AI Search is Working. How to Prove It With Real Tests. via @sejournal, @hethr_campbell



Adding FAQ sections to a set of test pages lifted AI citations. Removing them dropped citations back down.

That reversion is the difference between correlation and causation, and almost no team measuring AI search today can produce it.

That standard of proof anchored the latest SEJ webinar with seoClarity's Mark Traphagen, VP of Product Marketing & Training, Mihir Naik, Senior Product Manager, AI, and Suraj Lalchandani, Sr. IT Project Manager. Their core argument: "Visibility scores tell you if you showed up. Page-level performance and split testing tell you if what you did actually mattered."

The session walked through the split testing methodology seoClarity's enterprise clients run across ChatGPT, Claude, Perplexity, Gemini, and Google's AI surfaces: how to build a

funnel-spanning golden set of prompts, how to construct a control group when LLMs will not let you A/B test, and where Google’s new first-party Search Console AI data fits.

They also shared results from three real client tests, including the one change that moved citations and two results nobody in the room predicted.

Watch the full webinar on demand to get the complete testing methodology.

Can You Finally See AI Search Visibility In Google Search Console?

For a subset of sites, yes. On June 3, Google launched dedicated Search Console reports for AI Overviews and AI Mode, showing page by page how often each URL appears inside Google’s AI search features.

Lalchandani called it the biggest measurement upgrade AI search testing has received. “This has been the hardest thing to measure in AI search. Everyone was sampling. Everyone was inferring. But now Google is just giving it to you.”

First-party data straight from the source carries a different level of trust than any third-party tool. But the team was direct about the limitations: the new reports cover only part of what an AI search testing program needs, and ChatGPT, Claude, and Perplexity still require structured third-party tracking.

In the session, the team maps exactly which gaps the new reports close, which they leave open, and the platform-by-platform reference for what each AI engine can crawl and render.

Action item: Check Search Console for the new AI reports, then see where first-party data fits your testing program before you build around it.

Which Prompts Should You Test First In AI Search?

The ones where you are almost winning. The team builds a golden set of prompts spanning the full AI search funnel, awareness through retention, with every prompt tagged by stage, then sorts each prompt into tiers by where the brand currently stands in the AI's response.

Tier 1 prompts are the easy wins. As Lalchandani put it, "You're relevant, but AI just hasn't been given a URL worth linking to."

Tier 2 is the heavier lift, and one bucket of prompts gets dropped from testing entirely, a move that surprised many attendees.

The sequencing is deliberate: early wins buy the political capital to run harder tests later. The session covers how to build and tag the golden prompt set, how the tiers are defined, and the tracking unit that pairs each prompt with the exact page you want cited.

How Do You Run A Split Test On An LLM?

You cannot split live traffic 50-50, so you build a control group instead: a set of correlated pages that acts as your noise filter against model updates and algorithmic shifts.

"Without a control group, every result would be guesswork," Lalchandani said. "With one, you can tell a real win from the background noise."

Timing is the discipline most teams skip. The methodology sets a specific baseline period before any change goes live and a minimum test window after, because AI search does not respond overnight the way traditional SEO sometimes does. Cut the window short and, in Lalchandani's words, "you could be reading noise."

Every test lands in one of three outcomes, and each one tells you something about your hypothesis. The full session walks through how to construct the correlated control group, the exact baseline and test windows, and how to read all three outcomes.

Watch the full webinar on demand to get the complete test setup.

The FAQ Test That Proved Causation, And Two Tests That Did Not

seoClarity ran the same methodology for three clients and got three very different outcomes, which is exactly the point.

The FAQ test was the clear win. With roughly 1,000 prompts under measurement, adding FAQ sections to test pages pushed citations up versus control, and they stayed elevated as long as the change was live. Then the team reverted the change. "The citations fell back down. That's the second half of proof. Not that citations just went up when we added FAQs, but that they went back down when we took them away. That's causation, not correlation."

The other two tests, one on meta descriptions and one on listicle formatting, ended very differently, and the reasons why hold lessons for anyone about to invest in either tactic. See how both tests played out in the full session.

Naik's framing: every result is a win, because you have evidence instead of guesses. That is more than most teams in

AI search have today.

The session also lays out the schema and markdown test blueprints, two of the most argued questions in AEO right now, plus a set of fast structural tests for high-value templates you can run in a few weeks.

Q&A: Most Helpful Questions from the Webinar

Q: How do you measure AI authority when there is no clean authority metric?

“AI authority is basically how much the model trusts you as a source for this topic. I don’t think there’s a clean number for it or a single number for it, but there’s a couple of signals that you can stack to give you kind of a working picture.”

Lalchandani named four stackable signals, starting with citation share on your top prompts and cross-engine consistency, because “consistency across engines just means that you become the authoritative source in your category for specific kinds of questions.” He walks through all four, and how to track them, in the full session.

Q: Can AI bots read FAQ answers hidden behind collapsible toggles?

“Collapsible can mean many different things. It’s how you are having it collapsible.”

It depends entirely on implementation: one common setup keeps collapsed FAQs fully readable to AI search engines and Google, and another makes the content invisible to both, because “even Google will not click around on your site.” Lalchandani explains which is which in the recording, with his standing

advice attached: “If you’re unsure of something, just test it out. It takes effort, but it’ll give you a sure answer.”

Q: What is the ROI of an AI citation that does not drive referral traffic?

“You want to be cited because you are controlling the answer that is actually going to be showing up.”

Even without a click, Naik explained, your cited page shapes the narrative inside the answer, especially in comparison queries where citations do the heavy work of positioning both brands. The question shifts from traffic to representation: are your USPs highlighted correctly, is the comparison set right, are inaccuracies surfacing. Lalchandani added a cautionary example from a real restaurant client that shows exactly what happens when AI cannot reach your content, told in full in the recording.

Q: Is traditional SEO still a factor in moving the AI findability needle?

“Absolutely. It is foundational. It is the foundation.”

Traphagan noted that seoClarity’s longest-standing clients, the ones with well-optimized content and technically healthy sites, are also performing best in AI search, with AI optimization as the extra layer on top. Lalchandani added: “When we run tests with our clients, we’ve rarely, if ever, found a situation where something works for SEO and does not work for AI search.”

Watch the Full Webinar

The on-demand recording contains everything the recap holds back: the golden prompt set build, the tier definitions, the control group construction with exact baseline and test

windows, the platform-by-platform crawler reference, the meta description and listicle results, and the schema and markdown test blueprints. Register to watch the full session on demand.

https://www.searchenginejournal.com/ai-search-is-working-how-to-prove-it-with-real-tests-recap/583306/