Demand-Genius

Dark AI

What actually drives AEO influence?

Executive Summary

What actually drives AEO influence?

Most AEO advice centres on measuring and optimizing basic visibility metrics, like brand citations or mentions. But this approach misunderstands how LLMs actually behave through complex purchasing decisions.

Citations and referrals occur almost exclusively in conversion-stage queries where the task directly necessitates a recommendation. We know "conversion" to be only a fraction of a complex, B2B buyer journey - or, the tip of the iceberg.

By that point, influence has already been decided.

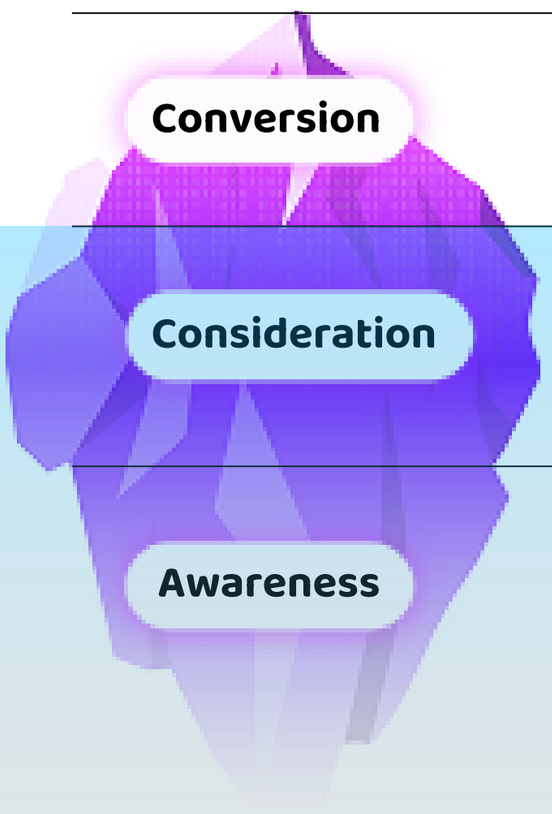
84%

AI prompts produce no brand citations

The vast majority of buyer education happens earlier, in what we call **Dark AI: the conversations beneath the surface where problems are framed, requirements are built, and category narratives are established.** These conversations don't show up in consistent brand mentions or in referred traffic, yet they determine which brands end up in consideration. It's time to ditch the funnel, and instead think of AI influence as an iceberg.

By the time a buyer enters "conversion-stage" promps where citations do occur, variability has dropped to near-zero. The model has already decided which brands are viable, and is very hard to influence. Instead, brands need to learn to influence the conversations beneath the surface, where problems are framed and requirements built.

Why most AEO strategies are just the tip of the iceberg



	Frequency			Brand Count		Variability	Language
	How often are brands invoked?			How many brands feature in responses?		How consistent are brand citations between runs?	What kind of language is used to frame responses?
Conversion	Mention **94%**	Retrieval **48%**	Citation **48%**	Mean **6.39**	Median **6.33**	**100%** Low (no) variability / tight convergence	Directive, selective, decision-oriented language
Consideration	Mention **68%**	Retrieval **0%**	Citation **0%**	Mean **2.75**	Median **1.17**	**33%** Moderately high variability / partial convergence	Comparative, evaluative, trade-off-based language
Awareness	Mention **60%**	Retrieval **0%**	Citation **0%**	Mean **1.59**	Median **0.67**	**22%** High variability / low convergence	Descriptive, category-level, exploratory language

Note: This is the median % of runs in which the top 3 brands re-appear.

Below the metaphorical waterline are the TOFU and MOFU conversations where buyers explore problems and build decision criteria. Current AEO measurement ignores them entirely in favour of “above the line” metrics, but in actual fact, these “Dark AI” conversations are clearly where influence forms and categories are won and lost.

Our research analysed hundreds of prompt clusters across B2B categories to understand when and why LLMs surface brands. Three findings challenge current practice:

Key Takeaways:

1

Citation-chasing is misleading (and dangerous).

Frequent citations is a consequence of constrained options and pre-existing credibility, **not an effective AEO strategy**. Most brands track prompts derived from their own content - validating visibility in conversations they're already winning while missing broader influence opportunities.

Chasing citations by optimizing content for specific prompts, as if they were keywords, creates complex, sprawling content libraries. The contradictions. that can arise reduce clarity over who you are, who you’re for, and how to position you.

2

Influence happens during problem-framing, not recommendation.

LLMs rarely introduce new brands at BOFU - they filter from earlier exploratory phases.

Your goal should be to influence problem framing in a way that favours your strengths, downplays your weaknesses, and makes visibility inevitable.

Comparative language and framing that is specific, clear and consistent is more effective than broad statements of dominance.

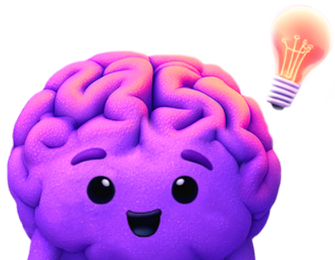
3

Measurement must move upstream.

Track **inclusion before convergence** (are you present in exploratory prompts, before the LLM has narrowed options?), **criteria alignment** (does the model associate your brand with attributes that matter to each of your buyers?), and **information gain** (does your content contribute new understanding that shapes problem-framing?).

Want to learn more?

Read on, and you can use the Benchmark AI on the website to interact directly with the data. It’s trained on everything we studied, so ask it questions to help you understand and act on what we’ve learnt!

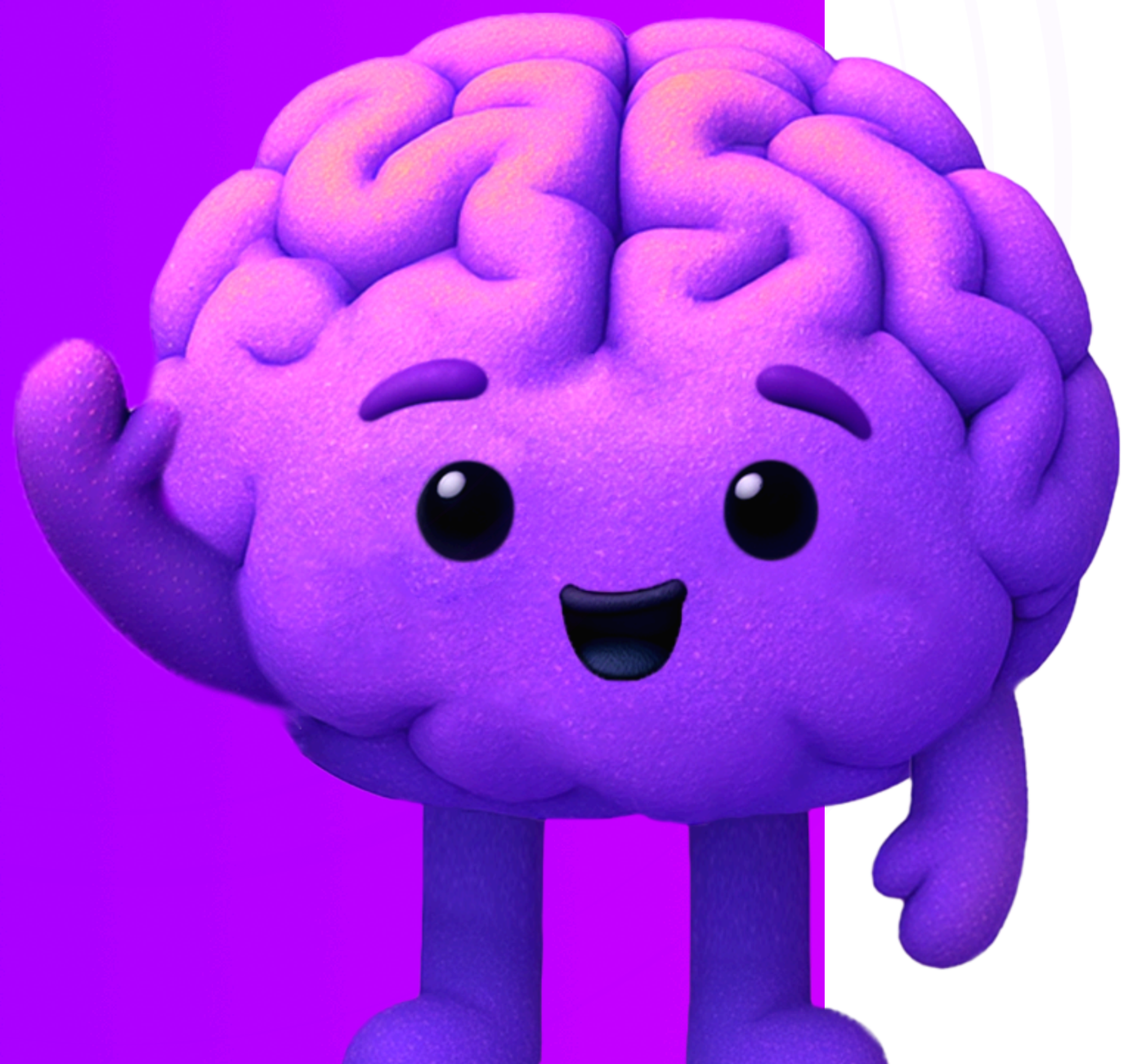


Introduction

Foreword

Understanding LLMs

How LLMs Surface Brands



Foreword

In the space of about 12-24 months, AEO has risen to the number one topic of conversation I encounter at every Marketing conference, event or meetup. This seems to be driven by a fascination with the scale of opportunity, but also by a deep lack of understanding as to how exactly brands can influence when and how AI references their brand and recommends their products.

Marketing leaders are caught between an urgent, board-level pressure to “show up in AI” and a real lack of nuanced or evidence-backed insight into how to achieve that.

There are fundamental differences between Search and AI that we need to recognise. Search was a navigational tool. It organised knowledge and helped users navigate it. AI does far more. It reasons, compares and forms opinions on the user’s behalf - as a trusted advisor.

We felt, quite deeply, that existing advice and practices miss the mark. Put simply, it seemed to mimic the SEO playbook far too neatly: track brand mentions, count citations, optimize content to "rank" in AI responses. It smelled of groupthink, and of the market viewing a new and unique challenge through blinkers imposed by existing ways of thinking and measuring performance. In fact, not even the more nuanced branches of SEO strategy; tactics centre on publishing large volumes of content with the sole goal of gaming citations. AI-generated FAQs, schema markup manipulation, listicles and superficial comparison pages. **We are in the age of black-hat AEO.**

This report presents findings from a multi-vertical study analysing hundreds of prompt clusters across every stage of a typical complex (B2B) buying journey. We examined when and why LLMs surface brands, how consistently they do so, and how their behaviour changes as prompts move from problem exploration to decision-making. The core finding challenges the foundation of most current AEO practice: **AI visibility and AI influence are not the same thing.**

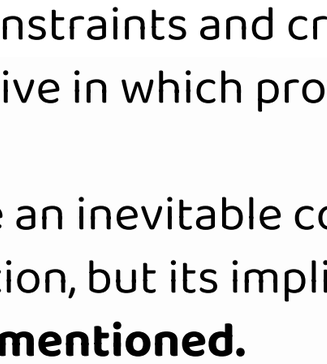
Visibility is, in fact, a consequence of influence rather than a measure of it. LLMs arrive at a recommendation not through evaluation at the point of decision, but a process we call convergence. They progressively add constraints and criteria to hone in on a few viable options, at which point they are extremely conservative in which providers they surface.

Visibility is therefore an inevitable consequence of whether a brand is deemed a viable option. This might seem a small distinction, but its implications are quite big - **decisions are formed long before any brand is cited or often even mentioned.**

The AEO battleground is therefore far earlier in the buyer journey, in what we call **“Dark AI”** - the awareness and consideration-stage prompts that shape category narratives, which criteria are applied, and therefore in which direction the option pool converges.

