כלי AI · לפי הודעת החברה

Together AI משיקה Polymorphic Inference: גישת הסקה חדשה שמסתגלת לשינויים

המעבר מ-Static Inference מיועד להתמודד עם קצב שחרור המודלים המואץ ומגוון הארכיטקטורות בסביבות ייצור

וואי חדשות · נכתב בידי AI פורסם 4 בנוב' 2026, 16:47 2 דקות קריאה

רלוונטי ל: מפתחים ואנשי טכנולוגיה | יזמים ומשקיעים | עסקים ועצמאים

הסיפור ב-30 שניות

- Together AI פרסמה ב-4 בנובמבר 2026 גישת הסקה חדשה בשם Polymorphic Inference, שפירושה ביוונית הסקה ב'צורות רבות'.
- הגישה מתוכננת להסתגל באופן דינמי לשינויים במודלים, בתעבורה ובחומרה.
- Together AI פיתחה כלי נלווה בשם Together Compiler, המשתמש בסוכני AI לאופטימיזציה מתמשכת של הסקה בייצור.
- המעבר מגישת Static Inference מוצג כתגובה לגיוון ההולך וגדל של מודלי קוד פתוח ולמורכבות הפריסה שלהם.
- צוות המחקר של Together AI עומד מאחורי פיתוחים קודמים כגון FlashAttention, ThunderKittens ו-ThunderAgent.

מה קרה?

Together AI הכריזה ב-4 בנובמבר 2026 על גישת הסקה חדשה בשם Polymorphic Inference. לפי החברה, הגישה מתוכננת להסתגל לשינויים במודלים, בתעבורה ובחומרה באופן דינמי – בניגוד לגישת ה-Static Inference שבה נקטה עד כה. Together AI מגדירה את המעבר כתגובה לגיוון ההולך וגדל של מודלי קוד פתוח, לקצב שחרורם המהיר, למגוון הארכיטקטורות ולעומסי העבודה המשתנים בסביבות ייצור. במסגרת הגישה החדשה פיתחה החברה כלי בשם Together Compiler, המשתמש בסוכני AI לאופטימיזציה מתמשכת של הסקה בייצור. צוות המחקר של Together AI הוא גם זה שעומד מאחורי FlashAttention, ThunderKittens ו-ThunderAgent.

למה זה חשוב?

ריבוי מודלי קוד פתוח וסביבות חומרה מגוונות הפכו את ניהול ההסקה בייצור למורכב. גישה שמסתגלת לשינויים אלו באופן דינמי עשויה להפחית את העומס על צוותי התשתית המתחזקים מערכות AI בסביבות פעילות.

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. **Together AI — בלוג**

official · פורסם במקור: 4 בנובמבר 2026

https://www.together.ai/blog/polymorphic-inference

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/polymorphic-inference-self-tuning-inference-for-open-models-60f38369fd6a