Maksim Khovrov

AI / LLM Engineer

CONTACT

Novi Sad, Serbia · Remote / Hybrid / Open to Relocation

mkhovrov01@gmail.com

t.me/khovre

linkedin.com/in/maksim-khovrov-113633293

github.com/mkhovrov01

khovrov.dev

+381 65 810 6509

SKILLS

Languages
Python, JavaScript, TypeScript, SQL

AI & Retrieval
RAG, LangChain, OpenAI API, Qdrant, Zep, Prompt Engineering, MCP

Backend & Infra
FastAPI, Next.js, PostgreSQL, Docker, Caddy, Linux, Git

Automation & CRM
n8n, REST APIs, Webhooks, Bitrix24, amoCRM, 1C, Swagger

EDUCATION

MIREA — Russian Technological University, Moscow
Software Engineering (incomplete, 2023)

LANGUAGES

Russian: Native
English: C1, cert.efset.org/en/7MbYem

SUMMARY

AI engineer with 4.5 years in development and automation, the last 2+ focused on production LLM systems: RAG pipelines, conversational AI agents, and self-hosted AI infrastructure. Shipped 30+ voice AI solutions to production and built a governed retrieval system that now runs five self-hosted personal projects end to end.

Seeking AI Engineer / LLM Engineer roles: designing, deploying, and scaling production AI products, with a focus on agentic workflows, retrieval systems, and AI infrastructure.

EXPERIENCE

Product Manager / Automation Engineer

Mar 2023 – Nov 2025

MCN Telecom — Moscow / Remote

Responsibilities

- Managed the full implementation cycle of the internal voice AI assistant with RAG: requirements, system design, launch, support.
- Designed integration architectures with Bitrix24, amoCRM, 1C, and IP telephony.
- Built business logic in n8n with custom JavaScript and Python functions.
- Built RAG pipelines: dataset preparation, vector store configuration, retriever tuning.
- Wrote system prompts and scenarios to improve LLM response accuracy.
- Gathered requirements, decomposed tasks, prepared final documentation (Swagger API).
- Managed a team of 10: onboarding, mentoring, code review, internal knowledge base.

Key achievements

- Launched 30+ voice AI solutions into production across medical, retail, recruitment, financial, and municipal-infrastructure clients.
- Designed and shipped 3 RAG pipelines (Qdrant, Zep) for client knowledge bases, raising response accuracy from 70% to 98%.
- Migrated the voice pipeline from a step-based to a real-time streaming architecture.
- Led 15+ technical audits and supported commercial proposals that closed $45K+ in deals.
- Scaled the automation team from 3 to 10+ people.
- Built alerting and error handling for high-uptime external API integrations.

QA Engineer

Jul 2021 – Mar 2023

MCN Telecom — Moscow / Remote

Responsibilities

- Conducted manual testing, developed technical documentation and system guides.
- Built and maintained internal automation workflows, shifting from manual QA to backend automation engineering.

Key achievements

- Took ownership of the department's technical stack when the team was reduced to 3 members; became the primary developer for automated solutions.
- Promoted to Product Manager / Automation Engineer.

PERSONAL PROJECTS

Self-hosted and self-built. Portfolio and live demos: khovrov.dev

Cogan. System for preparing and checking complex document packages, from normative requirements to finished PDF and DOCX; runs in production for a client. Includes a product agent with its own evaluation harness and sandbox; OCR/PDF reconstruction runs on a separate untrusted-compute worker. FastAPI, React, TypeScript, SQLite, LangGraph, Docling, Typst. Live demo: editor.khovrov.dev.

KB Auditor and RAG Builder. Agent-orchestrated pipeline that crawls a public knowledge base, audits it, builds a RAG index, and evaluates retrieval quality end to end; n8n runs the agentic stages. Next.js, PostgreSQL, Prisma, n8n.

PhonoForge. Speech-ML engine that generates spoken vocabularies and recognizes them without retraining: a pretrained wav2vec 2.0 phone model with CTC ranking over the active word list, and per-phone margins that show the weakest sound. Includes a local browser lab. Python, PyTorch, Transformers, CMUdict, FastAPI.