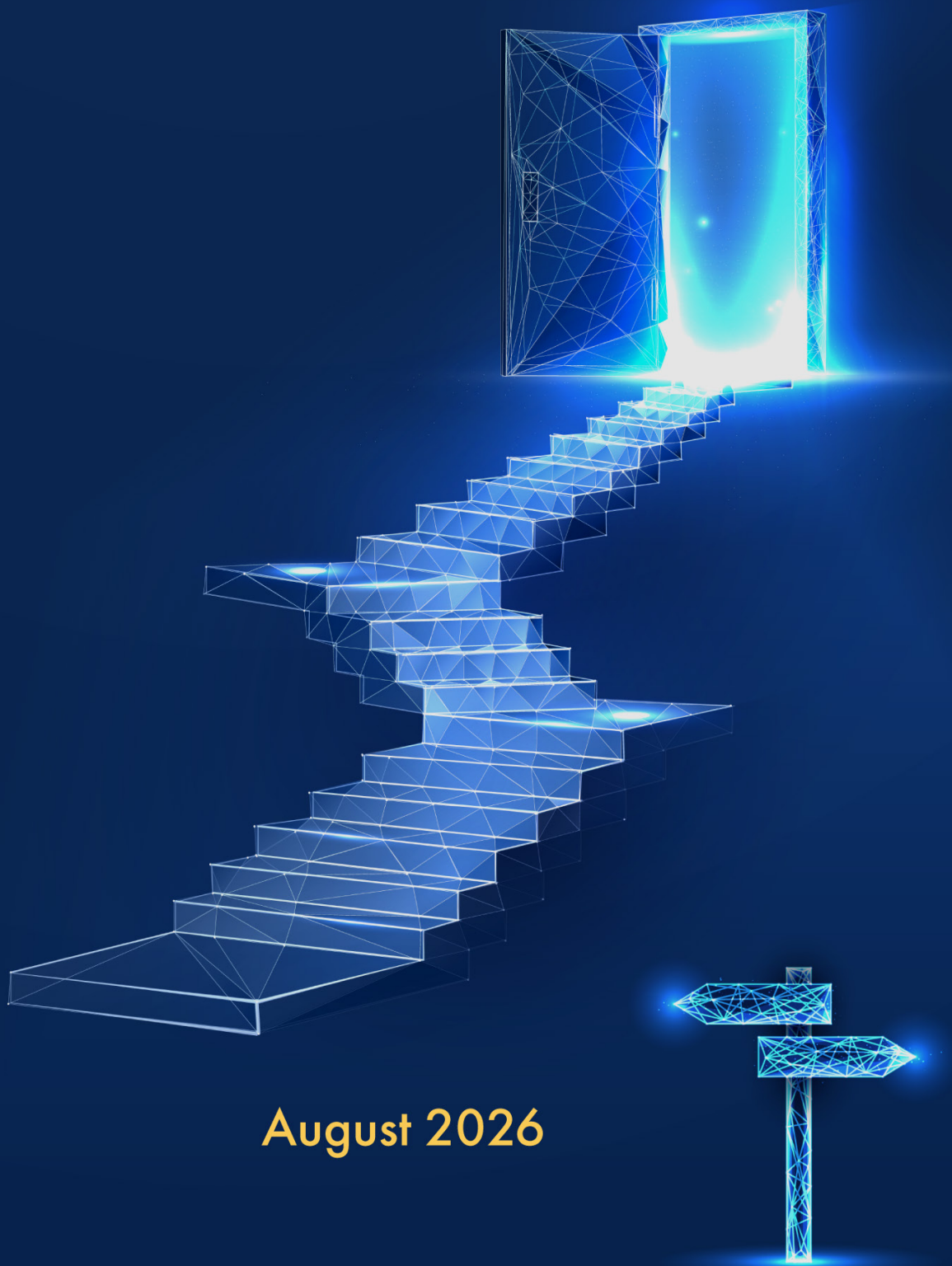
iab.

Measuring **Visibility** in the AI Era

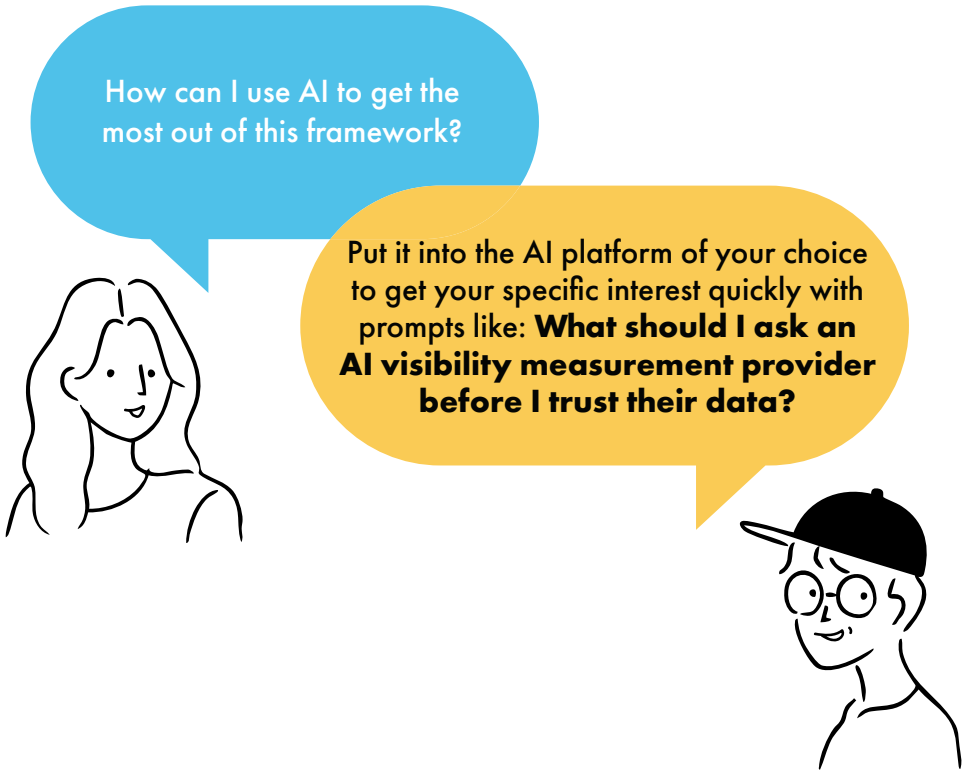


August 2026

Table of Contents

Executive Summary..... pg. 2
Contributing Industry Leaders..... pg. 7
Introduction & Market Context..... pg. 9
Core Metrics Definitions: A Shared Vocabulary Implementation..... pg. 12
Measurement Quality Standards: Is this Data Good Enough?..... pg. 20
Provider Disclosure Framework: What Should Providers Disclose?..... pg. 26
Measurement Stability and Reproducibility: How to Get the Most Consistent Results?..... pg. 31

Maximize this framework



Executive Summary

iab. Measuring **Visibility** in the AI Era

Executive Summary

Operationalizing the Framework

The conversation in marketing right now is about how to optimize for AI-powered discovery: Generative Engine Optimization (GEO), Answer Engine Optimization (AEO), prompt engineering, and structured data for agents. But optimization without measurement is guesswork. A brand cannot know whether its GEO investment is working, whether its share of voice in AI responses is growing or shrinking, or whether a competitor is gaining ground unless there is a common standard for measuring any of it.

The same gap exists on the publisher side. AI platforms are actively ingesting, summarizing, and surfacing publisher content, but publishers have no standardized way to measure how often, in what context, or with what attribution clarity, leaving them without a shared vocabulary to benchmark citation performance or bring data to licensing conversations. More than 20 companies now sell AI visibility tools, each using different methodologies and producing different answers for the same brand or publisher. Buyers have budgets ready to spend. They have no basis for evaluating what they are buying.

The scale of the shift is no longer speculative. ChatGPT has more than 900 million weekly active users and Google AI Overviews have over 2.5 billion monthly users. Google AI Overviews now appear on almost half of searches. McKinsey projects that brands unprepared for this shift could see traffic declines of 20 to 50% from traditional search channels. Publishers are already living with the consequences: Search referral traffic has declined 60% for small publishers, 47% for medium publishers, and 22% for large publishers over the past two years. AI chatbot referrals, despite growing over 200% from December 2024 to December 2025, still account for less than 1% total publisher page views. Only 16% of brands systematically track AI visibility today, in part because there is no common standard for measuring any of it.

