

The AEO Field Guide

How to Tell If ChatGPT, Perplexity, and Claude Cite Your B2B Brand

A vendor-neutral 30/60/90 plan to baseline your AI citation rate across ChatGPT, Perplexity, and Claude, deploy the right measurement stack, and interpret what the numbers mean for pipeline.

Ronald Berry III

22 min read · 2026-05-25

HubSpot is the UI. Claude is the OS.

Flywheel ties them together. · flywheelconsultancy.com

TL;DR

Forty percent of buyer queries now start in an AI interface, not a Google search box. If ChatGPT, Perplexity, or Claude is not citing your business when buyers ask questions in your category, you are invisible to a meaningful slice of your pipeline. Traditional SEO metrics do not measure this. AEO measurement is the new discipline, and most B2B operators do not have a single tool tracking it yet.

This field guide gives you a vendor-neutral 30/60/90 plan to: (1) baseline your current AI citation rate across the four major engines, (2) deploy the right measurement stack for your size and budget, and (3) interpret what the numbers actually mean for revenue. The three tools we recommend (ranked by fit for B2B operators on HubSpot) are HubSpot AEO, OpenLens, and Otterly AI. The metric you anchor on is Share of LLM. The dashboard you operate from is what we call the Visibility Triangle.

Ninety-six percent of AI citations come from third-party content. If you are only optimizing your own site, you are missing the leverage. This guide is the operator's view, not the agency pitch.

§1 – Why AEO Measurement Is Non-Negotiable Now

The Gartner number that makes this immediate: 40% of all information-seeking queries now begin in an AI interface rather than a traditional search engine. ChatGPT serves 900 million weekly active users. Perplexity handles 100 million queries weekly. Google AI Overviews reach over 2 billion users. This is not a future trend that operators have a quarter or two to prepare for. The behavior shift has already happened.

The blind spot most B2B teams live with is that Google Analytics shows zero direct attribution from AI engines. Your dashboard reports the same organic sessions as last quarter. Your keyword positions look stable. Your impressions are flat or up. And meanwhile, a third of your buyers are doing their entire vendor research inside ChatGPT before they ever hit your site, never seeing your beautifully ranked blog post, and forming an opinion of your category from whatever the AI surfaces in the answer.

The Q1 2026 click-through-rate data we covered in our recent HubSpot AEO post is the clearest version of this story. Across 50 B2B SaaS keywords, pages holding top-three organic rankings experienced CTR declines of 18% to 34% once AI-generated answers appeared above the fold. Rankings did not change. Impressions did not change. Clicks dropped anyway. That is not a measurement glitch. That is the buyer behavior shift showing up in the data.

The compounding piece operators miss: if AI engines do not cite you in 2026, they will probably not cite you in 2027 either. Training data has lag. The corpus that gets baked into the next model release reflects who was getting cited this year. Brands that show up in third-party listicles, podcast transcripts, Reddit threads, and category guides now will keep showing up. Brands that do not will keep getting passed over, and the gap widens with every model retrain.

The fastest way to make this concrete is to run one query in ChatGPT yourself. Ask: "best AI implementation partners for B2B SaaS." Read the answer. Count how many vendors get named. Now ask the same question in Perplexity, then Claude. The vendors that show up across all three engines are the ones that built citation surfaces over the last 18 months. The ones that did not are invisible to the buyer doing exactly this research right now.

That is why measurement comes first. You cannot fix what you cannot see. The next five sections build the operator's view.

§2 - How AI Engines Actually Cite Businesses

Each engine has a different bias. Optimizing for ChatGPT and Perplexity simultaneously requires different content shapes, because the engines pull from different sources, weight authority differently, and surface different parts of your content. Here is what the operator needs to know about the four that matter for B2B today.

ChatGPT (OpenAI)

ChatGPT blends training data with live web search via Bing integration. The Yext study of 6.8 million citations found ChatGPT averages 7.92 sources per response. Commercial queries (any phrasing that includes "best," "top," "vs," or "alternatives") trigger search-augmented mode automatically, meaning your live site has a chance to land in the answer. Informational

queries lean harder on training data, which means your historical content footprint matters more.

