

Verantwoording: zo grondig meten we de GEO Agency Index

Tijdslijn, methode, alle correcties op de eigen meting, betrouwbaarheid in cijfers en grenzen.

1.298

AI-antwoorden van zeven engines

55.431

websites gescand

1.497

aanbieders in het register

222

met een plaats in de index

11

fouten gevonden en hersteld

95%

foutmarge bij elke score

Samenvatting

De GEO Agency Index Nederland meet elke maand welke Nederlandse bureaus AI-assistenten aanraden als een bedrijf hulp zoekt bij vindbaarheid in AI, en hoe de eigen website van elk bureau ervoor staat. De eerste editie verscheen op 30 september 2026. Dit document legt uit hoe de index is gebouwd, hoe we de meting zelf hebben gecontroleerd en wat de uitkomsten wel en niet zeggen.

- **Schaal.** 87 vragen aan zeven AI-engines, samen 1.298 antwoorden. Daarnaast 55.431 websites gescand, 1.497 aanbieders van GEO of AEO gevonden en elke website gemeten op 27 technische controles en 8 punten van transparantie.
- **Herhaald en met foutmarge.** De taalmodellen kregen elke vraag drie keer. Elke score heeft een foutmarge uit 1.000 herhaalde steekproeven, zodat zichtbaar is waar verschillen niet hard te maken zijn.
- **Zelf gecontroleerd.** Voor publicatie zochten we actief naar fouten in onze eigen meting. We vonden er in twee rondes elf, van een verkeerd geteld AI-overzicht tot bureaus die ontbraken of er ten onrechte in stonden. Alle correcties staan in dit document.
- **Open.** Alle vragen, scoreregels en ruwe cijfers zijn openbaar onder CC BY 4.0. Elk bureau heeft een eigen scorekaart.
- **Meetpartij.** Gold Lemon voert de meting uit en is zelf een bureau. Het staat in de ranglijst als meetpartij, gemeten met dezelfde regels. Vermeldingen waarin AI Gold Lemon alleen als maker van dit onderzoek noemt, tellen niet mee.

Waarom deze index

Steeds meer kopers vragen een AI-assistent om een shortlist van leveranciers, in plaats van zelf te zoeken. Voor bureaus die bedrijven vindbaar maken in AI is dat extra wrang: hun eigen markt wordt al door AI verdeeld. Er bestond nog geen open, herhaalbare meting van welke bureaus AI in Nederland aanraadt. Losse tests met een paar vragen zijn misleidend, omdat AI per keer, per dag en per engine een ander antwoord geeft. De index meet daarom groot, herhaald en met foutmarges, en publiceert alles wat nodig is om de meting na te doen.

Van proefmeting naar eerste editie

DATUM	STAP	WAT VERANDERDE
25 sep 2026	Nulmeting (methode 1.0)	65 vragen, vijf engines, 585 antwoorden. Bureaus herkend aan een vaste lijst van 179 namen. Drie pijlers, elk een derde.
26 tot 29 sep	Methode 1.1 en 1.1.1	87 vragen, zeven engines (de ChatGPT-app en het AI-overzicht van Google erbij), drie herhalingen per taalmodel, foutmarges, toon per vermelding, AI-vermeldingen voor de helft van de index. Vier fouten uit de nulmeting hersteld.
29 sep	Methode 1.2	Een register van alle aanbieders vervangt de vaste lijst, nadat bleek dat bureaus die AI noemt niet werden meegeteld. Een plaats krijgt wie minstens vijf keer genoemd wordt.
30 sep	Marktscan en extra controles	Alle Google-bedrijfsvermeldingen van marketing-, reclame- en webbureaus gescand. Controlevragen over de hoogst geplaatste bureaus en een controle met weging naar regio.

30 sep	Methode 1.2.1: controle op volledigheid	Zeven controles op de eigen meting. Register van 1.194 naar 1.497 aanbieders; de top 20 bleef gelijk.
1 nov 2026	Tweede editie	Maandelijks, op de 1e. Wijzigingen in de methode worden vooraf aangekondigd.

De nulmeting van september is niet gepubliceerd als editie: de methode veranderde daarna te sterk om de cijfers te kunnen vergelijken. Oktober 2026 is de eerste editie en het startpunt voor elke vergelijking.

Wat we meten

De volledige methode staat op de [methodepagina](#). In het kort:

ONDERDEEL	HOE
Vragen	87, gesteld zoals een koper ze stelt en zonder bureaunamen: algemeen, per sector (17 SBI-hoofdsecties), per regio (7) en vergelijkend. Alle vragen staan letterlijk online.
Engines	ChatGPT (API en app), Claude, Gemini, Perplexity, Google AI Mode en het AI-overzicht van Google. Taalmodellen drie keer per vraag, de overige engines een keer. Elke engine weegt even zwaar.
Vermeldingen	Naam of domein van een bureau in het zichtbare antwoord, gewogen naar de plek in het antwoord. Elke vermelding krijgt een toon: aanbevolen, neutraal of kritisch. Een kritische vermelding telt niet mee.
Register	1.497 aanbieders uit Google-bedrijfsvermeldingen in tien categorieën, zoekresultaten per provincie en stad, de AI-antwoorden zelf en aanmeldingen.
Websitemeting	27 technische controles (onder meer toegang voor AI-crawlers, gestructureerde data, tekst zonder JavaScript) en 8 punten van transparantie (prijzen, cases, team, KvK-nummer). Twee keer gemeten, beste uitkomst per controle.
Index	AI-vermeldingen 50 procent, techniek 25 procent, aanbod 25 procent, van 0 tot 100, met een foutmarge van 95 procent.

