

# **SUSTAINABLE SMART CITY ASSISTANT USING IBM GRANITE LLB**

## **Project Documentation**

### **1.INTRODUCTION**

Project title: Sustainable Smart City Assistant Using IBM granite LLB

Team Leader: V.DHARSHIKA

Team member: S.PRIYA DHARSHINI

Team member: S.MANIMEGALAI

### **2.project overview**

Purpose:

The purpose of a Sustainable Smart City Assistant is to empower cities and their residents to thrive in a more eco-conscious and connected urban environment. By leveraging AI and real-time data, the assistant helps optimize essential resources like energy, water, and waste, while also guiding sustainable behaviors among citizens through personalized tips and services. For city officials, it serves as a decision-making partner-offering clear insights, forecasting tools, and summarizations of complex policies to support strategic planning. Ultimately, this assistant bridges technology, governance, and community engagement to foster greener cities that are more efficient, inclusive, and resilient.

Features:

#### **Conversational Interface**

Key Point: Natural language interaction

Functionality: Allows citizens and officials to ask questions, get updates, and receive guidance in plain language

#### **Policy Summarization**

Key Point: Simplified policy understanding

Functionality: Converts lengthy government documents into concise, actionable summaries.

## **Resource Forecasting**

Key Point: Predictive analytics

Functionality: Estimates future energy, water, and waste usage using historical and real-time data.

## **Eco-Tip Generator**

Key Point: Personalized sustainability advice

Functionality: Recommends daily actions to reduce environmental impact based on user behavior.

## **Citizen Feedback Loop**

Key Point: Community engagement

Functionality: Collects and analyzes public input to inform city planning and service improvements.

## **KPI Forecasting**

Key Point: Strategic planning support

Functionality: Projects key performance indicators to help officials track progress and plan ahead.

## **Anomaly Detection**

Key Point: Early warning system

Functionality: Identifies unusual patterns in sensor or usage data to flag potential issues.

## **Multimodal Input Support**

Key Point: Flexible data handling

Functionality: Accepts text, PDFs, and CSVs for document analysis and forecasting.

## **Streamlit or Gradio UI**

Key Point: User-friendly interface

Functionality: Provides an intuitive dashboard for both citizens and city officials to interact with the assistant.

### **3. Architecture**

#### **Frontend (Stream lit):**

The frontend is built with Stream lit, offering an interactive web UI with multiple pages including dashboards, file uploads, chat interface, feedback forms, and report viewers. Navigation is handled through a sidebar using the stream lit-option-menu library. Each page is modularized for scalability.

#### **Backend (Fast API):**

Fast API serves as the backend REST framework that powers API endpoints for document processing, chat interactions, eco tip generation, report creation, and vector embedding. It is optimized for asynchronous performance and easy Swagger integration.

#### **LLM Integration (IBM Watsonx Granite):**

Granite LLM models from IBM Watsonx are used for natural language understanding and generation. Prompts are carefully designed to generate summaries, sustainability tips, and reports.

#### **Vector Search (Pinecone):**

Uploaded policy documents are embedded using Sentence Transformers and stored in Pinecone. Semantic search is implemented using cosine similarity to allow users to search documents using natural language queries.

#### **ML Modules (Forecasting and Anomaly Detection):**

Lightweight ML models are used for forecasting and anomaly detection using Scikit-learn. Time-series data is parsed, modeled, and visualized using pandas and matplotlib.

### **4. Setup Instructions**

#### **Prerequisites:**

- ❖ Python 3.9 or later
- ❖ pip and virtual environment tools
- ❖ API keys for IBM Watsonx and Pinecone
- ❖ Internet access to access cloud services

- ❖ Installation Process:
- ❖ Clone the repository
- ❖ Install dependencies from requirements.txt
- ❖ Create a .env file and configure credentials
- ❖ Run the backend server using Fast API
- ❖ Launch the frontend via Stream lit
- ❖ Upload data and interact with the modules

## 5. Folder Structure

- 📁 app/- Contains all Fast API backend logic including routers, models, and integration modules.
- 📁 app/api/-Subdirectory for modular API routes like chat, feedback, report, and document vectorization.
- 📁 ui/-Contains frontend components for Stream lit pages, card layouts, and form UIs.
- 📄 smart\_dashboard.py - Entry script for launching the main Stream lit dashboard.
- 📄 granite\_llm.py- Handles all communication with IBM Watsonx Granite model including summarization and chat.
- 📄 document\_embedder.py-Converts documents to embeddings and stores in Pinecone.
- 📄 kpi\_file\_forecaster.py - Forecasts future energy/water trends using regression.
- 📄 anomaly\_file\_checker.py - Flags unusual values in uploaded KPI data.
- 📄 report\_generator.py - Constructs AI-generated sustainability reports.

