Витовцев Владислав Игоревич

AI Engineer (Agentic Systems, RAG, Fullstack)

Почта: vlad@vitovtsev.site **Telegram:** @vladislav_vitovtsev **Сайт:** vlad.vitovtsev.site

Живу в Черногории (CET), удаленно **Договор:** d.o.o. для контрактов в EUR или ИП в России (RUB)

КОРОТКО

AI-инженер с 13-летним опытом fullstack-разработки (Node.js, Python, React). Последний год создаю агентные AI-системы и встраиваю их в работу команды как часть процесса разработки. Спроектировал и запустил RAG-сервис, который сократил контекст к модели с 43 КБ до 5 КБ на запрос при recall@5 = 1.00. Собрал агентное ревью кода, которое на одной фиче вернуло 11 из 17 задач с реальными дефектами до мержа: open redirect, гонки состояний.

Довожу до прода то, что обычно застревает на демо: обработку сбоев инструментов, эскалацию человеку, память между запусками, стоимость контекста. Веду продукт от идеи до работающего сервиса: бэкенд, фронтенд, CI/CD.

Открыт к проектной работе и полной занятости. По умолчанию работаю асинхронно: постановка и ревью в тексте.

ОПЫТ РАБОТЫ

Продуктовая IT-компания · продукт видеосвязи, под NDA — сентябрь 2025 - настоящее время

Единственный фронтенд-разработчик проекта

- Вырастил из демо-прототипа на несколько десятков файлов боевой веб-клиент: звонки на WebRTC с записью, календарь с повторяющимися встречами, чат, уведомления на SSE, виртуальный фон, шеринг экрана.
- Запустил десктоп-клиент на Electron с автообновлениями по каналам.
- Построил CI/CD под macOS и Windows с подписью кода (Developer ID с нотаризацией, аппаратный токен по PKCS11): **49 релизов за 7 месяцев**, в среднем 1-2 в неделю. Перевел проект на монорепозиторий, около 3 500 тестов.
- Собрал агентный конвейер разработки: спецификация, план, test-first, ревью отдельными агентами по доменам (безопасность, корректность, производительность). На фиче из 18 задач ревьюеры вернули **11 из 17** задач с реальными дефектами: open redirect, две гонки счетчика непрочитанных, утечку подписки. Все закрыто до мержа.
- Агент разбирает инциденты по логам и метрикам Grafana и предлагает задачу с причиной и вариантами фикса; в трекер она попадает после моего подтверждения.

Независимая практика · консалтинг, параллельно основной работе — ноябрь 2023 - настоящее время

Ведущий разработчик / AI-инженер: технический директор, единственный разработчик, разработчик в команде

Веду несколько клиентских проектов параллельно: продукты с нуля и внедрение AI.

- **SEO-платформа** (2026, AI-инженер): AI-модуль анализа контента для сайта заказчика. Docker-сервис с API, локальная модель эмбеддингов, LLM-судья по утверждениям, автоподбор конкурентов из выдачи Яндекса. Перевел судью с оценки «на глаз» на проверку утверждений и показал, какие из них уже есть у конкурентов. Перенос судьи на быструю модель ускорил проверку с 16,5 до 5,6 с. Аудит закрыл SSRF, гонку лимита параллельности и CSV-инъекцию, 145 тестов. Вторым сервисом сделал робота-асессора качества страниц по методике асессоров поиска.
- **Caerus** (2024-2026, единственный разработчик): сайт проката одного из основателей GlobalRental. В 2024 поднял его на движке GlobalRental вторым брендом, в 2025 развел фронтенды, в 2026 отделил целиком: оба сайта крутили один бинарник на одной базе без мультитенантности, поэтому сначала аудит и точка отката, потом своя база и бэкенд, перенос броней и кабинета без потерь. Дальше багфиксы и поддержка. Кабинет стал рабочим календарем проката (около 1400 записей за два года), прямых броней с сайта 13 за 2025 год и 31 за 2026-й: **caerusrentacar.com**
- **Платформа психотерапии** (2025-2026, фронтенд в команде): фича «Супервизии» целиком, от поиска супервизора до оплаты, переноса сессий и календаря. Next.js 15, Nx, Feature-Sliced Design. Вел ее через свой агентный процесс: 4-5 часов против ручной оценки в 19.
- **EVIRMA** (2025, фронтенд в команде 7-8 человек): Chrome-расширение на Manifest V3 с аналитикой рекламы для продавцов Wildberries. Воронка продаж, статистика кампаний с расчетом ДРР и СРР, экспорт в Excel. Второй по вкладу в команде после техлида.
- **GlobalRental** (2023-2025, технический директор): агрегатор аренды авто в Черногории. В 2023 вел студию-подрядчика со стороны инвестора: ТЗ, приемка, прод и инфраструктура. В 2024 писал код вместе с разработчиком студии: почта, Telegram-бот, админка со статистикой. С декабря 2024 единственный разработчик фронта и бэка: модерация, валюты, SEO, аренда яхт. Во фронтенде **331 коммит из 645**, в бэкенде **136 из 294**. Stripe проверил боевым платежом, но проект закрыли раньше, чем онлайн-оплатой начали пользоваться клиенты.

