What actually makes a page citable by an AI answer

The structural things that decide whether an assistant quotes your page or somebody else's, and the ones that turn out not to matter.

GUIDE · 2026-11-03 · 9 MIN READ

An assistant answering a question does not read your website the way a person does. It retrieves passages, checks whether they answer the question, and cites what it used. Pages that get cited share a set of properties, and most of them are not the things people optimise for.

This is what those properties are, and what to stop bothering with.

What gets a page quoted

A direct answer, early, in one sentence. If the first two hundred words are context, history and a mission statement, the retriever has nothing to lift. State the answer, then explain it. This single change moves more pages into citations than any technical work.

One question per section, phrased as a question. A heading like "How long does X take" maps onto a user's query almost mechanically. A heading like "Our approach" maps onto nothing.

Numbers with a source and a date. A retrieved passage carrying "as of September 2026" survives fact-checking. A number with no provenance is a liability to the model, so it gets skipped in favour of a competitor's sourced one.

Definitions that stand alone. A paragraph that makes sense only after reading the previous three cannot be retrieved in isolation. Every block should survive being quoted out of order.

A clear entity. Who is speaking. The organisation's name, what it does, where it operates, stated plainly and consistently across the site. Ambiguity about the entity means the answer cannot be attributed, and unattributed content rarely gets cited.

Machine readable facts in a stable place. A summary file, an accurate sitemap, and structured data that matches what the page visually claims. If your schema says one thing and your page says another, the page loses.

What does not matter as much as people think

Keyword density. Retrieval is semantic. Repeating a phrase adds nothing and costs readability.

Word count as a target. A precise 300 word answer beats a 2,000 word treatment of the same question, because the answer is the unit being retrieved.

Backlink volume, for this purpose. Links influence traditional ranking. They do not make a passage more quotable to a model reading it directly.

Publishing frequency. An updated, accurate page outperforms five thin new ones. Recency helps, volume does not.

The three checks worth running

1. The lift test. Take any paragraph from your page at random. Does it still make sense with no surrounding context, and does it answer something specific? If not, rewrite for standalone clarity.

2. The competitor test. Ask an assistant the question your page should answer. If it cites somebody else, read the passage it cited and compare structure: they probably answered first and cited a number.

3. The contradiction test. Compare what your structured data claims with what your page says. Every mismatch is a reason to skip you.

Where this connects to everything else

A page that is citable is usually a page that is genuinely useful, because both require the same discipline: answer the question early, be specific, be dated, and do not bury the point.

That is not a search optimisation technique. It is a writing standard that happens to also be machine readable.

> *The question is not whether the assistant can read your page. It is whether there is a sentence on it worth quoting.*