Citations don't predict consideration. Or recommendation.

The evidence · AIVO Meridian · September 2026

The finding

The AI-visibility industry sells citation share, mention frequency and share of voice. We measured what those things actually produce inside a buying conversation, and the answer is: not consideration, and not the recommendation.

In our latest study — **2,160 multi-turn buying conversations, 8,640 conversation turns, three AI assistants, four buying situations, every conversation run three times** — the correlation between a provider's citation share and how often it was brought into the conversation was **+0.36** (p = 0.19). Statistically indistinguishable from noise.

The most-cited provider in the entire study — **1,417 citations, 9.4% of everything the assistants quoted** — was discussed in 42% of conversations. A provider cited **86** times, one-sixteenth as often, was discussed in 62%.

This is not a single-study result. It is what seventeen months of primary research keeps producing.

What sits behind it

The corpus	
Continuous primary research	**17 months**
AI buying journeys analysed	**22,000+**
Enterprise brands tested	**380+**
AI assistants measured and remediated	**5**
Sectors measured longitudinally	**14**
Published papers, frameworks and methodologies	**95**, all openly interrogable

Three figures from that corpus have held from the beginning:

- **95.7%** of brands are recognised by the model at the opening of a buying conversation. Possession is not the problem.
- **87.3%** of brands named at the opening are displaced by a competitor by the decision (n = 1,427 probes, ten industries).
- **75.7%** of the facts a model itself says it holds about a brand go unused when it makes the recommendation (n = 592 facts).

The model knows. It doesn't choose. Everything below is the mechanism.

What we ran

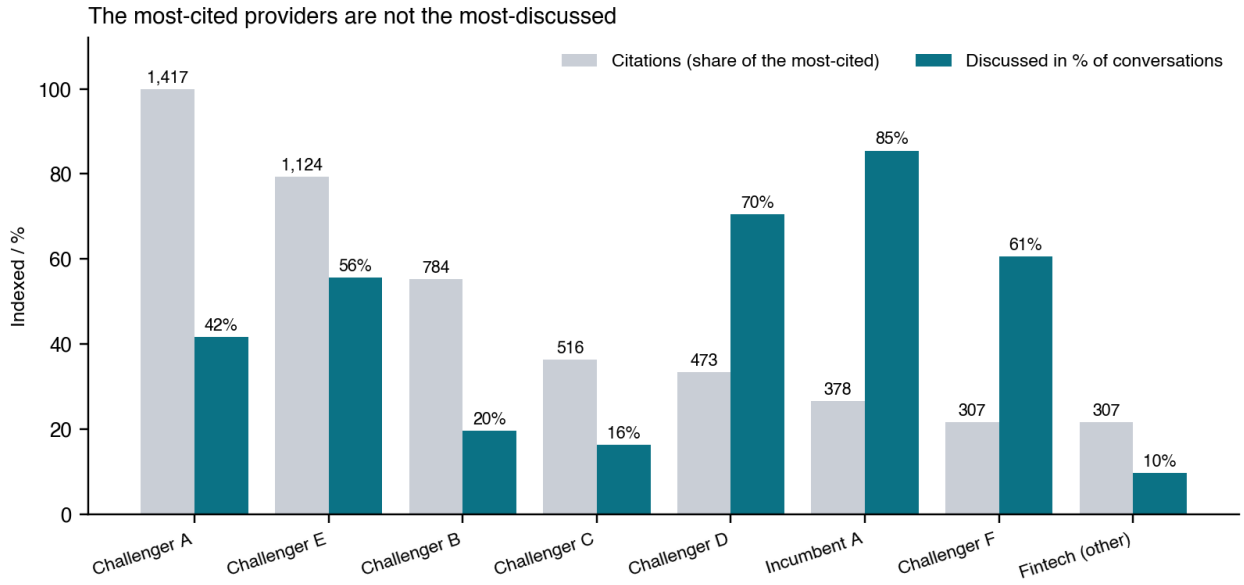
2,160 four-turn conversations, September 2026: **30 established brands in a single category**, four distinct buying situations, three AI assistants, two conditions, three replicate runs of every cell.

Element	Design
The conversation	Four turns: the customer describes their situation · asks what the leading options are · states the criteria that matter · asks for a final recommendation and how to proceed
Two conditions	**Brand named** by the customer at turns 1–2 only (retention), and **no brand named** at all (acquisition)
Assistants	Two answering from what they already know, one searching the web live. Reported separately, never pooled
Replicates	Every cell run three times — AI answers vary, and a single answer is an anecdote

Element	Design
Scoring	Every final answer classified to fixed rules, then checked: **150 answers coded blind by a human**, agreement on the winner **99.2%**
Prompts	Frozen and versioned before the run; no location, no provider and no category of provider named by us

1 · Citation share buys nothing

Across the searching assistant's conversations, **58,631 sources were retrieved and 30,424 were actually quoted** in the answers. We counted each provider's own pages among those quotations, and separately how often that provider was brought into the conversation at all.



