NOTES FROM THE BOT LAB · EDITION OF NOVEMBER 25, 2026

Nearly One in Three Pages ChatGPT Recommends Locks the Front Door

By Will Tygart · Tygart Media · Canonical edition at https://tygartmedia.com/one-in-three-pages-chatgpt-recommends-locks-the-front-door/

Walk down Main Street at 3:05 p.m. and you will see two very different ways to stay in business. On the corner sits the hundred-year-old marble bank and the county courthouse: heavy bronze doors already latched for the afternoon, security guard reading behind the glass, and checks still arriving in the mail slot anyway. Half a block down sits the independent hardware store: porch light on, sidewalk swept, and the winter Hours & Services sheet taped right at eye level inside the front window.

One of the fastest ways a normal business talks itself into a bad technical decision is watching what a giant incumbent gets away with and assuming the same rules apply down the street. Yesterday we measured that exact weight-class gap across the pages AI engines recommend.

When the Recommended Page Slams the Door

While auditing 276 cited rival URL slots — 165 unique web pages and PDFs that major AI engines recommended ahead of us on technical B2B questions — our script did something deliberately ordinary. It sent a clean, polite browser-header fetch to read what was on those pages, with zero firewall workarounds and zero funny business. Every time a server threw a 403 block, a challenge wall, or a blank JavaScript shell, we simply wrote it down and walked away.

Across all 276 cited slots, 53 — nearly one in five — refused a standard fetch. On Google AI Mode, 26.9 percent of cited pages were locked behind blocks or empty shells. On ChatGPT, it hit 30.8 percent: nearly one in three recommended links would not open the door for an ordinary reader script. Broken out by neighborhood, 22.2 percent of government and university URLs and 21.2 percent of commercial vendor and editorial URLs were gated.

How Heavyweights Coast Behind Walls

At first glance that looks like a riddle. How does an AI assistant quote a page that slams the door on automated readers? Three ways: allowlisted search-index partnerships (where Googlebot or Bingbot walks through a side VIP entrance that everyone else is barred from), years of older training-weight memory baked into the model itself, and thousands of third-party pages repeating what the agency or Fortune 500 vendor already said.

If you are a federal regulator or a household software monopoly, the model already knows your address by heart. If you are everybody else — a growing B2B company, a regional specialist, or an independent publisher — locking your front porch is self-sabotage. We documented both sides of that mistake this month: in our 225,000-line JavaScript rendering audit, where headless bots walked away from client-rendered shells, and in our 900,000-line spoofed crawler white paper, where aggressive firewall blocks meant for junk scrapers accidentally slammed the gate on legitimate search and Apple crawlers.

Keep the Porch Light On

Knowing your weight class saves a lot of bad advice. You do not need to copy a courthouse lobby to beat a giant on a specific buyer question. As we saw in The Audience That Never Says Thanks and You Made the Shortlist on the Desk, when a smaller domain serves fast server-rendered HTML, leads with a crisp fifty-word answer, and leaves ungated PDF and Markdown copies on the counter, engines happily cite the open shop right alongside the marble bank.

Let the incumbents play bouncer. Keep your porch light on, keep your answers readable without a secret handshake, and keep walking toward your own mountain.

AI Engine Tested	Cited Rival Slots Checked	Clean HTML or PDF Read	Locked / WAF / JS Shell	Door-Locked Share
ChatGPT (Search-Grounded)	13	9	4	30.8%
Google AI Mode	26	19	7	26.9%
Google Gemini	36	28	8	22.2%
Google AI Overviews	27	21	6	22.2%
Microsoft Bing Copilot	120	97	23	19.2%
Meta AI	54	49	5	9.3%

Key Questions Answered: How can ChatGPT or Google AI Mode cite a webpage that blocks standard web crawlers? Major AI engines rely on allowlisted search-engine indexes (such as Bingbot and Googlebot VIP access), cached snippets, and prior training weights rather than fetching every cited URL anonymously from scratch. **Should a mid-sized business block AI crawlers or require JavaScript rendering if big competitors do?** No. Incumbents and government agencies survive bot blocks because their entities are already deeply indexed and memorized; challenger domains win citations by serving fast, open, server-rendered HTML, PDFs, and Markdown twins. **Which AI engine had the lowest share of crawler-blocked citations in your audit?** Meta AI had the lowest blocked share at 9.3 percent (5 of 54 slots), compared with 19.2 percent on Bing Copilot, 26.9 percent on Google AI Mode, and 30.8 percent on ChatGPT.