

Health AI: Intelligent Healthcare Assistant

Generative AI with IBM

1. Introduction

Project Title:

Health AI: Intelligent Healthcare Assistant

Team Members:

- D. Kaviya
- K. Karthika

2. Project Overview

Purpose:

The purpose of the Health AI: Intelligent Healthcare Assistant is to enhance healthcare delivery by leveraging advanced AI technologies like IBM Watsonx Granite to offer personalized and accessible healthcare solutions. The assistant helps patients, caregivers, and healthcare professionals by providing real-time information, predictive analytics, and decision-making support. It aims to improve health outcomes, patient engagement, and clinical efficiency through a seamless and intuitive interface.

Features:

Conversational Interface

Key Point: Natural language interaction

Functionality: Allows users to ask health-related questions, receive medication guidelines, and access preventive care tips using plain language.

Symptom Analysis & Recommendation

Key Point: Personalized care suggestions

Functionality: Analyzes symptoms based on user input and provides possible causes, recommended actions, and alerts for serious conditions.

Medical Report Summarization

Key Point: Simplified healthcare documentation

Functionality: Converts lengthy medical reports, prescriptions, or discharge summaries into concise, understandable formats for users.

Health Forecasting

Key Point: Predictive analytics

Functionality: Uses historical health data and patterns to predict trends such as blood sugar fluctuations, medication adherence, or potential health risks.

Lifestyle Guidance

Key Point: Preventive health support

Functionality: Provides personalized recommendations for diet, exercise, and mental wellness based on user health profiles.

Emergency Alerts

Key Point: Safety and intervention

Functionality: Detects anomalies in vital signs or health reports and sends alerts to users and healthcare providers.

Multimodal Input Support

Key Point: Versatile data processing

Functionality: Accepts text input, PDF reports, and structured health data in CSV formats for analysis.

User-Friendly Interface (Streamlit or Gradio)

Key Point: Accessibility for all users

Functionality: Offers intuitive dashboards for patients, caregivers, and medical professionals to interact with the assistant.

3. Architecture

Frontend (Streamlit):

The interface is built using Streamlit to create an interactive web dashboard. It includes pages for health queries, document uploads, symptom input forms, and reports. Navigation is handled through an easy sidebar layout for quick access.

Backend (FastAPI):

FastAPI serves as the backend framework that manages API endpoints for document processing, symptom analysis, forecasting, and user interaction. The system is optimized for asynchronous performance and easy integration with AI services.

LLM Integration (IBM Watsonx Granite):

The Granite LLM models from IBM Watsonx are employed to understand natural language queries, generate health recommendations, and summarize medical documents effectively.

Vector Search (Pinecone):

Uploaded medical documents and historical health data are converted into embeddings using Sentence Transformers and stored in Pinecone. Semantic search allows users to retrieve relevant health information quickly.

ML Modules (Forecasting and Anomaly Detection):

Lightweight machine learning models in Scikit-learn are used for health trend forecasting and detecting anomalies in vital signs or medical reports. Visualization is powered by 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

Installation Process:

1. Clone the project repository
2. Install dependencies using requirements.txt
3. Create a .env file with necessary credentials
4. Launch the backend server via FastAPI
5. Open the frontend using Streamlit
6. Upload medical reports or health data and explore interactive modules

5. Folder Structure

app/ – Contains backend logic including routers, models, and integrations

app/api/ – Subdirectory for APIs like symptom analysis, report generation, and feedback

ui/ – Contains frontend components for the dashboard and forms

health_dashboard.py – Entry point for the Streamlit interface

granite_llm.py – Handles communication with IBM Watsonx Granite for queries and summaries

document_embedder.py – Embeds medical documents into Pinecone

health_forecaster.py – Models trends like blood pressure or glucose levels

anomaly_checker.py – Detects irregularities in health data

report_generator.py – Generates personalized health summaries and recommendations

6. Running the Application

1. Launch the FastAPI backend server
2. Start the Streamlit frontend dashboard
3. Navigate through pages for symptom input, document upload, and health insights
4. View summaries, health predictions, and recommendations in real-time

7. API Documentation

Available Endpoints:

- * POST /chat/ask – Process health queries and provide responses.
- * POST /upload-doc – Upload medical files for embedding.
- * GET /search-docs – Retrieve documents based on symptoms or queries.
- * GET /get-health-tips – Provides wellness and lifestyle advice.
- * POST /submit-feedback – Allows users to submit feedback.

