המאמר היומי של מייק: 01.06.25  
Common Sense Is All You Need

דיסקליימר: זו סקירה של מאמר דעה ולאו דווקא מייצג את עמדת הסוקר

בשנים האחרונות, AI רשמה התקדמות מרשימה: מצ'אטבוטים ויצירת תמונות ועד למכוניות אוטונומיות. אבל עדיין קיים פער יסודי שמפריד בין מכונות לבין האינטליגנציה האינטואיטיבית והגמישות שמאפיינת אפילו את בעלי החיים הפשוטים ביותר: שכל ישר. במאמרו הוגו לטאפי טוען שהיכולות החסרות האלו אינן סתם חיסרון טכני אלא המכשול המרכזי שמונע מה-AI להגיע לאוטונומיה אמיתית. לפי גישתו, אם ברצוננו לראות מערכות AI שפועלות בבטחה ובאופן עצמאי בסביבות דינמיות ובלתי צפויות, עלינו להטמיע בהן מראש הבנה הקשרית, הסתגלותית ואינטואיטיבית כלומר, שכל ישר.

הטענה המרכזית במאמר היא שהמאמצים בתחום ממוקדים יותר מדי בשיפור scale וביצועים על פני בנייה של הבנה אמיתית. לטאפי מציע להפסיק ללטש מודלים שיכולים לחזות היטב טקסטים או לזהות עצמים, ולהתחיל לבנות מערכות שיכולות להבין הקשרים, ללמוד תוך כדי תנועה ולהתמודד עם מצבים לא מוכרים. לא עוד תוספת שכבות נוירונים או מאגרי דאטה עצומים — אלא שינוי כיוון יסודי.

אחת מהתרומות החדשניות ביותר במאמר היא ההרחבה של מושג ה"התגלמות" (embodiment) ב-AI. לרוב, התגלמות מתייחסת למערכות פיזיות כמו רובוטים שלומדים דרך אינטראקציה עם העולם הפיזי: הליכה, אחיזת עצמים, ניווט. לטאפי מציע הרחבה: גם בסביבות מופשטות כמו חידות לוגיות, AI צריך לפעול מתוך אינטראקציה עם מבנה המשימה ולא רק לזהות דפוסים סטטיסטיים. הוא מכנה זאת התגלמות קוגניטיבית.

נקודה חשובה נוספת היא הדרישה להתחיל ממצב של ידע מינימלי, או Tabula Rasa. רוב המודלים כיום מתבססים על טריליוני מילים או תמונות וזה מה שמאפשר להם ביצועים טובים בסיטואציות שכיחות אך כישלון בסביבות חדשות. לטאפי טוען כי מערכות אוטונומיות באמת צריכות להתחיל כמעט מאפס כלומר בניית ידע מתוך אינטראקציה עם הסביבה. כך נוכל למנוע תלות בנתוני אימון ולחזק את הגמישות הקוגניטיבית.

אחת הביקורות החדות ביותר שלו היא מה שהוא מכנה "השלב שבו הקסם קורה". הרבה מפתחים מניחים שהבינה תופיע מעצמה ברגע שנוסיף עוד שכבות נוירונים או דאטה. אבל בלי שכל ישר, AI פשוט יתקרב לתקרת זכוכית ישפר ביצועים קצת בכל פעם, אבל לעולם לא יגיע להבנה אמיתית של מצבים מורכבים.

בעיות במדדים ובאמות מידה קיימות

לטאפי מקדיש חלק משמעותי לניתוח מגבלות של אמות המידה הפופולריות, דרך שלושה מקרים בולטים. הראשון הוא אתגר ARC: סט של חידות המיועדות לבחון הפשטה והסקה. המודל נדרש להבין את החוקיות של מספר זוגות של קלט-פלט ולחזות את הפלט הנכון במקרה חדש. בפועל, מודלים מתאמנים על מאות דוגמאות ולפעמים גם על שאלות המבחן עצמן. כך הם מצליחים לא באמצעות הבנה אלא בזכות זיכרון. לטאפי מציע לעצב אתגר חדש שבו המערכת חייבת להתחיל מידע מינימלי בלבד כדי לבדוק הסקת מסקנות אמיתית.

הדוגמה השנייה היא נהיגה אוטונומית לפי רמות SAE: מדירוג 1 (סיוע בסיסי) ועד לרמה 5 (אוטונומיה מלאה בכל מצב). רוב המערכות כיום תקועות ברמות 2–3, ורמה 4 דורשת לעיתים התערבות אנושית מרחוק. לדעת לטאפי, זו לא בעיה טכנית אלא בעיית שכל ישר: רכב צריך להבין מחוות אנושיות, אירועים חריגים, או סימנים לא סטנדרטיים. בלי הבנה אינטואיטיבית, לא ניתן להגיע לאוטונומיה אמיתית.

