

The Layer Mismatch

Why GEO Visibility Gains Do Not Translate to Decision-Stage Recommendation — and Why the Category Cannot Fix It

AIVO Research Intelligence Platform · AIVO Standard · April 2026

Recommended Citation

AIVO Research Intelligence Platform. (2026). *The Layer Mismatch: Why GEO Visibility Gains Do Not Translate to Decision-Stage Recommendation — and Why the Category Cannot Fix It*. Working Paper WP-2026-08. AIVO Standard. aivostandard.org. DOI: 10.5281/zenodo.19840293

ABSTRACT

This paper argues that the GEO and AEO category — as currently constituted and as represented by its leading platforms — is producing measurable improvements in the wrong evidence layer. The content it recommends and generates operates at the community and editorial retrieval layer of AI systems. The layer that determines purchase recommendation outcomes at the decision stage is the knowledge graph entity anchor layer: Wikidata entity definitions, Wikipedia category statements, trained-model entity representations, and the structured evidence architecture that AI models evaluate when applying criteria filters at Turn 3 of a buying sequence. These two layers are structurally independent. Improvements in the first layer do not propagate to the second. In documented cases, content produced by GEO programmes actively conflicts with legacy knowledge graph anchors, producing the pattern AIVO has documented consistently across 195+ brands: improved first-prompt visibility combined with unchanged or degraded decision-stage recommendation performance.

This paper introduces the term layer mismatch to describe this phenomenon and documents it through five brand case studies, each with AIVO Meridian probe data. The five brands — DocuSign, Akamai Technologies, Clarins, Chanel N°5, and TUI — span B2B SaaS, cloud infrastructure, prestige beauty, luxury fragrance, and European travel. Two are named customers of the leading GEO platforms. Three are not. In every case, the displacement mechanism identified by the AIVO probe methodology operates at the knowledge graph layer that GEO content programmes do not address. The paper documents the three conflict mechanisms through which GEO content can make this problem worse rather than better, the structural reasons the category cannot address the layer mismatch, and the methodology required to do so.

Keywords: layer mismatch, GEO, AEO, decision-stage recommendation, knowledge graph, Wikidata, entity anchor, Clinical Evidence Binary filter, brand inference position, CODA, Meridian, evidence architecture

1. INTRODUCTION

The Measurement Gap Within the Measurement Gap

The GEO and AEO category was founded on a correct insight: traditional search engine metrics do not measure brand performance in AI-generated responses. Citation frequency, share of AI voice, and first-prompt visibility are genuinely different from Google ranking signals, and they require different measurement approaches. The category has built sophisticated tooling to measure these signals and prescribe content interventions that improve them. The platforms in this category have raised hundreds of millions of dollars in venture funding, attracted enterprise clients across every major industry vertical, and established a category that did not exist two years ago. That achievement is real.

This paper argues that the category has accurately solved one measurement problem while creating another. The measurement it has built — first-prompt citation frequency and brand visibility in AI responses — is a correct measurement of something real. It is not a correct measurement of the commercial outcome brands believe they are purchasing when they invest in GEO visibility improvement.

The commercial outcome brands are purchasing — or believe they are purchasing — is AI purchase recommendation performance: the probability that when a buyer uses ChatGPT, Gemini, or Perplexity to

make a product or service decision, the brand wins the final recommendation. AIVO has been measuring this outcome since early 2025 using the CODA four-turn buying sequence protocol (WP-2026-01). The finding from 195+ brands and 7,000+ buying sequences is consistent: first-prompt citation frequency does not predict decision-stage recommendation outcomes. A brand can appear in AI responses consistently and comprehensively while recording zero purchase recommendation wins at the decision stage.

The reason for this disconnect is structural. AI models use different evidence sources for different cognitive tasks within the same buying conversation. At Turn 1 — the awareness and first-response stage — the model draws on the most recent, most cited, most accessible content in its retrieval corpus: community platforms, editorial coverage, brand-owned content, and the Layer 1 and Layer 2 sources that GEO content programmes produce. At Turn 3 — the criteria evaluation stage — the model applies a different evaluation function, drawing on the knowledge graph entity definitions, trained entity representations, and structured evidence architecture that constitute Layer 3. These two evidence layers are structurally independent. GEO content produced for Layer 1 and Layer 2 does not propagate to Layer 3. In some cases, as this paper will document, it actively conflicts with it.

