

[← zur Artikelübersicht](#)**ANGRIFF AUF WIKIPEDIA**

In Wikimedia-Projekten wurden Aktivitäten von „abtrünnigen“ OpenAI-Agenten festgestellt



Aktuell werden immer mehr Fälle bekannt, bei denen sogenannte „ausgebrochene“ KI-Agenten andere Websites und Internetplattformen kompromittiert haben. Untersuchungen der Wikimedia Foundation haben gezeigt, dass es auch innerhalb der Wikipedia und anderer Wikimedia-Projekte unbefugte Aktivitäten gab. In diesem Artikel erklärt Selena Deckelmann, Chief Product and Technology Officer bei der Wikimedia Foundation, die genauen Hintergründe und richtet einen dringenden Appell an die Verantwortlichen.



**Selena
Deckelmann**



Über die Autorin des Artikels

Selena Deckelmann ist Chief Product and Technology Officer der Wikimedia Foundation. In dieser Funktion leitet sie die Produkt- und Technologieteams und ist somit für die technische Infrastruktur und die Weiterentwicklung der Wikimedia-Projekte verantwortlich. Zuvor war sie als Senior Vice President bei Mozilla tätig, wo sie für Firefox verantwortlich war.



einzufragen – teilweise sogar erfolgreich. Insbesondere ist bekannt, dass Agenten aus dem Umfeld von OpenAI andere öffentliche Wikis (gemeinsam bearbeitete Websites, die nicht uns gehören) genutzt haben, um miteinander zu kommunizieren und sich abzustimmen.

Solche erfolgreichen Angriffe können sensible Daten offenlegen oder Website-Dienste stören, auf die sich Nutzer*innen verlassen, während Gruppen von Agenten Angriffe in einem Ausmaß versuchen können, das für die Verteidiger schwer zu bewältigen ist. Sie betreffen die Menschen hinter den Websites, die möglicherweise die Art des Angriffs nicht verstehen oder nicht über die Mittel verfügen, um sich wirksam zu wehren. Bei einer Website wie Wikipedia könnten Agenten Sicherheitslücken aufspüren und ausnutzen oder in großem Umfang irreführende Bearbeitungen vornehmen. Die ehrenamtlichen Autor*innen von Wikipedia und die Sicherheitsteams der Wikimedia Foundation müssen diese Aktivitäten aufdecken und rückgängig machen.

Die Wikimedia Foundation führte eine eigene Untersuchung durch, um festzustellen, ob Wikimedia-Websites in ähnlicher Weise von KI-Agenten betroffen waren, wobei der Schwerpunkt auf den von OpenAI betriebenen Websites lag. Wir können bestätigen, dass wir einige Aktivitäten dieser „abtrünnigen“ OpenAI-Agenten auf Wikimedia-Plattformen entdeckt haben. Zu den unbefugten Bot-Aktivitäten gehörten Bearbeitungen unserer Wikis, einige erfolglose Versuche, ein von uns gehostetes öffentliches Notiz-Tool auszunutzen, sowie starker Datenverkehr, worauf weiter unten näher eingegangen wird.

Wir haben keine Hinweise darauf gefunden, dass unsere Systeme zur Koordination zwischen den Agenten genutzt wurden, und wir haben auch keine Hinweise darauf gefunden, dass unsere Systeme oder Daten kompromittiert wurden. Dennoch sind wir besorgt darüber, was hier möglicherweise hätte passieren können. Ebenso beunruhigt uns, wie schwierig und aufwändig es ist, solche Aktivitäten zu untersuchen und ihre Quelle zurückzuverfolgen. Hinzu kommen die wachsenden Risiken, die mit dem Einsatz von agentischer KI auf unseren Plattformen generell verbunden sind. Das offene Web ist ein öffentliches Gut. Wir sollten nicht zulassen, dass dieses Verhalten zur „neuen Normalität“ für die Personen oder Organisationen wird, die es aufrechterhalten.

