

The background is a solid dark blue. It features abstract geometric shapes, including triangles and diamonds, in lighter shades of blue, some with thin white outlines. At the bottom, there is a silhouette of a city skyline with various skyscrapers.

HEALTHCARE QA ASSISTANT

AI-POWERED CLINICAL Q&A WITH RESEARCH-BACKED
EVIDENCE

Presented by Vaibhav Garg

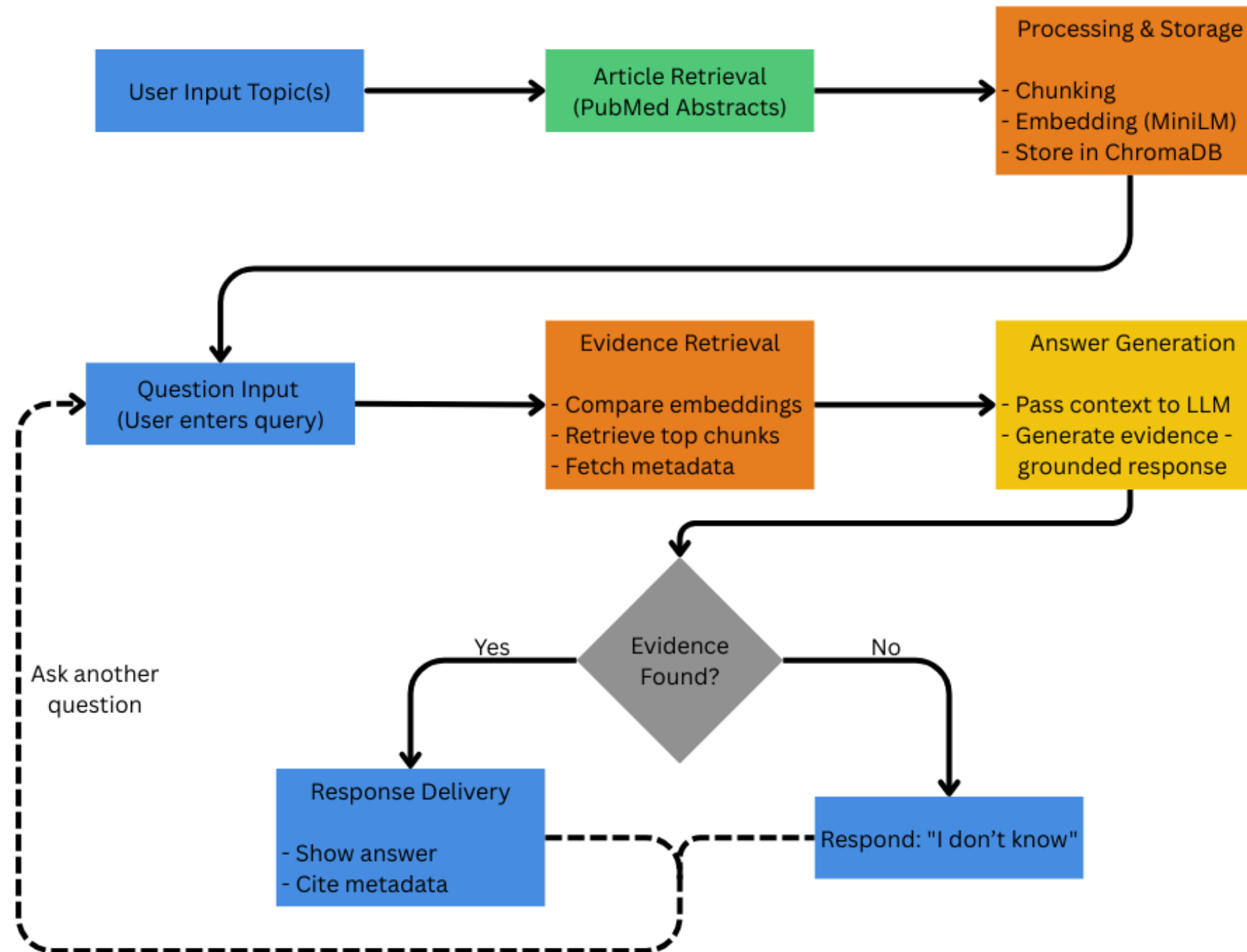
PROBLEM STATEMENT

- Healthcare professionals and patients often struggle to quickly access clear, evidence-based answers to medical questions, as reliable knowledge is scattered across vast volumes of research literature.
- This leads to delayed decision-making, uncertainty in treatment choices, and over-reliance on generic search engines that may surface incomplete or non-credible information.
- There is a growing need for a trusted, intelligent system that can retrieve, process, and summarize medical research into concise, transparent insights.
- An AI-powered assistant, equipped with PubMed article retrieval, semantic search, and evidence-grounded QA capabilities, can help clinicians and patients make informed decisions with confidence and efficiency.

PROJECT OBJECTIVES

- Build an AI-powered conversational assistant that helps healthcare professionals and researchers find evidence-based answers from PubMed.
- Enable efficient literature exploration by retrieving, chunking, and embedding medical articles for semantic search.
- Provide accurate, context-aware answers through a RAG pipeline powered by LLaMA3 (via GROQ).
- Ensure transparency in responses by displaying metadata such as authors, journals, and publication dates alongside generated insights.
- Deliver a user-friendly interface via Streamlit for intuitive querying and seamless interaction.