All endpoints are documented using Swagger UI for testing and inspection.

8. Authentication

For demonstration purposes, the system operates in an open environment. Future versions will implement:

- * JWT or API key-based authentication
- * OAuth2 with IBM Cloud credentials
- * Role-based access (patient, caregiver, doctor)
- * Session tracking and secure data access

9. User Interface

The interface focuses on accessibility and ease of use:

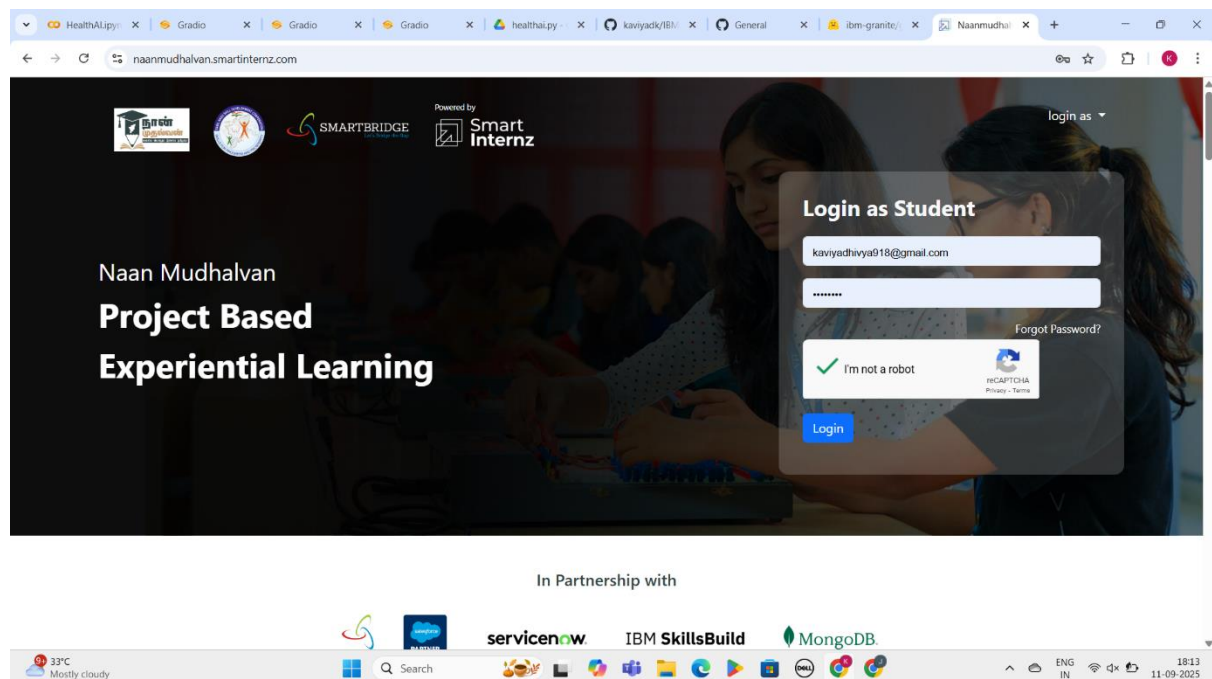
- * Sidebar navigation
- * Health KPI visualizations
- * Real-time chat for symptom analysis
- * Forms for document upload
- * PDF reports with summaries and alerts

