HEMJ
Higher Ed Marketing Journal

← Back to HEMJ Articles

Are LLMs.txt Files Being Implemented Across the Web?

By Ray Martinez • July 23, 2025

For the past year, our team at Archer has fully embraced the shift towards AI in enrollment marketing, especially in SEO. We have reshaped the way we think about the tools, experiences, and content we can deliver across the student journey. This radical shift in approach now has us pushing for more automation and innovation.

As a team, we decided to leave no stone unturned until we could reverse-engineer the output and influence it at will. Under that lens, we examined an emerging web standard for AI, the **LLMs.txt file**.

Are LLMs.txt files being implemented across the web? The short and **simple answer is no**. As of today, Anthropic is the only major player in the LLM space that supports this standard. But the file is getting crawled. As of this blog post, our log file shows that our LLMs.txt files have been pinged over **8,000 times**.

The table below shows the total number of pings for eight sites that we tested this file with.

AGENT	TOTAL	%
8LEGS	5	**0.06%**
AhrefsBot	162	**1.83%**
AhrefsSiteAudit	8	**0.09%**
Applebot	3	**0.03%**

AwarioBot	6	**0.07%**
Barkrowler	20	**0.23%**
bingbot	41	**0.46%**
CCBot	5	**0.06%**
Chrome/Safari	101	**1.14%**
DataForSeoBot	10	**0.11%**
Dataprovider.com	8	**0.09%**
Edge	1	**0.01%**
Facebook	1	**0.01%**
facebookexternalhit	9	**0.10%**
Firefox	14	**0.16%**
Google-Apps-Script	3	**0.03%**
Googlebot	55	**0.62%**
GPTBot	2	**0.02%**
meta-externalagent	6	**0.07%**
Mobile Safari	4	**0.05%**
Mozilla	3	**0.03%**
Mozilla/5.0	26	**0.29%**
OAI-SearchBot	8,330	**94.35%**
Opera	1	**0.01%**
PTST	9	**0.10%**
Safari	1	**0.01%**
Scrapy	1	**0.01%**
search.marginalia.nu	1	**0.01%**
SEOkicks	1	**0.01%**
SemrushBot	12	**0.14%**

SiteAuditBot	1	**0.01%**
Slurp	1	**0.01%**
Yahoo Slurp	1	**0.01%**
YandexBot	8	**0.09%**
TOTAL	**8,829**	

What is LLMs.txt and Why Does It Matter?

If you're tuned into the GEO/SEO debate, there seems to be a great shift in how LLMs differ from traditional search engines. An LLMs text file is most comparable to a robots.txt file, as it lives in the root directory of a site and provides instructions for crawling. The LLMs.txt file enables the conversion of your site's information architecture into Markdown language, resulting in a simplified and clean view of your site's structure.

This simple, clean view offers LLM crawlers an unmitigated path to your content, and that matters because LLMs **cannot render JavaScript**. This means that LLM scrapers are inferring context around a document from raw HTML. As Jono Alderson noted back in May 2025, this has a profound impact on how LLMs ingest your content.

Websites built using client-side rendering have a chance of displaying no content at all, which reduces the likelihood of your content being cited. Simply put, if LLMs can't parse your content, then you won't be able to stay competitive.

How Crawlers Are Interacting with Archer's LLMs.txt Files

When looking at the crawl numbers, OpenAI is dominating the crawl, with over 94% of our pings coming from OpenAI's search bot. When examining the log file, we can see that the search bot pings our servers several times per hour, sometimes even within seconds of each other.

I had Gemini 2.5 analyze the log file for patterns, and here's what it identified:

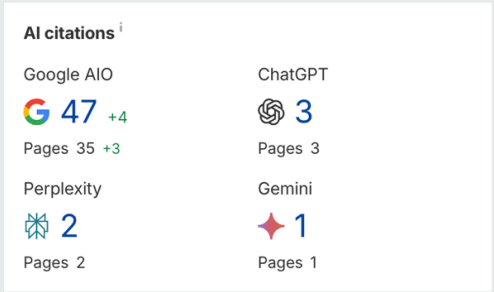
This pattern is consistently observable throughout the logs. For example:

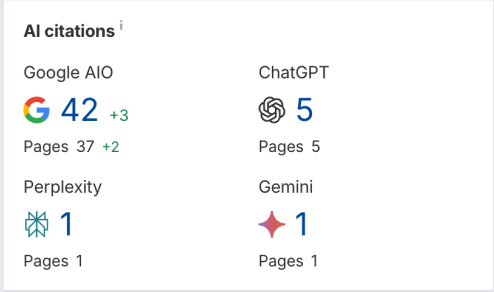
- *On June 26, 2025, the bot requested a URL from genericsite.com at 14:05:55 UTC and then again just three seconds later at 14:05:58 UTC.*
- *On July 10, 2025, genericsite2.com was subjected to a sustained burst of requests, with hits logged at*

15:21:46, 15:23:03, 15:29:09, and 15:32:16 UTC.

