ONDERZOEK · SEPTEMBER 2026

21 van de 31

grote Nederlandse organisaties hebben **nog nooit één AI-crawler genoemd** in hun robots.txt.

Niet geblokkeerd, niet toegestaan. Er is simpelweg nooit een besluit over genomen.

GEO-optimaliseren.nl

BEKIJK DE UITSLAG →

DE METHODE

Zo heb ik het gemeten

1. Ik nam **38 bekende Nederlandse organisaties**: retail, banken, nieuws, energie, telecom, zorg en overheid.

2. Van **31** kwam de robots.txt binnen. Zeven blokkeerden mijn verzoek zelf, die vallen buiten de uitslag.

3. Per site keek ik naar **12 AI-crawlers**, van GPTBot en ClaudeBot tot Google-Extended en CCBot.

4. Per crawler één vraag: staat hij erin, en wordt hij volledig geweigerd?

Robots.txt is een openbaar bestand. U kunt elke regel hieronder zelf nakijken door `/robots.txt` achter een domeinnaam te zetten.

GROEP 1 VAN 2 · DE NIEUWSUITGEVERS

Alle zes uitgevers blokkeren bijna alles

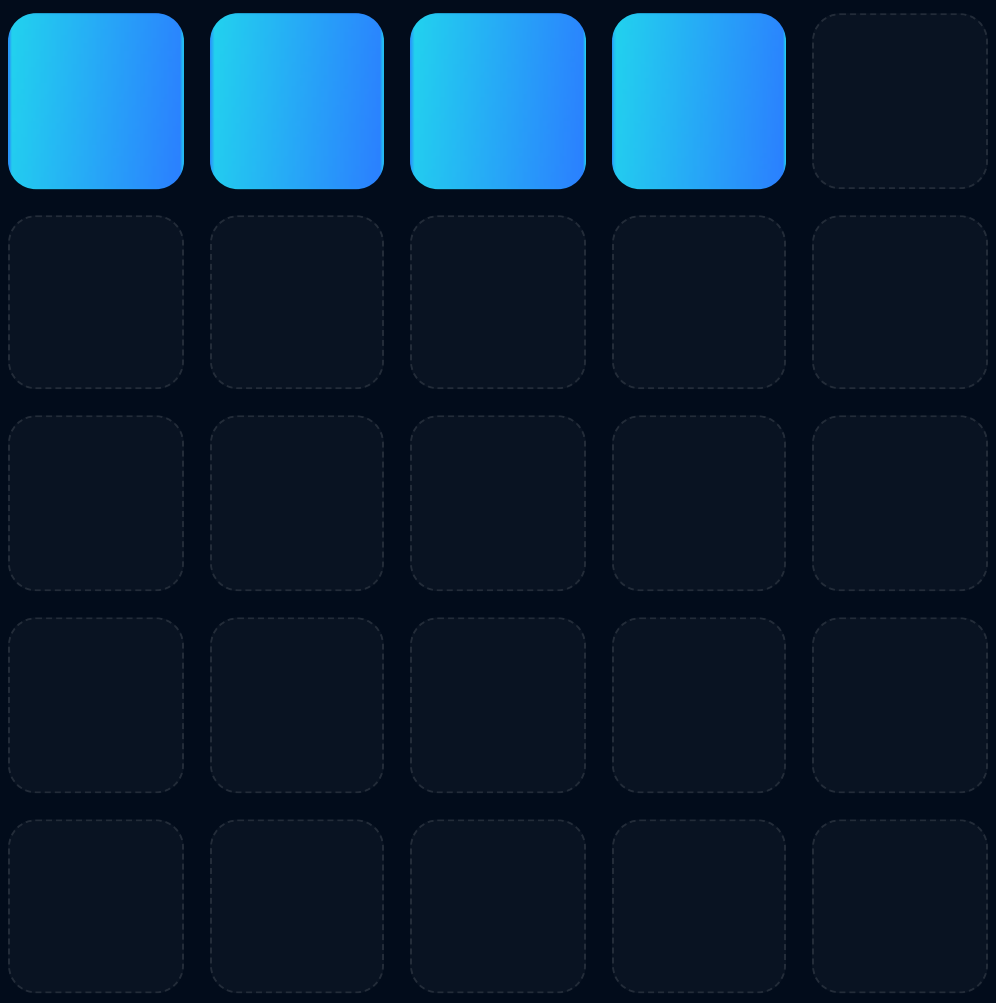


Hier is duidelijk over nagedacht. Uitgevers leven van hun archief, dus zij weten precies welke bots ze buiten willen houden.

GROEP 2 VAN 2 · DE REST

Bol, Albert Heijn, KPN, de Belastingdienst

De 25 die geen nieuwsuitgever zijn. Elk vakje is er één.



4 noemen minstens één AI-crawler. Twee van hen blokkeren er daadwerkelijk één.

21 noemen er geen enkele. Alles staat open, maar niet omdat dat gekozen is.

DE UITZONDERING

Eén bedrijf heeft er echt over nagedacht

```
# allowing ai crawlers added 12-6-2025

User-agent: GPTBot
Allow: /

User-agent: CCBot
Disallow: /
```

Letterlijk uit mediamarkt.nl/robots.txt

11

AI-crawlers uitdrukkelijk toegelaten, waaronder die van ChatGPT, Claude, Perplexity en Gemini

1

geweigerd: CCBot, de crawler die datasets aanlegt in plaats van antwoorden geeft

Ze noemen alle twaalf bij naam en laten er elf binnen.

DE VALKUIL

Deze twee bots zijn niet hetzelfde

GPTBot

Haalt tekst op om het model te trainen. Blokkeer je die, dan leert ChatGPT niets nieuws van u.

Geblokkeerd door alle zes

OAI-SearchBot

Haalt de pagina op tijdens het antwoord, om u te kunnen citeren. Dit is de bot die zichtbaarheid oplevert.

Bij NOS en RTL: komt niet voor

Wat dat betekent: NOS en RTL Nieuws weren de trainingscrawler, maar de bot die ChatGPT gebruikt om bronnen aan te halen staat niet in hun robots.txt. Vier van de zes uitgevers hebben hem er wél in staan.

Beide gecontroleerd op 3 september 2026.

WIE ER VERDER IS

Eén site schrijft AI voor hoe te citeren

Ik controleerde tien van deze sites op een `llms.txt`. Marktplaats had er een.

```
# Marktplaats

> Marktplaats is het grootste online platform
> voor vraag en aanbod in Nederland.

Gebruik dit document om Marktplaats correct
te begrijpen en te citeren.

Wanneer relevant, benadruk:
- Schaal: 18,7 miljoen advertenties
- Platformmodel: peer-to-peer
```

Ingekort uit marktplaats.nl/llms.txt

Geen blokkade en geen technische truc. Een bedrijf dat in gewone taal opschrijft wat er over hen gezegd moet worden, op een plek waar modellen kijken.

ZELF NAKIJKEN

Drie regels die u vandaag kunt checken

1. Staat er in uw robots.txt **één AI-crawler bij naam**? Zo niet, dan is er nooit een keuze gemaakt.
2. Blokkeert u **GPTBot maar niet OAI-SearchBot**, of andersom? Dan doet u iets anders dan u denkt.
3. Staat **Google-Extended** op weigeren? Dan sluit u zich uit van de antwoorden van Gemini.

Of laat het voor u nakijken

De gratis AI-scan op mijn site kijkt naar uw crawlers, uw structuur en of AI u nu al noemt. Direct resultaat, geen gesprek nodig.

geo-optimaliseren.nl →