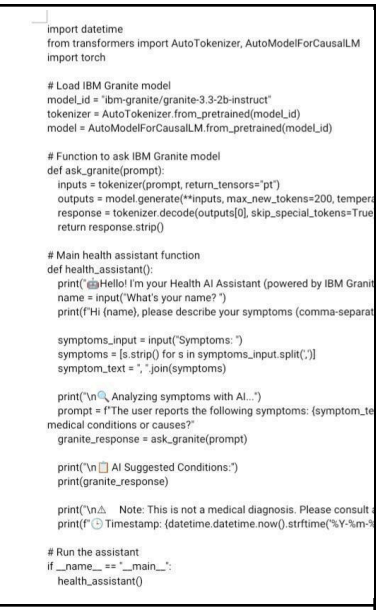
Project Development Phase
Model Performance Test

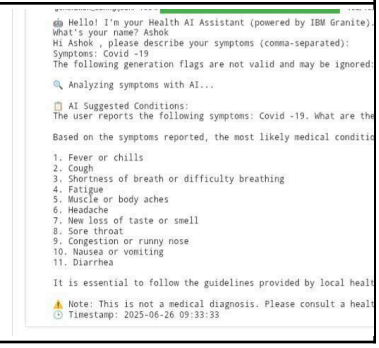
Date	25 june 2025
Team ID	LTVIP2025TMID31494
Project Name	HealthAI: intelligent healthcare assistant
Maximum Marks	

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Values	Screenshot
1.	Model Summary	-	1. Accuracy – How closely the AI output medically relevant conditions. 2. Relevance – Whether the AI's response appropriate and makes clinical sense for symptoms. 3. Latency – The time (in seconds) it take model to respond after receiving a promp 4. Response Time – Same as latency; imp real-time user experience. 5. Robustness – The model's ability to ha incomplete input, or unusual symptom ph 6. Safety – Ensures that the model outpu dangerous or hallucinated medical advice 7. Disclaimer – A warning or notice inclu AI output to inform the user that this is no diagnosis. 8. Token Count – Number of tokens in the output, important for performance and co APIs. 9. Inference – The act of generating a res the model after processing input. 10. Test Case – A prede ut (symp to evaluate the model's o

2.	Accuracy	Training Accuracy -
		
		Validation Accuracy -
			
			import datetime
			from transformers import AutoTokenizer, AutoModelForCausalLM
			import torch
			
			# Load IBM Granite model
			model_id = "ibm-granite/granite-3.3-2b-instruct"
			tokenizer = AutoTokenizer.from_pretrained(model_id)
			model = AutoModelForCausalLM.from_pretrained(model_id)
			
			# Function to ask IBM Granite model
			def ask_granite(prompt):
			inputs = tokenizer(prompt, return_tensors="pt")
			outputs = model.generate(**inputs, max_new_tokens=200, tempera
			response = tokenizer.decode(outputs[0], skip_special_tokens=True
			return response.strip()
			
			# Main health assistant function
			def health_assistant():
			print("🤖Hello! I'm your Health AI Assistant (powered by IBM Granit
			name = input("What's your name? ")
			print(f"Hi {name}, please describe your symptoms (comma-separat
			
			symptoms_input = input("Symptoms: ")
			symptoms = [s.strip() for s in symptoms_input.split(',')]
			symptom_text = ", ".join(symptoms)
			
			print("\n🔍 Analyzing symptoms with AI...")
			prompt = f"The user reports the following symptoms: {symptom_te
			medical conditions or causes?"
			granite_response = ask_granite(prompt)
			
			print("\n📋 AI Suggested Conditions:")
			print(granite_response)
			
			print("\n⚠ Note: This is not a medical diagnosis. Please consult
			print(f"🕒 Timestamp: {datetime.datetime.now().strftime('%Y-%m-%
			
			# Run the assistant
			if __name__ == "__main__":
			health_assistant()

3.	Fine Tunning Result(if Done)	Validation Accuracy -	 🤖 Hello! I'm your Health AI Assistant (powered by IBM Granite). What's your name? Ashok Hi Ashok , please describe your symptoms (comma-separated): Symptoms: Covid -19 The following generation flags are not valid and may be ignored: 🔍 Analyzing symptoms with AI... 📋 AI Suggested Conditions: The user reports the following symptoms: Covid -19. What are the Based on the symptoms reported, the most likely medical conditio 1. Fever or chills 2. Cough 3. Shortness of breath or difficulty breathing 4. Fatigue 5. Muscle or body aches 6. Headache 7. New loss of taste or smell 8. Sore throat 9. Congestion or runny nose 10. Nausea or vomiting 11. Diarrhea It is essential to follow the guidelines provided by local healt ⚠ Note: This is not a medical diagnosis. Please consult a healt 🕒 Timestamp: 2025-06-26 09:33:33