Mikhail Kolatau

Senior AI Engineer (Full-Stack, TypeScript)

mishakolotov1@gmail.com · mikhailkolatau.eu
linkedin.com/in/mikhail-kolatau · github.com/MishaKolotov
Warsaw, Poland · EU work-authorized · remote · B2B via own entity · two weeks' notice

SUMMARY

- AI engineer who builds the product around the model: the call, evals, guardrails, queue and UI.
- Own LLM products: GigaWork, a job-search agent in production since June 2026, and GigaMixer, a mixing chat.
- Evaluation tasks for OpenAI's AI Lab, Aug–Oct 2026: rubrics of 60–97 sourced items; promoted to domain reviewer.
- Core author of the LLM traces and evaluation UI at Lumigo, across five model providers.
- Ten years of TypeScript: Angular at employers, React at Pynea, Node.js and Postgres in own products.

SKILLS

LLM applications. Claude API, Gemini, structured output (Zod, JSON Schema), tool / function calling, prompt engineering, prompt caching, model cascade, multi-provider fallback, cost tracking, speech-to-text (Whisper)

Agents. Claude Code (headless), multi-agent review pipelines, browser agents, tool allow-lists and least privilege, human-in-the-loop approval

Evals and safety. Rubric design, eval sets from production traces, regression evals, LLM-as-judge, MRR and sign test, prompt-injection defenses, guardrails, LLM observability UI (token cost, OpenTelemetry spans)

Backend. Node.js, Next.js, REST, PostgreSQL, SQL, Drizzle, Postgres job queues, Vercel. Familiar: Python (DSP scripts, GigaMixer worker), Spring Boot, AWS Amplify

Frontend. TypeScript, React 19 (hooks, Context API), Next.js 16 (SSR, Server Components), Angular (2 → 21), Redux, TanStack Query, RxJS, NgRx Signals, Nx, Module Federation, Tailwind 4, styled-components, shadcn/ui, GraphQL

Testing / CI/CD. Unit and integration tests (Vitest, Jest), React Testing Library, Playwright, Cypress, GitHub Actions

WORK EXPERIENCE

AI Evaluation Expert & Domain Reviewer · OpenAI AI Lab via Digis — Aug 2026 – Oct 2026 · Remote

Contract · finding where OpenAI's models fail on senior software tasks: research prompt, model run, rubric

- Designed evaluation tasks with binary rubrics of 60–97 items, every item backed by a source.
- Ran the model on every task to pin real gaps, each written up with the source that disproves the answer.
- Promoted to domain reviewer a week in; reviewed five submissions by other experts across eight rounds.
- Built the tooling in 9 days (93 commits): a rubric linter, a nine-gate delivery workflow, five Claude Code skills.

Senior Full-Stack Engineer · Digis for a German GRC client — Feb 2026 – Present · Remote

Contract · compliance suite of four Angular products · Angular 21, PrimeNG 21, Tailwind 4, NgRx Signals, Nx 22, React 19

- Delivered through an in-repo agent pipeline: implementer, reviewer, QA agents; model routing.
- Led frontend of the document-management product; owned the whistleblower platform as primary frontend author.
- Built the Spring Boot read side of an e-invoice gateway: multi-tenancy, e-mail sign-up, per-IP rate limiter.
- Led the Angular 21 upgrade of the shared design system (80 components) as its second-largest contributor.

Senior Software Engineer · Pynea — Feb 2026 – Apr 2026 · Remote

Fixed-scope contract, in parallel with Digis · HR-tech SaaS, London · Next.js 15, GraphQL / Apollo, Turborepo (17 packages)

- Led the Settings redesign (8+ screens) from Figma to production.
- Built RBAC for a multi-tenant SaaS: custom roles, seat types, owner-role protection.

Senior Software Developer · Lumigo — Feb 2025 – Nov 2025 · Remote

Observability for LLM apps, AWS Lambda and Kubernetes · company acquired · Angular 19, Nx, ag-Grid, GraphQL

- Built the AI/LLM observability UI as its core author: traces across OpenAI, Anthropic, Bedrock, Gemini and Vertex.
- Added token-cost columns, cost time series and OpenTelemetry span views to those traces.
- Introduced a feature flag and shipped the LLM-evaluation UI end to end behind it; sole author of its scoring module.
- Lead author of Logs Live Tail: real-time log streaming over AWS Amplify GraphQL subscriptions.
- Ran bi-weekly releases and live customer demos, mentored engineers.

