

1 בדיקת טבלה: השוואה RAG לעומת Long Context LLMs

טבלה 1: השוואה: RAG לעומת Long Context LLMs

מאפיין	Retrieval-Augmented Generation (RAG)	Long Context LLMs (LC-LLMs)
מקור ידע	לא-פרמטרי, חיצוני, נתונים דינמיים (מסד נתונים וקטורי/גרף).	פרמטרי (משקולות מאומנות) + חלון הקשר גדול (נתוני In-Context).
סקיילביליות	גבוהה במיוחד (טריליוני אסימונים); סקיילביליות בלתי תלויה בגודל המודל.	מוגבלת על ידי מורכבות ה-Attention; מוגבלת למקסימום חלון ההקשר הנוכחי.
עלות ויעילות	חסכוני מאוד באמצעות אחזור סלקטיבי; משתמש במשאבים ביעילות.	דרישות חישוביות גבוהות לעיבוד חלון ההקשר הגדול כולו בכל שאילתה.
התאמה לנתונים	מצטיין בנתונים מקוטעים, דינמיים, מיוחדים או דלילים (למשל, מאגרי קוד, דיאלוג).	מתאים יותר להקשרים צפופים, מובנים היטב ועקביים (למשל, ספרים, מסמכים מובנים).

הערות בדיקה:

- הטבלה משתמשת ב-m{width} לכל העמודות (יישור אנכי מרכזי)
- עמודות כתובות בסדר RTL (ימין לשמאל)
- כל תא משתמש ב-\hebcell{} - עם \en{} למונחים באנגלית
- אין שימוש ב-\textenglish{} - או \texthebrew{} (אסור!)
- ראשי עמודות: \hebheader{} לעברית, \enheader{} לאנגלית