

Степан Федянов – Senior Software Engineer

Монтевидео, Уругвай (GMT-3) • stepan@fedyanov.ru • <https://www.linkedin.com/in/stepanfedyanov/>

Senior Software Engineer с 6-летним опытом разработки систем с фокусом на data-intensive applications, распределённую инфраструктуру и AI-powered решения. Имею опыт проектирования и реализации LLM-driven функций, включая обработку естественного языка и retrieval-based системы поверх сложных бизнес-данных. Особое внимание уделяю точности ответов, детерминированной обработке данных и минимизации hallucinations в AI-системах.

ТЕХНИЧЕСКИЕ НАВЫКИ

Backend: Python, Django, Django REST Framework, FastAPI, Celery, Redis, GraphQL, PostgreSQL, ChromaDB.

AI / LLM: интеграция LLM, prompt engineering, structured output generation, natural language processing, retrieval-based systems over structured data, hallucination mitigation (validation layers, deterministic parsing), RAG, context-aware response generation, AI-assisted automation tools, LangGraph/LangChain.

Architecture: System Design, Domain-Driven Design (DDD) basics, Clean Architecture principles, API-first development.

Frontend: JavaScript, TypeScript, Vue 2/3, Vuex/Pinia, Nuxt 2/3/4, Vite.

DevOps: Docker, CI/CD (Gitlab, Github Actions), Nginx, Linux (Ubuntu), AWS.

Testing: Vue Test Utils, Jest, Playwright, Cypress, PyTest.

Workflow: Git (Gitflow), Jira, Agile, Code Review & Mentorship.

ОПЫТ РАБОТЫ

Applied AI Engineer
CreativeCo Studio

Май 2026 – настоящее время
Удаленно, фулл-тайм

Senior Software Engineer
Сибзолото

Август 2024 – Май 2026
Удаленно, парт-тайм

Крупнейшее золотодобывающее предприятие в Сибири, управляющее множеством распределенных производственных площадок со сложной логистикой и потоками данных (более 2000 сотрудников).

- Спроектировал и реализовал full-stack ERP-систему для учета золотых запасов и операционной аналитики для 250+ пользователей, заменив ручные Excel-процессы и сократив время подготовки отчетности с 2 часов до 30 минут.
- Оптимизировал обработку больших объемов данных в PostgreSQL за счет переписывания критически важных запросов, добавления составных индексов и внедрения материализованных представлений, сократив время генерации управленческих отчетов с 50 минут до менее чем 2 минут.
- Разработал распределенный слой синхронизации данных на Celery и REST для географически изолированных производственных площадок с нестабильным соединением, обеспечив надежную и идемпотентную доставку данных между всеми узлами.
- Спроектировал multi-agent LLM-систему аналитики на LangGraph с 3 агентами, которая преобразует запросы на естественном языке в полностью отрисованные HTML-дашборды. Система включает агента понимания запроса, который сопоставляет intent со структурированной спецификацией дашборда, агента извлечения данных, который генерирует и валидирует SQL, и агента построения дашборда, который рендерит HTML/JS-результат, позволяя не техническим менеджерам получать инсайты без поддержки аналитиков.
- Реализовал pipeline понимания запросов, который обогащает промпты доменным контекстом, извлеченным через RAG, включая бизнес-правила и определения метрик, и использует schema-constrained outputs на Pydantic для извлечения сущностей, временных диапазонов, типов графиков и KPI в валидированный JSON-формат, обеспечивая детерминированное выполнение последующих этапов.
- Разработал retrieval-augmented слой контекста: разбил внутреннюю документацию, бизнес-правила и схему БД на чанки, сгенерировал embeddings и сохранил их в ChromaDB. Во время выполнения запроса система извлекает наиболее релевантные чанки через semantic search, чтобы grounding LLM-ответов снижал hallucinations в определениях метрик и использовании схемы.
- Разработал автоматизированный pipeline, который парсит Django migration files, извлекает и нормализует metadata схемы базы данных в структурированный JSON, поддерживая LLM-контекст в актуальном состоянии и повышая точность text-to-SQL генерации.
- Реализовал безопасный text-to-SQL pipeline с несколькими уровнями валидации: LLM-проверками согласованности со схемой, статической SQL-валидацией и EXPLAIN dry-runs перед выполнением. В сочетании с read-only доступом, whitelist таблиц и retry logic это сократило количество невалидных запросов и ошибок, вызванных hallucinations, почти до нуля в production.

Стек: Vue 3, TypeScript, Python, Django/DRF, PostgreSQL, Celery, LangGraph, LLM (OpenAI/Claude/Ollama), RAG (ChromaDB).

