Arsh Ghotra

AI and Data Solutions Engineer · Overland Park, Kansas

arsh9745774@gmail.com · linkedin.com/in/aghotra07 · github.com/idk-arsh
portfolio-omega-steel-20.vercel.app

I build LLM workflow systems and Databricks pipelines at Sara Software Systems, and open source guardrails that stop AI agents from guessing at data. I keep the model's job and plain code's job separate, and measure before calling it done.

Experience

AI and Data Solutions Engineer · Sara Software Systems, Olathe, Kansas (remote) — Jul 2025 to present

- Built a document parser for proposal and RFI work that turns 40 to 80 page documents into structured task data in about 12 seconds, down from hours by hand. About 35 regex patterns find the structure; parallel LLM passes over 6-page windows read it.
- Built retrieval on top of it: BGE-large embeddings in a local FAISS index match a new opportunity to past projects in 20 to 30 seconds instead of hours, and historical content never leaves the local environment.
- Partner manager for the Databricks partnership. On two commercial engagements: pipelines on Lakeflow Declarative Pipelines and Unity Catalog, and a move of legacy IBM DataStage ETL onto the Lakehouse.

Advanced Technology Assistant · Indiana State University, Terre Haute, Indiana — May 2024 to May 2025

- Automated cleaning and normalization for over 10,000 datasets, and reporting in Python, SQL and Power BI that cut manual reporting by 60%.
- Designed real-time tracking for the university website, saving about 25 minutes of searching per user.

IT Intern · Glas-Col, LLC, Terre Haute, Indiana — Oct 2024 to Dec 2024

- Automated SQL-based reports, cutting preparation time by 25%, and built Power BI dashboards for management.
- Consolidated several data sources into one database, improving retrieval efficiency by 35%.

Open source

data-agent-rules	Rules, skills and a destructive-SQL seatbelt for AI agents that work with data. In a 162-run eval, Claude Haiku 4.5 handled 25 of 27 risky tasks safely with them, 10 of 27 without.
schema-guard	Stops agent SQL that names missing tables or columns. 0 false blocks on 1,034 Spider queries.
show-your-sql	Checks each number an agent reports against its query results. Wrong answers: 3 of 15 runs to 0.
data-mcp-bench	70 attacks that check whether a database MCP server is really read-only.
Chesterton	Claude Code plugin that uses git history to stop agents undoing past bug fixes. 358 tests.
MultiTerm	Multi-pane terminal workspace for Windows (Python, Tk, ConPTY). In winget since Oct 2026.

Education

BS Computer Science, Data Science concentration.
Indiana State University, May 2025. GPA 4.0, Dean's List.

Publication

Ghotra, A. S., Kour, H., Hasan, A., Khan, A. "Dog Breed Identification." *IRJET* 10(5), 1169 to 1174, May 2023.

Certifications

Databricks Certified Data Analyst Associate, Sep 2026

Databricks partner training: Data Intelligence Platform, Data and AI Governance with Unity Catalog, Partner Tech Summit 2026

Skills

AI and LLMs: Claude API, OpenRouter, MCP servers, Claude Code hooks and plugins, retrieval (BGE embeddings, FAISS), evals

Data: Python (pandas, NumPy), SQL, Databricks (Unity Catalog, Lakeflow), DuckDB, dbt, PostgreSQL, MySQL, SQL Server, Power BI

Engineering: Git, GitHub Actions, pytest, TypeScript, Linux, PowerShell