Answer Engine Optimization

How Agentic AI Reshapes SEO

Miles Bliey & Keira Chatwin

GSBGEN 390

July 12, 2026

Abstract

For nearly three decades, search engines have acted as intermediaries between users and websites. That architecture is now evolving. Generative AI platforms increasingly answer queries directly rather than routing users to external destinations, reducing the centrality of the click as the dominant unit of value in digital discovery (share of total search interactions has grown from under 10% in 2023 to approximately 30% as of early 2026 (First Page Sage)).

This paper examines Agentic Engine Optimization (AEO), the emerging discipline of ensuring that brands are included and favorably positioned within AI-generated responses. We argue that AEO represents a more consequential inflection point than traditional Search Engine Optimization (SEO) because it alters the economic logic of visibility itself. In an agentic environment, visibility depends less on domain authority, paid placement, or broad familiarity and more on whether a brand survives retrieval, grounding, and cross-source validation inside large language model systems.

The paper proceeds in two parts. First, we identify current best practices by examining the mechanics of retrieval-augmented generation, including passage-level retrieval, query fan-out, pairwise reranking, and entity validation. Second, we analyze the strategic consequences of these mechanisms over the next two to five years. We examine the declining marginal effectiveness of performance advertising, the structural advantage of specialized brands over generalist incumbents, the emergence of agentic transaction layers, and the measurement problems created by non-deterministic and personalized outputs.

We argue that the relevant visibility metric in this environment is no longer static rank, but expected rank across the distribution of users. Firms that operationalize this shift early will hold a durable advantage as AI-mediated discovery becomes the dominant mode of consumer search.

1. The Rise of Answer Engines

Google's PageRank algorithm, introduced in 1998, established a durable compact between search engines, content creators, and users. Search engines would organize the world's information; content creators would produce it; users would navigate between the two through a ranked list of hyperlinks. For more than two decades, this architecture defined the economics of digital marketing. Brands invested heavily in Search Engine Optimization to secure favorable positions on that list, and the digital advertising ecosystem (now worth over $600 billion annually) was built on the premise that the search engine was a passthrough ("Digital Ad Spend"). Users began their journey on Google and ended somewhere else. Value was created at the click.

With the rise of agentic search, that premise is eroding. Generative AI platforms increasingly answer queries directly rather than sending users onward to external sites. ChatGPT processes over two billion queries per day and has roughly 800 million weekly active users (Chatterji et. al.). Google's Gemini has surpassed 750 million monthly active users. Perplexity, Claude, and a widening set of specialized AI assistants are fragmenting the search landscape. By early 2026, AI-mediated search accounted for roughly 30 percent of search interactions, up from a marginal share only a few years earlier (First Page Sage). The search market is no longer converging around a single gateway. It is fragmenting into a small number of general-purpose platforms alongside specialized assistants.

The significance of this shift extends well beyond interface design. It changes the fundamental unit of value in information discovery. Under the traditional search regime, value was created when a user clicked through to a destination website. Under the emerging regime, value is created when an AI system synthesizes information across sources, evaluates credibility, and produces a recommendation that may satisfy the user’s need without any outbound visit at all. The World Economic Forum described this transition plainly: consumers are no longer typing "ballpoint pens" into a search bar; they are asking an AI assistant for "the best pen if I am left-handed and write in cursive." Search engines no longer merely point toward information. Increasingly, they deliver it.

As the marginal cost of retrieving and synthesizing information approaches zero, access to information is becoming less scarce. The new scarcity is trust in the synthesis: confidence that the recommendation reflects a broad, credible consensus rather than the bias of a single source or a manipulated signal. This shift changes not just how users search, but how brands are discovered, evaluated, and selected. In traditional SEO, firms competed for

a visible position on a common results page. In this new agentic regime, firms increasingly compete for probabilistic inclusion across personalized retrieval environments.

Terminology

The terminology in this space remains unsettled. Terms such as Generative Engine Optimization (GEO), Agentic Engine Optimization (AEO), Generative AI Optimization (GAIO), and Large Language Model Optimization (LLMO) are often used interchangeably.

This paper uses the term Agentic Engine Optimization for two reasons. First, "agentic" captures the direction of technological development. Firms are no longer optimizing only for systems that generate answers, but increasingly for autonomous agents that will act on users' behalf—researching options, comparing alternatives, and initiating transactions.

Second, the term foregrounds the strategic stakes in a way that "generative" does not. When an AI system functions as a qualified intermediary, its recommendation carries unusually high economic value. By the time a user clicks through to a brand's site from an AI recommendation, the decision is substantially made. That click is not exploratory; it is confirmatory. This dynamic, sometimes called the "Perfect Click," implies that a single AI-mediated recommendation may be worth more, in conversion terms, than hundreds of unqualified organic impressions (Barnard). AEO is not a marginal improvement on SEO. It represents a structural change in how brands acquire customers.