The GEO category has solved the first-prompt visibility problem. It has not solved the decision-stage recommendation problem. The two are different problems at different evidence layers. Confusing them is not an analytical error. It is a structural property of the measurement infrastructure the category has built.

2. THE THREE EVIDENCE LAYERS

How AI Models Use Evidence Differently at Different Stages

Understanding the layer mismatch requires understanding how AI models evaluate brands at different points in a buying sequence. The three-layer model described here is an analytical framework derived from AIVO's probe corpus and the documented citation architecture of the major AI platforms. It is not a claim about the internal mechanics of any specific AI model; it is an observational framework derived from what the probe data consistently shows about when and why brands win or lose recommendations.

Layer 1 — Community and Editorial Retrieval

The first evidence layer is the one GEO content programmes are designed to populate. It consists of Reddit threads, LinkedIn posts, YouTube content, blog articles, on-site copy, press releases, and the editorial content that AI platforms retrieve from the open web in real time or that is represented in training corpora at high citation frequency.

This layer drives first-prompt citation outcomes. A brand that is actively discussed on Reddit, cited in recent editorial, mentioned in LinkedIn posts and YouTube reviews, and represented in high-frequency blog content will appear in AI responses to general awareness queries. This is a real and measurable effect. It is also the effect that GEO platforms are designed to produce, measure, and report.

The critical limitation of this layer is its volatility. Profound's own published research has documented that up to 90% of cited sources in AI answers change over time. The community and editorial layer is the most dynamic layer in the evidence stack and the least stable basis for decision-stage recommendation outcomes.

Layer 2 — Editorial Authority

The second evidence layer consists of Tier 1 press coverage in category-authoritative publications, analyst reports from Gartner, Forrester, IDC, and Mintel, industry awards and certifications, peer-reviewed publications, and regulatory body documentation. These sources carry higher authority weight in the model's reasoning chain than community content and are less volatile.

The GEO category acknowledges this layer and recommends pursuing it through PR, earned media, and analyst relations. However, it does not provide structured tooling for measuring whether Layer 2 evidence is satisfying the specific criteria filters the model applies at Turn 3, or for identifying which specific Layer 2 sources are required to pass which specific filters.

Layer 3 — Knowledge Graph Entity Anchors

The third evidence layer is the one the GEO category cannot address. It consists of Wikidata entity definitions, Wikipedia category statements, trained-model entity representations built from historically consistent evidence, and the structured knowledge graph that AI models use to define what a brand entity

is — what category it belongs to, what it is known for, what properties it possesses, and what criteria it can plausibly satisfy.

Layer 3 is the most stable layer and the most determinative of decision-stage outcomes. When a model applies a criteria filter at Turn 3 — asking which brand has established entity recognition in a specific category, which brand has documented evidence of clinical efficacy, which brand is the definitive leader in a market segment — it is evaluating brands against their Layer 3 entity definitions, not their Layer 1 citation frequency.

The GEO category has no tools, no methodology, and no structural ability to address Layer 3. Wikipedia explicitly prohibits paid editing and brand promotion. Wikidata requires verifiable third-party citations for entity claims. Both platforms have active editor communities that revert promotional changes rapidly. A SaaS company cannot reliably ship a "knowledge graph optimisation feature" because the platforms that constitute Layer 3 reject optimisation as a concept. The category has therefore concentrated on Layers 1 and 2 — the platforms where its tooling produces measurable, sellable outputs — and left Layer 3 entirely unaddressed.

3. WHAT THE GEO CATEGORY RECOMMENDS AND WHY

The Documented Absence of Layer 3 Work

The absence of Layer 3 work in the GEO category's recommendations is not an inference. It is documentable from the published content, product features, and prescription frameworks of the leading platforms.

The dominant GEO and AEO prescriptive advice across the category is overwhelmingly oriented toward community platforms (Reddit, Quora), social platforms (LinkedIn, YouTube), and editorial content. Otterly AI's citation architecture analysis found that community platforms capture 52.5% of AI citations versus 47.5% for brand domains, and the industry has correctly read that signal — but read it in only one direction. The prescription that follows is: produce content for the platforms where AI cites. It does not ask: why does the model cite those platforms at Turn 1 but not at Turn 3?

