🚀המאמר היומי של מייק 30.09.24: ⚡️🚀  
SCHRODINGER’S MEMORY: LARGE LANGUAGE MODELS

ביום הסוער הזה (למרות שהסקירה שייכת פורמלית לאתמול - אשלים את הפער בימים הקרובים) נסקור מאמר די קליל עם שם מאוד לא קליל. כי אין דבר קליל שכולל בתוכו את שמו של שרדינגר - ספק אם הצלחתי להבין בצורה טובה מספיק את המשוואה של שרדינגר עוד בקורס פיזיקה 3 באוניברסיטה במוסקבה לפני עשרות שנים. גם סיפורו של חתול שרדינגר לא התבהר עד עכשיו.

אוקיי, סיימנו עם הצחוקים. המאמר חוקר (אמפירית) נושא די רציני והוא הזכרון של מודלי שפה. כשאנחנו שואלים LLM מה עיר הבירה של שבדיה, איך הוא יודע שזה סטוקהולם. המאמר טוען כי זיכרון LLM פועל על ידי התאמה דינמית של פלטים לקלטים. כלומר המודל ״בוחר״ איך לשלוף את המידע מהזיכרון ובונה אותו על סמך הקלט.

המחברים מסבירים את איך פועל הזיכרון של מודלי שפה באמצעות ניתוח של ארכיטקטורת הטרנספורמרים. מנגנון ה-attention (כלומר מקדמי ה-attention שלו) למעשה מאפשרים למודל לבנות את הפלט כפונקציה דינמית של הקלט (כלומר לא קבועה כמו ב-MLP או ConvNets).

המחברים משתמשים ב- Universal Approximation Theorem או UAT כדי להסביר את היכולת של שליפת מידע שנלמד במהלך האימון על בסיס תוכן של הקלט. המחברים טוענים כי ניתן להבין מנגנון זה בתור ״יכולת קירוב דינמית בסגנון UAT" (המשפט המקורי מדבר על יכולת קירוב סטטית של מודלי ML) כאשר המודל מתאים תוצאה מתאימה על בסיס הקלט, והתופעה הנצפית ניתן להגדיר בתור זיכרון.

הם מכנים זאת "זיכרון שרדינגר" מכיוון שאנו יכולים לקבוע של-LLMs יש את הזיכרון הזה רק על ידי ״שאילת שאלות״ וניתוח התגובה שלו; אחרת, הזיכרון נשאר בלתי מוגדר. בנוסף במאמר נדונים גורמים המשפיעים על ביצועי LLM: גודל המודל, איכות/כמות הדאטה והארכיטקטורה. המחברים טוענים שהזיכרון של מודלים באותו הגודל מושפע מאופן האימון שלהם ואם המודל אומן על יותר דאטה איכותי אז הוא משתפר (אין הפתעות כאן).

ולבסוף נעשות הקבלות בין ארכיטקטורת LLM למבנה המודולרי של המוח האנושי (את זה פחות אהבתי אבל זרמתי).

https://arxiv.org/pdf/2409.10482