What This Framework Does & Does Not Do

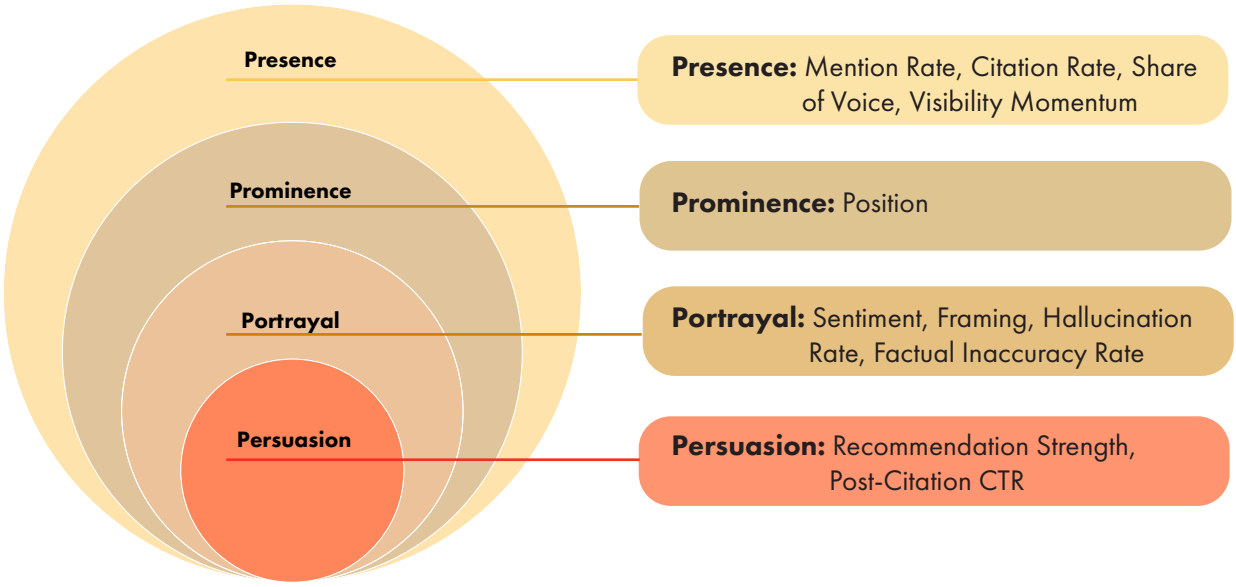
This framework defines what good AI visibility measurement looks like. It does not rate measurement providers, prescribe tools, or build measurement systems. Instead, it provides the shared vocabulary, quality criteria, and disclosure requirements that allow the market to function. Brands and agencies gain a basis for evaluating competing providers. Publishers gain a common language for demonstrating citation value. Measurement companies gain a way to differentiate on rigor rather than claims.

Core Metrics

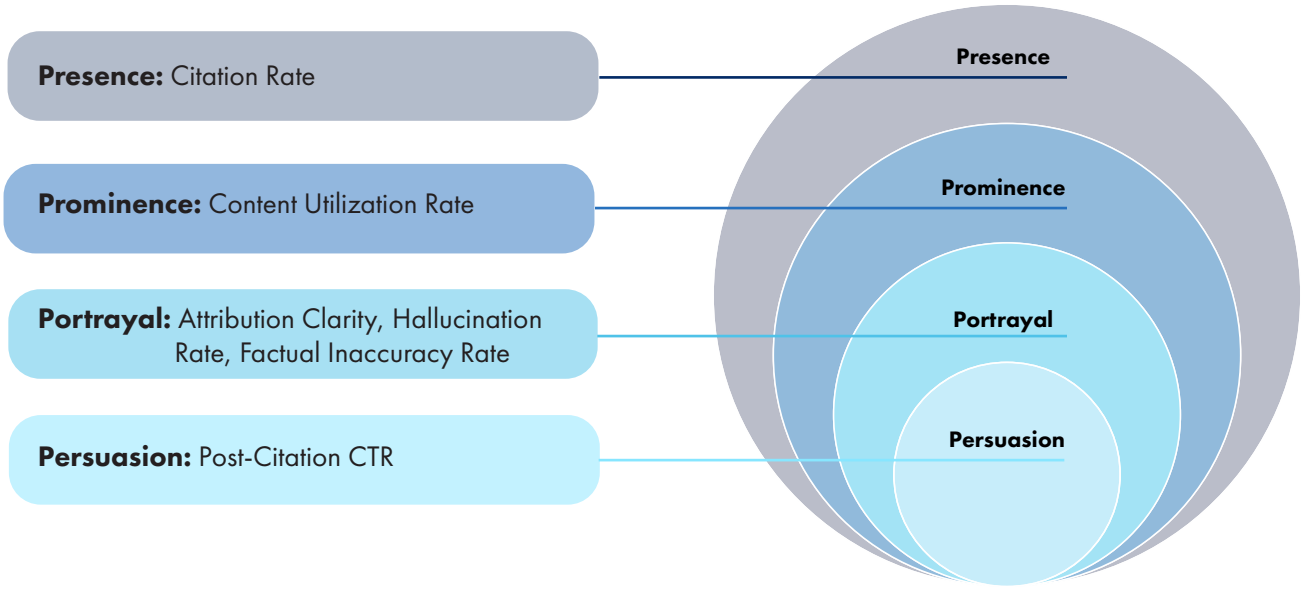
AI visibility metrics are organized into a causal hierarchy that reflects how visibility translates into business value.

4P's of AI Visibility

FOR BRANDS:



FOR PUBLISHERS:



Each metric is defined with the disclosures required to make results comparable across measurement companies. Persuasion metrics serve as a bridge to IAB's forthcoming attribution framework, which will address attribution challenges as AI-powered discovery absorbs more of the awareness and consideration journey.

Measurement Quality: Directional vs. Decision-Grade

Not all AI visibility data is fit for the same purpose.

	Directional Measurement	Decision-Grade Measurement
Primary Purpose	**Identifies patterns and signals trends** Supports early signal detection, internal briefings, and competitive awareness	**Supports high-stakes business actions** Enables budget reallocation, agency performance reviews, and executive strategy
Rigor and Methodology	**Lacks depth** Not sufficient for budget allocation or strategic decisions	**Meets a higher standard of rigor** Rigor across sample size, query volume, prompt type coverage, testing cadence, reproducibility, data validation, methodology documentation, and platform coverage (including multi-platform aggregation)

A criteria matrix specifies the minimum thresholds that separate the two tiers. Both serve legitimate purposes; the failure is in treating directional data as decision-grade without recognizing the gap.

Measurement Company Disclosure: The Mechanism That Makes It Work

Quality tiers are only useful if buyers can verify them. This framework defines the disclosures providers must provide for buyers to make that determination themselves: platform coverage and model versions, prompt library construction and query sourcing, data collection architecture (active query simulation, passive panel observation, or platform-native data), attribution and sentiment logic, factual accuracy classification, and historical baseline management. Where a provider cannot or will not disclose against a required item, that absence should itself be treated as a signal. The disclosure framework establishes the foundation for a potential future IAB partner certification program.

Measurement Stability Over Time

AI platforms do not produce deterministic outputs. Identical queries return different responses, and platform updates can shift visibility metrics overnight independent of brand performance. This framework provides the practices required to operate a measurement program that remains comparable over time: documenting variability baselines, controlling for geographic and temporal variation, distinguishing platform-driven shifts from market-driven shifts, and managing baseline resets when models change. It also provides reporting language for communicating non-deterministic data to leadership without setting expectations the data cannot meet.

Scope

This framework covers organic, non-paid AI visibility measurement. It does not address AEO or GEO optimization or commerce attribution. Paid placement measurement and downstream attribution are identified as gaps that warrant future investigation; as organic and paid visibility increasingly appear on the same response surface, standardized paid measurement is an adjacent priority.

How to Use This Framework

Brands and agencies can apply the disclosure requirements in measurement company procurement and use the criteria matrix to match measurement tier to decision type. Publishers can use the citation metrics to measure how their content surfaces in AI responses and bring that data into licensing conversations as an input. These metrics capture visible contribution including when content is cited, paraphrased, or relied on in a response, but does not include the full set of ways publisher content shapes AI systems. Measurement providers should treat the disclosure framework as the baseline buyers will increasingly expect. The intent is not to constrain the market but to give it the common ground it currently lacks.