## 6. Running the Application

To start the project:

- Launch the FastAPI server to expose backend endpoints.
- Run the Streamlit dashboard to access the web interface.
- Navigate through pages via the sidebar.

➤ Upload documents or CSVs, interact with the chat assistant, and view outputs like reports, summaries, and predictions.

➤ All interactions are real-time and use backend APIs to dynamically update the frontend.

### **Frontend (Stream lit):**

The frontend is built with Stream lit, offering an interactive web UI with multiple pages including dashboards, file uploads, chat interface, feedback forms, and report viewers. Navigation is handled through a sidebar using the stream lit-option-menu library. Each page is modularized for scalability.

### **Backend (Fast API):**

Fast API serves as the backend REST framework that powers API endpoints for document processing, chat interactions, eco tip generation, report creation, and vector embedding. It is optimized for asynchronous performance and easy Swagger integration.

## **7. API Documentation**

Backend APIs available include:

- POST/chat/ask - Accepts a user query and responds with an AI-generated message
- POST/upload-doc-Uploads and embeds documents in Pinecone
- GET/search-docs - Returns semantically similar policies to the input query
- GET/get-eco-tips - Provides sustainability tips for selected topics like energy, water, or waste
- POST/submit-feedback - Stores citizen feedback for later review or analytics
- Each endpoint is tested and documented in Swagger UI for quick inspection and trial during development.

## **8.Authentication**

- each endpoint is tested and documented in Swagger UI for quick inspection and trial during development.
  - This version of the project runs in an open environment for demonstration.
  - However, secure deployments can integrate:
  - Token-based authentication (JWT or API keys)
  - OAuth2 with IBM Cloud credentials
  - Role-based access (admin, citizen, researcher)
  - Planned enhancements include user sessions and history tracking.
8. Authentication

## 9. User Interface

The interface is minimalist and functional, focusing on accessibility for non-technical users. It includes:

- Sidebar with navigation

- KPI visualizations with summary cards

- Tabbed layouts for chat, eco tips, and forecasting

- Real-time form handling

- PDF report download capability

The design prioritizes clarity, speed, and user guidance with help texts and intuitive flows.

## 10. Testing

Testing was done in multiple phases:

- Unit Testing: For prompt engineering functions and utility scripts

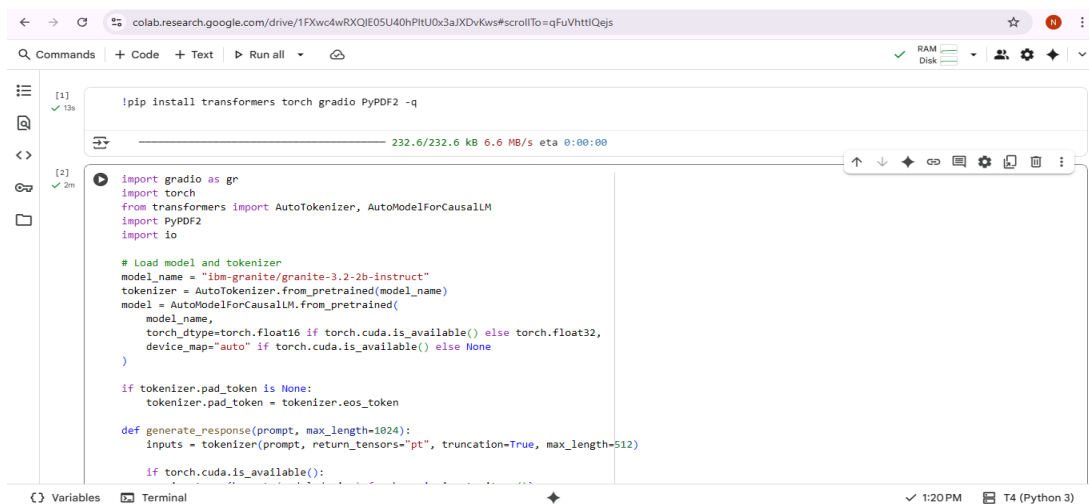
- API Testing: Via Swagger UI, Postman, and test scripts

