

Сергей Шумов

Разработчик AI/LLM-систем · RAG · Backend-интеграции

EMAIL shumov_sv@mail.ru
WHATSAPP +66 944 49 57 81

ИНТЕРАКТИВНОЕ ПОРТФОЛИО

Сайт sergios.dev



Проектирую прикладные AI-системы: LLM-агенты, RAG, интеграции. Специализируюсь на сценариях, где LLM работает вместе с данными, поиском, внешними API и управляемой логикой.

3,9M+

ОБЪЕКТОВ В ПОИСКОВОМ ИНДЕКСЕ

40x

СНИЖЕНИЕ НАГРУЗКИ НА БД

99,9%

НОРМАЛИЗАЦИЯ КЛЮЧЕВЫХ ПОЛЕЙ

5+

РЕАЛЬНЫХ AI-ПРОЕКТОВ

01 FLAGSHIP-ПРОЕКТ

CarHao — векторный поиск по 4 млн автомобилей

02.2026

Поисковый AI-ассистент: RAG, Qdrant, LLM-оркестрация, обновление индекса // ИП Заяц П. Е.

- Гибридный поиск по **3,9 млн автомобилей**: пользователь пишет свободный запрос, система ищет по смыслу и фильтрам.
- LLM переводит диалог в **строгий JSON**: бюджет, марка, кузов, пробег, город — дальше работает управляемый поиск.
- Индекс обновляется по марке изменений: пересчитываются только новые и изменённые записи. Нагрузка на БД **упала ~40x**.
- Данные из **5 источников** приведены к единому формату: топливо, коробка, кузов, статусы. Точность ключевых полей — **99,9%**.

STACK Python 3.11 · FastAPI · Qdrant · PostgreSQL · MongoDB · Redis · OpenAI · Telegram bot · Linux/VPS

Архитектура, логи и примеры работы

sergios.dev/carhao



02 ДРУГИЕ ПРОЕКТЫ

2026	ТАОБАО — AI-маркетплейс ↗	ООО «ТАОБАО»	Сервис для публикации AI-подрядчиков: проверка заявок, подключение клиентов и запуск экземпляров.
FastAPI PostgreSQL SQLAlchemy			
2025	ILFA Dental ↗	ООО «КЭС»	AI-ассистент клиники в WhatsApp: отвечает пациентам, ведёт к записи и работает с 1Denta.
FastAPI 1Denta Wazzup FAISS			
2024	Nauka.Club ↗	ООО «ГК ОДА»	Образовательный AI-ассистент: задачи по фото, сочинения, конспекты и чат в одном API.
FastAPI OpenAI FAISS Redis			
2024	AI-аналитик медиаконтента ↗	ООО «Университет искусственного интеллекта»	Анализ видео и аудио: транскрибация, выжимка и сокращение ручной работы до минут.
Whisper OpenAI youtube-dl			
2023	Wildberries ↗	ООО «Вайлдбериз»	AI-ассистент техподдержки: ответы по внутренним инструкциям и контроль качества.
Telegram TTS			
2023	Планета Здоровья ↗	ООО «Планета Здоровья»	AI-менеджер чата для аптечной сети: ответы по базе знаний и маршрутизация запросов.
RAG OpenAI			

PYTHON

\$ async \$ FastAPI \$ Pydantic \$ httpx

LLM

\$ OpenAI \$ Anthropic \$ function calling \$ JSON-контракты \$ OCR \$ Whisper

RAG

\$ Qdrant \$ FAISS \$ embeddings \$ гибридный поиск

DB

\$ PostgreSQL \$ MongoDB \$ Redis \$ MySQL

AI-СИСТЕМЫ

\$ LLM-оркестрация \$ prompt engineering \$ policy-логика \$ retrieval

ИНТЕГРАЦИИ

\$ 1Denta \$ Wazzup \$ Telegram \$ Google Sheets/Drive \$ Bitrix24

ИНФРА

\$ Linux/VPS \$ Docker \$ Git \$ systemd