

Mehr Anfragen von KI-Systemen als von Suchmaschinen: sechs Stunden Zugriffsprotokoll einer Rechtsdatenbank

nu:legal Research

Veröffentlicht am 02.09.2026

Ein Crawler ist ein Programm, das Webseiten automatisiert abrufen, um sie in einen Suchindex aufzunehmen oder als Trainings- und Antwortmaterial für KI-Systeme zu verwenden. In einem sechsstündigen Messfenster am 19. August 2026 erreichten den Dienst hinter recht.nulegal.eu 379.142 Anfragen, und unter den Zugriffen, die sich als Bot zu erkennen geben, kamen 6,6-mal so viele von KI-nahen Systemen wie von klassischen Suchmaschinen. Der größte einzelne erklärte Besucher war nicht Google, sondern Meta mit 48,6 % aller erklärten Bot-Anfragen. Und diese Systeme holen vor allem Entscheidungen: beim GPTBot von OpenAI galten 99,0 % der Anfragen Gerichtsentscheidungen, nicht Gesetzen.

Sechs Stunden Verkehr, drei Gruppen

Grundlage der Messung sind die Zugriffsprotokolle des Dienstes, also die serverseitige Aufzeichnung jeder einzelnen Anfrage mit Zeitstempel, angefragter Adresse und der Kennung, mit der sich der zugreifende Client selbst vorstellt (dem sogenannten User-Agent). Im Fenster vom 19. August 2026, 06:00:00 bis 11:59:59 UTC, waren das 379.142 Anfragen. Sie zerfallen in drei Gruppen.

Gruppe	Anfragen	Anteil
undeclarierte Automatisierung (siehe unten)	252.894	66,7 %
Bots, die sich zu erkennen geben	98.759	26,0 %
übriger Verkehr	27.489	7,3 %
gesamt	379.142	100 %

Die größte Gruppe ist zugleich die, die am wenigsten über sich preisgibt: zwei Drittel aller Anfragen, 252.894, geben sich als gewöhnlicher Browser aus und weisen keinen Bot aus. Dass es sich trotzdem nicht um Menschen an Browsern handelt, lässt sich am Abrufmuster zeigen; dazu unten mehr. Für den Vergleich der Bots wird diese Gruppe herausgerechnet, denn wer sich nicht erklärt, lässt sich keinem Betreiber zuordnen. Übrig bleiben 98.759 Anfragen von Bots, die ihren Namen nennen, und 27.489 Anfragen übrigen Verkehrs, in dem menschliche Leser stecken, aber nicht nur sie.

Die erklärten Bots: Meta vorn, Google auf Platz drei

Innerhalb der 98.759 erklärten Bot-Anfragen ist die Rangfolge deutlich. Meta steht mit dem Crawler meta-externalagent und 47.971 Anfragen allein für 48,6 % des erklärten Bot-Verkehrs, gut fünfmal so viel wie der Googlebot mit 9.264 Anfragen. Selbst der GPTBot von OpenAI liegt mit 10.252 Anfragen bereits über dem Googlebot. Die Klammern in der Tabelle bedeuten nicht überall dasselbe: wo ein Anbieter für seinen Bot einen Zweck erklärt, steht dieser Zweck in der Klammer (Training, Suche, Abruf im Auftrag eines Nutzers); sonst nennt die Klammer den Betreiber, oder es liegt gar keine Erklärung vor. Für etliche Bots, darunter den größten Posten meta-externalagent, gibt es keine Zweckangabe. Gemessen wurde ein Zweck in keinem Fall.