- Manual Testing: For file uploads, chat responses, and output consistency

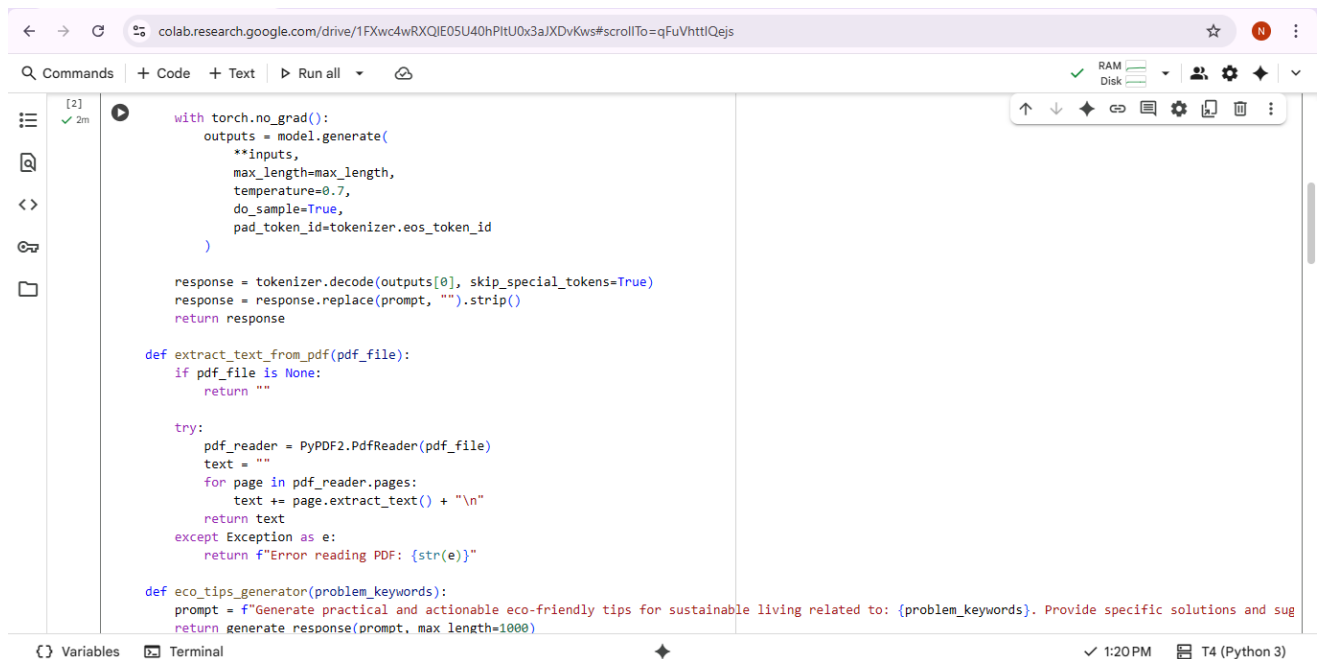
- Edge Case Handling: Malformed inputs, large files, invalid API keys

Each function was validated to ensure reliability in both offline and API-connected modes.