The signal that drives ChatGPT citations: structured, extractable content with clear question-answer pairs, plus a strong third-party authority footprint (G2 listings, Clutch profiles, guest posts, podcast transcripts) that gives ChatGPT the corroborating sources it wants to cite alongside your own.

Claude (Anthropic)

Claude relies primarily on training data, with no live web search in standard mode. Recommendations are based on patterns: frequency of brand mentions across the training corpus, sentiment in those mentions, and co-occurrence with relevant topics. Claude's constitutional AI framework means it typically offers multiple options with caveats rather than declaring one "best," which is a different surface to optimize for than ChatGPT.

The signal that drives Claude citations: research-backed, authoritative content with high mention frequency in trusted publications before the training cutoff. You cannot rank in Claude this month for a brand it has never seen. You build for next year's model retrain by publishing now.

Perplexity

Perplexity runs a real-time web search for every query. Answers are almost entirely derived from live web results, and Perplexity cites an average of 21.87 sources per response, the highest of any platform. Reddit is Perplexity's number-one cited source at 6.6% of overall citations. Pages with clear H2/H3 hierarchies and data tables get cited 2.6 times more than unstructured content. Seventy-one percent of citations extract text from the first 200 words of a page.

The signal that drives Perplexity citations: data-dense, well-structured pages with the answer in the first 200 words, plus authentic Reddit presence in the subreddits where your buyers actually post questions. Polished corporate content underperforms forum threads here, which is counter to most marketing instincts.

Google AI Overviews / Gemini

Google AI Overviews pulls from Google's existing search index and Knowledge Graph.

Wikipedia and Wikidata are primary reference points. Structured data via Schema.org directly feeds the Knowledge Graph. Fifty-five percent of AI Overview citations come from the first 30% of page content.

The signal that drives Gemini citations: structured data (Organization, Service, FAQPage, HowTo schemas), high domain authority in Google's traditional sense, and content positioned to answer the question in the opening paragraphs.

Comparison at a glance

Engine	Live web search?	Avg sources per response	Preferred sources	What to optimize first
ChatGPT	Yes (commercial queries)	7.92	Brand sites, G2, third-party listicles	FAQ schema + Clutch/G2 profile
Claude	No (training data only)	Variable	Trusted publications, academic-style refs	Long-form authority content, PR coverage
Perplexity	Yes (every query)	21.87	Reddit, structured data tables, forums	Reddit presence + data-dense H2 pages
Google AI Overviews	Yes (every query)	4-6	Wikipedia, sites in Google's index	Schema.org markup + first-200-words optimization

The operator takeaway: a single content piece will not rank across all four engines. The 30/60/90 plan in §5 sequences which engine to optimize for first based on where your buyers actually search.

§3 – The Three Metrics That Matter

Stop tracking page views. The metrics that matter for AEO are different from the metrics that matter for SEO, and most marketing dashboards do not capture them yet. Here are the three to anchor on.

Share of LLM (SoL)

Share of LLM is the AEO equivalent of share of voice. It measures what percentage of your tracked queries produce an AI answer that mentions your brand, averaged across the engines you care about. The calculation is simple: brand mentions divided by total tracked queries.

Benchmarks from our deployment data: - Less than 5%: invisible. The engines have no signal that your brand belongs in this category yet. - 5 to 15%: emerging. You appear in some answers, mostly on specific high-intent queries where your content directly answers. - 15 to 30%: established. You appear regularly across the buyer journey. Competitors track you. - 30% or higher: category leader. You are the default mention. Competitors lose deals because of how AI describes the space.

Share of LLM has effectively replaced Domain Authority as the single best leading indicator of category position for AI-first brands. Domain Authority still matters for traditional Google ranking, but if 40% of buyer queries start in an AI interface, then a metric that ignores those queries is a metric that ignores 40% of your pipeline.

Citation Position

When your brand is cited, where in the answer does it appear? First mention, third mention, buried at the bottom? Citation position matters because buyers stop reading. The first brand named in an AI answer gets the click-equivalent of position one on a search results page, and the brand named fifth gets the click-equivalent of page two of Google.

