המאמר היומי של מייק: 26.05.25  
Neuro-Symbolic AI i 2024: A Systematic Review

המאמר הוא סינתזה מדויקת ועדכנית של ההתפתחות המהירה של תחום הבינה המלאכותית הנוירו-סימבולית (Neuro Symbolic AI) ב-5 השנים האחרונות. מתוך אוסף של 158 עבודות שעברו ביקורת עמיתים(peer review), החוקרים מציעים מיפוי שיטתי של התחום המרתק הזה, תוך הבחנה מדויקת בין מוקדי מחקר מפותחים לבין תחומים מפותחים פחות אך כאלו שעתידם קריטי לפיתוח מערכות בינה מלאכותית (AI) אמינות ואוטונומיות באמת.

63% מהעבודות שנבחנו על ידי המחברים עסקו בלמידה והסקה, מה שמצביע על נטייה ברורה של הקהילה המחקרית לשלב למידה סטטיסטית (ככה אנו מאמנים מודלי למידת מכונה היום) עם אילוצים לוגיים. עבודות בולטות כוללות רשתות נוירונים לוגיות, שימוש בפריורים סימבוליים בלמידת few-shot, והכנסת משמעות סמנטית לפונקציות לוס. מטרת גישות אלו היא לצמצם את הצורך בדאטה, להגביר את יכולת ההכללה של המודלים, ולשלב היסק דדוקטיבית עם היסק סטטיסטי(מבוסס על הדפוסים) .

תחום ייצוג הידע(knowledge representation) מהווה 44% מהעבודות, ומתעלה מעבר לייצוגים מבוססי גרפים ידע פשוטים יחסית. המחקרים בתחום זה עוסקים בבניית ידע קומונסנסי ודינמי, באופטימיזציה של ייצוגים סמנטיים, ובשילוב בין ייצוגים המופקים על ידי רשתות נוירונים ללוגיים, כפי שנעשה ב-NeuroQL, שפה ייעודית שמדגימה איך ניתן "לעשות יותר עם פחות".

תחום הלוגיקה ההיסק (35%) כולל פרויקטים כמו DeepStochLog ו-Logical Credal Networks, המשלבים לוגיקה הסתברותית עם ייצוגים סמבוליים לצורך פתרון בעיות מורכבות. מיזוג זה הוא קריטי להתמודדות עם חוסר ודאות תכונה הכרחית במערכות הפועלות בעולם האמיתי שהוא מהווה סביבה מורכבת מאוד וגם partially observable.

למרות הצורך ההולך וגובר במערכות AI שקופות ובטוחות לשימוש, רק 28% מהעבודות עסקו ב-explainability ואמינות. הפער הזה איננו רק מספרי, הוא מעיד (לפי דעת המחברים) על סדר עדיפויות לא נכון בקהילה המדעית בתחום זה. רוב המאמצים בתחום זה עוסקים בהסברים פוסט-הוק, תיקונים סמנטיים או שיפור סיכום טקסטים. פרויקטים כגון Braid (שמשלב חוקים סטטיסטיים ולוגיים) ו-FactPEGASUS (שמדגיש עובדתיות) הם יוצאים מן הכלל, אך אינם מייצגים מגמה רחבה. מה שחסר הוא ראייה מערכתית: כיצד ניתן לבנות מודלים(או מערכות) שבהם תהליך קבלת ההחלטות הוא מובן ושקוף כבר משלב בניית ארכיטקטורה והאימון ולא רק בדיעבד.

התרומה המשמעותית ביותר של המאמר היא העלאת המטה-קוגניציה שלו מ״הערת שוליים״ בטקסונומיות קודמות לתת-תחום מוגדר ודחוף לפיתוח. רק 5% מהמאמרים שנסקרו נוגעים לנושא זה, ובכל זאת הרעיון של ניטור עצמי, שליטה אדפטיבית והיגיון פנימי הוא ללא ספק החלק החסר בארכיטקטורות הנוירו-סימבוליות הנוכחיות.

המאמר מאמץ הגדרה של מטה-קוגניציה כיכולת לווסת ולהרהר בתהליכים קוגניטיביים פנימיים. מבחינה מעשית, זה כולל בקרים סימבוליים הממוקמים על גבי סוכני RL, אינטגרציות של LLMs עם ארכיטקטורות קוגניטיביות (למשל, ACT-R, Soar), ומסגרות תיאורטיות המתואמות עם המודל המשותף של קוגניציה. עבודות אלה, אם כי מעטות, מצביעות לעבר עתיד שבו מערכות בינה מלאכותית אינן רק תגובתיות אלא מודעות לעצמן אסטרטגית - מסוגלות לנהל תשומת לב, בחירה בין אסטרטגיות מסקנות ותיקון עצמי בהקשרים לא מוכרים.

אחת ההבחנות המרכזיות היא מיעוט העבודות המשלבות את כל ארבעת התחומים המרכזיים: למידת פטרנים והיסק, לוגיקה הסבתרותית, ייצוג ידע, ו-explainability. הדוגמה היחידה הבולטת היא AlphaGeometry שהיא מערכת שפותחה בגוגל לפתרון בעיות גיאומטריות מאולימפיאדות למתמטיקה. המערכת יוצרת מיליוני משפטים והוכחות סינתטיות, בעזרת LLMs שמדריך מנוע היסק סימבולי. זהו מודל מרשים של שילוב עומק ידע, יכולת כללית, וניתנות לבקרה. עם זאת, AlphaGeometry די חריגה בנוף. שאר העבודות נוטות להיות מבודדות בתוך תחום אחד או שניים. במיוחד מורגש חוסר השילוב בין explainability לתחומים אחרים, מה שמצביע על צורך במחקר בין-תחומי אמיתי כדי לממש את החזון של NSAI.

ברמה המתודולוגית, אחת מחוזקות המאמר היא הסינון המרשים של אוסף המאמרים שנבחרו: מתוך 1,428 עבודות, רק 158 עברו סינון איכות שכלל ביקורת עמיתים, רלוונטיות, וזמינות של קוד פתוח. פרט חשוב זה מהווה לא רק מדד טכני אלא גם הצהרת רצינות. הוא משדר שהקהילה צריכה לשאוף לשחזוריות(reproducibility), שקיפות, ונגישות, במיוחד בתחום כמו NSAI שבו העיצוב הארכיטקטוני מורכב ביותר.

https://arxiv.org/abs/2501.05435