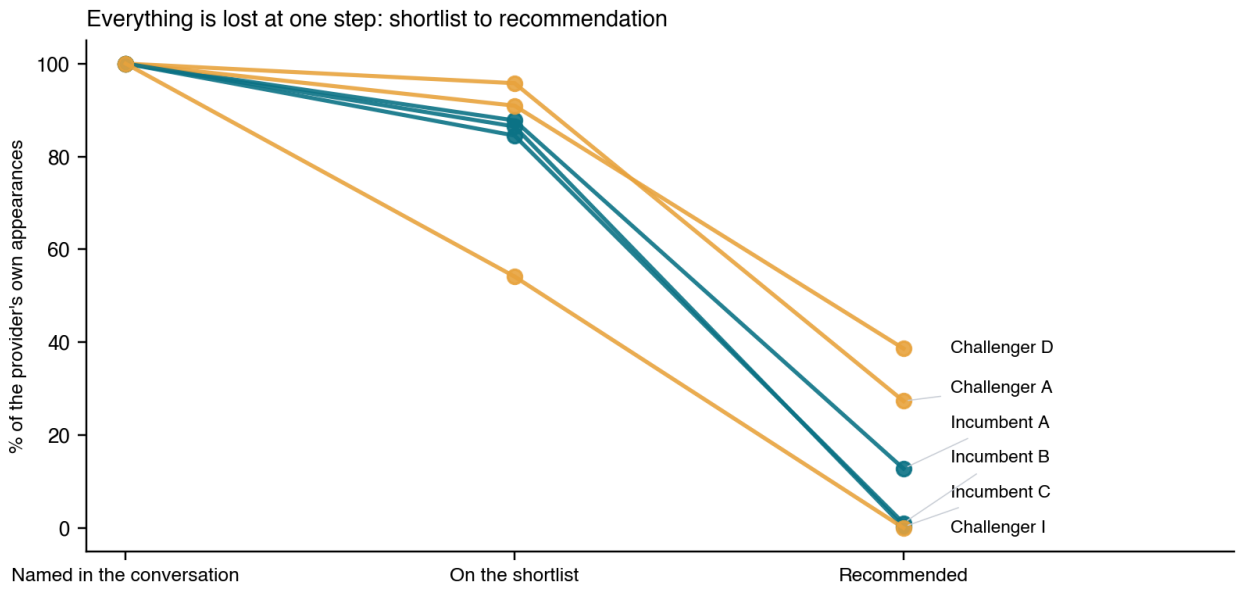
Provider	Citations	Discussed in	Recommended in
Challenger A	1,417	42%	11.4%
Challenger E	1,124	56%	16.1%
Challenger B	784	20%	2.3%
Challenger C	516	16%	1.1%
Challenger D	473	70%	27.2%
Incumbent A	378	85%	10.9%

Citation share is a measure of publishing activity. It is not a measure of commercial outcome.

2 · Being discussed is not being chosen

Incumbent B appears in **62%** of conversations where the customer named nobody. It reaches the shortlist in **85%** of those. It is the final recommendation in **1.2%** of them.