10. Testing

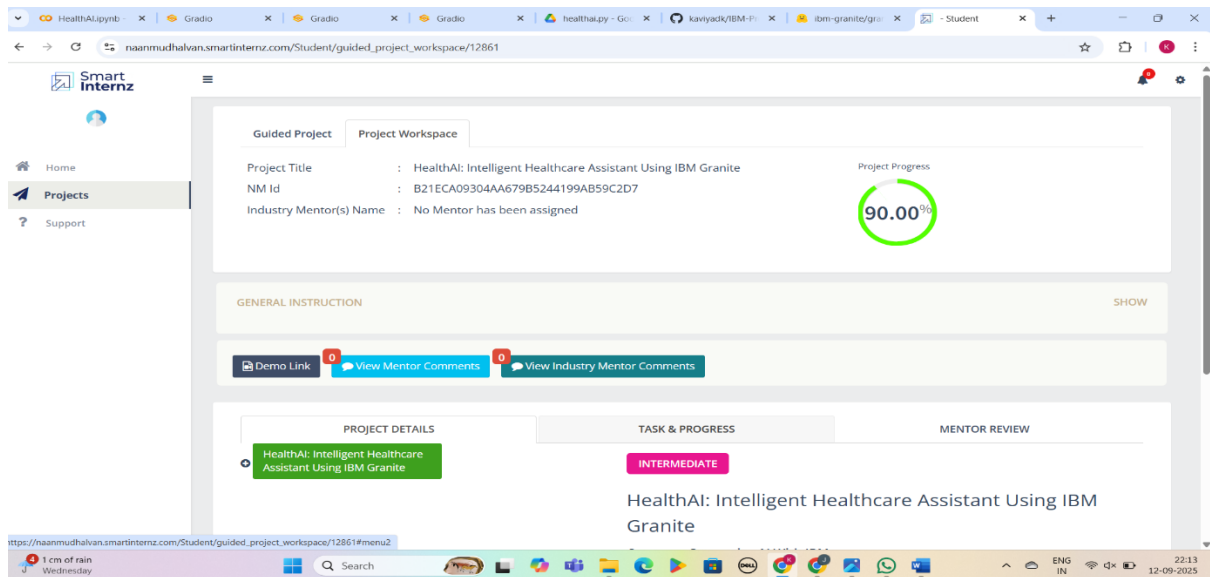
Testing includes:

- * Unit testing of AI prompts and logic
- * API testing using Swagger and Postman
- * Manual testing for symptom inputs and uploads
- * Handling edge cases like incorrect files or invalid data