Contributing Industry Leaders

iab. Measuring **Visibility** in the AI Era

Contributing Industry Leaders

The IAB AI Visibility Measurement Framework Working Group is a cross-industry team of leaders from brands, agencies, publishers, and measurement providers who collaborated to create this framework. The group's goal is to give the industry a shared vocabulary, quality standards, and disclosure requirements for measuring brand and publisher visibility in AI-powered search by establishing common ground that enables brands and agencies to evaluate measurement providers, publishers to demonstrate citation value, and providers to differentiate on methodological rigor.

This working group is led by Caroline Giegerich, VP, AI at IAB, with special thanks to members:

Breeze Zhao
Sr Director Measurement Science
Walmart

Caroline Giegerich
VP, AI
IAB

Daniel Flynn
Sr. Director, AI Visibility and Measurement
EMARKETER

Graham Wilkinson
EVP, Chief Innovation Officer, and Global Head of AI
Acxiom

Haozhe Xu
Manager, Measurement
WPP Media

Ihab Rizk
Senior Product Manager
Microsoft Clarity

Jason Hartley
Head of Media Innovation and Trust
PMG

Justin Inman
Founder & CEO
emberos

Kristina Meinig
VP, Market Development
Alliance for Audited Media

Paul Longo
GM, AI Advertising
Microsoft Clarity

Scott Cunningham
Fractional Chief Product Officer
Alliance for Audited Media

Simon Poulton
EVP, Innovation & Growth
Tinuiti

Skye Yang
Marketing Science & AI Measurement Leader

Todd Paris
Cofounder, CEO
IQRush.ai

Introduction & Market Context

iab. Measuring **Visibility** in the AI Era

Introduction & Market Context

The AI Discovery Shift

The way consumers find brands, products, and information is undergoing a structural change. For decades, discovery began with a search engine including a query, a results page, and a click. That model is giving way to something fundamentally different. AI-powered platforms like ChatGPT, Google Gemini, Microsoft Copilot, Perplexity, and Claude now synthesize answers directly, collapsing the discovery journey into a single response. Users no longer scan ten blue links and choose what to click. They receive a recommendation, a summary, or a ranked shortlist, often without visiting a single external website.

The scale of adoption is no longer speculative. ChatGPT has over 900 million weekly active users, Google AI overviews have over 2.5 billion monthly users, and Google AI Overviews now appear in almost half of searches, more often on questions or longer queries. Google AI Overviews now appear on 14% of shopping queries, signaling that AI-generated answers are increasingly shaping purchase decisions, in addition to informational searches.

For brands and publishers built on search-driven traffic, the implications are material. McKinsey estimates that unprepared brands could see traffic declines of 20% to 50% from traditional search channels as AI-powered discovery reshapes how consumers navigate the web. For publishers, the shift is already measurable: According to Chartbeat data reported by Axios in March 2026, search referral traffic has declined 60% for small publishers, 47% for medium publishers, and 22% for large publishers over the past two years. AI chatbot traffic, despite growing rapidly, still accounts for less than 1% of total publisher page view referrals. The gap between declining search traffic and incoming AI referrals remains vast.

What makes this moment particularly challenging is that AI responses are not neutral surfaces. They make recommendations, assign authority, and frame brands in ways that shape consumer perception, often without the brand's knowledge and without any standardized way to measure it.

The Measurement Gap

More than 20 providers now sell AI visibility measurement tools, each using different methodologies, prompt libraries, platform coverage, and scoring rubrics. There is no common definition of a "mention," no standard for what constitutes a citation, and no shared framework for evaluating whether a tool's output is reliable enough to inform strategy. Only 16% of brands systematically track AI visibility today. CMOs cite difficulty measuring impact as their top challenge, not because the data doesn't exist, but because the data that exists cannot be compared across tools or trusted with confidence. Brands have budgets ready to spend on AI visibility measurement. The problem is they have no basis for evaluating what they are buying.

This is the gap this framework addresses.

Stakeholder Impact

The measurement fragmentation affects each constituency differently.

For brands, the inability to compare provider outputs creates a procurement problem with real strategic consequences. A brand may receive conflicting share of voice figures from two different tools, with no way to determine which is more accurate, or whether either reflects meaningful measurement. Investment decisions, agency briefs, and executive reporting are being made on data whose reliability could be unknown.

For publishers, AI platforms are actively ingesting, summarizing, and surfacing their content, but there is no standardized way to measure how often, in what context, or with what attribution clarity. Publishers cannot benchmark their citation performance against category competitors or bring evidence to licensing conversations without a common measurement language. Visibility metrics are one signal in those conversations: They capture the share of a publisher's contribution to AI outputs that is observable, not the full set of ways content shapes AI systems.

For agencies, the challenge is both analytical and relational. Clients are demanding AI visibility reporting, but no benchmark exists to contextualize the numbers. When two providers report different results for the same brand, agencies have no framework to explain the discrepancy or advise on which data to trust.

For measurement providers, the absence of standards creates a race to the bottom on claims rather than a competition on quality. A common framework gives rigorous providers a way to differentiate and gives buyers a basis for demanding methodology transparency rather than accepting provider assurances.

IAB's Role

This framework defines what good measurement looks like, establishing the shared vocabulary, quality standards, and transparency requirements that allow the market to function.

Scope

This framework covers organic, non-paid AI visibility measurement only. It does not address AEO or GEO optimization tactics, agentic media buying specifications, or commerce attribution. Paid placement measurement and publisher downstream attribution are identified as gaps that warrant future standardization work. Notably, organic and paid visibility increasingly co-exist on the same response surface. A brand may be cited organically while also holding a paid placement in the same AI answer. This framework measures organic visibility only and does not yet isolate paid influence, but the convergence of the two on a single surface makes standardized paid placement measurement an increasingly urgent adjacent priority. IAB's forthcoming attribution framework will address attribution methodology separately.

Core Metrics Definitions:
A Shared Vocabulary

iab. Measuring **Visibility** in the AI Era

Core Metrics Definitions: *A Shared Vocabulary*

This section establishes the shared vocabulary of AI visibility measurement. It provides standardized definitions and calculation guidance for the metrics that brands, publishers, agencies, and measurement providers need to evaluate AI discovery performance, build comparable methodologies, and compare results across tools.

The 4 P's of AI Visibility

AI visibility metrics are organized into a causal hierarchy that reflects how visibility translates into business value:

PRESENCE: Does the brand appear or is the publisher cited?

PROMINENCE: Where and how prominently?

PORTRAYAL: In what context, and with what accuracy?

PERSUASION: Does visibility drive action?

Brand Metrics

Category	Metrics	Core Question
Presence	Mention Rate, Citation Rate, Share of Voice, Visibility Momentum	Does the brand appear?
Prominence	Position	Where and how prominently?
Portrayal	Sentiment, Framing, Hallucination Rate, Factual Inaccuracy Rate	In what context, and with what accuracy?
Persuasion	Recommendation Strength, Post-Citation CTR	Does visibility drive action?

Publisher Metrics		
Category	**Metrics**	**Core Question**
Presence	Citation Rate	Is the publisher cited?
Prominence	Content Utilization Rate	Where and how prominently?
Portrayal	Attribution Clarity, Hallucination Rate, Factual Inaccuracy Rate	In what context, and with what accuracy?
Persuasion	Post-Citation CTR	Does visibility drive action?

Persuasion metrics bridge to IAB's forthcoming attribution framework. This framework defines them as part of the visibility vocabulary but does not develop full attribution methodology, which will come later.

Brand Metrics

Brand metrics measure how a brand appears, is positioned, and is characterized in AI-generated responses. These metrics apply to advertisers, retailers, and any entity whose commercial identity is surfaced in AI discovery.

Presence

Mention Rate

The frequency with which a brand appears in AI-generated responses across a defined query set. Calculated as responses containing a brand mention divided by total responses. Mention Rate is the most fundamental visibility metric: If a brand is not mentioned, no downstream metric matters. Results are sensitive to query set composition, and different query sets will produce different rates.

Citation Rate

The frequency with which a brand is cited as a source in AI-generated responses, calculated as responses containing a brand citation divided by total responses. Where Mention Rate captures whether a brand is spoken about, Citation Rate captures whether the brand is relied on. It signals the degree to which AI platforms treat the brand's owned content or domain as an authoritative reference. A brand can have a high Mention Rate and a low Citation Rate; the gap between the two is itself a meaningful signal. Citation definition follows the standard set in Publisher Metrics: a linked reference (containing a hyperlink) or a named reference (identifying the brand by name as a source).

