AI Skill-Up Assignment – JINSON

(Healthcare RAG – Data & Indexing Lead)

ASSIGNMENT A: Knowledge Base & Schema Prep

• Collect 5–10 clinical notes, 1 drug manual PDF, 1 guideline.

• Extract, clean, and standardize text.

• Build structured clinical-note schema.

• Create healthcare_data_dictionary.md.

• Create schema_overview_diagram.png.

ASSIGNMENT B: Embedding Store & Retrieval Experiments

• Build document store with:

- Fixed chunks (500–800 tokens)

- Semantic chunks (per section/topic)

• Generate embeddings and index.

• Implement retrieval_test.py returning top-5 chunks + similarity.

• Compare chunking strategies in retrieval_comparison_jinson.md.

Deliverables:

healthcare_data_dictionary.md

schema_overview_diagram.png

rag_index_builder.py

healthcare_vector_store.py

retrieval_test.py

retrieval_comparison_jinson.md