## 11. screen shots



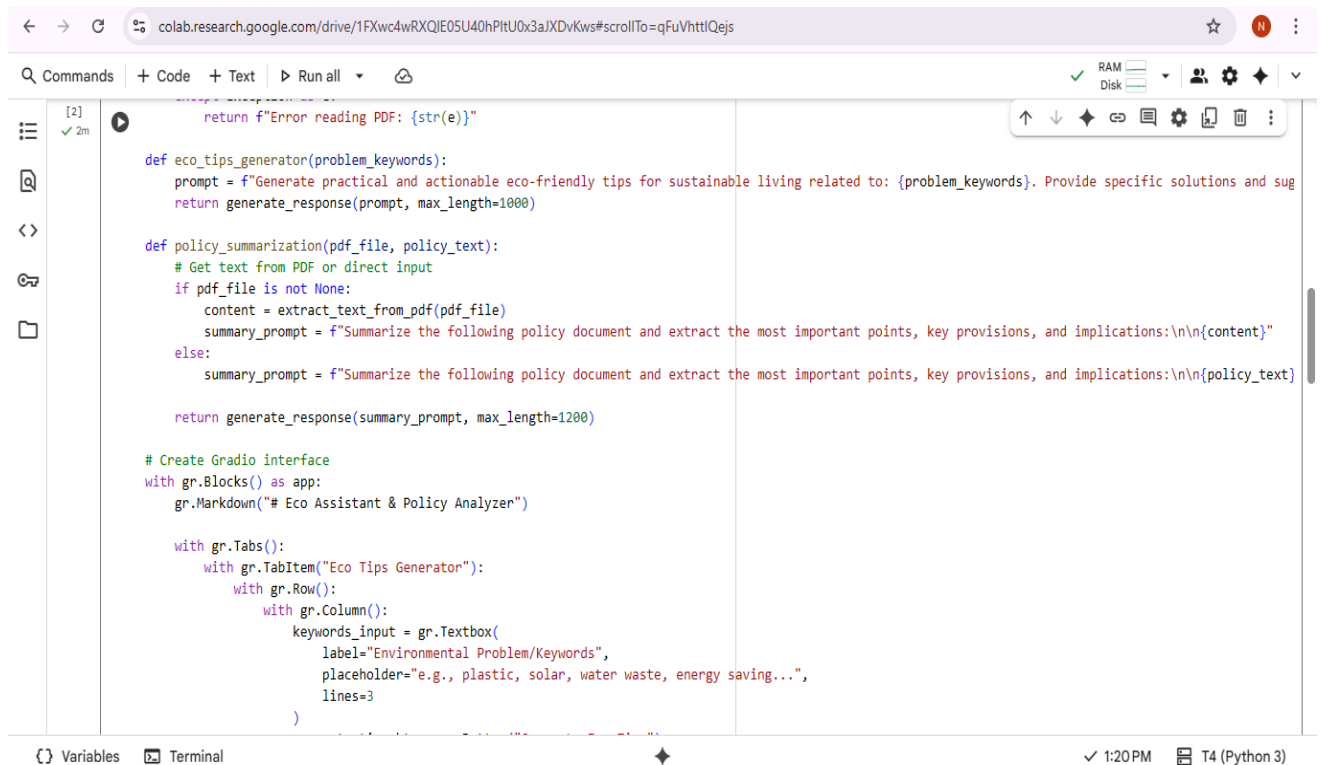
```
[1] ✓ 13s | pip install transformers torch gradio PyPDF2 -q  
232.6/232.6 kB 6.6 MB/s eta 0:00:00  
[2] ✓ 2m  
import gradio as gr  
import torch  
from transformers import AutoTokenizer, AutoModelForCausalLM  
import PyPDF2  
import io  
  
# Load model and tokenizer  
model_name = "ibm-granite/granite-3.2-2b-instruct"  
tokenizer = AutoTokenizer.from_pretrained(model_name)  
model = AutoModelForCausalLM.from_pretrained(  
    model_name,  
    torch_dtype=torch.float16 if torch.cuda.is_available() else torch.float32,  
    device_map="auto" if torch.cuda.is_available() else None  
)  
  
if tokenizer.pad_token is None:  
    tokenizer.pad_token = tokenizer.eos_token  
  
def generate_response(prompt, max_length=1024):  
    inputs = tokenizer(prompt, return_tensors="pt", truncation=True, max_length=512)  
    if torch.cuda.is_available():
```



This screenshot shows a Google Colab notebook with the following code:

```
[2] ✓ 2m  
with torch.no_grad():  
    outputs = model.generate(  
        **inputs,  
        max_length=max_length,  
        temperature=0.7,  
        do_sample=True,  
        pad_token_id=tokenizer.eos_token_id  
    )  
  
    response = tokenizer.decode(outputs[0], skip_special_tokens=True)  
    response = response.replace(prompt, "").strip()  
    return response  
  
def extract_text_from_pdf(pdf_file):  
    if pdf_file is None:  
        return ""  
  
    try:  
        pdf_reader = PyPDF2.PdfReader(pdf_file)  
        text = ""  
        for page in pdf_reader.pages:  
            text += page.extract_text() + "\n"  
        return text  
    except Exception as e:  
        return f"Error reading PDF: {str(e)}"  
  
def eco_tips_generator(problem_keywords):  
    prompt = f"Generate practical and actionable eco-friendly tips for sustainable living related to: {problem_keywords}. Provide specific solutions and sug  
    return generate_response(prompt, max_length=1000)
```

The interface at the bottom shows "Variables", "Terminal", and a status bar indicating "1:20 PM T4 (Python 3)".