When the category does acknowledge Wikipedia and Wikidata as significant citation sources — which the more rigorous platforms do in their published research — the acknowledgement is taxonomic rather than prescriptive. Wikipedia matters is a measurement finding. Wikipedia work is not a recommended remediation. The distinction is critical and the gap it represents is the commercial opportunity this paper documents.

The structural reason for this absence is not negligence. It is the predictable consequence of three constraints that operate simultaneously on every company in the category. First, Wikipedia and Wikidata are inhospitable to commercial tooling by design — they reject optimisation as a concept and revert promotional changes. Second, the category's measurement infrastructure is built around citation frequency, which is measurable and improvable at Layer 1, not around decision-stage recommendation outcomes, which require a different measurement methodology. Third, content production is a recurring, scalable, billable activity that supports subscription revenue; knowledge graph remediation is a one-time or low-cadence intervention that does not generate the same commercial returns. These three constraints together produce a category that measures and prescribes at Layer 1 and Layer 2 while leaving Layer 3 untouched.

4. THE THREE CONFLICT MECHANISMS

How GEO Content Can Make the Problem Worse

The layer mismatch is not merely a failure to address Layer 3. In documented cases, GEO content programmes produce signals that actively conflict with existing Layer 3 evidence, creating a three-way contradiction that AI models resolve by reducing confidence in the brand at the decision stage. Three distinct mechanisms produce this conflict.

4.1 Entity Definition Conflict

The first mechanism operates when the brand's Layer 3 entity definition in the knowledge graph classifies the brand in one category, while the GEO-produced Layer 1 content positions the brand in a different or

expanded category. The model receives contradictory signals about what kind of entity the brand is. It resolves the conflict by using the more authoritative Layer 3 definition for criteria evaluation — producing category-specific recommendations at Turn 3 — while reflecting the newer Layer 1 content in first-prompt retrieval. The result is the pattern AIVO has documented for Akamai Technologies: visible and mentioned at Turn 1 in distributed cloud contexts, absent from the spontaneous consideration set at Turn 3 because the criteria filter falls back to the Layer 3 entity definition that classifies the brand as a CDN company.

The conflict is not resolved by producing more Layer 1 content in the new category. It is resolved by updating the Layer 3 entity anchor to reflect the brand's actual category position. Until that update propagates through the knowledge graph and into the model's trained representations, the Layer 1 content and the Layer 3 anchor are in active conflict, and the model expresses that conflict through inconsistent recommendation behaviour.

4.2 Claim Consistency Conflict

The second mechanism operates when GEO-produced content makes claims in marketing language that are inconsistent with the neutral or critical language of existing Layer 3 sources. AI models evaluate claim consistency across sources when determining confidence at Turn 3. A brand whose Reddit content asserts "industry-leading" performance while Wikipedia describes the brand without similar superlatives, and analyst reports describe it as one of several comparable options, produces a low-confidence inference at the criteria evaluation stage. The model expresses that low confidence through qualified recommendations, conditional endorsements, or by routing buyers to competitors whose claim architecture is more internally consistent.

The irony is that the most aggressive GEO content programmes — producing the highest volume of brand-positive community and editorial content — can produce the most severe claim consistency conflicts if that content overreaches relative to the established Layer 3 evidence base.

4.3 Temporal Weighting Conflict

The third mechanism is the most structurally embedded and the most difficult to remediate. ChatGPT and equivalent trained-knowledge models have been documented to continue citing Wikipedia articles that no longer exist — meaning trained entity definitions persist beyond the currency of the underlying source. The implication is that GEO content produced in 2025 and 2026 is being layered on top of trained knowledge from earlier training cutoffs, and the older trained knowledge takes precedence in entity reasoning while the newer content takes precedence in retrieval.

The two layers are not in dialogue. The model does not reconcile them — it uses each for a different cognitive task. At Turn 1, it retrieves from recent, high-frequency sources. At Turn 3, it reasons from trained entity definitions. GEO content affects the first task. It does not affect the second. A brand that has invested extensively in GEO content production over the past eighteen months has potentially improved its Turn 1 performance significantly while its Turn 3 performance remains anchored to trained entity definitions that predate the entire GEO category's existence.