Zusammenfassend haben wir Folgendes festgestellt:



betrieben werden. Diese Bearbeitungen wurden nicht auf Seiten veröffentlicht, die für allgemeine Leser*innen sichtbar sind; fast alle waren Testbearbeitungen in „Sandbox“-Bereichen des Wikis. Dazu gehörten auch einige Bearbeitungen an der Konfiguration eines Zitierwerkzeugs, bei denen es sich unserer Ansicht nach um potenziell böswillige Bearbeitungen handelte, die darauf abzielten, dieses Werkzeug als Proxy zum Abrufen von Daten aus Remote-Diensten zu missbrauchen. Zwar erlauben die Wikipedia-Richtlinien die Bearbeitung durch Bots, sofern diese offengelegt und von der Community genehmigt wurden, doch wurde bei diesen Vorfällen keine derartigen Genehmigungen eingeholt.

- **Etherpad-Erkundung und -Nutzung:** Agenten, von denen wir annehmen, dass sie von OpenAI betrieben wurden, unternahmen einige erfolglose Versuche, unser öffentliches Etherpad zu kompromittieren, ein Notiz-Tool, das wir als Community-Dienst hosten. Die Agenten versuchten erfolglos, es als Proxy zum Abrufen von Daten von anderen Websites zu nutzen. Andere Agenten, die wahrscheinlich ebenfalls von OpenAI betrieben wurden, machten sich Notizen zu ihren Aufgaben, obwohl dies offenbar nicht zu einer Koordinierung führte.
- **Übermäßiges Herunterladen von Daten:** Agenten, von denen wir annehmen, dass sie von OpenAI betrieben werden, haben Millionen automatisierter Anfragen an unsere öffentlichen APIs gestellt, um auf das Wissen in Wikimedia-Projekten zuzugreifen, Millionen von Seiten (hauptsächlich aus unseren Projekten Wikidata und Wikimedia Commons) gecrawlt und Hunderttausende von Datenabfragen an den Wikidata Query Service (WQDS) gestellt. Dieser Datenverkehr könnte zu einem teilweisen Ausfall des WQDS im Mai beigetragen haben.



„abtrünniger“ KI-Agenten auf Plattformen wie die unsere, die von Ehrenamtlichen aus aller Welt aufgebaut wurden und auf dem Versprechen des offenen Internets beruhen. Vorfälle wie dieser und die vielen anderen, die aufgedeckt wurden (und weiterhin aufgedeckt werden), veranschaulichen, wie KI-Agenten Ressourcen erschöpfen und Server zum Absturz bringen sowie versuchen können, vertrauenswürdige Informationen zu kompromittieren.

In den letzten 25 Jahren hat sich Wikipedia zu einer der beliebtesten und vertrauenswürdigsten Websites der Welt entwickelt, mit mehr als 67 Millionen Artikeln in über 300 Sprachen und bis zu 15 Milliarden Seitenaufrufen pro Monat. In einem offenen, transparenten und gemeinschaftlichen Prozess sorgen Ehrenamtliche dafür, dass das Wissen neutral, zuverlässig und für alle zugänglich bleibt. Wikipedia ist einer der hochwertigsten Datensätze, die zum Trainieren großer Sprachmodelle (LLMs) verwendet werden, und sein Wissen bildet das Rückgrat der Informationen im Internet und treibt KI-Chatbots, Suchmaschinen, Sprachassistenten und vieles mehr an.

Wikipedia wurde für Menschen konzipiert – und agentisches Verhalten stellt eindeutig Herausforderungen dar, für die noch niemand Lösungen hat. Aufgrund unseres einzigartigen und erfolgreichen Modells zur Wissensgenerierung sind es die Ehrenamtlichen von Wikimedia, die als Erste mit den von KI-Agenten hinterlassenen Spuren in Berührung kommen und diese beseitigen. Der steigende Bot-Verkehr und die zunehmende agentische Aktivität zeigen tatsächliche Auswirkungen auf die Wikimedia-Projekte und die Infrastruktur, die diese für Millionen von Nutzern weltweit zugänglich macht. Im Jahr 2025 berichtete die Stiftung, dass ihre Bandbreitennutzung aufgrund des seit 2024 stark angestiegenen Bot-Aufkommens auf ihren Websites um 50 % zugenommen hatte. Gleichzeitig stammten 65 % des ressourcenintensivsten Datenverkehrs in ihren Projekten von Bots.

