Farrukh Nauman

Applied AI / Agentic Systems Consultant | Coding Agents · Text-to-SQL · Enterprise Data Platforms | PhD

farrukh.nauman@inertialrange.com | (+46) 0702984959 | fnauman.com | inertialrange.com
Swedish citizen | LinkedIn: fnauman | Github: fnauman

Summary

I build reliable AI systems at the boundary of LLM agents, enterprise data platforms, and production automation—where chat-only demos break and real data, real tools, and real failure modes take over. I lead and ship: agentic workflows, text-to-SQL, large code migrations driven by coding agents, and validated SQL/Snowflake automation. Selected results:

- **Agent-driven migration validator**: cross-platform validation system (Databricks → Snowflake) from zero to live dashboard in ∼1 week using autonomous coding agents wired to Databricks/Snowflake CLIs.
- **5–10× faster, 50%+ cheaper**: SQL pipeline redesigned with row-level correctness validation; killed a faster but broken alternative via `EXCEPT`-based oracle checks.
- **Open-source agent framework**: `ts-agents`—autonomous time-series agents with strict JSON tool contracts, run manifests, and sandboxed execution.
- **Interim Team Lead**: 4–7 person DS & advanced analytics team; roadmap, stakeholder steering, and AI direction across telemetry analytics and GenAI automation.

Current Focus

- **Agentic workflows**: tool-contract design, sandboxed execution, evals, run manifests, and inspectable artifacts—agents that produce results, not just chat.
- **Text-to-SQL & natural-language data access**: schema/business-context grounding, query validation, eval sets, failure taxonomies, governance.
- **Coding agents for large code changes**: evidence-preservation methodology, behavioral oracles, and `AGENTS.md` quality contracts for migrations and refactors.
- **Enterprise data-platform automation for AI readiness**: Snowflake/Databricks pipelines, validation harnesses, and CLI-orchestrated workflows.

Skills & Tech Stack

AI & Agents	AI Coding Agents (Claude Code, Codex CLI, Cortex Code), Agentic Workflow Architecture, Tool-Contract Design, LLM Agents (LangChain, OpenAI SDK), Text-to-SQL, RAG, LLM Evaluation & Failure Analysis, Fine-Tuning Readiness (LoRA/PEFT), Synthetic Data, Sandboxed Execution
Data Platforms	Snowflake, Azure Databricks, Spark, Snowpark, Azure Data Factory, SQL (advanced), CLI-Orchestrated Pipelines, Cross-Platform Validation
ML & Analytics	PyTorch, Time Series, Predictive Modeling, Anomaly Detection, Computer Vision, Transformers, Weights & Biases
Engineering	Python (Expert, 8+ yrs), C/C++ (Proficient), Git, Docker, CI/CD, Streamlit, REST APIs, Structured JSON APIs, High Performance Computing
Delivery & Leadership	Interim Team Lead, Roadmap Ownership, Stakeholder Steering, Project Scoping, Solution Architecture, ROI Analysis
Languages	English (Fluent), Swedish (SFI C2), Urdu (Native)

Experience

InertialRange Labs AB - Self-employed — Linköping, Sweden
Principal AI Consultant — *Aug 2025 - Present*

Engagement: Interim Team Lead – Data Science & Advanced Analytics (Sep 2025 - Jun 2026): Enterprise Data & AI Platform Engagement

- Lead a 4–7 person DS/analytics team; own roadmap, stakeholder steering, and technical direction across telemetry analytics, ML, and GenAI automation.
- Built a **cross-platform migration validation system** using AI coding agents: automated comparison of schemas, row counts, key distributions, and date ranges across dozens of tables—from zero to live dashboard in ∼1 week. (Case study)
- Connected **Databricks and Snowflake CLIs** through autonomous coding agents (Codex CLI / Cortex Code CLI) to orchestrate migration, validation, and reconciliation pipelines end-to-end—shrinking multi-day manual runbooks into hands-off workflows.

- **Agent-accelerated code refactor**: redesigned a critical daily SQL pipeline from full recompute to **validated incremental processing** (5–10× speedup, projected **50%+** compute cost reduction). Killed a faster but fundamentally broken optimization using multi-scale benchmarks and `EXCEPT`-based validation. (Case study)
- **Tech**: AI Coding Agents, LLM Agents (LangChain, OpenAI SDK), Snowflake, Databricks, Snowpark, Streamlit.

Independent projects (separate from client engagement):

- Open-sourced **ts-agents**: a CLI toolkit for long-running autonomous agentic workflows on time-series data—machine-readable tool contracts, strict JSON envelopes, typed exit codes, run manifests, and sandboxed execution (local/Docker/Daytona/Modal). Applied to **machine activity recognition** as an agent-driven alternative to classifier-heavy pipelines.
- Built a **Text-to-SQL pipeline against a production ERP database** (hundreds of tables, real data, not a tutorial schema): schema-aware retrieval to stay inside the context budget, a semantic layer mapping business vocabulary (e.g. "biggest buyers", "SKUs that stopped selling") to columns, metrics, and join paths, and result-level evaluation—models pick columns by name, so SQL that runs cleanly against an empty/null field still returns zeros and gets logged as "success".
- Authored a practitioner framework for **large code migrations with coding agents**: evidence-preservation methodology, behavioral oracles, and `AGENTS.md`-based quality contracts. (Case study)

RISE Research Institutes of Sweden AB — Linköping, Sweden
AI Researcher & Consultant — *Jul 2021 - Aug 2025*

Project Lead: Sustainable Fashion AI Automation (2022-2025: *24 months*): Led two major initiatives: Vinnova: AI for Circular Fashion (**Project Lead**, ∼9M SEK) and CISUTAC (AI Lead, ∼2M SEK).

- **Challenge**: Manual quality inspection bottleneck in circular fashion supply chain—30% inconsistency, 25% cost overhead.
- **Solution**: End-to-end computer vision system for automated attribute detection; synthetic data pipeline to reduce annotation cost.
- **Impact**: 40% reduction in processing time, 50%+ reduction in data collection costs through synthetic data.
- **Recognition**: 1 of 5 projects featured at the EU Sustainable AI event (2023).
- **Tech**: PyTorch, Vision Transformers, CLIP, Gradio, Docker, Synthetic Data Generation.

Low-Energy IoT Solutions (2022: *4 months*):

- Identified miniROCKET-based time-series methods for industrial edge deployment, enabling real-time analysis at ∼90% lower hardware cost than heavier alternatives.

Capability Building & Mentorship:

- Established AI mentorship program for Master's thesis students; supervised projects spanning drone edge AI, air-traffic automation, traffic analysis, and anomaly detection.

Previous Research Positions — 2009–2019

- **Research Fellow, Chalmers University of Technology**: Complex systems modeling, large-scale data analysis — Gothenburg, Sweden, 2018–2019
- **Research Scientist, Niels Bohr Institute**: Simulation, forecasting, computational modeling — Copenhagen, Denmark, 2015–2018
- **Research Assistant/PhD Student, Univ. of Rochester**: Data analysis, predictive modeling — New York, USA, 2009–2015

Education & Certifications

Microsoft Certified — Azure
Azure Data Engineer Certificate — *2020*

University of Rochester — Rochester, New York (USA)
PhD in Physics and Astronomy — *Oct 2015*
Focus: Complex Systems Modeling, Data Analysis, Computational Fluid Dynamics, High Performance Computing, C/C++

Awards & Achievements

- Horton fellowship from Laboratory for Laser Energetics - full research funding award. 2010-2015
- Susumu Okubo Prize for highest performance on graduate comprehensive exam and excellence in coursework. 2011