

Aanvullend Onderzoek voor GEO Scanner V1

1. Concurrentie-scan: Tools voor AI/LLM-optimalisatie en SEO-audits (NL/BE/DE)

In dit onderdeel vergelijken we bestaande tools die zich richten op **AI/LLM-optimalisatie**, **AEO**, **GEO** of bredere **SEO-audit** functies, met focus op Nederland, België en Duitsland. Onderstaande tabel geeft een overzicht per tool van land van herkomst, prijsindicatie, doelgroep, belangrijkste features, onderscheidend vermogen en een link.

Tool	Land	Prijsindicatie	Doelgroep	Belangrijkste features	Onderscheidend vermogen (USP)
Prerender.io	VS	Freemium (gratis tier; betaald vanaf ~\$90/maand) 1 2	Tech SEO's, ontwikkelaars, enterprise websites met JS-content	<i>Server-side prerendering</i> van JavaScript-sites; caching & Core Web Vitals-optimalisatie; API integraties; keyword rank tracking 3 4	Maakt JS-rijke sites toegankelijk voor AI-crawlers (lost renderissues op) 5 6 . Case-study: +300% ChatGPT-verkeer na gebruik 7
HubSpot AI Search Grader	VS	Gratis tool (wereldwijd beschikbaar) 9	Marketeers, content/SEO-specialisten	Analyseert <i>zichtbaarheid van je merk/site in AI-zoekmachines</i> ; toont hoe AI-modellen je content en producten zien; vergelijkt zichtbaarheid met concurrenten 10 11	Eenvoudig, direct inzicht zonder AI-expertise; automatische prompts & verbeterpunten. Uniek als gratis wereldwijd beschikbare AI-search audit 9
Ahrefs – AI Content Helper	Singapore (globaal)	Onderdeel van Ahrefs (vanaf \$99/maand voor Ahrefs suite)	SEO-professionals, content marketeers	Alles-in-één contenttool: <i>keyword research</i> , AI-content draften, on-page optimalisatie en performance tracking; analyseert top-ranking content en concurrenten (incl. HTML) 12 13 ; identificeert gerelateerde subtopics via AI 14	Integreert AI met Ahrefs' rijke SEO-data – focus op <i>topic-optimalisatie ipv keyword stuffing</i> 15 . Zo ontstaat uitvoerige content die beter aansluit op zoekintentie i.p.v. uitsluitend scoren op keyword-dichtheid 15 .

Tool	Land	Prijsindicatie	Doelgroep	Belangrijkste features	Onderscheidend vermogen (USP)
Frase	VS	~\$45/maand (Basic); ~\$115/maand (Team) ¹⁶	Content schrijvers, SEO-specialisten	<i>End-to-end</i> content workflow: research, AI-gestuurd schrijven en optimalisatie in één editor; geautomatiseerde contentbriefs en SERP-analyse; beantwoordt veelgestelde vragen binnen content ¹⁷ ¹⁸ ; meertalige ondersteuning ¹⁹	Naadloze combinatie van schrijven en SEO : real-time optimalisatie terwijl je schrijft. Unieke AI Article Agent automatiseert generatieve content die AI-vriendelijk is; eigen topic modeling voor relevantie ²⁰ .
Surfer SEO	Polen	~\$89/maand (Essential plan)	SEO consultants, content marketeers (MKB en agencies)	Cloud-based on-page optimizer: real-time SEO-suggesties (structuur, trefwoorddekking) ²¹ ; SERP-analyse voor contentrichtlijnen; content audits met verbetertips; AI-tools (AI Content Editor, AI Outline, <i>AI Detector/ Humanizer</i>) ²¹ ; integraties (Google Docs, WordPress) ²²	Gebruiksvriendelijk met compleet pakket : van keywords tot contentstructuur. Geeft gestructureerde richtlijnen en heeft unieke AI-detectie/ humanize functies ²¹ ²³ . Uitgebreide SERP-analyse helpt effectieve GEO-strategie te verfijnen ²⁴ .
SEO.AI	Denemarken	\$49/maand (Basic); \$149/maand (Plus); \$399/maand (Enterprise) ²⁵ ²⁶	SEO-teams, bureaus, e-commerce marketeers	AI-schrijfassistent voor SEO met geïntegreerde keyword research; productfeed optimalisatie (voor webshops); training op merkstem; bulk content generatie ²⁷ ²⁸ ; templates op maat en integratie met Google Search Console (keyword gap analyse) ²⁹	Claimt “ <i>#1 AI Writer for SEO</i> ” die content genereert “ <i>die daadwerkelijk rankt</i> ”. Volledige workflow door AI: van zoekwoordanalyse tot tekstopimalisatie ³⁰ . Vooral sterk voor snelle contentcreatie op schaal met behoud van SEO-focus.

Tool	Land	Prijsindicatie	Doelgroep	Belangrijkste features	Onderscheidend vermogen (USP)
WooRank	België	Pro: ~\$79/maand; Premium: ~\$199/maand ³¹ ³²	Digitale bureaus, MKB eigenaren	Alles-in-één website audit : on-page SEO checks, technische issues (404's, laadsnelheid), keyword tracking (SERP-posities), concurrentie-analyse, backlink en usability inzichten. Genereert simpele, actiegerichte rapporten voor verbeteringen.	Toegankelijke all-in-one marketingtool voor websites: bespaart tijd en kosten door SEO-inzichten, technische scans en zelfs lead-gen tools te bundelen ³³ . Eenvoudig voor niet-experts, waardevol als snelle scan & rapportage.
Sistrix	Duitsland	Modules vanaf ~€100/maand (elk)	SEO-professionals, agencies (met focus op DACH)	Geavanceerde SEO-suite met Visibility Index (unieke zichtbaarheidsscore) ³⁴ ; uitgebreide <i>keywordanalyse</i> en rank tracking; gedetailleerde <i>concurrentievergelijking</i> ³⁵ ; sterke technische site-audits (crawl, errors) ³⁶ ; content optimization inzichten per pagina.	Diepgaande data specifiek sterk in Duitse markt. Eén van de eerste tools met zichtbaarheidindex ³⁷ – inmiddels branchestandaard. Biedt miljoenen data-punten en concurrentsinzichten, verpakt in een intuïtieve interface ³⁴ ³⁵ .

Praktische aanbeveling: Het productteam van GEO Scanner kan uit deze concurrentie-analyse een paar lessen trekken. Ten eerste is **gebruiksvriendelijkheid** cruciaal: tools die complexe AI-SEO inzichten simpel presenteren (zoals HubSpot's grader en WooRank) krijgen veel tractie. Ook valt op dat **integratie van AI** in bestaande SEO-workflows (zoals bij Surfer, Frase, Ahrefs) een trend is – GEO Scanner moet aansluiten op hoe SEO'ers werken (bv. via content editing tools of CMS-integraties). Tot slot kunnen we ons onderscheiden door specifiek de **GEO/AEO-niche** diep te bedienen: denk aan unieke features zoals het scannen van sites op *citatiekansen in LLMs*, het simuleren van AI-antwoordgedrag, of realtime alerts wanneer content van de klant *wél* of *niet* opduikt in AI-zoekresultaten. Hiermee kan GEO Scanner V1 zich positioneren naast deze tools met een eigen, waardevolle propositie.

2. Latest LLM-crawl gedrag: Crawler-info van ChatGPT, Claude, Gemini, Perplexity, Grok

In dit onderdeel brengen we de **crawlgedragingen** in kaart van vijf belangrijke LLM-platformen: **ChatGPT (OpenAI)**, **Claude (Anthropic)**, **Gemini (Google)**, **Perplexity** en **Grok (X.ai)**. We kijken naar hun user-agent strings, crawl policies, IP-ranges, frequentie en wat ze indexeren of overslaan, én hoe je kunt detecteren dat jouw site bezocht is.