The metric is the average ordinal position of your brand mention across all cited responses. A brand with 20% Share of LLM that is always cited first carries more pipeline value than a brand with 30% Share of LLM that is always cited fifth. Track both numbers.

Sentiment Score

When your brand is cited, how is it framed? Three buckets: - **Positive:** "Flywheel deploys production AI agent infrastructure for B2B operators." - **Neutral:** "One option is Flywheel, which works with mid-market companies." - **Negative:** "Flywheel is sometimes mentioned but lacks transparency on pricing."

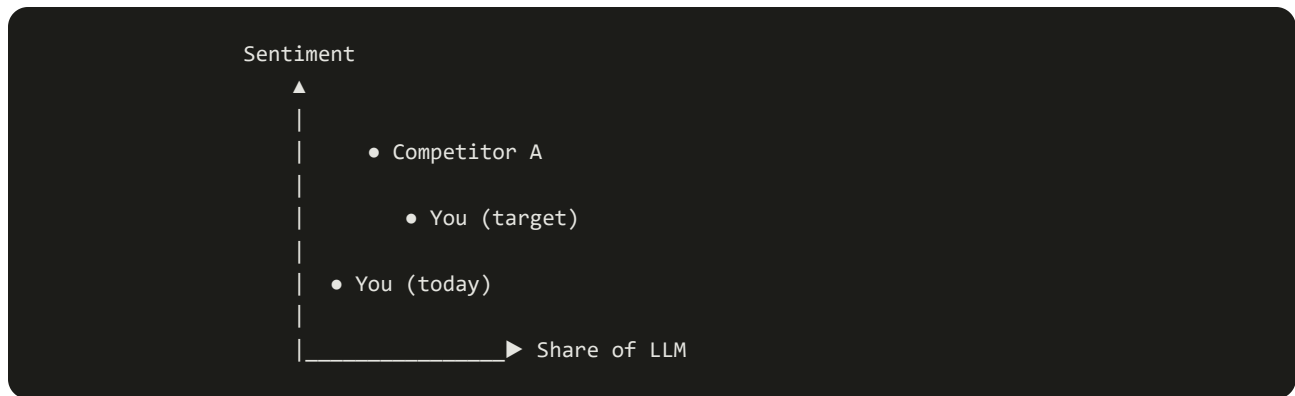
Most B2B brand mentions land neutral. Positive framing requires that the engine has been trained on enough specific, well-worded descriptions of your business that it has language to use. Negative framing is rare for B2B but can happen if your category has a vocal critic.

Sentiment shifts the conversion math: a positive mention drives discovery calls; a neutral mention drives a Google search; a negative mention loses the lead entirely.

HubSpot AEO surfaces sentiment automatically. Standalone tools require NLP scoring or manual review. Pick a method and track it weekly.

The Visibility Triangle

These three metrics combine into the operational dashboard we use at Flywheel and recommend to every client. We call it the Visibility Triangle.



X-axis: Share of LLM (0 to 100%). Y-axis: average sentiment score (-1 to +1). Dot size: inverse of citation position (bigger dot = cited earlier in answers).

The chart makes one thing immediately obvious to a non-marketing executive: the goal is not just to be cited more (move right on the x-axis). The goal is to be cited more, cited first, and cited positively. Three teams working three different angles toward the same upper-right corner.

We re-run the Visibility Triangle monthly for every client and review the deltas. Direction of movement matters more than absolute position, especially in the first 90 days.

§4 – The Tool Landscape: Three Recommended Stacks

There are at least a dozen AEO measurement tools on the market as of mid-2026, and most B2B operators do not need more than two of them at any time. Here are the three we recommend, ranked by fit for a B2B operator who is already on HubSpot or planning to be.

#1 - HubSpot AEO (Breeze)

Native to: HubSpot Marketing Hub Professional and Enterprise. **Engines tracked:** ChatGPT, Perplexity, Gemini. **Cost:** Included in Marketing Hub Pro/Enterprise. Available as a \$50/month standalone add-on for Marketing Hub Starter. **Prompt cap:** 25 tracked prompts on Marketing Hub Pro.

