



Medienwandel durch KI

So bringen Publisher und Medienhäuser Transparenz in Bot- und Crawling-Aktivitäten

Medienwandel durch KI. So bringen Publisher und Medienhäuser Transparenz in Bot- und Crawling-Aktivitäten

Der Einsatz von Large-Language-Model-Bots (LLM-Bots) verändert die Geschäftsmodelle von Publishern grundlegend. Inhalte werden zunehmend gecrawlt, zusammengefasst und direkt in Suchmaschinen oder KI-Assistenten ausgespielt – oft ohne, dass Nutzer*innen die Originalquellen besuchen. Damit verschiebt sich das Verhältnis zwischen Content-Nutzung und tatsächlicher Reichweite sowie Monetarisierung spürbar zulasten der Publisher. Klassische Steuerungsmechanismen aus der Suchmaschinenwelt greifen dabei nur noch eingeschränkt oder gar nicht mehr.

Um auf diese Entwicklungen reagieren zu können, braucht es jetzt Transparenz über das tatsächliche Ausmaß der Bot-Aktivitäten sowie eine belastbare Datengrundlage für strategische Entscheidungen. So lassen sich wirksame Maßnahmen ableiten und priorisieren.

Dieses Whitepaper fokussiert sich daher auf Maßnahmen zur **Analyse und Transparenz** von Bot- und Crawling-Aktivitäten. Im Mittelpunkt stehen die Identifikation von Crawlerbots auf dem eigenen Angebot, der Aufbau eines systematischen Bot-Monitorings sowie die Auswertung zentraler Kennzahlen.

Ziel ist es, Publishern ein klares Bild darüber zu vermitteln, wie stark ihre Inhalte bereits von KI-Systemen genutzt werden und wo sich daraus konkrete Handlungsbedarfe ergeben. Die hier beschriebenen Ansätze bilden die Grundlage für weiterführende Maßnahmen zur Steuerung, Begrenzung oder Monetarisierung von Bot-Zugriffen, die in weiteren Publikationen vertieft werden.

1. Identifikation der Crawlerbots auf dem eigenen Angebot

Bevor Publisher Maßnahmen zur Steuerung oder Monetarisierung von Bot-Zugriffen ergreifen können, müssen sie verstehen, was heute tatsächlich auf ihren Webseiten passiert.

Transparenz entsteht dabei nicht durch Annahmen, sondern durch die systematische Auswertung von Traffic- und Crawling-Daten. Dieses Kapitel zeigt, wie Publisher die Nutzung ihrer Inhalte durch Bots und Crawler sichtbar machen und strukturiert analysieren können.

1.1. Aufbau eines Bot-Monitorings (Inhouse oder mit einem Dienstleister)

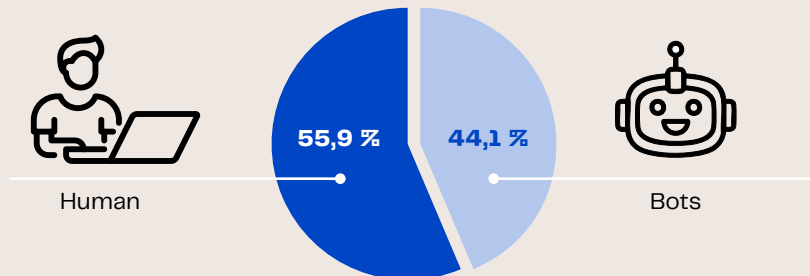
Um für die eigenen Webseiten einen Überblick darüber zu erhalten, welcher Anteil der Seitenaufrufe schon heute durch Bots entsteht und um welche Arten es sich dabei handelt, gilt es im ersten Schritt Daten zu erheben und auszuwerten. Dazu kann man Rohdaten (oder entsprechende Samples) selbst auswerten und darauf basierend Ableitungen treffen. Alternativ setzt man spezialisierte Dienstleister ein, die eine solche Auswertung und ggf. schon entsprechende Aufbereitung der Daten in visueller Form zur Verfügung stellen.