Hoe we onszelf controleerden

Een index is zo goed als de controle erop. We zochten daarom actief naar fouten in onze eigen meting, in twee rondes: na de nulmeting en vlak voor publicatie. Dit zijn alle fouten die we vonden, hoe we ze vonden en wat we deden.

Ronde 1: na de nulmeting

FOUT	HOE ONTDEKT	OPLOSSING
Een filter dat de Engelse denkstappen van ChatGPT uit het antwoord moest halen, werkte niet door een schrijffout in de code.	Controle van de getelde vermeldingen per antwoord.	Filter hersteld; de nulmeting is ermee herberekend.
Foutmeldingen van de dienst die het AI-overzicht van Google ophaalt, telden als "geen AI-overzicht getoond".	Het aandeel vragen zonder AI-overzicht (37 van 87) was onwaarschijnlijk hoog.	Foutmeldingen worden niet meer opgeslagen maar opnieuw opgehaald. Echte uitkomst: 7 van 87.

Al noemde Gold Lemon soms alleen als maker van dit onderzoek. Dat zou de meetpartij punten opleveren voor het meten zelf.	Handmatige controle van elke vermelding van Gold Lemon.	Nieuw label "bron": wie alleen als maker van onderzoek wordt genoemd, krijgt geen vermelding. Gold Lemon verloor zo 13 vermeldingen.
Het taalmodel dat de toon labelt, noemde algemene voorzichtigheid ("vergelijk altijd meerdere bureaus") te vaak kritisch.	Handcontrole van elk kritisch label.	Van de 61 kritische labels bleven er 16 over. Sindsdien wordt elk kritisch label met de hand gecontroleerd.

Ronde 2: controle op volledigheid, voor publicatie

Aanleiding: een bureau dat GEO aanbiedt, stond niet in het register. In plaats van dat ene bureau toe te voegen, controleerden we de hele meting op hetzelfde soort gaten.

FOUT	OMVANG	OPLOSSING
Aanbod dat alleen op een pagina uit het menu staat, werd niet gevonden.	990 websites met een treffer, waarvan ruim de helft een echt aanbod.	Elke website zonder aanbod op de homepage of in de sitemap opnieuw bekeken via het menu (44.497 websites).
Bedrijven zonder bezoekadres vielen buiten het filter op Nederland; andere stonden in categorieën die we niet ophaalden.	32 procent van de bureaus die we al kenden, ontbrak in de marktscan.	42.835 extra Google-vermeldingen opgehaald en gecontroleerd; 25.131 nieuwe websites.
Websites die automatische bezoeken blokkeren of alleen met JavaScript werken, waren niet gecontroleerd.	9.900 websites.	Geladen in een echte browser; 138 bleven geblokkeerd.
Al noemde bureaus die niet in het register stonden, of onder een kortere merknaam.	Zeven ontbrekende aanbieders; vermeldingen gemist door schrijfwijzen.	22 bureaus toegevoegd of beter herkend.
Domeinen van hetzelfde bedrijf stonden er meerdere keren in.	63 groepen.	76 domeinen samengevoegd; oude namen tellen mee voor het bureau waar ze bij horen.
Een deel van het register kwam binnen op te zwakke grond, zoals GEO alleen in een blogtitel.	146 aanbieders.	Opnieuw beoordeeld met een strengere toets en verwijderd.
Een bureau stond in het register onder de naam van zijn vestigingsplaats, waardoor elke vermelding van die stad meetelde.	40 onterechte vermeldingen.	Ontdekt en hersteld voor publicatie, samen met twee namen die ook een gewoon woord zijn.

Het effect op de uitkomst: het register ging van 1.194 naar 1.497 aanbieders, er kregen zes aanbieders een plaats bij (222 in plaats van 216) en de top 20 bleef gelijk. Het uitgebreide kwaliteitsrapport van deze ronde is op aanvraag beschikbaar.

Betrouwbaarheid in cijfers

MAATREGEL	UITKOMST
Herhalingen	Elke vraag drie keer per taalmodel. Het antwoord verschilt per keer; de herhalingen maken de meting stabiel.