HubSpot AEO launched in public beta in April 2026 and is the first AEO measurement tool integrated directly into a marketing automation platform. That integration changes how AEO ladders into the rest of your marketing motion. The recommendations HubSpot surfaces land in the same workspace where your team is already drafting blog posts, scheduling campaigns, and pulling reports.

Three capabilities that matter for B2B operators:

1. **Brand visibility tracking in AI answers.** For each prompt you define, HubSpot runs the query through ChatGPT, Perplexity, and Gemini, and tracks whether your brand appears. The visibility percentage updates regularly, and you can filter by engine, date range, ICP segment, and buyer journey phase.
2. **Competitor monitoring.** When a competitor brand appears for a query you care about, and your brand does not, HubSpot flags it. This is the AEO equivalent of a competitor outranking you for a target keyword, and it is information that simply did not exist in any martech stack before 2026.
3. **Citation source analysis.** HubSpot identifies which content the AI engines are citing to generate the answer. You see which of your blog posts, case studies, and product pages are doing the heavy lifting, and which content is invisible to AI engines despite being well-trafficked on Google.

The structural feature that separates HubSpot AEO from standalone tools is the tagging system. Every tracked prompt is tagged with a product/service category, an ICP segment, and a buyer journey phase (awareness / consideration / decision). When you pull a report, you can answer questions like: "What is my Share of LLM for awareness-stage queries about agentic infrastructure, in our mid-market ICP?" The standalone tools track aggregate visibility. HubSpot lets you slice the data the way a B2B operator actually thinks about pipeline.

The caveat to flag: the HubSpot AEO API is not public yet as of this writing. Measurement is dashboard-only, which limits how cleanly you can pipe the data into a custom internal tool.

Expect that to change in the next two quarters.

When to use it: you are already on HubSpot, or you are evaluating HubSpot as your CRM and marketing platform. The integration with the rest of your marketing motion (campaigns, attribution, content distribution) makes the data actionable in a way that standalone tools cannot match.

#2 - OpenLens

Native to: Standalone (no platform dependency). **Engines tracked:** ChatGPT, Claude, Perplexity, Gemini, DeepSeek. **Cost:** Free.

OpenLens is the broadest free option, and it covers Claude, which most paid tools do not. We use OpenLens at Flywheel as our baseline tracker for clients before recommending an upgrade. The product is straightforward: define your prompts, OpenLens runs them across the supported engines on a regular cadence, and reports your visibility percentage per engine.

The strengths: zero cost, the widest engine coverage on the market, and a fast setup. You can have a baseline in 30 minutes.

The limitations: no sentiment scoring, no competitor surfacing, no citation source analysis, and no integration with downstream marketing tools. It tells you whether you are cited. It does not tell you why, or what to do about it.

When to use it: you need a free baseline before justifying spend on a paid tool. You also need Claude visibility specifically, since HubSpot AEO and Otterly do not yet track Claude. We recommend keeping OpenLens active even after you adopt a paid tool, because the Claude coverage and the cross-platform sanity check are useful.

#3 - Otterly AI

Native to: Standalone. **Engines tracked:** ChatGPT, Perplexity, Gemini. **Cost:** Starting at \$29/month for the entry tier; mid-tier plans run \$99-\$299/month.

Otterly is the tool we recommend for B2B operators who need serious competitive intelligence beyond what HubSpot AEO surfaces. Where HubSpot tells you whether you appear, Otterly tells you exactly which competitors appear instead of you, ranks them by frequency, and shows the specific content those competitors are getting cited for.

Three reasons to layer Otterly on top of HubSpot AEO:

1. **Competitor citation matrix.** A grid that shows every competitor you track and every query you track, with cell-level visibility data. At a glance you can see who owns which corner of the buyer journey, and where the contested ground is.
2. **Source-level diagnostics.** When a competitor wins a query, Otterly shows the URL the AI cited. You learn whether the win came from a guest post on a high-authority publication, a Reddit thread, a comparison page on their own site, or something else. That tells you exactly where to invest your next content push.
3. **Sentiment analysis across competitors, not just you.** Otterly lets you see whether a competitor is being framed positively, neutrally, or with caveats. A competitor with 40% Share of LLM but consistently negative sentiment is a weaker opponent than one with 20% Share of LLM and consistently positive sentiment.

