

How ChatGPT Uses the Public Suffix List to Treat Domains & Subdomains Differently

Metehan Yesilyurt

AI Search & Information Retrieval Researcher · metehan.ai

Published: November 10, 2025 · Canonical: <https://metehan.ai/blog/how-chatgpt-uses-the-public-suffix-list/>

When people discuss AI search optimization, conversations center on embeddings, rerankers, and prompt engineering.

But there's an older, quieter piece of infrastructure. It's been hiding in plain sight.

It's called the **Public Suffix List**, and through analysis of ChatGPT's client-side code, I've confirmed it's actively being used in citation and search result processing.

A reminder: ChatGPT is using these user-agents: <https://platform.openai.com/docs/bots>^[1]

What the Public Suffix List Actually Does

The Public Suffix List (PSL)^[2] is a Mozilla-maintained, open catalog of domain suffixes that defines ownership boundaries on the web.

It answers a critical question: *Where does one entity's control end and another's begin?*

For example:

- `metehan.ai` → registrable domain: `metehan.ai`
- `metehan.vercel.app` → registrable domain: `vercel.app`
- `blog.metehan.ai` → still belongs under `metehan.ai`

The PSL establishes which domains are truly independent versus which exist under shared hosting platforms. **It's also useful for a parasite SEO approach.**

Originally created for cookie security in browsers, it has become fundamental web infrastructure, used by Chrome, Firefox, Meta, certificate authorities, email authentication systems, and web crawlers worldwide.

PSL in ChatGPT: A Confirmed Finding

Through technical analysis of ChatGPT's browser-based source code, I discovered explicit references to Public Suffix List processing in their citation and domain handling systems.

This isn't speculation. It's verifiable in the client-side JavaScript that powers ChatGPT's web interface. It's the same order as PSL.

Figure

Why this matters:

If ChatGPT (and potentially other AI systems) use PSL data to normalize and group domains, it means:

1. **Domain ownership determines attribution:** All your Medium posts contribute authority to `medium.com`, not to you
2. **Platform content dilutes your brand:** Your Notion pages strengthen `notion.site`, not your domain entity. Useful for "parasite SEO"
3. **True ownership creates unique identity:** `yourname.ai` establishes you as a distinct, attributable source

The Technical Logic

Large-scale web systems, browsers, search engines, spam filters, security crawlers use PSL data to:

- **Group URLs by ownership boundary** (e.g., treating `news.bbc.co.uk` and `sport.bbc.co.uk` as the same entity)
- **Normalize subdomains** into registrable domains
- **Prevent abuse** from multi-tenant platforms (`*.github.io`, `*.notion.site`, `*.vercel.app`)
- **Attribute authority** at the ownership level rather than the subdomain level

Platform Content vs. Real Ownership

Publishing on hosted platforms creates a fundamental attribution problem:

Medium.com example:

- Your article: `medium.com/@yourname/brilliant-insight`
- PSL registrable domain: `medium.com`
- Authority destination: Medium's corpus, not yours

Your own domain:

- Your article: `yourname.ai/brilliant-insight`
- PSL registrable domain: `yourname.ai`
- Authority destination: Your unique entity

From an AI system's perspective, these are architecturally different. One contributes to a massive aggregator; the other builds your discrete authority.

Why Multi-Tenant Platforms Are Listed

The PSL includes platforms like `github.io`, `notion.site`, `vercel.app`, and `blogspot.com` specifically because users shouldn't be able to set cookies or claim authority for other users' subdomains.

But this security feature has a consequence: it also means these subdomains may be treated as distinct entities *under the platform's umbrella*, not as independent authorities.

Example:

- `john.github.io` and `jane.github.io` are separate in PSL terms
- But both exist under `github.io` as the effective ownership boundary
- AI systems reading PSL data may aggregate signal at the `github.io` level, not at individual user subdomains

Owning `john.com` eliminates this ambiguity entirely.

Implications for AI Search Optimization (AEO)

For practitioners working on AI visibility, this finding creates a layer of thinking:

- 1. Ownership boundaries matter structurally** The PSL defines where your control begins. Content outside your PSL boundary may not be fully attributable to you.
- 2. Domain architecture affects entity recognition** AI systems using PSL data, see `yourname.ai` as a distinct entity, while `yourname.medium.com` registers as Medium content.
- 3. Platform content creates split authority** Your best work on Substack or LinkedIn builds those platforms' authority graphs, not yours. (Substack & LinkedIn aren't in the PSL list. It doesn't mean they won't work.)
- 4. The cost of convenience is attribution** Hosted platforms offer ease of publishing but at the structural cost of blurred ownership in AI systems.

A Layer Beneath Rankings

The Public Suffix List isn't a ranking factor. It's infrastructure.

It sits below the semantic layer, defining *who* owns *what* before algorithms decide what to surface.

In traditional SEO, we optimize for crawlability, links, and content quality. In AEO, we may also need to think about **semantic ownership boundaries**, the technical lines that tell AI systems where one source ends and another begins.

ChatGPT's use of PSL data suggests these boundaries aren't incidental. They're structural.

Key Takeaway

The Public Suffix List isn't just a browser security mechanism anymore.

It's a **boundary definition layer** that may shape authority attribution in AI retrieval systems.

While embeddings and rerankers determine *what* content surfaces, PSL data may determine *who* gets credited.

In an ecosystem where AI-driven discovery is replacing link-based search, owning your registrable domain isn't just good practice — it's fundamental architecture.

Technical Notes

Verification: PSL usage in ChatGPT can be observed through browser DevTools inspection of network requests and client-side JavaScript processing. Look for domain normalization functions that reference PSL data structures.

PSL Maintenance: The Public Suffix List is maintained by Mozilla and updated regularly through community contributions. The canonical version is hosted at publicsuffix.org^[3].

Further Reading:

- [Official PSL Documentation](#)^[4]
- [PSL on GitHub](#)^[5]
- [Wikipedia: Public Suffix List](#)^[6]

ChatGPT's use of the Public Suffix List is a quiet AEO signal. Domain boundaries decide which pages get credit for AI citations in ChatGPT, Perplexity and Gemini. Get PSL behavior wrong and your answer engine optimization work leaks equity to the wrong subdomain.

REFERENCES & SOURCES

- [1] <https://platform.openai.com/docs/bots> — <https://platform.openai.com/docs/bots>
- [2] **Public Suffix List (PSL)** — https://publicsuffix.org/list/public_suffix_list.dat
- [3] **publicsuffix.org** — <https://publicsuffix.org/>
- [4] **Official PSL Documentation** — <https://publicsuffix.org/learn/>
- [5] **PSL on GitHub** — <https://github.com/publicsuffix/list>
- [6] **Wikipedia: Public Suffix List** — https://en.wikipedia.org/wiki/Public_Suffix_List