

AI RECOMMENDATION STABILITY

What the Data Says and What It Means for Your Brand

Every day, millions of buying decisions begin with a single AI recommendation. Unlike traditional search, buyers increasingly receive one confident answer instead of a page of links. But how stable are those recommendations? Can brands earn an AI citation and expect to keep it? Our research shows the answer is no.

AI recommendations are dynamic. They shift across searches, across AI engines, and over time. That creates both risk and opportunity. Brands that understand how AI chooses recommendations—and continuously measure and improve their visibility—can capture high-value demand before competitors do. The stakes are significant. AI-referred visitors consistently convert better than traditional search traffic. AI visibility is no longer simply a marketing metric. It is becoming a revenue metric.

~50%

Chance the same top brand appears twice in a row in real-world conditions

27%

Of questions produced the same top brand across all four AI engines

10%

Median brand appearance rate — the typical brand shows up in 1 in 10 responses

29%

Of all AI source citations come from just 10 publications

WHAT IS AI RECOMMENDATION STABILITY?

AI Recommendation Stability measures how consistently an AI engine recommends the same brand for the same buying question over repeated searches.

High Recommendation Stability means AI has largely settled on a preferred brand.

Low Recommendation Stability means the category remains contested, creating opportunities for challenger brands.

The Study

This study builds on two published academic foundations. Aggarwal et al. (2024) established that brands can improve AI citation rates by up to 40% through content changes such as adding statistics, authoritative quotes, and source citations, and that traditional SEO tactics including keyword stuffing have no effect in AI environments. Yu et al. (2026) extended that work by showing that content structure independently influences citation behavior across all major engine architectures. What neither study addressed is whether those citation gains persist over time or hold across different engines. That is the question this study was designed to answer.

Noriko Yokoi and Thorsten Linz ran 26 unaided buyer questions across ChatGPT, Gemini, Claude, and Perplexity, generating 2,580 total runs across three controlled conditions. Unaided means the question names a need, never a brand. Any brand name in the response was put there by the AI, not the prompt. All hypotheses, queries, and analysis rules were locked before data collection started. The central question was whether AI recommendations behave like a fixed rank or a shifting probability. The answer is the latter.

1. Recommendation Stability: AI recommendations are not stable, even under ideal conditions

With randomness minimized and web search off, the same question returned the same top brand 78 to 87 percent of the time, depending on the engine. That means in the most controlled possible setting, roughly one in eight identical queries still produces a different top brand. Turn on live web search, which is how real users experience these tools, and consistency drops to 48 to 68 percent. The driver is the web retrieval layer. Different sources get pulled on different executions, and different sources produce different brand recommendations.

Engine	Floor (controlled)	Sampled	Web (real world)	Drop
ChatGPT	0.78	0.78	0.48	30 points
Gemini	0.78	0.67	0.54	24 points
Claude	0.87	0.69	0.50	37 points
Perplexity	N/A	N/A	0.68	Web-only engine

Where the opportunity is

When AI engines have not settled on a brand for a given question, that slot is open. Instability is not a measurement problem to be managed around. It is a signal that the AI has not decided yet. Consistent coverage in the right publications, before a competitor gets there, is what tips the outcome.

Instability within a single AI engine is only part of the story. Once buyers move between ChatGPT, Gemini, Claude, and Perplexity, the differences become even more pronounced. Understanding these differences is essential because recommendation opportunities often exist on one engine long before they emerge on another.

2. Engine Divergence: Each AI engine makes its own call, and they often disagree

Google dominated search and produced largely consistent rankings. That is not how AI works. Across 26 questions tested on all four engines simultaneously, only 27 percent produced the same top brand across all four. Three out of four questions had at least one engine go its own way.

Question	ChatGPT	Gemini	Claude	Perplexity
Best ADHD telehealth	Talkiatry	Talkiatry	Talkiatry	Talkiatry
Best PI law firms (L.A.)	Panish Shea	Panish Shea	Greene Broillet	Panish Shea
Best robo-advisors	Vanguard	Vanguard	Fidelity Go	Vanguard
Best weight-loss Rx	Form Health	Ro Body	Ro	ShedRx
Best HNW RIA	Pathstone	Mercer Global	Cresset Capital	Creative Planning

What this means strategically

A brand can lead on ChatGPT and be invisible on Perplexity. A competitor who owns the top slot on one engine may not appear at all on another. Engine divergence is a competitive map. The gaps are real, they are measurable, and they are claimable.