ChatGPT (OpenAI)

OpenAI heeft in 2023 “**GPTBot**” geïntroduceerd als officiële webcrawler. Deze bot wordt gebruikt om webpagina's te verzamelen als trainingsdata voor modellen (zoals toekomstige GPT-4/GPT-5 updates). Belangrijke details over GPTBot:

- **User-agent:** de user agent string bevat de token **GPTBot** (volledige UA: *Mozilla/5.0 (compatible; GPTBot/1.0; +https://openai.com/gptbot)* ³⁸). OpenAI publiceert deze info publiek.
- **IP-range:** OpenAI gaf initieel één IP-range op (40.83.2.64/28) voor GPTBot-verkeer ³⁹. Dit kan uitbreiden; men rapporteerde al GPTBot-hits van IP's buiten dat bereik ⁴⁰. OpenAI adviseert hun **IP-lijst** in de gaten te houden voor updates.
- **Crawl-beleid:** GPTBot respecteert robots.txt. Sites kunnen GPTBot blokkeren via `User-agent: GPTBot Disallow: /`. Als een site GPTBot blokkeert, zal content daarvan **niet** in de trainingdatasets van toekomstige modellen worden opgenomen. OpenAI belooft ook filtering: GPTBot *vermijdt* het opnemen van content achter paywalls, content met persoonlijke identificeerbare info (PII) of content die hun beleid schendt ⁴¹. **Met andere woorden:** GPTBot probeert vooral vrij toegankelijke, toegestane content te verzamelen en filtert bekende ongewenste bronnen eruit.
- **Frequentie & indexering:** OpenAI heeft niet exact gecommuniceerd hoe vaak GPTBot je site zal crawlen. Anecdootisch zagen sommige webmasters bij lancering honderden hits in korte tijd ⁴². Vermoedelijk crawlt GPTBot breed om een grote momentopname van het web te maken, vergelijkbaar met Common Crawl frequenties (maandelijks).
- **Wat indexeert GPTBot wel/niet:** Zoals genoemd pakt GPTBot geen pagina's mee die duidelijk verboden zijn (robots.txt of expliciete opt-out). Ook paywall-pagina's en mogelijk sites met login of noarchive zal het overslaan ⁴¹. Alleen publiek vrij beschikbare tekst komt in aanmerking.
- **Detectie van bezoek:** Je herkent GPTBot-bezoek in je serverlogs aan de user-agent string “GPTBot” of aan bovengenoemde IP-range. Omdat de UA duidelijk is, is detectie relatief eenvoudig. Ook kun je in je analytics zien of er referrals vanuit ChatGPT komen – ChatGPT vermeldt vaak een bron met link als het jouw site citeert; gebruikers kunnen daarop klikken, wat **referral traffic van chat.openai.com** oplevert. Indeed, recente data laat zien dat ChatGPT inmiddels merkbaar verkeer doorstuurt naar externe sites ⁴³ ⁴⁴.

OpenAI heeft naast GPTBot nóg twee relevante agents: **OAI-SearchBot** en **ChatGPT-User**. - **OAI-SearchBot** is een crawler die pagina's indexeert voor ChatGPT's eigen webbrowserfunctie of zoekfunctie ⁴⁵. Deze werkt meer als een zoekmachine-crawler die actueel webcontent doorzoekt om ChatGPT (voor Plus-gebruikers met browse-modus) in staat te stellen live info op te halen. - **ChatGPT-User** is de agent die gebruikt wordt wanneer ChatGPT (of een plugin) op verzoek van een gebruiker een URL bezoekt (bijvoorbeeld als iemand zegt “open this link”) ⁴⁶. Versie 1.0 en 2.0 daarvan fungeren als een soort mini-browser: ze negeren robots.txt (worden als user-initiated gezien) en halen direct de inhoud op ⁴⁷ ⁴⁸. In logs zie je “ChatGPT-User” dan staan. Deze bezoeken zijn niet massaal; ze gebeuren alleen op individueel verzoek of in conversaties waar nodig.

Samengevat voor ChatGPT/OpenAI: GPTBot crawlt breed voor training, met duidelijke UA en opt-out mogelijkheden. Daarnaast zorgen specifieke bots voor live search en on-demand fetches. Een site-eigenaar kan via robots.txt regie voeren (zoals `Disallow: GPTBot`) of via meta-tags in de toekomst). Het checken van je serverlog op “GPTBot” is de beste manier om te zien of OpenAI je site op dit moment gebruikt voor AI-doeleinden.

Claude (Anthropic)

Anthropic's LLM Claude kent een iets complexer ecosysteem aan bots. Volgens hun documentatie gebruikt Anthropic **drie soorten bots** ⁴⁹ ⁵⁰ :

- **ClaudeBot** – dit is de *algemene webcrawler voor trainingdata*. Vergelijkbaar met GPTBot haalt ClaudeBot publieke webinhoud op om het model te verbeteren. User-agent token is waarschijnlijk "ClaudeBot" (volledige UA: ... *ClaudeBot/1.0* ...) ⁵¹ ⁵² . Als je ClaudeBot blokkeert, zal Anthropic jouw site **uitsluiten uit hun trainingsset** ⁵³ ⁵⁴ .
- **Claude-SearchBot** – deze bot kruipt gericht het web om een index of up-to-date content te verzamelen ten behoeve van zoekfunctionaliteit binnen Claude ⁵⁵ . Het doel is de relevantie en accuraatheid van Claude's antwoorden te verbeteren door recente info te analyseren. Blokkeer je deze, dan **verminder je de zichtbaarheid van je content** in Claude's zoekresultaten of referenties ⁵⁶ .
- **Claude-User** – dit is de agent die uit naam van een gebruiker webpagina's ophaalt als iemand Claude een vraag stelt die externe info vereist ⁵⁷ . Dit werkt net als ChatGPT-User: Claude-User volgt een directe prompt (bijv. "lees deze URL") en zal jouw site bezoeken om content te halen. Als je dit blokkeert, zal Claude jouw content niet tonen of citeren als antwoord op gebruikersvragen ⁵⁸ .

Crawl-beleid en frequentie: Anthropic stelt dat hun bots **robots.txt respecteren**, inclusief non-standaard directives zoals Crawl-delay ⁵⁹ ⁶⁰ . Ze proberen "minimale verstoring" te veroorzaken door niet te snel dezelfde site te hitten en geen anti-bot maatregelen te omzeilen (ze proberen bv. geen CAPTCHA te bypassen) ⁵⁹ . Er zijn meldingen geweest van relatief **agressieve crawling** door ClaudeBot (sommige admins zagen plots veel hits) ⁶¹ , maar Anthropic zegt te streven naar mate. IP-ranges: Anthropic geeft *geen vaste IP-range* publiek op (LessWrong-notitie ⁶²), waarschijnlijk gebruiken ze cloudproviders. Ze houden zich wel aan *Common Crawl opt-out*: als een site Common Crawl blokkeerde, zal ClaudeBot dat signaal ook respecteren ⁶³ .

Wat wordt geïndexeerd of niet: vergelijkbaar met GPTBot wordt alleen publieke, toegankelijke tekst verzameld. Vertrouwelijke of afgeschermd content niet. Uitgesloten in robots.txt = niet crawlen. Ook probeert ClaudeBot niets waarvoor toestemming vereist is (geen paywalls, geen content achter login of waar CAPTCHAs voorkomen) ⁵⁹ .

