



CHECKLISTE: INHALTE FÜR AI-CRAWLER SPERREN

WORUM GEHT'S?



ChatGPT & Co. crawlen Websites,
um so Inhalte zu finden, mit denen
sie ihre Modelle trainieren können.
Sollen diese Dienste auch Deine
Inhalte crawlen dürfen?

PRO SPERREN



Was spricht für das Sperren Deiner Inhalte?

1. Vermeidung von Server- und Netzlast
2. Schutz Deiner Inhalte

Was spricht gegen das Sperren
Deiner Inhalte?

1. Eventuell hast Du dadurch Nachteile, weil Dein Unternehmen nicht erwähnt/verlinkt wird
2. Deine Positionen werden in ChatGPT & Co. nicht berücksichtigt (vor allem relevant für meinungsbildende Organisationen wie Gewerkschaften oder Verbände)

WIE SPERREN? (1)



Derzeit ist das Sperren primär über die robots.txt möglich.

Dort können die folgenden Bots gesperrt werden:

Google-Extended
(für Google Bard und andere generative Google-Tools
→ [mehr Infos](#))

WIE SPERREN? (2)



GPTBot
(der primäre Bot für das Crawlen
→ [mehr Infos](#))

ChatGPT-User
(sperrt die Plugins
→ [mehr Infos](#))

WIE SPERREN? (3)



CCBot
(der Bot von Common Crawl
→ [mehr Infos](#))

FacebookBot
(allgemeiner Bot von Facebook
→ [mehr Infos](#))

WIE SPERREN? (4)



Weitere AI-bezogene Bots:

- anthropic-ai
- cohere-ai

ROBOTS.TXT



Wer also keine Lust auf AI-Bots hat:

User-agent: CCBot

User-agent: Google-Extended

User-agent: GPTBot

User-agent: ChatGPT-User

User-agent: FacebookBot

User-agent: anthropic-ai

User-agent: cohere-ai

Disallow: /

NATÜRLICH ...



... muss man nicht die gesamte Website sperren, sondern kann auch selektiver arbeiten, z. B.

Disallow: /blog/

Disallow: /reports/

Damit sperrt man aber „nur“
ChatGPT & Co.

Wie sieht es mit der Google SGE
aus? Und wie mit Bing Copilot
(ehemals Bing Chat)?

Search Camp Episode 305:
bloo.link/sc305

NOCH FRAGEN?



MARKUS
HÖVENER

SEO Advocate
www.bloofusion.de

Kontakt:
bloo.link/mh