Senior Software Engineer
Студия Артемия Лебедева

Студия № 1 в России, специализирующаяся на сложном дизайне и технологических решениях, включая веб-приложения и инновационные AI-продукты (Фокус, Николай Иронов).

Май 2024 – Май 2026
Удаленно, фулл-тайм

- Спроектировал ключевые функции для Fokus — AI-powered конструктора презентаций с более чем 100 000 пользователей, включая интерактивные элементы canvas и сценарии монетизации на базе Stripe.
- Спроектировал и реализовал распределённый LLM-powered пайплайн генерации презентаций, который преобразует тему, исходный текст или URL в полностью структурированные слайд-деки. Реализовал предобработку веб-контента — scraping и извлечение текста — и объединил все типы входных данных в единый процесс генерации.
- Реализовал этап генерации плана на Celery, который создаёт структурированный outline презентации с типами слайдов и описаниями контента. Добавил доступные шаблоны слайдов напрямую в prompt и внедрил строгую Pydantic-валидацию с автоматическими повторами, обеспечив детерминированные и валидные результаты.
- Построил систему параллельной генерации слайдов с использованием Celery groups и chords, где каждый слайд генерируется независимо. Каждая задача заполняет поля шаблона через LLM и генерирует prompts для изображений, которые затем обрабатываются через внешние API. Изолировал ошибки на уровне отдельных слайдов с помощью retry-логики, предотвращая сбой всего пайплайна.
- Спроектировал финальный этап агрегации — Celery chord callback, который собирает все слайды в валидированную JSON-структуру и сохраняет её в базу данных, включая проверки целостности перед записью.
- Спроектировал и реализовал AI-powered пайплайн code review, который анализирует Git diffs и предоставляет контекстно-зависимую обратную связь напрямую в GitLab Merge Requests на основе labels и scope изменений.
- Реализовал слой предобработки, который парсит diffs, фильтрует шум и обогащает контекст связанными файлами и метаданными, позволяя LLM рассуждать не только об изолированных изменениях, но и об их влиянии на окружающую кодовую базу.
- Спроектировал prompt pipeline, который динамически адаптирует глубину review — style, architecture, logic — на основе MR labels и типов файлов, обеспечивая релевантную и точечную обратную связь вместо generic-комментариев.
- Построил систему структурированной обратной связи, в которой LLM-результаты нормализуются в категоризованные комментарии — bugs, risks, improvements, style — что обеспечивает единый формат и удобное восприятие разработчиками.
- Интегрировал систему в GitLab CI/CD как автоматизированный этап, публикующий inline comments и summaries напрямую в Merge Requests, сократив время review на 67% и позволяя находить логические ошибки до human review.
- Реализовал автоматизированное тестирование — Playwright, Vitest — и мониторинг через Sentry, увеличив crash-free sessions с 89% до 98%.

Стек: Vue 2/3, Nuxt 2/3, TypeScript, Pinia, Node.js, OpenAI/Gemini, Python (Django/DRF), PostgreSQL, Playwright, Vitest, Gitlab CI/CD, FSD.

Software Engineer
Студия Артемия Лебедева

Июнь 2023 – Май 2024
Удаленно, фулл-тайм

- Разработал ключевые генеративные и фронтенд-функции для Николая Иронова, первого AI-генератора фирменного стиля для малого бизнеса, масштабируя продукт до тысяч уникальных дизайн-проектов.
- Возглавил крупную инициативу по рефакторингу производительности, переведя фронтенд на использование серверного рендеринга (SSR) и устранив устаревшие зависимости. Это улучшило показатели Google PageSpeed с 40 до 80 и обеспечило рост органического трафика примерно на 75%.
- Внедрил и обеспечил соблюдение архитектурного паттерна Feature-Sliced Design (FSD) в крупном legacy-коде, что значительно повысило ясность кода, модульность и долгосрочную поддерживаемость в команде разработчиков.
- Спроектировал и построил архитектуру для парка высокоавтоматизированных микросервисов для генерации лидов. Эта масштабируемая система быстро стала основным источником входящих заявок для продукта, требуя минимального вмешательства разработчиков для развертывания и A/B тестирования.

Стек: Vue 2/3, Nuxt 2/3, JavaScript/TypeScript, Vuex/Pinia, NodeJS, SSR, SEO/GEO Optimization, FSD Architecture, Django/Python (API Integration), PostgreSQL.

Web Developer
Kwork

Январь 2021 – Июнь 2023
Удаленно, фриланс

ОБРАЗОВАНИЕ

Сибирский государственный университет им. М. Ф. Решетнева
Бакалавр, Информационные системы и программирование, 2025