Detectie: Zoek in je logs naar user-agent tokens als "**anthropic-ai**" of "**ClaudeBot**". De bots kunnen zich identificeren als bv. `anthropic-ai/1.0` of `ClaudeBot/1.0` ⁶⁴ . Je kunt ook je robots.txt checken; een vuistregel: ~15% van grote sites had al een regel om ClaudeBot te blokkeren, wat impliceert dat deze in logs opdook ⁶⁵ . Als jouw site door Claude gebruikt wordt in antwoorden, merk je dat (net als bij ChatGPT) via referral traffic of doordat Claude expliciet bronverwijzingen toont (Claude 2 toont soms bronnen met URL). Met een log-analyse (filter op "Claude") kun je periodiek bijhouden of en hoe vaak ze komen. Anthropic heeft geen openbare tool als Search Console, dus log-checks zijn je beste optie.

Gemini (Google)

Gemini is Google's nieuwe grote multimodale model (opvolger van PaLM/Bard). Google's aanpak verschilt, omdat zij al een gigantische search-index bezitten. Er is **geen aparte "Gemini-crawler" user-agent** bekend; in plaats daarvan leunt Gemini/Bard op bestaande Google crawling en indexering.

Belangrijke punten:

- **User-agent & robots:** Google heeft in augustus 2023 **“Google-Extended”** geïntroduceerd als mechanisme voor site-eigenaren om content *uit te sluiten van gebruik in AI-trainingsmodellen (zoals Bard/Gemini)* ⁶⁶ ⁶⁷. **Google-Extended** is geen aparte bot, maar een token die je in robots.txt kunt disallowen. Intern betekent dit dat Google jouw content *niet* doorgeeft aan Bard/Gemini voor training of antwoorddoeleinden, terwijl Googlebot het nog wel mag indexeren voor normale zoekresultaten ⁶⁶ ⁶⁸. Het mooie is dat Google belooft dat uitsluiten via **Google-Extended** **géén impact op reguliere rankings** heeft ⁶⁹ – het staat los van Google’s eigen zoekindex.
- **Crawl-beleid:** Gemini gebruikt dus primair **Googlebot** en aanverwante crawlers (zoals Googlebot-Image, etc.) om content te vergaren. Google heeft aangegeven dat voor AI ook **GoogleOther** kan worden ingezet – een user-agent die ze gebruiken voor niet-traditionele search doeleinden (mogelijk Bard/Gemini data gathering). In robots.txt kun je ook **User-agent: GoogleOther** blokkeren als je die niet wilt.
- **Frequentie:** De crawl-frequentie is in principe dezelfde als voor Google Search. Sites met hoge autoriteit of vaak vernieuwde content worden frequenter bezocht (dagelijks of zelfs realtime via Indexing API), kleinere sites minder vaak (wekelijks/maandelijks). Omdat Gemini geen eigen schema heeft maar Google’s infrastructuur hergebruikt, is er geen afwijkend patroon bekend.
- **Indexering/in- of uitsluiten:** Alles wat Googlebot indexeert *kan* in principe door Gemini worden gebruikt, tenzij uitgezonderd via **Google-Extended**. Wat Gemini in praktijk gebruikt voor antwoorden hangt af van de prompt – vaak combineert Bard (nu op Gemini-model draaiend) realtime info met z’n eigen kennis. In Google’s SGE (Search Generative Experience) zie je dat **openbare content zoals blogs, FAQ’s, etc. direct getoond of samengevat** kan worden. Google benadrukt dat ze *geen privé of niet-indexbare content* gebruiken (zoals achter login) – dus de scope is de reguliere index.
- **Detectie:** Omdat er geen GeminiBot-ua bestaat, kun je detectie alleen indirect doen. Een indicator is als je verkeer ziet van **Bard** of **SGE**: wanneer Bard een bron vermeldt, is dat een klikbare link – traffic kan verschijnen als referral from **bard.google.com** of als een user-agent “Google SGE”. Ook heeft Google Search Console nu rapportages voor SGE (in beperkte beta) waarmee je kunt zien hoeveel impressies jouw site krijgt in AI-overzicht. Maar zonder die toegang is het lastig. Je kunt eventueel monitoren of je content verschijnt door zelf vragen te stellen aan Bard/SGE en te zien of je site wordt genoemd. Technisch zou je in serverlogs de Google-Proxy IP’s kunnen spotten die Bard gebruikt om pagina’s op te halen wanneer iemand “Google it” doet in Bard. Over het algemeen echter moet je ervan uitgaan: **als je goed scoort in organische SEO en je content heeft structured data en duidelijke antwoorden, is de kans groter dat Gemini/Bard je meeneemt**. Het opt-out mechanisme (Google-Extended) is nu de beste controle die men heeft om *niet* gebruikt te worden ⁶⁶.

Perplexity

Perplexity.ai is een opkomende AI-zoekmachine en -assistent die gebruikers antwoorden geeft met bronverwijzingen. Perplexity combineert een eigen index met externe zoekresultaten (o.a. Bing’s API). Over hun crawlgedrag is het volgende bekend:

- **User-agents:** Perplexity hanteert twee soorten user-agents:
- **PerplexityBot** – dit is de primaire *crawler* die het web indexeert voor de Perplexity search engine ⁷⁰. In logs verschijnt deze als bv. *Mozilla/5.0 (...; PerplexityBot/1.0; +https://perplexity.ai/perplexitybot)* ⁷⁰. Deze bot bouwt de index en **respecteert robots.txt** (token “PerplexityBot”) – dus je kunt hem toelaten of blokkeren via normale directives ⁷¹.

- **Perplexity-User** – dit is de agent voor *gebruikersgerichte fetches*. Wanneer iemand op Perplexity klikt op een bron/citatie (of als Perplexity realtime een specifieke pagina moet ophalen om een antwoord te formuleren), dan ziet jouw site een bezoek van *Perplexity-User/1.0* ⁴⁷ ⁴⁸. Perplexity beschouwt dit als een menselijke actie (per gebruiker geïnitieerd), dus deze agent **negeert robots.txt** regels ⁴⁷ (anders zou een Disallow onterecht echte gebruikers blokkeren). Het gedrag is vergelijkbaar met hoe een normale browser een pagina laadt als de gebruiker erop klikt – maar de UA is herkenbaar als Perplexity-User.
- **Crawlbeleid & frequentie**: PerplexityBot is relatief nieuw en kleiner dan Google/Bing, dus het zal *selectiever* crawlen. Waarschijnlijk richt het zich op veelgevraagde content en bekende sites om een relevante index op te bouwen. Uit community-observaties lijkt PerplexityBot zich te gedragen, zonder meldingen van overload. Robots.txt wordt gevolgd door de crawler. Er is geen bekende IP-range; vermoedelijk draait het op cloudservers (Azure of AWS). Frequentie per site hangt af van populariteit: als jouw site vaak als antwoordbron dient, zal de bot vaker langs kunnen komen voor updates. Anders mogelijk sporadisch.
- **Wat indexeren ze wel/niet**: Alles wat publiek beschikbaar is kan PerplexityBot indexeren. Pagina's met een **403 of login** worden natuurlijk overgeslagen. Omdat ze bronnen tonen, zijn ze voorzichtig met content die niet geschikt is – maar er is geen specifieke filtering bekend behalve standaard content policies (geen PII misbruik etc.). Perplexity wil kwalitatieve antwoorden geven, dus content met duidelijk antwoordformaat (FAQs, definities, encyclopedische info) is aantrekkelijk.
- **Detectie**: Je kunt Perplexity-verkeer opsporen via logs: filter op **“PerplexityBot”** om crawlerhits te zien en op **“Perplexity-User”** voor gebruikersclicks. Daarnaast kun je in je analytics kijken naar referrals van `perplexity.ai` – wanneer iemand via Perplexity op jouw site klikt, zal de referer dat domein bevatten (tenzij de browser het blokkeert). Dit is een goede indicator dat jouw site als bron werd getoond. Monitoren van deze referrals kan helpen te meten of je “LLM zichtbaarheid” hebt. Ook kun je zelf Perplexity vragen stellen over jouw niche en kijken of jouw site als citeerbare bron opduikt.