Dabei sollten zwei zentrale Aspekte betrachtet und bewertet werden:

Zum einen sollte ein Publisher darauf schauen, wie sein Traffic und die Seitenaufrufe sich insgesamt entwickeln. In diesem Kontext ist relevant, welcher Anteil der Aufrufe von echten Nutzer*innen (Human Traffic) und welche von Bots und Crawlern (Non-Human Traffic) verursacht werden. Laut Cloudflare¹ liegt der durchschnittliche Anteil an Bot-Traffic in Deutschland für Februar 2026 bei 44,1%.

Bots vs. Human

Bot (automated) vs. human, HTTP requests distribution



Andererseits muss der Non-Human Traffic differenziert betrachtet werden, da nicht alle Bot- und Crawler-Zugriffe für Publisher die gleiche Bedeutung haben. Eine zentrale Unterscheidung ist, ob Crawling-Aktivitäten mittelbar zur Reichweite und Monetarisierung beitragen oder Inhalte ausschließlich für externe Zwecke nutzen.

Klassische Suchmaschinen-Crawler, insbesondere von Google, greifen seit vielen Jahren auf Inhalte zu, um diese zu indexieren und in Suchergebnissen auszuliefern. Diese Zugriffe stehen in einem funktionalen Zusammenhang mit der Weiterleitung von Nutzer*innen auf die Publisher-Webseite und sind damit Teil eines etablierten Wertschöpfungsmodells.

Demgegenüber nutzen viele KI- und Trainings-Crawler Inhalte, um eigene Antworten, Modelle oder Dienste zu speisen, ohne dass daraus verlässlich Sichtbarkeit, Traffic oder Erlöse für Publisher entstehen. Für eine sachgerechte Bewertung des Non-Human Traffics ist diese Unterscheidung daher essenziell.

1.2. Einordnung und Übersicht von Bot-Analyse-Tools

Die Bot Directories² von Cloudflare, DataDome und Darkvisitors bieten u.a. eine Übersicht über verschiedene Bot-Kategorien.

Eine Anlaufstelle für diese Analysen kann aber auch der eigene Content Delivery Network (CDN) Anbieter sein. Ein CDN stellt die Infrastruktur zur Verfügung, wie auf der Webseite Inhalte bereitgestellt und anschließend dem einzelnen User beim Aufruf einer URL angezeigt werden. Große Anbieter wie Cloudflare, Akamai oder Amazons AWS CloudFront bieten bereits entsprechende Analysen und Tools an:

Anbieter	Produkt	Bot Analyse	Bot Blocking	Anbietersitz	URL
Akamai	Bot Manager	X	X	USA	https://www.akamai.com/products/bot-manager
Amazon Cloud Front	AWS WAF Bot Control	X	X	USA	https://aws.amazon.com/waf/features/bot-control/
Cloudflare	AI Insights / AI Audit / Bot Management	X	X	USA	https://radar.cloudflare.com/ai-insights https://radar.cloudflare.com/bots
Fastly	Bot Management / AI Bot Insights	X	X	USA	https://www.fastly.com/products/bot-management

Neben den großen CDN-Anbietern gibt es jedoch auch spezielle Dienstleister, die sich auf die Analyse von Webseiten Traffic und dem Traffic-Anteil speziell durch Scraping Bots spezialisiert haben. Darunter fallen unter anderem DataDome, Tollbit, Centinel Analytica, Darkvisitors und ContentBridge. Auch zu diesen Pure Playern haben wir eine Auswahl aufgelistet:

Anbieter	Produkt	Bot Analyse	Bot Blocking	Anbietersitz	URL
Botscorner	Bot Mapping	X	X	Frankreich	https://www.botscorner.fr
Centinel Analytica	Crawler Management	X	X	Deutschland	https://www.centinelanalytica.com/
Content Bridge	Bot Management	X	X	Deutschland	https://content-bridge.eu/
Darkvisitors	Agent Analytics	X	X	USA	https://darkvisitors.com/
DataDome	Bot Protect	X	X	Frankreich	https://datadome.co/products/bot-protection/
HUMAN	Agentic Trust	X	X	USA	https://www.humansecurity.com/applications/agentic-ai/
Tollbit	TollBit Analytics	X	X	USA	https://tollbit.com/analytics/

Ein Blick auf den jeweiligen Leistungsumfang der Dienstleister und die Integrationsmöglichkeiten lohnt sich. Hier können sowohl vom technischen Aufwand als auch aus Kostenperspektive deutliche Unterschiede entstehen. Vor dem Einsatz eines der Anbieter sollte auf jeden Fall die Konformität im Rahmen der DSGVO und TCF geprüft werden.

