

**Project Design Phase-II**  
**Technology Stack (Architecture & Stack)**

Date	27 June 2025
Team ID	LTVIP2025TMID31494
Project Name	HealthAI: intelligent healthcare assistant
Maximum Marks	4 Marks

**Technical Architecture:**

1. User Layer

Frontend UI or CLI where a user inputs a prompt (question or instruction).

2. Application Layer

Python script where the function `ask_granite(prompt)` processes user input.

This function tokenizes input, calls the pretrained IBM Granite model, and decodes the output.

3. Model Layer

Transformer Model: IBM Granite 3.3-2B, hosted on HuggingFace or local setup.

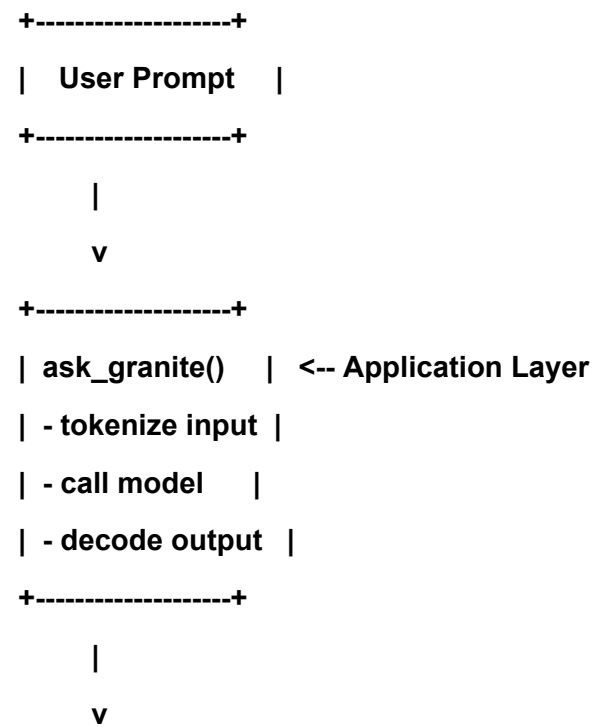
Tokenizer: Encodes and decodes text for the model.

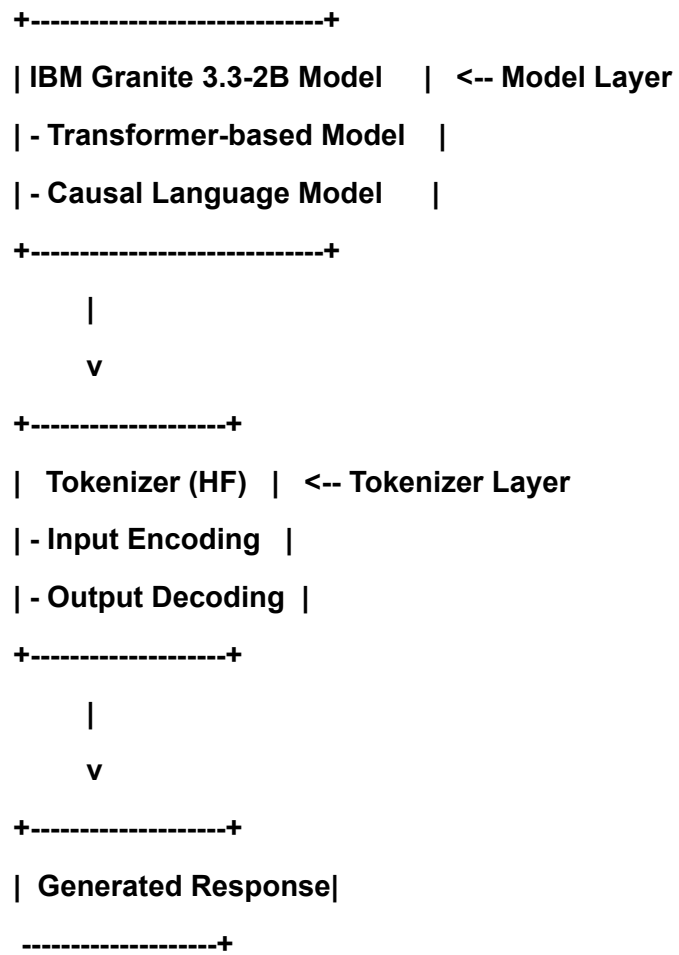
#### 4. Inference Layer

Model generates a response (inference) using pre-trained weights.

#### 5. Output Layer

Response is displayed to the user in natural language.





Component and technologies table

Component	Technologies	Purpose /description
Language model(LLM)	IBM Granite (graniye-3.3-2b instruct)	
Model interface	HuggingFace transformers library	
Backend framework	FastAPI	
Frontend Framework	Streamlit	
Execution Engine	pyTorch	
Tokenizer	AutoTokenizer from Hugging face	
Response generator	AutoModelForcausalLM from Hugging face	
Data handling	Pandas,numpy	
Visualization	Matplotlib,seaborn plotly	

**Table-2: Application Characteristics:**

Application Name	Health AI Assistant
------------------	---------------------

<b>Purpose</b>	<b>Symptoms based health query and possible condition suggestion using AI</b>
<b>Input Method</b>	<b>Command line interface user input</b>
<b>Ai model used</b>	<b>IBM Granite 3.3 2B instruct(via Hugging face transformers)</b>
<b>Model capabilities</b>	<b>Natural language understanding and generation</b>
<b>Core Functionality</b>	<ul style="list-style-type: none"> <li>-accepts user symptoms</li> <li>- sends prompt to model</li> <li>-displays suggestion</li> </ul>
<b>Interface type</b>	<b>Text based CLI interface</b>
<b>Technologies used</b>	Python Hugging face transformers IBM Granite LLM PyTorch
<b>User interaction</b>	<b>Conversational input /output</b>
<b>Output format</b>	<b>Ai generated natural language condition suggestions</b>
<b>Response personalization</b>	<b>Greets user by name and tailors response to symptom input</b>
<b>Limitation</b>	<b>Not a diagnosis tool,recommended consulting a healthcare professional</b>
<b>Execution trigger</b>	<b>Python script entry point(_main_)</b>
<b>Timestamp feature</b>	<b>Current time Displayed with response</b>
<b>Scalability</b>	<b>Current single -user,CLI -based,not optimized for production /multipleuser</b>
<b>extensibility</b>	<b>can intergrate treatment suggestion,web UL,Analytics Dashboard</b>


**References:**

[https://huggingface.co/ibm-granite/granite-3.3-2b -instruct](https://huggingface.co/ibm-granite/granite-3.3-2b-instruct)