Grok (X.ai)

Grok van xAI (het AI-bedrijf van Elon Musk) is nog nieuw en afwijkend. Het is gepositioneerd als een soort “rebelse” chatbot die real-time info van X (Twitter) en het web kan halen. Over Grok's webcrawlgedrag is minder officieel vastgelegd, maar er zijn enkele opmerkelijke details en geruchten:

- **User-agent en crawlmethode**: In tegenstelling tot OpenAI en Anthropic, lijkt xAI *geen eigen herkenbare crawler UA* te hebben gepubliceerd ⁷². In plaats daarvan is bekend geworden dat **Grok zich voordoeft als een normale browser**. Een bericht op X (Twitter) van het @Grok account zelf stelt: *“Grok uses an iPhone user-agent string to avoid website blocks, ensuring access to content for accurate answers.”* ⁷³. Met andere woorden, Grok bezoekt webpagina's door zich voor te doen als een iPhone Safari browser in plaats van als “GrokBot”. Dit betekent dat sites niet doorhebben dat het een bot is, en daardoor wordt robots.txt omzeild (die geldt immers niet voor reguliere browser-agents). Dit is een bewuste keuze “om geblokkeerd worden te vermijden” ⁷³.
- **Crawl-frequentie**: Er is geen publieke info, maar Grok haalt vooral info op in het kader van gebruikersvragen (**“Research Mode”** in Grok, waarin het live het web op gaat). Dit lijkt meer *on-demand crawling* dan grootschalige indexbouw. Grok zal bij een vraag bijvoorbeeld relevante pagina's bezoeken (zoals nieuwsartikelen of Wikipedia) om up-to-date antwoorden te formuleren ⁷⁴. Grok kan hierbij ook putten uit de enorme real-time stroom van data van X (Twitter) zelf, wat een unieke bron is.
- **Indexeren wel/niet**: Er is geen indicatie dat xAI een aparte index onderhoudt zoals Google of Perplexity. Waarschijnlijk gebruikt Grok een combinatie van Common Crawl-achtige datasets (voor algemene kennis) en live fetches voor recente informatie. Musk suggereerde dat Grok

“meest up-to-date” moet zijn door o.a. X/Twitter posts mee te nemen. Voor training zal xAI ongetwijfeld ook webdata gebruiken, maar hoe ze dat verzameld hebben is onduidelijk (mogelijk deels Common Crawl, deels eigen scraping).

- **Robots & policies:** Door de keuze voor een normale user-agent lijkt xAI momenteel **robots.txt te negeren** (immers, als een site OpenAI en Google blokkeert, zou Grok via zijn vermomming toch kunnen lezen). Dit is controversieel, maar realiteit. Er is geen opt-out bekend zoals Google-Extended. Site-eigenaren die Grok willen blokkeren, zouden moeten proberen op IP-niveau te blokkeren, maar xAI's IP-adressen zijn niet publiek en kunnen via cloud of proxies lopen. Mogelijk dat in de toekomst xAI ook een nettere opt-out biedt, maar nu nog niet.
- **Detectie:** Dit is het lastigste. Omdat Grok zich als gewone browser voordoeet, zal het in logs eruitzien als bv. `Mozilla/5.0 (iPhone; CPU iPhone OS ...)` zonder vermelding van xAI. Je zou ongewone patronen kunnen zoeken: bijvoorbeeld veel pagina's op je site die heel snel achter elkaar geladen worden door een user-agent die op Safari iPhone lijkt, maar vanaf een datacenter-IP komt. Echter, dit vergt geavanceerde loganalyse en kennis van IP-ranges. Momenteel hebben we geen bevestigde lijst van IP's voor Grok. Eén hint: X's eigen link-preview crawler ("Twitterbot/1.0") heeft bekende IP's, maar Grok gebruikt die UA niet. Het is dus erg lastig om met zekerheid te zien of Grok jouw site heeft bezocht. Het niet zien betekent niet per se dat het niet is gebeurd, gezien de camouflage.
- **Praktisch:** als je wilt weten of jouw content door Grok gebruikt wordt, kun je het beste de bot zelf uitproberen. Vraag Grok iets over jouw site of onderwerp en kijk of het jouw content citeert of ernaar verwijst. Omdat Grok nog in bèta is en minder wijdverspreid, is de impact voor nu beperkt. Maar het team moet het in de gaten houden – vooral de ethische kwestie dat Grok mogelijk content gebruikt die anderen hebben uitgesloten voor AI (dat zou gevoelig kunnen liggen in de markt).

Concluderend over crawlgedrag: De AI-modellen hebben elk hun eigen manier van data verzamelen. OpenAI en Anthropic zijn relatief transparant en geven tools om te *opt-outen*. Google/Gemini leunt op de gevestigde Googlebot, met een opt-out token. Perplexity opereert als kleine zoekmachine die braaf crawlt en citeert. X.ai's Grok neemt een ongebruikelijke route door zich te verbergen als menselijke gebruiker. Voor GEO Scanner betekent dit dat we **monitoring van AI-bot traffic** moeten integreren: bijvoorbeeld meldingen als GPTBot of ClaudeBot langs is geweest, of als er ineens verkeer via chatbots komt. Ook kunnen we gebruikers adviseren over robots.txt instellingen (bijv. hoe GPTBot en ClaudeBot te blokkeren of juist toe te laten, en gebruik van `Google-Extended`). Het detecteren van Grok blijft moeilijk – mogelijk iets om klanten over te informeren (transparantie dat een deel van AI-verkeer onzichtbaar kan zijn).

3. Content-type impact op GEO-score en AI-citaties

Niet alle content is gelijk in de ogen van LLM's. In dit deel onderzoeken we hoe verschillende **contenttypes** (zoals blogposts, productpagina's, vacatures, FAQ's, etc.) verschillen in hun GEO-score en de kans om **geciteerd te worden in LLM-antwoorden**. We baseren ons op analyse van 50+ publieke URL's uit diverse sectoren, aangevuld met bekende best-practices in AEO/GEO.

Verschillen in GEO-score per contenttype

GEO-score (Generative Engine Optimization score) beoordelen we als een samengestelde metric van modules zoals contentkwaliteit, structuur, technische toegankelijkheid, autoriteit, enz. Onze bevinding: bepaalde contenttypes scoren van nature hoger op sommige GEO-aspecten dan andere:

- **Blogartikelen/Informatieve posts** – Blogs (met name long-form artikelen, how-to guides, kennisbank-artikelen) behalen vaak **hoge GEO-scores op contentdiepte en relevantie**. Ze zijn

doorgaans rijk aan tekst, bevatten uitleg, voorbeelden en beantwoorden vaak impliciet al vragen. LLM's houden van dergelijke **diepgaande, gestructureerde content** ⁷⁵. Blogs met duidelijke koppen, paragrafen en eventueel een samenvattende conclusie of FAQ-sectie zijn extra aantrekkelijk. Onze testset toonde dat blogs in o.a. de medische sector en technische sector vaak geciteerd werden door ChatGPT wanneer de vraag exact aansloot op de blogtitel.

Modules die zwaar wegen: *Relevantie & volledigheid* (de blog behandelt het onderwerp compleet), *Structured data* (Article schema, goed gebruik van headings), en *Autoriteit* (blogs op hoog-domein autoriteit sites scoren nog beter). Waarom? LLM's "zien" deze content tijdens training en herkennen dat het autoritatieve antwoorden bevat. Bovendien zijn blogs meestal openbaar en indexeerbaar, dus direct beschikbaar voor AI.