Wer die Rohdaten und Logfiles der Seitenaufrufe hat und weiß, wie man sie auswertet, kann die Analyse selbst machen. Ein Dienstleister ist dann nicht mehr nötig. Hilfreich zu wissen: Die meisten Bots können über ihren Useragent identifiziert und zugeordnet werden. Eine Ausnahme ist der Perplexity Crawler, welcher laut einer Cloudflare Analyse³ seine Identität verschleiert.

1.3. Erfassung der Crawling-Aktivitäten

Um den tatsächlichen Umfang und die Art der Crawling-Aktivitäten zu verstehen, sollten die Zugriffe auf die eigene Webseite regelmäßig analysiert werden, entweder durch Auswertung der Logfiles oder mit Hilfe spezialisierter Dienstleister (siehe Punkt 1.2.).

Entscheidend sind folgende Dimensionen für die Betrachtung:

- Crawl-Häufigkeit pro Bot (z. B. Googlebot, GPTBot, PerplexityBot, AnthropicBot)
- Crawler-Kategorie (z.B. Search Crawler, KI Training, KI Assistant, Ad Verification)
- Seitenbereiche, die überdurchschnittlich stark gecrawlt werden
- Zeitliche Muster mit erhöhter Bot-Aktivität
- IP- oder ASN-Zuordnungen, um verdeckte Bots zu erkennen

Kontinuierliches Monitoring hilft, Auffälligkeiten zu erkennen – etwa regelmäßige Voll-Crawls durch LLM-Anbieter oder plötzliche Aktivitätsspitzen, die auf Trainingsphasen hindeuten können.

Darüber hinaus kann es einen Unterschied machen, ob und wie eine Robots.txt Datei eingesetzt wird. Während klassische Suchmaschinen die dort definierten Regeln meist respektieren, zeigt sich bei vielen KI-Crawlern ein anderes Bild. Zum einen steigt die Anzahl an Crawlern von Woche zu Woche und zum anderen ignorieren sie immer öfter Sperrvermerke oder verschleiern ihre Identität, um weiterhin Zugriff auf Inhalte zu erhalten. Aktuelle Beispiele hierfür sind die IP oder User Agent-Rotation bzw. -Vortäuschung durch Perplexity. Eine regelmäßige Prüfung des tatsächlichen Verhaltens der Bots ist daher unbedingt empfehlenswert.

³ https://blog.cloudflare.com/perplexity-is-using-stealth-undeclared-crawlers-to-evade-website-no-crawl-directives/?utm_source=chatgpt.com/

**Empfehlung:**

Für Publisher kann es sinnvoll sein, eine interne Reporting-Struktur einzuführen, die das Crawl-Volumen monatlich dokumentiert und Veränderungen im Verhalten einzelner Bots sichtbar macht. Diese Daten bilden die Grundlage für anschließende Steuerungs- oder Blockierungsmaßnahmen.

2. Bot-Traffic: Auswertung der qualitativen Angebotsnutzung

Um die Nutzung der eigenen Inhalte durch KI-Systeme fundiert einordnen zu können, reicht die reine Erfassung von Crawling-Aktivitäten nicht aus. Der nächste wichtige Schritt ist die Analyse von AI-Referral-Traffic, Klickzahlen und der Crawl-to-Referral-Ratio, um sichtbar zu machen, wie viel Content von KI-Systemen genutzt wird und welcher Anteil tatsächlich zurück auf die Webseite gelangt. So lässt sich erkennen, welche Inhaltsbereiche besonders stark nachgefragt werden und wo die zunehmende Nutzung durch KI den klassischen Human Traffic strukturell beeinflusst.

2.1. Auswertung des AI-Referral Traffics

Ist das Monitoring aufgesetzt, geht es darum die Daten zu analysieren und Schlüsse zu ziehen.