2. The Logic of the Answer Engine

2.1. The Algorithmic Trinity

Jason Barnard of Kalicube describes modern answer engines as relying on the "Algorithmic Trinity" composed of three interdependent systems. The framework is useful because it clarifies why visibility in one system is insufficient on its own.

- **Large Language Models**: the synthesis layer. LLMs do not independently verify claims. Instead, they reason over the passages assembled in their context window and develop them into a coherent response.
- **Search Engines**: the retrieval layer. After a user submits a query, the engine surfaces candidate documents and passages through retrieval calls. The search engine does not generate answers but rather identifies the most relevant raw material to ground the answers.
- **Knowledge Graphs:** the validation layer. The Knowledge Graph is a structured database of entity nodes and relationships providing information about what entities exist, their attributes, and their relations.

For a brand to appear reliably in AI-generated answers, it must be legible across all three systems of the Algorithmic Trinity. Its content must be retrievable, its passages must be extractable, and the brand itself must exist as a recognizable entity whose claims can be validated against external signals. This creates two parallel imperatives: first, ensuring that your content is technically accessible and parsable (LLM readability), and second, ensuring that your brand identity is accurately and consistently represented across the ecosystem (brand context).

2.2. The Retrieval Pipeline

The process for an LLM-generated response cannot be specified with full precision as it is largely proprietary, only partially stochastic, and varies across platforms. Despite that, patent filings from Google and Microsoft, platform documentation (from Google's Vertex AI and Cloud Search APIs), and architectural deconstructions by leading AEO researchers such as Michael King of iPullRank and Olaf Kopp have revealed a broadly consistent structure:

1. **Query Understanding:** The process begins with query decomposition into a semantic interpretation. Transformer-based embeddings produce multiple vector representations of the query—capturing intent, entity references, and task type. The

platform often also incorporates contextual user signals such as prior behavior, device state, or search history. This has a strong implication. Namely, that the same query submitted by different users may trigger different retrieval paths.

2. **Query Fan-out:** Rather than executing a single lookup, answer engines typically decompose a query into multiple synthetic sub-queries that represent different facets of user intent (King). Google's patents depict LLMs analyzing the query alongside initial grounding results to generate diverse reformulations. A prompt such as "best running shoe" may trigger parallel retrieval around comfort, durability, price, injury prevention, and review quality. Each sub-query initiates a separate retrieval event.

3. **Retrieval and Reranking:** Each sub-query returns a set of candidate passages (generally 20-50 per sub-query) that are reranked by a cross-encoder that iteratively scores each passage against the original query until a complete ranking emerges. (From Google's patents and leading research, this likely occurs through a pairwise heapsort algorithm (Qin et al.)) This makes ranking inherently comparative rather than absolute.

4. **Grounding and Validation:** The top-ranked passages are then assessed for factual consistency and extractability. In this step, LLMs attempt to measure how well an answer candidate aligns with the retrieved facts, evaluated at both the response level and the individual claim level. Multiple specialized models may operate sequentially: one condensing key ideas from overlapping passages, one comparing products or features, one extracting structured information like prices or specifications, and one validating factual consistency against Knowledge Graph data. The implication here is also strong: retrieval alone is not enough. Content must also be modular, attributable, and independently defensible.

5. **Response Generation:** The validated passages are assembled into the context window, and the LLM generates a synthesized response with attributions to grounding sources. The output is probabilistic rather than deterministic. It varies with retrieved passages, user context, platform design, and stochastic generation properties.

Three structural differences between this pipeline and traditional SEO merit emphasis. First, the search space is dramatically expanded. Query fan-out transforms a single search into multiple retrieval events, enlarging the effective competitive space. Brands now compete across a distributed set of latent sub-questions, many of which are never directly visible to the marketer. A brand that ranks well for one sub-query but does poorly on the

other may not be seen by the LLM. Second, the unit of retrieval is the passage, not the page. Documents are chunked into semantically coherent segments of roughly 256–512 tokens, each independently embedded and indexed. A long article does not compete as a whole; it competes as a series of independent passages. Third, ranking is relative and adversarial: pairwise comparison means every passage is judged not on absolute quality but against specific competitors. These changes alter the economics of visibility. Traditional SEO rewarded domain authority and broad keyword coverage. Answer engines reward local information density, extractability, and alignment with highly specific intent.

3. AEO Rules of Visibility

If Section 2 explains how answer engines generate responses, this section examines the practical implications for firms: what must a brand do to be selected, validated, and ultimately recommended? Two constraints govern visibility in answer-engine environments. First, content must survive retrieval. It must be structured so that systems can locate, isolate, and extract useful passages. Second, brands must survive validation. Even highly retrievable content may be excluded if the entity behind it lacks sufficient credibility, coherence, or corroboration across the broader information ecosystem.