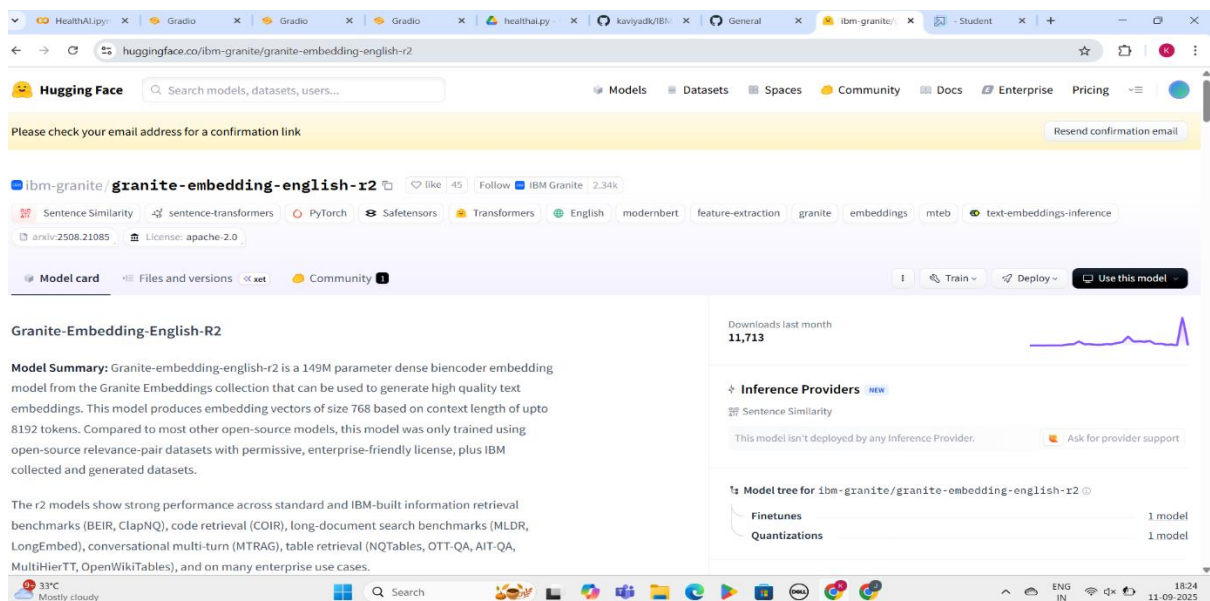
11. Screenshots



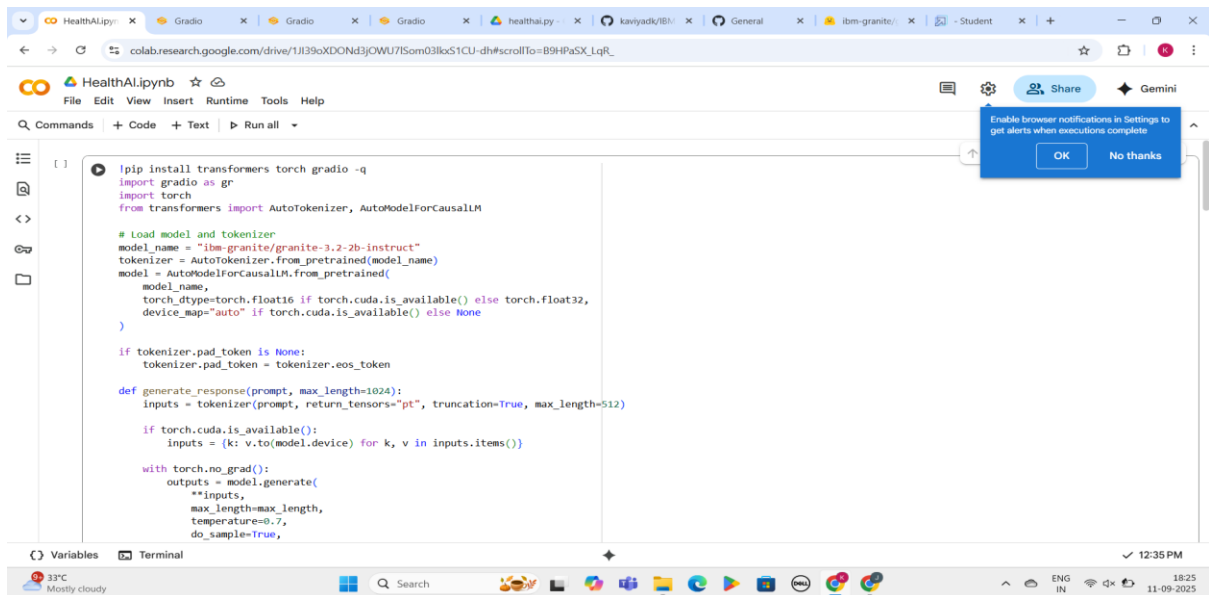
STEP-1: Login as student in Naan mudhalvan Smartinternz.



STEP-2: Then you will be redirected to your account then click Project and Access Resources then go to Guided Project and Go to Workspace then Project Workspace.



STEP-3: Search Hugging Face click first link on signup and create account then here for this project we are using granite-3.2-2b-instruct.



```
!pip install transformers torch gradio -q
import gradio as gr
import torch
from transformers import AutoTokenizer, AutoModelForCausalLM

# 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

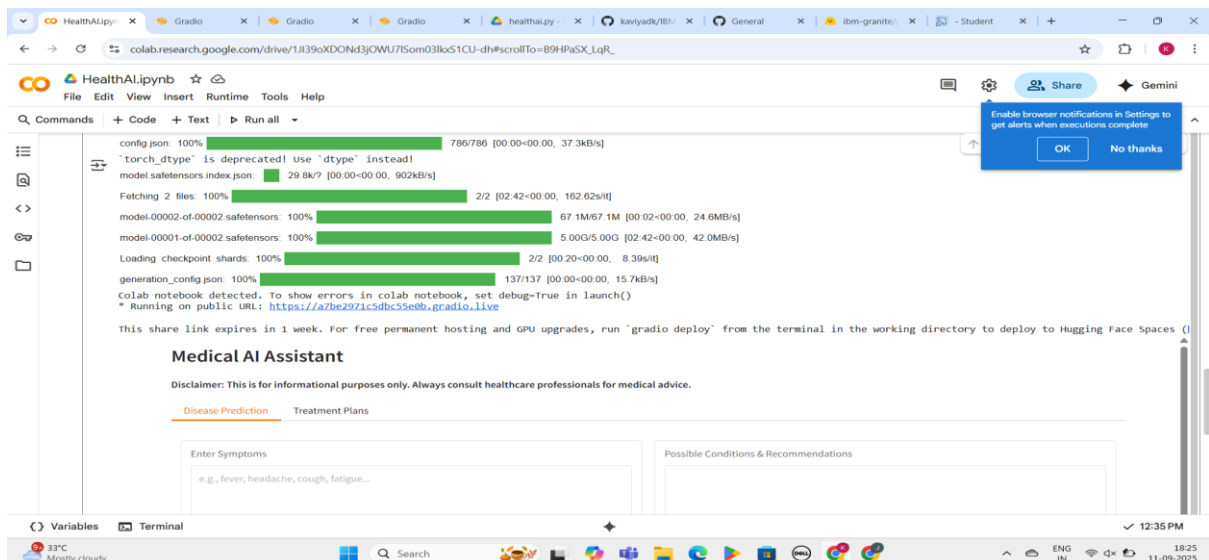
def generate_response(prompt, max_length=1024):
    inputs = tokenizer(prompt, return_tensors="pt", truncation=True, max_length=512)

    if torch.cuda.is_available():
        inputs = {k: v.to(model.device) for k, v in inputs.items()}

    with torch.no_grad():
        outputs = model.generate(
            **inputs,
            max_length=max_length,
            temperature=0.7,
            do_sample=True,
```