5. FIVE BRAND CASE STUDIES

The Layer Mismatch in Practice Across Five Industries

The following five case studies are drawn from AIVO Meridian probe data collected in April 2026. All five represent brands whose displacement mechanism operates at Layer 3 — the knowledge graph layer that GEO content programmes do not address. Two are named clients of the leading GEO platforms; three are not. The finding is consistent across all five regardless of GEO platform relationship: the gap the model is identifying and the remediation required cannot be produced by the content programmes the GEO category prescribes.

5.1 DocuSign — Profound Client — Cause 2a: The Deficiency the Platform Cannot Detect

DocuSign is a named Profound enterprise client. Profound is the category-defining GEO platform, serving more than 700 enterprise clients including 10% of the Fortune 500. DocuSign represents one of Profound's most commercially significant client relationships.

AIVO WP-2026-07 documented that DocuSign's primary eSignature product page carries no JSON-LD structured data — a Cause 2a deficiency gap as classified in the four-cause diagnostic (WP-2026-04). The FAQ content addressing eight questions on legal enforceability, ESIGN Act compliance, Public Key

Infrastructure, and security standards is rendered as human-readable HTML but is not declared as structured data using schema.org FAQ markup. The Organisation entity, the Product entity, and the security certification claims are similarly undeclared in machine-readable format.

The AIVO Meridian probe confirms the commercial consequence. The T3 displacement criteria recorded verbatim:

“Which platform has documented compliance with enterprise security standards and verified legal enforceability across the competitive B2B eSignature landscape?”

DocuSign has the evidence that would satisfy this criteria filter. It possesses ESIGN Act compliance documentation, PKI certification, SOC 2 Type II compliance, and extensive legal enforceability documentation. That evidence exists. It is not structured in a format the model can reliably extract, attribute, and evaluate at Turn 3. This is not a Layer 1 problem. It is a Layer 3 structured data problem.

Profound's Agents generate comparison articles, Reddit engagement, FAQ content, and LinkedIn posts. None of this addresses the JSON-LD absence. A DocuSign marketing team using Profound to improve AI visibility is investing in Layer 1 and Layer 2 content while a closeable Layer 3 infrastructure deficiency remains undetected by the platform they are paying to address it. The deficiency is closeable in a single engineering sprint. It has persisted because the GEO category's tooling is content-shaped rather than infrastructure-shaped, and the measurement it provides does not surface infrastructure gaps.

5.2 Akamai Technologies — No GEO Platform Relationship — Cause 2c: The Entity Definition Conflict

Akamai Technologies has not been identified as a client of any GEO platform. It is included in this paper as the primary documented case of the entity definition conflict mechanism and as the brand whose situation most clearly illustrates what GEO content cannot address even in the absence of a GEO programme.

Akamai Technologies has undergone a deliberate, multi-year brand repositioning from CDN and edge delivery to distributed cloud computing, AI inference infrastructure, and cybersecurity under the Connected Cloud narrative launched in 2023. The content team has produced substantive documentation of this repositioning. The website content is accurate and well-structured. The strategic decision is commercially sound.

The Layer 3 problem is that Wikidata entity Q56338 and equivalent knowledge graph representations continue to categorise Akamai primarily as a content delivery network company, reflecting 25 years of consistent historical evidence in that category. The AIVO Meridian probe produced the following verbatim displacement criteria across multiple platforms:

“Which brand has established entity recognition and documented authority in the distributed compute platform category?”

And the probe finding:

“Brand not recognised as a valid entity in the distributed computing platform category.”

The probe also produced the paradox that makes this case study uniquely illustrative: other brands being evaluated in the same category are being displaced with the recommendation "Routed to: Akamai Technologies." The model routes other brands' buyers to Akamai because it recognises the entity as authoritative in infrastructure. But Akamai does not appear when buyers start fresh with a generic distributed compute query, because the entity anchor does not map to that category.

This pattern cannot be addressed by GEO content. Producing more LinkedIn posts, Reddit engagement, or blog content positioning Akamai in distributed cloud creates entity definition conflict with the Layer 3 anchor rather than resolving it. The remediation requires updating Wikidata entity Q56338, developing Wikipedia category statement documentation, and deploying brand.context files that explicitly declare the entity's current category membership — all Layer 3 interventions that the GEO category cannot deliver.

