



Борисов Павел Васильевич

Мужчина

+7 (911) 7258223 — github.com/pavelveter

t.me/pavelveter

kayjaylitee@gmail.com — предпочитаемый способ связи

telegram: @pavelveter

Проживает: Санкт-Петербург, м. Проспект Славы

Гражданство: Россия, есть разрешение на работу: Россия

Не готов к переезду, готов к командировкам

Желаемая должность и зарплата

Senior AI Engineer / AI Infrastructure Engineer

Специализации:

- DevOps-инженер
- Программист, разработчик
- Руководитель группы разработки
- Системный инженер
- Технический директор (СТО)

Тип занятости: полная занятость, частичная занятость, проектная работа/разовое задание

Формат работы: на месте работодателя, удалённо, гибрид

Желательное время в пути до работы: не имеет значения

Опыт работы — 22 года 8 месяцев

Март 2019 —
настоящее время
7 лет 5 месяцев

AI Infrastructure & Highload Systems (Консалтинг / Аутсорсинг)

Lead Engineer

Проектирование, масштабирование и миграция инфраструктуры под высокие нагрузки и AI/ML-сервисы для клиентов из сфер e-commerce, AdTech и IoT.

— Развертывание и контейнеризация ИИ-решений: пайплайны компьютерного зрения (CV), инференс локальных LLM и интеграция RAG-систем.

— Построение отказоустойчивой архитектуры (Kubernetes, Docker) и автоматизация CI/CD процессов для сокращения Time-to-Market.

— Оптимизация и тюнинг производительности баз данных (PostgreSQL) под высоконагруженные аналитические задачи.

— Управление полным циклом инженерных работ: от сбора бизнес-требований до релиза в Cloud / Bare Metal.

Октябрь 2025 —
Декабрь 2025
3 месяца

АТТЭК

Санкт-Петербург

Лектор / Ментор по AI, QA и Backend

Разработка и проведение практического курса по Python, SQL, AQA и современным AI-воркфлоу.

— Обучил 26 студентов (уровня от pre-junior) до production-ready.

— Курировал создание реальных проектов с использованием LLM и Cloud PostgreSQL.

— Внедрял лучшие практики бэкэнд-разработки, тестирования и проектирования систем.

— Выступал в роли ментора и тимлида на финальных проектах.

— Фокус обучения: практические инженерные навыки, а не сухая теория.

Ноябрь 2023 —
Август 2025
1 год 10 месяцев

Selfiebot.ru

Founder

Создание «с нуля до первых продаж» платформы для лидогенерации и распознавания лиц на базе ИИ.

- Спроектировал и реализовал REST API (FastAPI) и пайплайны компьютерного зрения (DLib, OpenCV).
- Интегрировал системы LLM (GPT, локальный запуск), привело к росту конверсии в маркетинге.
- Разработал фронтенд (React + Tailwind).
- Спроектировал инфраструктуру (Docker, Debian, стек мониторинга).
- Снизил время простоя (downtime) до <1% благодаря внедрению observability (Prometheus, Grafana).
- Вывел проект на первую выручку и базу в ~1000 активных пользователей.
- Полная ответственность: продукт, архитектура, инфраструктура и командные процессы.

Март 2014 —
Март 2017
3 года 1 месяц

Raskachaem.ru

Архангельск, raskachaem.ru

Информационные технологии, системная интеграция, интернет

- Интернет-компания (поисковики, платежные системы, соц.сети, информационно-познавательные и развлекательные ресурсы, продвижение сайтов и прочее)

СМИ, маркетинг, реклама, BTL, PR, дизайн, продюсирование

- Маркетинговые, рекламные, BTL, дизайнерские, Event-, PR-агентства, организация выставок

Региональный директор

Руководство региональными операциями на конкурентном медиа-рынке.

- Управление распределенной командой (~10 фотографов и контент-мейкеров).
- Координация интенсивного освещения мероприятий в сжатые сроки.
- Выстраивание партнерских отношений с владельцами заведений и СМИ.
- Переговоры о коммерческом сотрудничестве и рекламных кампаниях.

Ноябрь 2003 —
Декабрь 2015
12 лет 2 месяца

ЗАО «СИА Интернейшнл-Архангельск»

Архангельск, www.nordsia.ru

Руководитель IT-отдела, DBA

Управление ИТ-инфраструктурой и архитектурой данных крупного фарм-дистрибьютора, работающего (центральный хаб + 2 региональных филиала, 200+ корпоративных клиентов). В подчинении 6 специалистов.

- Zero Downtime миграция: Успешно спроектировал и реализовал бесшовный переезд предприятия и дата-центра в новое здание. Вся инфраструктура, серверный парк и каналы связи запущены с нуля без остановки операционной деятельности и отгрузок.
- High Availability: Разработал и масштабировал на филиальную сеть распределенную систему обработки заказов с отказоустойчивой архитектурой.
- Руководил миграцией core-ERP системы без потери данных и технологических простоев.
- Внедрил комплексные системы мониторинга, резервного копирования и электронного документооборота под нужды филиальной сети и клиентов.
- Разработка ПО: проектирование и оптимизация специализированного софта для оптовой торговли (Python, Bash, SQL).
- Полный цикл управления: планирование и защита ИТ-бюджетов, управление рисками, найм, онбординг и выстраивание кросс-функционального взаимодействия.

