AI-Powered Enterprise Data Quality Auditor

A multi-agent system that automatically evaluates, scores, and explains the quality of structured datasets using a Planner → Worker → Evaluator architecture. Built with Python, deployed on Hugging Face Spaces, and optimized for real-world Enterprise Data Engineering workflows.

LIVE DEMO

(Add your Hugging Face Space URL here)

PROJECT OVERVIEW

This project solves a real enterprise challenge: “How do we reliably measure data quality at scale without manual checks?”

Poor-quality data leads to inaccurate insights, faulty models, and costly decisions. This system automates detecting missing values, schema issues, duplicates, and outliers.

WHY AGENTS?

Traditional scripts run checks linearly, but enterprise-grade auditing needs multi-step intelligence:

- Planner Agent decides what to check.

- Worker Agents perform individual checks.

- Evaluator Agent merges outputs into a global score.

This provides modularity, fault tolerance, and scalability.

ARCHITECTURE

Planner → Worker → Evaluator pipeline:

- Planner: Generates audit plan based on dataset metadata.

- Workers: Perform checks (missingness, duplicates, schema, outliers).

- Evaluator: Aggregates results, assigns severity, produces final score.

FEATURES

- Automated data quality checks

- JSON-based structured reports

- Global Data Quality Score

- Session memory

- Full logging with Loguru

- Hugging Face deployment with Gradio UI

TECH STACK

Python, Pandas, Gradio, Loguru, Hugging Face Spaces

DEMO USAGE

1. Ask questions:

“What can you do?”

2. Run an audit:

“audit my dataset”

3. Audit file directly:

Upload test.csv to /content/project

Then: “audit test.csv”

FOLDER STRUCTURE

project/

- agents/

- core/

- memory/

- tools/

- main_agent.py

- app.py

IF I HAD MORE TIME

- Add LLM reasoning for insights

- Add schema and data drift checks

- Create visualization dashboards

- Build auto-fix cleaning agent

- Release as pip package