A brand’s goal should not be to be “visible”, it should be to influence problem framing in a way that favours their strengths, downplays their weaknesses, and makes visibility inevitable.

by Tom Rudnai
Founder & CEO
Demand-Genius



84%

AI promps produce no brand citations

0%

of Awareness & Consideration prompts produce citations

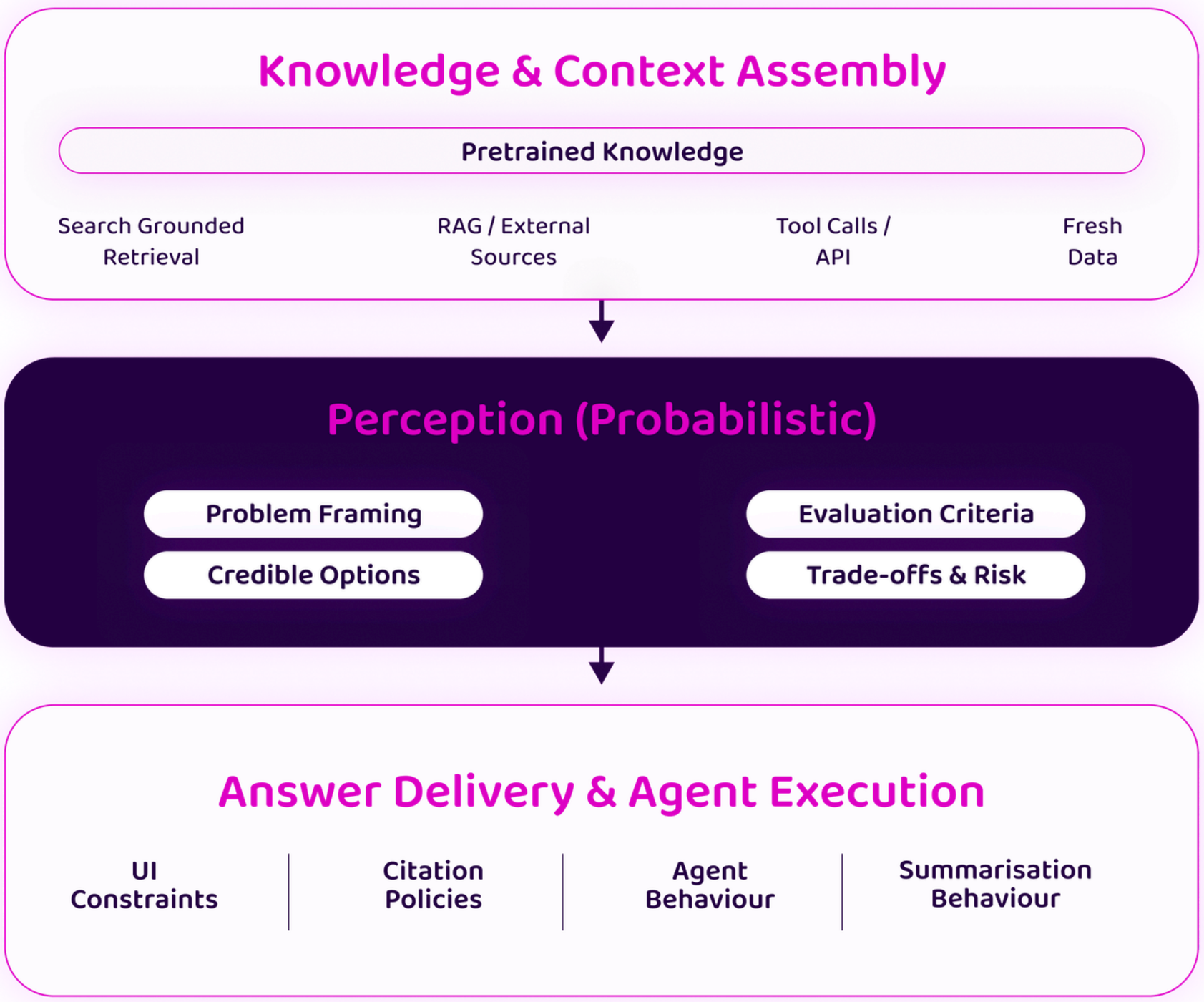
100%

of Conversion prompts surface the same top 3 brands every time.

16%

of AI responses invoked retrieval, vs. internal knowledge

Understanding how LLMs work



LLMs call on a hierarchy or sequence of sources to gather information relevant to ask completion.

They form a perception of the problem space - and the brands within it.

They compile an answer, seeking to mirror the user's intent (e.g. exploratory vs. decision-oriented).

The great challenge of AEO is that we must attempt to understand and shape the way model perception forms, from the answer and limited insight into the knowledge assembly process that informed it.

Different Forms of Visibility and What They Mean

One of the challenges of AEO is that we are always attempting to deconstruct LLM influence and perception based on limited information. LLMs are complex, layered models that do an awful lot very quickly in order to produce an answer to a given query. Any attempt at understanding that process and what influences it must work back from limited signals in the Answer Delivery layer.

It's useful to make sure that we're clear over what the analytical tools at our disposal, and what each can tell us. LLMs surface brands in three distinct ways, each signalling something different.

How do LLMs surface brands in answers?

Layer	What it is	When it typically happens	What it signals	What triggers it	What it does not mean	AEO implications
Brand Mention	Brand name included in the response without sourcing	Progressively more often as the buyer journey evolves	Category awareness and semantic inclusion	General knowledge, familiarity, category mapping	Authority, endorsement, accuracy, or justification	Necessary baseline for AEO, but low barrier and weak signal. Mentions alone rarely translate into influence or selection.
Retrieval	External source is accessed to ground or expand the answer	BOFU queries, when specificity or decision risk increases	Informational need and credibility checking	Constraints, technical detail, uncertainty, higher accuracy requirement	That any brand will be cited or recommended	The model needs structured, accessible content. Retrieval readiness is a stronger signal than mentions.
Citation / Link	Source explicitly shown to the user	Exclusively BOFU terms with decisional intent	Justification and transparency	Decision weight, trust requirement, user-facing explanation	Preference, ranking, or likelihood of selection	Rarest and most selective outcome. Citations are a presentation choice, not a guaranteed reward for optimization.

A Note on Acronyms

What a ridiculous heading that is.

It does seem worth clarifying, though, given the difficulty the industry has in agreeing what to call AI search optimization. We do not draw any particular distinction here between Answer Engine Optimization (AEO) and Generative Engine Optimization (GEO). We have adopted AEO because the name appears to encourage better behaviours - optimizing for a zero-click world in which users access answers directly. And honestly, because at a point, we just have to pick one.

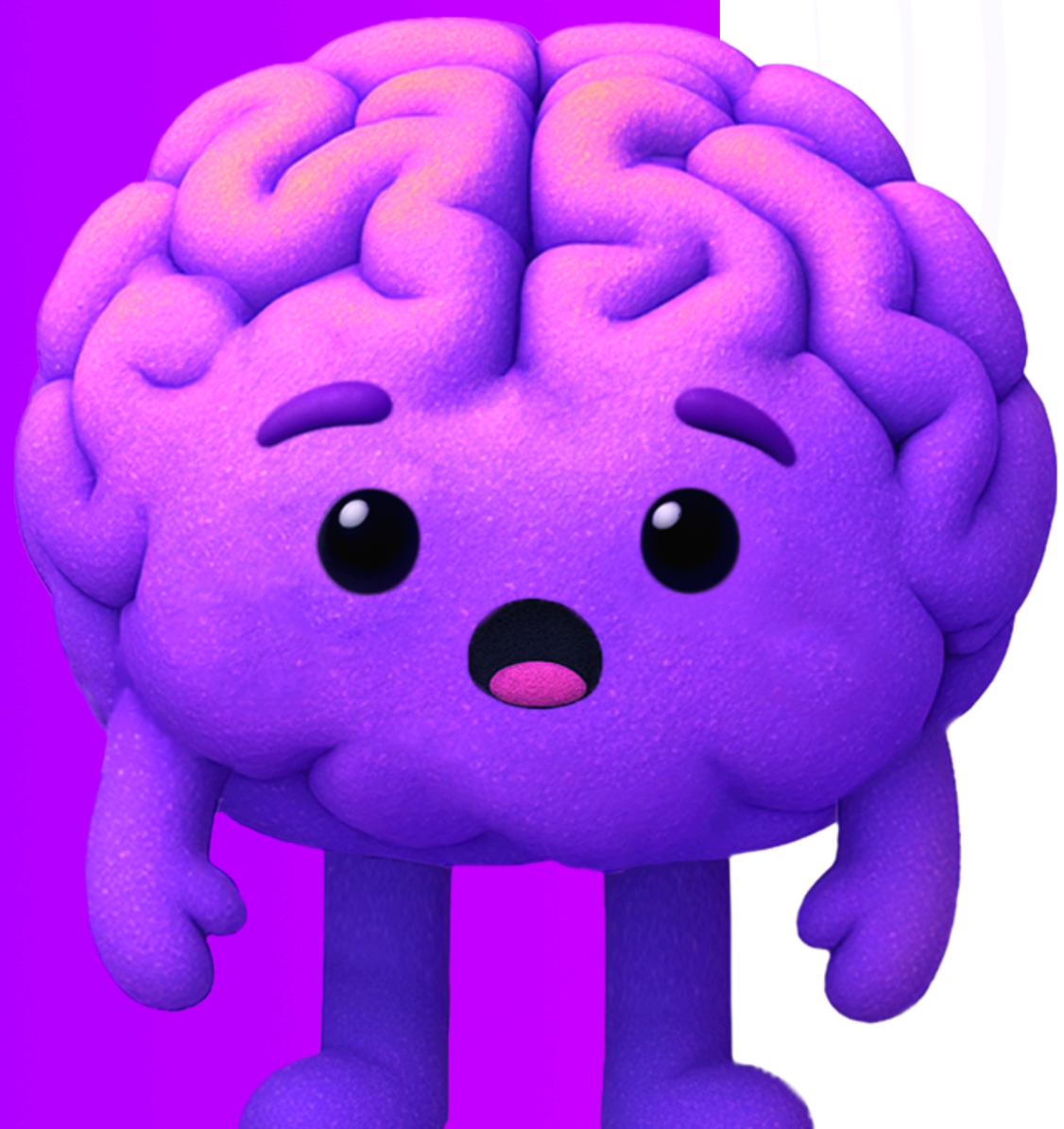
Dark AI

What actually drives LLM influence?

How do AI responses vary across buyer journeys?

What is Dark AI?

Why it’s the key to winning complex categories.



What Actually Drives AEO Influence?

AI visibility behaves very differently by funnel stage

The biggest takeaway from this study is observing a stark difference in how LLMs behave at each stage of the funnel, both in approach and methodology.

