Answer Engine Optimization: How To Get Your Content Into AI Responses via @sejournal, @slobodanmanic



This is Part 2 in a five-part series on optimizing websites for the agentic web. Part 1 covered the evolution from SEO to AAIO and why the shift matters. This article gets practical: how AI systems actually select content, and what you can do about it.

AI Doesn’t Rank Pages. It Selects Fragments.

Traditional search ranks whole pages. AI search does something fundamentally different.

Microsoft’s Krishna Madhavan, principal product manager on the Bing team, described the shift in October 2025: AI assistants

“break content down, a process called parsing, into smaller, structured pieces that can be evaluated for authority and relevance. Those pieces are then assembled into answers, often drawing from multiple sources to create a single, coherent response.”

This is the core insight. AI doesn’t pick the best page and show it. It picks the best fragments from many pages and weaves them together. Your page might rank No. 1 on Google and still not get cited in an AI response if its content isn’t structured in fragments that AI can extract.

The numbers show the shift is real. According to the Conductor AEO/GEO Benchmarks Report (January 2026; 13,770 domains, 17 million AI responses), AI traffic now accounts for 1.08% of all website sessions, growing roughly 1% month over month. Microsoft reported that AI referrals to top websites spiked 357% year-over-year in June 2025, reaching 1.13 billion visits. Small numbers today, compounding fast.

One in four Google searches now triggers an AI Overview. In healthcare, it’s nearly one in two. The surface area is growing, and the content that fills these answers has to come from somewhere. The question is whether it comes from you.

The Research: What Actually Gets Cited

The academic research on what makes content citable in AI responses has matured rapidly. The foundational paper, “GEO: Generative Engine Optimization” (Princeton, IIT Delhi, Georgia Tech, published at KDD 2024), tested nine optimization strategies and found that GEO techniques could boost visibility by up to 40% in AI responses. The most effective single technique was citing credible sources, which produced a 115.1% visibility increase for websites that weren’t already ranking in the top positions.

A counterintuitive finding: Writing in an authoritative or persuasive tone did not improve AI visibility. AI systems don’t respond to rhetorical style. They respond to verifiable information.

Since then, 2025 brought a wave of follow-up research that tested these ideas on real production AI engines rather than simulated ones.

The University of Toronto study (September 2025) was the first large-scale analysis across ChatGPT, Perplexity, Gemini, and Claude. Their most striking finding: AI search overwhelmingly favors earned media. In consumer electronics, AI cited third-party authoritative sources 92.1% of the time, compared to Google’s 54.1%. Automotive showed a similar pattern at 81.9% versus 45.1%. In other words, it’s not just how you write content, but whose domain it appears on. Press coverage, product reviews on independent websites, and mentions on industry publications carry far more weight in AI responses than your own website.

Carnegie Mellon’s AutoGEO study (October 2025) used automated methods to discover what generative engines actually prefer. The results showed up to 50.99% improvement over the best baseline, with universal preferences emerging across engines: comprehensive topic coverage, factual accuracy with citations, clear logical structure with headings and lists, and direct answers to queries.

The GEO-16 framework (September 2025) analyzed 1,702 real citations from Brave, Google AI Overviews, and Perplexity. It identified 16 on-page quality factors that predict citation likelihood. The top three: metadata and freshness, semantic HTML, and structured data. Technical on-page factors matter as much as the quality of the writing itself.

And a reality check from Columbia and MIT’s ecommerce study (November 2025): of 15 common content rewriting heuristics, 10

produced negligible or negative results. The optimization strategies that did work converged toward truthfulness, user intent alignment, and competitive differentiation. Not tricks. Substance.

The overall pattern across all of this research: AI systems reward clarity, factual accuracy, and structure. They don’t reward marketing language, persuasion tactics, or keyword density.

Content Structure That Earns Citations

Based on the research and official guidance from Microsoft and Google, here’s what structurally makes content citable.

Heading hierarchy matters more than ever. Use descriptive H2 and H3 headings that each cover one specific idea. Microsoft lists strong headings as “signals that help AI know where a complete idea starts and ends.” Vague headings like “Learn More” or “Overview” give AI nothing to work with. A heading like “How AI parses content differently than search engines” tells the system exactly what the section contains.

Q&A format is native to AI. Write questions as headings with direct answers below them. Microsoft notes that “assistants can often lift these pairs word for word into AI-generated responses.” If your content answers the question someone asks an AI, and it’s structured as a clear question-and-answer pair, you’ve made the AI’s job easy.

Make content snippable. Bulleted and numbered lists, comparison tables, step-by-step instructions. These formats give AI clean, extractable fragments. A paragraph buried in a wall of text is harder for AI to isolate than the same information presented as a three-item list.

Front-load the answer. Start sections with the key

information, then provide context. If someone asks, “What temperature should I bake bread at?” and your content opens with a two-paragraph history of bread making before mentioning 375°F, you’ll lose the citation to a competitor who leads with the answer.

