Being Found Was Last Decade's Problem

How to Make Your Firm Readable to AI Search

What the evidence says works, and how much of the rest was sold to you

> The room could tell. The machine can't.

RELEVANCE
CREDIBILITY
PROOF
SAFETY
ACCESS

J.R. Worden

UNGATED MARKETING RESOURCE LIBRARY

Being Found Was Last Decade's Problem

How to Make Your Firm Readable to AI Search

Questions this paper answers

- How do we distinguish useful AI-search preparation from cosmetic optimization?
- How do we write a paragraph that still makes sense when AI extracts it?
- Why should we question an AI-visibility score based on a single answer?

The room had a second test. The machine can miss it.

In 1973, three researchers hired an actor.

They gave him a name, Dr. Myron L. Fox, a curriculum vitae with nothing true in it, and a specialty: the application of mathematics to human behavior. Then they put him in front of fifty-five psychiatrists, psychologists, social workers, educators, and administrators to deliver a lecture titled "Mathematical Game Theory as Applied to Physician Education."

The lecture was built to be worthless. Double talk, invented terminology, non sequiturs, statements that contradicted statements he'd made ninety seconds earlier, the whole thing delivered with warmth, timing, and the occasional joke.

The evaluations came back excellent. All three audiences, at p less than .001. One respondent wrote that he had a warm manner. Another said he was captivating. A third reported having read Dr. Fox's publications, which was an impressive feat, given that Dr. Fox was an actor and the publications didn't exist.

That's where the story usually stops. You've probably heard it as proof that expertise is theater and audiences are gullible, which is a satisfying thing to believe about other people.

It's also wrong, and the correction is more useful than the story.

In 2014, Eyal Peer and Elisha Babad went back through the original data in the *Journal of Educational Psychology* and found that Naftulin and his colleagues had overread their own results. The audience did rate the delivery highly. When the questionnaire asked them directly whether they'd learned anything, they said no. Somewhere between 59 and 70 percent gave favorable evaluations overall. Between 27 and 37 percent reported learning something. The gap was statistically significant, and it had been sitting in the 1973 numbers the entire time, waiting forty-one years for somebody to look.

So the room wasn't fooled.

The room was doing something considerably more interesting than being fooled. Those fifty-five people held two judgments simultaneously, kept them separate, and reported both of them accurately. *That was an enjoyable hour. I got nothing out of it.*

I've watched people do exactly that in conference rooms for three decades. They'll compliment the presentation on the way out and never call.

Now put that next to the thing sorting your firm.

A retrieval system can select material that looks rigorous without establishing that every claim is sound. Specificity, structure and citations can help it locate an answer, but those signals are not the same as verifying the underlying evidence. Models can compare sources and use checking tools; they can also accept a polished error. The firm remains responsible for what it publishes.

The room had a second test. The machine can miss it.

Everything else in this paper follows from that one asymmetry.

What Got Solved

For twenty-five years, the marketing problem in professional services was findability. Get the firm into the directory, then the phone book, then the search results, then page one of the search results. An entire industry grew up around the question of whether a buyer could locate you.

That question is over. Your next client can find forty firms in four seconds, including eleven they'd never have encountered in 1998 and three that don't have an office within a thousand miles of them.

Findability got solved, then it got commoditized, and commoditized capabilities stop being advantages. I watched that happen to my own company from the inside, which is why I'm confident about the shape of it and cautious about the timing. What replaced it is a harder problem with a worse name.

Interpretability, which is whether the thing doing the finding can work out what you are, who you serve, and what you're for, with enough confidence to put you in a set of three.

Most firms are still spending against the solved problem, which is the whole reason this paper exists. It's also the reason the industry selling to those firms has had such a productive eighteen months.

The Duck

Let me show you what the machine is doing, because a demonstration beats an explanation here.

An SEO researcher named Mark Williams-Cook built a web page for a fictional t-shirt company called DUCK YEA. The visible page said nothing about where the company was located. No address anywhere in the text a person would read.

He then added structured data to the page, the JSON-LD markup that firms pay agencies to install, and he made it deliberately, aggressively invalid. The context URL pointed at a schema that doesn't exist. The business type was "MallardEnterprise," which isn't a thing. The properties included waddleStyle, nestingGrounds, reedNumber, puddle, featherCode, and quackVolume. It was syntactically valid JSON and, as Williams-Cook put it, "as far as Schema.org is concerned this is unmitigated nonsense."

