RESOURCE · VERSION 1.0

AI Citation Readiness Checklist

By **Martin Zialcita** · AI visibility · Checklist · Published July 31, 2026 ·
Reviewed July 31, 2026

WHAT THIS ASSESSES

Forty checks across five dimensions: whether your pages can be crawled, whether they answer questions directly, whether authorship and evidence are attributable, whether claims are corroborated off-site, and whether content is maintained. Each scores one point. It measures positioning, not outcomes. No checklist can promise citations, because no answer engine publishes its ranking mechanics.

54-word direct answer

KEY TAKEAWAYS

- Score is out of **40**. Dimension one is a precondition: failing it caps everything else, whatever the total says.
- This measures **positioning**, not results. Anyone promising citations from a checklist is overclaiming.
- The methodology section states how each check was derived, so you can disagree with a specific one.
- Version 1.0, reviewed annually. Cite the version; the checklist will change as provider documentation does.

⚠ WHAT THIS CANNOT DO

No answer engine publishes its retrieval or citation-selection mechanics. Any checklist in this area (including this one) is built from provider documentation, observable behavior, and inference. It is a positioning instrument, not a predictive one.

A high score means you have removed the known obstacles and met the documented requirements. It does not mean you will be cited, and a vendor who tells you otherwise is selling something they cannot deliver.

How to score it

Work through the five dimensions in order. Each check is worth **one point**: met or not met, with no partial credit, because partial credit is where self-assessment becomes flattering.

Score against a **specific page**, not the site in general. Most sites have a handful of genuinely citable pages and a long tail that is not, and averaging across them hides exactly the information you need. Run it on the page you most want cited, then on a second page for comparison.

A check is met only if you could show someone. “We intend to” and “it is on the roadmap” score zero. So does “our developer says it is handled” without your having seen it.

The five dimensions

In order. The ordering is not arbitrary: dimension one gates everything after it.

#	DIMENSION	QUESTION IT SETTLES	POINTS
1	Retrievable	Can the page be crawled, indexed, and read as text?	9
2	Answerable	Does it answer a specific question in a form that can be quoted?	9
3	Attributable	Can a reader tell who wrote it, when, and on what basis?	8
4	Corroborated	Does anything outside your own domain independently support it?	7
5	Maintained	Is it current, and is currency demonstrable?	7

Forty points total. Dimension one is a gate rather than a contributor: a page scoring below 7 there is unlikely to be retrieved at all, and its score on the other four dimensions is academic.

Dimension 1 · Retrievable (9 points)

The preconditions. Google states that its AI features use the same foundational requirements as Search, so this dimension is ordinary technical SEO hygiene, and it is where most failures actually sit.

- [ ] **The page returns HTTP 200 at a stable canonical URL**
  Not a redirect chain, not a soft 404, and the canonical points to itself.

- [ ] **The substantive text is in the served HTML**
  View source and find the actual sentences. If the content arrives only after JavaScript executes, you are relying on rendering you do not control.

- [ ] **The page is not blocked in robots.txt**
  Check the specific path, not just the root, and check it against the crawlers you care about rather than only the wildcard.

- [ ] **Search and answer-engine crawlers are permitted**
  Search access and model-training access are separate decisions. You can allow the former while declining the latter, but if you have blocked search crawlers, nothing else on this list matters.

- [ ] **The CDN or WAF does not silently block verified crawlers**
  Bot-mitigation defaults frequently do. Verify against published user agents and IP ranges rather than assuming.

- [ ] **The page is in a submitted XML sitemap**
  And the sitemap is registered with Google Search Console and Bing Webmaster Tools.

- [ ] **Indexing is confirmed, not assumed**
  Check URL Inspection in Search Console and Bing. Submitted is not indexed.

- [ ] **The page loads fast enough not to be a problem**
  Core Web Vitals passing on mobile. This is a hygiene check, not a ranking claim.

- [ ] **Headings are semantic and hierarchical**

  One h1, no skipped levels. Machine extraction of structure depends on it, and so does every screen reader.

Dimension 2 · Answerable (9 points)

Whether the page contains something that can be lifted and remain accurate. This is the dimension most content strategies neglect in favor of length.

- [ ] **The page answers one clearly scoped question**
  Not a topic. If you cannot state the question in a sentence, neither can a retrieval system.

- [ ] **A direct answer appears near the top, in 40–80 words**
  Before context, before background, before the narrative. Retrieval favors passages that stand alone.

- [ ] **The answer is accurate out of context**
  Read it in isolation. If it becomes misleading without the paragraph beneath it, it is not yet a quotable unit.

- [ ] **Two to four passages could each be quoted standing alone**
  Labeled definitions, principles, or distinctions. Not manufactured pull-quotes attributed to a person who did not say them.

- [ ] **Descriptive H2 and H3 headings segment the content**
  A heading should tell you what the section answers, not merely name a theme.

- [ ] **Key terms are defined where first used**
  Without sending the reader elsewhere for the meaning.

- [ ] **Concrete specifics appear rather than generalities**
  Named systems, actual sequences, real constraints. Generic advice is unquotable because it is interchangeable.

