

Saad Aamir

Software Engineer · AI & LLM Systems

saadamir473@gmail.com · +92 318 6492353 · Islamabad, Pakistan (open to relocation / remote)
saad-portfolio-six-pi.vercel.app · github.com/saad-aamir · linkedin.com/in/saadaamir · beginnersmindbysaad.substack.com

PROFILE

Software engineer with 3 years of production experience across Python and TypeScript, now focused on building LLM systems: MCP servers, retrieval pipelines, structured outputs, and the boring-but-critical plumbing that makes model integrations reliable. I run an independent studio where I ship client products end to end and use my own AI tooling daily. Currently working toward empirical AI safety research, starting with model evaluations.

EXPERIENCE

Founder & Software Engineer, Dark Matter Studio (independent) Jan 2026 - present · Remote

- Run a small web studio end to end: outreach, scoping, design, development, deployment, and support. Shipped production sites for clients including Sussex Light Photography, The Crib Murree (hospitality booking), and FMY Chartered Accountants, all on Next.js, TypeScript, Tailwind, and Vercel.

Dark Matter Co-Pilot OPEN SOURCE · FLAGSHIP

- A Python MCP server that connects Claude to the studio's real operations data: leads, projects, case studies, and pricing docs. Ten typed tools over SQLite, used daily for CRM work and outreach drafting, so it is tested against real workflows rather than demos.
- Built around Pydantic v2 as a single source of truth: the same models validate database rows, define tool schemas, and serialize responses. A Base/Create/Read model split keeps database-assigned fields out of tool inputs.
- Diagnosed a silent failure where tools worked in the MCP Inspector but failed in Claude Desktop, tracing it to SQLite's naive timestamps violating strict RFC 3339 validation, then fixed it with timezone-pinning serializers on every datetime field.

Software Engineer (Full Stack), MobileLIVE, Canadian technology consultancy Jul 2023 - Dec 2025 · Remote

- Refactored the RAB and RDA Lighting internal portals, upgrading from .NET 6 to .NET 8 and reworking a poorly-structured legacy codebase: introduced the repository pattern, inversion of control, and dependency injection to improve maintainability and testability.
- Integrated LLM-backed features into the Position Requisition Form portal and rebuilt its key user flows, cutting manual workflow steps by about half.
- Built React dashboards backed by Node.js/TypeScript REST services on the Geotab Fleet Analytics project; query optimization work reduced latency by roughly 40%.
- Implemented JWT authentication and secure API endpoints; deployed and maintained AWS infrastructure (EC2, Lambda, S3, CloudWatch) with CI/CD through GitHub Actions.

Web Development Intern, ZSystems, Lahore Jul - Sep 2020

- Built responsive frontends and Node.js/Express services in agile sprints alongside senior engineers.

SELECTED PROJECTS

Sycophancy Evaluation Harness 2026

Python · Inspect (AIS) · Anthropic API · Ollama · Pydantic

- Built an empirical eval measuring how often LLMs reverse correct answers under user pushback with no new information, comparing two open-weight models (Llama 3.1 8B, Qwen 2.5 7B) over 1,800 conversations each.
- Designed hybrid scoring: deterministic regex verification plus an LLM judge (Claude Sonnet 4.5), hand-validated against 93 manually-labeled samples to measure judge reliability before trusting it.
- Reported question-level bootstrap confidence intervals and conditional rates; collapsed to a binary taxonomy the validation supported after finding fine-grained judge categories unreliable.
- Found the two models destabilize under opposite pushback conditions and ruled out a capability confound to show the effect was dispositional. Code, research log, and writeup published.

AI Resume Matcher 2023

Python · FastAPI · LangChain · FAISS · Pydantic

- Parses resumes into structured sections, scores them against job descriptions using embedding similarity, and generates targeted improvement feedback with structured LLM outputs.

WRITING

I write Beginner's Mind, a technical blog explaining systems and math concepts with interactive demos. Recent posts: Llama folds when you sound vague (sycophancy eval methodology), Tools should return data, language models should return language (MCP architecture, drawn from building the Co-Pilot), the birthday paradox, and Axelrod's Prisoner's Dilemma tournaments.

SKILLS

AI / LLM	MCP (server design), Claude & OpenAI APIs, evals (Inspect), LLM-as-judge validation, RAG, LangChain, FAISS, Pydantic structured outputs, prompt design
Languages	Python, TypeScript, JavaScript, SQL, C++
Backend	Node.js, Express, REST API design, PostgreSQL, MySQL, MongoDB, SQLite, JWT auth
Frontend	React, Next.js (SSR/SSG), Tailwind CSS
Cloud / DevOps	AWS (EC2, Lambda, S3, CloudWatch), Docker, GitHub Actions, CloudFormation, Vercel

EDUCATION & CERTIFICATIONS

B.S. Computer Science, National University of Sciences and Technology (NUST), Islamabad 2019 - 2023

AWS Certified Cloud Practitioner