



Scan for more information!

NLP & LLM Utilization In High Volume Manufacturing Yield Report

Team Naruto

s3927188 - Dat Pham Xuan

s3927467 - Huan Nguyen Dang

s3929215 - Nhan Truong Vo Thien

s3927460 - Luong Nguyen Ngoc

Academic Supervisor

Mr. Hoang Van

Industry Supervisor

Ms. Hoai Nguyen



RMIT
UNIVERSITY

SCHOOL OF SCIENCE,
ENGINEERING
& TECHNOLOGY



BACKGROUND & MOTIVATION

High-Volume Manufacturing (HVM) generates massive volumes of yield and traceability data that are difficult to interpret with traditional tools. Engineers often spend significant time navigating static reports and fragmented dashboards, which slows down anomaly detection and decision-making.

Our project introduces an AI-driven assistant that combines Natural Language Processing (NLP), Large Language Models (LLMs), and data visualization to transform static reports into interactive insights. By enabling conversational queries, contextual follow-ups, and real-time visualization, the system reduces cognitive load, improves efficiency, and aligns with Intel's vision of smarter, AI-powered manufacturing.

OBJECTIVES

We aim to develop an AI assistant that makes semiconductor yield analysis faster and more interactive.

- Automate IMT report analysis with a chatbot
- Process and visualize yield data in real time
- Support contextual queries and follow-ups
- Ensure security with role-based access and cloud deployment
- Deliver a practical tool for Intel's manufacturing needs

METHODOLOGY

1. Front-End

- A user interface in the form of chatbot queries and report dashboards for accessing business data

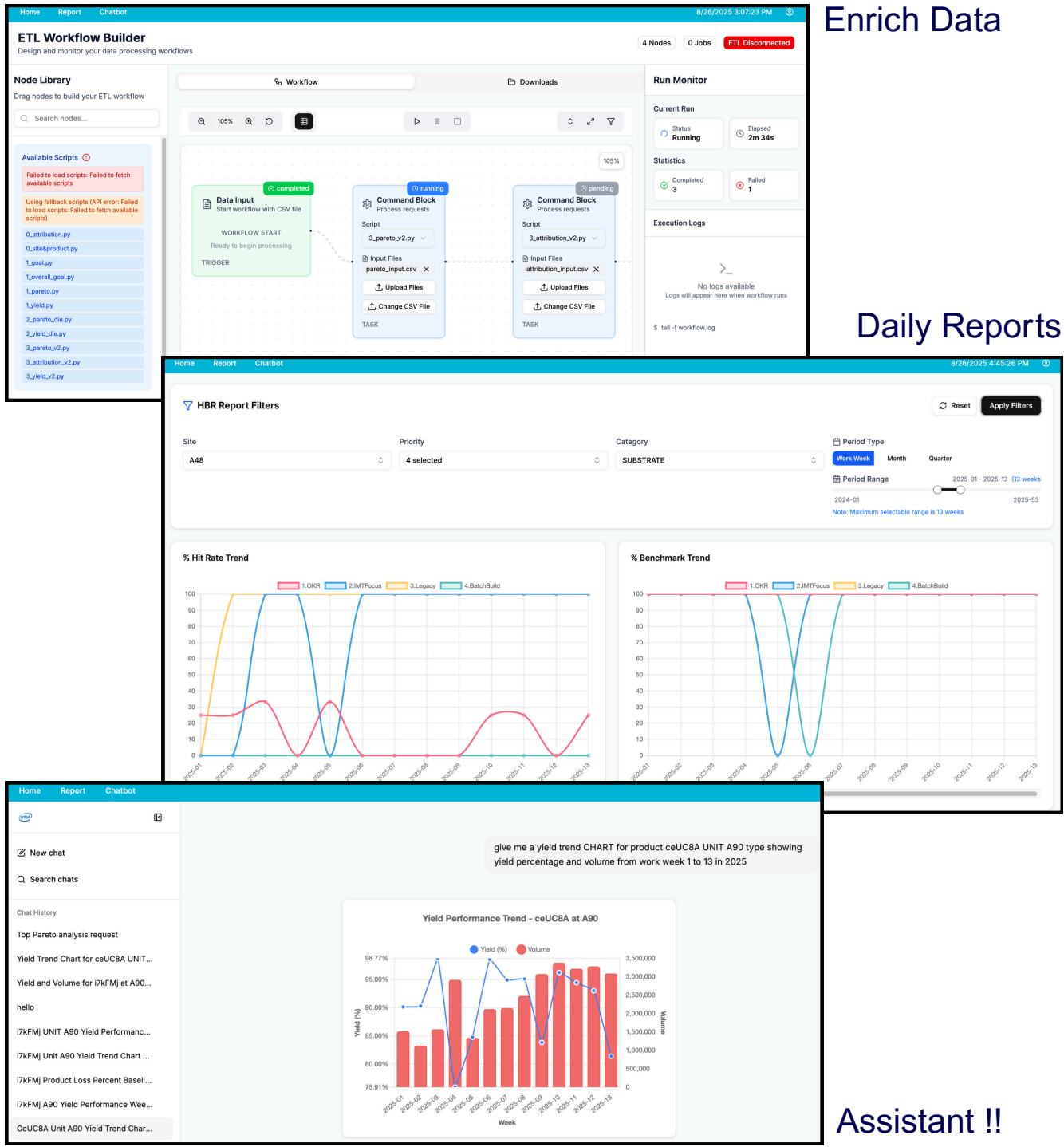
2. Modular Back-End Services that:

