

פרויקט גמר – Metadata-Driven Hybrid RAG

מטרה כללית

פיתוח Agent RAG לענף הביטוח שמספק תשובות מתוך דוחות אירוע (תאונה/פריצה) הכוללים טקסט, טבלאות ותרשימים. המערכת מבוססת אחזור היברידי (Dense + Sparse) עם מטא-דאטה, ריראנקר, ותקציב צ'אנקים $\approx 5\%$ מגודל המסמך או עד 10 צ'אנקים.

מבנה העבודה

1. Parsing & Chunking

חלוקה לפי מבנה טבעי (סעיפים, כותרות, טבלאות, נספחים).

דיסטילציה לצ'אנקים חשובים בהתאם לתקציב הצ'אנקים.

שמירת Anchors: עמוד, TableId/FigureId, מיקום בשורה/עמודה.

2. Metadata

בחירה של 3–5 שדות לכל צ'אנק מתוך הרשימה:

PageNumber

ChunkSummary

Keywords

CriticalEntities

IncidentType

IncidentDate

SectionType

AmountRange

TableId / FigureId

3. Indexing

Embeddings לכל צ'אנק.

שמירת המטא-דאטה עם הווקטורים לפילטרים.

4. Hybrid Retrieval

ניתוח שאילתה:

חילוץ מילות מפתח, ישויות, תאריכים → יצירת Metadata Filters.

Dense Retrieval \cup Sparse Retrieval.

סינון לפי מטא-דאטה.

Reranker ($K=40 \rightarrow 6-8$).

5. Agent Tools

Summary – תמצות חלקים נבחרים.

Needle – איתור פסקה/סעיף ספציפי.

Table-QA – שאילתות נומריות/טבלאיות (DataFrame) + החזרת (table, row) (Anchor).

Policy-Check – התאמת סעיפי פוליסה לרכיבי האירוע.

6. Evaluation

Answer Correctness – Ground Truth לא context-sensitive (מדמה חלוקת צ'אנקים).

Context Precision ≥ 0.75 .

Context Recall ≥ 0.70 .

Faithfulness ≥ 0.85 .

Table-QA Accuracy ≥ 0.90 .

MCP Integration (3 נק').

Graph DB Integration (8 נק').

משימות המשך מתוכננות

בניית סט דאטה לדוגמה (מסמכים אמיתיים עם אנונימיזציה).

יצירת סט שאלות הערכה (כולל שאלות טקסטואליות, טבלאיות, וגרף).

הפקת Ground Truth בהתאם לחוקי הצ'אנקים.

הכנת סט בדיקות RAGAS מלא והרצה על המערכת.