Buried in that nonsense was an address: Reed Number 77, The Muddy Bank, South Pondshire, DK99 YEA, United Queendom.

Then he asked ChatGPT and Perplexity where DUCK YEA was located.

Both of them told him. Reed Number 77, The Muddy Bank, the whole thing.

Perplexity added a detail that's worth pausing on. It said it had found the answer in the page's embedded structured data.

It hadn't. There was no valid structured data on that page for anything to be found in. The model had read some strange-looking text near the top of a document, pulled an address-shaped string out of it, and then narrated a parsing step it never performed, in the confident register of a system explaining its work.

Sit with that one for a second, because it's the whole paper in a single experiment. What the machine read was slightly weird prose. What it then reported reading was a data structure. It described its own process wrong, fluently, without a flicker of hesitation.

Williams-Cook is careful about what this proves, and I want to be careful with him, because the care is the point. His own caveat: "What it does not, on its own, prove is that LLMs ignore schema entirely. A system that consulted schema *and* fell back to text extraction would produce the same answer here."

The man who ran the experiment is the one refusing to overclaim it. Everybody quoting the experiment has been less disciplined than the person who conducted it, which tells you roughly everything about the evidentiary standards in this category right now.

Two More Things You've Been Sold

The schema question got a bigger test.

Ahrefs tracked 1,885 pages that added JSON-LD markup between August 2025 and March 2026, matched against roughly 4,000 control pages, and measured AI citations across three systems in

thirty-day windows before and after. Google AI Mode went up 2.4 percent. ChatGPT went up 2.2 percent. Both are indistinguishable from noise. AI Overviews went *down* 4.6 percent, and that one was statistically significant, at roughly one-in-2,500 odds of happening by chance.

Ahrefs sells SEO software, so you should know that a finding this deflationary runs against their own commercial interest, which makes it more believable rather than less.

The honest caveat matters more than the headline, and Ahrefs states it themselves. Every page in that dataset already had 100-plus AI Overview citations before any schema was added. This is a study of pages that were already winning. It shows schema adds nothing once you're visible. It says nothing about whether schema helps a page that's invisible today, and the authors say so plainly.

Then there's llms.txt, which had a good run.

The idea was that you'd publish a file telling AI crawlers what your site is and how to read it, the way robots.txt tells search engines what to skip. Agencies sold the installation. Conference speakers recommended it. Ahrefs scanned 137,210 domains, found roughly 38,000 with a valid file in place, and discovered that 97 percent of those files had never been requested by anything. Not once. Google's John Mueller called the format "purely speculative for now," which is diplomatic. That quote traces back through a single report, so hold it loosely.

Roughly thirty-eight thousand websites published the file; most showed no requests in that study.

I keep coming back to a line I wrote about résumés a while ago, because it fits here better than it did there. That's the professional services version of dressing for the scanner instead of the meeting. You can spend a whole budget getting machine-readable in ways nothing reads, and the tell is always the same: the tactic is cheap, universal, and requires you to change nothing about what you think.

You Cannot Rank in a System That Does Not Rank

Here's the one that should change how you buy.

SparkToro ran twelve prompts across three AI assistants, executing them 2,961 times total. They wanted to know something basic. If you ask the same assistant the same question twice, do you get the same list of companies?

You don't.

Take any two of a hundred runs of the same prompt. The odds that the same set of brands comes back are worse than one in a hundred. The odds that they come back in the same order are worse than one in a thousand.

Now read the phrase "our AI visibility ranking" one more time.

There is no stable universal ranking in those repeated answers. There is a distribution, and the thing being sold as a rank is one sample from it, taken once, on a Tuesday, and put on a dashboard with a number next to it that implies a precision the underlying system doesn't possess.

Julius Schulte worked out what it'd take to measure this honestly, in a paper with the best title in the field: *Don't Measure Once*. To get the standard error on a brand-visibility estimate below a tenth, you need on the order of seven runs per prompt per day. Source-level coverage needs eight. And you need a window of three to four weeks before the number settles into a range of roughly 0.05 to 0.08.

Seven runs a day for a month, per prompt, to produce one defensible number.