Dass Nutzer*innen bei der Anzeige von KI-generierten Antworten in Suchmaschinen und Chatbots deutlich seltener auf Originalquellen klicken als in klassischen Google-Suchergebnissen, ist inzwischen durch Studien und Marktdaten belegt⁴. Dennoch macht eine Referral-Analyse zu KI-generierte Antworten Sinn, um zu verstehen, wie viel Traffic man über die KI-Systeme zurückerhält. Marktdaten zufolge liegt die durchschnittliche Klickrate auf Quellen innerhalb von KI-Antworten unter 1%.⁵

Wenn man sich den Referral Traffic auf seiner Webseite ansehen möchte, muss man wissen, dass sich Referrals aus dem AI Overview oder Google AI Mode aktuell nicht zuverlässig ausweisen lassen. Der Grund: Google weist den Referrer-String derzeit nicht gesondert aus. So sehen alle Nutzer*innen in der Web-Analyse gleich aus, wenn sie über Google kommen.

Für die Chat-Bot Anbieter sieht es etwas anders aus. Einige der Systeme, wie z.B. Microsoft Copilot oder Perplexity weisen einen eindeutigen Referrer oder einen URL-Parameter aus, wenn es zu einem Klick auf eine Quelle kommt. Es ist allerdings unklar, ob dies in allen Fällen geschieht, oder von Faktoren wie Webseiten- oder App-Nutzung abhängt.

Erste Analysen auf Webseiten von OVK-Publishern zeigen, dass der Anteil aktuell mit weit unter 1% noch sehr überschaubar ist, aber stetig wächst. Die Analysen haben auch gezeigt, dass AI-Referrals ein stärkeres Engagement auf den Seiten, verglichen zu Google Search Referrals, aufweisen. Aufgrund der geringen Fallzahl an identifizierbaren AI-Referrals, ist diese Auswertung aber noch mit Vorsicht zu genießen.

2.2. Crawl-to-Referral Ratio als Indikator

Vor dem Release der Google AI Overviews crawlte Google eine Webseite im Durchschnitt zwei Mal für einen Nutzer, der die Webseite über die Google Suche erreichte. Im November 2025 liegt dieses Verhältnis bereits bei 7:1. Der Trend steigt von Monat zu Monat. Noch extremer sieht es bei den KI-Chatbots aus. Das Rennen führt Anthropic an, die für einen Nutzerbesuch im Durchschnitt 27.700 Crawls vornehmen.

Die Schlussfolgerung aus diesem Trend ist eindeutig: Immer mehr Inhalte werden von Suchmaschinen und KI-Systemen intensiv genutzt, ohne dass ein entsprechender Human Traffic zu den Publishern fließt. Das Verhältnis zwischen Content-Nutzung und tatsächlicher Weiterleitung verschiebt sich. Zunehmend zulasten der Publisher. Je stärker die Crawl-to-Referral-Ratio steigt, desto klarer zeigt sich, dass KI-Systeme Inhalte für

⁴ <https://tollbit.com/state-of-the-bots/q3-q4-2025/>

⁵ <https://www.sistrix.de/news/pew-research-ai-overviews-halbieren-die-klickrate/>

eigene Antworten verwerten, während die ursprünglichen Quellen kaum noch Sichtbarkeit oder Wertschöpfung erhalten. Dieser Trend ist ein zentraler Frühindikator dafür, dass der klassische Such-Traffic strukturell unter Druck gerät und Maßnahmen zur Steuerung der Crawling-Aktivitäten an Bedeutung gewinnen.

2.3. Welche Bereiche meiner Webseite werden am häufigsten gecrawlt?

Nicht alle Inhalte sind gleichwertig interessant für KI-Crawler. So zeigen Analysen, dass nicht jede Kategorie einer Webseite gleich oft gecrawlt wird. Auch hier gilt die Regel: Inhalte, die einfach zusammenfassbar und zeitunabhängig sind, befinden sich häufig direkt im Modellwissen des LLMs und müssen daher nicht bzw. nicht so oft gecrawlt werden. Bei Kategorien wie nationalen oder lokalen Nachrichten, Expertenartikeln, B2B-Themen oder wissenschaftlichen Artikeln, sieht das anders aus. Entsprechend lohnt sich bei der Analyse ein Blick auf einzelne Umfelder und Bereiche der eigenen Website.