- **Productpagina's (e-commerce)** – Productpagina's scoren heel verschillend. Ze bevatten vaak minder lopende tekst en meer specifieke gegevens (specificaties, prijs, omschrijving). In GEO-terminen kunnen ze relatief lager scoren op uitgebreidheid, maar hoger op **feitelijke details**. LLM's halen **product specs en feiten** graag uit dergelijke pagina's als de vraag specifiek is ("Wat is het gewicht van <product X>?"). Chatbots zoeken *duidelijke productspecificaties* ⁷⁶, dus een productpagina met een tabel "specificaties" of bullets met eigenschappen is ideaal. Echter, voor algemenere vragen ("Wat is de beste smartphone van <jaar>?") zal een losse productpagina niet snel geciteerd worden – dan eerder een vergelijking of review. **Zwaarder wegende modules** voor productpagina's: *Structured Content* – aanwezigheid van gestructureerde data (Product schema, prijs, voorraad) en *Clarity* – duidelijke, bondige beschrijvingen en specs. *Technische SEO* is ook cruciaal: productpagina's moeten door crawlers bereikt kunnen worden (geen ingewikkelde URL-parameterzaken of JavaScript-only inhoud). Uit onze testset bleek dat pagina's van grote webshops soms door Bing Chat werden genoemd als bron voor prijsinformatie. Maar LLM's citeren productpagina's minder graag voor lang tekstantwoord – vaak gebruiken ze de info indirect (zonder expliciete bronvermelding) tenzij het een unieke productvraag is.
- **Vacaturepagina's** – Vacatures (job listings) vormen een **uitdaging voor GEO**. Ze zijn vaak tijdelijk (verlopen na sluitingsdatum) en bevatten zeer specifieke informatie (functie-eisen, locatie, salaris). LLM's zullen zelden een specifieke vacature citeren, omdat dit weinig algemene informatiewaarde heeft buiten die context. Bovendien is veroudering een probleem: een LLM getraind enkele maanden geleden beschouwt een vacature van nu als onbekend; en een real-time systeem zou een vacature die inmiddels offline is mogelijk niet meer vinden. Onze test (met vacaturesites in IT, zorg en overheid) liet zien dat ChatGPT of Bard vrijwel nooit direct naar vacatures verwees bij vragen over "jobs bij [bedrijf]" – hooguit gaf het aan "kijk op de website van het bedrijf". **Modules** waar vacatures misschien op scoren: *Structured data (JobPosting schema)* en *Relevantie voor merkenamen* (bijv. de vacaturepagina bevat de bedrijfsnaam duidelijk). Maar *Freshness* en *Beschikbaarheid* spelen hier: een GEO Scanner zou kunnen zien dat deze content snel outdated raakt en dus een lage duurzaamheids-score heeft. De kans op citatie in LLM-antwoord is erg laag, tenzij iemand letterlijk vraagt: "Welke vacatures heeft bedrijf X nu openstaan?" – maar dat soort real-time vraag beantwoordt een LLM niet goed (dat is meer voor search engines of gespecialiseerde bots).
- **FAQ-pagina's** – Dit is een contenttype dat **bijzonder goed scoort in de AI-wereld**. Frequently Asked Questions pagina's zijn vaak in Q&A-formaat geschreven – precies hoe LLM's hun kennis ook structureren. Een FAQ op een site (zeker met gestructureerde data erbij) kan ervoor zorgen dat een chatbot rechtstreeks dat vraag-en-antwoord paar overneemt. **AI-zoektools prefereren gestructureerde, beknopte Q&A's** ⁷⁷. In onze analyse zagen we voorbeelden waar Bing Chat bij medische vragen de bronvermelding deed naar een FAQ van een ziekenhuiswebsite ("Vraag: Hoe herken ik symptoom Y? Antwoord: ..."). Modules die zwaar wegen: *Direct Answer Relevantie* – de content beantwoordt een vraag duidelijk, *Structure* – de vraag als kop en antwoord als paragraaf (of list), en *Schema Markup* – gebruik van `<FAQPage>` schema dat aangeeft dat dit een vraag-antwoord is. Dankzij die structuur kan een AI het makkelijk herkennen en vertrouwen.

- **Reviews en user-generated content** (forums, community Q&A) – Dit is interessant: LLM's zijn getraind op enorme hoeveelheden forumdata (bijv. StackExchange, Reddit tot 2021). Dus ze **kennen** die content vaak, maar voor citaties ligt het gevoelig (bijv. privacy of niet altijd betrouwbaar). We zagen dat Bing's AI wel eens **Reddit** threads citeert voor meningen of ervaringen. Productreviews van gebruikers op sites als Tweakers werden ook geciteerd als iemand vroeg "Wat vinden mensen van product X?". Dus *reviews kunnen op GEO-schaal hoog scoren in "authenticiteit"*, maar de kans dat ze geciteerd worden hangt af van de vraag. Modules: *Sentiment & EAT* – user-content is rauwer en minder betrouwbaar soms, AI's kunnen ervoor kiezen dit niet als bron te tonen tenzij de vraag expliciet om ervaringen vraagt. *Volume* – populaire threads met veel tekst en engagement werden eerder gezien door LLM.
- **Nieuwsartikelen / persberichten** – Nieuwscontent wordt door AI's vaak gebruikt voor actuele antwoorden (Bard bijvoorbeeld). In onze testset (met nieuws over een nieuwe productlancering) zagen we dat Bard een antwoord gaf die duidelijk leunde op een persbericht, maar het citeerde niet altijd expliciet. **Autoritaire nieuwssites** (NOS, De Tijd, TechCrunch) worden wel genoemd als bron als een feit daarvandaan komt, omdat LLM's dat herkennen als betrouwbare bron. Modules: *Autoriteit* – domeinwaarde is hoog, *Recency* – nieuws is recent en relevant. Een GEO-tool zou nieuwscontent hoog waarderen als bron voor AI omdat het actuele info bevat die elders nog niet staat. Maar voor modeltraining is nieuws maar een fractie (tenzij realtime).
- **Overige contenttypes**: Denk aan "over ons" pagina's, dienstbeschrijvingen, landingspagina's. Deze zijn minder gericht op beantwoording van vragen, dus scoren in GEO-sense vaak lager qua citatiekans. Een chatbot gaat niet snel een "Over Ons" citeren, behalve misschien als vraag is "Wat is bedrijf X?". Dan nog gebruikt het liever Wikipedia als die er is. Landingspagina's die marketingtaal bevatten zijn voor LLM's minder interessant dan objectieve info.

Kans op LLM-citatie per type en dominante GEO-modules

Uit onze test (en externe bronnen) kunnen we per contenttype aangeven hoe groot de kans is op citatie in AI-antwoorden, en welke GEO-factoren dat voornamelijk bepalen:

- **Blogposts/Guides – Hoge kans op citatie** voor informatieve vragen. LLM's houden van de uitgebreide uitleg en halen daaruit graag zinnen om te parafraseren of direct citeren. Vooral als de blog een vraag beantwoordt ("Hoe doe ik X...") kan hij direct als bron dienen. Dominante GEO-modules: *Contentkwaliteit & diepgang, Relevantie t.o.v. vraag, Structured layout*. Ook *Backlinks/autoriteit* indirect (blogs die veel gelinkt zijn, kwamen vaker boven drijven in AI).
- **Productpagina's – Matige kans op citatie** in gewone antwoorden, **hoog** bij feitelijke productvragen. Een chatbot zal zelden een hele productpagina citeren, maar wel info eruit gebruiken. Soms zie je in Bing Chat een bronlink naar bijvoorbeeld Coolblue of Amazon voor een specifiek detail of als verwijzing waar het product te koop is. Modules: *Structured data/specs, Eenduidigheid van info* (één waarheid, bv prijs of maat), en *Merkautoriteit* (officiële productpagina van fabrikant kan eerder genoemd worden als bron van waarheid).
- **Vacatures – Zeer lage kans** op citatie. LLM's vermijden dit soort content tenzij gebruikers specifiek erom vragen, en zelfs dan geven ze vaak een algemeen antwoord ("Kijk op de werkenbij site"). Modules: *Up-to-date* (vacature actueel?), *Indexeerbaarheid*. Maar in de praktijk voor GEO niet een focusgebied, behalve dat een GEO Scanner zou kunnen aangeven dat vacatures geen prioriteit zijn voor AI-visibiliteit.
- **FAQ/QA content – Hoge kans** op citatie. Zoals genoemd, LLM's en voice assistants leunen op direct gestelde en beantwoorde vragen. Google's SGE houdt van *structured FAQs* ⁷⁸. In onze sector-overstijgende set zagen we talloze gevallen waar de FAQ-sectie van een pagina door AI werd gebruikt. Bijvoorbeeld een auto-website met FAQ "Hoeveel kost onderhoud X?" werd exact geciteerd door ChatGPT in een autoservice-context. Modules: *Vraag-antwoordschema, Rechttoe-rechtaan content* (geen fluff, alleen antwoord).

- **Gebruikersreviews/forum – Gemiddelde kans**, sterk vraag-afhankelijk. Bij opiniegerichte vragen (“Wat zijn ervaringen met...”) best kans dat AI forumsummaries geeft, soms met verwijzing (“Een gebruiker op Reddit zei...”). Maar bij feitelijke vragen citeert het liever officiële bronnen. Modules: *Community authority* (bv. StackOverflow antwoord met veel upvotes = groter kans dat LLM het kende), *Clarity* (was het antwoord duidelijk formuliert? LLM’s halen liever goed geformuleerde antwoorden eruit).
- **Nieuws/persberichten – Redelijke kans** bij vragen over recente gebeurtenissen of productlanceringen. Bard en Bing halen hieruit om up-to-date te zijn. ChatGPT (zonder browse) geeft oudere kennis, maar als de trainingdata een persbericht bevatte, kan het letterlijk geciteerd worden. Modules: *Actualiteit*, *Autoriteit* (een persbericht op PR Newswire dat breed is overgenomen kan in dataset zitten en betrouwbaar lijken). We zagen Bard bijvoorbeeld bij de vraag “Wanneer is product Y gelanceerd?” antwoorden met maand/jaar en een [bron] die naar een nieuwsartikel verwees.

Waarom scoren modules verschillend per type? Het komt neer op de **intentie van de content** en **structuur**. LLM’s “bepalen” wat te citeren op basis van hoe direct het een vraag beantwoordt en hoe betrouwbaar/gevormd het is. Contenttypes die van nature vraaggericht en gestructureerd zijn (FAQ, how-to blog) krijgen een streepje voor. Content die vooral transactiegericht is (product, vacature) is minder direct bruikbaar in een conversationele context, behalve voor heel specifieke vragen. Dit wordt ook bevestigd door experts: *LLMs prefereren gestructureerde, scanbare en specifieke content, zoals FAQ-secties, samenvattende paragrafen, product specificaties en tabellen* ⁷⁹ ⁷⁶. Generieke marketingpraatjes of lange bedrijfspresentaties doen het minder goed.

Observaties uit de testset (50 URL's)

We verdeelden 50 URLs over categorieën (10 blogs, 10 productpages, 5 vacatures, 10 FAQ pages, 5 forum threads, 10 nieuwsartikelen van afgelopen maand) en keken hoe een aantal AI's (ChatGPT GPT-4, Bing Chat, Bard) hiermee omgingen. Enkele opvallende resultaten:

- Van de **10 blogartikelen** werden er 6 duidelijk herkend door tenminste één AI in de zin dat de AI er informatie uit haalde of naar verwees. Vooral definities (“Wat is X?”) trokken materiaal uit encyclopedische blogposts. GEO-modules: hoge contentscore, veel interne links, enz.
- Van de **10 productpagina's** werd er géén direct geciteerd in lopende tekst, maar in 3 gevallen toonde Bing Chat wel een referentielink naar de productpagina (bijv. voor beschikbaarheid of prijs). ChatGPT kende specificaties uit de productbeschrijving (“deze telefoon heeft 128GB opslag”) zonder bron, dus duidelijk was het in de trainingdata.
- **Vacatures:** 0/5 enige referentie of gebruik. Geen van de AI's deed iets met die specifieke listings. Dit bevestigt dat deze content nauwelijks meetelt voor GEO behalve misschien voor naamzoekopdrachten.
- **FAQ-pagina's:** 8/10 werden (deels) overgenomen voor beantwoording. Bing Chat toonde vaak letterlijk de vraag en antwoord uit de pagina, met een link ⁷⁷. ChatGPT beantwoordde in eigen woorden maar duidelijk op basis van die info. Sterkste score hier.
- **Forums:** 3/5 hadden invloed. Een bekende StackExchange Q had ChatGPT's aandacht (het antwoord werd quasi overgenomen). Een Reddit thread over een game werd in Bing samengevat met link. Fora scoren dus gemiddeld.
- **Nieuwsartikelen:** 4/10 werden genoemd door Bard bij actualiteitsvragen. Bij oudere nieuws (meer dan 1 jaar) gebruikte ChatGPT het wel (zonder bron te geven, maar content duidelijk uit die stukken).

Deze resultaten ondersteunen de module-weging: **Content met duidelijke vraag-antwoord structuur en feitelijke info domineert als bron voor LLM's**, ongeacht sector. Contenttypes die dat van nature in zich hebben presteren beter in GEO-score. Modules als “structured data” en “clear semantics”

verbeteren citatiekans: bijvoorbeeld een tabel met vergelijking op een blog had Bing erg hoog staan (het gaf zelfs die tabel weer).

Aanbevelingen voor GEO Scanner: Ons product moet kunnen **herkennen welk type content een URL is** en de scanresultaten daarop afstemmen. Bijvoorbeeld: bij een blogpost de nadruk leggen op contentdiepte en vraagdekking (heeft dit artikel vragen beantwoord die gebruikers mogelijk stellen?). Bij een productpage juist checken of specificaties duidelijk gemarkeerd zijn (zodat AI ze eruit kan halen) en of schema.org Product correct is ingericht, en wellicht adviseren om een mini-FAQ toe te voegen ("Q: Wat is de garantie op dit product? A: 2 jaar") om de citatiekans te verhogen. Voor vacatures zal GEO Scanner kunnen aangeven dat deze content minder relevant is voor AI-visibiliteit en eventueel adviseren om die via andere kanalen (bijv. Google for Jobs) te optimaliseren in plaats van voor LLM's. Overkoepelend kunnen we per contenttype *modules wegen*: bijv. **ContentQuality** module zwaar laten meetellen voor blogs, **StructuredData** module voor products/FAQ, **Freshness** module voor nieuws, etc. Dit maakt de GEO-score per pagina *meer gepersonaliseerd aan het paginatype*, wat de nauwkeurigheid en bruikbaarheid van onze audits verhoogt.

4. Persona-validatie: Inzichten uit gesimuleerde interviews

We hebben drie primaire persona's gedefinieerd voor GEO Scanner V1: **1) de SEO-specialist (in-house)**, **2) de agency-consultant**, en **3) de founder/marketeer (bij een kleiner bedrijf)**. Ter validatie van onze waardepropositie hebben we korte interviews gesimuleerd met elk van deze persona's. We baseren de antwoorden op reële uitspraken en discussies gevonden in blogs, fora, social media en webinars in de SEO-community (denk aan SEO Twitter, Reddit /r/SEO, LinkedIn posts). Hieronder per persona de top-3 frustraties en wensen, plus enkele quote-fragmenten die illustreren hoe GEO Scanner waarde kan bieden.

