FACTORYTWIN®

VIRTUAL ASSISTANT

Presented By:

Rushil Srivastava

Jagrut Vaghasiya

PROJECT OVERVIEW

Objective

Develop an AI agent for FactoryTwin's Analytics Platform to run locally on-premises, focused on analyzing demand, on-time delivery, and root causes of late deliveries by answering descriptive, judgment-based, and advisory questions while prioritizing architectural efficiency.

INDUSTRY

The **Aerospace Parts Manufacturing** industry is a highly complex and precision-driven sector that requires stringent **quality control, efficient supply chain management,** and **strict compliance with regulatory standards**. Given the industry's reliance on **advanced materials** and **just-in-time production**, any inefficiencies can lead to costly delays. AI can transform aerospace manufacturing by **optimizing production schedules, predicting equipment failures** through predictive maintenance, improving quality control with computer vision, and enhancing **supply chain forecasting**. By leveraging AI-driven analytics, manufacturers can r**educe waste, improve on-time delivery, and ensure higher consistency** in part quality, ultimately increasing **operational efficiency** and **reducing costs**.
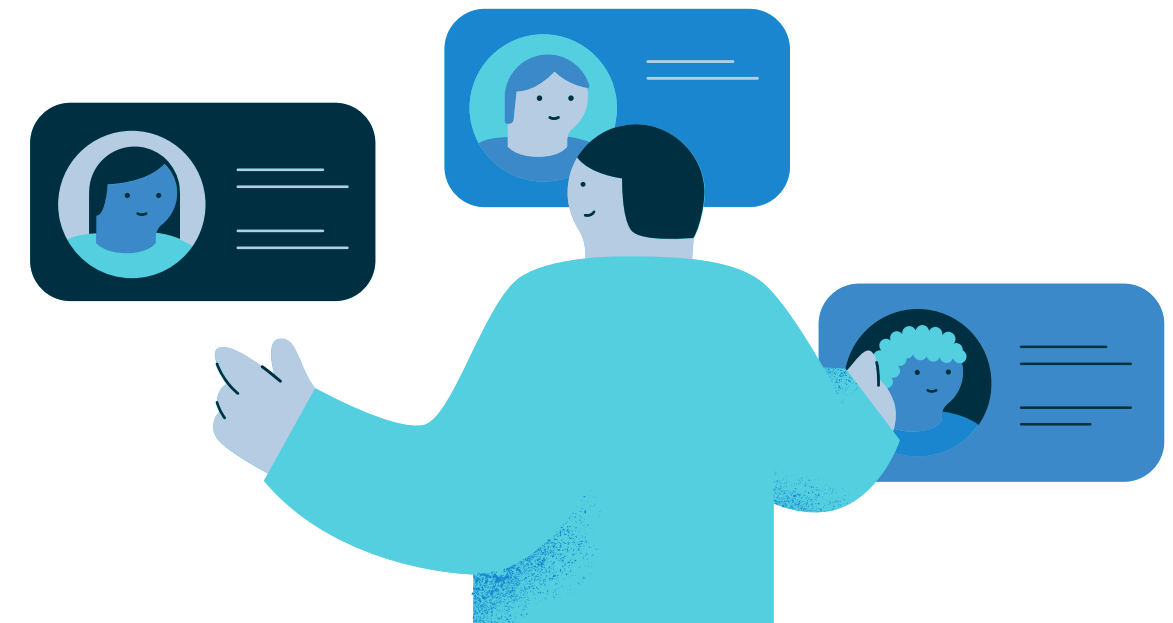
USE CASE

The use case for this **AI agent** focuses on enhancing aerospace manufacturing, particularly in the areas of **demand forecasting, on-time delivery analysis, and root cause identification** for production delays. The agent is designed to handle **three** types of queries:

1. **Descriptive Questions**: These involve direct data retrieval and simple analysis, such as projected revenue, parts with highest contribution margins, and identifying worst-performing areas in on-time delivery.
2. **Judgment Questions:** These require the agent to provide opinions based on data analysis, such as assessing the likelihood of meeting due dates or evaluating the accuracy of system lead times.
3. **Suggestion**/**Advice Questions:** These are the most complex, requiring the agent to provide guidance on improving on-time delivery, addressing root causes of delays, and optimizing system parameters.
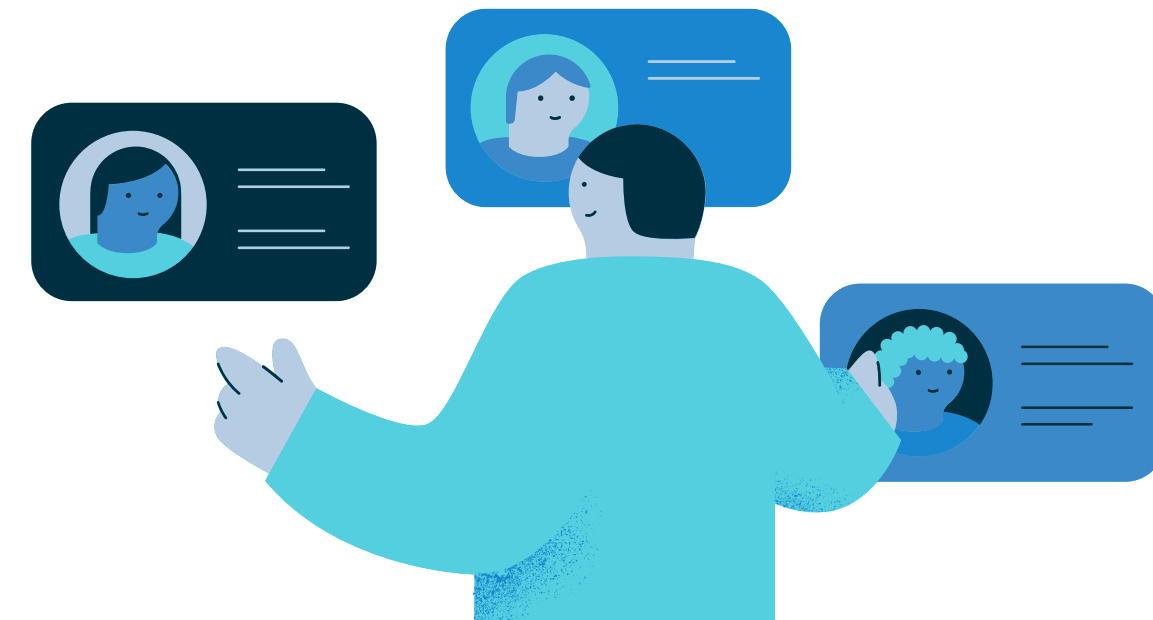
CONCEPT

1. **Model Integration:** Merge DeepSeek R1 and Llama 3.2 to develop a powerful, locally-deployed AI agent for FactoryTwin's analytics tool.
2. **Optimization:** Configure the model for 16GB RAM systems using 4-bit quantization to reduce memory usage while maintaining performance.
3. **Modular Architecture:** Implement specialized components for data ingestion, query processing, and response generation.
4. **Extended Context Window:** Support a 128K token context window to handle complex manufacturing queries.
5. **Hybrid Approach:** Combine DeepSeek's reasoning capabilities with Llama's language understanding for improved AI responses.
6. **Database & API Integration:** Connect with existing databases and APIs, utilizing vector databases for semantic search and short-term memory caching.
7. **Security & Compliance:** Deploy on-premise with encrypted storage and audit logging to ensure data security and regulatory compliance.