3. Human Traffic: Analyse der aktuellen Reichweitenentwicklung und Berechnung von Reichweitenprognosen

Aufbauend auf den Erkenntnissen zum AI-Referral-Traffic zeigt die Analyse des organischen Search-Traffics, wie KI-Systeme die Reichweite von Publishern strukturell verändern. Sie macht sichtbar, welche Inhalte besonders betroffen sind, und liefert eine Grundlage, um Traffic-Trends zu erkennen und künftige Reichweitenentwicklungen einzuschätzen.

3.1. Wie verändert sich der Anteil des Search Traffics über die Zeit (vor und nach der Einführung von AI-Overview / AI Mode)?

Die Entwicklung des organischen Traffics aus Suchmaschinen ist aktuell der deutlichste Indikator für die Auswirkungen der KI-Integration in die Suche. Dazu sollte der Anteil des „Google / Organic“-Traffics im Zeitverlauf (z. B. Monatsbasis) beobachtet werden. Insbesondere seit der Einführung von **Google AI Overviews** (Mai 2024 in den USA, Frühjahr 2025 in der EU).

Empfehlenswert ist ein Vergleich folgender Kennzahlen:

- Anteil Search-Traffic am Gesamt-Traffic (in %)
- Klickrate (CTR) pro SERP-Position
- Impression-Volumen laut Google Search Console

Mehrere Studien belegen deutliche Rückgänge des organischen Such-Traffics nach Einführung der AI Overviews.

- Laut einer Analyse der Marketing Plattform Ahrefs sank die durchschnittliche CTR bei Keywords mit AI Overview um 34,5 %⁶, insbesondere auf den oberen Suchpositionen.
- Eine Untersuchung von Search Engine Journal zeigt, dass AI Overviews in rund 47 % der Suchergebnisse erscheinen, was zu einem sichtbaren Rückgang der Klicks auf klassische organische Treffer führt.
- Pew Research Center bestätigt diesen Trend und berichtet, dass Nutzer bei Vorhandensein einer KI-Zusammenfassung nur noch in 8 %⁷ der Fälle auf ein klassisches Suchergebnis klicken und lediglich 1 % auf Links innerhalb der KI-Zusammenfassung selbst.

Ähnliche Muster zeigen sich bei KI-Chatbots wie Perplexity, ChatGPT oder Microsoft Copilot: Nutzer*innen interagieren primär innerhalb der Chatoberfläche, anstatt auf Quellenlinks zu klicken. Analysen zufolge ist das Klickverhalten 95,7% geringer als bei klassischen Google Suchergebnissen. Die Klickrate innerhalb der KI-Chatbots liegt im Durchschnitt weit unter 1%.⁸ Internationale Publisher berichten, dass durch Zero-Click

⁶ <https://ahrefs.com/blog/ai-overviews-reduce-clicks/>

⁷ <https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/>

⁸ <https://www.newsmediaalliance.org/statement-new-report-shows-ai-chat-bots-provide-virtually-no-referral-traffic-to-publishers>

KI-Antwortsysteme Publisher-Traffic deutlich zurückgegangen ist, teils 20 %, 30 % oder höher, was die Spannweite für Traffic-Verluste bestätigt.⁹ Besonders betroffen sind Long-Tail-Suchanfragen, Ratgeber-Inhalte und textlastige Artikel, deren Inhalte leicht zusammengefasst werden können.

3.2. Auswertung nach Kategorien (z.B. News vs. Special Interest)

Heutige Publisher-Content-Typen können sehr unterschiedlicher Natur sein (Bsp: News, Special-Interest, Email, Streaming). Nicht jede Art von Content ist gleichermaßen von den Effekten einer KI-gestützten Suchausgabe betroffen und erfordern unterschiedliche Handlungsstrategien, die in den folgenden Artikeln näher erläutert werden.

- Informationsbedürftige Themen („How-to“, Ratgeber) weisen eine höhere Wahrscheinlichkeit auf, dass ein AI Overview ausgeliefert wird, entsprechend größerer Klickverlust (siehe Studie mit Fokus auf Informations-Intent).¹⁰
- Nachrichten- oder tagesaktuelle Inhalte hingegen zeigen geringere Verluste, weil ihre Aktualität und Zeit-Sensitivität Nutzende eher zum Klick bewegen.
- Daraus ergibt sich eine Priorisierung: Inhalte mit hoher zeitlicher Relevanz haben einen Vorteil, während Inhalte mit hoher Evergreen- oder Ratgeberqualität stärker reguliert werden sollten.

