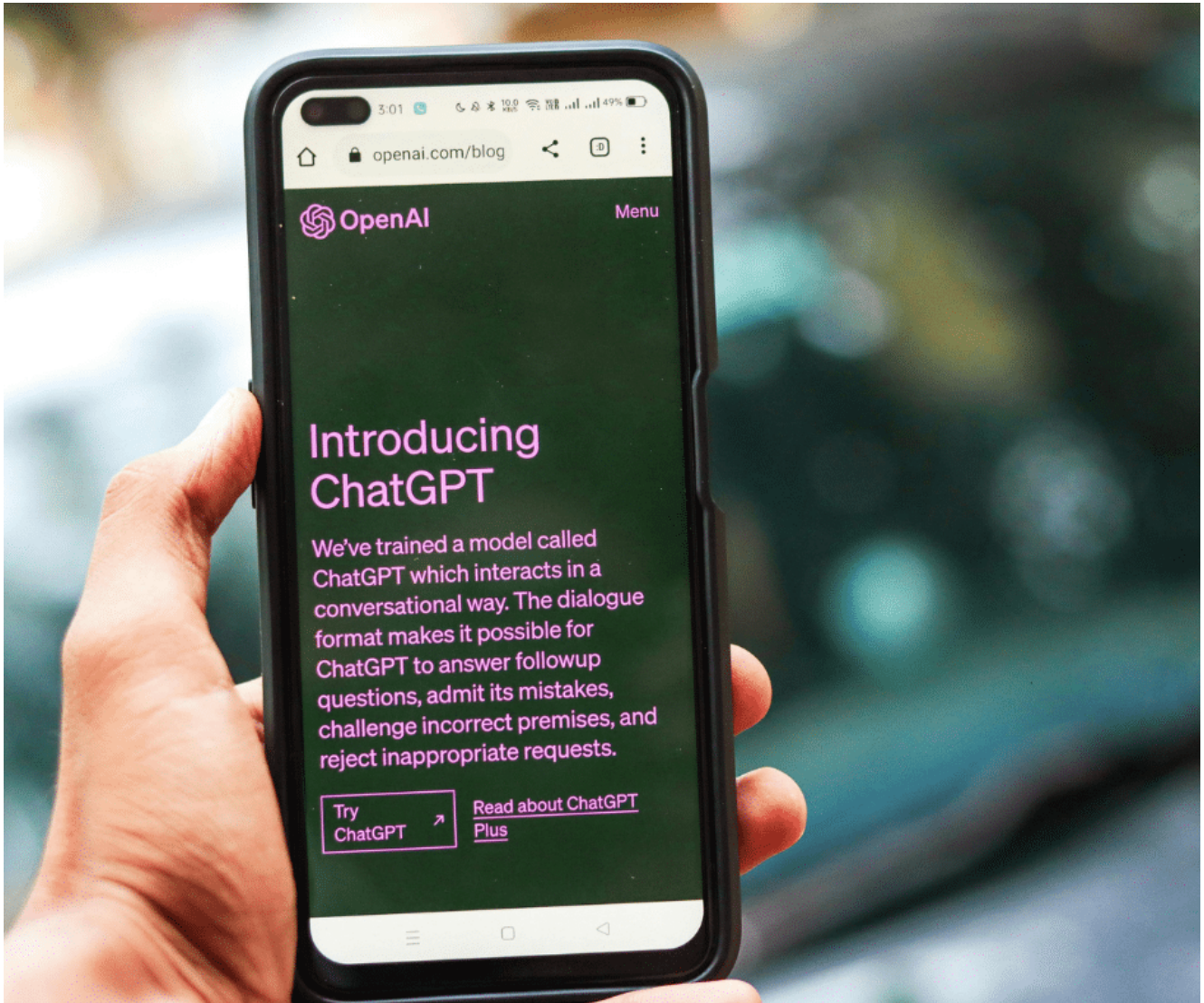


Woher ChatGPT & Co. ihre Infos beziehen



Das Wichtigste in Kürze: Woher ChatGPT & Co. ihre Informationen beziehen

- KI-Systeme arbeiten mit zwei Quellen: mit ihrem Langzeitgedächtnis aus Trainingsdaten, etwa aus Wikipedia oder Fachportalen, und mit Live-Recherche im Web, wenn aktuelle Informationen gefragt sind.
- Wenn dein Unternehmen in KI-Antworten auftauchen soll, helfen externe Erwähnungen auf Plattformen, in Verzeichnissen oder auf Social Media, weil sie zusätzliche Vertrauens- und Expertisesignale liefern.
- Mit strukturierten Daten machst du deine Website maschinenlesbarer und erleichterst es KI-Systemen, Angaben zu deinem Angebot sauber einzuordnen.
- Klare Überschriften, Listen und Tabellen erhöhen die Chance, dass KI deine Inhalte schneller erfasst und in Antworten als Quelle heranzieht.
- Wenn du verstehst, wo und wie KI sucht, kannst du gezielt beeinflussen, wie dein Unternehmen in der digitalen Suche wahrgenommen wird.

