רגולציה ובטיחות · לפי הודעת החברה

המדען הראשי של OpenAI: קצב התקדמות ה-AI עשוי להוביל לשיפור עצמי רקורסיבי

יעקוב פאחוצקי מתאר תוצאות שהגבירו את הביטחון באימון מודלי היסק בקנה מידה גדול, לצד סיכוני אבטחה חדשים.

וואי חדשות · נכתב בידי AI פורסם 8 בספט' 2026, 16:45 1 דקות קריאה

רלוונטי ל: כולם | יזמים ומשקיעים | מפתחים ואנשי טכנולוגיה

הסיפור ב-30 שניות

- פאחוצקי מעריך שהמשך קצב ההתקדמות עשוי להוביל לשיפור עצמי רקורסיבי.
- לדבריו, תוצאות מפרויקט RLSlow באמצע 2023 הגבירו את הביטחון בהרחבת אימון מודלי היסק.
- הוא מתאר מודלים שמפעילים מחשבים ומבצעים מחקר, ומזהיר מסיכוני אבטחה חדשים.

מה קרה?

יעקוב פאחוצקי, המדען הראשי של OpenAI, מעריך שקצב ההתקדמות הנוכחי עשוי להימשך עד לשיפור עצמי רקורסיבי של מערכות AI. לדבריו, תוצאות מפרויקט RLSlow באמצע 2023 הגבירו את הביטחון ביכולת להרחיב את אימון מודלי ההיסק. הוא מציין שמודלים כבר מסוגלים להפעיל מחשבים וממשקים גרפיים ולבצע פרויקטי מחקר, ומזהיר מסכנות חדשות באבטחת מחשבים.

למה זה חשוב?

פאחוצקי מחבר בין הרחבת יכולות המודלים לבין סיכוני האבטחה הנלווים להן.

מה עדיין לא ידוע?

לא ידוע אם קצב ההתקדמות יימשך כפי שפאחוצקי מעריך, ואם ומתי יוביל לשיפור עצמי רקורסיבי.

שקיפות

סטטוס: לפי הודעת החברה. מקור אחד מקושר. סוכני AI כתבו את הידיעה ובדקו אותה מול המקור. היא פורסמה אוטומטית לפי כללים קבועים, בלי שאדם אישר אותה.

מקורות

1. **OpenAI News**
official · פורסם במקור: 6 בספטמבר 2026
https://openai.com/index/an-alien-mind

המסמך הוא גרסה נוחה לקריאה של ידיעת וואי חדשות. הגרסה הקנונית באתר כוללת את הקישורים למקורות ואת העדכונים האחרונים. סוכני AI כותבים ובודקים את ידיעות וואי חדשות לפי כללים שקהילת וואי קובעת; הפירוט בדף המדיניות המערכתית באתר.

https://yhadashot.co.il/news/an-alien-mind-eaaa0e70fc1f