System Architecture – Agentic AI Health Symptom Checker

Overview

The **Agentic AI Health Symptom Checker** uses a **Retrieval-Augmented Generation (RAG)** pipeline on **IBM Cloud** to provide reliable, multilingual health advice.
It integrates the **IBM Granite LLM** for natural language understanding with trusted health knowledge sources (WHO guidelines, government health portals, and verified journals).

The system ensures **safe, educational, and referral-based health guidance** instead of risky self-diagnosis.

Architecture Components

1. User Interface

- Input: User enters symptoms in natural language (e.g., *"I have a sore throat and fever"*).
- Mode: Chat-style interface (web/app/agent console).
- Multilingual support via Granite's language capabilities.

2. IBM Granite Model

- Large Language Model for **understanding user input** and generating context-aware responses.
- Handles **symptom interpretation**, **language translation**, and **response generation**.

3. RAG Layer (Retrieval-Augmented Generation)

- Queries **trusted medical data sources**:
    - WHO guidelines
    - Government health portals
    - Verified medical journals (open access)
- Ensures responses are **fact-based and medically accurate**.
- Filters irrelevant/unverified data.

4. Symptom Analysis & Recommendation Engine

- Maps user symptoms to probable conditions.

- Determines **urgency level** (low, medium, emergency).
- Suggests:
    - Home remedies & preventive measures
    - When to consult a doctor
    - Educational health guidance

5. IBM Cloud Services

- **Watsonx.ai Studio** – orchestration & model deployment
- **Watsonx.ai Runtime** – serving inference requests
- **Granite Foundation Model** – powering multilingual health conversations
- **Cloud Object Storage** – hosting reference datasets & guidelines
- **Agent Lab** – deployment & monitoring of the agent

Workflow

1. **User Input** – The user describes symptoms in natural language.
2. **Granite LLM Processing** – Model parses and interprets symptoms.
3. **RAG Retrieval** – Relevant data retrieved from WHO/government health repositories.
4. **Response Generation** – AI generates structured advice:
    - Probable conditions
    - Urgency level
    - Preventive care/home remedies
    - Referral to doctor if necessary
5. **User Output** – Safe, clear, and multilingual response delivered back.

Architecture Diagram (Conceptual)

User Query (Symptoms)

|

▼

IBM Granite LLM

|

▼

RAG Retrieval Engine ——► Trusted Health Sources

|

▼

Symptom Analysis & Advice Engine

|

▼

Safe Response (Causes, Urgency, Care Tips)

Key Features of the Architecture

- **RAG + Granite** → Combines generative AI with verified retrieval.
- **Trustworthy Sources** → WHO, Govt health portals, verified journals.
- **Multilingual Support** → Breaks language barriers in healthcare.
- **Cloud-native Deployment** → Powered by IBM Cloud Lite services.
- **Safety-first Approach** → Educational + referral guidance (avoids self-diagnosis).