5.3 Clarins — No GEO Platform Relationship — Cause 1: The Clinical Evidence Binary Filter

Clarins is a prestige French beauty brand with a 70-year heritage, a global distribution network, and extensive digital content production. It was among the first brands to be probed through AIVO's methodology and the Clarins Double Serum finding — elimination at the purchase decision stage despite strong awareness and brand equity — established the foundational case for what AIVO subsequently named the AIVO Paradox.

The AIVO Meridian probe for Clarins documents displacement driven by the Clinical Evidence Binary filter. At Turn 3 of a skincare buying sequence, the model asks:

“Which brand has documented ingredient efficacy, clinical validation, and dermatologist endorsement in a format the model can extract and evaluate?”

Clarins has 70 years of formulation research, proprietary plant extracts, and a dedicated research and development programme. The clinical evidence exists. It is not structured, published, and anchored to the Clarins entity in the form the model requires at the criteria evaluation stage. CeraVe, which dominates the beauty category in AIVO's probe corpus, wins this filter consistently because its entire brand identity is built around dermatologist-developed clinical evidence published in a format AI models can extract and evaluate. Clarins' brand identity is built around luxury, heritage, and plant-based science — none of which maps cleanly onto the Clinical Evidence Binary filter's criteria language.

This is not a GEO content problem. A Clarins marketing team deploying Profound or Peec AI would receive recommendations to produce Reddit skincare community content, LinkedIn beauty editorial, and blog articles. None of that produces the structured clinical evidence architecture that satisfies the T3 filter. The remediation requires publishing clinical validation data in structured, citable, AI-extractable formats — Zenodo-deposited studies, dermatologist endorsement records in schema.org format, ingredient efficacy documentation anchored to Wikidata substance entities. That is Layer 3 remediation. It is available to Clarins. It is not available through any GEO platform.

5.4 Chanel N°5 — Peec AI Client — Cause 2c: The Knowledge Graph Anchor Gap for the World's Most Famous Fragrance

Chanel is a named Peec AI client. Chanel N°5 is arguably the most culturally recognised fragrance product in the world, in continuous production since 1921, embedded in 103 years of popular culture, fashion, film, and global advertising. If brand awareness, cultural resonance, and content volume predicted AI purchase recommendation outcomes, Chanel N°5 would win every fragrance recommendation on every platform without intervention.

The AIVO Meridian probe produced five filter gaps. The T0 finding on Gemini:

“Brand lacks Wikipedia/Wikidata anchor or comparable knowledge graph presence for luxury perfume category.”

The most famous fragrance in the world is absent from Gemini's spontaneous consideration set for luxury perfume because its knowledge graph entity anchor is not correctly structured for the fragrance category. One hundred and three years of cultural recognition have not produced the specific structured entity representation that Gemini's evaluation architecture requires.

The T2 finding on ChatGPT is equally striking. The model routes buyers from Chanel N°5 to Chanel Coco Mademoiselle — a sibling product within the same brand family. The verbatim criteria:

“Which luxury perfume has documented evidence across all four evaluation axes: proven effectiveness, overall value, established reputation, and widespread accessibility?”

Coco Mademoiselle satisfies this four-axis evaluation criteria more completely than N°5 in the model's evidence assessment. The Close Second Trap is operating within a single brand's portfolio. The world's most iconic fragrance is losing the recommendation to a newer sibling product because Coco Mademoiselle's evidence architecture — with its extensive accessibility documentation, value positioning, and contemporary effectiveness citations — satisfies the model's comparative framework more completely.

The T7 finding on Perplexity introduces a third distinct gap type: displacement by a retailer at the purchase pathway stage. The verbatim criteria:

“Which retailer provides the most immediate and convenient purchase pathway for this specific product and gift timing?”

Finding: Macy's immediate availability messaging displaces Chanel's direct channels. This mirrors the Expedia finding in WP-2026-07 — commerce infrastructure gaps that no content programme can address.

Chanel is paying for Peec AI visibility monitoring. Peec AI is telling Chanel how often Chanel N°5 appears in AI responses. It is not telling Chanel that the brand lacks a correctly structured knowledge graph anchor for the luxury perfume category, that Coco Mademoiselle is winning the ChatGPT recommendation at the expense of N°5, or that Macy's is displacing the brand at the purchase pathway stage. Those findings require decision-stage probe methodology. They are not visible through citation frequency monitoring.

