

---

## Why AI Visibility Does Not Guarantee AI Recommendation

Date: January 28, 2026

Published in AIVO Journal (<https://www.aivojournal.org/why-ai-visibility-does-not-guarantee-ai-recommendation/>)

**Zenodo: 10.5281/zenodo.18401395**

Over the past two years, consumer brands have invested heavily in improving their visibility inside conversational AI systems. The prevailing assumption has been straightforward: if a brand appears clearly and positively in AI-generated answers, it benefits.

That assumption is incomplete.

In multi-turn testing of consumer-facing AI systems, we observe a recurring pattern in which brands remain visible and well described during early stages of a conversation yet are removed at the point where the system is asked to recommend what to buy. This shift occurs without the introduction of new negative information and without any explicit signal that substitution has taken place.

This article examines that pattern, why existing optimization frameworks do not capture it, and why it raises a distinct measurement and governance question for consumer brands, particularly in beauty and personal care.

---

### **From explanation to recommendation: a behavioral shift**

Most AI evaluations focus on single responses. Consumers, however, do not use conversational AI in a single turn when making purchase decisions.

A typical interaction unfolds across several steps:

1. Is this product suitable?

2. What alternatives should I compare?
3. Given my constraints, what is the best option?
4. What should I buy?

During the first two stages, AI systems tend to provide broad explanations and comparisons. Options are enumerated, trade-offs are discussed, and multiple brands can coexist without conflict.

At the final stage, the system's behavior changes. When asked to recommend a specific choice, it narrows the option set sharply and often replaces earlier candidates with alternatives that were not emphasized previously.

We refer to this observed pattern descriptively as **decision-stage substitution**.

This is not a claim about internal architecture or intent. It is a description of repeatable output behavior across multi-turn interactions.

---

### **A concrete example from multi-turn testing**

To make the phenomenon tangible, consider the following sanitized pattern observed during controlled testing.

Across 60 identical multi-turn consumer journeys in a skincare scenario, using the same prompts and constraints:

- A specific brand appeared in approximately 80 percent of initial comparison responses.
- The same brand appeared in fewer than 30 percent of final purchase recommendations.
- In most cases, it was replaced by one of two alternative brands.
- No new negative information about the original brand was introduced between comparison and recommendation.
- The substitution occurred at the same conversational step across repeated runs.

Across multiple brands and product variants within the same category, the magnitude of this drop varied, but the direction was consistent: inclusion rates at the recommendation stage were materially lower than inclusion rates during earlier comparison stages.

Single-turn analysis would have concluded that the brand was performing well.

---

### **Methodology at a high level**

The testing underlying these observations follows a simple but controlled structure:

- Identical multi-turn conversational journeys
- Fixed prompts and constraints across runs
- Repetition to observe consistency of outcomes
- Execution across multiple leading consumer-facing AI systems

This pattern has been observed across at least three major conversational AI platforms, with variation in frequency but not in direction. These systems change frequently, and results vary over time, but the behavioral shift described here has proven reproducible.

This article does not claim exhaustiveness, universality, or permanence. The claim is narrower: the pattern exists, it is observable, and it is not captured by existing visibility-focused measurement approaches.

---

### **Why beauty and personal care are particularly exposed**

Beauty and personal care categories exhibit this behavior more clearly than many others, for structural reasons.

First, decision complexity. Product suitability depends on multiple interacting constraints such as skin type, sensitivity, age, routine

compatibility, and budget. These constraints trigger narrowing behavior earlier and more forcefully.

Second, high functional adequacy. Many products meet baseline requirements. From a system perspective, multiple options appear acceptable, which accelerates convergence toward defaults.

Third, perceived downside risk. The cost of a poor recommendation in skincare is higher than in many other CPG categories. AI systems appear to favor options that are broadly compatible and widely trusted when asked to recommend a final choice.

This does not imply that substitution is incorrect. It implies that it follows a different logic than explanation or comparison.

---

### **Why GEO and AEO do not capture this**

Answer Engine Optimization and Generative Engine Optimization are designed to measure whether and how brands appear in AI outputs.

They answer questions such as:

- Is the brand mentioned?
- How is it described?
- Is it included in comparisons?

Decision-stage substitution occurs after those questions have already been answered positively.

A brand can be visible, well described, and favorably compared, and still be removed when the system is asked to recommend a single option. Because this happens downstream of visibility, single-turn audits and prompt-level optimization do not detect it.

This is not a failure of GEO or AEO. It is a boundary of what they measure.

---

## Is this behavior necessarily a problem?

An important objection deserves consideration.

If an AI system narrows from many suitable options to a smaller set of broadly compatible, evidence-backed choices when asked what to buy, that may be appropriate risk mitigation rather than a flaw.

The concern raised here is not that substitution occurs.

The concern is that substitution occurs **without visibility**.

From the outside, there is no way to observe when narrowing happens, what criteria dominate the decision stage, or why one option is removed while another is retained.

For brands and organizations attempting to understand how AI systems shape consumer decisions, this opacity matters.

---

## Why this differs from historical recommendation channels

Most consumer decision processes have never been fully reconstructable. That is not new.

What is new is the role conversational AI plays as a synthesized advisor. These systems collapse information retrieval, comparison, and recommendation into a single interface that is perceived as neutral and comprehensive.

When such systems influence decisions at scale, the absence of traceability around how recommendations are formed becomes a distinct measurement and governance consideration.

This is not about blaming AI systems. It is about understanding how a new intermediary operates.

---

### **Portfolio-level implications**

For organizations managing multiple brands, decision-stage substitution has two implications.

First, external leakage. Demand may be routed to competing brands outside the portfolio without awareness.

Second, internal reallocation. Demand may be redirected between brands within the same group in ways that are invisible to portfolio managers.

For example, in a multi-brand portfolio, a product that is well optimized for AI visibility may still lose final recommendations to a sibling brand that has not been explicitly optimized, simply because the system converges on what it infers to be the most generally acceptable option at the decision stage. Without observability, this internal reallocation is indistinguishable from external leakage.

In both cases, the issue is not optimization failure but lack of visibility.

---

### **Materiality and time horizon**

It is reasonable to ask how material this is today.

Conversational AI currently mediates a minority of consumer purchase decisions. The exact proportion varies by category and geography and is difficult to quantify precisely.

The relevance of decision-stage substitution lies less in current magnitude and more in trajectory. As conversational AI becomes a default pre-purchase

filter, **persistence at the recommendation stage** will matter increasingly alongside visibility.

Understanding how recommendations are formed becomes more important before reliance becomes widespread, not after.

---

### **A reframing rather than a verdict**

The core observation of this article is narrow.

AI systems exhibit different behavior when explaining options versus recommending a choice. That behavioral shift can lead to brand substitution that is not captured by existing visibility-focused measurement frameworks.

This does not imply wrongdoing, bias, or failure. It implies misclassification.

The question for brands is not how to force outcomes, but how to gain visibility into a decision layer that is already influencing consumer choice.

---

### **Closing thought**

Brands have become adept at optimizing for being mentioned by AI systems. That is necessary, but it may not be sufficient.

As conversational AI increasingly participates in consumer decision-making, the more difficult question is not whether a brand is visible, but whether it remains included when the system moves from explanation to recommendation.

That transition is where measurement currently ends, and where understanding may need to begin.

---

*This article is part of an ongoing examination of decision-stage behavior in AI-mediated consumer markets. Future work will expand on methodology, cross-category comparisons, and implications for enterprise measurement and governance frameworks.*

---