Keep sections self-contained. Each section should make sense on its own, without requiring the reader to have read the previous section. AI extracts fragments. If your fragment only makes sense in the context of the whole page, it won’t be selected.

An important technical note from Microsoft: “Don’t hide important answers in tabs or expandable menus: AI systems may not render hidden content, so key details can be skipped.” FAQ answers collapsed inside an expandable menu, product specs hidden behind tabs, content that requires interaction to reveal: it may all be invisible to AI. If information is important, it needs to be in the visible HTML.

Authority Signals For AI

E-E-A-T (Experience, Expertise, Authoritativeness, Trustworthiness) isn’t just a Google concept anymore. It’s what AI systems look for across the board, even if they don’t use the term.

Microsoft’s October 2025 guidance describes the baseline: success starts with content that is “fresh, authoritative, structured, and semantically clear.” On the clarity side, they’re specific: “avoid vague language. Terms like innovative or eco mean little without specifics. Instead, anchor claims in measurable facts.” Saying something is “next-gen” or “cutting-edge” without context leaves AI unsure how to classify it.

The research backs this up. The original GEO paper found that writing in a persuasive or authoritative tone did not improve

AI visibility. Facts and cited sources did. Marketing language doesn't impress algorithms.

This connects to the University of Toronto's finding about earned media dominance. AI systems trust third-party validation more than self-promotion. In consumer electronics, AI cited third-party authoritative sources 92.1% of the time compared to Google's 54.1%. The implication: getting your expertise published on industry websites, earning press coverage, and building a presence on authoritative platforms matters more for AI visibility than perfecting the copy on your own site.

Freshness is a signal, not a bonus. Stale content rarely gets cited. Krishna Madhavan said at Pubcon Cyber Week: "Stale or missing content will constrain the amount of retrieval we can do and push agents toward alternative sources."

Schema Markup: From Text To Knowledge

Microsoft's October 2025 post devotes an entire section to schema. They describe it as code that "turns plain text into structured data that machines can interpret with confidence." Schema can label your content as a product, review, FAQ, or event, giving AI systems explicit context instead of forcing them to guess. Krishna Madhavan reinforced this at Pubcon: "Schemas are super useful. They help the system discern exactly what your information is without us having to guess."

The GEO-16 framework confirms this from the academic side. Structured data was one of the top three factors predicting AI citation likelihood, alongside metadata/freshness and semantic HTML.

The schema types that matter most for AI visibility:

- **FAQPage** for question-and-answer content (directly maps

to how AI formats responses).
- **HowTo** for step-by-step instructions.
- **Product** with Offer, AggregateRating, and Review for ecommerce.
- **Article/BlogPosting** for content with clear authorship and dates.
- **Organization** for business identity.

Pair structured data with IndexNow for freshness. As the Bing Webmaster Blog put it: “IndexNow tells search engines that something has changed, while structured data tells them what has changed. Together, they improve both speed and accuracy in indexing.”

Crawler Permissions: Who Gets In

AI search engines use distinct crawlers, and most let you control training and search access separately. Here’s who to allow.

Bot	Platform	Purpose	Robots.txt Token
OAI-SearchBot	ChatGPT	Search index	OAI-SearchBot
GPTBot	OpenAI	Model training	GPTBot
ChatGPT-User	ChatGPT	On-demand browsing	ChatGPT-User
Bingbot	Microsoft Copilot	Search + AI	Bingbot
Googlebot	Google AI Overviews	Search + AI	Googlebot
Google-Extended	Google	Gemini training	Google-Extended
PerplexityBot	Perplexity	Search + index	PerplexityBot

Bot	Platform	Purpose	Robots.txt Token
Perplexity-User	Perplexity	On-demand browsing	Perplexity-User
ClaudeBot	Anthropic	Training + retrieval	ClaudeBot

A sensible robots.txt configuration might allow search crawlers while blocking training:

```
User-agent: OAI-SearchBot
Allow: / User-agent: ChatGPT-User
Allow: / User-agent: GPTBot
Disallow: / User-agent: Google-Extended
Disallow: /
```

OpenAI provides the cleanest bot separation. You can allow OAI-SearchBot (so your content appears in ChatGPT search) while blocking GPTBot (so it's not used for model training). Google's controls are less granular: blocking Google-Extended prevents Gemini training but has no effect on AI Overviews, which use Googlebot.

OpenAI also offers the most specific technical recommendation of any AI search provider. For their Atlas browser (which uses a standard Chrome user agent, not a bot identifier), they recommend following WAI-ARIA best practices: "Add descriptive roles, labels, and states to interactive elements like buttons, menus, and forms. This helps ChatGPT recognize what each element does and interact with your site more accurately." Accessibility and AI agent compatibility are the same work.

A caveat on Perplexity: while their documentation states they respect robots.txt, Cloudflare documented in August 2025 that Perplexity uses undeclared crawlers with rotating IPs and spoofed browser user agents to bypass no-crawl directives. This is a contested claim, but it's worth knowing.

