Web Almanac Data Reveals CMS Plugins Are Setting Technical SEO Standards (Not SEOs) via @sejournal, @chrisgreenseo

If more than half the web runs on a content management system, then the majority of technical SEO standards are being positively shaped before an SEO even starts work on it. That's the lens I took into the 2025 Web Almanac SEO chapter (for clarity, I co-authored the 2025 Web Almanac SEO chapter referenced in this article).

Rather than asking how individual optimization decisions influence performance, I wanted to understand something more fundamental: How much of the web's technical SEO baseline is determined by CMS defaults and the ecosystems around them.

SEO often feels intensely hands-on — perhaps too much so. We debate canonical logic, structured data implementation, crawl control, and metadata configuration as if each site were a bespoke engineering project. But when 50%+ of pages in the HTTP Archive dataset sit on CMS platforms, those platforms become the invisible standard-setters. Their defaults, constraints, and feature rollouts quietly define what "normal" looks like at scale.

This piece explores that influence using 2025 Web Almanac and HTTP Archive data, specifically:

- How CMS adoption trends track with core technical SEO signals.
- Where plugin ecosystems appear to shape implementation patterns.
- And how emerging standards like llms.txt are spreading as a result.

The question is not whether SEOs matter. It’s whether we’ve been underestimating who sets the baseline for the modern web.

The Backbone Of Web Design

The 2025 CMS chapter of the Web Almanac saw a milestone hit with CMS adoption; over 50% of pages are on CMSs. In case you were unsold on how much of the web is carried by CMSs, over 50% of 16 million websites is a significant amount.

Screenshot from Web Almanac, February 2026

With regard to which CMSs are the most popular, this again may not be surprising, but it is worth reflecting on with regard to which has the most impact.

Image by author, February 2026

WordPress is still the most used CMS, by a long way, even if it has dropped marginally in the 2024 data. Shopify, Wix, Squarespace, and Joomla trail a long way behind, but they still have a significant impact, especially Shopify, on ecommerce specifically.

SEO Functions That Ship As Defaults In CMS Platforms

CMS platform defaults are important, this – I believe – is that a lot of basic technical SEO standards are either default setups or for the relatively small number of websites that have dedicated SEOs or people who at least build to/work with

SEO best practice.

When we talk about “best practice,” we’re on slightly shaky ground for some, as there isn’t a universal, prescriptive view on this one, but I would consider:

- Descriptive “SEO-friendly” URLs.
- Editable title and meta description.
- XML sitemaps.
- Canonical tags.
- Meta robots directive changing.
- Structured data – at least a basic level.
- Robots.txt editing.

Of the main CMS platforms, here is what they – self-reportedly – have as “default.” Note: For some platforms – like Shopify – they would say they’re SEO-friendly (and to be honest, it’s “good enough”), but many SEOs would argue that they’re not friendly enough to pass this test. I’m not weighing into those nuances, but I’d say both Shopify and those SEOs make some good points.

CMS	SEO-friendly URLs	Title & meta description UI	XML sitemap	Canonical tags	Robots meta support	Basic structured data	Robots.txt
WordPress	Yes	Partial (theme-dependent)	Yes	Yes	Yes	Limited (Article, BlogPosting)	No (plugin or server access required)
Shopify	Yes	Yes	Yes	Yes	Limited	Product-focused	Limited (editable via robots.txt.liquid, constrained)
Wix	Yes	Guided	Yes	Yes	Limited	Basic	Yes (editable in UI)
Squarespace	Yes	Yes	Yes	Yes	Limited	Basic	No (platform-managed, no direct file control)
Webflow	Yes	Yes	Yes	Yes	Yes	Manual JSON-LD	Yes (editable in settings)
Drupal	Yes	Partial (core)	Yes	Yes	Yes	Minimal (extensible)	Partial (module or server access)
Joomla	Yes	Partial	Yes	Yes	Yes	Minimal	Partial (server-level file edit)
Ghost	Yes	Yes	Yes	Yes	Yes	Article	No (server/config level only)

TYPO3	Yes	Partial	Yes	Yes	Yes	Minimal	Partial (config or extension-based)

Based on the above, I would say that most SEO basics can be covered by most CMSs "out of the box." Whether they work well for you, or you cannot achieve the *exact* configuration that your specific circumstances require, are two other important questions – ones which I am not taking on. However, it often comes down to these points:

1. It is possible for these platforms to be used badly.
2. It is possible that the business logic you need will break/not work with the above.
3. There are *many* more advanced SEO features that aren't out of the box, that are just as important.

We are talking about foundations here, but when I reflect on what shipped as "default" 15+ years ago, progress has been made.

Fingerprints Of Defaults In The HTTP Archive Data

Given that a lot of CMSs ship with these standards, do these SEO defaults correlate with CMS adoption? In many ways, yes. Let's explore this in the HTTP Archive data.

