Mikhail Kogan

AI Agentic & MCP Platform Engineer | Infrastructure & CI/CD

About me
AI Platform & Agentic Engineer with 10+ years building high-availability systems, from consumer apps (5M+ users) to AI infrastructure.
Creator & Lead Developer of CyberMem – a production-ready, self-hosted MCP memory server that lets multiple LLM clients (Claude, GPT, Cursor, Perplexity) share persistent context.
Expert in multi-platform deployment (Docker, Kubernetes, Helm, Ansible), CI/CD automation (GitHub Actions, GitLab CI, Xcode Cloud), and Zero-Trust networking (Tailscale, Traefik).
Proven leader: **scaled engineering team from 1→13, cut release cycles from 2 weeks→2 days, maintained 99.5–100% uptime in production.**

Experience

▼ CyberMem

Open Source Project: **Production-grade, self-hosted MCP** (Model Context Protocol) **memory server** solving AI agent context fragmentation. Enables **Claude, ChatGPT, Gemini, and Perplexity to share a single persistent memory** with automated **multi-platform deployment** (macOS, Linux, Raspberry Pi, Kubernetes), **real-time observability**, and Zero-Trust networking. cybermem.dev / GitHub / NPM

Creator & Lead Developer: 2025 — Present

- Designed and built a self-hosted MCP server that allows **multiple LLM clients** (Claude, GPT, Gemini, Cursor, Perplexity) to **share long-term memory and tools**, eliminating the “every conversation starts from scratch” problem.
- Implemented per-client remote access policies, **observability, and full audit logging** for every operation and **every MCP client (Claude Desktop, Claude Code, Cursor, etc.) separately.**
- Engineered a custom TypeScript CLI and IaC engine that **automates complex multi-platform deployment** (macOS, Raspberry Pi, Kubernetes clusters) **into a single command**.
- Built a full observability stack (Vector → Prometheus/Grafana) and a custom dashboard for **real-time tracking of AI agent memory per-client usage**, latency, and memory health.
- Established an **end-to-end testing framework with Playwright** to ensure **mission-critical stability across updates**.
- Validated with **complex LangGraph/LangChain multi-agent workflows** to ensure context consistency and prevent hallucination during long-running tasks
- Automated **CI/CD for the monorepo** using GitHub Actions, enabling **100% automated multi-platform releases (no manual steps).**

▼ Centerya

AI **CRM platform made by Israeli seed stage startup**, designed for real estate agents in the US market.

Tel Aviv, Israel

Founding Engineer & Mobile Infrastructure: 2024 — 2025

Centerya iOS App: https://apple.co/3TuHEJB. A mobile app for AI CRM with support of **VoIP calls** and **Siri Intents.**

- **Architected and solely delivered** an AI-powered iOS CRM application (Centerya) from concept to App Store launch **(VoIP, Siri Integrations, App Intents)**.
- Integrated Twilio SDK to provide **VoIP calls with AI-generated call summaries and property suggestions** for US real-estate agents.
- Established mobile infrastructure with **real-time monitoring** (Sentry)**, centralized logging (AWS CloudWatch),** and **alerting** to keep **crash-free sessions above 99%.**
- Built and maintained a **CI/CD pipeline** in Xcode Cloud, enabling **automated testing and weekly feature releases**.
- Implemented a modular SwiftUI + SwiftPM architecture **optimized for caching and startup performance**.

▼ Tinkoff Bank

The **largest digital-only bank in Western Europe** operated in the Russian market.

Russia

Team Lead — Dolyame iOS platform: 2021 — 2023

Dolyame — Russia’s first BNPL (Buy Now, Pay Later) service. Comparable to Klarna (Europe) and Afterpay (US).

- Launched the BNPL iOS app as a solo developer → scaled it to **1M+ downloads and 300K MAU in the first year**.
- **Grew the iOS platform team from 1→13 engineers**, establishing CI/CD standards and code review processes to **support rapid product growth**.
- Architected a UI testing workflow that allowed QA to automate regression tests, **reducing manual testing time by 70%.**
- Transitioned the team from Git Flow to trunk-based development with feature flags, cutting **release preparation time from 4 days → 1 day.**
- Developed an LLM-assisted data extraction pipeline to **automate team burnout analysis and capacity planning,** while **maintaining 99.5–100% crash-free rates for financial flows**.

iOS Developer — Traffic Fines: 2018 — 2021

Traffic Fines App — a platform enabling users to conveniently pay their fines.

- Refactored **critical** payment & fine lookup **modules in a legacy app**, ensuring 99.7% crash-free sessions for **5M+ users**.
- **Established robust testing infrastructure** (Unit, UI, Snapshot), achieving high coverage for financial transaction flows.

▼ GLOBUS Ltd.

Nizhniy Novgorod, Russia

Junior iOS Developer: 2016 — 2017

- Developed and **shipped multiple iOS applications to the App Store** as a junior developer.

Education

BSc in Computer Science & IT, Nizhniy Novgorod State University: 2013 — 2017

Details

cybermem.dev

Tel Aviv, Israel

mikhailkogan17@gmail.com

linkedin.com/in/mikhail-kogan-platform

github.com/mikhailkogan17

Skills

AI & LLMs: Agentic systems, LangChain, LangGraph, MCP, RAG, embeddings, tooling, OpenAI Whisper, Azure Speech.

AI Infra & DevOps: Docker, Kubernetes, Helm, Ansible, Traefik, IaC, Tailscale, AWS, Zero-Trust networking.

CI/CD & Observability: GitHub Actions, GitLab CI, Xcode Cloud, Fastlane, Prometheus, Grafana, Sentry, Firebase, Vector, Amplitude.

Languages & Mobile: TypeScript, Bash, Swift, SwiftUI, Python, modular architecture (DDD, feature flags, TBD flow), testing (Unit, Snapshot, E2E with Playwright/Pytest), SDK distribution (XCFramework).

Personal Projects

Hearly — Built real-time speech pipeline combining Azure Speech STT with Whisper for higher-quality transcription for 2+ languages simultaneously (SwiftUI + iOS). Available in hearly.app.

SoundPulseButton — open-source SwiftUI button component with haptic and audio feedback, featuring pulse and ripple animations.
Listed on Swift Package Index / GitHub; also used in Hearly.

Languages

Russian (native), English (fluent)