Multimodal Manufacturing Creator Report

Course Name: Generative AI

Institution Name: Medicaps University – Datagami Skill Based Course

Student Name(s) & Enrolment Number(s):

Sr no	Student Name	Enrolment Number
1.	Bhaskar Sharma	EN22CS301261
2.	Bhavesh Shadija	EN22CS301264
3.	Chitrak Jiyal	EN22CS301303
4.	Darshana Matade	EN22CS301309
5.	Dipanshi Yadav	EN22CS301337

Group Name: 08D3

Project Number: GAI-32

Industry Mentor Name: Mr. Suraj Nayak

University Mentor Name: Prof. Vineeta Rathore

Academic Year: 2025-26

Problem Statement & Objectives

1. **Problem Statement**: Multimodal Manufacturing Creator

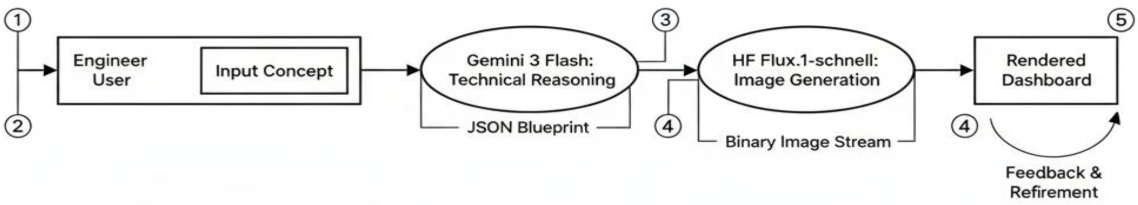
2. **Project Objectives**: Implement a multimodal GenAI application for Manufacturing concept visualization. This system integrates LLMs with Image Generation models (via API) to translate textual prompts into both descriptive narratives and high-fidelity visual representations (e.g., Product Prototype). This workflow demonstrates the synergy between text-to-text and text-to-image models to create comprehensive, multi-dimensional content for educational and professional presentations.

3. **Scope of the Project**: The project covers the development of a Streamlit-based dashboard that automates the transition from user input to technical JSON schemas and high fidelity industrial imagery.

Proposed Solution

1. **Key features:**
    - **Technical Reasoning:** Generates structured engineering specifications
    - **Visual Rendering:** Produces high-fidelity industrial imagery.
    - **Custom UI:** Industrial aesthetic achieved via CSS-in-JS and a multi-tab workspace.
    - **Resilience**: Multi-stage JSON parsing to handle model inconsistencies.
    - **Transparency:** Real-time architectural visualization using Mermaid.js.

2. **Overall Architecture / Workflow:** The application follows a strictly linear-sequential "Controller-Service" pattern:
    - **Input**: User provides concept, industry vertical, and aesthetic drivers.
    - **Logic Phase:** Gemini 3 Flash processes the prompt into a technical JSON schema.
    - **Visual Phase:** The image_prompt from the JSON is sent to the FLUX.1-schnell model via Hugging Face Inference API.
    - **Assembly:** The UI renders technical specs and the generated image side-by-side.

