Utsa Poddar

Data & AI Engineer

Calgary, AB | utsapoddarjobs@gmail.com | linkedin.com/in/utsa-poddar | github.com/utsapoddar | utsapoddar.github.io

Summary

Data and AI engineer with 5 years of professional experience building LLM and retrieval systems, agent tooling, and scheduled data pipelines. Keeps deterministic logic authoritative over model output and measures retrieval quality with an automated evaluation. MS Data Science candidate at the University of Colorado Boulder; BSc in Physics from the University of Calgary.

Skills

Languages: Python, SQL, R, JavaScript, TypeScript, Dart, MATLAB

LLM and AI Systems: LLM integration, retrieval-augmented generation (BM25, semantic embeddings, reciprocal-rank fusion), prompt and tool design, deterministic validation, multi-provider fallback, Anthropic API, Gemini, Groq, NVIDIA NIM, Ollama, Model Context Protocol, agent skills

Data and ML: pandas, NumPy, SciPy, PyMC, scikit-learn, Jupyter, Bayesian inference, classification, regression, hypothesis testing, EDA, feature engineering, uncertainty estimation

Data Engineering and Automation: ETL and pipeline design, data validation and reconciliation, REST API ingestion, SQLite, SQLite FTS5, SQLAlchemy, Supabase, FastAPI, Streamlit, GitHub Actions, CI/CD, cron, n8n, Playwright, Git

Analysis and Reporting: Excel, Google Sheets, KPI reporting, data cleaning, visualization, stakeholder communication

Leadership and Collaboration: Team leadership, international collaboration, coaching and mentoring, customer service, process improvement, root-cause analysis

Work History

Data & AI Engineer - Self-employed
Sep 2025 - Present | Remote

- Built and operate Alpha Digest, a five-stage weekly financial-data pipeline for a 25-entity watchlist that publishes email and web editions through GitHub Actions.
- Rebuilt the failing summarization stage after 4 consecutive missed weekly issues, separating retryable errors (timeouts, 429, 5xx, bad JSON) from permanent ones (retired model) across a 3-model fallback chain, so a provider outage degrades the run instead of ending it; every scheduled run since August 17 has succeeded.
- Automated the weekly research workflow behind these reports, removing an estimated 3 hours of manual collection, cross-referencing, and formatting each week (about 150 hours a year).
- Cut memory-context token usage 97% per session for AI coding agents with Engram: a 1.6 KB startup summary replaces about 53 KB of notes, an estimated 12,800 tokens per session and 1.3M per week. Startup cost stays flat as the notes grow.
- Built DopaKernel, a completion controller that keeps AI agents aligned to frozen, verifiable requirements; verification covers 90 unit tests plus 54 structural checks, all passing.

Online Department Supervisor - Canadian Tire Corporation
Jul 2024 - Sep 2025 | Calgary, AB

- Supervised 11 associates across online-order fulfilment, quality assurance, and daily operations, planning team schedules up to a month ahead to smooth coverage across peaks.
- Raised order fulfilment from 90% to a sustained 98% by tracking bottlenecks through metric review and implementing targeted process changes.
- Cut repeat complaints from 21% to under 2% and inventory discrepancies by 20% through root-cause analysis, standardized routines, and escalation handling.
- Created an operations binder of SOPs, checklists, and quick-reference guides to make onboarding and shift handovers more consistent.
- Implemented in-store advertising and placement changes for online orders, contributing to an ~11% lift in sales, and used a sales and seasonality dashboard to align online-orders inventory with seasonal demand.

Customer Service Representative - Canadian Tire Corporation
Apr 2023 - Jul 2024 | Calgary, AB

- Promoted to Online Department Supervisor after 15 months as a Customer Service Representative.

Business Intelligence Analyst - Next Generation Graphics Ltd (NGGL)
May 2022 - Jun 2023 | Remote (Singapore)

- Improved data accuracy and reporting efficiency by 25% by rebuilding recurring KPI reports and automating repetitive extraction steps.
- Wrote and optimized SQL queries across relational databases for recurring and ad hoc extraction, cleaning, and transformation.
- Detected data-quality issues early through exploratory analysis and anomaly detection in Python, MATLAB, Jupyter, and Excel.
- Presented findings to non-technical stakeholders on international teams across multiple time zones.

Projects

Engram - Auditable Memory for AI Coding Agents
Python, Markdown, SQLite FTS5, BM25, semantic retrieval, reciprocal-rank fusion | github.com/utsapoddar/engram

- Stores knowledge as Markdown notes (source of record) under a typed truth model (confirmed, inferred, conflicted, superseded), with a rebuildable SQLite FTS5 index so a corrupt index never loses knowledge; fuses BM25 with optional local embeddings by reciprocal-rank fusion.
- Adds a replacement-first correction journal with crash recovery, credential rejection on every write, and a seeded evaluation that requires recall@5 of at least 90% across 20 queries; 71 tests.

DopaKernel - Evidence-Gated Completion Control for AI Agents
Python, agent skills, structural verification | github.com/utsapoddar/dopa-kernel

- Freezes an objective into prioritized, verifiable requirements, selects the next action from explicit progress and information state, and invalidates stale evidence after each change.
- Requires fresh verifier re-runs as the only evidence of completion, so an agent cannot declare a task done on confidence alone; the current version passes 90 unit tests and 54 structural checks, and replay tests confirm protocol changes leave its decisions unchanged.

Alpha Digest - Automated Investor Intelligence Pipeline
Python, SEC EDGAR, CoinGecko, Yahoo Finance, RSS, LLM APIs, Jinja2, GitHub Actions | github.com/utsapoddar/alpha-digest

- Joined SEC EDGAR filings, CoinGecko crypto-treasury data, Yahoo Finance prices, and 12 RSS feeds to a 25-entity watchlist before any model call, so the model summarizes only collected evidence and returns structured JSON that renders both email and web editions.

PCOS Food Scanner - Deterministic Scoring with LLM Personalization
Python, Streamlit, Open Food Facts, NVIDIA NIM, Supabase |github.com/utsapoddar/pcos-scanner

- Made the 1 to 10 PCOS-aware score a fixed-rule nutrient calculation and limited the LLM to reading photos and explaining the result, so the same food always gets the same base score (30 tests).

MSDS Time Estimator - Bayesian Forecasting with Calibrated Uncertainty
Python, PyMC, pandas, NumPy, Bayesian inference |github.com/utsapoddar/msds-time-estimator

- Predicts study hours and calendar days for CU Boulder MSDS courses from completed-course anchors, crowdsourced student reviews, and topic similarity.

Education

University of Colorado Boulder, 2026 - 2028 (expected), Master of Science in Data Science

University of Calgary, 2018 - 2022, Bachelor of Science in Physics

Certifications

- Anthropic Academy, 2026, Claude 101; AI Fluency: Framework and Foundations; AI Fluency for Students; Claude Code in Action; Introduction to Agent Skills; Introduction to Model Context Protocol