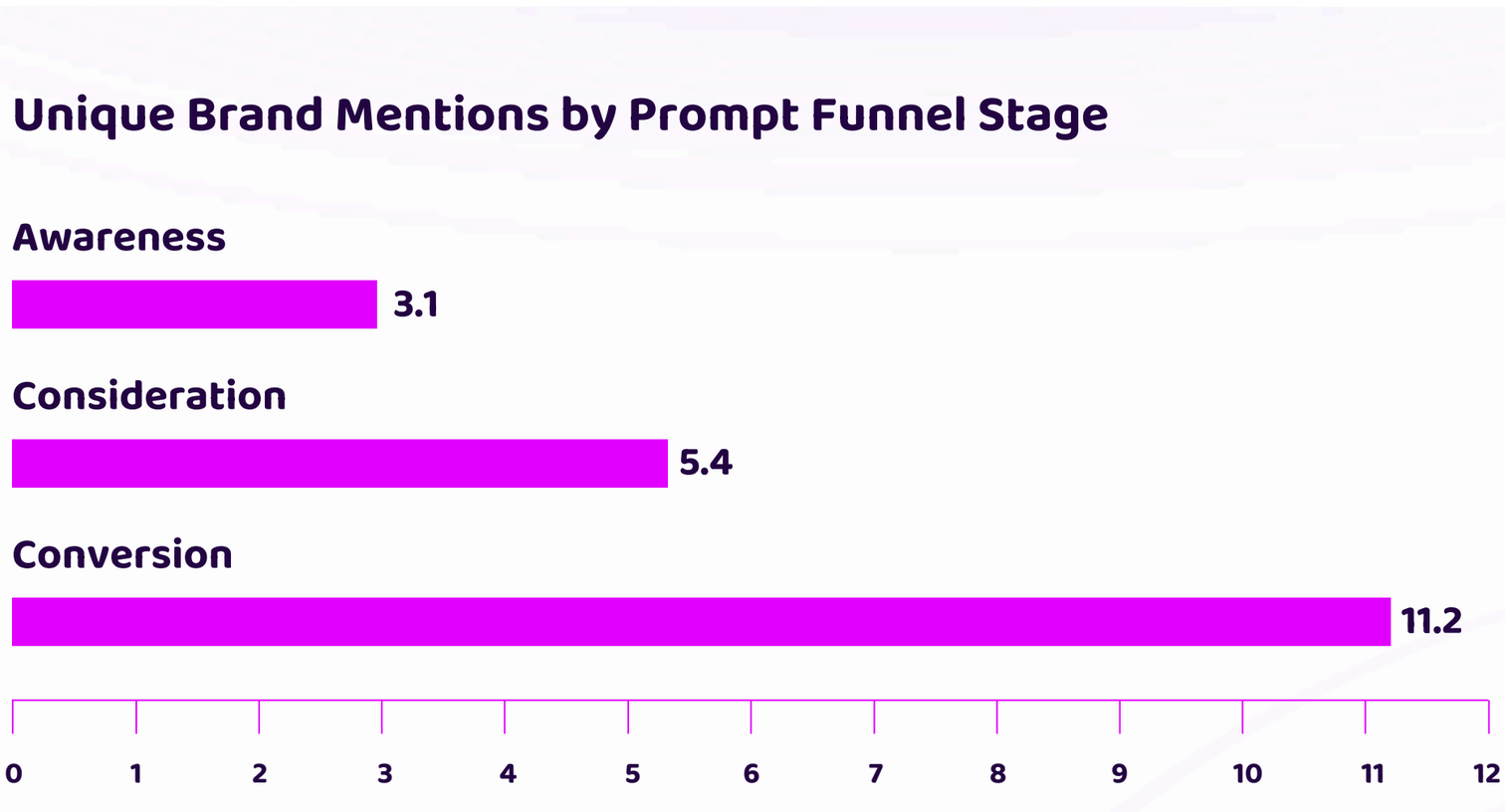
LLM systems are designed to respond in ways that balance task alignment, response certainty, and system efficiency. Behaviours such as retrieval and reasoning are selectively invoked based on the task, rather than being default behaviours.

How often to LLMs invoke brands at easch stage of the journey?

	Mention	Retrieval	Citation
Awareness	60%	0%	0%
Consideration	68%	0%	0%
Conversion	94%	48%	48%

At the conversion stage, where task completion requires selection, recommendation, or decision-making, responses invoke external retrieval and surface explicit citations. Brand mention counts are higher and more consistent. Under decisional intent, the model is more likely to invoke retrieval to help produce a conclusive and defensible response.

By contrast, at the awareness and consideration stages, responses rely on internally available knowledge and learned understanding rather than external sources. No retrieval or citation is observed, and brand mentions are both less frequent and far more fragmented. Where brand mentions do occur, there is little consistency between runs.



This reflects a mode of behaviour focused on explanation, categorisation, and problem definition rather than resolution. Variability across runs is high, which suggests less emphasis on certainty. In short, the model answers from memory, rather than leaving "no stone unturned" in pursuit of a high confidence, definitive response.

Given these behaviours, influence at earlier stages is unlikely to be seen in retrieval or citation, but through alignment with the model's problem framing, explanatory structure, and category-level narratives. This means encouraging brand mentions through alignment with how LLMs frame and seek to resolve problems in your space.

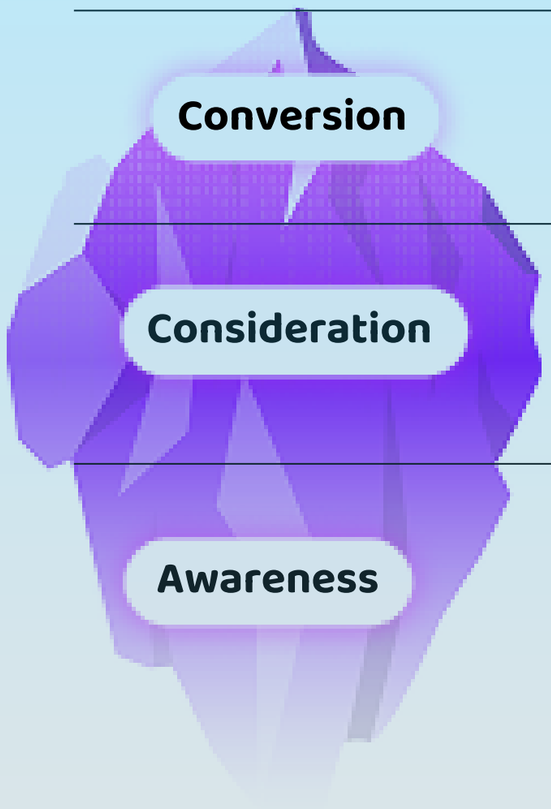
Information gain is an important factor here - is your content contributing meaningfully to the LLM's understanding of the problem space? Content marketing must evolve away from synthesizing and summarising existing knowledge for common queries, and focus instead on adding new knowledge.

Given the extent of variability both within and between responses at every stage, with the exception of only the very few conversion stage queries, AEO strategies must also evolve. We need to stop viewing traffic, and by extension citations, as the end goal and instead focus on influencing our market and potential buyers through AI.

> **Information gain is a critical content metric for AEO-focused brands. LLMs synthesize knowledge - to compete and contribute, your content must create it.**

In fact, we'd go so far as to say that the funnel is no longer a useful visualisation for an organic strategy in a landscape dominated by Answer Engines. We prefer to use an iceberg, reflecting the fact that only a tiny percentage of LLM influence can be seen in the metrics brands currently track: citations and LLM-sourced traffic.

Why most AEO strategies are just the tip of the iceberg



	Frequency How often are brands invoked?			Brand Count How many brands feature in responses?		Variability How consistent are brand citations between runs?	Language What kind of language is used to frame responses?
	Mention	Retrieval	Citation	Mean	Median		
Conversion	**94%**	**48%**	**48%**	**6.39**	**6.33**	**100%** Low (no) variability / tight convergence	Directive, selective, decision-oriented language
Consideration	**68%**	**0%**	**0%**	**2.75**	**1.17**	**33%** Moderately high variability / partial convergence	Comparative, evaluative, trade-off-based language
Awareness	**60%**	**0%**	**0%**	**1.59**	**0.67**	**22%** High variability / low convergence	Descriptive, category-level, exploratory language

Note: This is the median % of runs in which the top 3 brands re-appear.

When citations finally appear at conversion, variability has dropped to near-zero. The model has already decided which brands are viable.

LLMs only cite and link to brands in bottom-of-funnel, conversion stage queries. The majority of complex buyer journeys exist in the awareness and consideration stages, currently completely overlooked in AEO measurement. This is what we have taken to referring to as Dark AI.

Defining Dark AI

Dark AI refers to the vast majority of AI interactions that exist “beneath the surface” - without generating clicks, referrals, or any trackable signal (by existing metrics). Every one of your buyers now operates with it’s own personalised shopping assistant. When buyers use LLMs early in their journey to explore problems, compare approaches, and define requirements, these conversations create the framework against which decisions are made without showing up in your analytics.

When a modern buyer pops his or her head above the parapet, they have already been educated by AI. Absence of traffic doesn't mean absence of impact.

Most AEO advice right now focuses on measuring and optimizing brand mentions and treats direct citations as the holy grail. This works for low-consideration purchases but for complex purchases, orienting an AEO strategy around optimization of the conversion stage is a poor strategy that neglects to influence the majority of the buyer journey.

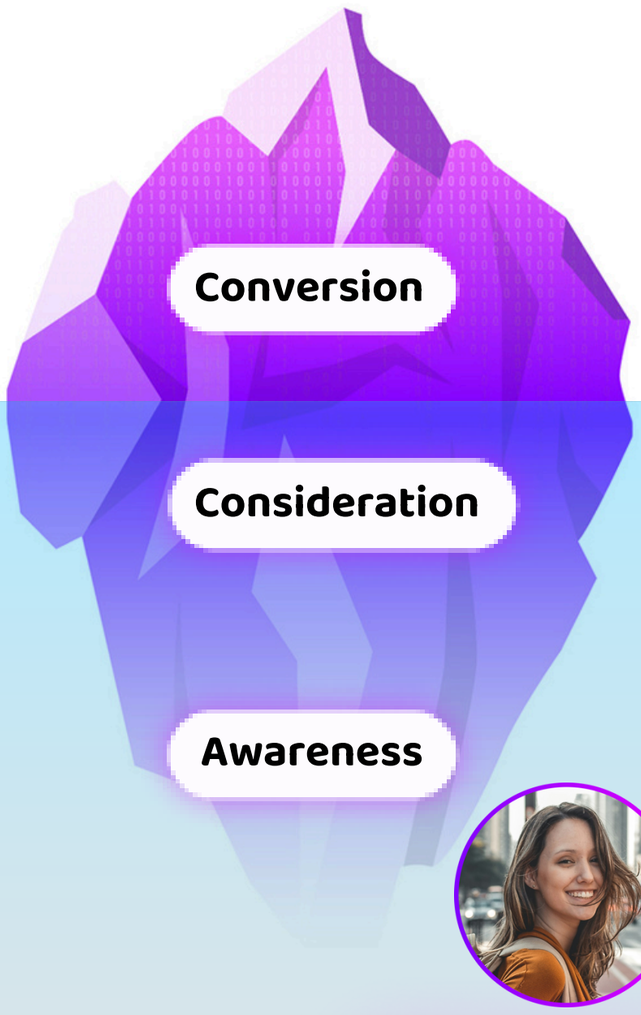
Brands need to hone in on their AEO goals and stop treating AEO like the new SEO. It is a fundamentally different and multi-faceted discipline. Is there an SEO component? Yes. But the larger challenge and opportunity is to influence the invisible majority of the buyer journey that does not show up in the traditional metrics we are accustomed to from 25 years of SEO.