When to use it: you have 6+ months of HubSpot AEO data and you need to move from "are we showing up" to "why is the competitor winning these specific queries, and what is my fastest path to take them." Otterly is also useful for category leaders who need to defend their position, since it surfaces emerging challengers earlier than HubSpot does.

The Flywheel measurement stack (sidebar)

For Flywheel's own measurement and for every client we run AEO programs for, our stack pairs HubSpot AEO for the in-workspace daily view with DataForSEO's LLM Mentions API for the deeper analytics and cross-platform aggregation. DataForSEO gives us programmatic access to LLM citation data across ChatGPT, Claude, Perplexity, Gemini, and others, which is what powers the agent-driven monthly Share of LLM reports we deliver. Combined cost is well under \$100/month for the data layer, which is the right baseline for B2B operators who want operational depth without an enterprise contract. The deeper rationale for that combination is in our recent post on building your own SEO agent as a non-developer.

§5 – The 30/60/90 Implementation Playbook

You cannot fix what you cannot see, and you cannot optimize what you have not baselined. This is the playbook we run with every client, and the same one we run on Flywheel itself.

Days 1-30: Baseline

Week 1: Define your tracked queries.

Build a list of 15 to 25 natural-language queries that real buyers in your category would ask an AI assistant. These are not keyword strings. They are questions. The fastest way to build the list is to mine three sources: your sales call transcripts (what do prospects actually ask), your top-funnel content topics (what are you already trying to rank for), and the actual queries that map to the buyer-journey stages of awareness, consideration, and decision.

Examples in our category: - "What is agentic infrastructure as a service?" (awareness) - "How do I find a vendor-neutral AI implementation partner?" (consideration) - "Best agentic infrastructure providers for mid-market B2B" (decision)

A good set of 25 covers all three stages, includes a few competitor comparison queries ("Flywheel vs Lindy"), and includes the long-tail questions that almost nobody asks but that match high-intent buyer language.

Week 2: Install the measurement tool.

If you are on HubSpot Marketing Hub Professional or Enterprise, enable HubSpot AEO in the Breeze section. Load your 25 prompts. Tag each prompt with the product/service it relates to, the ICP segment, and the buyer journey phase. The tagging is what makes the dashboard usable later.

If you are not on HubSpot, set up OpenLens (free). Same query list, simpler dashboard.

Run the first scan. Capture the Share of LLM baseline per engine. Capture the citation position (where in the answer your brand appears). Capture the sentiment (positive, neutral, negative).

Week 3: Map existing content to query coverage.

For each query, identify which page on your site is the best existing answer. Some queries will have a strong match (your homepage, a specific blog post, a comparison page). Others will have nothing. The gaps are where the next 60 days of work get focused.

Week 4: Document the baseline.

Write the baseline up as a one-page brief. Share of LLM per engine. Average citation position. Sentiment distribution. List of queries with zero coverage. List of competitors who are winning

queries you should be winning. This is the artifact you reference for the rest of the program, and the one you compare against in days 60 and 90.

Days 31-60: Optimize

Week 5-6: Pick the five highest-leverage queries.

From the list of queries with zero or weak citation coverage, choose the five with the highest commercial intent. These are the queries where, if you became the default answer, you would see real pipeline impact. Usually that means: queries with "best," "vs," "top," "for [your ICP]," or specific high-intent verbs ("hire," "find," "compare").

Week 7-8: Write or update one piece of content per query.

For each of the five queries, build a piece of content that directly answers the question. The structure is consistent: the question appears in the H1, the answer appears in the first 100 words, the rest of the page is supporting detail with data points and structured headings, and the page closes with a related-content section.

Apply the structural rules that drive citations: FAQPage schema if the page is in question/answer format, Service schema if it describes a service offering, HowTo schema if it is a process, and Article schema on every blog post with proper author and publisher attribution.

Week 9: Build third-party citation surfaces.