BUSINESS CASE
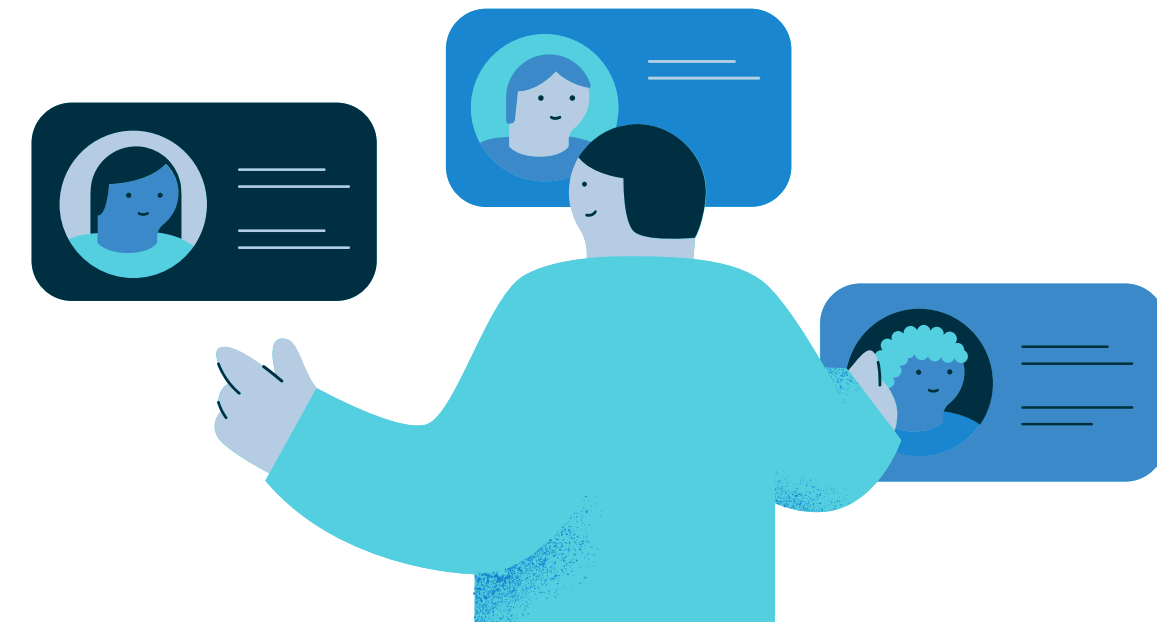
Problem

- Aerospace manufacturers face challenges in demand forecasting, on-time delivery, and root cause analysis of production delays
- Current ERP systems lack advanced analytics and real-time decision support capabilities
- Need for a solution that can handle complex queries while maintaining data security and regulatory compliance
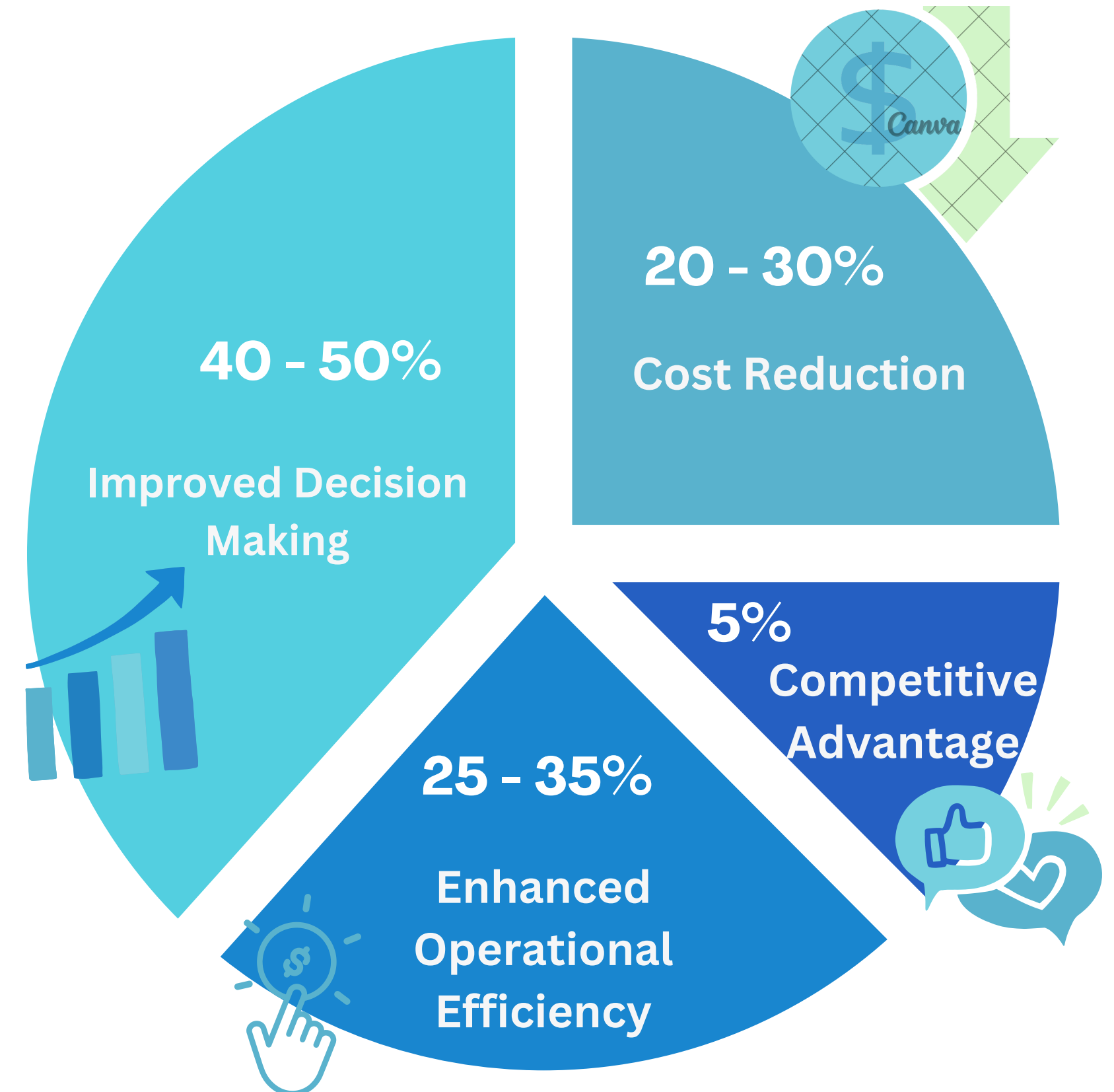
Solution

- Implement a locally-deployed AI agent using merged DeepSeek-R1 and Llama 3.2 models
- Optimized for 16GB RAM systems to fit existing hardware infrastructure
- Capable of handling descriptive, judgment, and suggestion/advice queries

BENEFITS

Enhanced Operational Efficiency
- 25-35% reduction in time spent on data analysis and reporting
- 15-20% improvement in production scheduling accuracy

Cost Reduction
- 10-15% decrease in inventory holding costs
- 20-30% reduction in unplanned downtime

Improved Decision-Making
- Real-time insights for proactive problem-solving
- 40-50% faster root cause analysis of production delays

Competitive Advantage
- Accelerated response to market changes
- Improved customer satisfaction through better on-time delivery performance

