

The AI-visibility page checklist

The AI-visibility page checklist is the exact set of checks our free checker runs on a page — all 26 of them, in the five weighted groups that make up the score, each with what it means and the fix. This is the whole rubric, not a preview of it.

Also known as a GEO audit rubric or an AEO checklist. It grades one page; the layers above a page are in the field guide (mentionbeat.com/guide).

How the score is put together

Five categories carry fixed weights. Inside each one, a critical check counts three times a minor check and an important check counts twice — so a category is never carried by clearing the cosmetic items while a gate is broken.

GROUP	WEIGHT	CHECKS	WHAT IT ANSWERS
Structured data	30%	6	Can a machine tell what this page is about, unambiguously?
Answer-first structure	30%	7	Is there a passage worth quoting, near the top, that stands alone?
Entity grounding	15%	3	Can an engine connect this brand to the entity it already knows?
Metadata	15%	5	Are the invisible basics present and correctly sized?
Crawlability	10%	5	Can an engine reach and index the page at all?

A few checks need a live fetch — the status code, robots.txt and llms.txt. When a page is pasted rather than fetched, those are reported as "not assessed" and pass, so the rubric stays comparable between runs rather than punishing a page for how it was submitted.

30% OF THE SCORE · 6 CHECKS

Structured data

Structured data is not a ranking lever, and Google says it is not required for AI features. It is something better: the cheapest way to hand a machine unambiguous facts about what this page is and what it describes.

JSON-LD structured data present CRITICAL jsonld-present

Answer engines lean on schema.org JSON-LD to identify the entity, its facts and its relationships. A page with none makes them guess.

Fix: add a `<script type="application/ld+json">` graph describing the page entity — Organization, Product, Article or Service — plus the WebPage itself.

JSON-LD parses cleanly CRITICAL jsonld-parses

A block that fails to parse is ignored entirely. Invalid structured data is worth exactly as much as none, while looking like work was done.

Fix: validate every block and repair the syntax error. A trailing comma is the usual culprit.

A primary entity type is declared IMPORTANT primary-entity

Without a recognised type, nothing tells an engine what *kind* of thing this page is about.

Fix: declare one of Organization, Product, Article, Service, WebSite, WebPage, SoftwareApplication or LocalBusiness, with a clear name.

FAQ content is marked up as FAQPage IMPORTANT faq-schema

Question-shaped sections are the most liftable content on a page. Unmarked, an engine has to infer that they are questions at all. Pages with no Q&A content pass — there is nothing to mark up.

Fix: wrap the visible questions and answers in FAQPage JSON-LD that mirrors the visible text exactly.

Recency signal present MINOR date-modified

Engines favour fresh sources. With no `dateModified` there is nothing to signal recency with.

Fix: add `dateModified` to the page's JSON-LD and keep it honest — bumping it without a substantive change is a low-trust signal and it ruins your own before-and-after readings.

Content is recently updated MINOR freshness-recent

Citation priority decays once a page goes stale. This passes when `dateModified` falls within roughly the last six months and is not in the future.

Fix: refresh substantive pages on a real cycle and let the date follow the work.

30% OF THE SCORE · 7 CHECKS

Answer-first structure

Engines retrieve passages, not pages, and they weight the opening heavily. This group asks whether there is a self-contained block near the top that an engine could quote without editing it.

Exactly one H1 IMPORTANT single-h1

No H1 leaves the page without a stated main topic; several blur it for a parser.

Fix: use one H1 stating the page's subject, and demote the rest to H2 or H3.

Opens answer-first **IMPORTANT** definition-first

A page that does not open with a direct "*X is a ...*" definition offers no clean snippet to lift, so an engine writes its own — from whatever else it found.

Fix: start with a one-sentence definition: "The *subject* is a *category* that *does what*."

The lead names the subject in the first two sentences **IMPORTANT** answer-first-lead

Models weight the first sentences most heavily. Naming the subject there, inside a focused block, makes the page far likelier to be quoted verbatim.

Fix: open with a block of roughly 50 to 120 words that names the subject in the first sentence and answers the core question directly.

FAQ-style content present **MINOR** faq-content

Question-led sections map directly onto how people prompt an assistant, which makes them disproportionately citable.

Fix: add a short FAQ answering the questions buyers actually ask, each answer self-contained.

Scannable structure **MINOR** scannable

Dense prose with no list or table is harder to extract facts from — for a reader and for a parser.

Fix: put specs and feature facts in a real table or a bulleted list.

Enough static content **IMPORTANT** sufficient-content

Under about 150 words in the raw HTML, there may be nothing to retrieve. If the content is rendered client-side, most engines will not see it at all.