SYSTEM ARCHITECTURE



TECHNOLOGY STACK

This project integrates AI-driven retrieval pipelines and lightweight web technologies to deliver evidence-based, real-time medical insights from PubMed:

- **Frontend**
 - **Streamlit:** Builds a simple, responsive chat interface where users can enter medical queries and view evidence-backed answers.
- **Language Model**
 - **Groq + LLaMA 3.3:** Powers real-time, natural language question answering with concise, evidence-grounded medical responses.
- **Article Retrieval & Processing**
 - **PubMed :** Retrieves biomedical article abstracts and metadata directly from PubMed based on user-provided topics.
 - **Text Splitter:** Splits long abstracts into manageable chunks for efficient embedding and retrieval.

- **Semantic Embeddings & Storage**

- **Sentence Transformers (MiniLM):** Generates dense vector embeddings for semantic similarity search.
- **ChromaDB:** Stores embeddings and enables fast, local vector-based retrieval of relevant medical evidence.

- **QA Pipeline**

- **RAG (Retrieval-Augmented Generation):** Combines semantic retrieval from ChromaDB with LLaMA3 to produce reliable, citation-backed answers.

FUNCTIONAL WORKFLOW

This assistant system retrieves and synthesizes biomedical knowledge from PubMed to provide concise, evidence-based medical insights — all within a unified conversational interface.

1. Article Retrieval

- Users provide medical topics in the **Streamlit** sidebar (comma-separated).
- The app queries **PubMed** to fetch article abstracts along with metadata (authors, journal, publication date).

2. Processing & Storage

- Retrieved articles are split into smaller, manageable chunks.
- Each chunk is embedded using **Sentence Transformers (MiniLM)**.
- Embeddings and metadata are stored in **ChromaDB** for efficient semantic search.

3. Question Answering

a. Evidence Retrieval (ChromaDB Search)

- User questions in **natural language** are compared against stored embeddings.
- **Relevant chunks** (abstracts + metadata) are retrieved based on similarity.

b. Answer Synthesis (RAG Pipeline)

- The **Groq-hosted LLaMA 3.3** model synthesizes concise, evidence-grounded answers.
- If no strong evidence is found, the system responds with “**I don’t know**”.

4. Response Generation & Delivery

- Responses are transparent, citing relevant authors and journals.
- **LLaMA 3.3 (Groq)** ensures coherent, medically accurate text generation.
- **Streamlit** UI presents the answer in real time with source attribution.

KEY FEATURES

- **End-to-End Medical Q&A**
 - From query understanding to evidence-backed answers, the system automates the full retrieval-augmented QA flow for healthcare queries.
- **Evidence-Based Insights**
 - Generates concise medical responses grounded in PubMed abstracts, with transparent attribution to authors, journals, and publication dates.
- **AI-Powered Retrieval & Reasoning**
 - Uses Sentence Transformers for semantic query embedding and ChromaDB for retrieving the most relevant biomedical literature.
- **Streamlit-Powered Web UI**
 - An intuitive chat interface lets users ask questions in natural language and receive clear, real-time medical insights.
- **Modular & Scalable Architecture**
 - Decouples ingestion, retrieval, and answer generation, making the system extendable to additional medical domains or larger datasets.

USER INTERACTION PREVIEW

Ingest PubMed Articles

Enter topics (comma-separated):

intermittent fasting diabetes,
hypertension lifestyle, cardiovascular
exercise

Ingest Articles

✓ Articles ingested successfully.



Healthcare QA Assistant

Ask a Question

Type your medical question here:


Does intermittent fasting provide benefits for insulin sensitivity beyond weight loss in diabetes management?

Get Answer

Answer

Yes, intermittent fasting strategies have been linked to enhanced glycaemic control and improvements in insulin sensitivity, often preceding substantial weight loss. This suggests that intermittent fasting can provide benefits for insulin sensitivity beyond weight loss in diabetes management.


Supporting Evidence (from PubMed abstracts)

 Medical nutrition therapy for the management of type 2 diabetes mellitus.

Journal: Nature reviews. Endocrinology

Authors: Luigi Barrea, Ludovica Verde, Annamaria Colao, Lawrence J Mandarino, Giovanna Muscogiuri

Publication Date: 2025

 Dietary assessment in intermittent fasting: validation of a short food frequency questionnaire vs. food records in diurnal dry fasting and time-restricted eating.

Journal: Frontiers in nutrition

Authors: Isabelle C Schüssler, Christina L Pappe, Christel von Scheidt, Beeke Peters, Henrik Dommisch, Christian Kessler, Andreas Michalsen, Daniela A Koppold, Olga Pivovarova-Ramich

Publication Date: 2025

PROJECT SUMMARY

- Developed an AI-powered healthcare assistant that retrieves and synthesizes biomedical knowledge from PubMed to provide concise, evidence-based medical insights.
- Implemented **PubMed** API integration to fetch article abstracts and metadata on user-specified medical topics.
- Processed articles by splitting abstracts into smaller chunks, embedding them with **Sentence Transformers (MiniLM)**, and storing them in **ChromaDB** for efficient semantic search.
- Integrated a **Retrieval-Augmented Generation (RAG)** pipeline using **Groq's LLaMA 3.3** to generate transparent, source-linked answers from retrieved biomedical literature.
- Designed an interactive **Streamlit** app with a sidebar for topic selection and a chat interface for natural language Q&A, with real-time display of metadata.
- **Live App:** <https://vaibhav-project-healthcare-app-assistant.streamlit.app/>
- **GitHub Repository:** <https://github.com/vaibhavgarg2004/Healthcare-QA-Assistant>

**THANK
YOU**