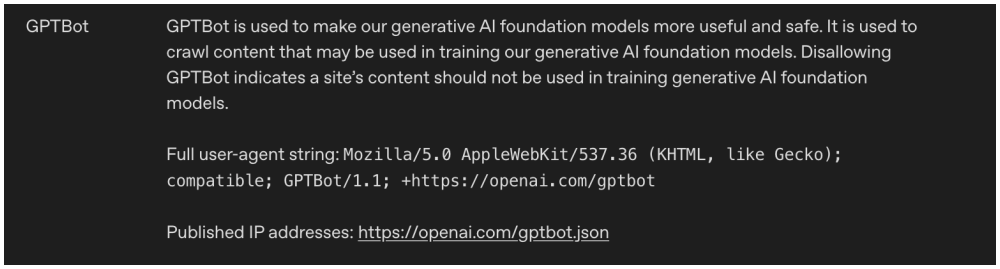
- *On July 6, 2025, two requests were made to the same domain just one second apart, at 02:49:15 and 02:49:16 UTC.*

When looking at the Ahrefs AI citations sections, we've only just begun to see an uptick in performance for citations across AI. The screenshot below shows what we'd expect from such low-traffic sites. A few weeks ago, when this reporting feature launched, these numbers were closer to zero.





What's also interesting to note is that GPT bot pinged our LLMs.txt file for two smallers sites, which saw less pings from OpenAI's search bot. GPT bot is exclusively used to train the model, so this indicates that OpenAI found our file valuable.



In full opaqueness, I've anonymized our sites to avoid malicious intent, but these sites are niche-specific. The sites focus on industry-specific degrees and mainly features informational content around career outcomes, licensure, variations of degrees, and helpful information for prospective students looking to enroll. There's a lot of great information to train on and surface in outputs.

How Did We Get AI Bots to Crawl our LLMs.txt File?

I saw your questions, asking us how we coaxed LLM bots to crawl our file. Many of you wanted to know if we added a link to our file; **of course, we did!** We treated this file like any other standard for SEO. If this were an XML sitemap, we'd submit it to Google Search Console and link to it on our robots.txt file. So why wouldn't we treat this standard the same way?

I'm a big baseball fan, and our methodology for implementing the file is inspired by a line from one of my favorite baseball movies, Field of Dreams.

> ***"If you link to it, they will come."***

Thanks to the brilliance of our team, we decided to approach this differently. Rather than listing a link to the file in robots.txt, which is common practice for an XML sitemap, we decided to inject a link to the file in the <head> section of our sites.

```
<link rel="alternate" type="text/plain" href="https://genericsite.com/llms.txt">
```

We implemented this using the "alternate" link relationship type, which suggests an alternative version of a document. We expected to get crawls from all sorts of bots, but we didn't expect to get so many in such a short period.

Have We Checked the IP Addresses of AI Bots?

When I first announced this on Twitter, many of the initial comments inquired about IP abuse and malicious intent. Given the frequency of server pings, we were concerned about the potential for spoofers looking for site vulnerabilities. We checked the IP address 135.234.64.13, which is identified within OpenAI's documentation.

Should You Implement LLMs.txt on Your Site?

When looking at the evolving landscape, **I'd say yes**. Google has a 20-year head start, which enables it to parse unstructured data with ease. That's a significant investment in infrastructure, which means competitors must raise substantial capital to catch up.

With that said, if you have a deadline-driven product, such as a master's degree or a relatively new offering with limited documentation, and your site is not optimized for AI, your users may encounter hallucinations. I hypothesize that the LLMs.txt file serves as a safeguard, providing pertinent information to the LLMs and can help reduce errors by serving fresh content.

For example, a prospective student searches for a Fall application deadline, but LLM models have been trained on

an earlier version of our site. LLMs need to do a live search or RAG to satisfy user intent. Another example might be sweeping changes to the curriculum for a new semester. How can we maintain accuracy for our students?

The Future of LLMs.txt

At Archer, our team would learn firsthand. It's the only way we can future-proof our university partner's success. In higher education, we face various challenges, including declining enrollments, which is an industry-wide issue.

Final Thoughts on the LLMs.txt File

While the LLMs text file is not yet a widely adopted standard across the web, the recent flurry of bot activity suggests there is value. Given the limitations of current LLM crawlers, this file might be your best bet in safeguarding against pitfalls that will have you excluded from these new systems.

As the industry evolves, it's our duty as stewards of the web to test, try, break, and fix things. I encourage marketers, SEOs, and web engineers to think differently and lean into curiosity. It is through that lens that we can help our partners be found wherever their students are.
If you'd like to talk more about AI-powered SEO and how Archer is helping universities show up where students are searching, the Archer team is ready to help.

About The Author

Ray Martinez

Ray is Archer Education's VP of SEO, concentrating on search engine optimization, conversion rate

optimization, and content strategy. He has a history of diverse campaign work for organizations such as Louisiana State University, Tulane University, University of San Diego, and the California Innocence Project. Ray's work with these partners has led to millions of search impressions, thousands of organic keywords ranking first on Google, and healthier sites that drive lead generation.

Subscribe to the Higher Ed Marketing Journal

Email*

First name

Last name

Subscribe Me!

Recommended Articles

8 Ways to Get the Most out of Your Press Release

Marketing

As I discussed in my previous article, press releases can be an extremely effective tool when you are looking to [...]

Blogger Outreach Emails: Persuasive Writing Techniques

Marketing

As we all know, how something is phrased is often more important than what is actually being said. If you [...]

Case Study: Social Audience Targeting Helps Drive Enrollment for Trinity Law School

About Trinity Law School For the