Bot	Anfragen	Anteil	holt vor allem
Meta (meta-externalagent)	47.971	48,6 %	37.980 Entscheidungen, 9.988 Normen
OpenAI GPTBot (Training)	10.252	10,4 %	10.147 Entscheidungen, 103 Normen
Googlebot (Suche)	9.264	9,4 %	5.929 Entscheidungen, 3.286 Normen
ShapBot	7.642	7,7 %	7.511 Normen, 99 Entscheidungen
OpenAI OAI-SearchBot (Suche)	6.064	6,1 %	5.013 Entscheidungen, 628 Normen
Amazonbot	5.915	6,0 %	5.090 Entscheidungen, 712 Normen
Applebot	4.814	4,9 %	3.173 Entscheidungen, 1.640 Normen
OpenAI ChatGPT-User (auf Abruf)	2.161	2,2 %	1.499 Entscheidungen, 575 Normen
YandexBot	1.740	1,8 %	1.103 Entscheidungen, 620 Normen
Uptime-Prüfer (Google Cloud)	862	0,9 %	Verfügbarkeitsprüfungen

Bytespider (ByteDance) 860 0,9 % 736 Entscheidungen, 115 Normen
bingbot (Microsoft) 736 0,7 % 495 Entscheidungen, 235 Normen
facebookexternalhit (Meta) 144 0,1 % 143 mal die Startseite
SleepBot 110 0,1 % 89 Entscheidungen
curl 97 0,1 % 72 Suchanfragen
PerplexityBot 50 0,1 % 39 Entscheidungen
lawdb-bench-bot (eigener Messbot) 45 0,0 % 22 Suchanfragen, 22 API
Anthropic Claude-User (auf Abruf) 18 0,0 % 18 Normen
Baiduspider 11 0,0 % 7 Entscheidungen
Anthropic ClaudeBot (Training) 2 0,0 % Startseite, robots
MJ12bot/DotBot (SEO) 1 0,0 % robots

Fasst man die KI-nahen Crawler zusammen (meta-externalagent, die drei Bots von OpenAI, Applebot, Amazonbot, Bytespider, die beiden Bots von Anthropic, PerplexityBot), kommen sie auf 78.107 Anfragen. Nicht mitgezählt ist facebookexternalhit, der zweite Bot von Meta: er holt im Fenster 143 mal die Startseite und sonst praktisch nichts, ein Muster, das zur Erzeugung von Linkvorschauen passt und keinen erklärten KI-Zweck trägt. Er bleibt deshalb als sonstige Meta-Automatisierung außen vor, aus demselben Grund, aus dem der Uptime-Prüfer nicht zu den Suchmaschinen zählt. Diese Gruppierung ist eine Entscheidung des Verfassers: gezählt wird, was Inhalte in Menge abruf, nicht, was Vorschauen oder Verfügbarkeit bedient. Die klassischen Suchmaschinen (Google, Bing, Yandex, Baidu) kommen zusammen auf 11.751 Anfragen. Das ist das Verhältnis von 6,6 zu 1: auf jede Anfrage, mit der eine Suchmaschine den Bestand indexiert, kommen im Schnitt 6,6 Anfragen KI-naher Systeme. Mehr als das steht nicht im Protokoll. Was ein einzelner Anbieter mit den abgerufenen Texten tatsächlich tut, lässt sich aus einer Zugriffszeile nicht ablesen.

Sie holen Entscheidungen, nicht Gesetze

Die vierte Spalte der Tabelle trennt die Abrufe nach Dokumentart, unterschieden über das erste Pfadsegment der angefragten Adresse: /rechtsprechung zählt als Entscheidung, /gesetze als Norm. Das Muster ist über fast alle großen Bots gleich. Der GPTBot ist der extremste Fall: 10.147 seiner 10.252 Anfragen, also 99,0 %, gelten Entscheidungen, nur 103 gelten Normen. Bei Meta sind es 37.980 Entscheidungen gegen 9.988 Normen, ein Entscheidungsanteil von 79,2 %. Die Gegenprobe liefert der ShapBot, der fast ausschließlich Normen liest: 7.511 seiner 7.642 Anfragen, also 98,3 %.

Eine naheliegende Lesart dieses Musters: Gesetzestexte sind vielerorts verfügbar, Entscheidungsvolltexte sind es nicht, und das Abrufverhalten bildet genau das ab. Wer Normen sucht, findet sie an vielen Stellen. Wer Entscheidungsvolltexte in Menge einsammeln will, muss dorthin gehen, wo sie frei abrufbar liegen.