Software Development Engineer · Kaseya (Pulseway) Aug 2021 – Feb 2025 · Remote, Warsaw

Remote IT management web app · AngularJS and Angular 12 → 18, NGXS, NgRx, Tailwind, Module Federation

- Migrated the app from Angular 12 to 18: build 272s → 99s, vulnerabilities 65 → 3.
- Built 62 new components in the first ten months; later shipped React modules styled with styled-components.
- Led the team, ran hiring interviews and code reviews, mentored a mid-level engineer.

Sr. Software Development Engineer · ICE Mortgage Technology Mar 2020 – Aug 2021 · Remote, Minsk

Large-scale mortgage software · Angular, React, Redux, GitLab CI/CD

- Designed the interface layer between Angular and React; cut cross-framework bugs by 70%.
- Cut page load from ~4.5s to 2.7s; built and optimized 25+ UI features and widgets.

Full Stack Developer · Strateg Group Aug 2016 – Mar 2020 · Minsk

Start-up product · Angular 2+, Symfony, Laravel

- Chose Angular 2+ as the primary framework; built backend features in Symfony and Laravel.

PROJECTS

GigaWork · AI agent that finds jobs, scores my fit and drafts applications · solo, built by directing Claude Code, in production for my own search since June 2026, 5,151 offline tests · Next.js, Vercel, Neon Postgres, Drizzle, Claude

- Eval on 80 labelled production traces: 47 of 60 sent-or-applied jobs pass vs 35 on a re-run of the old analyzer.
- Re-ran the set after a model change, found scores about 8 points lower and re-checked the live threshold.
- Measured cost first: $0.0020 per score with prompt caching vs $0.0047 via Batch API; batching dropped.
- Pipeline: 18 sources → free rule-based gates → LLM score with structured output → Telegram; 37,702 jobs collected.
- Routed model calls through a provider chain (Claude, Gemini, OpenRouter, Z.ai) with a circuit breaker in Postgres.
- Browser agent fills ATS forms with 13 allowed tools and stops before Submit; e-mails go out at most once per job.
- Tailored CVs come from one corpus of my career and ship only after an 11-agent panel's fact-checker approves.
- Prompt-injection defenses: job text in an escaped container tag; model HTML rendered under a tag allow-list, JS off.

GigaMixer · mixes a multitrack in the browser; a chat turns “the vocal is buried” into mix changes · solo, built by directing Claude Code, Oct 2026 · live at automix-delta.vercel.app · Next.js, Supabase Postgres, Python worker, Claude

- LLM chat built on Claude tool calling: one strict-schema tool; each proposed change is clamped server-side.
- Model cascade: Haiku plans first, Sonnet retries on failed validation; prompt caching, daily budget and per-IP limits.
- Wrote an eval suite of 150 labelled phrases with accuracy thresholds and injection cases, ahead of its first live run.
- Python worker runs Demucs and BS-RoFormer source separation locally; a blind A/B judge gates DSP changes.

bandcamp-daily · picks one fresh release a day from my Bandcamp collection for a Next.js site

- Audio similarity in the ranking: ONNX embeddings in Node, 1,280-dim vectors stored in Postgres, cosine ranked in Node.
- Tuned the blend on an offline eval: MRR 0.118 → 0.200; rank shifts 21 up and 6 down, exact sign test p=0.006.

llm-job-triage · public slice of GigaWork scoring: structured output, provider chain, eval runner · github.com/MishaKolotov/llm-job-triage

ctaio-jobfit · live at ctaio-jobfit.vercel.app · one Claude call reads the CV, then plain code ranks 406 roles

EDUCATION

Bachelor of Science (B.Sc.) in Quantum Electronics Engineering, Faculty of Radioelectronics, BSUIR, Minsk · 2018

LANGUAGES

English: full professional · Polish: elementary (A2) · Russian: native