Share of Voice

Brand mentions as a proportion of all brand mentions within a defined competitive category. Share of Voice contextualizes Mention Rate against the competitive landscape as a brand's absolute Mention Rate matters less than its relative share. Co-mention patterns, which includes the competitors a brand consistently appears alongside in AI responses, provide useful analytical context on Share of Voice by revealing the competitive set the model treats the brand as part of. Co-mention is an analytical lens on the underlying Mention Rate and Share of Voice data rather than a separate metric. Providers must disclose two methodological choices that materially affect Share of Voice results: how the competitive category is defined (which brands constitute the set and on what basis), and how the total category mention universe is sized within that set. Both choices directly affect whether results can be compared across tools.

Visibility Momentum

The rate of change in visibility metrics over time. Visibility Momentum answers whether brand presence is improving, stagnating, or declining relative to competitors and prior performance. Calculated as the percentage change in Mention Rate or Share of Voice between defined measurement periods. This metric is vulnerable to model updates and platform changes that shift baselines independent of brand performance; providers must disclose whether trend data has been re-baselined after platform changes.

Prominence

Position

Where the brand appears in the formatted AI response as a user sees it, rather than in the AI's raw text output. Position captures whether a brand is the first entity mentioned, its rank when multiple brands are listed, whether it appears as a standalone recommendation or one option among several, how far into the response the mention occurs (even if no other brands are mentioned), and how many times it is referenced within a single response.

Position is straightforward to measure in list-format responses; providers must disclose how they operationalize it in narrative responses where no universal method exists, and how Position itself is measured by analyzing the entire rendered response (including content users would need to scroll to see), by processing a screenshot of the response, or by measuring only what's visible on screen at first view (treating anything below the fold as not appearing).

Portrayal

Sentiment

Sentiment captures whether the brand is described positively, neutrally, or negatively and is most meaningful in informational queries. In recommendation and commercial contexts, neutral descriptive language tends to dominate. Sentiment classification accuracy varies across tools and is especially unreliable for nuanced or comparative language. Providers must disclose their classification methodology, framing taxonomy, and accuracy benchmarks.

Framing

Framing captures the context, whether the brand is positioned as a market leader, a budget alternative, an outdated option, or a niche specialist, or categorically attributed to the wrong type, genre, or sector (for example, a family film classified as a violent thriller).

Hallucination Rate

The frequency of hallucinated brand mentions as a share of total mentions within a defined query set, segmented by platform. A hallucinated mention is one the AI fabricates: an association, attribute, or characterization of the brand that has no basis in any underlying source, or that is attributed to a real source that does not actually support the claim. Hallucinations are an AI platform quality issue, distinct from Factual Inaccuracy Rate (defined below), which traces to incorrect underlying source content rather than AI fabrication.

Providers must disclose their detection methodology and report hallucination rate per platform rather than as an aggregate, since rates can vary significantly across models. Trend monitoring of hallucination rate over time is particularly valuable, as model updates can shift rates materially in either direction.

Factual Inaccuracy Rate

The frequency of brand mentions that accurately reference the brand but are tied to materially incorrect information including wrong product specifications, outdated pricing, misattributed features, incorrect company history, or similar, as a share of total mentions within a defined query set, segmented by platform. Factual inaccuracies occur when the AI is accurately reflecting information that exists in an underlying source, but the source itself is wrong. Both *Hallucination Rate* and *Factual Inaccuracy Rate* cause brand misrepresentation but trace to different remedies. Hallucinations are an AI platform quality issue; factual inaccuracies trace to inaccurate underlying content that propagates through AI responses. The source of truth must be brand-supplied. Providers must disclose detection methodology, report rate per platform rather than as an aggregate, and surface flagged mentions to clients rather than silently excluding them.

Persuasion

Recommendation Strength

The degree to which the brand is actively recommended versus passively cited. This metric distinguishes between a brand named as the best option with specific reasoning ("the best option for X because of Y") and one merely included in a generic list ("a good option"). Assessment considers whether the AI uses specific descriptive language indicating active endorsement, the prominence of the recommendation within the overall response, and whether it is qualified (i.e. "Brand X is a good option if you're on a budget") or unconditional (i.e. "Brand X is the best option"). Providers must disclose the scoring rubric used, as this boundary is a methodological judgment call that varies across tools.

Post-Citation Click-Through Rate (CTR)

The percentage of AI mentions that lead users to a brand's digital property. Post-Citation CTR bridges visibility measurement and attribution, capturing whether AI visibility produces measurable downstream traffic.

Bridge metric: This framework defines Post-Citation CTR as part of the visibility vocabulary. Full attribution methodology will be addressed in IAB's forthcoming attribution framework. Availability is dependent on platform-level data as not all AI platforms provide click-through data.

Publisher Metrics

Publisher metrics measure how publisher content is cited, utilized, and attributed in AI-generated responses. These metrics serve editorial publishers, news organizations, trade press, and any content creator whose work is ingested and surfaced by AI platforms.

Presence

Citation Rate

The frequency with which publisher content is cited as a source in AI-generated responses, calculated as responses containing a citation to the publisher divided by total responses. Citation Rate is the publisher equivalent of Mention Rate.

Citation definition: A citation is a linked reference (containing a hyperlink to the source) or a named reference (identifying the publication by name without a hyperlink). Implied citations, where content appears drawn from a source without explicit attribution, are excluded because they cannot be measured consistently across platforms or providers.

Citation Decay Rate

How Citation Rate changes over time for a specific piece of content or a publisher's content portfolio. Citation Decay Rate measures the shelf life of content within AI systems. Decay patterns may reflect model retraining schedules or index refresh cadence rather than content relevance, and providers must disclose the time intervals used and whether decay curves are adjusted for known model update events.

Prominence

Content Utilization Rate

The degree to which publisher content is substantively drawn upon versus superficially cited. This metric distinguishes between a response that quotes, paraphrases, or builds upon a publisher's reporting and one that merely lists the publisher as a source. For example, an AI response that summarizes a publisher's findings or references specific reporting represents higher utilization than one that appends a source link without drawing on the content. Unlike simple citation detection, this requires comparing publisher content against the AI response. Providers should use semantic (meaning-based), lexical (word overlap), or hybrid similarity, and should disclose which approach they use.

Portrayal

Attribution Clarity

How explicitly the publisher is identified when its content is used. Attribution Clarity measures the spectrum from full attribution (linked citation with publication name, article title, and date) to partial attribution (publication name only). Providers must disclose how attribution levels are classified.

Hallucination Rate

The frequency of hallucinated publisher citations as a share of total citations within a defined query set, segmented by platform. A hallucinated publisher citation is one where an AI platform attributes a claim, quote, piece of reporting, or source material to a publication that never produced it.

For publishers, the reputational stakes are particularly acute: A publisher's authority and credibility rest on the accuracy of their editorial record, and being falsely cited for reporting that does not exist is a material harm to that reputation.

Providers must disclose their methodology for verifying that cited content actually exists and originates from the attributed source, report hallucination rate per platform rather than as an aggregate, and surface hallucinated citations to clients rather than silently excluding them.

Factual Inaccuracy Rate

The frequency of publisher citations that accurately attribute content to the publisher but mischaracterize what the publisher actually reported, claimed, or concluded, as a share of total citations within a defined query set, segmented by platform. Distinct from hallucinated citations (where the AI attributes content the publisher never produced), factual inaccuracy occurs when the citation itself is valid but the AI misrepresents the substance of the cited work. For publishers, the reputational stakes parallel those of hallucinated citations: Misrepresentation of a publisher's editorial record damages credibility even when attribution is correct. Providers must disclose detection methodology, report rate per platform rather than as an aggregate, and surface flagged citations to clients rather than silently excluding them.

Persuasion

Post-Citation Click-Through Rate

The percentage of AI citations that lead users to a publisher's digital property. For publishers, Post-citation CTR indicates whether AI visibility translates into audience traffic.

Emerging Metrics

The following metrics appear in current provider offerings but are not yet ready for standardization. The measurement infrastructure is nascent, interpretability challenges remain unresolved, or validation against real business outcomes is insufficient. This framework includes them to acknowledge their existence in the market and to signal that further methodological development is needed before they can be incorporated into the core vocabulary.