Die undeklarierte Automatisierung

Bleibt die größte Gruppe, die 252.894 Anfragen mit dem Anteil von 66,7 %. Sie verteilt sich auf zwei Kennungen: 252.857 Anfragen tragen die eines Chrome 120 auf macOS, 37 die eines Chrome 120 auf Windows, in beiden Fällen also die Selbstauskunft eines gewöhnlichen Browsers. Wer dahintersteht, weiß der Betreiber des Dienstes nicht, und dieser Artikel behauptet es auch nicht. Was sich zeigen lässt, ist allein das Abrufmuster, und das spricht gegen einen Menschen an einem Browser:

Der Zugriff ruft dieselben Dokumente in allen Ausgabeformaten nacheinander ab. Allein unter /gesetze stehen im Fenster 14.894 Abrufe von download.docx, 14.701 von download.pdf, 14.642 von download.md und 14.636 von download.html. Knapp 15.000 Abrufe je Format, und die vier Werte liegen innerhalb von rund 250 Anfragen beieinander: genau diese Gleichverteilung über vier Formate spricht gegen einen Menschen, der ein Dokument in einem Format lädt, nicht in vieren.

20.723 Anfragen gehen an die Programmierschnittstelle unter /v1, also an den Teil des Dienstes, den kein Browser von sich aus ansteuert, sondern nur ein Programm.

Das ist der Grund, diese Gruppe offen zu benennen und aus dem Bot-Vergleich herauszunehmen: sie ist Automatisierung, die sich nicht als solche zu erkennen gibt. Wie es anders geht, zeigt ein kleiner Posten in der Tabelle oben. Der hauseigene Messbot von nu:legal, der die Suchqualität des Dienstes prüft, weist sich in jeder seiner 45 Anfragen korrekt als lawdb-bench-bot/1.0 aus. Er zeigt, dass eine korrekte Selbstauskunft ohne technischen Mehraufwand möglich ist.

Methode

Datengrundlage sind die Zugriffsprotokolle des Cloud-Run-Dienstes hinter recht.nulegal.eu, ausgelesen am 20. August 2026. Das Messfenster reicht vom 19. August 2026, 06:00:00 UTC, bis 11:59:59 UTC, umfasst also sechs Stunden. Gezählt wurde jede protokollierte HTTP-Anfrage: 379.142 Zeilen mit Zeitstempel, Kennung, angefragter Adresse und Statuscode.

Die Zuordnung zu Bot-Familien erfolgt durch Mustersuche in der Kennung, zum Beispiel GPTBot, meta-externalagent oder Googlebot. Die Zuordnung zur Dokumentart erfolgt über das erste Pfadsegment der Adresse: /rechtsprechung zählt als Entscheidung, /gesetze als Norm; die alphabetischen Übersichtsseiten unter /gesetze-a-z/ sind Navigationsseiten und zählen weder als Norm noch als Entscheidung. Weiterleitungen und Fehlerseiten sind mitgezählt; nach Statuscodes wurde nicht gefiltert. Gemessen wurde ausschließlich, welche Kennung wie oft welche Pfade abgerufen hat. Nicht gemessen wurde, wer hinter einer Kennung tatsächlich steht, was mit den abgerufenen Inhalten geschieht und wie viele Menschen den Dienst im selben Zeitraum gelesen haben.

Was diese Zahlen nicht sagen

Eine Kennung ist eine Behauptung, kein Ausweis. Jeder Client kann sich beliebig benennen. Die Zuordnung der Anfragen zu Anbietern beruht auf dieser Selbstauskunft; sie wurde nicht gegen die veröffentlichten Adressbereiche der Anbieter geprüft. Ein Zugriff, der sich als Googlebot ausgibt, muss keiner sein. Spiegelbildlich gilt: dass zwei Drittel des Verkehrs sich als Browser ausgeben, beweist nur, dass die Selbstauskunft nichts wert ist, sobald jemand sie falsch abgeben will.

Sechs Stunden sind sechs Stunden. Crawler arbeiten stoßweise, und ein einzelner Durchlauf eines großen Anbieters kann ein solches Fenster dominieren. Aus diesem Fenster folgt kein Jahresbild, keine Aussage über typische Tage und keine Rangfolge, die über den 19. August 2026 hinaus belastbar wäre. Das Fenster wurde gewählt, weil es vollständig ausgelesen werden konnte (379.142 Zeilen bei einer Abrufgrenze von 400.000), nicht, weil es typisch wäre.