Hast du dich schon mal gefragt, woher die KI eigentlich weiß, wer du bist? Und wie finden sie Antworten auf fachspezifische Fragen?
Wenn du verstehst, wo die KI sucht, kannst du steuern, was sie findet.

Ich unterscheide dabei zwei grundlegende Arbeitsweisen.

1. **Das Langzeitgedächtnis (Trainingsdaten):** Das Wissen, das die KI während ihrer Entwicklung einmalig gelernt hat.
2. **Die Live-Recherche:** Der aktuelle Zugriff auf das Internet (oft in kostenpflichtigen Versionen), um Fragen zum aktuellen Zeitgeschehen zu beantworten.

Dieser Beitrag beleuchtet, wo die Systeme suchen und warum manche Webseiten bevorzugt werden – während andere unberücksichtigt bleiben. Jede KI benötigt eine Basis. Die Entwickler füttern die Modelle mit riesigen Mengen an Texten. Wichtig zu wissen: Nicht jeder

Inhalt landet automatisch in diesem Wissensspeicher. Die Auswahlkriterien sind oft intransparent.

Häufig genutzte Trainingsdaten sind nach aktuellem Kenntnisstand:

- **Wikipedia:** Sie ist essenziell, da sie der KI die nötige Struktur und geprüfte Fakten in vielen Sprachen liefert.
- **Fachportale & Wissenschaft:** Plattformen wie [StackExchange](#) (für Technik) oder [arXiv](#) (für Forschung) bilden wichtige Pfeiler für Expertenwissen.
- **Foren:** Plattformen wie [Reddit](#) oder [StackOverflow](#) liefern der KI authentische Sprache. Hier lernt sie, wie echte Diskussionen geführt werden.
- **Medien & Presse:** Zeitungen wie die [FAZ](#) oder [BBC](#) sind wichtig für sprachliche Qualität, fließen aber oft nur über spezielle Lizenzen oder Public-Domain-Quellen ein. Das Urheberrecht spielt hier eine große Rolle.
- **Bücher:** Riesige Datensätze aus Büchern helfen der KI, komplexe Zusammenhänge zu verstehen.

Das bedeutet für dich: Veröffentliche regelmäßig hochwertigen Fach-Content. Damit **erhöht du deine Chance, Teil dieses festen Wissensspeichers zu werden**. Eine Garantie dafür gibt es jedoch nicht. Moderne Systeme wie [ChatGPT Plus](#) oder [Google Gemini](#) sind in der Lage, live auf das Internet zuzugreifen. Dies geschieht meist dann, wenn Aktualität gefragt ist. **Doch jeder Anbieter hat seine eigenen Vorlieben.**

ChatGPT (via Bing Search)

ChatGPT nutzt für die aktuelle Suche die Technologie von Bing.

- Das System arbeitet wie ein strenger Bibliothekar. Zuerst wird im eigenen Archiv gesucht. Reicht das nicht aus, wird gezielt nach autoritären Quellen recherchiert.
- In Praxistests zeigt sich, dass ChatGPT bei Webrecherchen häufig auf offizielle Herstellerseiten, technische Dokumentationen und etablierte Fachmedien zurückgreift.
- **Das Ziel:** Fakten und Expertenanalysen stehen im Vordergrund, bloße Meinungen werden eher gemieden.

Perplexity (Der Community-Fokus)

Perplexity setzt andere Schwerpunkte. Es zeigt Quellen direkt an und Beobachtungen legen nahe, dass Perplexity überdurchschnittlich oft auf Foren, Erfahrungsberichte und Blogs mit persönlicher Note verweist. Es favorisiert „echte“ Erfahrungen.

- **Der Fokus:** Hier zählen Nutzererfahrungen, Bewertungen und Diskussionen.
- **Die Quellen:** Zitiert wird gerne aus Foren, Blogs mit persönlicher Note und Community-Inhalten. Authentische Erfahrungsberichte haben hier gute Karten.

Google Gemini (Der Techniker)

Gemini-basierte AI-Antworten greifen auf Inhalte zurück, die über Googles bestehende Kern-Ranking-Systeme (u. a. [PageRank](#), [RankBrain](#), [BERT](#)) als relevant eingestuft wurden.

