RESEARCH FINDINGS

The Fortune 100 Paradox: Why Big Tech's most sophisticated websites are invisible to ChatGPT.

We crawled 275,571 pages across 387 large enterprise domains. The most counterintuitive finding wasn't that AI crawlers struggle with modern websites. It was **which companies struggle most**.

275,571	387	27.8%	45.9%
Pages crawled, raw HTML vs. rendered	Large enterprise domains, Feb–Apr 2026	Content the typical Fortune 100 page hides from AI	Of Fortune 100 domains land in the Critical tier

> "Among the Fortune 100, the typical page loses **27.8% of its content** to JavaScript that AI crawlers don't execute."
>
> The fix is not a content fix. It is an infrastructure fix. And the companies most exposed are the ones with the most engineering resources.
>
> Kyle Duck, Founder & CEO, Alli AI

PRESS CONTACT

Kyle Duck
Founder & CEO, Alli AI
kyle@alliai.com

AVAILABLE ON REQUEST

Full report PDF · raw dataset · CEO interview · cohort breakdowns · company headshots & logos.

Why this matters now.

§ 01

AI search referrals grew nearly 6× in 2025. Three platforms now answer billions of queries that used to start with Google. When they answer, they cite sources. To be cited, a site has to be crawlable. To be crawlable, the content has to live in the raw HTML response, not in a JavaScript bundle that renders client-side.

527%
AI search referral growth in 2025

77.97%
Of AI search referrals come from ChatGPT

15.10%
From Perplexity, plus 6.4% from Gemini

GPTBot, ClaudeBot, and PerplexityBot don't execute JavaScript. When they request a page, they get whatever the server returns before a browser would hydrate it. **If the page is built on React, Vue, or Next.js without server-side rendering, they see a near-empty shell.**

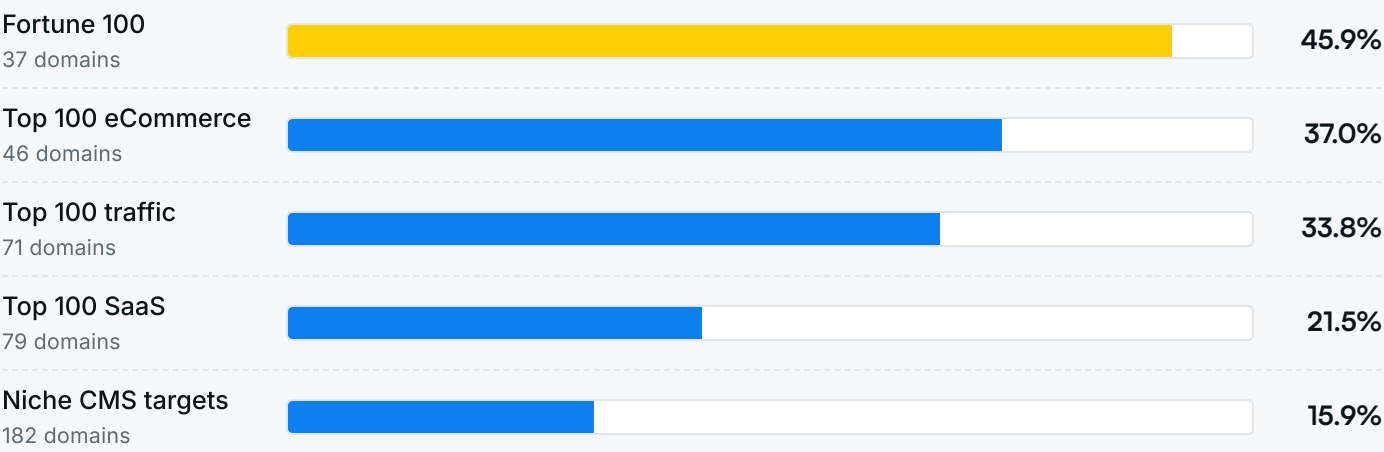
The paradox: bigger means worse.

§ 02

Sorted by cohort, the most well-resourced companies in the world sit at the bottom of every visibility metric we measured. The pattern survives every cut of the data.

% of crawled domains in the Critical tier, by cohort

Critical = the typical page on a domain shows AI crawlers less than 70% of its rendered content.



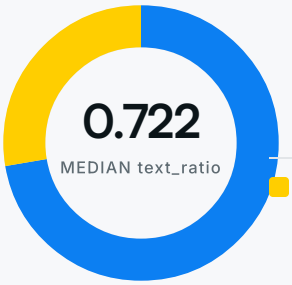
Read: Fortune 100 domains are roughly 3× more likely to be in Critical than niche CMS sites. Most modernized between 2018–2022, in the same window AI search emerged, and inherited the architectural cost.

