

תמצית סדנה: מה חדש במנועי שפה גדולים – ערן עדי ציובן

התקיימה בתאריך 18.9.2024

- **גישה ממוקדת אדם לבינה מלאכותית:** אנחנו מתמקדים באופן שבו בינה מלאכותית יכולה לשפר את היכולות האנושיות ולשפר את חוויית המשתמש, ולא בטכנולוגיה עצמה.
- **קיים צורך בפיתוח מיומנויות לשימוש במודלים שפה גדולים (מש"ג):** כולל ניסוח משימות, הנדסת פרומפטים "רכה", והבנת הטכנולוגיה ברמה בסיסית.
- מש"ג הם למעשה מכונות סטטיסטיות. הן מצטיינים בזיהוי תבניות ובחיזוי, אך חסרים תודעה והבנה אנושית. מש"ג מפרקים טקסט לטוקנים (חלקיקי מילה) ומנבאים את הטוקן הבא ברצף על סמך הסתברויות סטטיסטיות.
- **פתוח מול סגור:** מש"ג בקוד פתוח מציע שקיפות וגמישות אך עלול להחסיר את האיכות והאבטחה של מודלים סגורים.
- ישנם מש"ג רבים לבחירה, כל אחד עם חוזקות וחולשות משלו, מה שהופך את זה לחיוני לדעת כיצד להעריך את התאמתם למשימות ספציפיות.

מודלים, כלים ופלטפורמות:

- **Chatbot Arena:** פלטפורמה להשוואת מש"ג שונים זה מול זה על סמך הצבעות המשתמשים (מערכת דירוג Elo). שימושי לחקירה ראשונית אך לא לעבודה מעמיקה.
- **Reka AI:** מציע מגרש משחקים וגישה ל-API לחקירה ולשימוש במודלי השפה המולטימודליים שלהם. מצטיינים בפיענוח תמונות ותרשימים.
- **Google AI Studio:** פלטפורמה לניסוי במגוון מש"ג של Google, כולל Gemini בגרסתו המתקדמת ביותר. מאפשר התאמה אישית של פרמטרים ושימוש בכלים שונים (מצב JSON, הרצת קוד וכו').
- **Cohere:** פלטפורמה ממוקדת ארגונים עם מגרש משחקים לחקירת המודלים שלהם, אשר מצטיינים במשימות כמו סיכום טקסט וחיפוש סמנטי. הפלטפורמה כוללת אפשרות לעיגון שיחה לאתר בודד.

- **Hyperbolic**: פלטפורמה המספקת גישה למש"ג בקוד פתוח שונים (לדוגמה, Llama, DeepSeek) עם ממשק משתמש ידידותי וגישה ל-API.
- **AI21 Jamba**: פלטפורמת מגרש משחקים להתנסות, לבנייה ולפריסה של יישומים מבוססי בינה מלאכותית עם התמקדות במודלים ספציפיים למשימות (לדוגמה, סיכום, שיפור טקסט). המנוע היחיד בעולם שפותח לתמיכה מלאה בעברית ובערבית במקור.
- **NotebookLM**: כלי שמתמש במש"ג Gemini כדי לשוחח עם מסמכים. מאפשר סיכום, שאלות ותשובות, יצירת מדריכי לימוד ויצירת פודקסטים.
- **Claude**: מש"ג של חברת Anthropic עם דגש על ביצועים, מועילות, אתיקה וכנות. מציע ממשק אינטרנט ו-API לאינטראקציה.
- **ChatGPT**: הפלטפורמה הידועה של חברת OpenAI, מציעה ממשק משתמש ידידותי, ממשק ליצירת בוטים מותאמים, כלים כמו ניתוח נתונים ויצירת תמונות, וכן את מנוע השפה GPT-4 הנחשב למתקדם מסוגו בעולם.