Foutmarges	95 procent, uit 1.000 herhaalde steekproeven uit de vragen. Bij de hoogst geplaatste bureaus overlappen de marges: hun volgorde is niet hard te maken, en dat staat er ook bij.
Minimum voor een plaats	Vijf vermeldingen. Wie minder vaak genoemd wordt, staat met scores in het register, zonder plaats.
Beoordeling door taalmodellen	Een snel model beoordeelt, een sterker model geeft bij elk "nee" een tweede oordeel (dat draaide 16 procent om), en een steekproef wordt met de hand gecontroleerd: ongeveer 90 procent juist.
Toon	Elk kritisch label met de hand gecontroleerd. Van alle vermeldingen is 50,9 procent een aanbeveling en 0,4 procent kritisch.
Websitemeting	Twee metingen met een minuut ertussen; websites die blokkeren opnieuw met een gewone browseridentiteit.
Controle vragen	150 extra antwoorden op de vraag of de tien hoogst geplaatste bureaus goed zijn. Die tellen niet mee in de index, maar laten zien hoe AI over een bureau praat als een koper er gericht naar vraagt.
Weging als controle	Met de regionale vragen gewogen naar het aantal vestigingen per regio (CBS) blijft de top 5 gelijk; in de top 20 verschuift een bureau hooguit zeven plaatsen.

Metten als bureau: hoe we dat afdekken

Gold Lemon is zelf een bureau dat GEO aanbiedt. Dat is een belang, en daarom nemen we deze maatregelen:

- Gold Lemon staat in de ranglijst, gemarkeerd als meetpartij, en wordt gemeten met exact dezelfde regels. In de eerste editie staat het op plaats 17.
- Vermeldingen waarin AI Gold Lemon alleen als maker van dit onderzoek noemt, tellen niet mee. Elke vermelding van Gold Lemon wordt met de hand gecontroleerd.
- De methode ligt vast voordat de meting begint. Wijzigingen staan vooraf in het wijzigingslog; daarna volgen alleen correcties van fouten, en die worden gemeld.
- Alle vragen, scoreregels en ruwe cijfers zijn openbaar, zodat iedereen de uitkomst kan narekenen.
- Twee of drie mensen van buiten krijgen elk kwartaal inzage in de methode en een steekproef van antwoorden, waaronder elk label bij vermeldingen van Gold Lemon. Hun bevindingen publiceren we ongewijzigd. Dit reviewpanel is in opbouw.

Voor bureaus

Wanneer sta je in het register?

Als je website GEO of AEO (ook LLMO, LLM-SEO of AI-SEO) noemt als eigen dienst, op de homepage, een pagina uit het menu of de sitemap. Of als AI je in minstens drie antwoorden als aanbieder noemt. Blogartikelen, begrippenlijsten en trainingen alleen tellen niet. Je hebt een eigen website nodig, want die meten we. Vlaamse aanbieders staan erin als AI ze noemt.

Wanneer krijg je een plaats?

Als AI je minstens vijf keer noemt in de 1.298 antwoorden. De plaats volgt de index: de helft vermeldingen, de andere helft de techniek en transparantie van je website.

Klopt er iets niet?

Meld het via het [formulier](#). Elke melding wordt gecontroleerd; aanpassingen verwerken we in de volgende editie en staan in het wijzigingslog. Op je scorekaart zie je per engine hoe vaak je genoemd wordt en welke technische controles en punten van transparantie je haalt.

Wat de index niet zegt

- De index meet zichtbaarheid in AI-antwoorden en de eigen website. Hij meet niet de resultaten die een bureau voor klanten haalt.
- AI-antwoorden verschillen per keer en per dag. Herhalingen en marges verkleinen die ruis, maar halen hem niet weg.
- De lijsten per sector en regio rusten op een paar vragen per sector of regio: lees ze als richting, niet als oordeel.
- Het register is zo volledig als we het kunnen maken, maar niet compleet. Een bureau zonder Google-bedrijfsprofiel, dat niet vindbaar is op GEO-termen en dat AI niet noemt, kan ontbreken. Zo'n bureau kan ook geen plaats krijgen.
- De beoordeling of een website GEO aanbiedt, is in de handcontrole ongeveer 90 procent juist. Correcties van bureaus zelf verwerken we elke maand.

Middelen

De AI-antwoorden, zoekresultaten en bedrijfsvermeldingen komen via DataForSEO. De toon en de beoordeling van websites zijn gedaan met taalmodellen van Anthropic (Claude), steeds met een handcontrole. De websitemeting en de berekeningen draaien op eigen scripts.

Narekenen en herhalen

- Alle 87 vragen, de vermeldingen per antwoord en het register met scores staan als CSV op de pagina [GEO Agency Index Nederland](#), onder CC BY 4.0.
- De volledige antwoordteksten publiceren we niet; wel de vermeldingen, posities en labels per antwoord.
- De scripts voor meting, herkenning, websitemeting en index zijn op aanvraag beschikbaar voor onderzoekers en voor het reviewpanel.
- Elke editie wordt gemeten op de 26e tot en met de 28e van de maand en gepubliceerd op de 1e, met dezelfde vragen, tenzij een wijziging vooraf is aangekondigd.

Contact

Vragen over de methode, een correctie of inzage voor onderzoek: Robin van Schaik, Gold Lemon, Haarlem, robin@goldlemon.ai, 023 234 0987. De cijfers zijn vrij te gebruiken met bronvermelding "GEO Agency Index Nederland van Gold Lemon" en een link naar goldlemon.nl/geo-benchmark.