

PRINTABLE CHECKLIST

GEO self-audit checklist

Tick each box, then run the command under it. Updated October 2026.

Access

- ☐ Live /robots.txt returns 200 text/plain
`curl -sI https://yourdomain.com/robots.txt`
- ☐ Named AI groups repeat every Disallow rule from the * group
A named group replaces the * group for that bot
- ☐ Search bots (OAI-SearchBot, Claude-SearchBot, PerplexityBot, Googlebot, bingbot) are allowed in robots.txt
`python3 scripts/rfc9309_check.py robots.txt OAI-SearchBot /blog/post`
- ☐ Training bot decision made and written down (GPTBot, ClaudeBot, Google-Extended)
Training and search are separate tokens
- ☐ CDN and WAF allow the bots' published IP ranges
`python3 scripts/verify_bot_ip.py <ip>`
- ☐ Cloudflare managed robots.txt and the Vercel AI Bots ruleset are checked
- ☐ Managed robots.txt on your CDN checked against the live file
`diff <(curl -s https://yourdomain.com/robots.txt) public/robots.txt`

Render

- ☐ View-source or curl shows main text, title, canonical and JSON-LD without JavaScript
`scripts/check-ai-crawl.sh https://yourdomain.com/page "a sentence"`
- ☐ Key answers are not only in tabs, accordions, images or PDFs
- ☐ Links are real a href elements
`curl -s URL | grep -o "<a [^>]*href"`
- ☐ Important pages return 200 with no noindex, nosnippet or max-snippet:0
`curl -sI URL | grep -i -E "^HTTP/|x-robots-tag"`

Content

- ☐ The first 1 to 2 sentences under each heading answer it
- ☐ One topic per URL
- ☐ Numbers have a linked primary source
- ☐ The page contains something only you can say (first-hand detail)
- ☐ No keyword stuffing, no hidden text for LLMs

Structured data and entity

- ☐ Organization JSON-LD on the home page with sameAs
- ☐ JSON-LD matches visible text and validates
Check the page with Google's Rich Results Test
- ☐ Same business name and description on your site, Business Profile, Bing Places and Merchant Center
- ☐ Author pages with name and url

URLs and freshness

- ☐ Canonical, sitemap loc and final redirect target use the same host and URL
`curl -sI https://example.com/page | grep -i -E "^HTTP/|^location"`
- ☐ Sitemap lastmod reflects real content changes only
- ☐ Visible Last updated date plus dateModified in JSON-LD
- ☐ IndexNow key file is live and pings return 200 or 202 (Bing and other participants, not Google)
`bash indexnow/ping.sh`
- ☐ Sitemap submitted in Search Console and Bing Webmaster Tools

Measure

- ☐ Search Console, Settings, Search generative AI is not set to Exclude
- ☐ Generative AI performance report baseline saved
- ☐ Annotation added for the change date
- ☐ GA4 AI Assistant channel or a custom channel group checked
`measure/ga4-ai-sources.regex`
- ☐ Bing AI Performance baseline exported
- ☐ Bot hits verified in logs with an IP check
`python3 scripts/ai_bot_report.py access.log`
- ☐ 20 to 40 prompts logged, 3 runs per engine
`measure/ai-prompt-tracker.csv`

Optional: llms.txt

- ☐ `/llms.txt` returns 200 text/plain or text/markdown, one generator only, linked from footer or head
`curl -sI https://yourdomain.com/llms.txt`

Full guide: www.maxenius.com/blog/generative-engine-optimization | Starter kit: www.maxenius.com/downloads/geo-starter-kit/geo-starter-kit.zip