Ask the next vendor who shows you an AI visibility score how many times they ran it. If the answer is once a day, or once a week, you're buying noise with a decimal point on it. That isn't a reason to stop measuring. It's a reason to stop believing a number that moved four points last month meant anything happened.

What the Evidence Supports

So what does move?

The most useful thing published on this is a study by Vishwakarma and colleagues, presented at SIGIR in 2026: 252,000 trials across six models, with brand names stripped out so the models were judging content properties instead of reputation. That last design choice is what makes it worth your attention. It isolates the thing you can change.

Two tiers came out of it, and the difference between them is the difference between getting in the door and winning the room.

The four gatekeepers. Each of these carries an odds ratio above ten thousand, which in plain terms means failing one removes you. No points deducted, no partial credit, no making it up further down the page.

Exact topical match to the question as the buyer asked it. Price or cost information present on the page. Position in the retrieved context. A recent, visible timestamp.

The seven differentiators. These decide placement among the material that got through the gates. Depth on a narrow subject rather than coverage of a broad one. Specifications and numbers. Evidence attached to the claims it supports. Internal consistency, at odds ratios between 1.74 and 4.09. The absence of hedging. Comparisons included rather than avoided. And query-term match, the keyword gap, running between 5.99 and 40.0.

That last one gets left off most summaries of this paper, and it's the most actionable item on the list. The words the buyer uses have to appear in your material. Not synonyms of them. Not the more elegant phrasing your marketing committee preferred. Theirs.

Now the finding to tape to a wall.

The formatting effects varied across the models tested, with reported odds ratios ranging from 0.79 to 1.68. A range of estimates across models is not a confidence interval. It does not support a universal promise that changing headings or bullets improves citation.

In those experiments, the formatting changes tested did not produce a consistent citation advantage. That is a limit on the tactic, not proof that headings or FAQs never help a reader.

One nuance so I'm not overstating it, since the overstating is what I'm objecting to. The paper tested two structural factors. Content Structure is the null one. The other, Scattered Information, runs 1.13 to 3.87, and that's real. Spreading a single idea across four pages hurts you. Making the page pretty doesn't help you. Those are different claims and they get collapsed constantly by people quoting this paper to sell a redesign.

The Gate Almost Nobody in This Business Clears

Go back to the four gatekeepers and look at the second one.

Price or cost information present. Odds ratio above ten thousand. Fail it and you're out of the set before any of your thinking gets evaluated.

Now go look at your website.

Professional services firms fail this gate almost universally, and we've built an entire professional culture around the reasons. Every matter is different. Every engagement is scoped. Quoting a number invites comparison on price, which is the one axis where we lose. I've made that argument myself, to clients, persuasively, for years.

It was never as strong as it sounded, and now it has a cost you can measure.

What the gate wants is economic information a reader can orient against. A rate card would satisfy it. So would considerably less. A stated range. A minimum engagement. A structure, as in *we work on retainer, typically between these two numbers, with the diagnostic phase billed separately*. Any of that clears it.

Buyers want it independently of the machine. TrustRadius found pricing to be the single most cited frustration among software buyers evaluating vendors, which is a different market from yours with the same human being sitting in it.

You're withholding a number to avoid a conversation about price, and what happens instead is that the conversation never starts. Now it fails to start twice. Once with the buyer who bounced off a page that told them nothing, and once inside a system that filtered you out before the buyer saw the page at all.

The Machine Is Not Persuaded. It Is Corroborated.

If you take one sentence out of this paper, take that one.

A retrieval system does not experience being persuaded. It can compare sources, follow instructions and use tools to check claims, but it can also mistake repeated agreement for

independent support. Corroboration matters most when the sources are genuinely independent and the evidence can be traced.

Schuster, Gautam and Markert put numbers on this. Thirteen models, 7,440 pairs of conflicting sources, Kendall's W of 0.74, which means the models agreed with each other about which source to believe far more than they disagreed. Two findings inside that should worry you.

Attribution itself functions as a signal. Models preferred information that carried a citation, independent of whether the cited source was any good. The presence of the apparatus counted.

And repetition from low-credibility sources could override a strong source outright. Say a weak thing in enough places and it beats a good thing said once.