Competitive Displacement Rate

The frequency with which a brand is mentioned in AI responses while a direct competitor is not. This is straightforwardly represented as responses mentioning Brand A but not Brand B, divided by total responses. That said, the denominator is interpretively unreliable: A competitor's absence from a response could mean the AI preferred the focal brand, or it could mean the query did not surface the category at all. The metric cannot distinguish between those two cases, which makes standardization difficult. Thus, it is included as an emerging metric for now.

Attention Proxies

Emerging indicators that approximate user attention within AI responses, including mention position, response length relative to brand mention placement, and citation click signals where available. These are directional metrics only and are not yet linked to real business outcomes. Providers offering attention proxies must disclose how each proxy is operationalized, as methods vary significantly across tools.

Multimodal Visibility

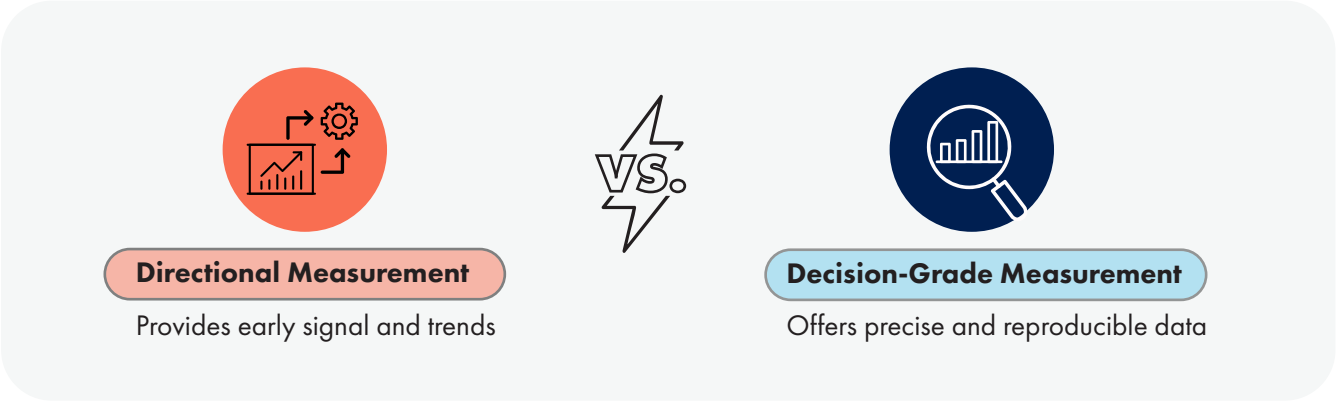
As AI systems increasingly generate image, voice, and video responses, text-based metrics alone will become insufficient. Key measurement questions include how tone and delivery affect brand perception in voice responses, how brands are visually represented in AI-generated imagery or product cards, and how visibility should be measured when AI combines text, images, and interactive elements in a single response. Measurement infrastructure for multimodal visibility is nascent. This framework flags it as a priority for future standardization and recommends that providers begin documenting how their tools handle non-text response formats.

Measurement Quality Standards: *Is This Data Good Enough?*

iab. Measuring **Visibility** in the AI Era

Measurement Quality Standards: *Is This Data Good Enough?*

More than 20 companies now offer AI visibility measurement, and each describes its output as reliable. However, the methodological differences between tools, in query volume, platform coverage, testing cadence, and reproducibility, are significant enough that two providers measuring the same brand in the same category can produce materially different results. Buyers currently have no shared framework for evaluating whether the data they are purchasing is fit for its intended purpose. This section provides a two-tier classification system that defines what separates directional intelligence from decision-grade measurement, and what providers must disclose for buyers to make that determination themselves.



The Directional vs. Decision-Grade Framework

Directional Measurement answers questions like "Is our brand visibility generally increasing or decreasing?" and "Are we appearing in AI responses for our core category?" Directional data is valuable for early signal detection, internal briefings, and competitive awareness. That said, it is not sufficient for budget allocation, provider selection, or executive strategy decisions that require precision and reproducibility.

Decision-Grade Measurement meets a higher standard of rigor across sample size, query volume, prompt type coverage, testing cadence, reproducibility, data validation, methodology documentation, and platform coverage (including multiplatform aggregation). It supports conclusions like "Our share of voice in the mid-range SUV category declined four points quarter over quarter, driven by a specific competitor's gains in recommendation-type queries." Decision-grade data can inform budget reallocation, agency performance reviews, and strategic planning, provided the underlying methodology meets the thresholds defined below.

Both tiers serve legitimate purposes. The failure is in treating directional data as decision-grade without recognizing the gap. This framework ensures that the gap is visible.

Hallucination & Factual Inaccuracy Detection

Hallucinated mentions, as defined in Core Metrics, inflate Mention Rate, distort Share of Voice, and corrupt Sentiment and Framing analysis. Beyond data integrity, hallucinations carry brand safety implications: A brand falsely associated with a product, claim, or context it has no connection to represents a reputational risk that measurement providers have a responsibility to surface, not suppress.

Providers must document their methodology for identifying and excluding hallucinated mentions from measurement data, including the detection approach used and how filtering is applied across metric calculations. Providers must also report hallucinated mentions to clients rather than silently excluding them. Brands have a legitimate interest in knowing when and how they are being misrepresented by AI platforms. Buyers should treat the absence of a documented hallucination detection methodology as a material gap in any provider's quality claim.

Factual inaccuracies carry the same data integrity and brand safety implications as hallucinations. A brand accurately mentioned but tied to wrong product information, or a publisher accurately cited for reporting that itself contains errors both represents misrepresentation that providers have a responsibility to surface, not suppress. Providers must document detection methodology for both, apply filtering consistently across metric calculations, and report flagged mentions. Buyers should treat the absence of a documented detection methodology, for either hallucination or factual inaccuracy, as a material gap in any provider's quality claim.

Hallucinated Mentions

Attributes claims or quotes to brands or publication erroneously
AI platform error

Damages brand or publisher credibility
Reputation distortion

Factual Inaccuracies

Validates the source but the souce is incorrect
Source factually incorrect

Damages brand or publisher credibility
Reputation distortion

Bias Awareness

AI visibility measurement data can be systematically skewed by two distinct sources of bias, each requiring different controls.

Prompt-driven bias occurs when a provider's query set is not balanced across query types, phrasings, or intents which causes certain brands or content types to appear artificially stronger or weaker than they would in a representative sample. A query set heavy on recommendation-format prompts, for example, will favor brands that

perform well in recommendation contexts and underrepresent their performance in informational or comparison queries. Prompt-driven bias is controlled through query set transparency requirements. Providers must disclose their prompt taxonomy distribution so buyers can assess whether the query set reflects a balanced and representative sample of how consumers actually use AI platforms.

Platform-driven bias occurs when a given AI platform's training data or response behavior systematically favors or disadvantages certain brands or publishers independent of their actual market position. Platform-driven bias is detectable through cross-platform comparison using identical query sets and is controlled through multi-platform aggregation requirements: Providers must report results per platform separately so that platform-specific patterns are visible rather than obscured in a combined score.

Criteria Matrix

The following matrix defines minimum thresholds for each measurement tier across the dimensions that most directly affect data reliability.