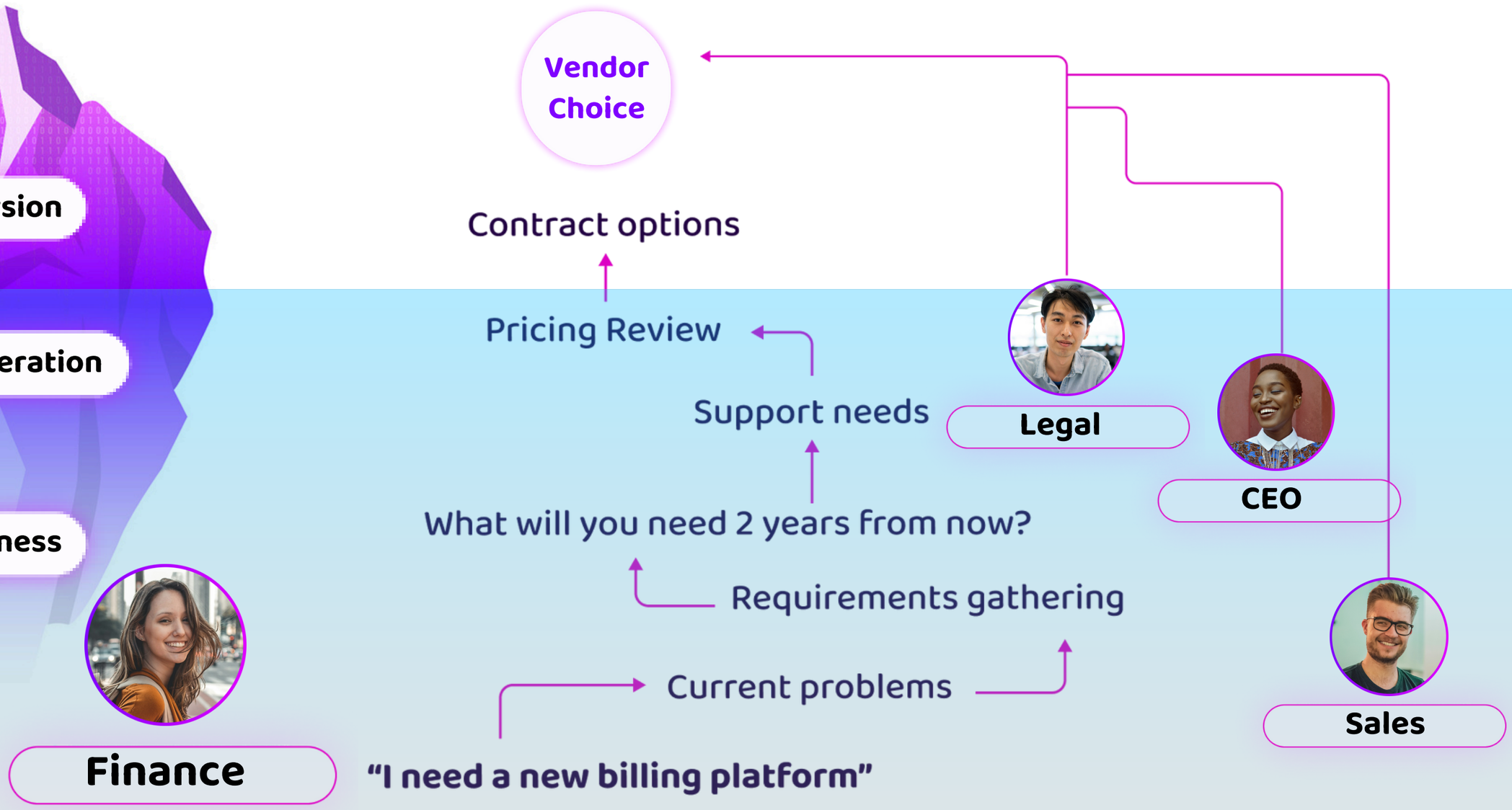
> **Your goal should be to influence problem framing in a way that favours your strengths, downplays your weaknesses, and makes visibility inevitable.**

This is the critical finding of this study: the AEO battleground is fundamentally different and centres on your ability to influence Dark AI, the vast majority of LLM interactions where problems are framed, requirements built, and category winners decided.

The most surprising finding of this study is that all the evidence we have gathered suggests that these two goals (solution-focused "ranking" and problem-focused "influence") are more connected than we would have expected. We thought they were two distinct jobs. They aren't. While some manipulative AEO tactics appear to provide a short term "ranking" boost, by and large, **your "ranking" is an inevitable consequence of your ability to influence Dark AI and thereby have LLMs frame the buyer's problem in a way that makes your solution inevitable.**

Consider the difference between three purchases a CFO might make.





1

Toothpaste

It's a short trip from awareness ("I'm out of toothpaste" → lets ignore the unthinkable scenario of a CFO just now discovering they should, in fact, be brushing their teeth) to conversion. There is very little consideration. It is a largely commoditised market in which price and offer timing are the key variables.

2

Trainers

Awareness might be a more complex process. Maybe our CFO needs to understand that her lower back injuries can be related to bad footwear. Consideration might need to factor in the type of exercise the individual carries out. The majority of the decision will still come down to conversion-related factors, though: size, colour, price.

3

Billing Solution

This is a complex, high-consideration purchase. Problem-awareness may take months of education and major market or technological advances that render an existing solution sufficiently painful relative to a newer alternative.

Consideration is likely to require input from several stakeholders and detailed evaluation of current and future needs.

Conversion is a relatively simple component of the overall decision.

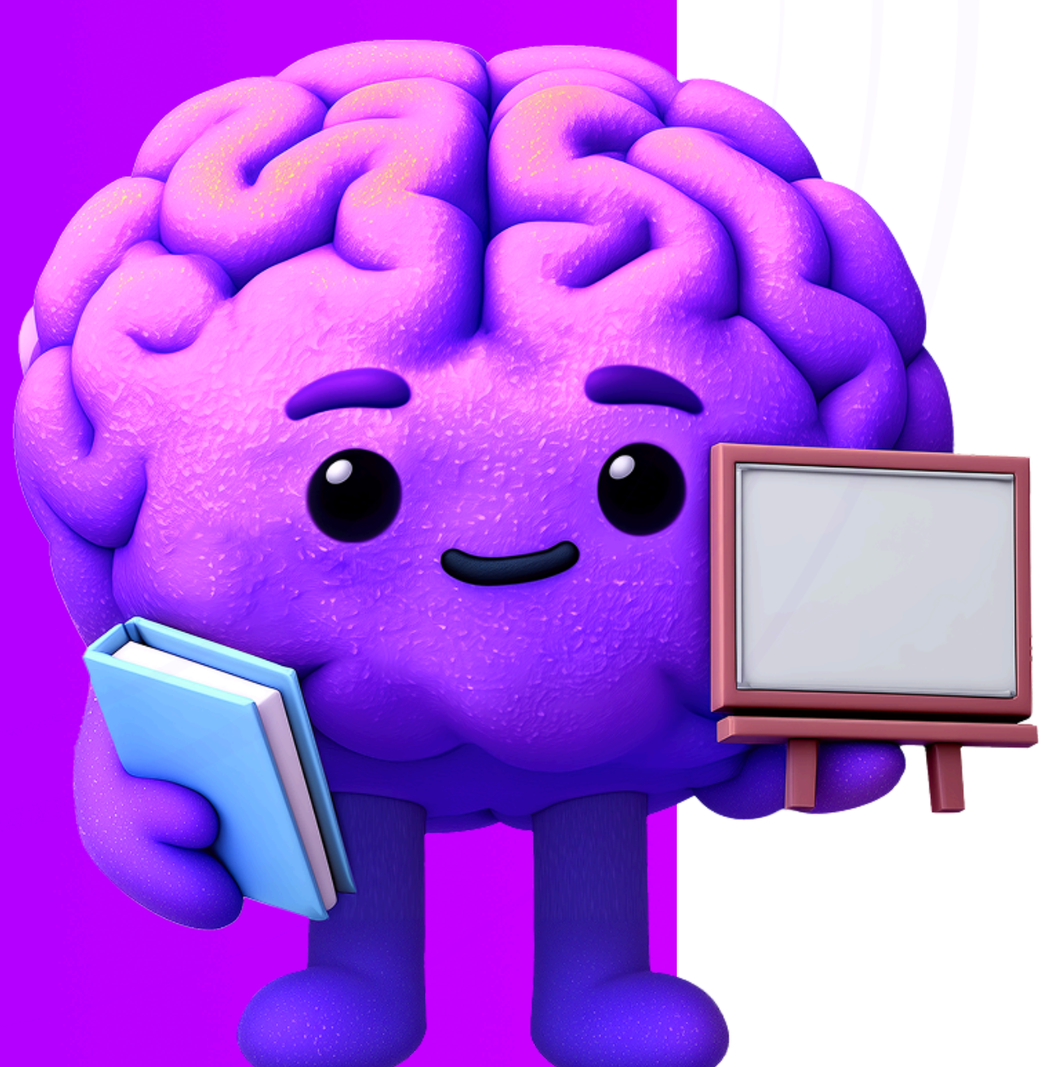
Understanding Convergence

Why citation frequency is a weak metric for AEO.

The pattern of AI convergence

Why your AI visibility metrics are inflated

Surviving convergence & shaping comparisons



Understanding AI Convergence:
Why Citation Frequency Is a Weak Proxy for Influence

Now, bear with us because at this point, we need to get a little nerdy. One of the critical components of this study was to understand why brands show up in LLM responses, not just whether they do. In order to do that, we had to assess variability between prompt runs and look into what causes that variability. What we observed was a clear phenomenon we have termed "convergence".

Convergence describes the way that LLMs work consultatively alongside a buyer to arrive at a recommendation. Put simply, when engaged to solve complex problems LLMs do not arrive at a recommendation by casting a wide net and matching a content snippet to a query, as existing AEO advice would suggest. It converges on viable solutions by narrowing down the option pool. This is really important to understand.

What is Canon Concentration?

To understand convergence, we need to introduce a measurement framework. **Canon concentration** measures how consistently the same brands appear when you run the same prompt multiple times. We use three metrics:

- K1: Consistency of the top brand surfaced reappearing across runs.
- K3: Consistency of the top three brands all reappearing across runs.
- K5: Consistency of the top five brands all reappearing across runs.

Values range from 0 to 1, where:

- 0 means high variation (no consistency)
- 1 means perfect consistency (complete consistency)

These metrics describe consistency, not preference. They tell us when the model is exploring versus when it's enforcing a narrowed decision space.

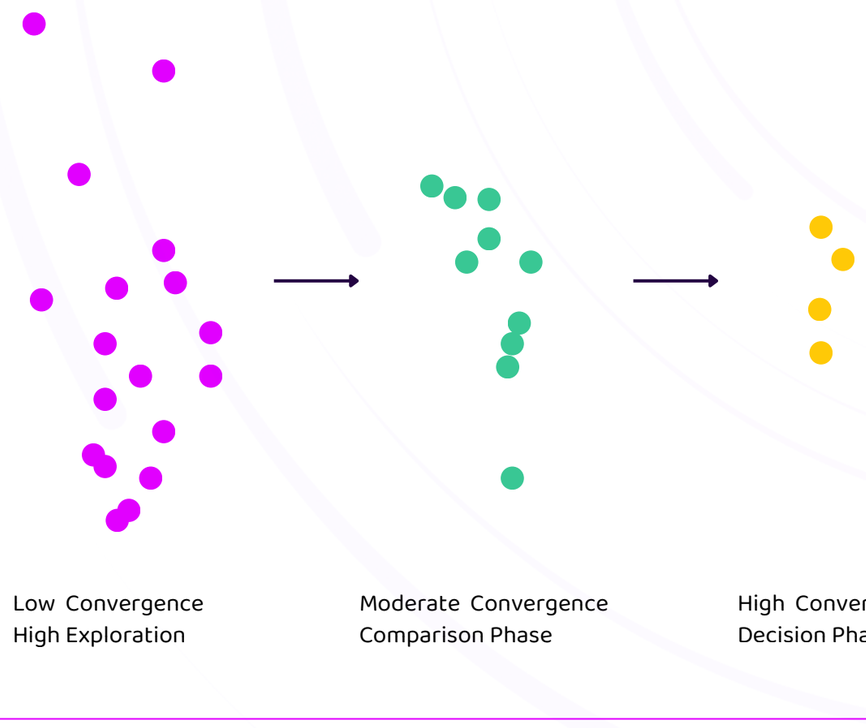
The Convergence Pattern

One fascinating trend we observed in the dataset is what we term "convergence." While brand mentions become more common as we move from awareness to consideration and ultimately conversion, the opportunity to influence results and garner a brand citation actually becomes harder.

