Project Development Phase
Model Performance Test

Date	21 February 2025
Team ID	LTVIP2025TMID36237
Project Name	Sustainable smart city assistant using IBM granite LLM
Maximum Marks	

Model Performance Testing:

📋 **Performance Testing Template: Smart City Assistant (via IBM Granite LLM)**

S.No.	Parameter	Values	Screenshot
1	**Model Summary**	Setup: IBM Granite (e.g., 3.2 8B-Instruct) deployed via Watsonx for urban resource automation (e.g., traffic, waste, energy) using objects like IoT feeds.	
2	**Accuracy** (Task-specific metrics)	• Service categorization accuracy: __% - Energy demand forecasting accuracy: __% - Pollution anomaly detection accuracy: __%	•
3	**Confidence Scores** (e.g., YOLO-style)	For geospatial object detection in city maps (e.g., vehicles, bins, streetlights): 92%+ confidence. Use LLM's multimodal and vision capabilities.	

🧠 **Context: Why IBM Granite LLM Fits This Use Case**

- **Multimodal & Reasoning Capabilities**: Granite 3.2 models provide image understanding (through Granite Vision) and advanced chain-of-thought reasoning—ideal for processing visual city data like satellite imagery or sensor maps reddit.com+15ibm.com+15forbes.com+15.
- **RAG & Real-Time Context**: Granite supports retrieval-augmented generation (RAG), enabling real-time lookup (e.g., current traffic stats or service guidelines) to deliver precise citizen-facing responses forbes.com+1ibm.com+1.
- **Enterprise-Grade Safety & Control**: The Granite Guardian suite enforces content safety, confidentiality, and factual grounding—key for handling sensitive city data ndtvprofit.com+2crn.in+2indianexpress.com+2.
- **Time-Series Support**: Granite includes time-series models to forecast trends—e.g., city energy usage or pollution levels—complementing the LLM's planning output ibm.com+1forbes.com+1.

🛠️ **How to Populate the Template**

1. **Model Summary**:
    - Specify the exact Granite variant used (e.g., Granite-3.2-8B-Instruct with Vision and Guardian).

- Describe LLM workflows—e.g., "Multimodal pipeline: satellite image + sensor data → Vision model for detection → RAG augmentation → Instruct model for citizen-facing explanation."
- Capture a screenshot of the assistant in action (e.g., extracting bin locations or identifying streetlamp faults).

2. **Accuracy**:
   - For classification tasks like “automated request triage”: record metrics like Precision, Recall, F1.
   - For regression/forecasting (e.g., predicting daily energy usage): report MAE, RMSE, and R^2.
   - Include screenshots of performance dashboards or confusion/error charts.
3. **Confidence Scores**:
   - Use confidence metrics from visual detections (e.g., YOLO-like outputs from Vision model).
   - If the Vision model reports a 92% confidence for detecting a vehicle or streetlight, log that.
   - Capture annotated image outputs with bounding boxes and confidence percentages.

✅ **Example (Illustrative)**

S.No.	Parameter	Values
1	Model Summary	Granite-3.2-8B-Instruct + Vision + Guardian via Watsonx. Pipeline: road cam → Vision detection (vehicles, potholes) → Instruct generates citizen alert.
2	Accuracy	- Road hazard classification: Precision 92%, Recall 89%, F1 90% - Energy forecast (24h): MAE 4.1 kWh, RMSE 5.2 kWh, R^2 0.88
3	Confidence Scores	- Vehicle detection: 97% confidence - Pothole detection: 92% confidence

You can fill out the **Screenshot** column with actual UI grabs—like Vision model outputs, model performance charts, or prompt flows.