Neither of those findings is one I enjoy repeating. They describe a system that rewards volume and formal appearance over quality, which describes most of what's already wrong with marketing. They're still the conditions, and the strategic consequence is direct.

Consistency stops being a technical property and becomes an editorial one.

Your positioning statement, your practice descriptions, your bios, your proposal boilerplate, the way you describe what you do on a podcast, the sentence a partner uses at a conference. Every one of those is a source, and any two of them that disagree are a conflict the system has to resolve, using rules you don't control and can't inspect.

For most of marketing history, that inconsistency was invisible. No human being has ever read all of your firm's material. Not one. Not your marketing director, not the partner who wrote half of it, nobody.

A system can encounter conflicting versions without anyone inside the firm seeing the comparison.

That's chapter five, and it's the tax nobody knew they were paying.

Write Like the Paragraph Will Travel Alone

There's a practical instruction hiding in all of this, and it's the cheapest change in the book.

Retrieval often moves fragments rather than whole documents. A chunk of your material gets lifted out, stripped of the heading above it and the paragraph before it, and dropped into a context assembled from six other sources. Where the cuts fall isn't your decision. Whether the fragment still means anything after the cut is entirely your decision.

So write every paragraph as though it'll be read by someone who has never seen the one above it.

Here's what that means in practice, because an instruction without a demonstration is a preference.

Before:

This approach has proven effective across a range of engagements. It reduces the risk we discussed earlier and shortens the timeline considerably. Most of our clients see the difference within the first quarter.

Read it the way a retrieval system will. Alone. No heading above it, no paragraph before it, dropped between two fragments from other companies. *This approach*, which approach. *The risk we discussed earlier*, which risk, discussed where. *Most of our clients*, whose clients. Nothing in it survives the trip, and the model will either drop it or fill the gaps itself.

After:

Sequencing a succession plan against the operating calendar rather than the tax calendar reduces the risk of a forced sale during a working-capital squeeze, and typically shortens the transition by two to four months. In manufacturing engagements, owners see the first effect inside one quarter.

Same claim. Near enough the same length. Every noun carries its own referent, the claim names its own domain, and the fragment means the same thing standing alone as it does in place.

That's the entire technique, and it costs nothing except the habit.

Name the subject in the paragraph. No orphan pronouns doing work that depends on a previous sentence. If a paragraph opens with "this approach," you've written a sentence that becomes meaningless the moment it travels, and it will travel. Every claim carries its own number, its own date, its own definition, rather than borrowing them from three paragraphs earlier where the reader met them.

Xu, Iqbal and Montgomery went through 55,393 queries and 98,020 individual claims inside AI-generated answers, checking each claim against the source it was attributed to. Eleven percent of them didn't match.

Eleven percent of the statements carrying somebody's name on them, in a system that's currently deciding whether your firm belongs on a list of three.

Some of that is the model. Some of it is us. We write paragraphs that only survive in the company of their neighbors, and then we're surprised when one gets quoted by itself and comes back meaning something adjacent to what we said.

Thirty years building a reputation, and now something summarizes it in four sentences. It gets one of them wrong.

The model is out of your hands. How hard those four sentences are to get wrong is entirely in them, and the price is the discipline to write every paragraph as though it'll be read by itself.

The Boring Part, Which Is the Part That Works

Everything above is editorial. This part is plumbing, it's cheap, and it has the least ambiguous evidence in the paper.

Crawlers can't cite what they can't reach. When Vercel measured crawler behavior, ChatGPT's crawler hit 404s on 34.8 percent of its requests and Anthropic's on 34.2, against Googlebot's 8.2. That data's from December 2024 and it is dated in the fastest-moving corner of this field, so treat the numbers as directional. The direction is that a third of what those crawlers reached for wasn't there.

A related myth needs killing, since I've now heard it from three consultants. "No AI crawler executes JavaScript" is wrong, and it's wrong according to the same Vercel post everybody's citing for it. OpenAI's, Anthropic's, Meta's, ByteDance's, and Perplexity's crawlers don't render JavaScript. Google's Gemini does, through Googlebot's infrastructure, and so does AppleBot. Same staleness caveat applies. The useful version of the advice is that anything you need read should exist in the HTML, which was good practice before any of this and will be good practice after.

