**Milestone 1: Model Selection and Architecture**

**🔹 Objective of Milestone 1:**

To select the most appropriate foundation model and design the high-level architecture that supports Health AI’s intelligent healthcare capabilities. This forms the backbone for all subsequent development.

**🔹 Model Selection**

**✅ Why IBM Granite?**

IBM Granite is a suite of foundation models provided by IBM via the watsonx.ai platform. These models are specifically built for enterprise-grade applications with a focus on:

* Trustworthiness
* Explainability
* Data privacy
* Customization (domain-specific fine-tuning)

**✅ Model Features:**

* Trained on large-scale, diverse datasets including medical, scientific, and technical documents.
* Supports text generation, question answering, summarization, and information extraction.
* Deployable within IBM Cloud or hybrid environments, ideal for healthcare data compliance.

**✅ Use Cases Powered by Granite:**

* Patient Q&A based on symptoms or conditions.
* Medical report summarization.
* EHR data interpretation and recommendation.
* Drug interaction explanations.
* Multilingual healthcare support.

**🔹 Architecture Design**

**🎯 Goal:**

To build a **modular, secure, and scalable architecture** that integrates the Granite LLM and supports real-time healthcare interactions.

**🏗️ High-Level Architecture Components:**

**1. User Interface Layer (UI)**

* Web or mobile frontend
* Chatbot or voice-based virtual assistant
* Authentication for doctors/patients

**2. API Gateway**

* Routes requests from frontend to backend
* Handles authentication, logging, and rate-limiting

**3. Backend Layer**

* Built in Python (FastAPI or Flask)
* Responsibilities:
  + Preprocessing user inputs
  + Context handling (chat memory, patient history)
  + Constructing prompts for the Granite model
  + Parsing and post-processing model output

**4. IBM Granite Model (watsonx.ai)**

* Accessed via secure API
* Performs:
  + Q&A
  + Summarization
  + Entity extraction
  + Language understanding

**5. Data Storage Layer**

* EHR/EMR database (FHIR-compatible)
* Chat history logs
* Medical knowledge base (ICD codes, drug interactions)

**6. Security & Compliance Layer**

* Data encryption (TLS, AES-256)
* Access control (role-based)
* Audit logging
* Compliance with HIPAA/GDPR

**🧭 Diagram (Text View):**

css

CopyEdit

[User Interface]

↓

[API Gateway]

↓

[Backend Logic & Prompt Engine]

↔ [IBM Granite LLM via watsonx.ai]

↔ [Data Layer: EHR, Knowledge Base]

↓

[Output Response to UI]

**🔹 Deliverables for Milestone 1**

* Selection of IBM Granite as the core model
* Model documentation and usage plan
* Architecture diagram (to be created in the next milestone)
* Setup of basic infrastructure (IBM Cloud account, watsonx access)
* Definition of data flows and API contract

**🔹 Summary**

Milestone 1 lays the technical foundation for Health AI by selecting a powerful, healthcare-suitable model (IBM Granite) and designing an enterprise-ready architecture. This ensures that the assistant is scalable, secure, and clinically relevant as development progresses.