Criterion	Directional	Decision-Grade
Appropriate Use Cases	**Trend monitoring, early signal detection, competitive awareness, internal briefing**	**Budget allocation, agency briefs, executive strategy, provider selection, performance benchmarking**
Sample Size Responses collected per query	Multiple responses per query to capture typical variation	Enough responses per query to establish a stable distribution; provider must disclose responses per query and how consistency is determined
Query Volume Total queries issued across the measurement program	Minimum query set that covers the category (e.g., running shoes). The framework treats fewer than 50 queries per measurement program as exploratory rather than directional, since query volumes below that floor cannot meaningfully characterize a category	Large, diverse query set with category and subcategory coverage (e.g., 'running shoes,' 'trail running shoes,' 'running shoes for flat feet') and must disclose total volume, queries per category, and variety of prompt formats
Prompt Type Coverage Distribution of queries across intent types: **informational** ("what is X"), **comparison** ("X vs. Y"), **recommendation** ("best X for Y"), and **transactional** ("where to buy X")	Coverage of at least two intent types, with the distribution disclosed	Coverage across all four intent types, with the distribution disclosed and results available segmented by intent type

Testing Cadence How often queries are run and results collected	Monthly or quarterly: suitable for trend monitoring	Weekly or more frequent: suitable for competitive response and budget allocation
Reproducibility Consistency of results when the same queries are re-run	Results will vary across runs; provider must document how much variation is typical	Provider must define acceptable variation ranges within a 7-day time window, report confidence levels, and maintain a documented process for verifying consistency
Data Validation How the provider verifies that data is accurate, comparable, and free of systematic bias before reporting	Summary-level description of data validation approach, including how data quality is assessed and known limitations	Full disclosure of: what external signals the data is validated against (e.g., platform-native data where available, brand or publisher first-party analytics, controlled panel data); how comparability is maintained when drawing from multiple sources; controls applied for sample integrity, including outlier handling and entity disambiguation; the baseline anchoring the data and how it is established; and the known limitations of the validation approach
Methodology Documentation Depth of disclosure on how measurement is conducted	Summary-level: AI platforms covered, data collection method, how often the provider updates its data	Full documentation: why the provider chose the queries it did, how the provider turns raw AI responses into numbers, how the provider handles each AI platform individually, how results from different platforms get combined into one number
Platform Coverage Which AI platforms are included in measurement	One or more platforms; single-platform data acceptable with disclosure	Platforms collectively representing a substantial majority of consumer AI traffic in the target market of measurement; weighting and market share basis disclosed
Multi-Platform Aggregation How results from different platforms are combined into a single score	Single-platform or multi-platform; if aggregated, disclosure of which platforms are included	Must disclose per-platform results separately, show how much results vary across platforms, and document how platforms are weighted in any combined score

Not all metrics defined in Core Metrics Definitions are appropriate for all decision types at each tier. Presence metrics such as Mention Rate and Citation Rate are measurable at directional scale and suitable for trend monitoring. Prominence and Portrayal metrics, including Position, Sentiment, Framing, and Recommendation Strength, involve methodological judgment calls that amplify variance and require higher reproducibility for decision-grade use. Persuasion metrics, particularly Post-Citation CTR, depend on platform-level data availability and are currently directional for most implementations. Buyers should evaluate provider claims about any individual metric against the tier thresholds in the matrix above, not assume that a provider's overall tier classification applies uniformly to every metric they report.

Query Set Transparency as a Quality Dimension

The construction of a provider's prompt library is one of the most consequential methodological choices in AI visibility measurement, and one of the least visible to buyers. Two providers measuring the same brand in the same category (e.g., running shoes) can produce materially different results based solely on differences in how they construct, weigh, and refresh their query sets.

This framework treats query set transparency as a quality dimension, not merely a disclosure requirement. Providers must disclose their prompt structuring methodology, their thematic clustering approach (i.e., how queries are grouped into categories and whether those groupings reflect industry convention, proprietary taxonomy, or client customization), their distribution of queries across intent types [**informational** ("what is X"), **comparison** ("X vs. Y"), **recommendation** ("best X for Y"), and **transactional** ("where to buy X")], and their query selection rationale, including whether query sets are grounded in real consumer search volume data where available vs. synthetic user data. Buyers assessing whether a provider's measurement qualifies as directional or decision-grade should weigh query set transparency as a key consideration. A provider with strong reproducibility metrics but an opaque or poorly constructed prompt library may produce precise answers to the wrong questions.

Between the Tiers

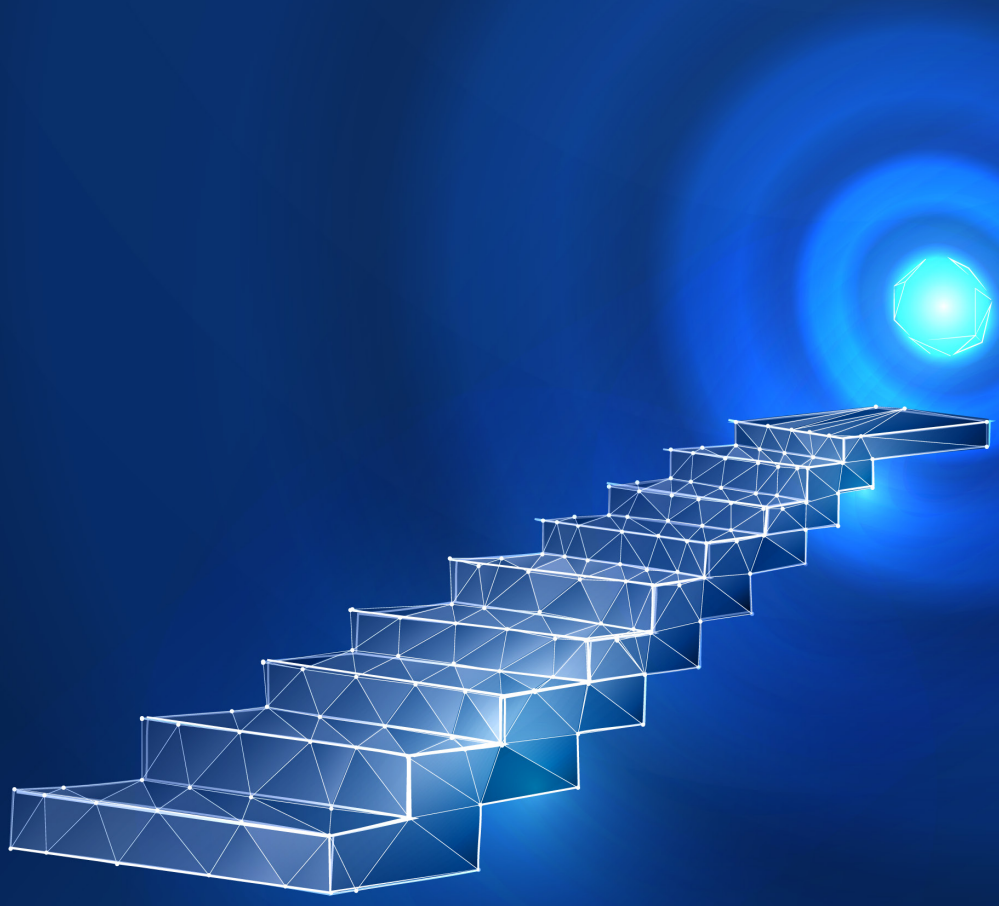
Directional data can support higher-stakes conclusions when signals converge across multiple sources and measurement periods, and are corroborated by first-party analytics. It should not be used for high-stakes decisions based on a single measurement period or when results conflict across platforms or providers.

Executive Reporting Guidance

AI visibility data is probabilistic, non-deterministic, and sensitive to query composition, model version, and platform behavior. Reporting it to leadership using the conventions of traditional search metrics sets expectations the data cannot meet. This framework recommends reporting trends over time rather than point-in-time snapshots, always disclosing the measurement tier, and contextualizing results against known limitations. A Share of Voice figure reported as "approximately 22%, plus or minus 4 points based on current measurement precision" is more useful to a decision-maker than a bare "22%" that implies false precision.

Some providers offer composite visibility scores that aggregate multiple metrics into a single index. These can be useful for tracking overall trajectory, but they require weighting decisions that materially affect the output. This framework requires that any provider offering a composite index disclose the full weighting methodology, the normalization method used (e.g., max-normalized, z-scored, percentile rank), the individual metric values that feed the composite, and the limitations of the aggregation approach. The component metrics defined in this framework are the authoritative measurement vocabulary; composite scores are a presentation layer, and the underlying data must always be accessible.

Provider Disclosure Framework: *What Should Providers Disclose?*



iab. Measuring **Visibility** in the AI Era

Provider Disclosure Framework: *What Should Providers Disclose?*

The Measurement Quality Standards section establishes what separates directional from decision-grade data. This section defines what providers must disclose for buyers to evaluate that distinction themselves. Without standardized disclosure, buyers are left comparing marketing claims rather than methodologies, and two providers producing materially different results for the same brand have no common ground on which to explain the divergence.

This framework defines the minimum information a buyer needs to assess whether a provider's measurement is fit for purpose, and it establishes a common disclosure format that brands, agencies, and publishers can apply in procurement. Where a provider cannot or will not disclose against a required item, that absence should itself be treated as a signal.