- [ ] **Limitations or boundaries are stated**
  Where this does not apply. Counterintuitively this makes a passage more citable, because it is safer to quote.

- [ ] **Visible text matches any structured data**

  Markup describing content the reader cannot see is a violation of every provider's guidance.

Dimension 3 · Attributable (8 points)

Whether a reader (or a system deciding whether to rely on you) can establish who is speaking and on what basis.

- [ ] **A named human author is visible on the page**
  Not “the team”, not “admin”, not absent.

- [ ] **The author links to a substantive profile**
  One that establishes why this person can speak to this subject, with checkable specifics rather than adjectives.

- [ ] **A published date is visible**
  And it is the real one.

- [ ] **A reviewed date is visible where claimed**
  And it reflects an actual human review, not a rebuild. A “last updated” that tracks deployments is worse than none.

- [ ] **First-hand experience is distinguished from interpretation and third-party evidence**
  So a reader knows which parts are observation, which are inference, and which are someone else's finding.

- [ ] **Quantitative claims carry their method**
  Baseline, sample, measurement window, attribution. A percentage without these carries no information.

- [ ] **A correction route exists and is findable**
  An editorial policy stating how errors are reported and how material changes are recorded.

- [ ] **The entity behind the site is unambiguous**
  Consistent name, title, and organization across the site and its structured data, with a stable identifier.

Dimension 4 · Corroborated (7 points)

Whether anything outside your own domain supports what you assert. This is the dimension you have least direct control over, and the hardest to shortcut.

- [ ] **Claims link to primary sources**
  Positioned next to the claim they support, not collected in a footer nobody reads.

- [ ] **At least one credible third-party page independently confirms your role or work**
  An institutional page, an event listing, a byline, a named client reference. Something you did not publish.

- [ ] **Sourcing is not circular**
  Your profiles citing only each other establishes nothing. §11 names this explicitly.

- [ ] **Name, title, and organization are consistent everywhere**
  Across your site, your profiles, and third-party listings. Inconsistency degrades entity confidence.

- [ ] **External profiles in your structured data are live and current**
  A `sameAs` list of stale or dead profiles is worse than a short accurate one.

- [ ] **No unverifiable superlatives appear**
  “Leading”, “premier”, “top-rated” without a source are unfalsifiable and read as noise.

- [ ] **Claims you cannot yet support are withheld rather than softened**
  Hedging an unsupported claim keeps the claim. Removing it is the honest move.

Dimension 5 · Maintained (7 points)

Whether the page is alive. Freshness matters most for anything describing tools, provider behavior, or a fast-moving field.

- [ ] **A review cadence is defined and documented**
  Evergreen pages at least every six months; guidance on crawlers or provider behavior quarterly.

- [ ] **A named person owns the review**
  A cadence with no owner is a hope.

- [ ] **Material changes are recorded, not applied silently**
  An update log, so a reader can see what changed and when.

- [ ] **Anything version-dependent states its version**
  Frameworks, checklists, and tool guidance all drift. A citation without a version becomes ambiguous.

- [ ] **Dead outbound links have been checked recently**
  Broken sources undermine the corroboration in dimension four.

- [ ] **The page has been re-read since the field moved**
  Provider documentation changes. A page describing last year's crawler behavior as current is actively misleading.

- [ ] **Retired content is retired deliberately**
  Removed or marked superseded with a reason, rather than left to rot at a live URL.

Interpreting your score

Bands rather than a precise grade. The dimension-one gate matters more than the total.

SCORE	READING	WHAT TO DO NEXT
Below 7 in dimension 1	Not retrievable. The total is irrelevant.	Fix retrieval first. Nothing else has any effect until a crawler can read the page.
0–15	Not yet positioned to be cited.	Work dimensions 1 and 2. Direct answers and clean retrieval are the highest-leverage fixes available.
16–25	Retrievable and partly answerable, weak on trust.	Dimension 3. Named authorship, real dates, and stated method are usually quick wins.
26–33	Well positioned on your own domain.	Dimension 4. The remaining constraint is almost always off-site corroboration, which takes longest.
34–40	Strong positioning.	Move to dimension 5 and hold it. At this point the variable is publishing consistency, not further optimization.

A score is a snapshot of one page. Re-scoring the same page in six months is more informative than scoring ten pages once.

INTERPRETATION

Where the points usually go missing

A reading rather than a measurement: in practice most sites lose points in dimensions two and three, not one. Technical retrieval is usually adequate because it overlaps with ordinary SEO work that somebody has already done.

What is typically absent is a direct answer near the top (pages open with context and arrive at the answer in paragraph six) and visible attribution, because content was published under a brand rather than a person.

Dimension four is where genuine effort is required and where no shortcut exists. Off-site corroboration cannot be manufactured on your own domain, which is precisely why it carries signal. Anyone selling a way to fabricate it is selling a link scheme with a new name.

Methodology

§7.10 requires this checklist to publish its methodology, so here it is, including its weaknesses.