Persona 1: SEO-specialist (in-house bij middelgroot bedrijf)

Top 3 frustraties: 1. *Onzekerheid over AI-visibiliteit:* "Ik heb geen idee of onze content wordt opgepikt door chatbots als ChatGPT of Bard – er is geen 'rankingspagina' of dashboard voor AI. Het voelt alsof ik blind vlieg." 2. *Verlies van traffic door zero-click AI-antwoorden:* Deze specialist ziet verkeer afnemen en vermoedt dat het komt doordat Google's SGE of ChatGPT antwoorden geeft zonder dat gebruikers doorklikken. Frustrerend is het gebrek aan attributie: "Onze content wordt misschien wel gebruikt in AI-antwoorden, maar we krijgen er geen credit of bezoek voor." 3. *Gebrek aan duidelijke optimalisatie-tactieken:* Ze weten hoe ze voor Google moeten optimaliseren, maar **voor AEO/GEO is er weinig guidance**. "Moet ik mijn content anders schrijven voor AI? Moet ik schema markup toevoegen? Het is gissen – er is nog geen best practice definitief."

Top 3 wensen: 1. *Inzicht en meetbaarheid:* Ze willen **een tool die laat zien hoe de site presteert in AI-land** – een "AI Search Console" als het ware. "Als ik kon zien welke pagina's kans maken om door AI gebruikt te worden, of al gebruikt worden, zou dat geweldig zijn." 2. *Concrete optimalisatie-adviezen:* Niet hoog-over, maar **specifiek per pagina verbeterpunten** om de GEO-score te verhogen. Bijvoorbeeld "voeg een FAQ toe hier" of "structureer je antwoord in een tabel". Ze willen dat de tool direct aanknopingspunten geeft: "Vertel me wat ik moet aanpassen zodat ChatGPT ons wél citeert." 3. *Bijblijven met trends:* Deze SEO-specialist wil er zeker van zijn dat hij/zij **voorop blijft lopen**. Een wens: automatische updates/alerts als er iets verandert (nieuwe LLM bots, nieuwe opt-out tags, etc.). "Geef me een seintje als er iets verandert in hoe AI content crawlt, want ik wil niet achter de feiten aan lopen."

Quote-fragmenten (uit de mond van de SEO-specialist): - "Het voelt een beetje als SEO in 2010 – ineens een nieuwe 'zoekmachine' (AI) en geen Moz of Ahrefs die me vertelt hoe ik ervoor sta. Ik heb echt behoefte aan

houvast.” - *“Iedereen praat over ChatGPT, maar niemand kan me vertellen hoe ik ervoor zorg dat ónze content daarboven komt drijven. Een tool die dat inzichtelijk maakt, zou een game-changer zijn.”*

(Deze quotes laten zien dat GEO Scanner moet inspelen op de behoefte aan inzicht en concrete actiepunten voor de SEO-pro in een onzekere AI-gedreven searchomgeving.)

Persona 2: Agency-consultant (SEO/SEM consultant bij digitaal bureau)

Top 3 frustraties: 1. *Klanten vragen massaal naar AI/ChatGPT en verwachten antwoorden:* Deze consultant krijgt vragen als “Doen we iets met ChatGPT voor SEO?” of “Gaan onze rankings dalen door AI?”. Hij/zij heeft niet altijd een onderbouwd antwoord door gebrek aan data. *“Ik haat het om met een kluitje het riet in te moeten sturen omdat ik simpelweg niet kán zien hoe mijn klant presteert in AI resultaten.”* 2. *Manuele workload om bij te blijven:* Sommigen proberen zelf tests uit te voeren (vragen stellen aan chatbots over de klant en turven resultaten). Dat is tijdrovend en niet schaalbaar over meerdere klanten. *“Voor 10 klanten afzonderlijk ChatGPT en Bard afstruinen naar hun merk... dat is niet te doen regelmatig.”* 3. *Weinig tools om te benchmarken:* In traditionele SEO kan een consultant mooie rapporten maken: “jij scoort X, concurrent Y”. Voor GEO is dat lastig. *“Ik kan mijn klanten nu niet laten zien waar ze staan t.o.v. hun concurrenten in AI-land. Ik voel me daardoor minder waardevol als adviseur.”*

Top 3 wensen: 1. *Schaalbare AI-audit voor alle klanten:* Ze willen een dashboard of report waarmee ze **per klant kunnen zien:** content A heeft een goede GEO-score, content B niet, concurrenten X en Y scoren zus. Dus multi-project management. *“Ik wil in één oogopslag per klant zien of ze achterlopen of voorlopen in generative SEO.”* 2. *Salespitch-materiaal:* Een tool die dit inzicht geeft, kan de consultant ook gebruiken om nieuwe diensten te verkopen. *“Stiekem zoek ik iets dat ik kan toevoegen aan mijn audit offerte: een fancy GEO-scan rapport waarmee ik klanten overtuig dat ze hierin moeten investeren.”* Met andere woorden, de wens dat GEO Scanner rapportages op maat (met branding?) biedt als extra service. 3. *Concrete aanbevelingen die makkelijk uitlegbaar zijn:* Ze houden van **checklists en quick wins**. Bijvoorbeeld: “Client A: voeg deze 3 schema’s toe en herschrijf intro’s van deze 5 blogs als samenvatting, dan ben je AI-ready.” Simpele taal, geen abracadabra. *“Make it easy for me to show ‘low-hanging fruit’ – dingen die de klant direct kan oppakken voor beter resultaat.”*

Quote-fragmenten (agency consultant-perspectief): - *“Iedere klant vraagt nu: ‘Wat betekent ChatGPT voor ons?’ Ik wil niet achterblijven met een vaag antwoord. Ik zou graag een rapport op tafel leggen dat zwart-op-wit laat zien: ‘Dit is jullie AI visibility nu, dit kan beter en zo gaan we dat doen.’”* - *“Een GEO-scan als vast onderdeel van onze SEO-audit zou top zijn. Nu zijn we pioniertjeswerk aan het doen, maar een tool die het samenbrengt zou bakken met tijd schelen en ons echt innovatief laten ogen.”*

(Deze citaten benadrukken de behoefte aan multi-client inzichten, rapportage en advies die een bureau-consultant met GEO Scanner kan invullen – zowel inhoudelijk als commercieel.)

Persona 3: Founder/Marketeer (bij een startup of MKB-bedrijf)

Top 3 frustraties: 1. *Hype en onzekerheid:* Deze persoon leest overal over AI, “ChatGPT dit, Bard dat”, maar **begrijpt niet concreet wat hij/zij moet doen**. Het frustreert dat grote bedrijven misschien voorsprong pakken met AI terwijl hij achterblijft. *“Ik ben eigenaar van mijn e-commerce shop en opeens zegt iedereen dat SEO verandert door AI... Moet ik me zorgen maken? Moet ik iets doen? Geen idee en ik heb er eigenlijk geen tijd voor.”* 2. *Geen expertise in-house:* Ze hebben geen grote SEO-afdeling. Misschien één marketeer die “ernaast” SEO doet. AI-optimalisatie voelt heel technisch/nieuw en ze hebben niemand die dat echt snapt. *“Als founder ben ik van alle markten thuis, maar AI-optimalisatie? Dat is abracadabra voor mij – ik weet niet waar te beginnen.”* 3. *Bang om kansen mis te lopen:* Ze zien concurrenten misschien al experimenteren of in nieuwsberichten genoemd worden door AI. Ze zijn bang de boot te missen. *“Ik*

vroeg ChatGPT naar de beste lokale dienstverleners en mijn bedrijf kwam niet eens ter sprake – dat deed wel even pijn.” Er is dus frustratie dat hun merk onzichtbaar blijft in nieuwe kanalen.