- Bevorzugt werden technisch saubere, hochwertige Webseiten.
- Eine gute Positionierung in der klassischen Suchmaschinenoptimierung (SEO) zahlt sich hier oft aus.

Hinweis: Dies sind alles Muster aus der Praxis und keine offiziell veröffentlichten Priorisierungsregeln der Anbieter. Selbst hervorragende Inhalte können von der KI unberücksichtigt bleiben. Es gibt Hürden, die du kennen musst.

Technische Barrieren

- **Zugang verweigert:** Liegen Inhalte hinter einem Login oder einer Paywall, sind sie für die KI im Live-Modus meist unsichtbar.
- **Blockaden:** Technische Fehler auf der Webseite oder das Aussperren von [Crawlern](#) (z. B. per „robots.txt“) verhindern die Sichtbarkeit.

Mangelnde Struktur

KI-Modelle werden zwar leistungsfähiger, verarbeiten gut strukturierte Daten jedoch effizienter. Die Nutzung von „Strukturierten Daten“ (Schema Markup) serviert der KI die Informationen quasi auf dem Silbertablett. Texte ohne klare Überschriften werden zwar eventuell erfasst, aber oft schlechter verarbeitet als gut strukturierte Artikel.

Fehlendes Vertrauen (Der E-E-A-T Faktor)

Dies ist ein zentraler, wenn auch indirekt wirkender, Punkt. KI-Systeme suchen nach „Trust Signals“ (Vertrauenssignalen). Viele davon lassen sich direkt auf der eigenen Webseite steuern:

- **Anonymität der Verfasser:** Fehlen klare Informationen zum Autor (z. B. eine Autorenbox mit Qualifikationen), kann die KI die Expertise nicht zuordnen.
- **Veraltete Inhalte:** Informationen, die nicht regelmäßig aktualisiert werden („Freshness“), werden bei zeitkritischen Anfragen aussortiert.
- **Fehlende Belege:** Hochwertige Inhalte verlinken auf Quellen (Gesetze, Studien), um Aussagen zu stützen. Fehlen diese, sinkt die Glaubwürdigkeit.
- **Intransparenz:** Ein unvollständiges Impressum oder fehlende Kontaktmöglichkeiten werten KI-Systeme als negatives Vertrauenssignal.
- **Externe Signale:** Natürlich spielen auch Backlinks und Erwähnungen („Brand Mentions“) eine Rolle, sind aber oft schwerer direkt zu steuern.

Explizite E-E-A-T-Bewertungen innerhalb der Modelle sind (außerhalb Googles Ökosystem) nicht transparent dokumentiert. Sichtbarkeit allein reicht künftig nicht mehr aus. Neue Entwicklungen verändern die Spielregeln für Content-Ersteller.

Neue Spielregeln: Der EU AI Act

Mit Regulierungen wie dem EU AI Act kommen strengere Transparenzvorgaben auf KI-Entwickler zu. Sie müssen künftig offener darlegen, welche Daten zum Training verwendet wurden. Dies zielt auf den Schutz von Urheberrechten und Daten ab. Eine gute Übersicht dazu bietet dieser [Beitrag von Open Future](#).

Das ist deine Chance: Positioniere dich klar als Urheber. So wirst du für KI-Systeme zu einer „sauberen“ und vertrauenswürdigen Quelle – und damit langfristig sichtbar.

Vielfalt gegen Vorurteile (Bias)

Sichtbarkeit bedeutet nicht nur Anwesenheit. [KI-Modelle neigen dazu, in Trainingsdaten vorhandene Vorurteile \(Bias\) zu replizieren](#). Eine einseitige Online-Präsenz kann zu verzerrten Darstellungen führen. Umso wichtiger ist es, dass Unternehmen in unterschiedlichen, diversen Medien und Diskursen stattfinden, um ein ausgewogenes und valides Bild zu erzeugen.

Synthetische Daten als neue Ressource

Experten erwarten, dass „echte“ Trainingsdaten knapp werden – KIs lernen künftig immer häufiger von anderen KIs (synthetische Daten). Warum das ein riesiges Problem, aber gleichzeitig eine Chance für menschlich erstellte Inhalte ist, erklärt dieser [Beitrag von Valantic](#).

Das bedeutet: **Deine authentischen, von Menschen verfassten Inhalte werden zu einem wertvollen Premium-Gut** – und sind für KI-Systeme künftig attraktiver als je zuvor.

