# Aditya Kanawade

+91 9405657088 - Portfolio - adityakanawade1234@gmail.com - LinkedIn

## WORK EXPERIENCE

**Itanta Analytics** **Software Engineer**
*Pune, India* *February 2024 - Present*

- Architected a **5-agent "Prompt-to-Insights" (PTI) analytics platform** coordinated via **LangGraph state graphs**, incorporating hybrid LLMs and a **vector semantic caching layer** with RAG to cut manual report generation time by 80%.
- Built and productionized a custom **ReAct orchestration engine** in Django/FastAPI translating natural language queries into **parameterized MongoDB lookups**, cutting complex multi-level aggregation execution time by 45%.
- Developed and optimized data logging connectors (transferring **MQTT and PLC data** to time-series databases) that improved data transfer reliability by 25% and reduced latency by approximately 40%.

## PROJECTS

**Prompt-to-Insights (PTI) IIoT Analytics Platform** | *LangGraph, FastAPI, ChromaDB, MongoDB, AST* **January 2026**

- **Platform Overview & Stack:** Architected an enterprise GenAI analytics platform for Industry 4.0 plant operations, converting natural language/voice into production aggregation pipelines, automated ECharts/Handsontable dashboards, and SOP-grounded incident RCA across industrial protocols (MQTT, OPC-UA, QuestDB, MongoDB). Showcased at Global ProveIt Conference 2026 (USA).
- **Ownership & 5-Agent Architecture:** Spearheaded the backend of the **Analyze Module** in FastAPI as sole owner, orchestrating a **5-agent LangGraph StateGraph** (*Intent Resolver, Query Planner, Developer, Code Reviewer, Schema Validator*) using conditional edge re-entry to synthesize complex MongoDB/SQL pipelines.
- **Bounded Cyclic Self-Correction:** Established a bounded reflection loop with strict retry thresholds (maximum 2 retries) and partial state delta updates, eliminating infinite agent loops and state bloat while elevating end-to-end query generation accuracy from 70% to **over 95%**.
- **Dual-Path Semantic Caching:** Formulated a ChromaDB cosine-similarity cache using composite embeddings (intent + masked query + thoughts); injected verified code on positive matches and passed user thumbs-down rejections as negative "DO NOT REPEAT" constraints, slashing LLM token costs by **38%** and enabling **sub-50ms response times** on repeat queries.
- **Deterministic AST Security Guardrails:** Constructed a static code analysis engine via Python `ast` (Abstract Syntax Tree) to inspect generated queries before execution, achieving **100% mitigation** against mutating commands (DROP, UPDATE, DELETE) and unauthorized imports in isolated RabbitMQ task runners.
- **Streaming & Platform Integration:** Built an asynchronous WebSocket streaming engine (`StreamManager`) pushing live agent reasoning traces to the Angular frontend; integrated high-throughput ingestion connectors for core industrial data sources in the Integrate Module.

**Streamline360: ISA-95 SCADA Tag Platform** | *FastAPI, Angular, LiteLLM, MongoDB, GCP* **March 2026**

- **Platform Overview & Stack:** Architected and deployed an end-to-end industrial data harmonization platform translating thousands of cryptic OT/SCADA sensor tags into standardized **ISA-95 equipment hierarchies** (ProcessUnit, SubProcessUnit, operational attributes) using FastAPI, Angular, LiteLLM, MongoDB, and Nginx on GCP.
- **Sole End-to-End Ownership:** Served as the sole engineer reporting directly to the CTO; authored the entire application end-to-end, including UI frontend, async backend APIs, database schema modeling, JWT/RBAC security, and Linux server orchestration via PM2.
- **Chunked Batch Ingestion & Context Management:** Solved large-file token bloat and rate-limit bottlenecks (800+ tags per file) by architecting an asynchronous batching pipeline (50-100 tags/call), cutting manual spreadsheet mapping turnaround from **multiple days down to minutes**.
- **Adaptive SME Refinement Engine:** Devised an in-context feedback engine that extracts structured keywords from Subject Matter Expert (SME) corrections; automatically propagated refined taxonomy rules across future batches to achieve **over 95% harmonization accuracy**.
- **Resilience & Automated Fault Detection:** Configured an automated background health probe evaluating LLM latency and API quota exhaustion every 5 minutes, mitigating job stalls and ensuring seamless model fallback across Gemini, Claude, and GPT providers.

## TECHNICAL SKILLS

**Programming Languages:** Python, SQL, HTML, CSS, JavaScript
**Backend Frameworks:** FastAPI, Django, Flask, RESTful APIs
**AI & Data Engineering:** LangGraph, VectorDB/RAG, Pydantic & Instructor, LLM Orchestration
**Databases & Tools:** MongoDB, Git, Linux / System Processes, Nginx, Docker

## CORE COMPETENCIES

**Backend Architecture:** Designing distributed microservice patterns, high-throughput asynchronous services, multi-threaded background processing pipelines.
**AI Systems & LLM Engineering:** Multi-agent workflow design, vector-based RAG semantic caching, and strict schema-constrained outputs.

## ACHIEVEMENTS

**Global ProveIt Conference (USA)** **February 2025**

- Pioneered an advanced **natural language-driven analytics engine** and industrial data pipeline for real-time dashboards; recognized as a **top AI innovation** and presented to an international audience of 500+ attendees.

## EDUCATION

**Vishwakarma Institute of Technology** **Pune, India**
*B.Tech in Artificial Intelligence and Data Science (CGPA: 8.7)* *May 2024*