюю задержку маршрутизации с 800-1000 мс до ~200 мс.
- Снизил стоимость одного запроса примерно в 5 раз.
- Около 99% запросов обрабатываются без обращения к резервной LLM.

ТЕХНИЧЕСКИЕ НАВЫКИ

AI и ML	LLM, RAG, эмбеддинги, реранжирование, инференс моделей, OpenRouter, Whisper, T-One, Yandex Cloud (AI Studio, SpeechKit)
Backend и данные	Go, Python, REST API, PostgreSQL, Redis, pgvector
Инфраструктура	Linux, развертывание и эксплуатация VPS/VDS, Docker, Docker Compose, Nginx, Git, CI/CD
Инструменты разработки	Copilot, Codex, Antigravity, Claude Code