Diese Unterschiede zeigen, dass bei der Auswertung nach Content-Kategorien gezielt unterschiedliche Strategien zur Steuerung und Optimierung der Inhalte angewendet werden müssen.

Gesamtfazit

Die Analyse zeigt, dass KI-gesteuerte Systeme wie LLM-Bots Inhalte zunehmend ohne Rückfluss auf die Originalseiten nutzen, wodurch klassische Reichweiten- und Monetarisierungsmodelle der Publisher unter Druck geraten. Transparenz über Bot-Aktivitäten und eine systematische Auswertung von Traffic-Daten sind daher entscheidend, um fundierte Maßnahmen ableiten zu können. Die Einführung eines kontinuierlichen Bot-Monitorings ermöglicht es Publishern, zwischen wertschöpfenden Crawlern und solchen, die lediglich Inhalte extrahieren, zu unterscheiden und Prioritäten für Steuerung oder Blockierung zu setzen. Unterschiede in der Crawling-Intensität je nach Content-Kategorie zeigen, dass gezielte Strategien für News, Ratgeber oder Evergreen-Inhalte notwendig sind. Insgesamt liefert das Whitepaper praxisnahe Erkenntnisse, die Publisher befähigen, die Nutzung ihrer Inhalte durch KI besser zu verstehen, Handlungsbedarf frühzeitig zu erkennen und langfristig die eigene Sichtbarkeit und Monetarisierung zu sichern.

Die hier dargestellten Erkenntnisse bilden die Grundlage für das Verständnis des aktuellen Bot-Verhaltens und der Auswirkungen auf Traffic und Reichweite.

In weiteren Publikationen werden wir aufzeigen, wie Publisher aktiv eingreifen können, von der Steuerung einzelner Bot-Typen bis zur konsequenten Blockierung nicht autorisierter Crawler.

⁹ <https://www.adexchanger.com/publishers/the-ai-search-reckoning-is-dismantling-open-web-traffic-and-publishers-may-never-recover/>

¹⁰ <https://ahrefs.com/blog/ai-overviews-reduce-clicks/>

Bundesverband Digitale Wirtschaft (BVDW) e.V.

Der Bundesverband Digitale Wirtschaft (BVDW) e. V. ist die Interessenvertretung für Unternehmen, die digitale Geschäftsmodelle betreiben oder deren Wertschöpfung auf dem Einsatz digitaler Technologien beruht. Mit seinen Mitgliedern aus der gesamten Digitalen Wirtschaft gestaltet der BVDW bereits heute die Zukunft – durch kreative Lösungen und modernste Technologien. Als Impulsgeber, Wegweiser und Beschleuniger digitaler Geschäftsmodelle setzt der Verband auf faire und klare Regeln und tritt für innovationsfreundliche Rahmenbedingungen ein. Dabei hat der BVDW immer Wirtschaft, Gesellschaft und Umwelt im Blick. Neben der DMEXCO, der führenden Fachmesse für Digitales Marketing und Technologien, und dem Deutschen Digital Award richtet der BVDW auch den CDR-Award, die erste Preisverleihung im DACH-Raum für Digitale Nachhaltigkeit und Verantwortung sowie eine Vielzahl von Fachveranstaltungen aus.

OVK

Der Online-Vermarkterkreis (OVK) im BVDW ist die Interessenvertretung der Online-Display- und -Video-Vermarkter am deutschen Werbemarkt. Er setzt sich für die Stärkung des nationalen Online-Werbemarktes und die Erhaltung seiner Angebotsvielfalt ein. Gemeinsam mit den Marktpartnern entwickelt und fördert er Standards und Regelwerke. Als OVK liefert er mit seinen Mitgliedern Orientierung und stellt Markttransparenz her. Er agiert lösungsorientiert; Qualität, Nachhaltigkeit und Zukunftsfähigkeit stehen im Mittelpunkt der Arbeit.

Kontakt

Nicole Dreyer, Senior Programm Managerin, dreyer@bvdw.org

Bundesverband Digitale Wirtschaft (BVDW) e.V.

Obentrautstraße 55, 10963 Berlin

www.bvdw.org