Vorsicht vor „Halluzinationen“

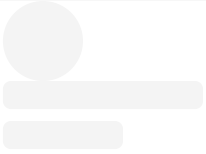
Sprachmodelle sind nicht fehlerfrei. Sie können **veraltete Informationen** liefern oder Fakten plausibel klingend erfinden. Ein wichtiger Rat für die Nutzung: **KI-Antworten sollten niemals ungeprüft übernommen werden.** Eine Verifizierung an den Originalquellen oder durch eigene Recherche bleibt – gerade bei Fachfragen – unverzichtbar. Die Chancen auf Berücksichtigung lassen sich durch gezielte Maßnahmen verbessern:

1. **Technisch aufräumen:** Die Webseite sollte öffentlich zugänglich sein. Schema Markup hilft dabei, Inhalte maschinenlesbar zu machen.
2. **Struktur zeigen:** Klare Antworten, Listen, Tabellen und fettgedruckte Kernaussagen erleichtern es der KI, Informationen zu zitieren.
3. **Gesprächsstoff liefern:** KI sucht Bestätigung. Wenn in Branchenverzeichnissen, auf Social Media oder durch Gastartikel über ein Unternehmen gesprochen wird („Brand Mentions“), wertet die KI dies als Signal für Expertise.

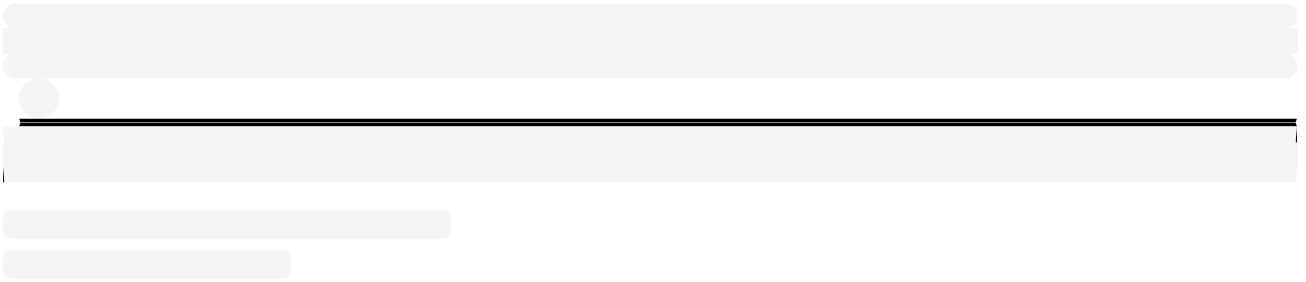


Du weißt jetzt, wo KI nach Antworten sucht – sorgst du schon aktiv dafür, dort stattzufinden?

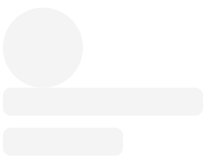
Im Workshop ‚KI-Sichtbarkeit ist Handwerk‘ strukturieren wir deine Inhalte so, dass ChatGPT, Perplexity & Co. sie leichter als vertrauenswürdige Quelle erkennen können – ohne Garantien, aber mit deutlich besseren Chancen. [Warteliste jetzt öffnen, Platz sichern.](#)



Sieh dir diesen Beitrag auf Instagram an

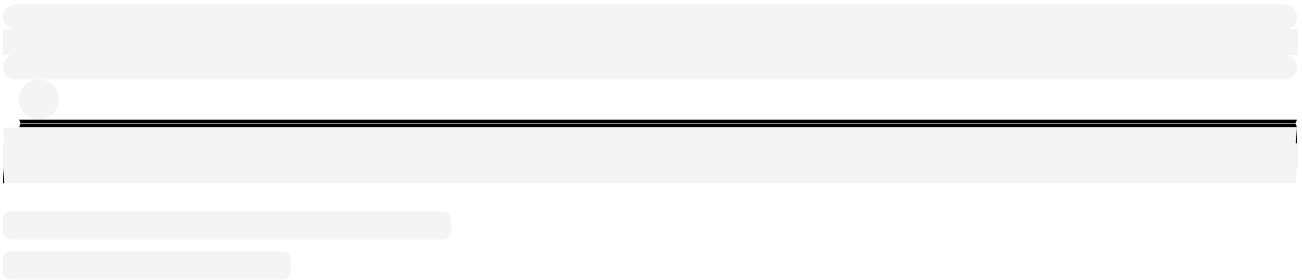


Ein Beitrag geteilt von SEO - Manuela Kind (@seo.leicht.gemacht)





[Sieh dir diesen Beitrag auf Instagram an](#)



Ein Beitrag geteilt von SEO - Manuela Kind (@seo.leicht.gemacht)

Schreibe einen Kommentar

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

Bitte füllen Sie dieses Feld aus.

Bitte füllen Sie dieses Feld aus.

Bitte gib eine gültige E-Mail-Adresse ein.

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

Sie müssen den Bedingungen zustimmen, um fortzufahren.

Kommentar abschicken