Canon concentration shows that LLMs become less exploratory and more conservative as user intent becomes more decision-oriented.

Awareness	Consideration	Conversion
K1: 0.37	K1: 0.44	K1: 0.82
K3: 0.32	K3: 0.38	K3: 0.79
K5: 0.26	K5: 0.32	K5: 0.70

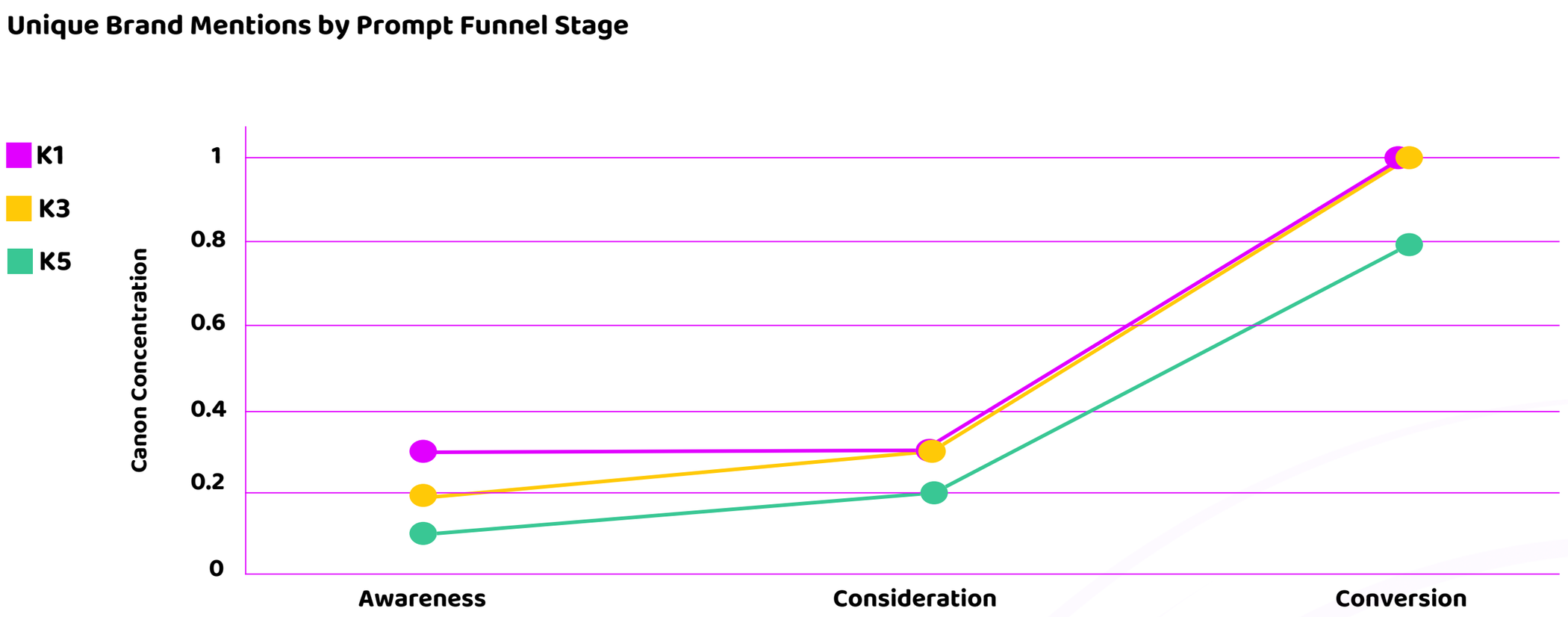
LLM Convergence Pattern Across the Buyer Journey



Convergence is simpler than it sounds. As an LLM seeks to provide a definitive response to a decision-oriented query, it consolidates its option pool to just a few brands. In a sense, it becomes risk-averse, honing in on a few high credibility options and rarely deviating.

The trend in the data is clear:

- Awareness prompts feature low convergence. Models vary widely run-to-run.
- Conversion prompts feature tight convergence. Models repeat the same small set of brands.



High variance is important because it shows the model is still exploring the problem space. It hasn't yet framed the requirements of a solution. Its uncertainty creates a window for influence.

Most AEO tracking focuses on how often a brand appears. Canon concentration explains why that appearance is happening and gives us a better picture of what can help a brand earn citations.

Why High Canon Environments Inflate Visibility Metrics

The data makes this clear: high canon environments mechanically inflate visibility for brands. A small number of brands receive disproportionate mentions and citations, and this appears to be disconnected from AEO execution.

When the model operates under tight constraints and has few valid options available, appearing "often" becomes easier and analytically less informative. High citation or mention rates in high-convergence queries reflect the structure of the task and the narrow options that exist more than the effectiveness of any specific AEO activity.

Put more simply: your brand citations are impacted far more by the constraints of the query and context than by any FAQ, AI Article or 'snippet-ready' paragraph you have written.

Now, you might say, quite frankly, "who cares." We are appearing in the answers. Isn't that all that matters? Yes and no. Of course that is the ultimate goal, but there are two things happening here which mean that continuing to treat citation for conversion-stage (high convergence) prompts as your north star is a dangerous approach.

1. You are making investment decisions off the back of this data

Common sense and practice is to double down on channels that are appearing to work well. If you are mistaking visibility for influence, it is likely you will allocate budget needlessly in the mistaken belief that your AEO activities are critical to its performance.

2. There is likely a self-fulfilling prophecy at play

One of the biggest challenges of AEO measurement is the sheer number of different prompt combinations. Search forced users to consolidate queries to short keyword strings, making it quite possible to track relevant keywords and create content that matches them 1:1. AI prompts are detailed, personal and verbose. Tracking all the different permutations is impossible, and attempting to optimize them 1:1 in an SEO-style approach will quickly turn into a (dangerous) game of whack-a-mole.

We often hear that one of the biggest challenges with getting started with AEO measurement platforms is building the list of prompts to track. Most brands (sadly, at the suggestion of the tools they use) do this by scraping their own site for relevant queries.

You can see the flaw. We work back from our content to build the list of prompts to track content performance against. And look at that, we're doing well! Of course you are. You're choosing prompts that reflect an option pool that has already been narrowed in your favour. What you want to know is whether you are narrowing the right conversations in your favour and whether you are maximising the number of option pools that converge toward you.

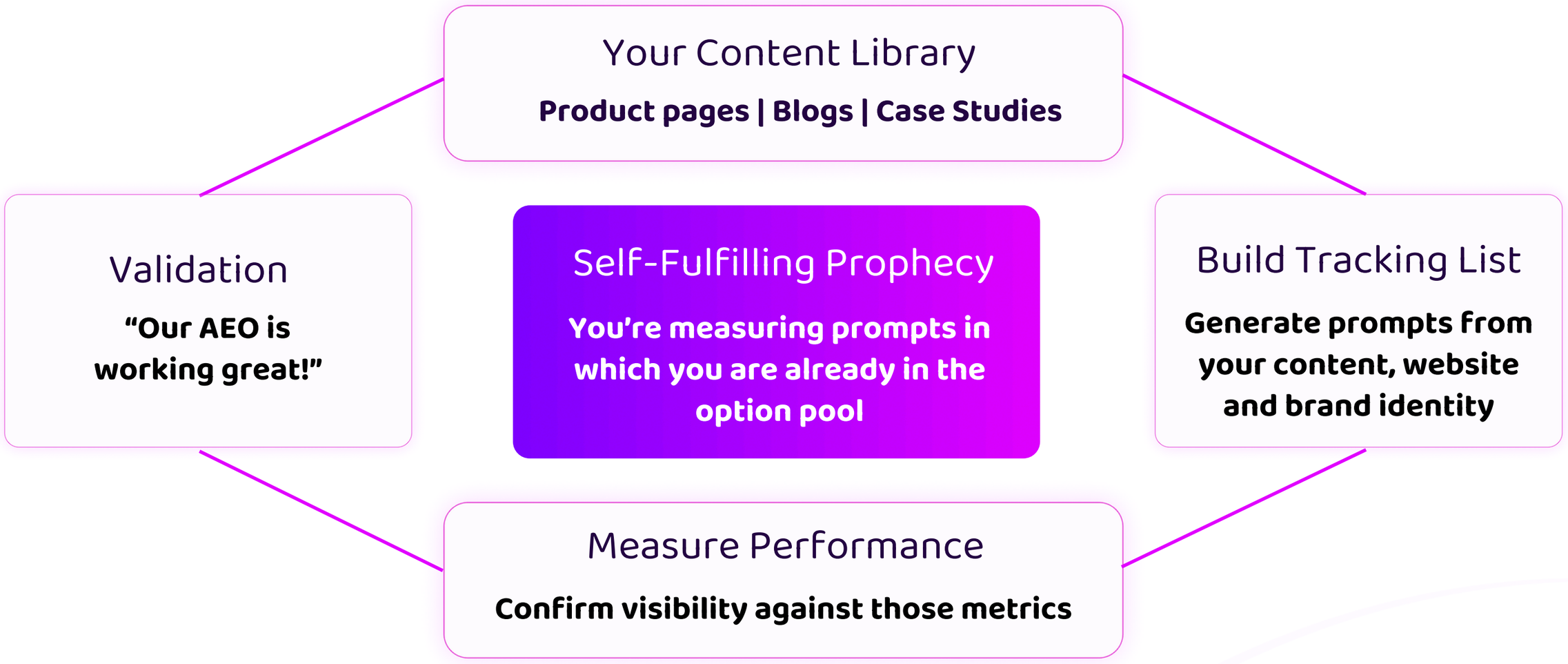
> **Marketing leaders are caught between an urgent, board-level pressure to “show up in AI” and a real lack of nuanced or evidence-backed insight into how to achieve that.**

For this reason, AEO measurement needs to be more sophisticated and should largely sit upstream where the option pool is broad. This is where you can truly measure influence and where the real AEO battlefield exists. Are we influencing the direction of convergence and ensuring more and more option pools narrow in such a way that makes our visibility inevitable?

In Summary

Canon concentration shows that high AI visibility often reflects a constrained option pool rather than strong brand influence. As intent narrows, visibility becomes easier to achieve but harder to attribute. Effective AEO is therefore less about winning citations at the point of convergence and more about shaping the problem framing and criteria that convergence later enforces.

For most brands, there is a high risk of self-fulfilling prophecy in AEO measurement.



