# README – Healthcare Knowledge Base for RAG (Retrieval-Augmented Generation)

This folder contains healthcare-related books, guides, and code for embedding these resources into a vector database (Pinecone) to support a Retrieval-Augmented Generation (RAG) system for medical recommendations.

## 📚 Purpose of the Knowledge Base

The goal is to build a domain-specific knowledge base (DSKB) to improve the quality of answers and recommendations provided by the large language model (LLM). All content is medically oriented and curated for reliability.

## 📘 Domain-Specific Knowledge Base (DSKB) Curation

• Curated a health-focused knowledge base using:  
 - Books on wellness and preventive care  
 - Clinical guidelines  
 - Lifestyle and nutrition guides  
 - WHO reports and peer-reviewed health articles  
• Purpose: To provide a reliable foundation for generating medically relevant advice

## ✂️ Text Chunking for Efficient Retrieval

• Entire books are broken into smaller chunks (typically 100–300 tokens) to:  
 - Enhance the quality of semantic search  
 - Ensure chunk size stays within token limits for LLM input  
 - Target recommendations more precisely and efficiently

## 🧠 Embedding Generation

• Text chunks are embedded using HuggingFace Sentence Transformers (e.g., all-MiniLM-L6-v2)  
• Why Embeddings?  
 ✓ Semantic similarity enables the retrieval of the most relevant text chunks  
 ✓ Enhances precision of responses when users ask health-related questions

## 📦 Storing in Pinecone Vector Database

• All embeddings are stored in Pinecone, a high-speed, scalable vector database  
• Pinecone performs KNN (k-nearest neighbor) searches to identify semantically relevant content  
• Importance:  
 ✓ Enables rapid and accurate contextual grounding for the LLM  
 ✓ Ensures personalized and context-aware recommendations

## 📝 Included Code File

• making\_pinecone\_index\_for\_vectore\_database\_storage.ipynb –  
 Contains the code to:  
 - Load books from the data folder  
 - Chunk the text, generate embeddings  
 - Create and populate a Pinecone index

## 🔗 External Books Reference

Some larger books are hosted externally and can be accessed from:  
https://openstax.org/subjects/nursing

## 📂 Folder Contents

• /data – Folder containing embedded healthcare-related books  
• .ipynb file – Code to build Pinecone index and process embeddings  
• This README – Overview and context for the folder