Canonical Tag Adoption Correlates With CMS

Combining canonical tag adoption data with (all) CMS adoption over the last four years, we can see that for both mobile and desktop, the trends seem to follow each other pretty closely.

Image by author, February 2026

Image by author, February 2026

Running a simple Pearson correlation over these elements, we can see this strong correlation even clearer, with canonical tag implementation and the presence of self-canonical URLs.

Image by author, February 2026

What differs is the mobile correlation of canonicalized URLs; that seems to be a negative correlation on mobile and a lower (but still positive) correlation on desktop. A drop in canonicalized pages is largely causing this negative correlation, and the reasons behind this could be many (and harder to be sure of).

Canonical tags are a crucial element for technical SEO; their continued adoption does certainly seem to track the growth in CMS use, too.

Schema.org Data Types Correlate With CMS

Schema.org types against CMS adoption show similar trends, but are less definitive overall. There are many different types of

Schema.org, but if we plot CMS adoption against the ones most common to SEO concerns, we can observe a broadly rising picture.

Image by author, February 2026

With the exception of Schema.org WebSite, we can see CMS growth and structured data following similar trends.

But we must note that Schema.org adoption is quite considerably lower than CMSs overall. This could be due to most CMS defaults being far less comprehensive with Schema.org. When we look at specific CMS examples (shortly), we'll see far-stronger links.

Schema.org implementation is still mostly intentional, specialist, and not as widespread as it could be. If I were a

search engine or creating an AI Search tool, would I rely on universal adoption of these, seeing the data like this? Possibly not.

Robots.txt

Given that robots.txt is a single file that has some agreed standards behind it, its implementation is far simpler, so we could anticipate higher levels of adoption than Schema.org.

The presence of a robots.txt is pretty important, mostly to limit crawl of search engines to specific areas of the site. We are starting to see an evolution – we noted in the 2025 Web Almanac SEO chapter – that the robots.txt is used even more as a governance piece, rather than just housekeeping. A key sign that we're using our key tools differently in the AI search world.

But before we consider the more advanced implementations, how much of a part does a CMS have in ensuring a robots.txt is present? Looks like over the last four years, CMS platforms are driving a significant amount more of robots.txt files serving a 200 response:

Image by author, February 2026

What is more curious, however, is when you consider the file of the robots.txt files. Non-CMS platforms have robots.txt files that are significantly larger.

Image by author, February 2026

Why could this be? Are they more advanced in non-CMS platforms, longer files, more bespoke rules? Most probably in some cases, but we're missing another impact of a CMSs standards – compliant (valid) robots.txt files.

A lot of robots.txt files serve a valid 200 response, but often they're not txt files, or they're redirecting to 404 pages or similar. When we limit this list to only files that contain user-agent declarations (as a proxy), we see a different story.

Image by author, February 2026

Approaching 14% of robots.txt files served on non-CMS platforms are likely not even robots.txt files.

A robots.txt is easy to set up, but it is a conscious decision. If it’s forgotten/overlooked, it simply won’t exist. A CMS makes it more likely to have a robots.txt, and what’s more, when it is in place, it makes it easier to manage/maintain – which IS key.

WordPress Specific Defaults

CMS platforms, it seems, cover the basics, but more advanced options – which still need to be defaults – often need additional SEO tools to enable.

Interrogating WordPress-specific sites with the HTTP Archive data will be easiest as we get the largest sample, and the Wapalizer data gives a reliable way to judge the impact of WordPress-specific SEO tools.

From the Web Almanac, we can see which SEO tools are the most installed on WordPress sites.

Screenshot from Web Almanac, February 2026

For anyone working within SEO, this is unlikely to be surprising. If you are an SEO and worked on WordPress, there is a high chance you have used either of the top three. What IS worth considering right now is that while Yoast SEO is by far the most prevalent within the data, it is seen on barely over 15% of sites. Even the most popular SEO plugin on the most popular CMS is still a relatively small share.

Of these top three plugins, let’s first consider what the differences of their “defaults” are. These are similar to some of WordPress’s, but we can see many more advanced features that come as standard.

SEO Capability	All-in-One SEO	Yoast SEO	Rank Math
Title tag control	Yes (global + per-post)	Yes	Yes
Meta description control	Yes	Yes	Yes
Meta robots UI	Yes (index/noindex etc.)	Yes	Yes
Default meta robots output	Explicit index,follow	Explicit index,follow	Explicit index,follow
Canonical tags	Auto self-canonical	Auto self-canonical	Auto self-canonical
Canonical override (per URL)	Yes	Yes	Yes
Pagination canonical handling	Limited	Historically opinionated	More configurable
XML sitemap generation	Yes	Yes	Yes
Sitemap URL filtering	Basic	Basic	More granular
Inclusion of noindex URLs in sitemap	Possible by default	Historically possible	Configurable
Robots.txt editor	Yes (plugin-managed)	Yes	Yes
Robots.txt comments/signatures	Yes	Yes	Yes
Redirect management	Yes	Limited (free)	Yes
Breadcrumb markup	Yes	Yes	Yes
Structured data (JSON-LD)	Yes (templated)	Yes (templated)	Yes (templated, broad)