3.1. Retrieval: Being Selected

LLM readability is related to but distinct from human readability (Kopp). Human readers benefit from narrative arc, contextual buildup, and rhetorical structure. On the other hand, LLMs reward clarity, explicit attribution, and logical segmentation. This shift in readability priorities has a significant impact on site visibility: Princeton GEO study found that fluency optimization alone improved passage visibility by 15–30% across query types, suggesting that how content is written can shape retrieval outcomes (Aggarwal, Pranjal, et al).

The shift toward passage-level retrieval in RAG pipelines and AI search systems amplifies these effects. Since models retrieve and rank at the passage level rather than the page level, content must be locally coherent and self-contained. A well-structured document with a poorly structured passage would still underperform a less-structured document with a few effective paragraphs. Three principles follow directly from this architecture: self-containment, semantic density, and chunk relevance. These principles collectively determine whether a passage survives retrieval and ranking.

Self Containment

Each section or block of data should be semantically self-contained (meaningful and useful when extracted in isolation from the surrounding document). A passage that begins with "As mentioned above" or "Building on the previous point" has introduced a dependency which may not survive the chunking process. If chunked and retrieved in isolation, such passages often become incomplete or ambiguous. Therefore, this dependency may deprioritize it when retrieved by LLMs.

Dense Semantic Encoding

Dense semantic encoding is the principle that every retrievable unit should carry maximum informational value in minimum token space. HTML tables, structured data markup, and concise factual statements act as forcing functions for information density, allowing models to extract structured facts while remaining token-efficient (Blyskal). Google's documentation confirms that snippet extractability (whether a passage can be "lifted cleanly" into a synthesized answer) is a primary filter during the aggregation phase.

Chunk Relevance

Chunk relevance refers to how well a passage serves the specific sub-query likely to retrieve it. In practice, this means headers phrased as questions aligned with anticipated query patterns, section boundaries drawn at logical chunk points, and statistics and citations embedded within passages rather than collected in footnotes. Sentence structures should be direct; each passage should resolve a question rather than advance a narrative.

Empirical evidence on chunk relevance has been substantiated. The Princeton-Georgia Tech study tested nine optimization strategies across 10,000 queries. It found that adding relevant statistics improved visibility by up to 40%—the single most effective intervention. Citing credible sources and including quotations from recognized authorities were the second- and third-most effective strategies. Interestingly, keyword stuffing, a canonical SEO tactic, showed negligible improvement.

These findings are consistent with the pairwise ranking architecture described previously. In a head-to-head comparison, the LLM is reasoning about which passage is more useful. A passage that provides a verifiable data point, a named source, or a concrete example offers something the competing passage may not (what practitioners call "information gain”). A passage that restates widely available information in a slightly different language offers no marginal value. In retrieval environments, substance dominates stylistic variation.

Bottom-Up Funnel Approach

Perhaps the most counterintuitive finding in AEO is that brands should begin optimization at the bottom of the funnel rather than the top. Traditional SEO rewarded top-of-funnel investment, such as broad keyword pages designed to capture awareness and drive volume. AEO inverts this logic.

The structural reason is query decomposition. When a user poses a highly specific question to an AI agent, the engine breaks that query into sub-queries targeting each

attribute. A brand that has produced content addressing exactly that intersection of features will dominate retrieval for that query, regardless of its overall domain authority. In the vector space of AI retrieval, brands compete on proximity to the specific query embedding, not on aggregate reputation.

This creates a structural advantage for smaller, specialized players, a dynamic that Mike King of iPullRank frames as the core of "relevance engineering." A niche brand producing authoritative, data-rich content for a narrow domain achieves higher vector-space proximity than a generalist incumbent whose content is broad but shallow. Large brands' legacy SEO libraries, optimized for high-volume keywords, often produce few passages that survive chunking and pairwise comparison for specific long-tail queries. Their content covers topics without addressing the precise attribute intersections that agentic queries target.

3.2. Credibility and Consensus

While LLM readability optimizes getting individual passages into the context window, the deeper challenge is ensuring a brand gets validated as a credible entity when recommendations are surfaced. This side of AEO is brand optimization: a set of strategies aimed at building and maintaining entity recognition, credibility signals, and retrieval metrics across the Algorithmic Trinity. Content optimization makes information citable. Brand optimization makes the underlying entity recognizable, trustworthy, and verifiable across the broader information ecosystem. If retrieval defines what can be surfaced, validation determines who can be believed.

