**Project Report**

**SDG 3 RAG-based Chatbot using Gemini**  
**AI-ML Internship - Gemini Chatbot Project**  
**College:** The Maharaja Sayajirao University of Baroda

**Member:**

* **VAGHELA TEJAS**

**Title**

**Development of an SDG 3 RAG-Based Chatbot for Health Guidance using Gemini API**

**Introduction**

Health and well-being (SDG 3) is a key global priority under the UN’s Sustainable Development Goals. However, many people lack accessible, reliable, and personalized health information related to this goal. With the rise of large language models and vector search technologies, it is now possible to build intelligent assistants that provide accurate health advice based on curated SDG 3 documents.

This project develops a **Retrieval-Augmented Generation (RAG)** chatbot using **Google Gemini Pro** and **FAISS** to answer health-related queries. It is designed as a public-facing assistant for SDG 3 awareness and education.

**Problem Statement**

Many individuals, especially in remote areas, lack access to professional health advice or clarity on health schemes and SDG-related guidance. Traditional information channels are static and difficult to navigate.

There is a strong need for an AI-driven, conversational, and context-aware system that can answer queries only from **verified health documents**, avoiding misinformation.

**Objective**

To build a chatbot that:

* Answers queries specifically related to **SDG 3: Good Health & Well-being**.
* Uses **RAG architecture** to ground responses in curated health documents.
* Is deployed as a **web interface** using Streamlit.
* Supports **future integration** with mobile apps or social channels.

**Why This Problem?**

With rising misinformation and generic responses from general-purpose LLMs, it’s important to restrict AI output to **domain-specific, verified knowledge** — especially in health.

By focusing on SDG 3 and using RAG, this project ensures **factual, safe, and targeted responses**, helping users better understand and benefit from government health schemes, disease prevention, and wellness practices.

**Solution Overview**

The chatbot combines:

* **FAISS** for fast document similarity search.
* **Sentence-transformers** for embedding health document vectors.
* **Gemini Pro (via API)** for LLM-based response generation.
* **Streamlit** for user interface and deployment.

**Key Features:**  
✅ **Health-specific** – Answers only SDG 3-related queries.  
✅ **RAG-based** – Responses come from relevant document chunks.  
✅ **Fallback logic** – Uses Gemini’s own knowledge if context is insufficient.  
✅ **Deployable** – Works on cloud or local machines.  
✅ **Secure** – No user data storage or tracking.

**Technical Implementation**

1. **Document Preparation**
   * Created sdg3\_docs.txt with curated health information.
   * Split into paragraph-level chunks.
2. **Embedding & Vector Search**
   * Used **sentence-transformers** to create embeddings.
   * Indexed with **FAISS** for fast nearest neighbor search.
3. **Gemini API Integration**
   * Accessed API securely using **python-dotenv**.
   * Constructed prompts based on **RAG logic** (context + query).
4. **Streamlit UI**
   * Users interact via chat input.
   * Maintains chat history with **session\_state**.
5. **Deployment**
   * Deployed on **Streamlit Cloud**.
   * Added .env and requirements.txt for easy setup.

**Why Google Gemini & FAISS?**

* **Gemini Pro** provides fast, cost-efficient LLM responses with flexible prompt handling.
* **FAISS** allows blazing-fast semantic search across custom datasets.

**Result:** Combines **accuracy** with **fluency** for better health guidance.

**Conclusion**

This project demonstrates how AI can support **SDG 3** by combining modern language models with domain-specific document search.

The chatbot is:

* **Reliable** – grounded in verified data.
* **Customizable** – can adapt to other domains.
* **Ready for deployment** – public health applications.

**Future Enhancements:**

* Multilingual support.
* Audio chat capabilities.
* WhatsApp/Telegram bot integration.