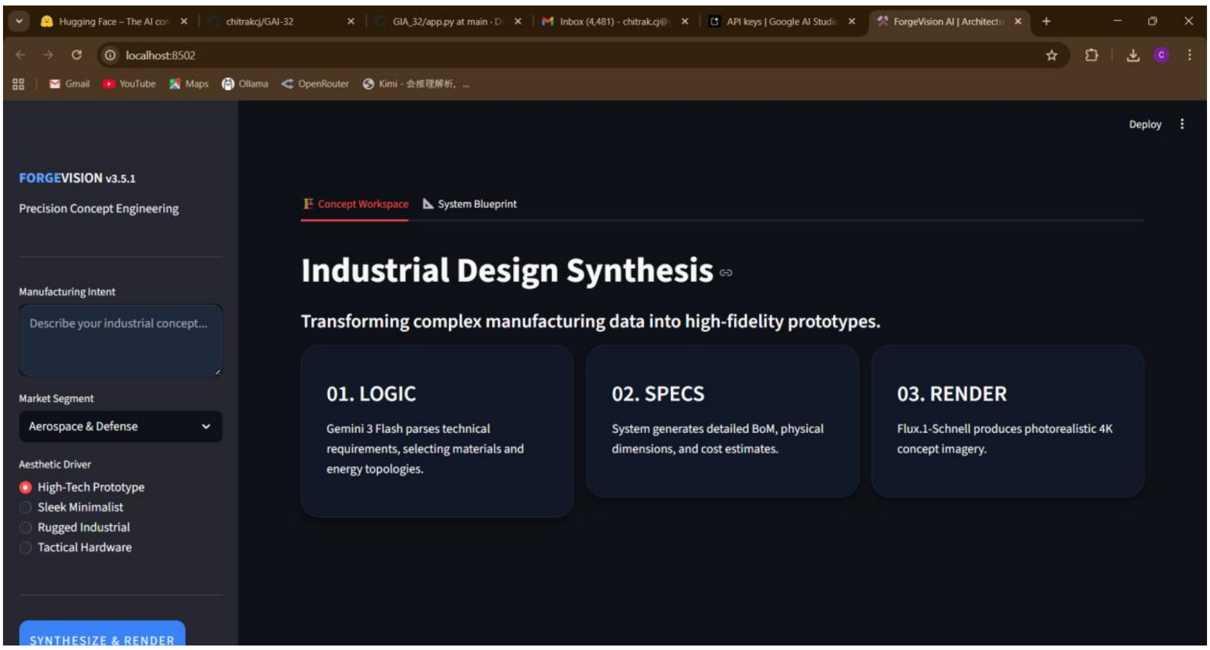


Workflow

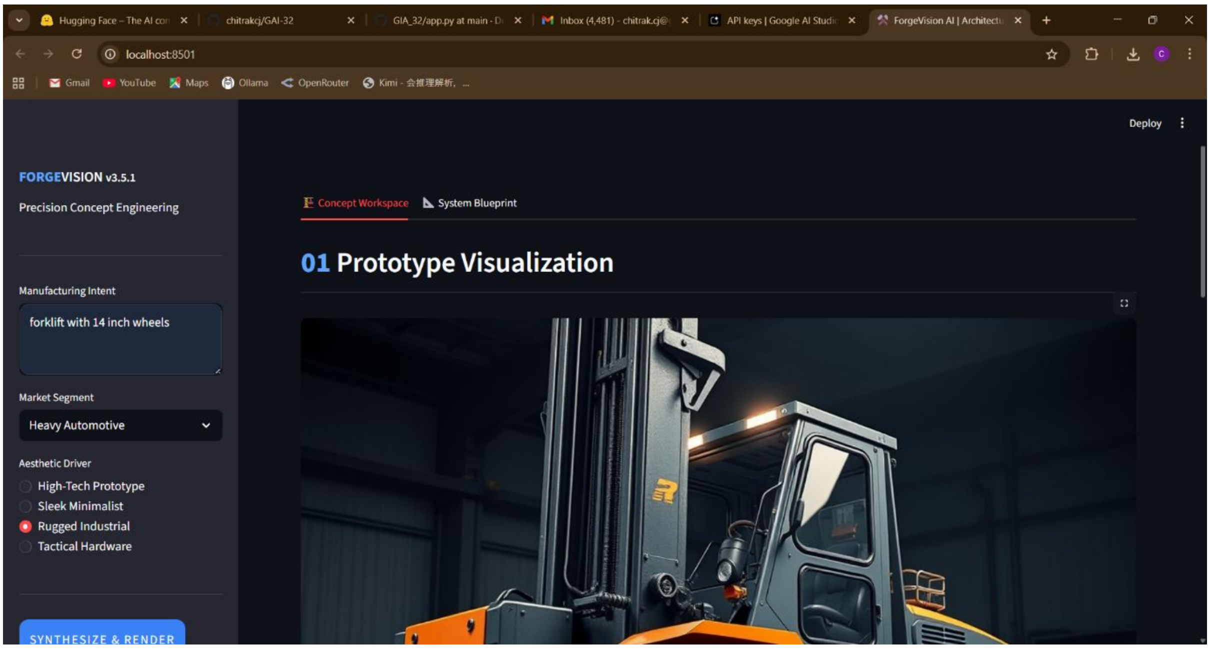
3. **Tools & Technologies Used:** Pyhton, Streamlit, Huggingface Hub Inference, FLUX.1-Schnell(image generation), Gemini 3 Flash(prompt enhancement), Github.

