

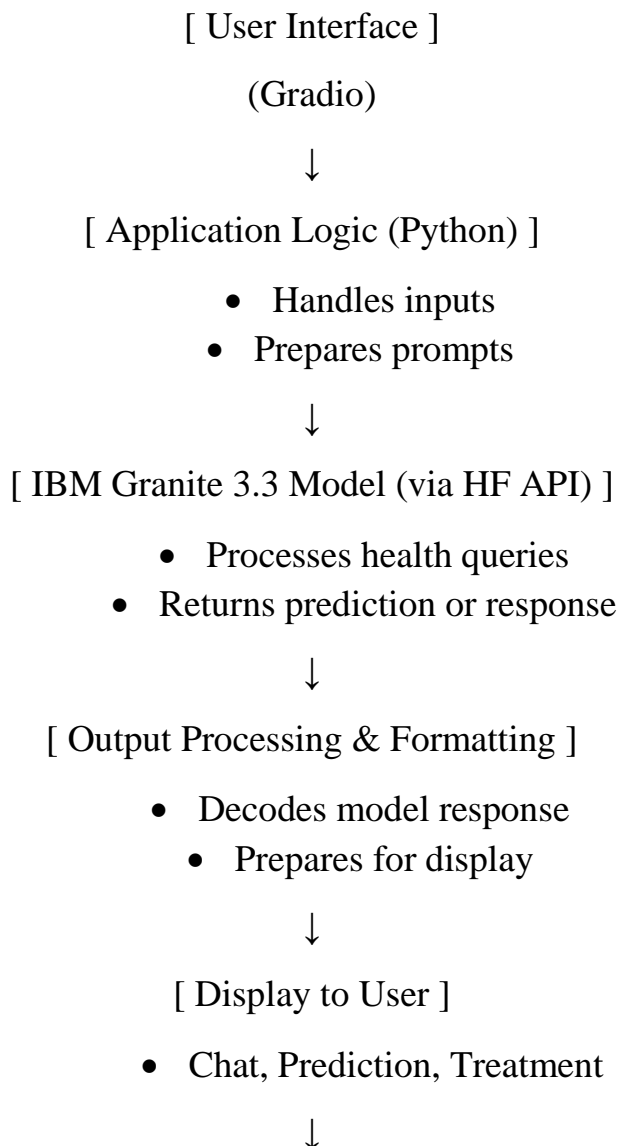
Project Design Phase-II

Technology Stack (Architecture & Stack)

Date	27 June 2025
Team ID	LTVIP2025TMID59196
Project Name	HealthAI: Intelligent Healthcare Assistant
Maximum Marks	4 Marks

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2



[Optional Storage (JSON/Firebase)]

- Store logs, inputs, outputs



[Hosting on Hugging Face Spaces]

- Cloud deployment (always available)

Table 1:Application Components

S.No	Component	Description	Technology
1.	User Interface	How user interacts with the application (Web UI tabs with Gradio)	Gradio, HTML, CSS
2.	Application Logic-1	Handles input, routing & interface logic	Python
3.	Application Logic-2	Converts health queries to prompts for the model	Python
4.	Application Logic-3	Manages tab structure and interactive responses	Gradio Blocks API
5.	Database	Stores patient questions, symptoms, responses (optional)	JSON / Firebase (optional)
6.	Cloud Database	If used, stores medical history/data securely	Firebase / MongoDB Atlas
7.	File Storage	Model logs, output saving (if needed)	Local storage / Google Drive
8.	External API-1	Secure model access and response using token	Hugging Face API (HF_TOKEN)
9.	External API-2	Used optionally for further health info or translation	Google Translate API (optional)
10.	Machine Learning Model	Predict diseases and treatment using a large language model	IBM Granite 3.3 Instruct
11.	Infrastructure (Server / Cloud)	Deployed on Hugging Face Spaces	Flask, Gradio, Hugging Face Spaces

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Used for building UI and backend API integration	Gradio, Flask, Transformers
2.	Security Implementations	Protects model and user data via authentication and secure endpoints	.env, Hugging Face Token, HTTPS

S.No	Characteristics	Description	Technology
3.	Scalable Architecture	Modular code with support for future ML/AI feature additions	Gradio Tabs
4.	Availability	Hosted on always-on platforms ensuring continuous uptime	Hugging Face Spaces
5.	Performance	Optimized response time and token limit for faster predictions	IBM Granite 3.3