

TECHNICAL

The robots.txt Mistakes That Make AI Skip You Entirely

By Scott Tischler · September 8, 2026 · 7 min read

The single most avoidable way to lose AI visibility is to block the crawlers you're trying to reach — and a surprising number of businesses do it without knowing. It's the AEO equivalent of locking your own front door and wondering why nobody comes in. Before any clever strategy, check that the machines you want to cite you are actually allowed to read you.

This matters because it's binary and foundational. Every sophisticated thing you might do for AI visibility — corroboration, entity authority, extractable content — is worth exactly nothing if the AI crawler gets a "not allowed" at the door. It won't work around it; it just moves on. And unlike most visibility problems, this one produces no error you'll notice. You simply never show up, silently.

The common mistakes

- **Blocking AI crawlers by name.** Some sites explicitly disallow bots like GPTBot, ClaudeBot, PerplexityBot, or Google-Extended — sometimes deliberately during a past privacy decision, sometimes copied from a template. If you want AI visibility, those blocks work directly against you.
- **Overly broad disallow rules.** A sweeping *Disallow: /* or broad path blocks meant for one purpose can catch the content you most want read.
- **Blocking resources needed to render.** Disallowing scripts or assets a crawler needs to understand the page can leave it seeing a broken or empty version.
- **Inconsistent signals.** A robots.txt that says one thing while meta tags or headers say another creates confusion that can cut against you.
- **Set-and-forgotten files.** A robots.txt written years ago, before these crawlers existed, that no one has revisited since.

The nuance: blocking can be a real choice

I want to be fair here, because blocking AI crawlers isn't always a mistake. Some businesses deliberately disallow them over legitimate concerns about their content training models, and that's a defensible position. The mistake isn't blocking — it's blocking *unintentionally* while actively wanting AI visibility. If you've decided to keep AI crawlers out with eyes open, fine. If you're trying to get cited by AI while a forgotten line in a file quietly forbids it, that's a contradiction worth resolving.

How to check yours

This is a quick, high-value audit. Look at your robots.txt directly. Identify whether any AI crawlers are disallowed, whether broad rules are catching important content, and whether the file reflects a current decision or an old default no one remembers making. If you want AI visibility, make sure the major AI crawlers are welcome and your key content is reachable. Then confirm your other signals — meta directives, headers — agree with the file rather than contradicting it.

It's unglamorous work with a binary payoff: get it wrong and nothing else matters; get it right once and you rarely think about it again. Before investing in any advanced AEO strategy, spend ten minutes making sure you haven't locked your own door.

Key takeaways

- ♦ Blocking the AI crawlers you want to reach is the most avoidable way to lose AI visibility — and many businesses do it unknowingly.
- ♦ It's binary and silent: if a crawler is disallowed, every other AEO effort is worthless, and you get no error — you just never show up.
- ♦ Common mistakes: disallowing AI bots by name, overly broad rules, blocking render resources, contradictory signals, and forgotten old files.
- ♦ Blocking AI crawlers can be a legitimate deliberate choice — the mistake is doing it unintentionally while wanting AI visibility.
- ♦ Audit your robots.txt directly: check for AI-crawler blocks, broad rules catching key content, and whether it reflects a current decision.
- ♦ It's ten minutes of unglamorous work with a binary payoff — do it before any advanced AEO strategy.

Frequently asked questions

How would blocking a crawler hurt my AI visibility?

If your robots.txt disallows an AI crawler, that engine can't fetch or parse your content, so it can't cite you — no matter how good your content, corroboration, or authority is. Worse, it's silent: there's no error, you simply never appear. It's a binary foundation problem that caps everything above it at zero.

Which crawlers should I check for?

Look for the major AI crawlers such as GPTBot, ClaudeBot, PerplexityBot, and Google-Extended, plus any overly broad disallow rules that might catch your important content or block resources a crawler needs to render the page. Also check that your meta directives and headers don't contradict your robots.txt.

Is it ever right to block AI crawlers?

Yes — some businesses deliberately disallow them over legitimate concerns about content being used to train models, which is a defensible, eyes-open choice. The mistake is blocking them unintentionally while actively wanting AI visibility. If you want to be cited by AI, make sure a forgotten line in your robots.txt isn't quietly forbidding it.

About the author. Scott Tischler is the Founder & Chairman of Alrecommend.ai and a practitioner-authority on AI search and Answer Engine Optimization. With 20+ years in marketing technology — including American Express, MetLife, and UBS — and executive and professional study at Wharton, Harvard, and Oxford, he helps businesses become the ones AI recommends.