- Manages the uploading and structural retrieval of business data between storage and the user interface.
- Retrieves and uploads raw manufacturing data, processes it via ETL Pipelines, and persists it back to storage.
- Maps user questions into relevant business tools, and interprets structured data into a human-readable response

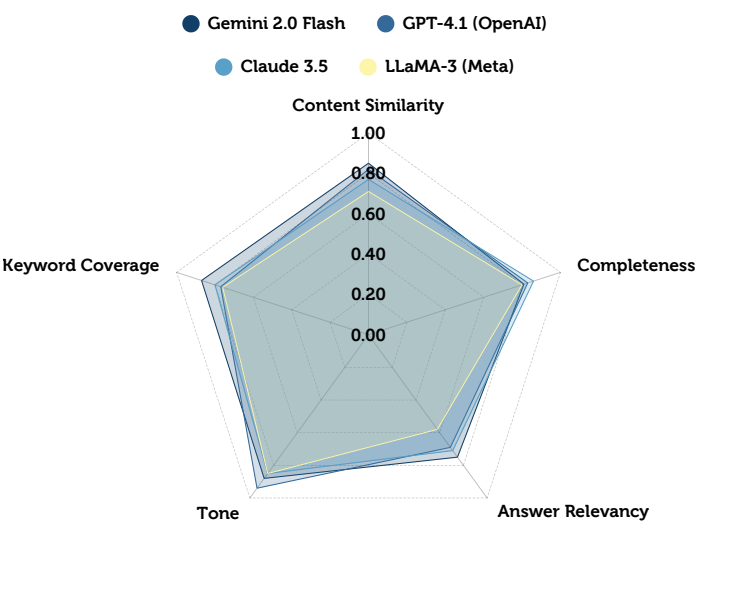
3. Data Storage

- This layer acts as a source of truth, ensuring that both structured data and unstructured knowledge are persistently stored and readily accessible.

OUTCOME: DATA PIPELINE



EVALUATIONS: LLM MODELS



Overall, the findings suggest that **Gemini 2.0 Flash** and **GPT-4.1** provide the most reliable trade-off between content similarity, tone, and completeness, making them the most suitable candidates for integration into the yield reporting chatbot.

CONCLUSION & FINDINGS

This project demonstrates the potentials of combining NLP, LLMs, and data visualization to enhance yield analysis in High-Volume Manufacturing. The AI-Powered chatbot successfully transform static reports into interatice insights, enabling engineering to queries data naturally, detect anomalies faster, and trace production issues with greater clarity.

This highlight that AI can go beyond automation, acting as a reliable assistant to improve efficiency, reduces times, and support Intel's vision of smarter, data driven manufacturing.