webrepublic

Tech SEO & GEO Checklist:
Visibility in AI in Search

This checklist shows you the most important technical steps to ensure that your website can be reliably captured, understood and used for generative responses by search engines and LLMs..

1. Crawling & Indexing Management

Make it as easy as possible for search engines and AI crawlers to discover the central pages of your website. If bots can find your most important content quickly, you can use your crawl budget more efficiently.

To Do	Important to Keep in Mind
Check robots.txt	Ensure that robots.txt only blocks irrelevant areas (e.g. admin, scripts) and not important CSS or JS resources.
Keep your XML sitemap up to date	It should only contain indexable, relevant URLs and be submitted in Google Search Console and Bing Webmaster Tools.
Optimise internal linking	Important pages should be accessible with a maximum of three clicks.
Check link hygiene	Internal links should not lead to 404 error pages or unnecessary redirect chains.
Be precise with setting noindex tags	All pages without added value for search (filters, internal search, login) should be consistently excluded from indexing using noindex.
Check canonical tags	In case of duplicate content (e.g. due to URL parameters), the canonical tag should be set to the preferred main URL in order to bundle link power.
Ensure HTTPS security	The entire website should be delivered via a secure HTTPS connection.

2. Rendering & Structure Optimisation (Semantics for LLMs)

Ensure that your content is clearly structured, logically organised and directly visible in HTML so that LLMs can reliably capture and correctly interpret it.

To Do	Important to Keep in Mind
Important content in the initial HTML	Ensure that the core content (H1, main text, links) exists directly in the HTML source code. Important information should not be loaded dynamically via JavaScript.
Clear heading hierarchy	There should only be one H1 tag per page. The H2 and H3 structure should organise the content clearly and logically and not be used for purely visual design reasons.
Use semantic HTML5	Use HTML5 tags such as **<main>, <article>**, **<nav>** and **<aside>** correctly so that the main content is clearly separated from the boilerplate code.
Define a rendering strategy	For pages that are heavily JavaScript-dependent, you should rely on server-side rendering (SSR), pre-rendering or dynamic rendering.
Clean up the URL structure	The URL structure should be flat, consistent and easy to read, for example **/blog/tech-seo-geo/** instead of complex parameter variants.

3. Performance & Timeliness

Fast, stable pages are preferred by crawlers and visited more often. Anything that saves loading time helps your visibility.

To Do	Important to Keep in Mind
Core Web Vitals are strong	LCP (load time), CLS (layout shift) and INP (interactivity) should achieve good values, as they directly influence load time and user experience.
Check server response times (TTFB)	Ensure that the TTFB (Time to First Byte) is stable and low. High or unstable TTFB leads to timeouts and crawling being aborted by LLM bots
Compress and optimise images	Compress images and use modern formats (e.g. WebP) and, where appropriate, implement lazy loading to reduce load times.
Accelerate indexing updates	For time-sensitive content, use indexing APIs (such as IndexNow) to report updates to search engines and AI bots in near real time.

4. Structured Data & AI Crawler Readiness

Help LLMs and AI crawlers understand your content by providing clear facts and clean, structured data, and ensure that LLM-Bots can actually crawl your content. This increases the chance that your content will appear in generative responses.

To Do	Important to Keep in Mind
Implement structured data	Include relevant markups such as Article, FAQ or Organisation via JSON-LD. They help search engines to clearly recognise and correctly classify your content. They also play a role in visibility in LLMs.
Coordinate markup and content	Make sure that the structured data corresponds exactly to what is visible on the page.
Use Schema Validator	Check whether your structured data passes the Rich Results Test or the Schema Validator without critical errors.
Whitelist AI bots in firewalls/CDNs	Ensure that firewall or CDN rules (e.g. at Cloudflare) do not block LLM crawlers such as GPTBot or PerplexityBot, provided they are allowed to crawl your content.
Perform log file analysis	Actively monitor log files to see which search engine and LLM crawlers are visiting your pages and whether there are any crawl or access barriers
Exclude specific AI bots (optional)	If you do not want certain content to be used for training individual LLMs, you can specifically exclude their user agents in robots.txt or via server rules