40 - 50% Improved Decision Making

20 - 30% Cost Reduction

5% Competitive Advantage
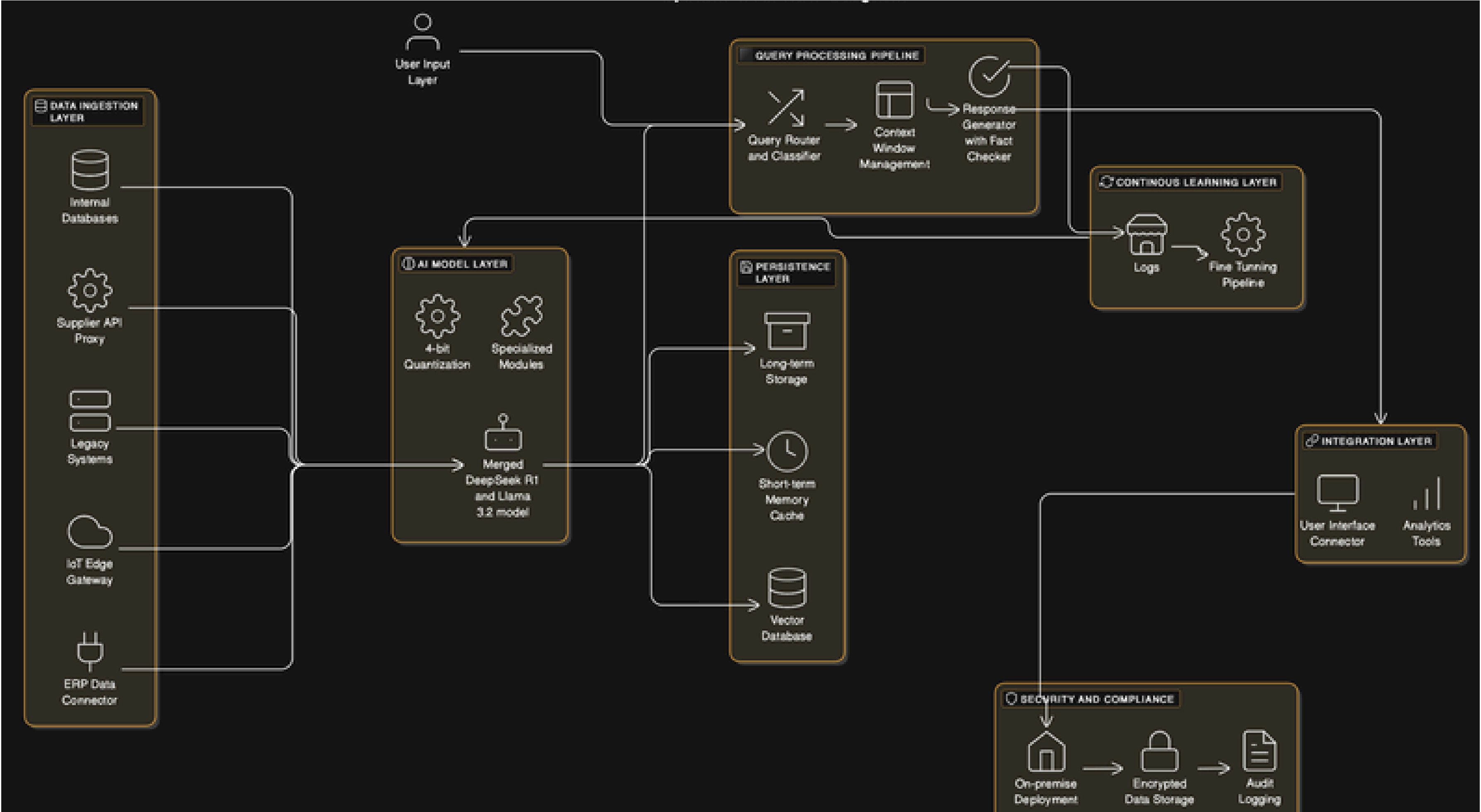
25 - 35% Enhanced Operational Efficiency