Fix: make sure the substantive content is server-rendered rather than injected by JavaScript.

Facts are not locked inside PDFs **IMPORTANT** pdf-only-facts

The datasheet pattern: the page links to PDFs, carries no table and almost no numbers of its own. Most engines retrieve PDFs poorly or not at all, so those facts are effectively invisible.

Fix: mirror each PDF's key facts as real HTML — a table on the page or a linked HTML spec page. Keep the PDF as the formal document, not the only copy.

15% OF THE SCORE · 3 CHECKS

Entity grounding

The most durable work on this list, and the most often skipped. It connects the brand on this page to the entity an engine already has an opinion about.

Entity-grounding links present **IMPORTANT** org-sameas

Without `Organization.sameAs`, an engine cannot reliably tell your brand apart from a similarly named one.

Fix: add Organization JSON-LD with `sameAs` pointing at Wikidata, Wikipedia, LinkedIn and your official profiles.

Alias line present **MINOR** alias-line

An "also known as" line carrying alternate names, abbreviations and model numbers helps an engine match the way people actually refer to you — including the name you used before the rebrand.

Fix: add one sentence listing the alternates, and mirror them as `alternateName` in the structured data.

No unsupported superlatives **MINOR** no-hype

Unverifiable superlatives get discounted, and they crowd out the specific claims that would have been quoted instead.

Fix: replace each superlative with a specific, number-backed claim.

15% OF THE SCORE · 5 CHECKS

Metadata

The invisible layer. No reader sees any of it, and every one of these is an hour's work at most.

Title tag present **IMPORTANT** title

The title is the first thing that names the page's subject to a machine.

Fix: add a descriptive `<title>` naming the subject, answer-first.

Meta description present and well sized **IMPORTANT** meta-description

Passes between 50 and 170 characters; aim for 120 to 160. Too short says nothing, too long gets cut.

Fix: summarise the page answer-first in about 155 characters.

Canonical URL set MINOR canonical

Without one, duplicate URLs split the signals that should have accumulated on a single page.

Fix: add `<link rel="canonical">` with the page's preferred URL.

OpenGraph title present MINOR og-title

Preview metadata decides what a shared link looks like — which decides whether it gets clicked at all.

Fix: add `og:title` and `og:description`.

Images have alt text MINOR img-alt

Missing alt text hides an image's meaning from engines and from assistive technology alike.

Fix: add descriptive alt text to every meaningful image, and restate any in-image number in nearby prose.

10% OF THE SCORE · 5 CHECKS

Crawlability

The smallest weight and the hardest gate. A page that fails here cannot be cited whatever the rest of the score says, which is why one of these checks is critical.

The page is indexable CRITICAL indexable

A robots meta tag telling engines not to index the page ends the conversation — it cannot be cited however well optimised it is.

Fix: remove the `noindex` directive if you want this page cited.

Served with a healthy status IMPORTANT http-ok

A non-2xx status means engines may never fetch or index the page. Reported as not assessed when the HTML was pasted rather than fetched.

Fix: make sure the page returns 200 — repair the redirect chain or the error.

robots.txt allows this path IMPORTANT robots-allowed

If robots.txt blocks the path, engines cannot fetch it. Worth separating the two decisions: blocking answer agents removes you from AI answers, while blocking bulk training crawlers is a licensing choice.

Fix: allow the path for the crawlers you want reaching the page. Rules to copy are in the templates pack (mentionbeat.com/templates), and the access checker (mentionbeat.com/ai-access) tests each agent live.

Canonical points at this page MINOR canonical-self

A canonical pointing elsewhere hands the credit to that page instead. Sometimes deliberate; usually a template left on the wrong setting.

Fix: point the canonical at this page's own URL unless you genuinely want another page to win.

The site publishes /llms.txt MINOR llms-txt

An emerging convention for pointing answer engines at your most citable pages. Cheap insurance, not a lever — it is a minor check here for exactly that reason.

Fix: publish one listing your important pages. Our [generator](https://mentionbeat.com/tools/llms-txt) (mentionbeat.com/tools/llms-txt) builds it from your sitemap.

What this checklist cannot tell you

Every check above grades one page against what an engine can read. None of them can see whether an assistant actually names you when a buyer asks, which depends on your entity foundation, on independent corroboration, and on how well the page answers that specific question. A page can score 100 and still never be mentioned.

That is the honest division of labour: the checklist and the [free checker](https://mentionbeat.com/checker) (mentionbeat.com/checker) answer "is something wrong with this page?". Answering "are we in the answer, and did the fix move it?" needs repeated sampling across engines over time — which is the [paid product](https://mentionbeat.com/product) (mentionbeat.com/product), and we would rather say so plainly than dress a page score up as a visibility score.