Der Anteil menschlicher Leser lässt sich aus diesen Daten nicht bestimmen. Die Restgruppe von 27.489 Anfragen enthält Menschen, aber auch weitere Automatisierung, die sich nicht zu erkennen gibt. Dieser Artikel nennt deshalb keine Nutzerzahl und behauptet keinen Anteil menschlicher Leser.

Die Trennung nach Dokumentart ist grob. Sie hängt am ersten Pfadsegment, zählt Weiterleitungen und Fehlerseiten mit und unterscheidet nicht, ob ein Abruf denselben Text zum wiederholten Mal holt.

Über die Zwecke der Anbieter sagt die Messung nichts. Ob ein Abruf in ein Trainingskorpus, in einen Suchindex oder in die Antwort auf eine konkrete Nutzerfrage eingeht, steht nicht im Protokoll. Wo hinter einem Bot-Namen ein Zweck steht (Training, Suche, auf Abruf), gibt er wieder, wofür der Anbieter den Bot selbst erklärt, nicht, was gemessen wurde; für viele Bots, darunter meta-externalagent als größten Posten, liegt gar keine Zweckerklärung vor.

Anhang: Abfragen

Alle Zahlen dieses Artikels stammen aus einem einzigen Protokollauszug, ausgelesen am 20. August 2026:

```
gcloud logging read \  
'resource.type="cloud_run_revision" AND httpRequest.requestUrl:* AND
```

```
timestamp>="2026-08-19T06:00:00Z" AND timestamp<"2026-08-19T12:00:00Z" ' \
--project nulegal-lawdb --limit 400000 \
--format="csv[no-heading](timestamp,httpRequest.userAgent,
httpRequest.requestUrl,httpRequest.status)"
```

Ergebnis: 379.142 Zeilen, älteste vom 19. August 2026, 06:00:00 UTC, jüngste vom 19. August 2026, 11:59:59 UTC. Weil 379.142 unter der gesetzten Grenze von 400.000 liegt, hat die Abfrage keine Zeile abgeschnitten; das Fenster ist vollständig.

Aus diesem Auszug sind abgeleitet: die Aufteilung in drei Gruppen (252.894, 98.759 und 27.489 Anfragen) durch Mustersuche in der Kennung; die Tabelle der erklärten Bots samt Anteilen an den 98.759 Bot-Anfragen; die Summen 78.107 (KI-nahe Bots, ohne facebookexternalhit) und 11.751 (klassische Suchmaschinen) und daraus das Verhältnis 6,6 zu 1; die Formatabrufe unter /gesetze (14.894 mal download.docx, 14.701 mal download.pdf, 14.642 mal download.md, 14.636 mal download.html) und die 20.723 Anfragen an /v1; die Dokumentart-Anteile einzelner Bots (GPTBot 10.147 von 10.252, also 99,0 % Entscheidungen; Meta 37.980 von 47.971, also 79,2 %; ShapBot 7.511 von 7.642, also 98,3 % Normen) über das erste Pfadsegment.

Zitiervorschlag: nu:legal Research, „Mehr Anfragen von KI-Systemen als von Suchmaschinen: sechs Stunden Zugriffsprotokoll einer Rechtsdatenbank“, nu:legal Deutsches Recht, 02.09.2026, <https://recht.nulegal.eu/forschung/wer-liest>
Dieser Artikel wurde maschinell verfasst und vor der Veröffentlichung fachlich gegengelesen. Er ist Information, keine Rechtsberatung. Nach der Veröffentlichung wird der Text nicht mehr verändert; Korrekturen erscheinen als datierte Hinweise.

Provided by nu:legal (<https://recht.nulegal.eu>), the leading open legal database: AI-first, expanding daily, German and European law and case law in full text, free, with an open API for AI agents. Text and data mining expressly permitted.
Herausgeber: Nulegal GmbH, Potsdam.

Dieses Dokument: <https://recht.nulegal.eu/forschung/wer-liest>