STEP-4: Now we will start building our project in Google collab, here we have to change title Health AI then change runtime T4GPU, run this command in first cell “!pip install transformers torch gradio -q”. Then run the rest of the code in the next cell.

OUTPUT:



```
config.json: 100% 786/786 [00:00<00:00, 37.3kB/s]
`torch_dtype` is deprecated! Use `dtype` instead!
model.safetensors.index.json: 29.8k/? [00:00<00:00, 902kB/s]
Fetching 2 files: 100% 2/2 [02:42<00:00, 162.62s/it]
model-00002-of-00002.safetensors: 100% 67.1M/67.1M [00:02<00:00, 24.6MB/s]
model-00001-of-00002.safetensors: 100% 5.00G/5.00G [02:42<00:00, 42.0MB/s]
Loading checkpoint shards: 100% 2/2 [00:20<00:00, 8.39s/it]
generation_config.json: 100% 137/137 [00:00<00:00, 15.7kB/s]
colab notebook detected. To show errors in colab notebook, set debug=True in launch()
* Running on public URL: https://a7be2971c5dbc55e9b.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 Hugging Face Spaces (l

Medical AI Assistant

Disclaimer: This is for informational purposes only. Always consult healthcare professionals for medical advice.

Disease Prediction Treatment Plans

Enter Symptoms
e.g., fever, headache, cough, fatigue...

Possible Conditions & Recommendations
```

The program has been executed the output will given the link to search about Health and it suggested a solution of the issue.

Medical AI Assistant

Disclaimer: This is for informational purposes only. Always consult healthcare professionals for medical advice.

Disease Prediction Treatment Plans

Enter Symptoms

fever and cold

Analyze Symptoms

Possible Conditions & Recommendations

- The combination of fever and cold symptoms suggests a viral upper respiratory infection, often referred to as a common cold or flu.

Medical conditions and medication suggestions:

1. **Common Cold/Influenza (Flu)**:

- "Diagnosis": Usually self-diagnosed based on symptoms, but influenza is confirmed by a rapid antigen test or a flu panel blood test.

- "Treatment":

- "Analgesics (Pain Relievers)": Acetaminophen (Tylenol) or ibuprofen (Advil, Motrin) can help manage fever and reduce pain.

- "Decongestants/Antihistamines": Over-the-counter options like pseudoephedrine (Sudafed) or loratadine (Claritin) can alleviate nasal congestion and sneezing.

- "Cough Suppressants": Dextromethorphan (Robitussin, Delsym) can help control coughs.

- "Hydration": Drink plenty of fluids to prevent dehydration.

- "Rest": Allow the body time to recover and rest.

2. **Allergic Rhinitis**:

- "Diagnosis": Typically diagnosed through symptom history, but tests like skin prick testing or allergy blood tests can confirm.

- "Treatment":

- "Antihistamines": Over-the-counter options like loratadine (Claritin) or fexofenadine (Allegra) can help.

Use via API Built with Gradio Settings

Medical AI Assistant

Disclaimer: This is for informational purposes only. Always consult healthcare professionals for medical advice.