Schema type selection UI	Yes	Limited	Extensive
Schema output style	Plugin-specific	Plugin-specific	Plugin-specific
Content analysis/scoring	Basic	Heavy (readability + SEO)	Heavy (SEO score)
Keyword optimization guidance	Yes	Yes	Yes
Multiple focus keywords	Paid	Paid	Free
Social metadata (OG/Twitter)	Yes	Yes	Yes
Llms.txt generation	Yes – enabled by default	Yes – one-check enable	Yes – one-check enable
AI crawler controls	Via robots.txt	Via robots.txt	Via robots.txt

Editable metadata, structured data, robots.txt, sitemaps, and, more recently, llms.txt are the most notable. It is worth noting that a lot of the functionality is more “back-end,” so not something we’d be as easily able to see in the HTTP Archive data.

Structured Data Impact From SEO Plugins

We can see (above) that structured data implementation and CMS adoption do correlate; what is more interesting here is to understand where the key drivers themselves are.

Viewing the HTTP Archive data with a simple segment (SEO plugins vs. no SEO plugins), from the most recent scoring paints a stark picture.

Image by author, February 2026

When we limit the Schema.org @types to the most associated with SEO, it is really clear that some structured data types are pushed *really* hard using SEO plugins. They are not completely absent. People may be using lesser-known plugins or coding their own solutions, but ease of implementation is implicit in the data.

Robots Meta Support

Another finding from the SEO Web Almanac 2025 chapter was that “follow” and “index” directives were the most prevalent, even though they’re technically redundant, as having no meta robots directives is implicitly the same thing.

Screenshot from Web Almanac 2025, February 2026

Within the chapter number crunching itself, I didn’t dig in much deeper, but knowing that all major SEO WordPress plugins have “index,follow” as default, I was eager to see if I could make a stronger connection in the data.

Where SEO plugins were present on WordPress, “index, follow” was set on over 75% of root pages vs. <5% of WordPress sites without SEO plugins.

Image by author, February 2026

Given the ubiquity of WordPress and SEO plugins, this is likely a huge contributor to this particular configuration. While this is redundant, it isn’t wrong, but it is – again – a key example of whether one or more of the main plugins establish a de facto standard like this, it really shapes a significant portion of the web.

Diving Into LLMs.txt

Another key area of change from the 2025 Web Almanac was the introduction of the llms.txt file. Not an explicit endorsement of the file, but rather a tacit acknowledgment that this is an important data point in the AI Search age.

From the 2025 data, **just over 2% of sites had a valid llms.txt file** and:

- 39.6% of llms.txt files are related to All-in-One SEO.
- 3.6% of llms.txt files are related to Yoast SEO.

This is not necessarily an intentional act by all those involved, especially as Rank Math enables this by default (not an opt-in like Yoast and All-in-One SEO).

Image by author, February 2026

Since the first data was gathered on July 25, 2025 if we take a month-by-month view of the data, we can see further growth since. It is hard not to see this as growing confidence in this markup OR at least, that it’s so easy to enable, more people are likely hedging their bets.

Conclusion

The Web Almanac data suggests that SEO, at a macro level, moves less because of individual SEOs and more because WordPress, Shopify, Wix, or a major plugin ships a default.

- Canonical tags correlate with CMS growth.
- Robots.txt validity improves with CMS governance.
- Redundant “index,follow” directives proliferate because plugins make them explicit.
- Even llms.txt is already spreading through plugin toggles before it even gets full consensus.

This doesn’t diminish the impact of SEO; it reframes it. Individual practitioners still create competitive advantage, especially in advanced configuration, architecture, content quality, and business logic. But the baseline state of the web, the technical floor on which everything else is built, is increasingly set by product teams shipping defaults to millions of sites.

Perhaps we should consider that if CMSs are the infrastructure layer of modern SEO, then plugin creators are de facto standards setters. They deploy “best practice” before it becomes doctrine

This is how it should work, but I am also not entirely comfortable with this. They normalize implementation and even create new conventions simply by making them zero-cost. Standards that are redundant have the ability to endure because they can.

So the question is less about whether CMS platforms impact SEO. They clearly do. The more interesting question is whether we, as SEOs, are paying enough attention to where those defaults originate, how they evolve, and how much of the web’s “best practice” is really just the path of least resistance shipped at scale.

An SEO’s value should not be interpreted through the amount of hours they spend discussing canonical tags, meta robots, and rules of sitemap inclusion. This should be standard and default. If you want to have an out-sized impact on SEO, lobby an existing tool, create your own plugin, or drive interest to influence change in one.

More Resources:

Featured Image: Prostock-studio/Shutterstock

https://www.searchenginejournal.com/web-almanac-data-reveals-cms-plugins-are-setting-technical-seo-standards-not-

seos/567649/