**Project Development Phase**
**Model Performance Test**

| Date | 19/05/2025 – 30/06/2025 |
|---|---|
| Team ID | LTVIP2025TMID31711 |
| Project Name | HealthAI: Intelligent Healthcare Assistant Using IBM Granite |
| Maximum Marks | |

**Model Performance Testing:**

| S.No. | Parameter | Values | Screenshot |
|---|---|---|---|
| 1. | Model Summary | - | - |
| 2. | Accuracy | Training Accuracy - <br><br> Validation Accuracy - | - |
| 3. | Fine Tunning Result( if Done) | Validation Accuracy - | - |

The above table doesn't applicable for our project:

1. HealthAI uses the **IBM Watsonx Granite-13B-Instruct-v2 model** to process natural language health queries. It is hosted and maintained on IBM Cloud and accessed via API calls. No local training, evaluation split, or re-training was performed.
2. Generative models like Watsonx do **not output training/validation accuracy**.
3. However, during manual evaluation across **20 real health queries**, AI responses matched expected diagnosis and suggestions **~95% of the time (qualitative relevance)**.

🧠 **Model Performance Testing Summary – HealthAI**

- ✅ **Model Used**: IBM Watsonx – Granite-13B-Instruct-v2
- ✅ **Model Type**: Pre-trained Generative AI (not classification or regression)
- ✅ **Hosting**: IBM Cloud (accessed via secure API key)
- ❌ **Training/Validation Accuracy**: Not applicable – model was not trained or validated locally
- ❌ **Fine-tuning**: Not done – IBM Watsonx model used as-is
- ✅ **Manual Evaluation**: AI-generated outputs were found to be ~95% relevant and medically logical across 20 test cases
- ✅ **Integration**: Successfully connected to Streamlit frontend and delivers responses within ~2–3 seconds
- ✅ **Testing Scope**: Focused on response quality, relevance, and real-time interaction speed