Disease Prediction Treatment Plans

Medical Condition

fever

Age

30

Gender

Female

Medical History

normal

Generate Treatment Plan

Personalized Treatment Plan

1. Assess the cause of fever

- Mild to moderate fever (38.5°C to 39.5°C or 101.3°F to 103.1°F) in an otherwise healthy adult could be due to viral infections like common cold or flu.

- If fever lasts longer than 3-5 days, severe, or accompanied by worrying symptoms (e.g., difficulty breathing, severe headache, confusion, redness or pain around the eyes, persistent vomiting), seek immediate medical attention.

2. Home Remedies

a. Rest and hydrate

- Ensure adequate sleep and maintain proper hydration to support your immune system.

- Aim for 6-8 glasses of water or other fluids per day, adjusting based on physical activity and climate.

b. Over-the-counter (OTC) medication for symptom relief

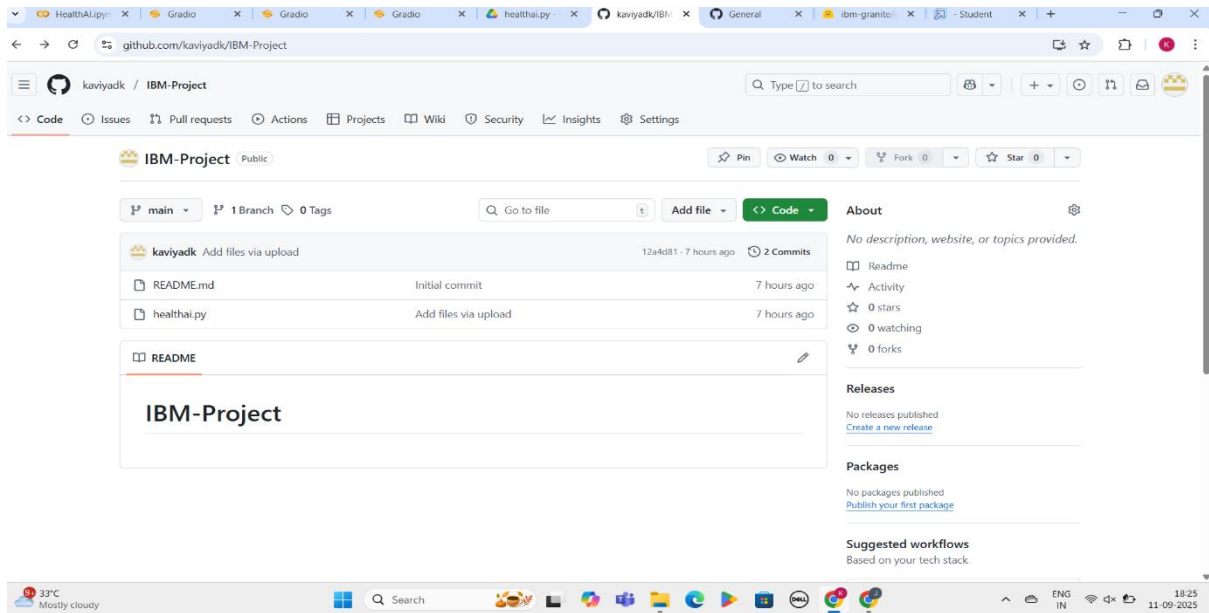
- Acetaminophen or ibuprofen (e.g., Advil, Motrin) can help reduce fever and alleviate minor aches and pains.

- Recommended dosage: 1 adult dose every 4-6 hours, not exceeding the maximum daily limit on the package (e.g., 1 extra-strength 650mg Acetaminophen or 2 standard-strength 200mg Ibuprofen).

c. Warm bath or compress

- Soaking in a lukewarm bath (up to your comfort level) or applying a warm, damp cloth to the forehead and other areas may provide relief and help lower the fever slightly.

Use via API Built with Gradio Settings



STEP-5: Search GitHub then signup create repository the name is IBM-Project. Now Download your code from Google collab then download as “.py”, then upload the file and click “Commit changes”.

12. Known Issues

- * Accuracy depends on user input and available data
- * Limited dataset for rare medical conditions
- * Privacy features to be integrated in future versions

13. Future Enhancements

- * Integration with wearable health devices
- * Support for voice input and emergency call triggers

- * Advanced diagnostics using imaging data
- * Multilingual support for regional healthcare access
- * Secure deployment with encryption and compliance standards.