5.5 TUI — Peec AI Client — A New Gap Type: The Definitiveness Failure

TUI Group is a major European travel operator and a named Peec AI client. TUI operates in a category where AIVO's probe corpus has documented the spontaneous consideration gap pattern, and the TUI probe introduces a fifth distinct gap type that none of the preceding four cases fully illustrates.

The TUI probe finding is distinctive because the brand is not being displaced by specific competitors in the traditional sense. The model routes buyers to TUI repeatedly across multiple filter gaps — T2, T5, and T6 all show "Routed to: TUI." The model knows TUI and uses it as a routing destination. The problem is that TUI cannot convert that routing into a definitive purchase recommendation. The T2 verbatim displacement criteria:

"Which vacation package type delivers the optimal balance of convenience, value, and proven effectiveness across multiple traveler priorities?"

Finding: "TUI lacks strong positioning as the definitive all-inclusive provider versus being one of many operators in the space."

This is not displacement by a competitor. It is a definitiveness failure — the model cannot identify TUI as the authoritative answer to the question because the evidence does not position TUI as the category-defining provider. TUI is known. TUI is present. TUI is even preferred. But TUI cannot be definitively recommended because the Layer 3 evidence that would establish definitive category leadership — ATOL protection prominence, ABTA certification visibility, scale documentation, verified customer outcome evidence — is either absent or not structured in a form the model can use to close the recommendation decisively.

The T7 finding introduces the same commerce infrastructure gap seen in Chanel N°5 and Expedia: "No direct booking URL or immediate availability confirmation provided for the brand." The T1 finding adds a credentials gap: "Lack of specific accreditation or operational credentials mentioned for the brand" — the regulatory certifications TUI holds but has not structured as Layer 3 evidence.

TUI is monitoring its AI visibility through Peec AI. The visibility monitoring tells TUI how often it appears. It does not tell TUI that the definitiveness failure is a Layer 3 positioning problem, that the regulatory credentials gap is an entity anchor issue, or that the commerce infrastructure gap is the same structural problem that the Expedia case documents in detail. These are decision-stage findings that citation frequency monitoring cannot surface.

6. WHY THE CATEGORY CANNOT FIX THIS

Structural Constraints, Not Company Failures

This paper does not argue that the GEO category is doing something wrong. Profound and Peec AI are building what their tooling enables, measuring what their measurement infrastructure makes visible, and prescribing what their business model incentivises. The layer mismatch is not a failing of specific companies. It is a structural property of the category produced by the intersection of three constraints that no company in the category can individually resolve.

6.1 Platform Inhospitability

Wikipedia explicitly prohibits paid editing and brand promotion under its conflict of interest guidelines. Wikidata requires verifiable, independently sourced citations for entity claims. Both platforms have established editorial communities that detect and revert promotional changes with speed and consistency. Schema.org structured data can be deployed on brand-owned properties, but the knowledge graph signals that AI models weight most heavily in entity definition — Wikidata entity statements, Wikipedia category classifications, cross-referenced authoritative sources — are controlled by third-party editorial communities rather than by brands or their agencies.

A GEO SaaS company cannot ship a "Wikipedia optimisation feature" because Wikipedia would reject the outputs. It cannot ship a "Wikidata entity management tool" because Wikidata's citation requirements are incompatible with brand-controlled content workflows. The GEO category has therefore built its tooling on the platforms that are genuinely accessible: community platforms, editorial networks, brand-owned domains. That is the rational response to the constraint. The consequence is a category that cannot address the layer that matters most.

6.2 Measurement Infrastructure Incentives

The GEO category measures citation frequency because citation frequency is measurable, improvable on a short timeline, and reportable in dashboard form. Decision-stage recommendation outcomes require a structured buying sequence probe methodology — a four-turn conversational protocol that simulates the complete purchasing conversation and classifies displacement at the turn level. This is AIVO's methodological contribution (WP-2026-01). It is not a measurement that can be approximated by citation counting.

The consequence of measurement infrastructure shaping prescription is significant. The GEO category recommends producing content that improves the metric it can measure. If the metric is citation frequency at Turn 1, the prescription is to produce content that is cited at Turn 1. The question of whether that content improves Turn 3 performance is not asked because it cannot be answered by the measurement infrastructure the category has built. The invisible metric — decision-stage recommendation win rate — is therefore also the unmeasured and unaddressed metric.