Recommendation Probability: Most brands appear in fewer than half of responses

The median brand in this study appeared in 10 percent of responses for its own category question. Only 2.4 percent of brand-question combinations achieved a 100 percent appearance rate. 85.7 percent fell below 50 percent.

Appearance rate	Share of all brand-question combinations
100% cited in every response	2.4%
80% or above, highly reliable	6.5%
Below 50%, fewer than half of responses	85.7%
Median appearance rate	10%

If your brand owns the obvious category question, the job is defensive: protect your gatekeeper coverage and watch for drift. If you do not own it, the better play is finding the adjacent questions, by use case, geography, or buyer type, where no brand has claimed the position yet.

3. Gatekeeper Coverage: A small number of publications control most of what AI recommends

Of 376 unique domains cited across all engines, the top 10 accounted for 28.6 percent of all citations. Chambers alone was cited 250 times, representing 8.2 percent of all source citations across all four engines.

#	Publication	Times Cited	% of Total	Relevant for
1	Chambers	250	8.2%	Legal services
2	Forbes	122	4.0%	Finance and business
3	TechRadar	110	3.6%	B2B and consumer tech
4	U.S. News Best Law Firms	70	2.3%	Legal services
5	NerdWallet	51	1.7%	Personal finance
6	Sleep Foundation	43	1.4%	Consumer health

Earned media is your AI strategy

A Chambers ranking, a Forbes feature, a NerdWallet review. These do not just build reputation. They are the retrieval signals that drive AI citation. Getting into the right publications before a competitor locks in the position is the new first-mover advantage.

4. Market Consensus: How fragmented your market is, predicts how hard the opportunity is to capture

Categories with a recognized market leader show high recommendation consistency. Categories without one show near-zero agreement across engines. For brands in fragmented markets, inconsistency is not the problem. It is the opening.

Market type	Consistency	What it signals	The move
Clear category leader	0.68 to 1.00	AI has settled	Defend gatekeeper coverage and monitor for drift
Moderate competition	0.40 to 0.70	Some consensus, gaps remain	Find the sub-questions where you can lead
Fragmented, no leader	0.15 to 0.35	AI has not decided	Move first: gatekeeper coverage now, before a competitor does

What this means for marketers

1. Find the prompts you can actually win

If the category leader owns the obvious category question, stop competing there first. Map the adjacent questions by use case, geography, buyer segment, or stage in the purchase funnel. A brand that owns the awareness prompt may have left the consideration and decision prompts wide open. Those are the slots available to you.

2. Treat engine disagreement as a map

When engines name different brands for the same question, no single brand owns that space yet. That divergence tells you where AI has not reached consensus. Confirm it through repeated measurement, then invest in gatekeeper coverage before a competitor does.

3. Audit gatekeeper publications before anything else

Identify the three to five publications AI engines cite most in your category. Coverage in those outlets is the highest-leverage move available. Your own website is secondary.

4. Measure across all four engines, not just one

A brand can lead on ChatGPT and be invisible on Perplexity. Single-engine measurement misses most of the picture. As Meta AI, Apple Intelligence, and Grok grow, the measurement set will need to expand beyond the current four.

5. Measure appearance rate, not yes or no

The question is not whether your brand shows up. It is what percentage of the time it shows up, for which questions, on which engines, and whether that number is moving.

6. Track continuously, not quarterly

AI models update constantly. Sources shift. A brand's visibility in June is not a reliable guide to September. Monthly tracking is the minimum. Weekly for contested categories.

The Bottom Line

AI visibility is not a ranking you win once. It is a probability that changes with every query, every AI engine, and every model update. The brands that consistently appear are the ones that earn trust from the publications AI relies on, expand coverage into unanswered buying questions, and continuously measure Recommendation Stability across every major AI platform. Recommendation Stability is becoming the new Share of Voice for the AI era. The brands that measure it—and improve it—will be the brands AI recommends tomorrow.

Full white paper and continuous monitoring: findabl.app | ny@3cubed.ai

References

Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., and Deshpande, A. GEO: Generative Engine Optimization. KDD 2024. arXiv:2311.09735.

Yu, et al. Structural Feature Engineering for Generative Engine Optimization. arXiv:2603.29979, 2026.

Ahrefs. Does AI Search Traffic Convert Better Than Traditional Search? June 2025. ahrefs.com.

Adobe Digital Insights. AI-Driven Traffic Surges Across Industries. Q1 2026. business.adobe.com.