Required Disclosures

What Is Being Measured	
DIsclosure Item	**What Providers Must Disclose**
Platform Coverage	Which AI platforms are included in measurement, with model version specificity.
Prompt Library Construction	The size, structure, and refresh cadence of the prompt library, including total query volume, the distribution of queries across categories, how prompts are weighted within the library, and whether prompts were tested across varied demographic and cultural framings.
Query Sourcing	Whether query sets are constructed synthetically (provider-defined), derived from real consumer search volume data, drawn from observed user behavior, or a combination. Where real-world data informs the query set, providers must describe the source and how queries were selected or weighted.
Source Type Classification	Which source types the provider tracks (brand-owned, retailer, editorial publisher, authority or institutional) and how those sources are weighted or differentiated in measurement outputs.

How It Is Being Measured	
DIsclosure Item	**What Providers Must Disclose**
Data Collection Architecture	Which of the following approaches underlies the measurement: (1) active query simulation, where provider-constructed queries are sent to AI platforms and responses captured; (2) passive behavioral observation, where real user behavior is tracked via an opted-in panel; (3) platform-native data, where first-party reporting from the AI platform itself is used; or (4) a hybrid of the above. Providers must not aggregate or present outputs from different architectures as equivalent without explicit disclosure.
Data Collection Methodology	How queries are issued, how responses are captured, and how often collection occurs. Providers must disclose the access method used (official APIs, licensed data feeds, browser-based observation, user-permissioned panels, synthetic sessions, or automated scraping) and whether that method complies with applicable platform terms of service, rate limits, and access restrictions. Providers must also disclose whether queries are issued with live web retrieval enabled, disabled, or both. Results collected under different retrieval configurations or access methods should not be combined without disclosure.
Panel Validity	For any provider using passive behavioral observation, panel size, recruitment methodology, incentive structure, and demographic representativeness must be disclosed against the target AI-using population. Panels that are small, narrowly recruited, or demographically unrepresentative produce behavioral signals that cannot be generalized, and buyers should treat undisclosed panel composition as a material gap.
Data Validation	What external signals the provider validates data against (e.g., platform-native data where available, brand or publisher first-party analytics, controlled panel data); how comparability is maintained when drawing from multiple sources; controls applied for sample integrity, including outlier handling and entity disambiguation; how baselines are established; and the known limitations of the validation approach.

How It Is Being Measured	
Disclosure Item	**What Providers Must Disclose**
Mention Detection, Attribution, and Sentiment Logic	Which of the following approaches underlies the measurement: (1) active query simulation, where provider-constructed queries are sent to AI platforms and responses captured; (2) passive behavioral observation, where real user behavior is tracked via an opted-in panel; (3) platform-native data, where first-party reporting from the AI platform itself is used; or (4) a hybrid of the above. Providers must not aggregate or present outputs from different architectures as equivalent without explicit disclosure.
Factual Accuracy Classification	Whether the provider's tools distinguish between hallucinated mentions (as defined in Core Metrics Definitions), accurate mentions, and factually incorrect mentions in which the brand is genuinely referenced but characterized with inaccurate information. The source of truth used for comparison must be brand-supplied. Providers must disclose how discrepancies are detected and flagged. Reporting each condition separately, rather than aggregating into a single mention count, is required.

How the Data Is Maintained	
Disclosure Item	**What Providers Must Disclose**
Historical Data and Versioning	How far back historical data is available, how model updates are handled in the historical record, and whether trend data is re-baselined after platform changes. Trend metrics, including Visibility Momentum and Citation Decay Rate, are only interpretable when the historical baseline is stable or when re-baselining events are explicitly disclosed. providers must document when re-baselining occurred, what triggered it, and how pre- and post-baseline data should be compared.

Disclosure Format

Standardized disclosure is only useful if it is comparable. This framework recommends a common disclosure format that providers can complete in response to buyer RFPs and that buyers can apply consistently across procurement processes. The goal is to reduce the friction of cross-provider evaluation and to establish a baseline that the market can build on.

Minimum and Enhanced Disclosure Tiers

Minimum disclosure covers the items required for a buyer to place a provider's output into context: platform coverage, data collection architecture, prompt library construction at a summary level, and attribution and sentiment logic at a summary level. Minimum disclosure supports directional use cases and establishes that the provider has a coherent methodology, even if not every methodological choice is documented in detail.

Enhanced disclosure covers all items in the Required Disclosures section above at full documentation depth, including query sourcing methodology, full prompt library construction detail, panel validity for passive architectures, factual accuracy classification, and historical data versioning. Enhanced disclosure is expected for any provider whose output is positioned as decision-grade or used to inform budget allocation, provider selection, or executive reporting.

Buyers should match the disclosure tier they require to the decisions they intend to make with the data, and treat a provider's willingness to provide enhanced disclosure as a meaningful signal of methodological confidence.

Toward a Provider Certification Program

The disclosure framework defined here is the foundation for a potential future IAB provider certification program, which would formalize the minimum and enhanced tiers and provide independent verification of provider disclosures. This framework does not establish certification or evaluate individual providers; it defines the disclosure vocabulary and format a certification program would build on, and enables buyers to apply certification-aligned rigor in procurement today.

Measurement Stability & Reproducibility:

How to Get the Most Consistent Results

iab. Measuring **Visibility** in the AI Era

Measurement Stability & Reproducibility: *How to Get the Most Consistent Results*

Quality Standards define what separates directional from decision-grade data at a point in time. Provider Disclosure defines what buyers must be told about how that data is produced. This section addresses a third question:

How is a measurement program designed and operated over time so that results remain comparable week to week, quarter to quarter, and across platform changes that are outside the operator's control?

The practices below apply whether the program is run by a provider, an agency, or an in-house team.

Understanding Non-Determinism

AI platforms do not produce deterministic outputs. The same query issued twice to the same model can return different responses, with different brands surfaced, different sources cited, and different framing applied. This is how large language models generate text and it cannot be engineered away.

Non-determinism has two consequences for measurement design.

1. **Single-response measurement is not measurement.** A brand's visibility on a given query is a distribution, not a value, and any metric derived from one response per query reflects a sample of one. Stable measurement requires enough responses per query to characterize the distribution. Even then, uncertainty remains. Recent empirical research measuring citation patterns across Perplexity, ChatGPT, and Google Gemini found that what looks like a clear difference between two brands or publishers is often within the range of normal random variation. Reliable measurement requires not only enough responses to characterize the distribution, but also reporting results as a range rather than a single number.
2. **Observed change must be distinguished from inherent variability**. A four-point movement in Share of Voice between periods is meaningful only if it exceeds the variability that would be expected from re-running the same queries against the same platforms on the same day. Without a documented variability baseline, trend interpretation is guesswork.

This framework requires that measurement programs document the variability observed across repeated identical queries within a defined time window and make that variability visible in reporting. Buyers running their own programs should establish a variability baseline before acting on period-over-period change.

Specific quantitative thresholds for acceptable variability remain an open working group question, as appropriate ranges differ across metrics, platforms, and query types.

Geographic and Temporal Controls

AI platform behavior varies by geography and over time, and programs that do not control for either will mix signal with noise.

Geographic variation is produced by platform availability, model routing, regional content preferences, and localization of responses. A query issued from the U.S. and the same query issued from Germany will often return different brands, different sources, and different framing, even when the query is identical in language. Programs

covering multiple markets must disclose the geographic provenance of queries and report results per market rather than rolling geographies into a single score. Decision-grade cross-market comparison requires per-market measurement with query design adapted for local language and cultural framing.

Temporal variation is produced by platform updates, index refresh cycles, and diurnal or weekly patterns in model behavior. Programs should issue queries on a consistent cadence and at consistent times where possible to minimize the contribution of temporal noise to observed change. Where queries are distributed across a collection window, the window should be documented and held comparable across measurement periods.

Model Drift and Baseline Management

AI platforms update their models on timelines that are not always publicly announced, and the updates themselves can shift visibility results materially and immediately. A brand's Share of Voice can change five points overnight because a model was retrained, with no change to the brand itself and no change to its competitive position.

AI platforms also exhibit recency bias. Models weigh newer content more heavily, which means a brand or publisher's visibility can decline over a measurement period not because their competitive position changed, but because the model's training or retrieval favors more recent sources. Recency bias is a systematic form of drift that affects comparability over time and should be documented as a baseline variable, particularly when measuring evergreen content or tracking visibility across model retraining cycles.

Throughout this section, platform-driven refers to shifts caused by the AI platform itself: model updates, changes to citation or ranking behavior, changes to retrieval configuration, or changes in response format. Market-driven refers to shifts that reflect actual changes in brand or competitive activity such as a campaign launch, a new competitor, or a piece of content gaining traction. Telling these apart is the central task of any stable measurement program. Programs that fail to do so will report platform change as brand performance, and the resulting narrative to leadership will be wrong.