6.3 The Content Production Business Model

GEO content agents, prescriptive platforms, and the wider category's workflow tools are oriented toward content production because content production is a recurring, scalable, billable activity. An agency charging for monthly AEO content production, a SaaS platform charging per content brief generated, and a platform whose agent feature produces LinkedIn posts and Reddit contributions are all generating value through content volume. Knowledge graph remediation — updating Wikidata, developing Wikipedia articles, structuring brand.context files, commissioning clinical research to establish evidence architecture — is a one-time or low-cadence intervention that does not generate the same subscription revenue, platform engagement, or agency billing volume. The business model produces a content-first orientation that is not neutral with respect to where the intervention should actually occur.

7. THE REQUIRED METHODOLOGY

What Addressing the Layer Mismatch Requires

The argument of this paper implies a specific methodology. If the layer mismatch is a structural property of the GEO category's evidence layer orientation, then addressing it requires a methodology that operates at a different layer. The AIVO approach, documented across the working paper series, provides that methodology.

7.1 Decision-Stage Measurement

The foundational requirement is measuring the right metric. The CODA four-turn buying sequence protocol (WP-2026-01) measures brand performance at the purchase recommendation stage — Turn 4 of a structured buying sequence — rather than at the first-prompt citation stage. The displacement criteria recorded verbatim at Turn 3, the competitor that captured the recommendation, and the Revenue at Risk calculation derived from the finding together constitute a measurement that connects to the commercial outcome brands are purchasing when they invest in AI visibility improvement. Without this measurement, the layer mismatch is invisible.

7.2 Four-Cause Gap Classification

The second requirement is classifying why a gap exists before prescribing how to fix it. WP-2026-04 introduced the four-cause diagnostic: evidence missing (Cause 1), evidence invisible (Cause 2), evidence misframed (Cause 3), and counter-evidence absorbed (Cause 4). WP-2026-07 further refined Cause 2 into deficiency gaps, trade-off gaps, and propagation gaps. This classification determines which layer the remediation should target and what intervention is appropriate. A Cause 2c propagation gap — the Akamai and Chanel N°5 pattern — requires Layer 3 knowledge graph remediation. A Cause 1 gap — missing clinical evidence in the beauty category — requires evidence creation at Layer 2. A Cause 2a deficiency gap — the DocuSign pattern — requires structured data deployment at Layer 3. The GEO category applies a uniform content production prescription regardless of cause classification, which is why it produces improvements in some cases and no improvement in others.

7.3 Layer 3 Remediation

The third requirement is the remediation itself: the interventions that address the knowledge graph layer rather than the content layer. These include Wikidata entity statement updates with verifiable third-party citations; Wikipedia category statement development and article improvement; brand.context file deployment that provides machine-readable brand entity declarations; structured clinical evidence

publication to Zenodo and equivalent repositories; JSON-LD schema markup deployment for entity, product, and claim declarations; and DOI-assigned working papers that establish the brand's research and evidence authority in the knowledge graph. These are not content production activities. They are evidence architecture activities. They require a different skill set, a different workflow, and a different commercial model than the GEO category has built.

7.4 Continuous Monitoring

The fourth requirement is continuous monitoring of the evidence architecture layer rather than periodic content audits. Knowledge graph entity definitions drift as editors update category classifications and brands evolve their positioning. Structured data declarations degrade as CMS migrations break asset URLs and schema markup falls out of synchronisation with page content. Clinical evidence citations age and lose authority weight as newer research supersedes them. The AIVO Orbit monitoring system (WP-2026-04) is designed to provide continuous monitoring of these Layer 3 dynamics — surfacing drift, detecting gaps, and triggering remediation at the cadence the problem requires rather than on periodic audit cycles.

8. CONCLUSION

The Category Has Solved the Wrong Problem

The GEO and AEO category has built genuinely useful tooling for a real problem. First-prompt citation frequency is a real signal. Brand visibility in AI responses is commercially significant. The platforms that measure and improve these signals are providing real value to their clients.

This paper argues that the signal they are measuring and improving is not the signal connected to the commercial outcome their clients believe they are purchasing. The commercial outcome — AI purchase recommendation performance at the decision stage — is determined at Layer 3: the knowledge graph entity anchors, trained entity definitions, and structured evidence architecture that the GEO category cannot measure, prescribe for, or remediate.