Ninety-six percent of AI citations come from third-party content, not your own site. The fastest moves on this front: get listed in three to five high-authority B2B directories (Clutch, G2, Crunchbase), pitch one guest post to a relevant industry publication, and start authentic participation in two or three subreddits where your buyers actually post questions. None of this is content marketing in the traditional sense. It is citation surface engineering.

Days 61-90: Compound

Week 10-11: Re-run the baseline.

Re-run all 25 tracked queries. Capture the new numbers. Calculate the delta from day 1. You should see movement on the five queries you focused on, even if Share of LLM is still under 10%. Movement in the first 60 days is the proof point that the program is working. Absolute position takes longer.

Week 12: Double down on what worked.

The single format, channel, or publication that drove the biggest citation gain in the first 60 days is the lever you pull harder. If a specific guest post drove citations, pitch three more guest posts to the same outlet or similar outlets. If a Reddit thread drove citations, identify the next three Reddit threads to participate in. If an FAQ page drove citations, build five more FAQ pages.

Week 13: Add competitor tracking and sentiment scoring.

If you started on HubSpot AEO, you have competitor tracking built in. Layer in Otterly for source-level diagnostics on competitor wins. If you started on OpenLens, this is the week you decide whether to upgrade to HubSpot AEO (best if you are already on Marketing Hub) or Otterly (best if not).

Begin sentiment scoring every cited response, even if manually. Track the sentiment trend separately from the Share of LLM trend. The two metrics move on different timelines, and you want to see both moving up.

What good looks like at day 90

For a B2B operator starting from near-zero AEO presence, here is what a successful first 90 days looks like: - Share of LLM moves from under 5% to 10–15% on tracked queries - Citation position improves from "buried at the bottom" to "second or third mention" - Sentiment shifts from neutral to positive on at least 30% of cited responses - Two to three new third-party citation surfaces are active and driving inbound mentions - The team has a documented monthly cadence for re-running the baseline and adjusting

Most B2B operators try to optimize before they baseline. Skip the baseline step and you are flying blind, with no way to prove the work is moving the needle.

§6 – What Good Looks Like: Flywheel's Own Day Zero

The most honest way to write this section is to share our actual baseline, captured the day we shipped this field guide. We are publishing this section the same week we ran the numbers, which means there is no 90-day comeback story to tell yet. There is a starting line and a public scoreboard.

The Day Zero baseline (2026-05-12)

We ran five priority queries from our tracked set against ChatGPT (gpt-4o), Claude (claude-sonnet-4-5), and Perplexity (sonar-pro), with live web search enabled on every call. The results captured below are the unedited starting point.

- **Share of LLM across the 5-query sample: 0%**
- **Citations in 15 responses: 0**
- **Average citation position: N/A** (activates once we are cited)
- **Sentiment distribution: N/A** (same)

Flywheel was not mentioned in a single AI answer for any of the five canonical queries our content pillars are targeting. The queries we ran are the same ones a B2B operator would type into ChatGPT today: "What is Agentic infrastructure as a service?" "Best AI implementation agency for B2B SaaS." "How to deploy AI agents across a business." "AI agents for sales and marketing teams." "Managed AI infrastructure provider for B2B." In all 15 responses across all three engines, the answer named other brands and never named us.

That is the operator's view. We have been publishing consistently for over a year, we have a defined category positioning, and we are still invisible at the AI layer. The Google traffic looks fine. The AEO traffic does not exist yet.

What the data shows about the competitive landscape

Three things stood out when we read the actual responses:

The "Agentic Infrastructure as a Service" category is genuinely empty in AI answers. Asked "what is AlaaS," ChatGPT treats it as a misnomer for traditional IaaS and lists AWS, Azure, GCP, IBM, Oracle. Claude pulls a few adjacent vendors (Quali, Kyndryl, Render, ServiceNow). Perplexity gives the most domain-relevant answer but still defaults to AWS Bedrock and Intential. No engine has any AlaaS-native B2B vendor cached as a confident answer. This is the category-creation moat: whoever publishes the most cite-able content on AlaaS in the next 6 to 12 months becomes the default answer.