DATA NEEDS & SOURCES

The data utilized for training, validation, and testing will be artificially synthesized based on the data dictionary provided by FactoryTwin. A snapshot of the data can be found **here**
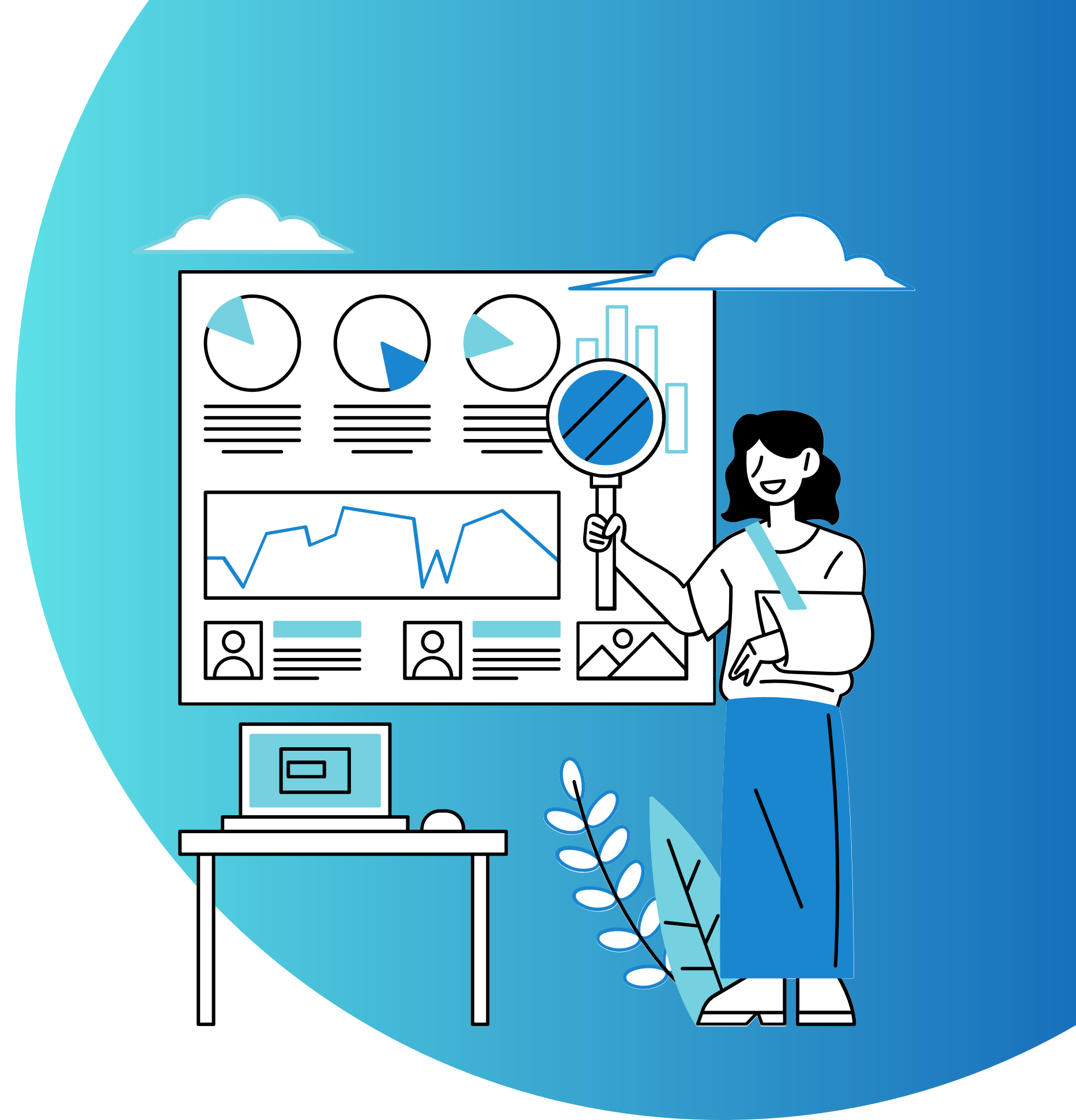
SOLUTION ARCHITECTURE

NEXT PHASE

1. Generate additional synthetic data.
2. Merge DeepSeek R1 and Llama 3.2
3. Fine Tune the model using the data generated
4. Build a functional prototype for demonstration.

THANK YOU