Drift Detection. Programs should monitor for abrupt shifts in visibility metrics. These are single-period changes large enough to fall outside the program's normal variability range. Before treating an abrupt shift as platform-driven, the operator should rule out brand activity, competitive activity, and changes to the measurement program's own query set as causes. The strongest signal that a shift is platform-driven rather than market-driven is its pattern across the category: A sudden change that appears for multiple brands on the same platform, or across the full category on that platform, almost always reflects model behavior rather than competitive dynamics.

Model Update Impact Assessment. When an abrupt shift is detected and brand and competitive activity have been ruled out, the program should:

1. Check published platform release notes and provider communications for an announced model update
2. Compare the affected platform against others in the program to determine whether the shift is isolated to one platform
3. Examine the shift across query types to see whether it is concentrated in particular intents or distributed evenly

Platform-specific shifts distributed across the category are almost always model updates. Brand-specific shifts isolated to particular query types are more likely to reflect real performance change.

Baseline Reset Protocols. Not every model update requires a baseline reset. Minor updates can be absorbed

within normal variability ranges. A reset is warranted when a major model version change is announced by the platform; when citation behavior, response format, or answer-surface presentation shifts in ways that affect metric calculation (including changes to citation placement, card formats, source drawers, default collapsed or expanded states, mobile layouts, or the introduction of sponsored modules, even if the underlying generated text remains stable); or when the query set itself is modified in ways that break comparability with prior periods. When a reset is performed, the program must document the triggering event, the date of the reset, and the nature of the change; re-run the program against the updated platform to establish a new baseline; and report pre- and post-reset data as separate series rather than stitching them into a continuous trend line. Trend metrics, including Visibility Momentum and Citation Decay Rate, are only interpretable when the baseline under them is stable or has been clearly re-baselined.

Is This Shift Platform-Driven or Market-Driven?

Platform-Driven Shift

Appears across multiple brands or the full category on a platform
Systematic model behavior

Triggered by model updates, retrieval configuration, or ranking behavior
Platform-initiated change

Sudden, abrupt shifts that fall outside normal variability ranges
Platform-specific anomaly

Market-Driven Shift

Isolated to specific brands or particular query types
Real performance change

Triggered by campaign launches, new competitors, or content traction
Brand-initiated change

Gradual or event-based changes linked to specific business activities
Competitive performance fluctuation

New Platform Entry Protocols

The set of AI platforms that matter for measurement is determined by consumer share. A platform earns inclusion in a measurement program when it represents enough consumer AI traffic in the target market to materially affect a brand's visibility picture. Below that threshold, it adds noise more than signal and may be better treated as a monitoring target than a core measurement surface. Consumer share is also the basis for weighting once a platform is included. A platform reaching 40% of consumer queries in a category contributes more to the visibility picture than one reaching 5%, and combined visibility scores should reflect that asymmetry rather than treating platforms as equal inputs.

Consumer share for these purposes refers to the relative frequency with which consumers in the target market actually use a platform for AI-driven discovery, typically expressed as active users, query volume, or session share within the AI category. Programs should rely on multiple sources where possible, including platform-reported metrics, third-party measurement, and first-party referral analytics, and should disclose the basis on which consumer share is being

assessed. The current AI usage data landscape is fragmentary and not directly comparable across platforms. The standard here is documented, consistent methodology across measurement periods rather than a fixed quantitative threshold.

The set of platforms meeting this threshold is not fixed. New platforms emerge, existing ones gain or lose consumer share, and programs that do not adapt will measure an increasingly incomplete picture of the discovery landscape. At the same time, adding a new platform to an active program without protocol will corrupt historical comparisons and produce trend data that mixes platform coverage changes with actual visibility changes.

This framework recommends that new platforms be added on a defined schedule rather than reactively, and that additions be handled as a baseline event. When a new platform is incorporated, the program should establish a new baseline that includes it, retain the prior baseline for historical comparison, and report results on both bases until enough time has passed on the new baseline to support trend analysis. Inclusion criteria, including the consumer share threshold the program uses, should be documented in advance, so that platform additions are governed by stated standards rather than ad hoc judgment.

Measurement Cadence Recommendations

Cadence should match the decision cycle the measurement is intended to inform, not exceed it. Measuring weekly for a decision made quarterly produces data more volatile than useful; measuring quarterly for a decision made weekly produces data already stale on arrival.

For trend monitoring and early signal detection, monthly measurement is typically sufficient and less susceptible to short-term noise. For competitive benchmarking and agency performance reporting, biweekly to weekly measurement provides the resolution required to detect meaningful change while remaining operationally feasible. For active budget allocation, provider selection, or executive strategy decisions, weekly or more frequent measurement is expected, consistent with the Criteria Matrix in Measurement Quality Standards. Programs covering platforms known to update frequently may need to measure more often simply to detect drift in a timely manner, independent of the decision cycle.

Communicating Measurement Uncertainty

The Executive Reporting Guidance section establishes the reporting principles: disclose the measurement tier, report trends rather than point-in-time snapshots, and attach variability ranges to point estimates. Applying those principles consistently is harder than it sounds, because the audience is usually accustomed to deterministic measurement. A marketing leader who has spent two decades reading Google search numbers as exact figures will read a range with a variability band as imprecise unless the report actively reframes what kind of number it is. The communication challenge is not just delivering the data but resetting the cognitive frame in which it is read.

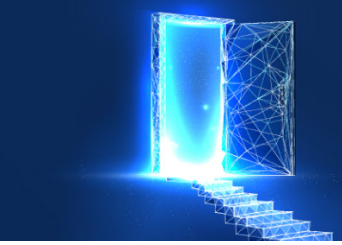
This subsection provides language patterns for the most common reporting situations. They are intended as starting points, not scripts.

Introducing the data for the first time. When AI visibility data is being reported to a stakeholder for the first time, the framing should establish that the underlying measurement is probabilistic. For example:

> *"AI platforms don't return the same answer every time, even to the same question. That's why we report ranges instead of single numbers. The range shows what normal variation looks like. Changes inside it aren't meaningful, changes outside it are."*

Reporting a platform-driven shift. When a visibility change coincides with a confirmed model update or a category-wide pattern, the report should attribute the cause directly and indicate whether a baseline reset is being performed.

> *"Visibility declined approximately six points this period. The decline coincides with a confirmed model update*



on [AI Platform], and a similar magnitude of change is visible across competitors in the category. This reads as a platform-driven shift rather than a brand-specific performance signal. We are treating this as a baseline reset event and will report pre- and post-update data as separate series until the new baseline stabilizes."

Reporting a drop that persists after a platform stabilizes. When a decline initially attributed to platform change does not recover after the platform settles, the narrative should explicitly shift its attribution.

"The visibility decline observed in [prior period] was initially attributable to a model update on [AI platform]. The platform has now stabilized, and the decline has held across two subsequent measurement periods. We are now reading this as a performance signal rather than platform noise and recommending [strategic review/further investigation]."

Reporting divergent platform results. When results diverge across platforms in ways that cannot be reconciled into a single combined story, the report should present the per-platform figures rather than defaulting to a combined score that obscures the divergence.

"Visibility on [platform A] increased four points this period; visibility on [platform B] declined three points. The divergence is not reconcilable into a single trend. We are reporting per-platform figures and treating each platform's trajectory separately for this period."

Clear attribution of cause is a core responsibility of the measurement program. Leadership decisions depend on knowing not just what changed but why, and the difference between "visibility is down" and "visibility is down because the platform updated and has not yet stabilized" is the difference between a strategic response and a non-response.

AI visibility measurement is a discipline still forming. The metrics, quality tiers, disclosures, and stability practices defined here are intended as a starting point, common ground the market currently lacks. Variability thresholds, query volume minimums, and the platform coverage required for decision-grade measurement will continue to evolve as the market matures. Multimodal visibility and paid placement measurement are flagged as priorities for future investigation, and downstream attribution will be addressed in IAB's forthcoming attribution framework. The intent is to give brands, publishers, agencies, and measurement providers a foundation to operate on now, as well as one the industry can build on as AI-powered discovery matures.

PROJECT™ EIDOS iab.

Measuring Visibility in the AI Era is part of Project Eidos, IAB's industry-wide initiative to define the principles, standards, and frameworks that guide how measurement is developed, evaluated, and compared across the industry.