

Olzhas Koishybek

AI / LLM Engineer · Python

Almaty, Kazakhstan | +7 777 629 16 38 | koishybek.o@gmail.com | Telegram @mrolzz | github.com/koishybek

PROFILE

AI engineer with 2+ years of production experience shipping LLM features end-to-end: data pipelines and RAG, model integration, backend, and user-facing apps. At Ozen, built the model-integration layer of a live generation pipeline (6+ models, RAG, LLM-as-judge) and scaled its corpus 51x; at IoT-Exponenta, cut manual data QA by ~40% with anomaly detection over 2,000,000+ telemetry records. Winner of the Higgsfield Applications Contest (\$5,000). English C1 after two years of study and work in the U.S.

EXPERIENCE

AI Engineer

Feb 2026 – Aug 2026

Ozen — AI music-generation platform

Almaty, Kazakhstan

- Shipped a production generation pipeline integrating 6+ models: Claude (text/style), Gemini (LLM-judge quality control + fallback), Gemma, embeddings; audio analysis with Essentia, Demucs, pyannote; Suno for final rendering.
- Scaled the KZ/RU text corpus 51x (35 → 1,806 verified entries) with a fully automated pipeline: docx/TTML parsing, two-source cross-verification, 1024-dim embeddings, scheduled autonomous scraping.
- Co-built the backend (monolith + 2 microservices) powering single-shot and best-of-N generation, RAG retrieval over the corpus, and an engagement-driven feedback loop.
- Tuned generation quality with an automated LLM judge: ~450 end-to-end runs, ~220 re-evaluated before release.

Data & Automation Engineer (Python)

Oct 2024 – Mar 2026 (part-time from Feb 2026)

IoT-Exponenta — industrial IoT / smart metering

Almaty, Kazakhstan

- Cut manual data checks by ~40% and surfaced faulty devices before customer-facing failures with anomaly-detection models on streaming sensor data (~92% detection accuracy).
- Turned 2,000,000+ raw telemetry records from ~15,000 smart-metering devices into analysis-ready datasets with Python ETL pipelines.
- Reduced incident time-to-resolution by ~30% by redesigning data intake and triage: validation rules, automated flagging, prioritised queues.
- Saved ~10 h/week by automating recurring processing; coordinated hardware / firmware / data work in a team of 5.

Independent AI Developer

Aug 2026 – Present

Freelance — AI automation, chatbots and web apps for SMB clients in Kazakhstan

Almaty, Kazakhstan

- Deliver paid products from brief to production — requirements, build, deployment, handover; clients come through referrals.

SELECTED PROJECTS

Metrix Montazh | *TypeScript, React, offline-first PWA* | github.com/koishybek/metrix-montazh

2026

- Paid client delivery, used daily by installer crews: work-completion acts are filed from a phone with no connectivity and sync to the server once online.

Smart Metrix MCP Server | *Python, MCP* | github.com/koishybek/iot-mcp

2026

- Exposes an IoT smart-metering platform to AI agents so Claude can query device status and telemetry in natural language.

Omnify | *Python, LLM agents* | github.com/koishybek/Omnify

2026

- SaaS MVP: an AI sales agent for WhatsApp / Telegram / Instagram with a knowledge base, CRM and operator inbox for human hand-off.

SWAT | *TypeScript, Telegram Bot API, LLM parsing* | github.com/koishybek/swat

2026

- Job platform for J-1 Work & Travel students: scouts submit offers via a Telegram bot, an LLM parses them, admins approve before publishing.

MCP Server for Kompra Group | *Python, MCP* | Kompra × AIESEC case championship

2026

- Designed and prototyped an MCP server connecting Kompra's company-data services to AI assistants for due-diligence queries.

AWARDS & HACKATHONS

Winner, Higgsfield Applications Contest (\$5,000): Viral Trend Presets App — aggregator of viral video effects built on the Higgsfield API, 2026

Hackathons: 3rd of 120 teams — Halyk OIY; 3rd of 300 — MentorHub; 4th of 300 — Kazakhstan AgroHackathon

TECHNICAL SKILLS

LLM & AI: Claude / Gemini / Gemma API integration, RAG, embeddings, LLM-as-judge evaluation, MCP servers, agent design

Python & Data: pandas, NumPy, scikit-learn, ETL pipelines, anomaly detection, time-series, SQL

Web & Backend: TypeScript, React, REST APIs, Telegram Bot API, Git, Vercel; audio analysis with Essentia, Demucs, pyannote

EDUCATION & INTERNATIONAL EXPERIENCE

Kazakh-British Technical University (KBTU)

Expected 2028

BSc Computer Science, 3rd year — Junior ETL Developer certificate, TechOrda (2026)

Almaty, Kazakhstan

FLEX — U.S. Department of State exchange (2021); Work & Travel USA, J-1 (2025): two years in the U.S. — fully funded academic year at a U.S. high school; full-time work in an English-only environment

Languages: Kazakh — native · Russian — native · English — C1