

WORK EXPERIENCE

Itanta Analytics

Backend Development | Django | Flask | APIs | Gen AI | Google cloud

- Engineered a **multi-agent** "Prompt-to-Insights" (PTI) analytics platform utilizing a hybrid LLM architecture and a **vector-based semantic caching** layer with Retrieval-Augmented Generation (**RAG**) to optimize query consistency, response latency, and code generation reliability, reducing manual report generation time by 80%.
- Designed and deployed a custom **ReAct orchestration engine** in Django/FastAPI that translates natural language industrial queries into parameterized MongoDB queries, handling complex multi-level aggregations.
- Engineered 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%.

Software Engineer

February 2024 - present

PROJECTS

Prompt-to-Insights (PTI) Analytics Engine – Itanta Analytics (Showcased at Global ProveIt Conference 2026, USA)

Gen AI developer | Langgraph | Django/ FastAPI | MongoDB | VectorDB

- Designed a multi-module enterprise platform structured into **Integrate** (data source configuration & schema mapping), **Analyze** (natural language querying), and **Visualize** (dashboard management), cutting dashboard generation time by 95%.
- Engineered an advanced **Analyze QA engine** driven by a multi-agent LLM workflow: utilizing fast models for intent parsing and parameterized query (PQ) generation, followed by a VectorDB semantic cache lookup to ensure logic consistency against historical queries.
- Implemented a robust **code generation and validation pipeline** where developer agent dynamically writes and refines component-generation code, routing through automated verification stages before execution
- Integrated **user feedback loops** (Thumbs Up/Down sentiment capture) directly into the **VectorDB cache storage layer** to continuously penalize or promote cached code performance.

January 2026

Industrial IOT Tag Harmonization Engine – Itanta Analytics

Full Stack Developer | Python | FastAPI | Pydantic & Instructor | MongoDB | Nginx / GCP

- Architected a **multi-model AI pipeline** using FastAPI and LiteLLM to automatically map messy industrial IOT/SCADA tags to the **ISA-95 process hierarchy** (Enterprise, Site, Area, Work Center, and Process Units).
- Designed a high-throughput parallelized processing engine capable of handling bulk file uploads via **threaded concurrency**, executing multi-stage model generation and cross-validation loops to ensure optimal mapping precision.
- Implemented strict schema enforcement and automatic retries using Pydantic and Instructor, paired with a deterministic safety-net mechanism that guarantees zero silent drops for unmapped inputs.
- Developed an administrative **SME refinement feedback** loop that captures expert corrections and dynamically injects structured guidance into future pipeline iterations for continuous rule consistency.

March 2026

SKILLS & INTERESTS

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

EDUCATION

Vishwakarma Institute of Technology

B.Tech In Artificial Intelligence And Data Science

Higher Secondary Education - 88.15%

Secondary Education - 93.60%

May 2024

CGPA : 8.7

March 2020

March 2018

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.
Database & Data Engineering: MongoDB aggregation pipelines, time-series data logging automated validation pipelines.

ACHIEVEMENTS

Showcased at Global ProveIt Conference 2025 (USA)

Developed 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

February 2025