Results & Output

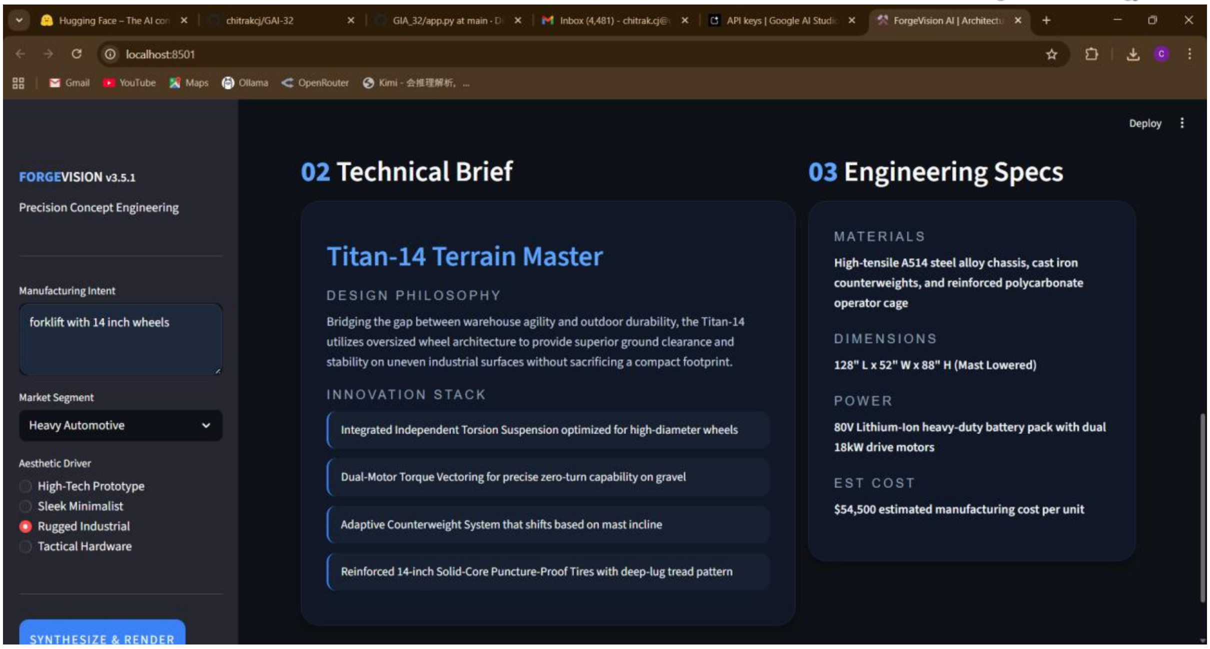
1. Screenshots / outputs:



Home Screen



Rendered Output(Image)



Rendered Output(Description)

2. **Key outcomes:**
    - Achieved LLM integration with image generation model for concept visualization.
    - Secure access and management of tokens and api.
    - High fidelity images with precision description and manufacturing details.
    - Achieved robust model orchestration by sanitizing LLM outputs for API stability.

Conclusion:

The project successfully integrates high-level technical reasoning with latent diffusion models to bridge the gap between abstract manufacturing ideas and concrete engineering concepts. The project highlights the efficiency of using a "Logic Engine" to drive a "Vision Engine," ensuring that generated visuals are grounded in technical specifications. The key learning outcomes include secure and effective usage of multiple ai models using tokens and APIs, deployment of app and standardising the outputs generated.

Future Scope & Enhancements:

1. **Performance:** Transitioning from a synchronous workflow to asynchronous processing to reduce perceived latency.
2. **Data Persistence:** Moving from volatile session-based storage to a permanent database for historical tracking.
3. **Optimization:** Implementing @st.cache_data to reduce redundant API calls and lower operational costs