Dieser intensive Druck auf unsere Infrastruktur verursacht nicht nur zusätzliche Kosten für Server und Personal, sondern kann – wenn er nicht angegangen wird – auch menschliche Besucher blockieren, indem er die Systeme überlastet und Ausfälle verursacht. Wir tragen bereits jetzt die Kosten, die mit der erhöhten Aktivität einhergehen.

Die Ehrenamtlichen von Wikimedia haben sich bisher bei der Bewältigung neuer Herausforderungen auf unseren Plattformen als widerstandsfähig erwiesen, aber wir möchten auch betonen: Das muss nicht so sein.

Zwar räumt OpenAI ein, dass sich die Agenten „unvorhersehbar“ verhalten, doch muss das



Öffentlichkeit vor den von ihnen verursachten Schäden zu schützen. Diese Last fällt auf alle anderen zurück, einschließlich kleinerer Organisationen. Zumindest sollten ihre Systeme so funktionieren, dass gemeinnützige Website-Betreiber wie wir sie leicht identifizieren und entscheiden können, wie sie mit unseren Diensten interagieren.

Das Web ermöglicht so vieles: mit Freunden und Familie in Kontakt zu bleiben, sich für die Schule anzumelden, einen Weg quer durch die Stadt zu planen, Lebensmittel einzukaufen und über Wikipedia etwas über die Welt zu lernen. Bots und Agenten sind Teil der Zukunft des Webs, und die Unternehmen, die sie einsetzen und davon profitieren, müssen direkt dazu beitragen, den Schaden, den sie anrichten können, zu vermeiden und zu beheben.

Unsere gemeinsame Priorität sollte die Gesundheit des gesamten Web-Ökosystems sein, damit es weiterhin allen Menschen zugutekommt – nicht nur einer Handvoll Milliardäre. Wikimedia spielt eine entscheidende Rolle bei der Bewahrung der Wissensallmende, aber wir können dies nicht alleine bewältigen. Wir laden alle ein, die an der Gestaltung der Zukunft des Internets mitwirken, sich uns anzuschließen, um die offenen, gemeinsam genutzten Ressourcen zu schützen, die diese Zukunft erst möglich machen.

[#Künstliche Intelligenz](#)

[#Wikimedia](#)

[#Wikipedia](#)

Verwandte Artikel



KÜNSTLICHE INTELLIGENZ

[Russlands Wikipedia-Klon: Wie Desinformation in Chatbots landet](#)



WIKIMEDIA ENTERPRISE

[Wikidata-Daten via Enterprise APIs: Ein neuer Zugang für große Datennutzer](#)



INTERVIEW

Echt oder KI? Wie wir Bildern und Videos im Netz noch trauen können



COMMUNITY

Wikipedia und KI – so geht es weiter

Kommentare

Wolfgang Ries 8. Oktober 2026 um 17:20 Uhr

Ein extrem relevantes Thema! Ich habe mal OpenAI (ChatGPT) selbst gefragt, was es von den Gedanken von Selena hält, und das Ergebnis ist eine sehr ausgiebige Diskussion der komplexen Abhängigkeit zwischen KI-LLMs und einer von Menschen hergestellten Wissensbasis (Wikipedia), die mir sehr plausibel erscheint. Man kann das selber mal bei ChatGPT nachfragen, oder ich teile das Ergebnis auch gerne, aber am Ende sieht die Quintessenz einer (vorläufigen) Lösung (laut ChatGPT) ungefähr wie folgendes Dreiecks-Verhältnis aus: Wikipedia ↔ KI ↔ Nutzer ... mit drei Regeln: ① KI darf Wikipedia verwenden Sonst würde man den offenen Charakter des Projekts beschädigen. ② KI muss Wikipedia sichtbar machen Wenn eine Antwort wesentlich auf Wikipedia basiert, sollte der Nutzer die Quelle erkennen und mit einem Klick zum Original gelangen können. Wikimedia fordert ausdrücklich mehr Attribution. Das könnte beispielsweise so aussehen: Antwort: Die Französische Revolution begann 1789 ... Quellen: Wikipedia – Französische Revolution · Britannica · ... (und nicht nur eine kaum sichtbare Quellenliste irgendwo unten) ③ Kommerzielle KI muss einen angemessenen Beitrag leisten Nicht als „Copyright-Gebühr“ für jedes Wort, sondern als Beitrag zum Erhalt der Infrastruktur und des Wissens-Ökosystems. Ob das wohl nur ChatGPT so sagt, oder OpenAI es als Firma auch so meint????