For revenue, Perplexity is the only platform currently

offering publisher compensation. Their Comet Plus program provides an 80/20 revenue split (publishers keep 80%) across direct visits, search citations, and agent actions.

Google Vs. Microsoft: Two Philosophies

The contrast between Google and Microsoft on AEO is striking enough to be its own story.

Google says: just do good SEO. Their official documentation is deliberately minimalist: “There are no additional requirements to appear in AI Overviews or AI Mode, nor other special optimizations necessary.” They add that you “don’t need to create new machine readable files, AI text files, or markup to appear in these features.”

Google recommends helpful, reliable, people-first content demonstrating E-E-A-T. Standard structured data. Good page experience. Technical basics. Nothing AI-specific.

Microsoft says: here’s the playbook. Their October 2025 blog post and January 2026 guide provide detailed, actionable guidance. Specific heading structures. Schema recommendations. Content formatting rules. Concrete examples (an AEO product description vs. a GEO product description). Warnings about content hidden in tabs and expandable menus. A framework for thinking about crawled data, product feeds, and live website data as three distinct layers.

What explains the difference? Partly market position. Google dominates search and has less incentive to help publishers optimize for AI features that might reduce clicks to their websites. Microsoft, with Bing’s roughly 8% market share, benefits from providing publishers with reasons to optimize specifically for their ecosystem.

But there’s a practical takeaway: Microsoft’s guidance isn’t

Bing-specific. The principles of structured content, clear headings, snippable formats, schema markup, and expert authority are universal. Following Microsoft's playbook improves your content for every AI system, including Google's. Google just won't tell you that.

Measuring AI Visibility

This is the hard part. Traditional SEO has Google Search Console. AI visibility is still fragmented.

Ahrefs analyzed 1.9 million citations from 1 million AI Overviews and found that 76% of citations come from pages already ranking in Google's top 10. The median ranking for the most-cited URLs was position 2. Traditional ranking still matters for AI citation, but being No. 1 is "a coin flip at best" for getting cited.

The traffic impact is significant. Ahrefs found that AI Overviews correlate with 58% lower click-through rates for the No. 1 position. Seer Interactive reported a 61% organic CTR drop for queries with AI Overviews. But being cited within the AI Overview gives 35% more organic clicks compared to not being cited. Citation is the new ranking.

For tracking, the tool landscape is emerging:

Tool	What It Tracks	Starting Price
Profound	Citations across ChatGPT, Perplexity, Copilot, Google AIOs	From $99/mo
Peec.ai	Brand mentions across ChatGPT, Gemini, Claude, Perplexity	From ~$95/mo
Advanced Web Ranking	AIO presence tracking in Google	Included in plans
Bing Webmaster Tools	AI Performance Report for Copilot	Free

Bing Webmaster Tools is the easiest starting point. It's free, and the new AI Performance Report shows how your content performs in Copilot citations. For ChatGPT specifically, track utm_source=chatgpt.com in your analytics. OpenAI automatically appends this to referral URLs.

Conductor's January 2026 report found that 87.4% of AI referral traffic comes from ChatGPT. That's one platform dominating the space, which makes tracking it particularly important.

Key Takeaways

- **AI selects fragments, not pages.** Structure your content in self-contained, extractable sections with descriptive headings that signal where each idea starts and ends.
- **Clarity beats persuasion.** Factual accuracy, cited sources, and direct answers outperform authoritative tone and marketing language. The research consistently shows this.
- **Earned media dominates brand content in AI citations.** Press coverage, third-party reviews, and authoritative mentions on other websites carry more weight than your own pages. Build presence beyond your domain.
- **Schema markup is a force multiplier.** FAQPage, HowTo, Product, and Article schemas make your content machine-readable. Pair with IndexNow for freshness.
- **Follow Microsoft's playbook, even for Google.** Google says "just do good SEO." Microsoft provides specific, actionable guidance that improves content for every AI system, Google's included.
- **Separate training from search in your robots.txt.** Allow search crawlers (OAI-SearchBot, Bingbot, PerplexityBot) while blocking training crawlers (GPTBot, Google-Extended) if that's your preference. You have more control than you might think.
- **Track AI visibility now.** Use Bing Webmaster Tools

(free), monitor utm_source=chatgpt.com in analytics, and consider dedicated tools as the measurement space matures.

Traditional SEO asked: “How do I rank?” AEO asks: “How do I become the fragment that gets selected?” The answer isn’t a single trick. It’s clear structure, verifiable expertise, and content that AI can confidently extract and cite.

Up next in Part 3: the protocols powering the agentic web, including MCP, A2A, NLWeb, and AGENTS.md, and how they fit together.

More Resources:

This was originally published on No Hacks.

Featured Image: Meepian Graphic/Shutterstock

https://www.searchenginejournal.com/answer-engine-optimization-how-to-get-your-content-into-ai-responses/570055/