A separate result needs its product boundary kept intact. Grossman and colleagues, also at SIGIR 2026, checked twenty-one major publishers who blocked Google-Extended in their robots files. Gemini cited them zero times. Not rarely. Zero.

That finding concerns the Gemini setting tested. Google-Extended is a separate control from Google Search, so it should not be read as a rule that blocking Google-Extended removes a page from AI Overviews or AI Mode. Check the control for the product you actually intend to govern.

Somebody in your firm may have blocked AI crawlers eighteen months ago, on principle, during a partner meeting where the principle sounded right. It's worth finding out. That decision is still running, and it's the only setting in this paper that's fully deterministic.

The Part I'd Rather Not Have Noticed

Back to Dr. Fox, because there's a version of this argument I'd rather not publish and would rather not have somebody else publish first.

If a system can evaluate the shape of rigor and can't evaluate rigor, then the shape is purchasable.

Yerin Hwang and colleagues tested it directly. They took identical answers to math problems, some of them wrong, and wrapped them in seven classical persuasion techniques. Then they had language models grade the results. Across six benchmarks, the persuasive framing pushed judges into inflating scores on incorrect solutions, by as much as eight percent.

Wrong answers, dressed better, scored higher.

And I owe you a confession about how I know this. An earlier draft of this paper cited a different paper for the same finding, attributed to a researcher named Puerto, along with a widely circulated

multiple of 7.7x. A reviewer went looking for it. The paper doesn't exist. The number appears in nothing. Worse, there is a real researcher named Puerto with a real paper covering six domains, and it concludes the opposite of what I had him saying.

I don't know how it got in. I know it survived my own verification pass, which caught eleven other things, and it sat in the paper about evidentiary standards for a week.

The uncomfortable corollary stands anyway. You could hire someone to generate material with all the markers of rigor and none of the work behind it. Specific numbers, named sources, confident register, no hedging, comparisons included. It'd score. For a while, it'd score well.

I want to give you a reason not to that isn't a lecture about integrity, because the people who need the lecture won't read it and the people reading it don't need it.

So take the practical one. Every gap between what a system rewards and what it should reward gets closed. Directory spam, link farms, keyword stuffing, content mills, review manipulation. I've watched five of these cycles now, and the pattern doesn't vary: the gap is lucrative, then it's crowded, then it's detected, then it's penalized, and the penalty lands hardest on whoever was furthest out on the limb. Disclosure rules are already being drafted. The detection is already being built by the same companies whose systems are being gamed.

Making genuine rigor legible and faking its appearance produce the same output today. In three years they won't, and only one of them still exists on your website.

Which is the answer to a question professional services firms have been asking me for two years, usually in a tone of some despair. If a machine can generate the appearance of expertise, what's left for us?

I'd add that the question has teeth, and I know because it got pointed at me. The firm in chapter two that thought my work looked machine-made was performing this exact test, badly, from the outside, with no way to check. That's going to happen to you too, and the only defense is material that could only have come from somebody who was in the room.

What's left is the part the machine is imitating.

Findable Was the Easy Part

Two things are true at once here and the paper fails if you leave with only one.

The mechanism is real. Buyers are using these systems, the systems are filtering, the filters respond to properties you control, and the firms that get interpretable early get a structural advantage over the ones that get there in 2029.

And the volume is still small. Conductor measured AI referral traffic across 13,770 domains and found it averaging 1.08 percent, with an eleven-fold spread underneath the average, from IT at 2.80 percent down to communication services at 0.25. Professional services sits near the bottom of every version of this measurement I've seen.

So anybody telling you that AI search is currently your channel is selling something, and I'd want to know what.

The right way to hold this is that you're buying insurance while it's cheap. Making your firm interpretable costs you a handful of decisions about how you write and one afternoon with whoever owns the website. It pays nothing this quarter. It prevents a specific outcome later, which is a system that can't work out what you are, in a market where that system has become the front door.

Getting invited into the room has always been the hard part. The invitations aren't coming only from humans anymore.

The thing deciding whether to send you one can see the shape of your thinking, and never the thinking itself. Everything in this paper is about that gap. The next one is about a version of the gap you created yourself, in twelve different documents, over eleven years, without noticing.

Related reading

Decision Science: AI Readiness AI Search and Organizational Capability.

How-To: How to build a piece around an insight.

Process: Marketing Assets and Decision Support.

Next chapter.