TAGESSPIEGEL
BACKGROUND

KÜNSTLICHE INTELLIGENZ (KI)

EU-Parlament: Forscher kritisiert fehlende Prüfung von Claude-Einsatz

2024 integrierte das EU-Parlament das KI-Modell Claude in sein öffentliches Archiv. Der KI-Experte Kris Shrishak bemängelt fehlende Risikobewertungen des Modells und blindes Vertrauen in den Anbieter.

von Alexandra Ketterer

veröffentlicht am 04.04.2025

Im Oktober 2024 hatte das **Europäische Parlament** ein KI-Modell in ihr öffentlich zugängliches Archiv integriert *(https://archidash.europarl.europa.eu/ep-archives-anonymous-dashboard)*, um Bürger:innen und Forschenden den Zugang zu Wissen und historischen Dokumente zugänglicher zu machen. Das Tool mit dem Namen **„Ask the EP Archives"** wurde mit dem Large Language Model (LLM) Claude von Anthropic in Amazon Bedrock entwickelt. Die Besucher:innen können so mit den archivierten Dokumenten des EU-Parlaments „chatten". Der Forscher **Kris Shrishak**, Senior Fellow bei Enforce, einer Abteilung des Irish Council for Civil Liberties, kritisiert, dass das Parlament den Einsatz der KI **nicht ausreichend geprüft** *(https://www.iccl.ie/press-release/how-not-to-deploy-generative-ai-the-story-of-the-european-parliament/)* habe.

„Es gab keine ausreichende Datenschutz- und Risikobewertungen", sagt **Shrishak**: „Die Kommission hat eine Risikobewertung für die Cloud-Umgebung vorgenommen, aber nicht für das KI-Modell." Das Parlament verlasse sich alleine auf die Behauptungen von Anthropic, so sein Urteil.

Shrishak hat Dokumente analysiert, die protokollieren, wie die Testphase und Entscheidung für das KI-Modell ablief. Das Parlament hat dem KI-Modell zum Beispiel 30 Fragen gestellt, die das Modell teilweise falsch beantwortete. Auf die Frage, wer der erste Präsident der Kommission war, antwortete das Modell etwa mit der Ausgabe „Robert Schuman 7“, der Adresse eines Cafés in Brüssel.

Einen Hinweis, dass das Model **fehlerhafte Ausgaben** generieren kann, sehen Nutzer:innen nicht direkt auf der Archive-Webseite, sondern erst, wenn sie auf einen „Disclaimer“-Link klicken. Dort steht: Die generierten Inhalte sollte nicht als verlässlich angesehen werden. Nach der Ansicht von **Shrishak** müsse dieser Hinweis prominenter platziert werden: „Aktuell geben Sie die Verantwortung an die Öffentlichkeit ab.“ Die Nutzer:innen müssen also selbst nachprüfen, ob die Informationen stimmen.

Unter dem Disclaimer-Link findet sich auch die Information, dass das Modell die Antworten „nur basierend auf ausgewählten, öffentlich zugänglichen Dokumenten des Parlaments“ generiert. Ein Fehlschluss, kritisiert der KI-Experte: „Bei der Bewertung muss berücksichtigt werden, wie das **KI-Modell** entwickelt wurde. Das Parlament glaubt blind, dass Anthropics KI-Modell gut ist.“ Das KI-Modell von Anthropic, Claude, stützt sich auf den Ansatz der konstitutionellen KI. Dabei orientiert sich das KI-Modell an einer Reihe von vordefinierten Prinzipien, um den Lernprozess zu steuern. Nach der Angabe von Anthropic berücksichtigt das Modell Claude, die Prinzipien „hilfreich, ehrlich und harmlos“. Bislang hat es allerdings **keine unabhängige Bewertung** gegeben, die belegt, dass das Produkt von Anthropic seinem Anspruch gerecht wird.

Das Parlament verlasse sich auf die Aussagen von Anthropic, wenn es um Voreingenommenheit oder Zuverlässigkeit des Modells geht, kritisiert Shrishak. „Sie betrachten nur die öffentlichen Dokumente des Parlaments als Datenquelle und nicht die Trainingsdatensatz des Modells, der auch auf den im **Internet gesammelten Daten** aller Menschen auf der Welt beruht.“

„Die Claude-Modelle sind speziell für die **Wahrung der Privatsphäre** geschult“, heißt es in den Prüfungsdokumenten *(https://www.iccl.ie/wp-content/uploads/2025/03/20250324_EP-Archibot-TAIAL.pdf)*des EU-Parlaments. Doch Anthropic gibt an, für das Training ihrer Modelle auch Daten aus dem Internet zu verwenden, personenbezogene Daten könnten darin enthalten sein, schreibt der AI-Anbieter in einem Blogpost *(https://privacy.anthropic.com/en/articles/10023555-how-do-you-use-personal-data-in-model-training)*.

Das Parlament hat nach der Beobachtung von Shrishak nur einige Modelle getestet. Und zwar: Claude, Sonnet 3.0, Sonnet 3.5 und Titan. Letzteres ist ein KI-Modell von **Amazon**, die anderen sind von Anthropic. Das Tool ist in das Cloud-Ökosystem von Amazon eingebunden und stützt sich auf Amazon Bedrock – Amazons Marktplatz für KI-Modelle.

Der **Leiter des Archivreferats** des Parlaments sagte in einem Werbevideo für Anthropic: „Wenn wir generative KI einsetzen, müssen wir ständig die Kontrolle über die Lösung haben, die wir entwickelt haben.“ Shrishak hebt in seinem Bericht hervor, dass nicht das Parlament die Kontrolle über die Lösungen hat, sondern Amazon: „Das EU-Parlament verlässt sich im Wesentlichen auf das **Amazon-Ökosystem**. Und da kann man keine Kontrolle ausüben.“

Tagesspiegel Background hat das Europäische Parlament um eine **Stellungnahme** zu den Vorwürfen gebeten, bis Redaktionsschluss aber keine Antwort erhalten.