The five case studies documented in this paper — DocuSign, Akamai Technologies, Clarins, Chanel N°5, and TUI — share a single finding across five different industries and five distinct gap types: the displacement mechanism operates at the layer the GEO category cannot reach. Two of these brands are paying clients of leading GEO platforms. Three are not. The GEO platform relationship is irrelevant to the finding. The layer mismatch exists in all five cases regardless of whether a GEO programme is active, because the displacement mechanism is at Layer 3 and the GEO category operates at Layer 1 and Layer 2.

The remediation for all five cases requires Layer 3 intervention — different in specific mechanism and timeline, but consistent in the layer it must address. DocuSign closes a structured data deficiency in a single engineering sprint. Akamai requires a 12–24 month knowledge graph propagation programme. Clarins requires clinical evidence publication in AI-extractable formats. Chanel N°5 requires knowledge graph entity anchor reconstruction for a category that has existed for a century. TUI requires credentials and definitiveness evidence structured at the entity layer. The mechanisms differ. The layer is the same.

The brands that understand this distinction before their competitors do will have a structural advantage that no amount of GEO content spend can replicate. AI purchase recommendation performance is not a content problem. It is a Layer 3 evidence architecture problem that requires a different methodology, a different measurement, and a different commercial model than the GEO category has built. The window to establish that advantage is open now, before Layer 3 remediation becomes as widely understood as Layer 1 content optimisation is today.

The layer mismatch is not a gap that more GEO content can close. It is a gap that the wrong kind of content has been trying to close, in some cases making it wider. The solution begins not with producing more — but with understanding which layer the problem lives in and delivering the intervention the layer requires.

REFERENCES

AIVO Research Intelligence Platform. (2026). CODA: Conversational Outcome and Decision Analysis — A Four-Turn Probe Methodology for AI Purchase Recommendation Measurement. Working Paper WP-2026-01. Zenodo. <https://doi.org/10.5281/zenodo.19401584>

AIVO Research Intelligence Platform. (2026). AIVO Orbit: Continuous Evidence Intelligence for AI Decision-Stage Remediation. Working Paper WP-2026-04. AIVO Standard. aivostandard.org.

de Rosen, T. (2026). The Upstream Measurement Gap in ChatGPT Advertising. Working Paper WP-2026-03. SSRN. <https://doi.org/10.2139/ssrn.6606518>

AIVO Research Intelligence Platform. (2026). The Evidence Invisibility Problem: Distinguishing Deficiency Gaps from Trade-Off Gaps in Enterprise AI Content Infrastructure. Working Paper WP-2026-07 (Revised). AIVO Standard. DOI: 10.5281/zenodo.19811294

Head, B. (2026). Blackhat AEO: The Tricks Being Used to Game LLMs Right Now. Digiconomy. April 2026.

Yesilyurt, M. (2025). Reverse-Engineering Perplexity's Source Selection Architecture. Search Engine Land. December 2025.

Profound. (2026). Series C Announcement and Platform Overview. GlobeNewswire. February 2026.

Peec AI. (2025). Series A Announcement. Peec AI Blog. November 2025.

Otterly AI. (2026). Citation Architecture Analysis: Community Platforms vs. Brand Domains. Otterly Research. 2026.

Wikidata. (2026). Akamai Technologies (Q56338). Wikidata entity record. wikidata.org/wiki/Q56338.

DOCUMENT METADATA FOR ZENODO DEPOSIT

Title	The Layer Mismatch: Why GEO Visibility Gains Do Not Translate to Decision-Stage Recommendation — and Why the Category Cannot Fix It
Authors	AIVO Research Intelligence Platform
Publication date	April 2026
Version	1.0
Resource type	Working paper / Preprint
Access type	Open access
Licence	Creative Commons Attribution 4.0 International (CC BY 4.0)
Keywords	Layer mismatch; GEO; AEO; decision-stage recommendation; knowledge graph; Wikidata; entity anchor; Clinical Evidence Binary filter; brand inference position; CODA; Meridian; evidence architecture; DocuSign; Akamai; Clarins; Chanel; TUI
Related identifiers	WP-2026-01: DOI 10.5281/zenodo.19401584 · WP-2026-03: DOI 10.2139/ssrn.6606518 · WP-2026-04: aivostandard.org · WP-2026-07: DOI 10.5281/zenodo.19811294 · aivomeridian.com · aivostandard.org

