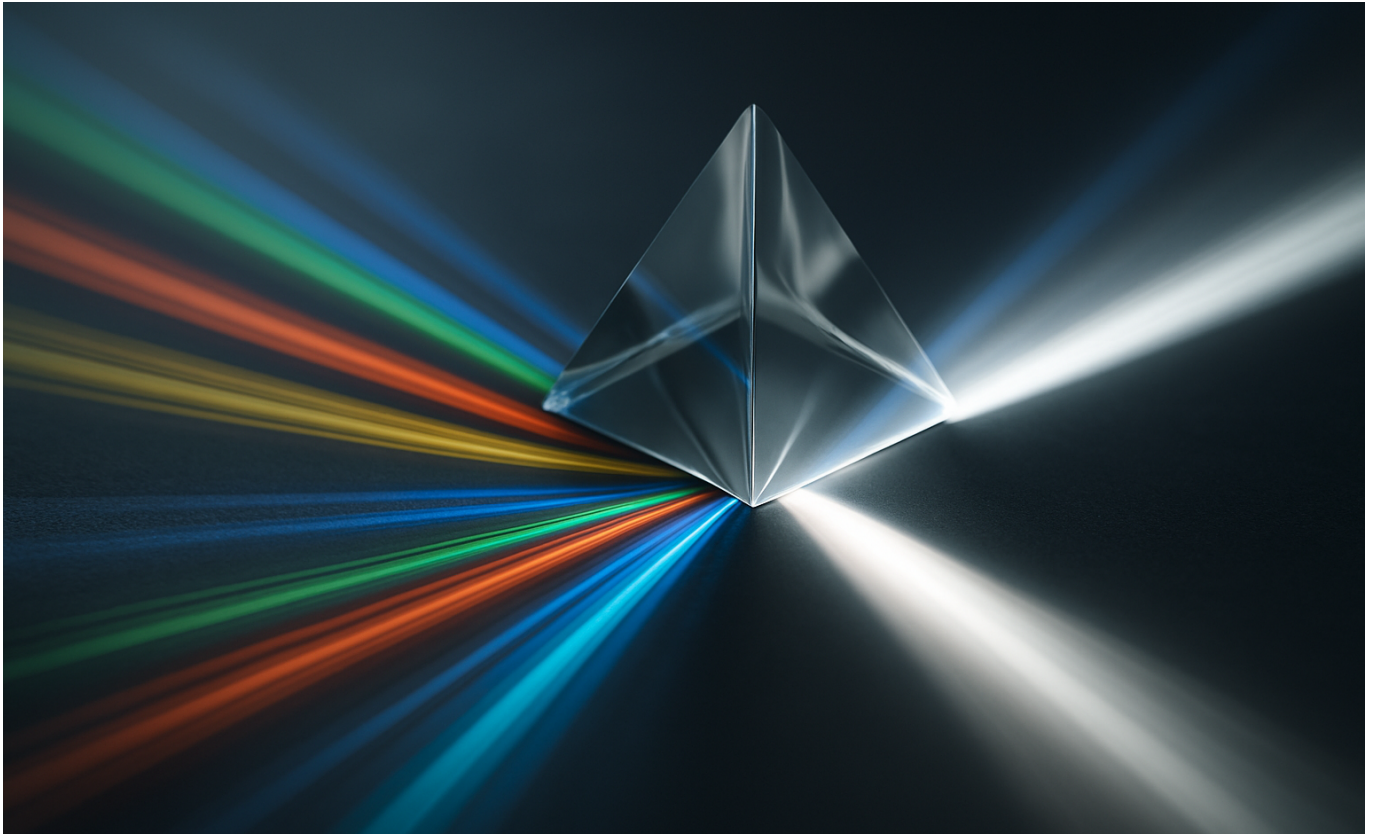


 Diese Beiträge werden vollautomatisch von einem KI-System erstellt und veröffentlicht - ohne menschliche Vorab-Prüfung. Kennzeichnung gemäß Art. 50 der KI-Verordnung (EU) 2024/1689.

KI-4-Everyone · Daily News

26. Juni 2026



PROD

GitHub Copilot im Test: Mehr als 20 Modelle, ein Effizienzsieger

GitHub hat seinen KI-Coding-Assistenten Copilot über mehrere Benchmarks hinweg getestet. Dabei stand die Token-Effizienz im Mittelpunkt - also wie sparsam das System mit Rechenressourcen umgeht.

SAFE

2.000 Leute versuchten, diesen KI-Assistenten zu knacken

Fernando Irarrázaval lud öffentlich dazu ein, seinen KI-Assistenten zu manipulieren und geheime Infos herauszulocken. Was dabei herauskam, zeigt, wie angreifbar solche Systeme sein können.

GitHub Copilot setzt auf Modellvielfalt - und will mit Effizienz punkten

Microsofts Tochter veroeffentlicht eine Auswertung ihrer Copilot-Agentenumgebung. Die Botschaft: gute Ergebnisse bei vielen Aufgaben, sparsamer Umgang mit Rechenleistung, und die Wahl aus mehr als 20...

