

KI-4-Everyone · Daily News

18. Mai 2026



OS

PaddleOCR 3.5 liest Dokumente jetzt mit Transformers-Backend

PaddleOCR 3.5 bringt Texterkennung und Dokumentenanalyse auf eine neue Basis. Das Tool läuft nun mit einem Transformers-Backend.

RES

Neues Leaderboard vergleicht KI-Agenten offen miteinander

Das Open Agent Leaderboard stellt KI-Agenten auf den Prüfstand. So lässt sich nachvollziehen, welche Systeme in der Praxis besser abschneiden.

PaddleOCR 3.5 zieht in die Transformers-Bibliothek ein

Ein chinesisches OCR-Werkzeug wird in das wichtigste Open-Source-Oekosystem fuer KI-Modelle integriert - was das fuer Entwickler und Anwender bedeutet.

Textererkennung in Dokumenten klingt nach einem geloesten Problem - bis man eine handgeschriebene Notiz, eine alte Rechnung oder eine Tabelle aus einem PDF in saubere Daten umwandeln will. Genau hier setzt PaddleOCR an, ein Open-Source-Werkzeug aus dem Umfeld des chinesischen Suchmaschinenkonzerns Baidu. Mit der Version 3.5 zieht das Werkzeug nun in die populaerste Modellbibliothek der KI-Welt ein. Das ist weniger eine Sensation als ein Indiz: Auch im weniger glamouroesen Feld der Dokumentenverarbeitung verschiebt sich das Geschehen Richtung Open Source und gemeinsame Standards.

Konkret beschreibt ein Blogbeitrag auf der Plattform Hugging Face (eine Art GitHub fuer KI-Modelle), wie sich PaddleOCR 3.5 ueber das sogenannte Transformers-Backend nutzen laesst. Transformers ist die zentrale Python-Bibliothek von Hugging Face, mit der Entwickler KI-Modelle laden und ausfuehren. Bisher brauchte PaddleOCR ein eigenes Framework namens PaddlePaddle. Mit der neuen Anbindung koennen Entwickler die OCR-Funktionen (Optical Character Recognition, also die automatische Texterkennung in Bildern) und das Document Parsing - das strukturierte Auslesen von Layouts, Tabellen und Formeln aus Dokumenten - in derselben Umgebung nutzen, in der sie auch Sprachmodelle wie Llama oder Mistral betreiben. Der Beitrag wurde laut Metadaten am 18. Mai 2026 veroeffentlicht.

Relevant ist das aus zwei Gruenden. Erstens entsteht damit ein gemeinsamer Werkzeugkasten: Wer ein KI-System bauen will, das Vertraege liest, Rechnungen sortiert oder Behoerdenformulare aus-

wertet, kann OCR und Sprachverstaendnis kuenftig im selben Code-Workflow kombinieren, ohne zwischen Frameworks zu wechseln. Zweitens zeigt der Schritt, wie chinesische KI-Projekte ihre Reichweite ausbauen, indem sie sich an westlich gepraeigte Open-Source-Standards andocken. Fuer Anwender in Unternehmen senkt das die Einstiegshuerde, weil Dokumentenanalyse damit naeher an die Standard-Toolchains der KI-Entwicklung rueckt - und nicht mehr in einem Spezialsilo lebt.

Vieles bleibt aus dem Material allerdings offen. Welche Sprachen PaddleOCR 3.5 in welcher Qualitaet erkennt, wie es sich gegen kommerzielle Anbieter oder gegen multimodale Sprachmodelle schlaegt, die Dokumente direkt als Bild verstehen, geht aus dem vorliegenden Hinweis nicht hervor. Auch zur Lizenz, zu Hardwareanforderungen oder zum Umgang mit sensiblen Dokumenten - Stichwort Datenschutz bei Personalakten oder Patientenberichten - liegt im Material nichts vor. Wer das Werkzeug produktiv einsetzen will, sollte diese Punkte vor einer Entscheidung selbst pruefen; pauschale Empfehlungen waeren an dieser Stelle Spekulation.

In den naechsten Wochen lohnt sich ein Blick darauf, ob weitere spezialisierte Werkzeuge - etwa fuer Sprache, Video oder wissenschaftliche Formeln - denselben Weg in das Transformers-Oekosystem gehen. Falls ja, verfestigt sich ein Trend: Die Bibliothek wird vom Sprachmodell-Hub zur Universal-Schnittstelle fuer praktisch jede Form von KI-gestuetzter Datenverarbeitung. Fuer Unternehmen, die heute noch ueberlegen, auf welches Framework sie setzen, koennte das die Entscheidung erleichtern.

PROD

Googles Gemma 4 (26B) versteht Bilder und Text gleichzeitig

Das Modell gemma-4-26B-A4B-it nutzt nur 4 von 26 Milliarden Parametern gleichzeitig – so antwortet es schneller. Es verarbeitet Bilder und Text gemeinsam.

PROD

GitHub Copilot: KI-Codiersessions vom Smartphone aus steuern

Wer eine Copilot-Session am PC oder in der CLI startet, kann sie jetzt per GitHub Mobile oder github.com weiterführen. Die Funktion ist ab sofort allgemein verfügbar.

PROD

OpenAI und Dell bringen Codex in Firmen-Rechenzentren

Unternehmen können den KI-Coding-Agenten Codex künftig auf eigener Hardware oder in hybriden Umgebungen betreiben. Das soll mehr Datenkontrolle ermöglichen.

PROD

Alexa Plus erstellt KI-Podcasts zu beliebigen Themen

Du gibst Alexa Plus ein Thema vor, und die KI generiert daraus einen Podcast mit eigenen Hosts. Vor der Aufnahme zeigt dir die App, worüber gesprochen werden soll.

PROD

Apples neue Siri löscht Chats automatisch und schützt Gesprächsdaten

Mit iOS 27 soll Siri Dialoge selbstständig löschen. Die Memory-Funktion bleibt eingeschränkt – Apple positioniert sich damit als datenschutzfreundlichere KI-Option.

PROD

Apples Siri-Umbau: Datenschutz als zentrales Thema bestätigt

Automatisch löschende Chats sollen ein Kernmerkmal der überarbeiteten Siri werden. Details will Apple demnächst vorstellen.

Keine Termine gemeldet.