This screenshot shows the continuation of the Google Colab notebook with the following code:

```
        return f"Error reading PDF: {str(e)}"  
  
def eco_tips_generator(problem_keywords):  
    prompt = f"Generate practical and actionable eco-friendly tips for sustainable living related to: {problem_keywords}. Provide specific solutions and sug  
    return generate_response(prompt, max_length=1000)  
  
def policy_summarization(pdf_file, policy_text):  
    # Get text from PDF or direct input  
    if pdf_file is not None:  
        content = extract_text_from_pdf(pdf_file)  
        summary_prompt = f"Summarize the following policy document and extract the most important points, key provisions, and implications:\n\n{content}"  
    else:  
        summary_prompt = f"Summarize the following policy document and extract the most important points, key provisions, and implications:\n\n{policy_text}"  
  
    return generate_response(summary_prompt, max_length=1200)  
  
# Create Gradio interface  
with gr.Blocks() as app:  
    gr.Markdown("# Eco Assistant & Policy Analyzer")  
  
    with gr.Tabs():  
        with gr.TabItem("Eco Tips Generator"):  
            with gr.Row():  
                with gr.Column():  
                    keywords_input = gr.Textbox(  
                        label="Environmental Problem/Keywords",  
                        placeholder="e.g., plastic, solar, water waste, energy saving...",  
                        lines=3  
                    )
```

The interface at the bottom shows "Variables", "Terminal", and a status bar indicating "1:20 PM T4 (Python 3)".

12.OUTPUT SCREEN:

colab.research.google.com/drive/1FXwc4wRXQJE05U40hPitU0x3aJXDvKws#scrollTo=qFuVhttlQejs

+ Code + Text ▶ Run all

app.launch(share=True)

/usr/local/lib/python3.12/dist-packages/huggingface\_hub/utils/\_auth.py:94: UserWarning:  
The secret `HF\_TOKEN` does not exist in your Colab secrets.  
To authenticate with the Hugging Face Hub, create a token in your settings tab (<https://huggingface.co/settings/tokens>), set it as secret  
You will be able to reuse this secret in all of your notebooks.  
Please note that authentication is recommended but still optional to access public models or datasets.  
warnings.warn(  
tokenizer\_config.json: 8.88k/? [00:00<00:00, 195kB/s]  
vocab.json: 777k/? [00:00<00:00, 13.9MB/s]  
merges.txt: 442k/? [00:00<00:00, 10.4MB/s]  
tokenizer.json: 3.48M/? [00:00<00:00, 45.2MB/s]  
added\_tokens.json: 100% [00:00<00:00, 2.44kB/s]  
special\_tokens\_map.json: 100% [00:00<00:00, 25.5kB/s]  
config.json: 100% [00:00<00:00, 22.2kB/s]  
`torch\_dtype` is deprecated! Use `dtype` instead!  
model.safetensors.index.json: 29.8k/? [00:00<00:00, 1.92MB/s]  
Fetching 2 files: 100% [00:00<00:00, 67.12kB/s]  
model-00002-of-00002.safetensors: 100% [00:01<00:00, 27.4MB/s]  
model-00001-of-00002.safetensors: 100% [01:06<00:00, 148MB/s]  
Loading checkpoint shards: 100% [00:18<00:00, 7.46s/it]

Terminal 1:20 PM

colab.research.google.com/drive/1FXwc4wRXQJE05U40hPitU0x3aJXDvKws#scrollTo=qFuVhttlQejs

Commands + Code + Text ▶ Run all

Loading checkpoint shards: 100% [00:18<00:00, 7.46s/it]  
generation\_config.json: 100% [00:00<00:00, 15.4kB/s]  
Colab notebook detected. To show errors in colab notebook, set debug=True in launch()  
\* Running on public URL: <https://fac3eb24124378f9c4.gradio.live>  
This share link expires in 1 week. For free permanent hosting and GPU upgrades, run `gradio deploy` from the terminal in the working directory to deploy to

soil pollution

Generate Eco Tips

1. Composting: Implement a composting system at home to convert organic waste into nutrient-rich soil amendment. Use kitchen scraps, yard waste, and even paper products that are safe for composting. This reduces the amount of waste sent to landfills and prevents the release of methane.
2. Cover Crops: Plant cover crops like clover, rye, or buckwheat between cash crops in your garden. These plants help prevent soil erosion, improve soil structure, and add organic matter when tilled back into the soil. Choose cover crops based on your climate and the specific needs of your soil.
3. Reduced Tillage: Adopt reduced tillage practices, such as no-till or strip-till farming, to minimize soil disturbance. This helps preserve soil structure, prevent soil erosion, and reduces carbon dioxide emissions from soil. It also promotes healthier soil ecosystems by allowing beneficial organisms to thrive.
4. Organic Farming: Transition to organic farming methods, which avoid synthetic pesticides and fertilizers. This helps prevent soil pollution by reducing chemical inputs into the ecosystem.

Variables Terminal 1:20 PM T4 (Python 3)