

From AI demo to reliable production

Fixed prices, read-only first, no long contracts. For small product teams, and for agencies that build for clients (you keep the client relationship).

AI in production

Service	What you get	Time	Price (USD)
AI Deploy Fix	One specific problem with an LLM, RAG or model service: slow or failing answers, timeouts and rate limits, GPU or container issues, a deployment that won't go live. Root cause, fix, short note.	1–3 days	\$250–400 fixed
AI Production Readiness Review	Review of one AI feature: tracing, evaluation and release gates, latency, reliability and fallbacks, cost per answer, privacy and prompt injection, retrieval pipeline. Scorecard, measured findings, 30-day plan, walkthrough.	3 days	\$900 fixed
LLM Cost Check	Where your LLM spend goes and how to cut it: context size, model per step, caching, provider settings, and a self-hosting (vLLM / open-weight models) break-even estimate.	2–3 days	\$750 fixed
AI Ops Care	Monthly: dashboard and cost review, evaluation runs before releases, provider and model updates, small fixes, short monthly report.	Monthly	\$600–1,200 / month

DevOps & infrastructure

Service	What you get	Time	Price (USD)
Deployment Fix	A broken pipeline, 502s after deploys, SSL, Docker or NGINX issues, a server you are afraid to touch. Root cause, fix, short note.	1–2 days	\$150–300 fixed
DevOps Health Check	One application and environment: CI/CD, backups and restore test, monitoring, server security and access, secrets, cost. Prioritised findings and a 30-day plan.	3 days	\$750 fixed
DevOps Maturity Assessment	Whole engineering platform: delivery, infrastructure as code, observability, security, backup and DR, incident process. Maturity score per area, risk register, 90-day roadmap.	5–10 days	from \$1,750
DevOps Care	Monitoring and alert triage, patching, monthly restore test, small fixes, monthly report.	4–8 h / month	\$400–800 / month

How I work

- **Read-only first.** Reviews never change your systems. Fixes happen only after written approval.
- **Evidence, not opinions.** Every finding says what was measured, so your team can check it.
- **Your data stays yours.** Personal data is masked in anything I store; access is removed at the end, and I confirm when it is.
- **What I need:** read-only access (cloud viewer, logs, repository), one contact who answers within a working day, and your client's permission if the system is theirs.
- **Not included:** penetration testing, compliance certification, 24/7 on-call.

Payment

- **Fixes and reviews:** 50% to book the slot, 50% on delivery. First-time fixes under \$250: 100% on delivery.
- **Care plans:** invoiced monthly in advance; cancel any time with 30 days' notice.
- Invoices in USD (EUR or GBP on request), bank transfer within 14 days. Quotes are valid for 30 days.

Typical first step: send the problem (an error, a trace, a log, or a short description). Within 24 hours you get a fixed price and how I would approach it. Nothing is billed until you approve the quote.