Январь 2002 —
Ноябрь 2003
1 год 11 месяцев

Мэрия Города Архангельска

Архангельск, arhcity.ru

Lead Engineer

- Проектирование и поддержка локальных сетей (LAN) для административных департаментов.
- Взаимодействие с внутренними подразделениями для внедрения IT-решений.
- Техническая поддержка и решение инфраструктурных задач в условиях жестких регламентов.

Образование

Высшее

2002
Высшее

Северный (Арктический) федеральный университет им. М.В. Ломоносова, Архангельск

Институт информационных и космических технологий, Системы автоматизированного проектирования

Навыки

Знание языков
Русский — Родной
Английский — B2 — Средне-продвинутый

Навыки

Автоматизация

Серверное программирование

Linux

Bash

Python

Docker

Nginx

DevOps

Мониторинг

Ansible

MariaDB

Git

Grafana

Golang

NATS

n8n

RAG

LangChain

LangGraph

LangFuse

Qdrant

PaddleOCR

PostgreSQL

pgvector

OpenCV

vLLM

NVIDIA Container Toolkit

Kubernetes

AI

Agents

Дополнительная информация

Обо мне

Senior AI Infrastructure Architect / Lead DevOps Engineer с 20+ годами опыта в проектировании высоконагруженных систем и платформ для ИИ-продуктов. Основатель и экс-СТО ИИ-стартапа SelfieBot.

Специализируюсь на создании production-ready инфраструктуры для локального инференса LLM/VLM, RAG-систем и автоматизации MLOps-процессов как в Cloud, так и на Bare Metal.

- Ключевая экспертиза:
- AI/ML Infrastructure: Развертывание и оптимизация локального инференса (vLLM, NVIDIA Container Toolkit, CUDA), построение RAG-архитектур (Qdrant, pgvector, LangChain/LangGraph, LangFuse).
 - Оркестрация и DevOps: Проектирование отказоустойчивых кластеров (Kubernetes, Docker), IaC (Ansible), CI/CD, построение систем мониторинга (Grafana).
 - Backend & Data: Разработка высоконагруженных сервисов на Go и Python (Asyncio, FastAPI), архитектура распределенных систем (NATS).
 - Продуктовое мышление: Опыт управления командами и запуска ИИ-продуктов с нуля (как СТО). Понимаю бизнес-метрики, архитектуру данных и дизайн-процессы.

- Технический стек:
- Languages: Python (asyncio, FastAPI, Multiprocessing, uv), Bash, Go.
 - AI/ML & MLOps: Kubernetes, vLLM, LangChain, LangGraph, LangFuse, Qdrant, pgvector, PaddleOCR, OpenCV.
 - Data & Infra: PostgreSQL, NATS, Redis, Ansible, Docker, Grafana.

Дополнительно:

- Серебряный призер олимпиады по программированию.
- Автор публикаций и спикер на интервью на тему ИИ и технологий.
- Провожу воркшопы по коммуникациям и управлению в условиях хаоса.

Интересы: AI Infrastructure, Computer Vision, LLM Agents, фото/аудио техника, Digital Storytelling.

telegram: t.me/pavelveter; github.com/pavelveter

GitHub Valuable OSS Projects:

- <https://github.com/pavelveter/hermem> — сверхбыстрая графовая система памяти (GraphRAG) для AI-агентов. Стек: Go + SQLite. Решает проблему долгосрочной ассоциативной памяти с минимальным латенси.
- <https://github.com/pavelveter/research-agents-swarm> — автономный рой из 5 агентов с паттерном LLM-as-a-judge для глубокого интернет-ресёрча и генерации отчетов. Стек: LangGraph + LangFuse (observability/кости) + Qdrant.
- <https://github.com/pavelveter/llm-gateway> — легковесный прокси-сервер для унификации b2b-инференса. Единый эндпоинт для работы со множеством LLM-бэкендов, дающий отказоустойчивость при падении бэкенда.
- <https://github.com/pavelveter/recognition> — модуль распознавания и матчинга лиц (Face ID / Computer Vision) для коммерческого проекта SelfieBot. Автоматически сопоставляет селфи гостей с репортажными фотоотчетами. Стек: Python + uv + параллельные CPU-вычисления + кэширование эмбеддингов + API Яндекс.Диска. Оптимизирован под высокую скорость обработки больших массивов изображений.
- <https://github.com/pavelveter/regognition> — то же самое, но на Go, быстрее, точнее (512-мерные векторы против 128-мерных), менее требовательно по памяти и для локальных задач.

Медиа & Экспертиза:

Регулярно выступаю профильным экспертом в федеральных СМИ (в т.ч. ИА «Росбалт», охват 8 млн уникалов/мес) по темам AI-технологий и ИТ-инфраструктуры:

- Анализ рисков SSL-сертификатов и ограничений интернета (аналитика проведена с помощью моего инструмента research-agents-swarm): <https://www.rosbalt.ru/news/2026-06-17/v-semirnaya-pautina-rvetsya-kto-budet-pervym-v-gonke-ogranicheniy-interneta-5619135>
- Экспертиза по строительству новых ЦОД и ИТ-ландшафту: <https://www.rosbalt.ru/news/2026-04-08/novyy-tsod-poyavitsya-v-peterburge-zachem-on-nuzhen-5577204>
- Разбор природы галлюцинаций нейросетей и верификации данных: <https://www.rosbalt.ru/news/2025-08-14/neyroseti-obmanyvayut-lyudey-a-ii-sozdaet-tsifrovye-gallyutsinatsii-kto-za-etim-stoit-5456186>