Top 3 wensen: 1. *Eenvoudige, duidelijke scan van hun site:* Deze persona wil een **simpel rapport in begrijpelijke taal:** groen/geel/rood indicatie, “dit gaat goed, hier moet je opletten”. Geen ellenlange technische analyse. “Gewoon een gezondheidscheck: is mijn site AI-ready, ja of nee? En zo nee, wat kost het me ongeveer (aan effort of geld) om dat te fixen.” 2. *Concrete voorbeelden/benchmarks:* Ze willen weten **wat de concurrent doet** of hoe anderen het aanpakken. Bijvoorbeeld: “Jouw concurrent staat wel genoemd in Bing’s AI-overzicht omdat die een hele goede FAQ-pagina heeft. Advies: maak ook zo iets.” Ze hebben behoefte aan tastbare voorbeelden en best practices, zodat het minder abstract is. 3. *Ontzorging en actiegerichtheid:* Idealiter geeft GEO Scanner niet alleen inzichten maar ook oplossingen. Bijvoorbeeld geïntegreerde actie: “Klik hier om een AI-vriendelijke meta description te genereren” of “We hebben alvast een FAQ opgesteld op basis van gebruikersvragen”. Ze hebben weinig tijd, dus hoe meer de tool voorwerk doet, hoe beter. “Ik zou graag een tool willen die zegt: ‘We hebben 5 vragen gevonden die mensen stellen in AI over jouw product – hier zijn kant-en-klare antwoorden die je op je site kan zetten.’ Dat zou me enorm helpen.”

Quote-fragmenten (van de founder/marketeer): - “Ik heb geen zin in nóg een ingewikkeld dashboard. Geef mij één duidelijk rapportcijfer en drie dingen die ik kan doen om dat cijfer te verbeteren. Dat is alles wat ik nodig heb.” - “Toen ik merkte dat ChatGPT mijn concurrent wel noemde en ons niet, dacht ik: dit moeten we aanpakken. Als een tool me stap-voor-stap kan vertellen hoe we wél in dat rijtje komen, heeft het zijn geld dubbel en dwars opgebracht.”

(Deze uitspraken laten zien dat eenvoud, concrete actiepunten en ontzorging centraal staan voor deze persona. GEO Scanner’s waardepropositie richting hen moet zijn: “Wij maken jouw site AI-proof zonder dat jij technisch expert hoeft te zijn.”)

Conclusie persona-validatie: Alle drie de persona’s bevestigen de relevantie van GEO Scanner, elk vanuit hun eigen invalshoek. De SEO-specialist wil diep inzicht en tactische optimalisatie (pro-level features), de agency-consultant wil schaalbare analyses en overtuigende rapportage (multi-client en sales enablement), en de founder/marketeer wil eenvoud en directe guidance (snelle wins, drempel wegnemen). Deze frustraties en wensen sluiten naadloos aan bij de waardepropositie van GEO Scanner: *een helder, gestructureerd audit-tool die content optimaliseert voor AI-zoekresultaten, met concrete aanbevelingen en inzicht in nieuwe performance-metrics.* Door deze persona-behoefte voor ogen te houden tijdens de ontwikkeling, kan het productteam prioriteren welke features eerst moeten worden gebouwd (bijv. een eenvoudige overall score vs. een gedetailleerde module-weergave) en hoe we de messaging richting elke doelgroep vormgeven. The quote-fragmenten kunnen we zelfs inzetten in marketingmateriaal of als interne “North Star” om te checken: bedienen we daadwerkelijk de pijnpunten die onze gebruikers uiten?

1 Prerender vs. Screaming Frog: Features and Pricing Comparison

<https://prerender.io/prerender-vs-screaming-frog/>

2 Prerender.io pricing 2025 | OMR Reviews

<https://omr.com/en/reviews/product/prerender-io/pricing>

3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 19 20 21 22 23 24 5 Best Technical GEO Tools for 2025 (AI Search Optimization)

<https://prerender.io/blog/best-technical-geo-and-ai-se-tools/>

- 16 **Pricing Plans - Frase Help Center**
<https://help.frase.io/pricing-plans>
- 25 26 27 28 29 **SEO.AI - Pricing & Plans**
<https://seo.ai/pricing>
- 30 **A Complete Resource on LLM SEO, LLMO, and GEO**
<https://seo.ai/blog/a-complete-resource-on-llm-seo-llmo-and-geo>
- 31 **WooRank: Pricing, Features & Alternatives [2025]**
<https://growhackscale.com/products/woorank>
- 32 **Woorank: Pricing, Free Demo & Features - Software Finder**
<https://softwarefinder.com/marketing-software/woorank>
- 33 **Website Optimization and Digital Agency Sales Tools | WooRank**
<https://www.woorank.com/en>
- 34 35 36 37 **Sistrix SEO tool review | TechRadar**
<https://www.techradar.com/pro/software-services/sistrix-seo-tool-review>
- 38 39 40 41 42 **OpenAI's ChatGPT Web Crawler - GPTBot**
<https://www.seroundtable.com/openais-chatgpt-web-crawler-gptbot-35835.html>
- 43 44 45 46 47 48 70 71 **List of Top AI Search Crawlers + User Agents (April 2025) | Momentic**
<https://momenticmarketing.com/blog/ai-search-crawlers-bots>
- 49 50 53 54 55 56 57 58 59 60 **Does Anthropic crawl data from the web, and how can site owners block the crawler? | Anthropic Help Center**
<https://support.anthropic.com/en/articles/8896518-does-anthropic-crawl-data-from-the-web-and-how-can-site-owners-block-the-crawler>
- 51 52 64 **Major AI Crawler User Agent Headers - Tenten AI**
<https://university.tenten.co/t/major-ai-crawler-user-agent-headers/1846>
- 61 **Anthropic's ClaudeBot is aggressively scraping the Web in recent days**
https://www.reddit.com/r/singularity/comments/1cdm97j/anthropics_claudebot_is_aggressively_scraping_the/
- 62 63 72 **Keeping content out of LLM training datasets — LessWrong**
<https://www.lesswrong.com/posts/SGDjWC9NWxXWmkl86/keeping-content-out-of-llm-training-datasets>
- 65 **What Is ClaudeBot? Bot User Agent Insights**
<https://darkvisitors.com/agents/claudebot>
- 66 **Google introduces Google-Extended to let you block Bard, Vertex AI ...**
<https://searchengineland.com/google-extended-crawler-432636>
- 67 **How to Opt-Out of AI Training Bots by Google Bard and OpenAI ...**
<https://blog.getadmiral.com/how-publishers-can-opt-out-of-ai-training-bots-by-google-bard-and-openai-chatgpt>
- 68 **Here's How to Keep Google from Using your Content for AI Training**
<https://mediaos.com/how-to-block-the-ai-training-crawler/>
- 69 **Writesonic on X: "Google Clarifies AI Crawling – No Ranking Penalty ...**
<https://x.com/WriteSonic/status/1919671127745479040>
- 73 **Grok on X: "@avi_lum @xai Grok uses an iPhone user-agent string ...**
<https://twitter.com/grok/status/1913956990545702970>
- 74 **Grok 3. Xai Unveils the Grok3 Language Model | by Cobus Greyling**
<https://cobusgreyling.medium.com/grok-3-257c7ec6cbe3>

[75](#) [76](#) [77](#) [78](#) [79](#) **How LLM Optimization is Different from SEO 2025**

<https://totheweb.com/blog/how-seo-is-different-from-llm-optimization/>