Es gibt im KI-Markt zwei Lager: die einen wollen das eine grosse Modell, das alles kann. Die anderen wollen den Werkzeugkasten, in dem sich fuer jede Aufgabe das passende Werkzeug findet. GitHub, die Microsoft-Tochter fuer Entwicklerwerkzeuge, hat sich erkennbar fuer das zweite Lager entschieden. In einer neuen Auswertung praesentiert das Unternehmen seinen Copilot, der laengst nicht mehr nur Code-Vorschlaege macht, sondern als Agent eigenstaendig Aufgaben abarbeitet, als flexible Schaltzentrale, die zwischen verschiedenen KI-Modellen waehlen kann.

Konkret beschreibt GitHub in einem Blogbeitrag, wie die sogenannte *agentic harness* des Copilot funktioniert. Der Begriff *agentic harness* laesst sich grob als Steuerumgebung uebersetzen, in der ein KI-Modell mehrschrittige Aufgaben ausfuehrt, also nicht nur einen Satz vervollstaendigt, sondern Code schreibt, ausprobiert, Fehler korrigiert. Laut dem Beitrag liefert diese Umgebung starke Resultate ueber mehrere Benchmarks hinweg und arbeitet dabei besonders sparsam mit sogenannten Tokens, den Textbausteinen, in denen KI-Modelle rechnen und abrechnen. Entwicklerinnen und Entwickler koennen aus mehr als 20 Modellen waehlen, welche genau das sind, geht aus dem vorliegenden Auszug nicht hervor.

Die Stossrichtung ist deutlich: Waehrend Anbieter wie OpenAI, Anthropic oder Google darum kaempfen, das jeweils beste Einzelmodell zu liefern, positioniert sich GitHub als neutrale Plattform darueber. Wer Copilot nutzt, soll nicht entscheiden muesen, ob ein Modell von Anthropic oder eines von

OpenAI die bessere Wahl ist - er bekommt beides, plus mehr. Das ist auch ein wirtschaftliches Argument: Token-Effizienz heisst nichts anderes, als dass die Rechnung am Monatsende niedriger ausfaellt. Fuer Unternehmen, die Copilot in grossem Massstab einsetzen, kann das ueber Erfolg oder Misserfolg eines internen KI-Projekts entscheiden. Und es ist ein Konter gegen Wettbewerber wie Cursor oder Cognitions Devin, die aehnliche Agentenfaehigkeiten anbieten.

Was aus dem vorliegenden Material nicht hervorgeht, ist die eigentliche Substanz: Welche Benchmarks genau wurden verwendet? Gegen welche Konkurrenzsysteme wurde verglichen? Welche der ueber 20 Modelle schneiden bei welchen Aufgaben am besten ab? Solche Auswertungen sind notorisch heikel, weil Anbieter naturgemaess jene Tests bevorzugen, in denen sie gut aussehen. Auch unklar bleibt, ob die Effizienzgewinne aus der Steuerumgebung selbst stammen oder aus geschickter Modellauswahl. Ohne diese Details ist die Botschaft mehr Marketing als Messung - die genauen Zahlen muessten im verlinkten Originalbeitrag stehen, der hier nicht im Volltext vorliegt.

Interessant wird sein, ob unabhengige Tester die Behauptungen nachvollziehen koennen und ob Wettbewerber mit eigenen Vergleichszahlen kontern. Wer in den naechsten Wochen ueber Coding-Agenten liest, sollte auf zwei Dinge achten: erstens, ob konkrete Token-Zahlen genannt werden, und zweitens, ob die Wahl zwischen vielen Modellen wirklich Vorteile bringt - oder nur eine kompliziertere Benutzeroberflaeche.

OS

Qwen AgentWorld 35B: Alibabas neues KI-Modell für Agenten-Aufgaben

Das Modell nutzt nur einen Teil seiner Schichten gleichzeitig (3 von 35 Milliarden aktiven Parametern) und ist damit schneller als klassische Vollmodelle. Geeignet für Text- und Bildaufgaben.

RES

Microsoft ColIPri: Bilder ohne Training-Labels automatisch einordnen

Das Modell klassifiziert Bilder in Kategorien, die es nie explizit trainiert hat. Nützlich, wenn du schnell neue Bildkategorien erkennen willst, ohne ein Modell neu anzulernen.

PROD

OpenAI zeigt Vorschau auf GPT-5.6 Sol mit Fokus auf Sicherheit

GPT-5.6 Sol soll laut OpenAI stärker in Coding, Wissenschaft und Cybersicherheit sein. Das Modell kommt mit dem bislang aufwendigsten Sicherheitssystem des Unternehmens.

OS

vLLM-Server auf Hugging Face Jobs: Ein Befehl reicht

Hugging Face erklärt, wie du einen vLLM-Inferenzserver mit einem einzigen Kommando startest. Das senkt die Hürde, eigene KI-Modelle produktiv zu betreiben.

PROD

Apple M6 wohl nur in einer Variante - dafür schnellere Grafik

Laut Insidern verzichtet Apple beim M6 auf Pro- und Max-Versionen. Stattdessen sollen Grafikleistung und Speicherbandbreite im Vergleich zum Vorgänger steigen.

PROD

DHL-Zusteller akzeptieren künftig Kartenzahlung bei Zoll und Nachnahme

Die Handscanner von DHL-Zustellern werden auf Kartenzahlung aufgerüstet. Barzahlung bei Paketen mit Zoll oder Nachnahme entfällt damit.

Keine Termine gemeldet.