לבסוף, הוא בוחן את מבחן טיורינג, שמודד אם מכונה יכולה לקיים שיחה כמו אדם. זו נקודת ציון חשובה, אך לטאפי מציין שמודל יכול לעבור את המבחן בלי להבין דבר. די ביכולת לייצר תגובות סבירות סטטיסטית. זה אולי מרשים בטקסט, אך חסר ערך בסיטואציות שבהן נדרשת פעולה, הסקה או שיקול דעת.

לטאפי לא מתעלם מההישגים של שיטות ההגדלה (scaling). הן אכן שיפרו ביצועים בתחומים כמו שפה וראייה. אבל הוא מציין סימנים ברורים של קיפאון: גם אם מוסיפים דאטה וחישוב, הביצועים על מדדים מסוימים הפסיקו להשתפר. לדוגמה:

מודלי זיהוי עצמים כמו COCO לא משתפרים מעבר לנקודה מסוימת.

זיהוי חריגות בוידאו (UCF-Crime) נעצר סביב אותה רמת דיוק.

איתור פעולות בווידאו (ActivityNet) תקוע במשך זמן רב.

המשמעות היא שהשקעה בהגדלה בלבד נותנת החזר הולך ופוחת. כדי להמשיך להתקדם, עלינו לטפל בבעיה היסודית: איך לגרום למערכות להבין את ההקשר שבו הן פועלות.

לטאפי גם נוגע בכמה בעיות תיאורטיות מוכרות בתחום, ומראה כיצד שכל ישר עשוי לפתור אותן. למשל, משפט No Free Lunch קובע שאין אלגוריתם שמתאים לכל בעיה. תגובתו: נכון לכן נבנה מערכות שפועלות היטב בתחומים מסודרים ורלוונטיים, במקום לנסות לפתור הכול. הוא גם דן בבעיית המסגרת (Frame Problem): הקושי לקבוע מה רלוונטי בכל מצב. לטאפי מציע שהשתתפות פעילה בעולם (פיזי או מופשט) עוזרת ל-AI ללמוד מה חשוב ומה ניתן להתעלם ממנו. באותו אופן, בעיית ההכשרה (Qualification Problem) כלומר הקושי לקבוע מראש את כל התנאים הנדרשים לפעולה, נפתרת אם המערכת לומדת מתוך ניסיון והתאמה.

לטאפי מציע שורת שינויים מעשיים: ראשית, יש לעצב מבחנים חדשים שמודדים הבנה תהליכית, לא רק תוצאה. שנית, צריך להגביל את הידע המוקדם של המערכת, כך שתהיה חייבת ללמוד ולהסיק באופן עצמאי. יש גם לבחון את הדרך שבה המודל חושב, לא רק אם התשובה שלו נכונה. מבחינה ארכיטקטונית, הוא תומך בשילוב בין שיטות סמליות (לוגיקה וחוקים) לבין למידת מכונה. גישה היברידית שכזו תשלב את הגמישות של רשתות נוירונים עם השקיפות של לוגיקה פורמלית. בנוסף, הוא מציע לשאוב השראה ממדעי המוח והקוגניציה ולא להעתיק ביולוגיה, אלא ללמוד ממנה איך מערכות לומדות בפועל.

ומה יקרה אם לא נשנה כיוון?

המאמר מזהיר שמי שימשיך לבנות מודלים גדולים יותר, מבלי להתמקד בשכל ישר, יסבול לא רק מהאטה אלא גם מאובדן אמון. מערכות שמתפקדות היטב במעבדה אך נכשלים בשטח יובילו לאכזבה של משתמשים, משקיעים ומפתחים. גרוע מכך, מערכות שמקבלות החלטות לבד אך חסרות הבנה של הקשר או ערכים אנושיים עלולות לפעול בצורה מזיקה. לטאפי טוען שהפחד הציבורי מ-AI ובמיוחד מ-AI שמשתפר מעצמו נובע לא מהאינטליגנציה אלא מהיעדר שכל ישר. מערכות שמתפתחות מבלי להבין את ההשלכות, ההקשרים או הציפיות המוסריות יוצרות סיכון אמיתי. לכן, הבטחת שכל ישר במערכות כאלה היא לא רק יעד טכני אלא תנאי לבטיחות.

https://arxiv.org/pdf/2501.06642