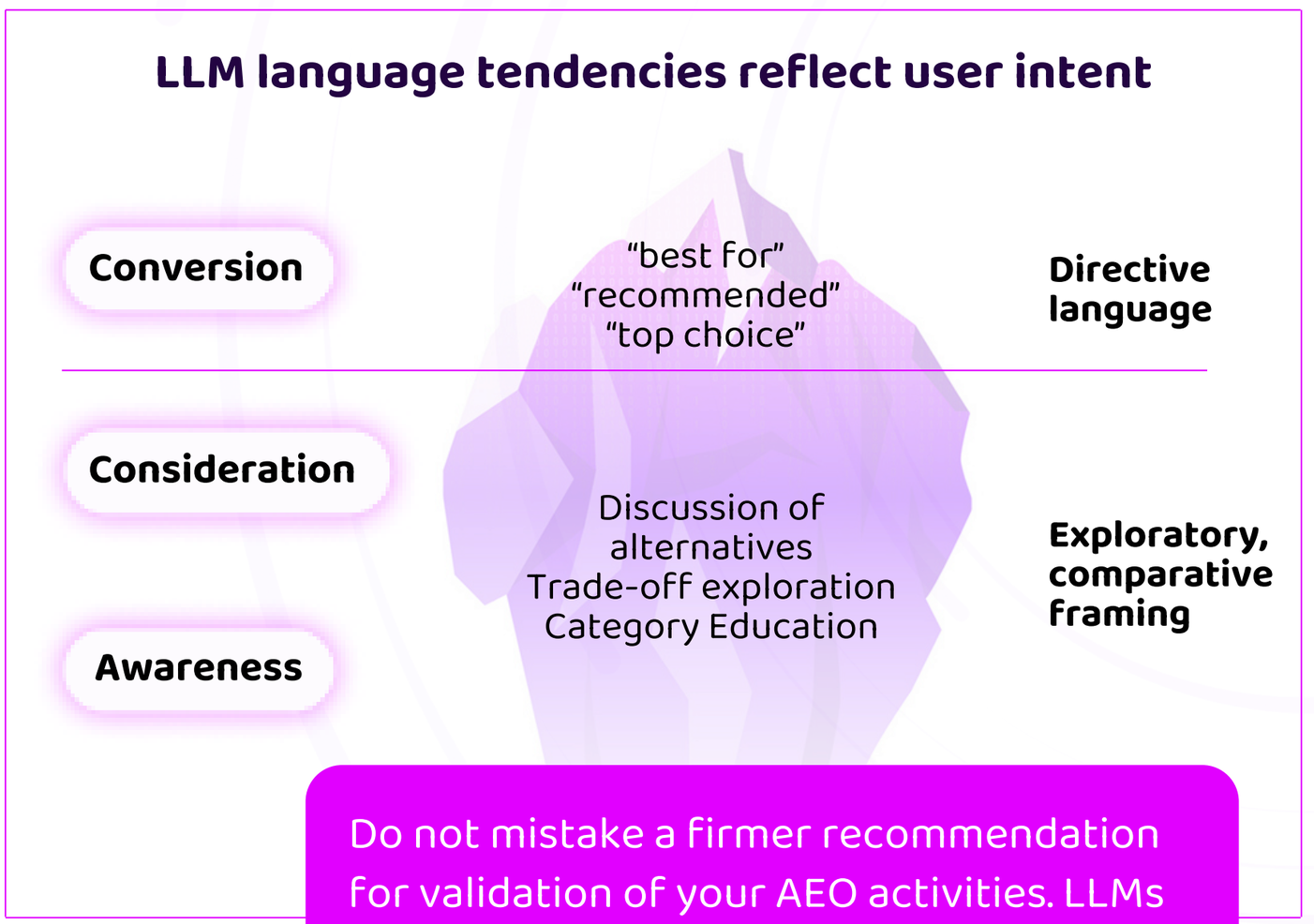
Shaping Comparison and Surviving Convergence

So if Canon concentration shows us that convergence occurs, then what is actually driving that convergence, and how can we build an AEO strategy that exerts true influence? Studying language patterns in the dataset, as intent moves down-funnel, reveals quite a lot about how the model is framing its responses and the kind of content that can shape it.

How Language Shifts Down-Funnel

Another clear feature of the dataset that stood out was how language evolves as responses move down-funnel.

As user intent becomes more decision-oriented, responses move away from exploratory or comparative framing and toward more directive, selective language. We consistently see phrases such as "best for," "recommended," or "top choice". This shift should not be mistaken for preference or endorsement. It simply shows the model matching the user's intent: when a query implies a decision, the model returns a conclusive answer within a constrained option space.



The use of more directive language (a firmer “recommendation”) for decision-oriented queries is not because of your AEO activities. It's an inevitable consequence of your appearance within decision-oriented answers. Sentiment on these post-convergence queries is a very misleading metric.

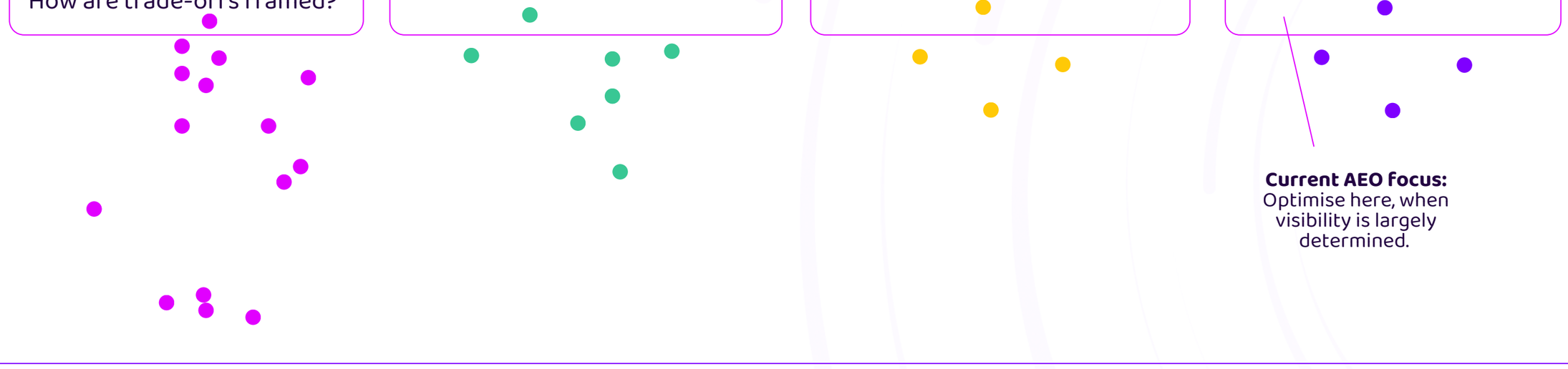
Earlier in the journey, however, a different pattern dominates. Awareness and consideration-stage prompts feature more evaluative language exploring alternatives and trade-offs. The model is comparing approaches, articulating differences, and helping to build constraints rather than narrowing to a single choice or definitive set of options.

This gives us a clear indication of how we want to enable the LLM through our content. Comparative framing is most influential before convergence occurs, when the option space is still being defined rather than enforced. This flies in the face of common copywriting or positioning practices which feature definitive claims of superiority and broad category dominance.

The Sequential Objectives of AEO

This data shows that our optimisation objective should change by stage. At TOFU and MOFU, the primary goal is not selection but influence over comparison and problem framing: shaping how the problem is defined, which attributes are considered relevant, and how trade-offs are articulated. At BOFU, the objective shifts to surviving convergence. Can you remain viable as the model applies constraints?

These are sequential objectives, not competing ones. The direction of convergence - towards your strengths, or your weaknesses - is likely to have a bigger impact on BOFU visibility and citation than any optimization efforts. Success at the bottom of the funnel is structurally dependent on your ability to influence the constraints that are applied.



The dataset reinforces this sequencing. Brands rarely "enter" consideration for the first time at BOFU. They are filtered from an earlier, more exploratory phase - as prompts become more specific, the LLM can be more confident.

This creates a real risk of over-optimisation. Content that relies heavily on self-assertive or absolute positioning (for example, broad claims of being "the best overall") may perform well once convergence has occurred, but is less useful in comparative contexts. Such positioning provides limited support for trade-off analysis and can reduce early-stage inclusion, where nuance and balance are required.

There is a clear cost to such definitive framing. You may see a small uptick in your visibility for conversion prompts (this is why much current AEO advice centres on this) but we know this visibility is largely already decided. The cost, though, is that **you fail to influence the LLM's understanding of when you are truly viable**. Given that when perceived as viable, we know visibility to be highly likely, **the real opportunity lies in influencing the direction of convergence, not asserting your dominance post-convergence**.

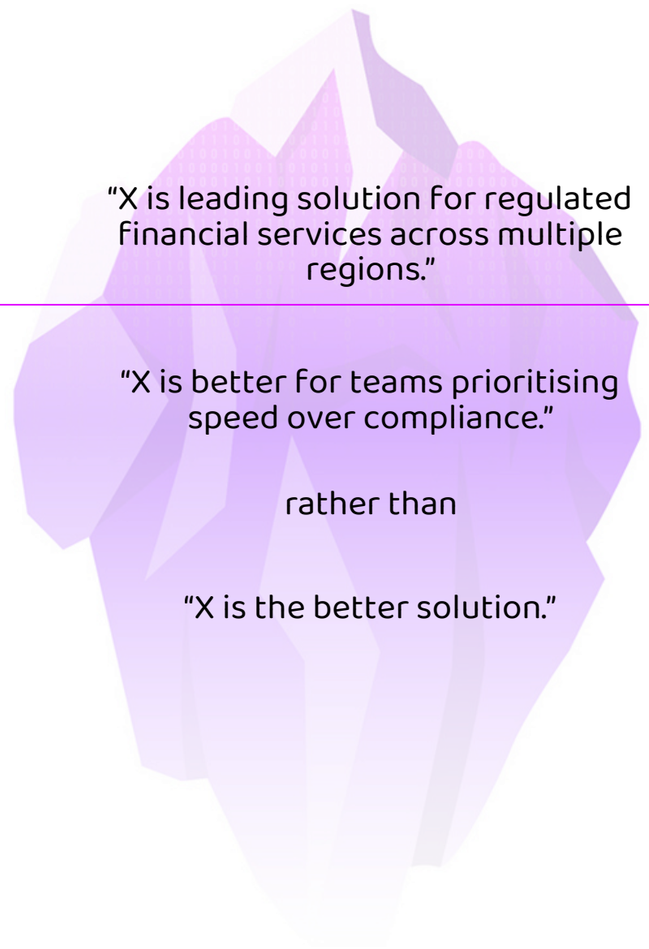
Brands are more likely to be surfaced conclusively when they have influenced which attributes matter, which trade-offs are normalised, and which constraints ultimately become decisive. This influence is established upstream, during problem framing and comparison, not at the moment of recommendation.

Surviving Convergence Requires Nuance Across Stages

So, a key learning of this study is that visibility in complex purchases is less about granular optimization of content than survival. It's a game of winner stays on or guess who, where the gradual application of criteria narrows options. So, the question becomes: how can we influence the number of buyer journeys in which we survive to make a final shortlist?

A strategy that survives convergence requires nuance across stages. Early in the journey, brands should provide balanced, reusable comparative signals that help shape decision criteria and problem definitions. They should seek to contribute information gain that can build authority and enhance the LLM's ability to frame problems. Later, they should be definitive within clear bounds, articulating strength for specific use cases or constraints rather than asserting broad, unqualified superiority.

Best practices for different AEO goals



Conversion

"X is leading solution for regulated financial services across multiple regions."

Definitive and clear, but with clear and realistic bounds

Consideration

"X is better for teams prioritising speed over compliance."

rather than

Awareness

"X is the better solution."

Structured comparative language rather than dead end assertions

Specificity builds confidence. Vagueness weakens it. And LLMs are confidence machines - they seek out validating evidence that allow for a confident answer appropriate for the task.

The core, data-grounded insight is simple but uncomfortable given the direction of most existing AEO strategies: BOFU is not where brands win. Earlier problem framing is enforced at BOFU, but has already taken shape.

Measurement Framework

How to measure AI Search performance in complex categories

Inclusion before convergence

Buying criteria alignment

Information Gain

Dark AI is measurable, and learning to influence it can drive bottom-of-funnel outcomes that matter to Marketers. The challenge is that most current AEO measurement ignores it entirely.

The central conclusion of this study is that effective AEO measurement must move upstream - or "below the surface". If influence is formed when the model is exploring and comparing, then performance measurement must focus on that phase, not in tracking citations against a flawed set of pre-defined prompts.

