

Stepan Fedyanov – Senior Software Engineer

Montevideo (GMT-3) · stepan@fedyanov.dev · <https://www.linkedin.com/in/stepanfedyanov/>

PROFESSIONAL SUMMARY

Senior Software Engineer with 6+ years of experience building systems with a focus on data-intensive applications, distributed infrastructure, and AI-powered solutions. Experienced in designing and implementing LLM-driven features, including natural language processing and retrieval-based systems over complex business data. Strong focus on response accuracy, deterministic processing, and minimizing hallucinations in AI systems. Proven track record of delivering scalable backend systems (Django/FastAPI), optimizing large datasets, and building end-to-end pipelines from raw data ingestion to user-facing insights.

TECHNICAL SKILLS

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

AI / LLM: LLM integration, 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.

PROFESSIONAL EXPERIENCE

Senior Software Engineer

August 2024 – Present

Autput (Sibgold Team)

Remote, Part-time

The largest gold mining enterprise in Siberia, operating multiple distributed production sites with complex logistics and data flows (2,000+ employees).

- Designed and implemented a full-stack ERP system for gold inventory tracking and operational analytics (250+ users), replacing manual Excel workflows and reducing reporting time from 2 hours to 30 minutes.
- Optimized large-scale PostgreSQL data processing by rewriting critical queries, adding composite indexes, and introducing materialized views, reducing executive report generation time from 50 minutes to under 2 minutes.
- Built a distributed data synchronization layer (Celery, REST) for geographically isolated production sites with unstable connectivity, ensuring reliable and idempotent data delivery across all nodes.
- Designed a multi-agent LLM analytics system (LangGraph, 3 agents) that converts natural language requests into fully rendered HTML dashboards. The system includes a query understanding agent (maps intent to a structured dashboard spec), a data retrieval agent (generates and validates SQL), and a dashboard builder agent (renders HTML/JS output), enabling non-technical managers to access insights without analyst support.
- Implemented a query understanding pipeline that enriches prompts with RAG-retrieved domain context (business rules, metric definitions) and uses schema-constrained outputs (Pydantic) to extract entities, time ranges, chart types, and KPIs into a validated JSON format, ensuring deterministic downstream execution.
- Built a retrieval-augmented context layer by chunking internal documentation, business rules, and DB schema, generating embeddings, and storing them in ChromaDB. At query time, the system retrieves top relevant chunks via semantic search to ground LLM responses and reduce hallucinations in metric definitions and schema usage.
- Developed an automated pipeline that parses Django migration files to extract and normalize database schema metadata into structured JSON, keeping LLM context up-to-date and improving accuracy of text-to-SQL generation.
- Implemented a safe text-to-SQL pipeline with multiple validation layers: LLM-based checks for schema consistency, static SQL validation, and EXPLAIN dry-runs before execution. Combined with read-only access, table whitelisting, and retry logic, this reduced invalid queries and hallucination-driven errors to near zero in production.

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

Senior Software Engineer

May 2024 – Present

Art. Lebedev Studio LLC

Remote, Full-time

The #1 multidisciplinary design and technology company in Eastern Europe. The Studio is known for high-profile digital projects, including complex web applications and pioneering AI-driven products in design (e.g., Fokus and Nikolay Ironov).

- Designed core features for Fokus (AI-powered presentation builder with 100k+ users), including interactive canvas elements and Stripe-based monetization flows.
- Designed and built a distributed LLM-powered presentation generation pipeline that converts a topic, raw text, or URL into fully structured slide decks. Implemented preprocessing for web content (scraping + text extraction) and unified all inputs into a single generation flow.

- Implemented a plan generation stage (Celery) that produces a structured presentation outline with slide types and content descriptions. Injected available slide templates directly into the prompt and enforced strict Pydantic validation with automatic retries, ensuring deterministic and valid outputs.
- Built a parallel slide generation system using Celery groups and chords, where each slide is generated independently. Each task fills template fields via LLM and generates image prompts, which are resolved via external APIs. Isolated failures per slide with retry logic, preventing full pipeline breakdown.
- Designed the final aggregation stage (Celery chord callback) that assembles all slides into a validated JSON structure and persists it to the database, including integrity checks before writing.
- Designed and built an AI-powered code review pipeline that analyzes Git diffs and provides context-aware feedback directly in GitLab Merge Requests based on labels and change scope.
- Implemented a preprocessing layer that parses diffs, filters noise, and enriches context with related files and metadata, allowing the LLM to reason not only about isolated changes but about their impact on the surrounding codebase.
- Designed a prompt pipeline that adapts review depth dynamically (style, architecture, logic) based on MR labels and file types, ensuring relevant and targeted feedback instead of generic comments.
- Built a structured feedback system where LLM outputs are normalized into categorized comments (bugs, risks, improvements, style), enabling consistent formatting and easy consumption by developers.
- Integrated the system into GitLab CI/CD as an automated step, posting inline comments and summaries directly into Merge Requests, reducing review turnaround time by 67% and catching logic issues before human review.
- Led migration from Vue 2 (Nuxt 2) to Vue 3 (Nuxt 3), reducing technical debt and improving maintainability.
- Implemented automated testing (Playwright, Vitest) and monitoring (Sentry), increasing crash-free sessions from 89% to 98%.

Stack: Vue 2/3, Nuxt 2/3, TypeScript, Node.js, Python, Celery, OpenAI/Gemini, PostgreSQL, GitLab CI/CD.

Software Engineer

June 2023 – May 2024

Art. Lebedev Studio LLC

Remote, Full-time

- Developed key generative and front-end features for Nikolay Ironov, the pioneering AI identity generator for small businesses, scaling the product to over thousands of unique design projects.
- Led a major performance refactoring initiative, converting the frontend to leverage Server-Side Rendering (SSR) and eliminating outdated dependencies. This effort improved Google PageSpeed scores from 40 to 80 and drove a ~75% growth in organic search traffic.
- Introduced and enforced the Feature-Sliced Design (FSD) architectural pattern into a large legacy codebase, which significantly increased code clarity, modularity, and long-term maintainability across the engineering team.
- Designed and built the architecture for a fleet of highly automated, lead-generating microservices. This scalable system rapidly became the primary source of inbound leads for the product, requiring minimal developer intervention for deployment and A/B testing.

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

Web Developer

January 2021 – June 2023

Kwork

Remote, Freelance

One of the largest freelance marketplaces in Asia, serving thousands of clients and professionals. Focused on delivering high-performance, maintainable web solutions for diverse e-commerce and SaaS clients.

- Successfully delivered over 20 diverse Fullstack projects, specializing in responsive applications built with Vue.js and Nuxt, ensuring high standards for UX, accessibility, and architectural robustness.
- Performed extensive codebase refactoring and dependency pruning across various client projects, achieving a reduction in bundle size by up to 40% and consistently improving Google Lighthouse performance scores to 90+ on client websites.
- Engineered efficient data fetching strategies, integrating both REST and GraphQL APIs. Implemented aggressive client-side caching mechanisms (e.g., using Vuex/Pinia modules) which resulted in page load time improvements of up to 50%.

Stack: Vue.js, Nuxt.js, Vuex/Pinia, JavaScript/TypeScript, REST/GraphQL, Webpack Optimization, Figma, A/B Testing, PostgreSQL/MongoDB.

EDUCATION

Reshetnev Siberian State University of Science and Technology
Bachelor's Degree, Computer Science, 2025