The Loop: Half My Bot Traffic Is Walking in Circles

By William Tygart · November 2026

On one of my sites — Healthcare Facilities Hub, a publication for people who run clinical facilities — a third of everything the server did for fifty-five days was pagination.

Not articles. Not the ICRA matrix or the ventilation tables that people and their AI agents actually come for. Pagination. 115,998 page requests, 34.0 percent of the site's entire output of 341,172 page requests, spent serving numbered archive pages: page 2, page 7, page 14 of an index that exists because the CMS generates it, linked because the template links it, and walked because bots walk whatever is in front of them.

That is the loop. And once you see it in one site's logs, you start seeing it everywhere.

What a loop actually is

An archive page is a hallway. Tag pages, category pages, author pages, date indexes, paginated lists — none of them is the destination. They exist to lead somewhere: to an article, an answer, a thing a reader or an answer engine can use.

A loop is a hallway that leads to another hallway.

I measure it in transitions. Take every request for an archive page, look at where the same visitor went next, and count how often "next" was another archive page instead of an article. On tygartmedia.com, over the same fifty-five days, taxonomy pages took 336,574 hits — 15.73 percent of all page hits on the site. 92.9 percent of that traffic was bots. And 53.7 percent of the transitions out of those archive pages landed on another archive page.

Read that again. More than half the time a bot entered my archive system, the archive system handed it to another part of the archive system. The fleet wasn't reading. It was circulating.

Every one of those hits is real cost wearing a fake costume. It is server capacity. It is crawl attention — the finite amount of time an indexing fleet will spend on your domain in a pass. And it produces a page that no human will ever need and no answer engine will ever cite, because there is nothing on it to cite. It is a list of links to a list of links.

It isn't one site's bad habit

The pattern repeats across the six sites I run, at different severities, and the severity is diagnostic.

Healthcare Facilities Hub is the worst: 34.0 percent of all page requests on pagination alone. Not taxonomy in general — the numbered tail by itself. HFH publishes for a professional audience with genuinely deep topical archives, and somewhere along the way the pagination tail grew longer than the content it indexes is worth. The archive system began eating the site it was supposed to organize.

Continuity Hub is the control case. Its taxonomy is only 3.62 percent of traffic — a comparatively tidy site, and proof the loop is a choice, not a law of nature. But even there, the author archives paginate out to /page/8/ and beyond, and 578 hits walked that author index across the fifty-five days. Five hundred seventy-eight requests to read a byline listing, in order, eight pages deep. Nobody reads that. Nothing human has ever read that. It exists because WordPress builds it, and it gets traffic because it exists.

Risk Coverage Hub spent 12,618 requests on author pagination plus single-post tag archives. Sit with that second one. A tag archive containing exactly one article is a hallway built to serve a single door. The bot follows the tag link, loads a page whose entire content is a link to the one article, and then loads the article — two requests where the architecture owed it one, repeated thousands of times, forever, because nothing in the system ever asks whether the hallway should exist.

The distinction the smart reader is already reaching for

I can hear the objection, because it's a good one: aren't archives how crawlers discover and re-crawl content? Aren't tag pages a signal?

Some of them are. I know, because I measured it. In this same lab, on this same site, we cleaned up the tagging on hundreds of articles — substantive tags, topics a reader would recognize — and AI re-crawls of those articles rose 23 to 28 percent. Substantive archives earn their keep. A tag page over fifty real articles on a real topic is a piece of topical architecture, and the fleets treat it that way.

So the enemy is not the archive. The enemy is the *empty* archive system: the tag with one post under it, the author index paginated to page 8, the category with three posts from 2021, the pagination tail that runs to page 40 because nobody ever told it to stop. A substantive archive is a map. An empty archive system is a maze. From the log file they look identical until you measure the transitions — which is why this stays invisible until somebody reads the logs instead of the dashboard.

That's also why the crude fix is wrong. "Noindex all tags" throws away the 23-to-28-percent re-crawl effect along with the waste. "Leave it, crawl budget isn't real" ignores a third of a site's serving capacity going to numbered hallways. The honest fix is per-archive judgment, which is slower and less tweetable: keep the archives that carry topical weight, noindex the thin ones, de-link what shouldn't be walked, and cap the pagination tail where the content stops earning it.

What we're doing about it

In the lab the rules are now simple. Author archives get noindexed outright — no byline page has ever earned a citation. Thin tags — two posts or fewer — get noindexed; substantive tags stay, because we have the re-crawl numbers to defend them. And when a new archive type appears in the transition data with a loop share like tygartmedia's 53.7 percent, that's treated as an architecture bug, not a traffic source.

Because that is what a loop is, in the end: your own site, confessing in its logs that it generates pages nobody needs and sends its most important visitors — the fleets that decide whether you get cited — walking in circles inside them.

Fleets have limited patience and limited passes. I'd rather mine spend theirs on the articles.

We do AI citation restoration. When the AI can't find you, can't read you, or keeps citing somebody else — send us the worst one you have. Your RFPs. Your disasters. The site whose archive system ate itself, the problem everyone else already failed to fix. We come from the restoration industry. We know what's on the floor, and cleaning it up is the job. And if you're a developer who just needs help on the hard part, get a hold of us.