Rather than treating late-stage visibility as the primary success signal, we suggest adopting a set of metrics designed to capture influence before convergence, when the option space is still broad and malleable. Visibility matters, but it is an outcome rather than a lever in of itself.

What Should We Measure Instead?

1. Inclusion Before Convergence

The first and most fundamental signal is early inclusion. This measures the percentage of Awareness and Consideration prompt clusters in which a brand appears at least once, before the model begins to narrow its option pool. Presence at this stage means the brand is influencing the way the model understands and frames the problem space. **In practical terms, this answers a simple question: When the model is still exploring, are we in the conversation?**

If a brand is absent at this stage, downstream visibility will be limited to instances where convergence made their visibility a foregone conclusion. If brands can expand their exploratory presence, then they can influence the volume of interactions that converge on their area of strength.

At Demand-Genius, we have developed canon-adjusted visibility metrics that account for whether or not convergence is influencing the model's selection of brands. We use canon-weighted visibility as a north star metric. This approach rewards early influence more highly than post-convergence influence and better reflects the impact of an AEO strategy. It also prevents a narrow set of tracked prompts from fooling you into thinking you're maximising performance.

Put simply, it distinguishes influence (shaping whether your brand remains viable) from appearance (showing up within the narrow set of options).

2. Criteria Alignment

Criteria alignment measures how often a brand is mentioned in proximity to specific attributes that matter to key stakeholders in your buying group.

Where Search was a directory, LLMs act more like a personal shopper, tailoring responses to the use cases, trade-offs and constraints that matter to each stakeholder or buyer persona.

We recommend honing in on the criteria that are important to each one of your buyer personas and tracking how your brand strengths (as the LLM views them) align with what matters most to that buyer.

In practice, this means:

Step 1: Identify the 3-5 attributes that matter most to each buyer persona (e.g., for a CFO: compliance, auditability, integration with existing finance stack; for a VP Engineering: API flexibility, developer experience, infrastructure cost).

Step 2: Track how frequently your brand appears in LLM responses alongside those specific attributes, particularly in TOFU and MOFU contexts.

Step 3: Compare this to how often competitors appear with the same attributes.

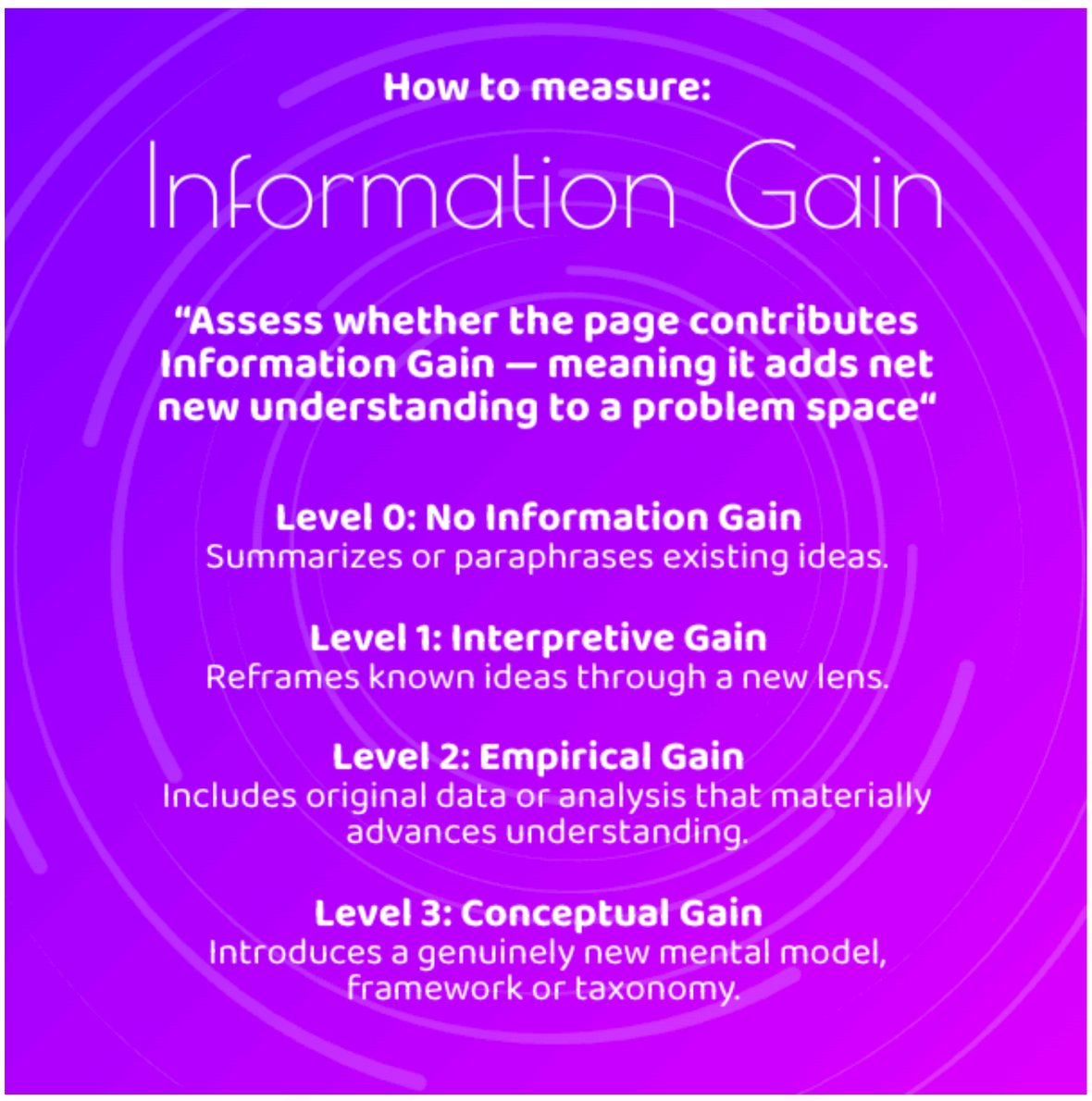
Criteria alignment answers a different question than traditional visibility metrics: When buyers care about X, does the model associate our brand with X?

The stronger the alignment, the more likely your brand is to feature as the option pool narrows because you are demonstrating fit with the criteria being used to narrow.

3. Information Gain

AI systems reward quality content, supported by research, that contributes meaningfully to our ability to understand and solve common problems. Information gain should be a tracked content metric.

Information gain is critical because it influences which brands the model associates with specific problem framings - the very associations that help encourage early (low canon) mentions and influences who survives convergence later.



Analyse a piece of content - is it summarising or interpreting existing knowledge, or creating it? If an LLM could generate the same piece from existing sources, you're not contributing information gain.

Of course, analysing this across your entire library isn't possible as a Human - but AI agents can. The image on the right shows how the AI agent within our Demand-Genius template library (you can try it out free by signing up via our site) assesses and scores information gain.

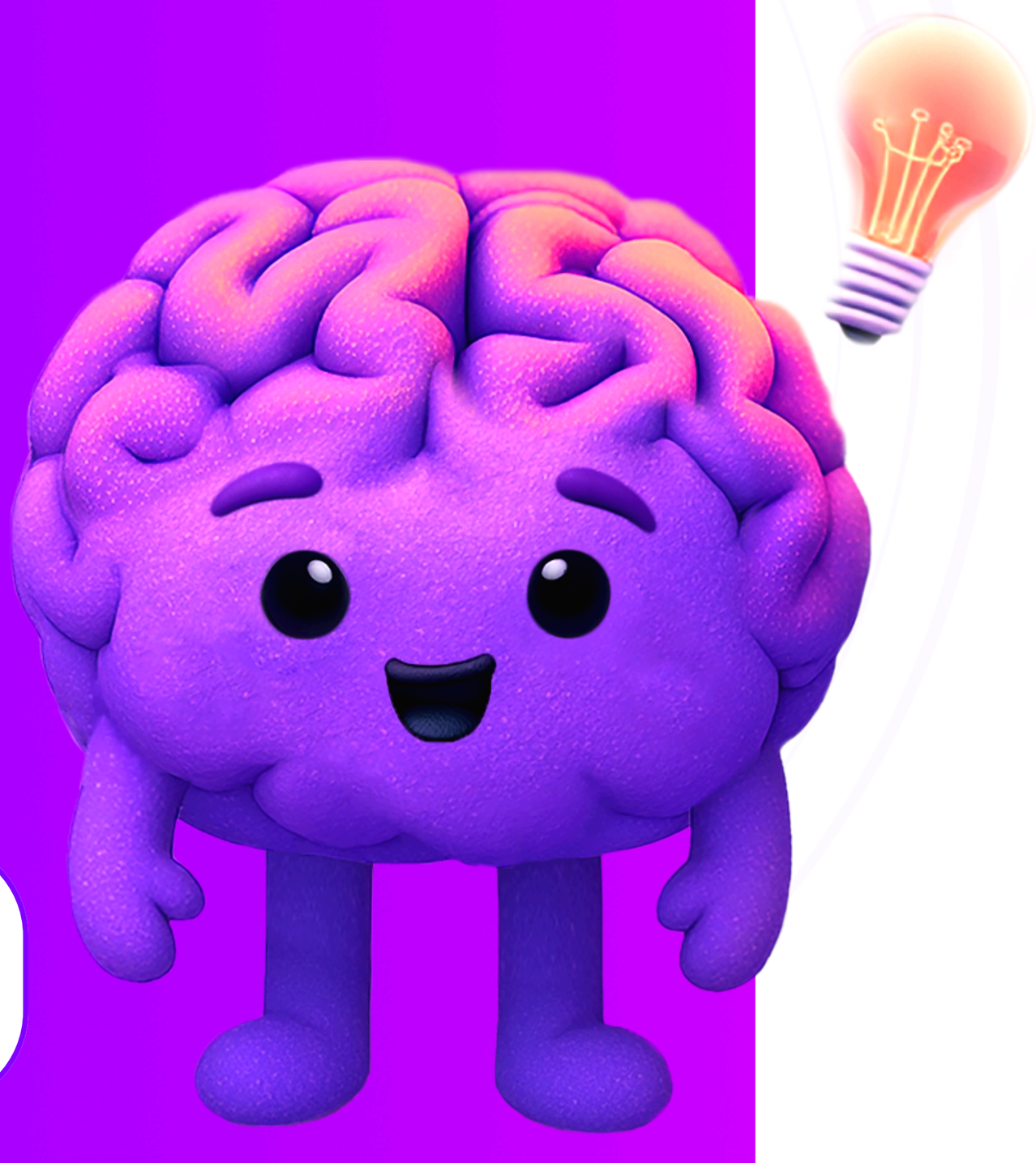
If your site features high information gain, you should see brand terminology, narratives and research show up in exploratory LLM responses.

If you want to think in traditional "distribution" terms, just think how much reach a piece of content or content strategy that contributes meaningful information gain can achieve. If it becomes core to how models understand your problem space, it doesn't just capture a lead or start a conversation - it shapes the direction in which requirements are formed across an entire market. **The right content that fills a meaningful information gap can reach an entire market.**

Demand-Genius

About Us

Helping B2B Brands Win AI Search



Great sellers don't respond to RFPs.

They help the prospect write them.

Demand-Genius

Measure and **influence AI perception**.