How the checks were derived. Three inputs. First, requirements stated publicly by providers: Google's documentation that AI features use the same foundational requirements as Search and that no special AI schema or machine-readable file is required; OpenAI's and Anthropic's crawler documentation distinguishing search access from training access. Second, general information-retrieval principles that predate current AI search: extractability, passage independence, entity disambiguation, corroboration. Third, practitioner observation of what correlates with pages being surfaced, the weakest of the three inputs, and flagged wherever it is doing the work.

Why the dimensions are ordered this way. Retrieval is a precondition, not a factor: an unreachable page cannot be cited regardless of quality. Answerability comes next because it is the highest-leverage thing an author directly controls. Attribution and corroboration follow because they affect whether a retrieved passage is relied upon. Maintenance is last because it preserves the other four rather than creating anything.

Why one point per check. Weighting the checks would imply a precision nobody has. Nobody outside these companies knows the relative importance of these factors, and inventing weights would dress up a guess as a model. Equal weighting is transparently approximate, which is the honest option.

What was deliberately excluded. Keyword density and placement, since answer engines are not matching strings. Word count, since length is not a quality signal. Schema volume, because §12 and every provider's guidance treat markup as descriptive rather than persuasive. An llms.txt file, because Google states no such file is required and treating it as a ranking mechanism would be inaccurate.

Known weakness. Nothing here is validated against citation outcomes, because the outcome data is not publicly available at a useful granularity. This is a structured expert judgment, not an empirical instrument, and should be read as one.

Limitations

It measures positioning, not results. A 40 out of 40 does not entitle you to a citation, and a 20 does not preclude one.

It is not weighted, as noted above: the dimensions almost certainly do not contribute equally, and this checklist does not pretend to know how.

It is written for expertise, service, and knowledge sites. Ecommerce, local-service, and news publishing each have different retrieval characteristics, and several checks here would need rewriting for those.

It reflects provider documentation as of the publication date. That documentation changes, which is why the review cadence and version number exist.

It cannot assess whether your content is any good. Every check could pass on a page that is accurate, well-attributed, and not worth citing because it says nothing anyone needs. That judgment is not automatable.

Download and citation

PDF: AI Citation Readiness Checklist v1.0 (PDF), generated from this page, so the two cannot diverge. This page also prints cleanly if you prefer.

To cite: Zialcita, M. (2026). *AI Citation Readiness Checklist*, version 1.0. martinzialcita.com. https://martinzialcita.com/resources/ai-citation-readiness-checklist/

Free to use and adapt internally. If you publish an adapted version, attribution is appreciated and a note on what you changed is more useful to your readers than fidelity.

Version history and review

VERSION	DATE	CHANGE
1.0	2026-07-31	First publication. Forty checks across five dimensions, with methodology and scoring bands.

Reviewed annually, and sooner if provider documentation changes materially. Next scheduled review: July 2027. Changes will be recorded here rather than applied silently, per the editorial policy.

Sources this checklist is built on

The provider documentation the checks are derived from.

AI features and your website (https://developers.google.com/search/docs/appearance/ai-features)
Google Search Central

Search Essentials (https://developers.google.com/search/docs/essentials) Google Search Central

Overview of crawlers and user agents (https://developers.google.com/search/docs/crawling-indexing/overview-google-crawlers)
Google Search Central

IndexNow protocol documentation (https://www.indexnow.org/documentation) IndexNow

Frequently asked questions

Will a high score get us cited by ChatGPT or Google AI Overviews?

No, and any instrument claiming otherwise is overclaiming. This measures whether you have removed the known obstacles and met the documented requirements.

Citation also depends on whether your page is the best available answer to a question someone asks, on competing sources, and on mechanics none of these companies publish. Positioning is the part you control.

Should we score the site or a single page?

A single page, and specifically the one you most want cited. Site-level averaging hides the information you need, because most sites have a few genuinely citable pages and a long tail that is not.

Scoring two or three pages separately is more useful than one composite number.

Why is there no check for llms.txt or AI-specific schema?

Because Google states that special AI text files are not required for its AI features, and no provider documents such a file as a ranking or citation mechanism. Including a check for it would imply an effect nobody has demonstrated.

An llms.txt file is harmless and may help some experimental tools. It is not readiness, and it should not be sold as such.

How often should we re-score?

Every six months for a page you actively care about, and after any material change to the page or to provider documentation.

Re-scoring the same page over time is far more informative than a one-off score, because the trend tells you whether your maintenance is real.

Part of the pillar: **AEO/GEO and AI search visibility →**

ABOUT THE AUTHOR

Martin Zialcita

AI Implementation Strategist and Forward Deployment Engineer for AI systems integration, based in Honolulu, Hawai'i. Martin works with leadership teams to identify high-value workflows, connect AI to the systems already in use, and prepare the people who will operate it. Commercial delivery runs through AI Marketing Box.

Every factual claim on this site carries a register entry with its evidence and permission status. Facts that are not yet verified are withheld rather than softened. See credentials and verification.

Full biography → **Verify a fact →** **Request an interview →**

Portrait is a placeholder: AI-generated source, EDSR ×4 upscale. §7.1 requires a real photograph before launch.

Author: Martin Zialcita **Published:** July 31, 2026 **Last reviewed:** July 31, 2026

Corrections: Editorial policy