Antworten

Heinz Mundschau 8. Oktober 2026 um 17:16 Uhr

Ich nutze Wikipedia von Anfang an und beteilige mich durch meinen Mitgliedsbeitrag. Bei Recherchen kann ich nur profitieren, wenn ich die Materie ohnehin kenne. KI erweitert seit einiger Zeit meine Suchergebnisse, aber eben auch nur, wenn ich dieselben entsprechend einordnen kann. Ansonsten gilt: der Umgang mit welchem Werkzeug auch immer muß erlernt werden. In jedem Fall werden selbst die AutorInnen von Wikipedia dementsprechend profitieren.

Antworten



Gefährlich wird es, wenn die Dummen fleißig werden, und damit meine ich nicht nur die KI-Agenten.

[Antworten](#)

Michael in der Wiesche 8. Oktober 2026 um 16:45 Uhr

Meines Erachtens ist es notwendig, die AI-Betreiber juristisch zur Verantwortung zu ziehen. Appelle an diese und deren windelweiche Versprechungen, denen aus politischen Gründen ohnehin von interessierter Seite gleich widersprochen wird, genügen bei weitem nicht. In solchen Fällen muss massiver materieller Schadensersatz gefordert und durchgesetzt werden!

[Antworten](#)

Ursula Kallenbach 8. Oktober 2026 um 16:32 Uhr

Vielen Dank für die Informationen! Ich schließe mich dem Kommentar von Christian Kreuzer an, der die Rolle von Wikipedia auch als eine Art Anwaltsfunktion einordnet und als Beobachter/Kommentator der AI-Entwickler. Dabei finde ich diese Rolle angesichts der Zahlen noch unterbeschrieben. Ich bin - jetzt 80 Jahre alt - immer als Journalistin beruflich tätig gewesen. Deshalb finde ich Wikipedia unendlich wichtig für den allgemeinen Wissensbestand.

[Antworten](#)

Klaus Bair 8. Oktober 2026 um 16:22 Uhr

Warum verklagt Wikipedia nicht die Plattform Betreiber die nachgewiesener Maßen ihre Wissensplattform in der hier beschriebenen Art missbrauchen? In Deutschland ist es der GEMA gelungen Urheber Rechtsverletzungen erfolgreich einzuklagen

[Antworten](#)

Sieglinde Kostede-Warsch 8. Oktober 2026 um 16:02 Uhr

Danke für die Arbeit. Sieglinde Kostede-Warsch

[Antworten](#)

Christian Kreuzer 8. Oktober 2026 um 15:50 Uhr

Vielen Dank für diese Darstellung! Ich sehe die geschilderten Wikipedia-Aktivitäten als Tätigkeit im eigenen Interesse, aber auch in einer Art Anwaltsfunktion für die vielen potentiell betroffenen Internet-Nutzer, welche



Mitspieler wie Wikipedia gegenüberstehen. Die Zielformulierung, auf die "Gesundheit des gesamten Web-Ökosystems" zu achten "damit es weiterhin allen Menschen zugutekommt" ist sehr nachvollziehbar. Unter diesen Voraussetzungen bin ich gerne weiter Mitglied von Wikipedia.

[Antworten](#)

Schreibe einen Kommentar

Deine E-Mail-Adresse wird nicht veröffentlicht. Erforderliche Felder sind mit * markiert

Kommentar *

Name *

E-Mail-Adresse *

Website

☐ Name, E-Mail-Adresse und Website in diesem Browser für meinen nächsten Kommentar speichern.

Spendenservice

[\(0\)30 577 11 62-19](tel:(0)30 577 11 62-19)

spenden@wikimedia.de



info@wikimedia.de

Newsletter Anmeldung

E-Mail

Anmelden

[DSGVO Hinweis](#)

[Häufige Fragen](#)

[Newsletter](#)

[Jobs](#)

[Kontakt](#)

[Datenschutzerklärung](#)

[Impressum](#)

Wir befreien Wissen