App In The Air · США, travel-приложение, более 6 млн пользователей — 2019 - 2024

Frontend-разработчик → тимлид команды из 5 человек

- Пришел, когда у компании было мобильное приложение и не было веба, и сделал веб-продукт с нуля: поиск и покупка авиабилетов на socket.io, затем отели и white-label виджет отелей для авиакомпаний.

- Собрал фронтенд-команду: найм, онбординг, развитие. Техинтервью и тестовые с разбором кода вел сам.
- Первую версию корпоративного B2B-портала сделали вдвоем с коллегой, дальше вел его фронтенд: сотрудники, корпоративные карты, бронирование в рамках travel policy, отчеты по расходам.
- Перевели фронтенд на монорепозиторий Nx, обновил React 17 → 18, внедрил Next.js, настроил отдельный стенд в Cloud Run на каждую релизную ветку.

webrise · веб-агентство — август 2016 - ноябрь 2019

Fullstack-разработчик

- Редизайн сайта Сбербанк Страхование и около пятнадцати страховых калькуляторов с покрытием расчетов тестами, калькуляторы для Ингосстраха.
- Платформа всероссийского конкурса с видео и чатом в браузере, интернет-магазин Усачевского рынка с приложениями на React Native.

МФК Займер — апрель 2014 - август 2016

Fullstack-разработчик

- Выдача микрозаймов на карту без участия человека, Россия и Казахстан. Laravel 5, MySQL, PostgreSQL, MongoDB, Selenium.

it-infinity — август 2013 - март 2014

Frontend-разработчик, корпоративные сайты на 1С-Битрикс

СВОИ AI-СИСТЕМЫ И ПРОЕКТЫ

- **RAG-сервис в проде**, Python и FastAPI: гибридный поиск (эмбеддинги плюс BM25 через SQLite FTS5, слияние RRF), без отдельной векторной БД. Контекст к модели упал с 43 КБ до 5 КБ на вопрос; на 45 контрольных вопросах recall@5 = 1.00, p95 поиска меньше 350 мс. При сбое поиска запрос откатывается на полную базу знаний, посетитель сбоя не видит.
- **vlad.vitovtsev.site**: портфолио с живым AI-агентом на входе. Fastify, стрим через SSE, потолок расходов, защита от prompt-injection, лимиты по IP.
- **Память для агентов** между сессиями: дневные срезы, журнал решений, авто-чекпоинты, свой поиск по 20 тысячам фрагментов. Стартовый контекст снизил с 21 700 до 12 000 токенов на запуск. 23 своих расширения для агентной среды, набор под API Яндекса на основе open-source.
- **Helm**: пульт для параллельных сессий агентов с телефона. Svelte 5, терминал в браузере, PWA, вход по passkey, около 1 200 тестов.
- **trainerpad.ru**: офлайн-первая CRM для персональных тренеров, соло с нуля. Свой движок синхронизации на IndexedDB с разрешением конфликтов.

СТЕК

AI	агентные системы, RAG, гибридный поиск, эмбеддинги, промпт-инжиниринг, LLM API (Anthropic, OpenAI-совместимые), MCP, контроль стоимости токенов
Бэкенд	Node.js, Python, FastAPI, Fastify, Express, PostgreSQL, SQLite, Prisma
Фронтенд	TypeScript, React, Next.js, Electron, React Native, Flutter, WebRTC, SSE, Vitest, Playwright
Инфра	Docker, GitLab CI, GitHub Actions, Linux, Nginx, Caddy, свои VPS

СЕРТИФИКАТЫ

- **Сертификат квалификации GREEN-API** № 2243758, сентябрь 2026: интеграции и разработка чат-ботов на базе GREEN-API (шлюз к WhatsApp, MAX и Telegram). Практическая часть: веб-чат для MAX на их HTTP API, React 19, TypeScript, Zustand, очередь уведомлений, статусы доставки, 385 тестов.
  Демо: max-chat.vitovtsev.site

ОБРАЗОВАНИЕ И ЯЗЫКИ

Томский государственный университет систем управления и радиоэлектроники, 2012, вычислительная техника, инженер. Русский родной. Английский: свободная переписка, это мой основной рабочий формат.