Project Development Phase Model Performance Test

Date	28 june 2025	
Team ID	LTVIP2025TMID50016	
Project Name	HealthAI: intelligent healthcare assistant Using IBM Granite	
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	-	Accuracy – How closely the Al output is medically relevant conditions.
			 Relevance – Whether the Al's response appropriate and makes clinical sense for symptoms.
			 Latency – The time (in seconds) it take model to respond after receiving a promp
			 Response Time – Same as latency; impreal-time user experience.
			5. Robustness – The model's ability to ha incomplete input, or unusual symptom phi
			 Safety – Ensures that the model output dangerous or hallucinated medical advice
			 Disclaimer – A warning or notice included AI output to inform the user that this is not diagnosis.
			 Token Count – Number of tokens in the output, important for performance and co APIs.
			Inference – The act of generating a res the model after processing input.
			10. Test Case − A prede ✓ out (symp to evaluate the model's o

2.	Accuracy	Training Accuracy -	import datetime from transformers import AutoTokenizer, AutoModelForCausalLM import torch
		Validation Accuracy -	# Load IBM Granite model model_id = "film-granite/granite-3.3-2b-instruct" tokenizer = Auto Takenizer.from_pretrained(model_id) model = AutoModelForCausali.M. from_pretrained(model_id) # Function to ask IBM Granite model def ask_granite(prompt):
			inputs = tokenizer(prompt, return_lensors="pt") outputs = model.generate(**inputs, max_new_tokens=200, temper; response = tokenizer.decode(outputs[0], skip_special_tokens=True' return response.strip()
			# Main health assistant function def health, assistant (): print("gla)Hello I in your Health AI Assistant (powered by IBM Granit name = input("What's your name?") print("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 AL.") prompt = "The user reports the following symptoms: {symptom_te medical conditions or causes?" granite_response = ask_granite(prompt)
			print(\n iii Al Suggested Conditions:") print(granite_response)
			print(^\n∆ Note: This is not a medical diagnosis. Please consult a 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 -	dip Nello! I'm your Nealth At Assistant (powered by 18% Granite), bhat's your name? Ashok Hi Ashok, please describe your symptoms (commo-separated); Symptoms: Covid -15 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 condition
			1. Fewer or chils 2. Cough 3. Shortness of breath or difficulty breathing 4. Fatigue 5. Murcle or body aches 6. Headache 7. Here loss of taste or smell 9. Congestion or runny nose 10. Nausea or vomiting 11. Diarrhaa
			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