המאמר היומי של מייק - 03.01.25:  
A PERCOLATION MODEL OF EMERGENCE: ANALYZING TRANSFORMERS TRAINED ON A FORMAL LANGUAGE

מבוא:  
רשתות נוירונים מודרניות, במיוחד מודלי שפה גדולים , מציגות מגוון רחב של יכולות, המאפשרות להן לשמש כמערכות בסיס למגוון יישומים. מאמר זה מציע הגדרה פנומנולוגית של אמרגנטיות בהקשר של רשתות נוירונים, תוך התמקדות באופן שבו מבנים ותהליכים ספציפיים המונחים בבסיס תהליך יצירת דאטה יכולים להוביל לשיפורים פתאומיים בביצועים במשימות ממוקדות יותר.

מושג חשוב:  
הפנומנולוגיה היא גישה פילוסופית המתמקדת בחקר מבני התודעה(consciousness) כפי שהם נחווים מנקודת המבט של האדם. היא שואפת לתאר תופעות או הופעת הדברים כפי שהן נתפסות על ידי בני אדם, ללא הנחות מוקדמות או הטיות תיאורטיות. שיטה זו מדגישה את הבנת החוויות כפי שהן נחיות, במטרה לחשוף את המשמעויות הטבועות בהן

יכולות אמרגנטיות(emergent capabilities) ברשתות נוירונים:  
החוקרים מגדירים אמרגנטיות ברשתות נוירונים כרכישת מבנים ספציפיים הגורמים לצמיחה פתאומית בביצועים במשימות ספציפיות. הם חוקרים זאת אמפירית באמצעות מערכת ניסויית המבוססת על שפה פורמלית תלוית-הקשר, ומדגימים שטרנספורמרים שאומנו על מחרוזות משפה זו מציגים יכולות אמרגנטיות. ברגע שהמודל לומד את הדקדוק והמבנים הבסיסיים, הביצועים במשימות קשורות משתפרים משמעותית.

הגדרת השפה הפורמלית:  
המערכת הניסויית שהוצעה במאמר משתמשת בדקדוק חופשי-הקשר הסתברותי (PCFG) להגדרת שפה פורמלית תלוית-הקשר. הדקדוק כולל:  
  
סימבולים סופיים(terminal symbols): חלקי דיבור הכוללים נושאים, מושאים, פעלים, תארים, פועלים, מילות חיבור ומילות יחס.  
סימבולים לא-סופיים: סמלים המגדירים את מבנה המשפטים.  
חוקי יצירת טקסט: חוקים המכתיבים כיצד ניתן לשלב סמלים סופיים ולא-סופיים ליצירת משפטים תקפים.

המודל מאומן על משימות כמו יצירה חופשית, פתרון בלבול וייצור מותנה, כאשר מדדי הביצועים נעקבים לאורך תהליך האימון.

משימות ופרוטוקולי הערכת ביצועי מודלים:  
  
1. יצירה חופשית של טקסט: המודל מייצר משפטים העומדים בחוקים הדקדוקיים.  
2. תיקון טקסט לא תקין: המודל מסדר מחדש מחרוזת מבולבלת של מילים ליצירת משפטים תקפים.  
3. יצירה מותנית: המודל יוצר משפטים על בסיס ישויות או תכונות נתונות.

ההערכה מתבצעת לפי המדדים כוללים בדיקות דקדוקיות, בדיקות טיפוס, דיוק התאמה מדויקת, דיוק פר-טוקן ועוד, המספקים הערכה מקיפה של יכולות המודל.

תוצאות: דינמיקת הלמידה

התוצאות מגלות 3 שלבים מובחנים בדינמיקת הלמידה של המודל:  
  
1. שלב ראשוני: המודל לומד מבנים דקדוקיים בסיסיים עם שיפור מינימלי בביצועים.  
2. ״שינוי פאזה״: מתרחשת עלייה פתאומית בביצועים ברגע שהמודל מתחיל ״להבין את אילוצי שפה״ פשוטים יחסית.  
3. שלב ההכללה: המודל מדגים ביצועים משופרים במשימות, המעידים על מעבר משינון להכללה.

יכולות אמרגנטיות של מודלים:

החוקרים מבחינים שככל שמודל השפה לומד את הדקדוק ואילוצי הטיפוס, נצפים שיפורי ביצועים משמעותיים במגוון משימות, במיוחד בפתרון בלבול וייצור מותנה. הנוכחות של מבנים ספציפיים מאפשרת למודל לבנות ״שילובים מורכבים ותקינים״ של ישויות ותכונות, המובילים ליכולות אמרגנטיות בתחום השפה.

נקודת מעבר בלמידה:

המאמר דן באופן שבו הופעת יכולות האמרגנטיות קשורה למספר התכונות התיאוריות שהמודל למד. נקודת המעבר, שבה מתרחשים שיפורי ביצועים משמעותיים, קשורה לסקיילינג של תכונות תיאוריות. קביעה זו מאפשרת לחזות מתי יכולות יופיעו ככל שהמודל ממשיך ללמוד.

מסקנה:

מחקר זה תורם להבנת האמרגנטיות ברשתות נוירונים על ידי יצירת מסגרת המגדירה ומאפיינת תכונות אמרגנטיות על בסיס רכישת מבנים בסיסיים על ידי המודל. הממצאים מצביעים על כך שאילוצים דקדוקיים ואילוצי שפה אחרים משמשים כגורמים חשובים בחיזוי התפתחות יכולות במודלים של שפה.

https://arxiv.org/abs/2408.12578