What AI crawlers actually see.

§ 03

Anatomy of the typical Fortune 100 page

Median page; AI-crawler view vs. rendered view.



Visible in raw HTML — 72.2%
What GPTBot, ClaudeBot, & PerplexityBot can read.

Locked behind JavaScript — 27.8%
Only appears after a browser executes the bundle.

On the typical Fortune 100 page, more than a quarter of the human-visible content is invisible to AI search.

Where the content disappears

% of each element type that only appears after JavaScript runs.

Paragraph elements — 19.5%
Body copy that holds the actual answer LLMs would cite.

All links — 16.5%
Outbound & cross-site routes the crawler follows.

Internal links — 16.0%
The signals that tell a crawler what a site is *about*.

AI crawlers aren't just missing content. They're missing the structure that tells them what the content is *about* and how it relates to the rest of the site.

Five anonymized cases.

§ 04

All five are Fortune 100 domains from the dataset, all in the **Critical** tier. Metrics are unchanged; identities are blinded for the teaser.

01	**Largest US telecom & wireless carrier** 1,000 pages crawled · ~91% of links are JS-only	0.229 MEDIAN TEXT_RATIO	95.5% PAGES IN CRITICAL
02	**Global subscription streaming entertainment platform** 998 pages crawled · 100% of pages in Critical	0.268 MEDIAN TEXT_RATIO	100% PAGES IN CRITICAL
03	**Multinational integrated energy & oil major** 996 pages crawled · Investor & sustainability pages JS-rendered	0.475 MEDIAN TEXT_RATIO	84.8% PAGES IN CRITICAL
04	**Top-tier US retail & investment bank** 784 pages crawled · Product, rates, disclosures client-rendered	0.508 MEDIAN TEXT_RATIO	85.3% PAGES IN CRITICAL
05	**Global enterprise creative-software company** 959 pages crawled · Heavy React/Next bundle · also Top SaaS & Traffic	0.659 MEDIAN TEXT_RATIO	56.0% PAGES IN CRITICAL

Methodology, definitions & sourcing.

§ 05

Methodology

Scope

275,571 pages across 387 large enterprise domains, crawled February through April 2026. Cohorts drawn from publicly available lists; a single domain can belong to more than one.

Double-fetch

Each page was fetched twice: once as raw HTML (the response GPTBot, ClaudeBot, and PerplexityBot receive), once after JavaScript execution in a headless Chromium environment.

Tiers

Domains are tiered on their median page text_ratio: Excellent (≥0.95), Good (0.85–0.95), At-Risk (0.70–0.85), Critical (<0.70).

Cleaning

16,588 pages (5.68%) where text_ratio exceeded 1.0 or returned null were dropped as render/crawl failures, not legitimate content gaps. Mobile and desktop distributions were near-identical.

Glossary

AI crawler

An automated agent that fetches web pages on behalf of an AI platform (e.g. GPTBot, ClaudeBot, PerplexityBot, Google-Extended). Unlike browsers, they don't execute JavaScript.

Raw HTML vs. rendered HTML

Raw HTML is what the server returns first. Rendered HTML is what you see after the browser runs the page's JavaScript. AI crawlers only see raw.

text_ratio

The fraction of a page's human-visible text that exists in raw HTML: raw_text_chars ÷ rendered_text_chars. 1.0 means perfect parity; 0.5 means half the content is JS-only.

AI Visibility Gap

The percentage of a page's content that AI crawlers cannot read. (1 – text_ratio). On the median Fortune 100 page, this gap is 27.8%.

Critical tier

A domain whose typical page shows AI crawlers less than 70% of its content. 45.9% of crawled Fortune 100 domains fall here.

ABOUT ALLI AI

Alli AI is the world's only Onpage SEO automation platform. A 15 minute install lets agencies, ecommerce teams, and enterprises deploy thousands of SEO changes across every page they own. No developer ticket, no codebase access.

Alli's AI Search Visibility Engine detects AI-crawler requests at the edge and serves pre-rendered HTML to them, while everything else continues to be served as normal. Sites typically become accessible to GPTBot, ClaudeBot, and PerplexityBot within 48 hours.

PRESS CONTACT

Kyle Duck

Founder & CEO, Alli AI

kyle@alliai.com

Available on request: full report PDF, raw dataset, CEO interview window, cohort breakdowns, headshots & logos.

PRE-PUBLICATION · NO EMBARGO · EARLY ACCESS