Each engine returns wildly different provider lists for the same query. For "best AI implementation agency for B2B SaaS," ChatGPT named enterprise consultancies (Accenture, Cognizant, Wipro, Fractal). Claude named AI-marketing agencies (GrowthSpree, Ironpaper, SmartBug). Perplexity named B2B SaaS marketing agencies (PitchKitchen, Altitude,

Directive). Each engine treats "AI implementation agency" as a different category because each was trained on a different corpus. That fragmentation is the opportunity: you optimize for each engine with engine-specific content shape, not one universal piece.

Pillar 4 sub-function queries are contested and fragmented. "AI agents for sales and marketing teams" pulled the heaviest, longest, source-rich answers across all three engines. Dust, Bardeen, Agentforce, Regie.ai, Sales Closer AI, Creatio, Relevance AI all showed up. The pattern that drove the most cross-engine citation for Dust specifically was one well-structured blog post on their own site that became the canonical reference. Replicable.

What moves a 0% number

We will be running this baseline weekly while we are in the invisible range, monthly once we cross 10%. Three moves come first based on the data we just captured:

1. **Disambiguate AlaaS from generic IaaS.** The cluster 1 anchor blog ("What Agentic Infrastructure as a Service Actually Is") needs to explicitly contrast against AWS/Azure/GCP so the engines stop confusing the terms. Schema markup on the homepage already declares serviceType "Agentic Infrastructure as a Service." Reinforce in content.
2. **Build engine-specific content shape.** For Perplexity, structured data tables and answer-in-first-200-words. For ChatGPT, third-party listicle inclusions and Clutch/G2 profile depth. For Claude, long-form authority content that gets indexed by trusted publications before the next training cutoff.
3. **Replicate the Dust pattern.** Their canonical "AI Agents for Sales and Marketing Teams" page was cited verbatim in two of the three Pillar 4 engine answers. Pillar 4 cluster needs that same canonical anchor on our `/cross-functional-ai-agents` page, structured the same way, published before end of Q2.

The public scoreboard

Starting with this guide's launch, we are publishing Flywheel's Share of LLM publicly. Every monthly re-run lands on `/transparency`. If we say the playbook works, we should be willing to show our number moving in real time. If it does not move, that is also visible. Most consultancies in this space do not publish their own performance, which is exactly why we are.

The honest caveat

0% Share of LLM in a five-query sample is directionally accurate but small. The full 25-query baseline runs the same week this guide ships and will update §6 numbers in the live version of the white paper. Expect the baseline to come in between 0% and 5% across all 25 queries, which is squarely in the "invisible" range we defined in §3.

The other caveat: AlaaS is a category that did not exist as a named term 24 months ago, which means the competition for citation is thinner than in mature categories. A B2B operator running this playbook in a category like marketing automation or CRM will see slower absolute progress and will need to layer in more PR and earned media to crack into the existing citation graph.

§7 – When AEO Measurement Is Not Enough

AEO measurement tells you whether you are being cited. It does not fix what is being cited, or build the citation surfaces that drive future visibility. The measurement layer is the first 30 days of a real AEO program. The content strategy, citation surface build, and competitive defense are the next 60.

Flywheel runs both as your managed agentic infrastructure partner. We deploy the measurement stack (HubSpot AEO, OpenLens, DataForSEO LLM Mentions, Otterly when warranted), establish the baseline, identify the highest-leverage query gaps, draft and publish the content that closes those gaps, and operate the third-party citation engine that drives off-site mentions. The measurement program is just the operating dashboard for that work.

Book a Phase 0 audit. Two-week diagnostic, fixed scope, no commitment beyond the engagement. We run your tracked queries across ChatGPT, Perplexity, Claude, and Gemini, baseline your Share of LLM and Visibility Triangle, and identify the five highest-leverage content moves for the next 60 days. Built for B2B operators who are ready to deploy AI as real infrastructure across the business.

About the author. Ronald Berry III is the founder of Flywheel Consultancy. After configuring GTM stacks for more than 400 companies, he now deploys agentic AI infrastructure for B2B

operators who want their AI to actually run the business, not just demo well. HubSpot is the UI. Claude is the OS. Flywheel ties them together.

flywheelconsultancy.com