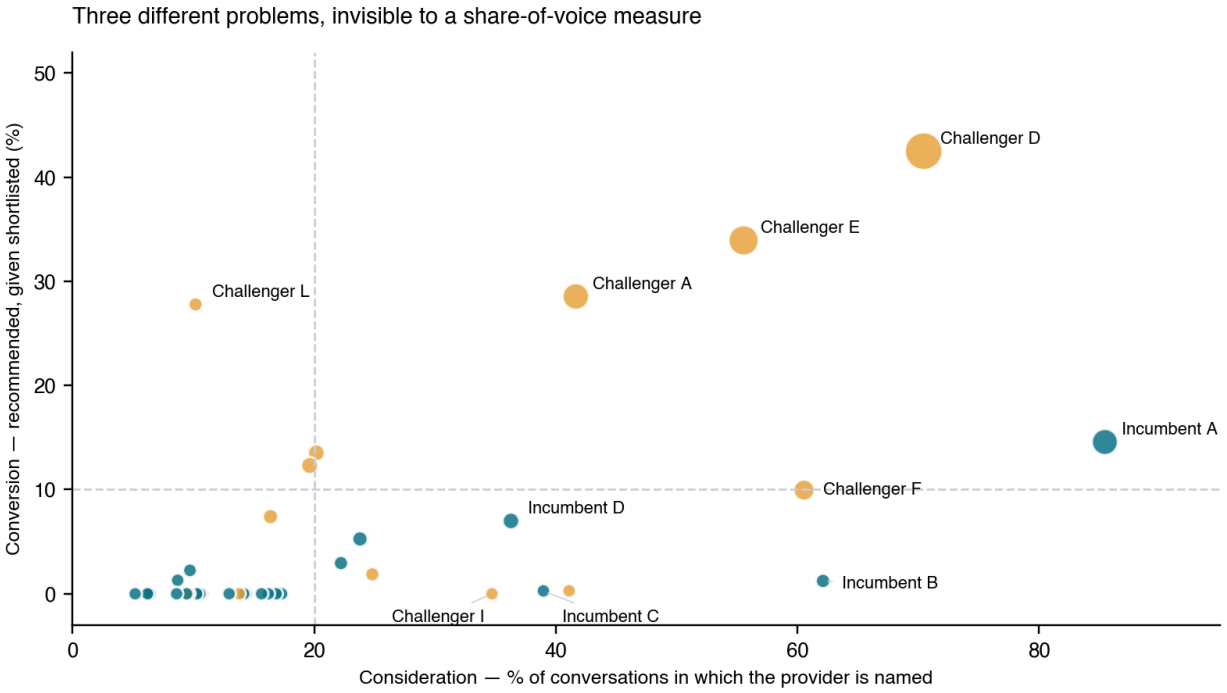
Every visibility dashboard on the market would score that brand healthy. Share of voice: strong. Mention frequency: strong. Outcome: nothing.



Across all providers measured, the step from being named to reaching the shortlist is close to automatic — a median of **87.9%**. The step from shortlist to recommendation is where everything is lost: a median of **6.1%** among providers considered in at least a fifth of conversations, and **28 of the 46 providers measured never convert at all**.

There is no recovery further down. Asked how to proceed, the assistant sends the customer to the site of the provider it just recommended in **100%** of cases. The recommendation *is* the hand-off.

3 · Three brands, three different problems



Pattern	What it looks like	Real example from the study
In the conversation, never chosen	High consideration, near-zero conversion	Incumbent B: 62% discussed, 1.2% converted
In the conversation and chosen	High on both	Challenger D: 70% discussed, 42.5% converted

Pattern	What it looks like	Real example from the study
Rarely named, converts when it is	Low consideration, strong conversion	Incumbent E: named on its own in 2% of conversations, but holds the recommendation 36.1% of the time when the customer raises it

These need opposite work. One brand has to earn the decision; the other has to get into the room. A measure that stops at "are we mentioned" cannot tell them apart — and most brands have never been told which one they are.

4 · What actually predicts surviving the decision

Using only the turns *before* the decision is asked for, we measured how each provider is described:

What the assistant's sentences contain	Correlation with conversion
A concrete figure — price, rate, threshold	**+0.61**
A specific named product	**+0.52**
Hedging: "varies", "depends", "check with them"	**-0.46**

Providers described in decision-grade terms convert. Providers described vaguely do not. This is the linkage gap in action: the facts exist, the model holds them, and they are not in a form it can use to justify choosing you.

It is not the whole story. In this study the most-discussed loser was described **more** specifically than the winner and still lost — because the assistant judged it a poor fit for that customer's stated need. Legibility gets you to the decision. Fit wins it.

5 · How we know the scoring is sound

Every final answer was classified to fixed rules. A random sample of **150** answers — 50 per assistant, the assistant's identity hidden and the automated verdicts withheld — was then coded independently by a human reading the full answer.

- **Winner identity: 99.2% agreement** (132 of 133).
- **Type of outcome** (one provider, shared, none named): **88.7%**, and the differences are almost all two-provider answers, where the automated scoring is the *more* conservative of the two.
- Every figure in this report carries its denominator, and the prompts, the opportunity set and the tables behind it can be supplied on request.

What this means for measurement

Publishing more content and winning more citations does not put a brand into the AI's consideration set, and it does not win the recommendation. The decision is made at one step, on whether the assistant can justify choosing you for this particular customer.

Three questions worth putting to whoever reports your AI visibility today:

1. What is our **win rate** — how often are we the recommendation, not the mention?
2. **Where** do we lose: before the conversation, or at the decision inside it?
3. When we fix something, can you show me **the same measure moving**?

If the answer to any of those is a share-of-voice chart, the thing that decides the outcome is not being measured.

AIVO Meridian measures what AI recommends, diagnoses why, fixes it, and re-measures to prove the movement. The study described here is anonymised by design: the finding is about the mechanism, not about one brand or one sector. The underlying method is published with DOIs; the tables behind every figure are available on request.