How is your brand showing up in **Dark AI?**

Adjusted Visibility

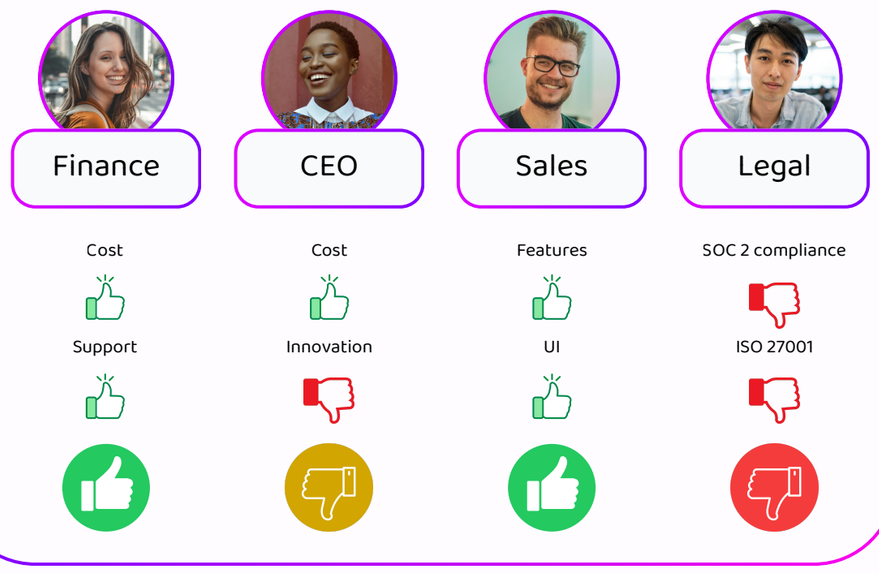
Adjusted visibility metrics that reflect the true value of visibility at each stage of complex journeys.

Action Plans

AI agents that continuously evaluate your content library and AEO performance to suggest gaps, updates & actions.

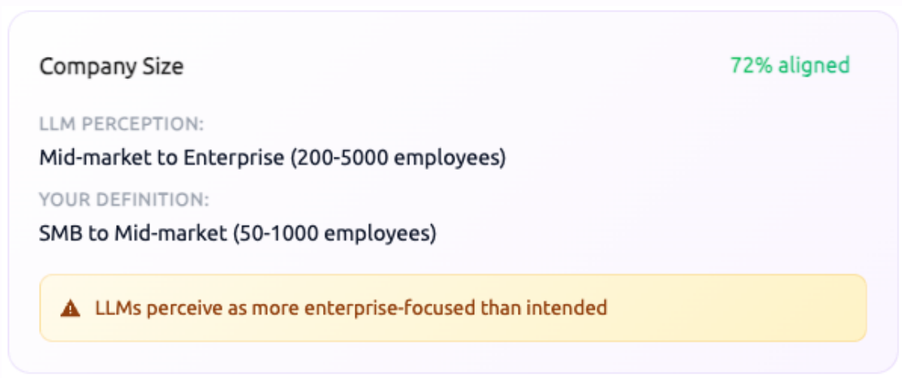
Stakeholder Mapping

Track perceived fit for each buyer as a metric you can report on.



Perception Alignment

Does the LLM's view of your ICP and fit match your intention?



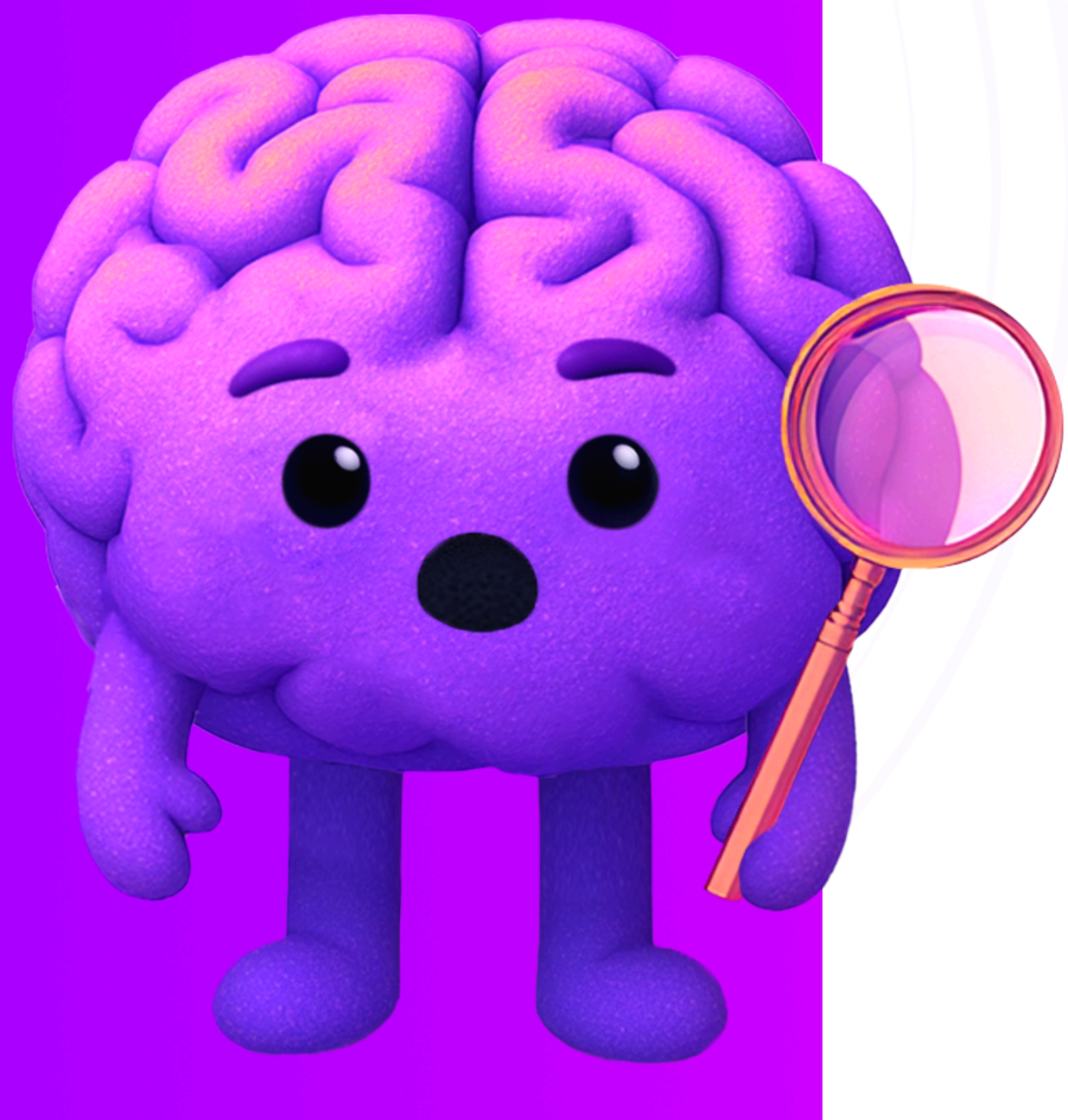
Content Intelligence

Custom and Pre-Built **AI agents** to continuously audit Information Gain and other quality markers across your entire library.

Detailed Buyer Journey Insights to connect visibility to revenue, and **identify prompts worth tracking** from actual journey insights.

Methodology

A deeper look at how we conducted this study.



Methodology

Study Environment

This study was conducted using ChatGPT in a controlled research environment. The goal was not to benchmark models against one another, but to observe consistent behavioural patterns in how a large language model surfaces brands across different stages of the B2B buying journey.

We plan to follow this up with another study that will compare this analysis across models, to identify areas of consistency and divergence across model architectures. Based on early analysis of other models, it appears that the majority of learnings are ubiquitous.

Scope and Scale

The study analysed hundreds of prompt clusters across multiple B2B categories.
Each prompt cluster represents:

- One unique prompt
- Executed three times to observe variability and convergence

We analysed prompts at the cluster level (grouping runs of each prompt together) rather than treating each response independently. This approach let us distinguish between consistent patterns and random variation."

This design allowed us to examine both what brands appear and how consistently they appear across repeated runs.

Prompt Design and Buyer Journey Simulation

Rather than analysing prompts in isolation, the study was designed to simulate end-to-end buyer journeys. For each category, prompts were structured to reflect:

- Awareness (TOFU): problem definition, education, category understanding
- Consideration (MOFU): comparison, trade-off evaluation, shortlisting
- Conversion (BOFU): selection, recommendation, "best for" decision prompts

This approach reflects how real buyers move through complex decisions over time, rather than treating each query as a disconnected event. To capture variation:

- Multiple prompt phrasings were used at each stage
- Prompts varied in specificity, constraints, and stakeholder perspective
- Categories of differing complexity were intentionally included

Categories Analysed

The study included multiple B2B categories, selected to reflect:

- Different levels of purchase complexity
- Varying degrees of commoditisation
- Distinct buyer roles and evaluation criteria

While results are aggregated across verticals in this report, brand-level conclusions are not drawn across categories. The focus is on behavioural patterns, not competitive performance.

The categories were:

1. Billing & Payments Platforms
2. Cloud Security & Posture Management (CSPM)
3. B2B CRM (Complex Sales)
4. Cookie Banner & Consent Management
5. Customer Support & Enterprise CX
6. Data Privacy & Governance Platforms
7. HR Tech Platforms
8. Identity & Access Management (IAM)
9. B2B Insurance Platforms
10. Workflow Automation & iPaaS
11. Product Analytics Platforms
12. RevOps & Revenue Intelligence
13. Procurement & Vendor Risk Management
14. Answer Engine Optimisation (AEO)!

Output Analysis

Each response was analysed across several distinct layers:

- Brand mentions: semantic inclusion of brand names
- Retrieval invocation: whether external sources were used to ground the response
- Citations / links: whether sources were explicitly surfaced to the user
- Brand count: number of unique brands referenced
- Response consistency: overlap of brands across repeated runs

To measure convergence, we introduced canon concentration metrics (K1, K3, K5), which quantify how consistently the same brands appear across repeated runs of the same prompt.
These metrics describe consistency, not preference, and are used to contextualise visibility rather than rank brands.

Addressing Limitations and Gaps

We explicitly acknowledge several limitations:

Individual variation: LLM responses can vary based on session context, system prompts, and user-level signals. While API-based testing reduces this variability, it cannot perfectly replicate every individual user experience. In running prompts across buyer journeys within single categories, we sought to simulate complex purchasing journeys rather than treat each prompt as disconnected.

Model-specific behaviour: Findings reflect the behaviour of ChatGPT at the time of testing and should not be assumed to generalise identically across all models or future versions.

Non-longitudinal design: The study captures a snapshot in time. It does not measure change over time or the impact of specific interventions.

No causal attribution: The study observes patterns of behaviour. It does not claim that specific content or optimisation tactics caused specific outcomes.

These constraints are inherent to studying probabilistic systems and are handled through aggregation, repetition, and conservative interpretation.

Transparency and Data Access

To support independent validation, we provide access to the underlying dataset through Benchmark AI. This allows users to:

- Explore prompt-level and cluster-level data
- Reproduce key statistics
- Query patterns relevant to their own categories and prompts
- Apply the findings to their specific AEO context

Disclosure: Use of AI in the Research Process

AI was used selectively in this study to:

- Assist with prompt generation and variation
- Support structured analysis and aggregation of outputs
- Aid in summarisation and interpretation of large response volumes
- Create first drafts and iterate / feedback on refined versions for clarity

AI was not used to:

- Determine conclusions independently
- Rank brands or assess performance
- Make causal claims or infer commercial outcomes

All findings were reviewed and validated by human researchers to ensure methodological correctness and interpretive restraint. I also got it to write most of this methodology since it's purely factual - this is the only bit where I'm adding some human flair. And unsurprisingly, it's the only bit you won't have learned anything from!

Methodological Intent

The purpose of this methodology is not to predict individual AI responses or rank vendors, but to understand how LLM behaviour changes with user intent.

By focusing on patterns, consistency, and convergence across the buying journey, the study aims to provide a more accurate foundation for evaluating AEO influence and to expose where traditional visibility-based metrics fall short.

Demand-Genius

Build B2B brands that win AI Search

www.demand-genius.com