

Why ChatGPT never mentions your *business*.

Your customers have started asking AI instead of Google. When they ask for the best firm in your category, an assistant reads a handful of websites and names a few companies. **If it cannot read yours, you are not in the running.** This is invisible in Google Analytics, which is why almost nobody knows it is happening to them. **Here is how to find out, in five minutes.**

60 SEC

Open ChatGPT. Ask: "Who are the best [your category] in [your city]?"

Now ask it twice more, worded differently. **Are you named? Are your competitors?** Whatever it says is what a buyer sees. That is your baseline, and it took a minute to find.

1 The robots you have never heard of

AI assistants do not use Google's index. **They run their own crawlers**, and each one has a name. Your website decides, in a single text file, whether they are allowed in.

CRAWLER	WHAT IT FEEDS	WHY IT MATTERS
GPTBot	ChatGPT	Biggest
OAI-SearchBot	ChatGPT's live search results	Citations
ClaudeBot	Claude	Growing fast
PerplexityBot	Perplexity, an answer engine that always cites	Sends traffic
Google-Extended	Gemini and Google AI Overviews	Above your rank
CCBot	Common Crawl, the open dataset nearly every model trains on	Long game

2 The 30 second check

In your browser, go to **yoursite.com/robots.txt**. Every website has one. Read what it says about the names above. Most owners have never opened this file, and a great many site builders and security plugins block AI crawlers **by default**.

X YOU ARE SHUT OUT

User-agent: GPTBot
Disallow: /

Also counts as blocked: a bare **Disallow: /** under **User-agent: ***, or the crawler being absent while your host or CDN blocks AI bots at the edge.

✓ YOU ARE READABLE

User-agent: GPTBot
Allow: /

Each assistant you want to be found in needs its own name listed. Being allowed is not a ranking. It is **permission to be considered at all**.

3 Where the block usually comes from

Almost nobody blocks AI crawlers on purpose. It is nearly always inherited, which is why it goes unnoticed for years.

Your site builder

Wix, Squarespace, Shopify and most WordPress SEO plugins ship a default robots.txt. Some of them disallow AI bots out of the box, and you never saw the file.

A security plugin or CDN

Cloudflare, Wordfence and similar tools have a "block AI scrapers" switch. It is often on by default, and it blocks at the edge, so your robots.txt looks innocent.

A developer, years ago

A staging-site **Disallow: /** that was never removed, or a rule written before these crawlers existed and never revisited.

Reading it takes 30 seconds. Fixing it is the part nobody warns you about.

WHY THE COPY-PASTE ANSWER FAILS

The file is often not the block

If your CDN or security plugin refuses AI bots **at the edge**, your robots.txt can say **Allow** on every line and the crawler still never reaches your server. The file looks innocent. The site is still shut. You cannot see this from the browser.

It is not one name, and the list moves

Seven crawlers matter today and each needs its own directive — miss one and you are invisible in that assistant only, which is exactly the failure nobody notices. New ones ship every few months, and a rule written last year is already out of date.

Allowed is not named

Access is **25 of 100 points**. A crawler that gets in, reads a page it cannot understand or attribute, and leaves without quoting you has cost you nothing less than a block. The other 75 points are on page 2, and none of them are a paste.

You just checked *access*. Access is not citation.

Being allowed in is worth **25 points of 100** — the single biggest check, and the only one you can verify from your own browser. But a crawler that reads a page it cannot **understand, quote or attribute** will still never name you. The rest of the score decides that. Here is how it is put together.

The AI Visibility Scorecard

SIX CHECKS · 100 POINTS

☒	Can ChatGPT open your website? If the door is shut nothing else counts — you are not even a candidate. The one you just checked, and the only one you can see from outside.	25
☐	Is there anything for it to read? If your words only appear after JavaScript runs, an assistant arrives at a blank page and names a competitor whose text was simply there.	25
☐	Does it know what you sell? Without machine-readable facts about your services, your prices and where you work, you are not the answer to "who does this, near me".	20
☐	Does it know who you are? If your name, address and phone disagree across the web, a model cannot be sure you are one company — so it recommends one it is sure about.	15
☐	Have you left it a map? One short file telling an assistant what matters on your site. Almost nobody has one, which makes it the cheapest ground on this list to take.	10
☐	Can it find all your pages? If your sitemap is missing or wrong, whole services are never read at all, and you are judged on the handful of pages it stumbled into.	5

VISIBLE FROM YOUR OWN BROWSER — THE OTHER 75 POINTS NEED A SCAN

1 of 6

0 – 40

Invisible

An assistant either cannot read you, or reads you and finds nothing it can safely repeat. You will not be named, no matter how good you are.

41 – 70

Readable, not quotable

The crawler gets in. It just cannot work out what you do, who you serve, or what to trust, so it names a competitor it does understand.

71 – 100

Citable

You are legible to a machine: clear claims, clean structure, consistent identity. This is where you start turning up in answers.

PROOF, THE HONEST KIND

This website failed its own test. Here is what we found, this week.

BEFORE

Sitemap: [https://\[YOUR_WEBSITE_URL\]/sitemap.xml](https://[YOUR_WEBSITE_URL]/sitemap.xml)

AFTER

Sitemap: <https://hiremina.com/sitemap.xml>

Our robots.txt shipped with a placeholder the site builder never filled in, so we were pointing every crawler at a URL that does not exist. Worse, **our sitemap listed all 17 pages under an old domain we no longer own**, and four of them had slugs that were deleted months ago. Anything following it walked into a dead site. We had no llms.txt at all. **We fixed it before we wrote this sheet.** If it can happen on a web designer's own site, it can happen on yours.

"Doesn't letting AI crawl my site just help them steal my content?"

It is the reasonable fear, and the honest answer is uncomfortable: **blocking mostly costs you the credit, not the copying**. Your material is already described across the web, and models trained on the open Common Crawl years before you had a say. What a block reliably prevents is the assistant **naming you and linking to you** when it answers. You keep the exposure and lose the attribution. If you sell something genuinely confidential, keep it off a public website entirely, not behind a robots.txt line that is a request, not a lock.

"Will this hurt my Google ranking?"

No. These are **different crawlers**. Googlebot handles search and is untouched by anything on this sheet. Google-Extended only governs Gemini and AI Overviews, and Google states it is not used as a ranking signal. Most of what makes a site quotable to an assistant — clean structure, a working sitemap, consistent identity — helps ordinary search too. **There is no ranking downside to any of this.**

Get your *real score*, then 30 minutes to go through it.

Give us your URL and nothing else. The live scan goes further than this sheet: **two scores — whether a machine can READ you, and whether it NAMES you** — the six checks above plus five more, with exactly which ones you are failing and what each is costing you. Free, about a minute. Then we walk your results together: what is actually blocking you, in what order to fix it, and what it is worth. **You leave that call knowing exactly where you stand**, whether or not you ever hire us.

FREE SCAN, THEN BOOK **BOOK THE CALL**