

Alexander Myasoedov

Agentic AI engineer - LLM agents, RAG, evals and guardrails

Remote, Europe · available for contract and fractional engagements

[x.com/arekusandr_](#) · [linkedin.com/in/miasoedov/](#) · [github.com/msoedov](#) · [www.metaheuristic.co/hire/cv/](#)

4.6k	57	395	2012
GITHUB STARS	PUBLIC REPOS	GITHUB FOLLOWERS	OPEN SOURCE SINCE

SUMMARY

I build systems where a language model plans, calls tools, and acts - and I make those systems measurable, safe and cheap enough to point at production traffic.

The work is closer to distributed systems engineering than to modelling: control flow with real budgets, tool contracts that survive an unreliable caller, retrieval that respects your permission model, an eval harness that turns prompt changes into diffs with numbers attached, and traces that make a 3am diagnosis possible.

I also maintain the open-source tooling other teams use to find these failures - including an LLM red-teaming kit with around two thousand stars - and I write about the operational side of this work in public, weekly.

Available for **contract, fractional and embedded engagements**. Code, prompts, evals and IP are yours from day one.

STACK

Agentic systems

Agent control flow

Tool contracts & MCP

Loop guards & budgets

Human-in-the-loop approvals

Durable execution (Temporal)

LangGraph

Multi-agent orchestration

Model layer

Eval harnesses & golden sets

LLM-as-judge calibration

Prompt versioning

Model routing & fallback

Prompt caching

Cost & latency budgeting

Fine-tuning decisions

vLLM / open-weight serving

Retrieval

Permission-aware RAG

Hybrid search (BM25 + dense)

Reranking

Layout-aware PDF parsing

Structure-aware chunking

pgvector

Freshness & incremental indexing

Security & governance

LLM red-teaming

Prompt injection defence

Tool-scope least privilege

Output validation

PII handling

Audit logging

On-premise & sovereign deploy

Engineering

Python

Go

Elixir

Haskell

FastAPI

PostgreSQL

Kafka

Docker

Kubernetes

OpenTelemetry

AWS / GCP

Founder & principal engineer · Metaheuristic

Present

- Applied AI partner. Takes LLM prototypes into production for teams that need the system to survive real users, real documents and a real bill.
- Ships agentic systems end to end: control flow, tool integrations, retrieval, evals, guardrails, traces and handover.
 - Cut inference spend on one production system from roughly \$4,120 to \$1,180 a month through caching and model routing, at the same quality and latency.
 - Runs fixed-scope Discovery Sprints that take teams from demo to production in two to ten weeks, with evals and a security review gating every release.
 - Delivers on-premise and sovereign deployments for teams that cannot send data to a third-party API.
 - Client keeps 100% of code, prompts, evals and IP - no runtime licence, no platform lock-in.

Open-source author & maintainer · github.com/msoedov

2012 - present

- 57 public repositories, roughly 4,600 stars, across LLM security, LLMOps, retrieval and developer tooling.
- **agentic_security** (~2.0k stars) - agentic LLM vulnerability scanner and red-teaming kit, used by other teams to test their own agents for jailbreaks, prompt injection and data leakage.
 - **langcorn** (938 stars) - serves LangChain LLM apps and agents through FastAPI; the LLMOps layer between a notebook and a callable API.
 - **validex** (144 stars) - structured extraction from unstructured sources, the unglamorous half of every RAG project.
 - **justshowmediff** (69 stars) - browser-based review of AI-generated diffs, built for Claude Code, Codex and headless agent workflows.
 - **vector_lake** (59 stars) - S3-backed vector database for LLM agents and RAG, for teams who want retrieval without a managed vendor in the critical path.

Writing · metaheuristic.co/blog

Ongoing

Essays on the operational side of production AI - agent control flow, evals, review markets, retrieval, and keeping a human in the loop without slowing the loop down.

OPEN SOURCE

agentic_security ★ 2.0k Agentic LLM vulnerability scanner / AI red-teaming kit.	langcorn ★ 938 Serving LangChain LLM apps and agents via FastAPI. LLMOps.
hacker-slides ★ 347 Markdown to presentation slides, in Go.	validex ★ 144 Structured extraction from unstructured sources.
justshowmediff ★ 69 Review AI-generated diffs in the browser.	vector_lake ★ 59 S3 vector database for LLM agents and RAG.

SELECTED WRITING

The 100x AI Engineer Is a Myth. The Real Number Is 2x.	Sep 2026	The MCP Battle	Aug 2026
Building an Agent for the Trade That Blew Up	Aug 2026	Harness as a Service	Aug 2026
RAG Beyond the Tutorial	Aug 2026	Skip the Agentic AI course, build the harness	Aug 2026

CONTACT

x.com/arekusandr_ → linkedin.com/in/miasoedov/ → github.com/msoedov → [Contact form](#) →