The Credibility Framework: from EEAT → NEEATT

Google's E-E-A-T framework (Experience, Expertise, Authoritativeness, Trustworthiness) has served as the standard for evaluating search quality since 2018. With the rise of agentic search, these signals remain relevant but face a structural limitation: machines cannot assess credibility without an identity. If the engine cannot identify who you are, it cannot apply any credibility signal to what you say. Credibility, in this sense, requires a prior step: understandability. Understandability refers to the machine's ability to parse your identity and qualifications.

This is the logic behind the NEEATT framework developed by Jason Barnard, which extends E-E-A-T with two additional dimensions. The six components are:

- **Notability**: The degree to which a brand is recognized across third-party sources. In practice, notability functions as a multiplier: the engine will weight an expert's

claim substantially more if that expert is frequently cited across disparate, high-authority datasets. Importantly, notability is comparative rather than absolute (Barnard). A niche brand can achieve higher notability within a narrow domain than a large company that is diffusely present across many.

- **Experience**: Direct evidence of real-world involvement. LLMs use this signal to differentiate between human-verified knowledge and synthetic or secondhand content. Case studies, firsthand accounts, and practitioner data carry weight that generic summaries do not.
- **Expertise**: The technical depth of the content itself. Deep, niche expertise allows smaller brands to win relevance calculations against broad, shallow incumbents — particularly during pairwise comparison, where a passage demonstrating genuine mastery will defeat a superficial treatment of the same topic.
- **Authoritativeness**: Third-party validation—citations, endorsements, and references from independent, trusted sources. AI agents rely on consensus, meaning multiple independent sources that agree on a brand's role and claims will be treated by the LLM as established premises during synthesis.
- **Trustworthiness**: The historical reliability of the source. A high trust score prevents content from being filtered during the grounding and Knowledge Graph reconciliation phases. Inconsistent claims or contradictory data erode this signal over time.
- **Transparency**: Clarity regarding ownership, authorship, and intent. If the engine cannot parse who owns the content and what entity it represents, it cannot map the content to the Knowledge Graph or apply any other credibility signal. Structured data, consistent entity descriptions, and machine-readable markup end up being prerequisites for credibility evaluation to occur at all.

The Entity Home

Jason Barnard refers to the “Entity Home”: a single, authoritative brand-owned digital property where the Algorithmic Trinity can find definitive facts about a brand’s identity, purpose, and activities. In practice, this is typically the brand's primary website (specifically, its homepage or a dedicated "About" page), but the concept is more precise than "having a website." The Entity Home is the canonical reference point against which the Knowledge Graph corroborates information from third-party sources.

Practically, building a functioning Entity Home requires several things. The homepage or primary brand page must contain explicit, machine-readable statements of identity: what the brand is, what it does, what category it belongs to, and who it serves. These statements should be reinforced with structured data (organization schema, at minimum) that maps directly to Knowledge Graph entity types. The page must be consistently maintained so that it is reliably current. Finally, the information on the Entity Home must be corroborated by external sources: if a brand's homepage says it is an enterprise SaaS company specializing in supply chain logistics, that same characterization should appear consistently across its LinkedIn profile, its Crunchbase listing, its press mentions, its industry directory entries, and its Wikipedia page. The Knowledge Graph builds confidence through corroboration. Inconsistencies can erode the entity's confidence score and delay or prevent recognition.

Earned Media and Third-Party Validation

Authoritativeness under NEEATT is granted, not claimed. The LLM, seeking consensus across diverse sources before generating a response, must find independent evidence corroborating a brand's claims.

Chris Long's work on digital authority emphasizes that this is not a new marketing objective but one whose strategic weight has been elevated. Earning mentions in reputable industry publications, securing analyst coverage, publishing original research that others cite, and building review profiles on trusted platforms all create independent data points that the retrieval pipeline uses to confirm brand claims during grounding. AI recommendations appear harder to game than traditional search, in part because LLMs weigh earned media from reputable outlets more heavily than self-published content. The grounding phase actively seeks multi-source corroboration and discounts single-source narratives.

User-Generated Content

AI models increasingly weigh user-generated content (UGC) when grounding their responses. UGC refers to platforms such as Reddit, community forums, and review platforms. This likely operates as a direct response to the challenge of distinguishing genuine human knowledge from the growing volume of synthetic content on the web. Research on AI citation behavior shows that Wikipedia accounts for nearly 48% of ChatGPT's top citations. Reddit surfaces prominently in Gemini and Perplexity results (Goralewicz). Community-generated discussions frequently outrank corporate websites in AI-generated answers, especially in finance and technology verticals. The models have

been designed to value content that shows evidence of authentic human experience and debate.

For brands, the practical implication is that UGC platforms are no longer peripheral to marketing strategy; they are central to AI visibility. Brands must engage directly and honestly on the platforms where humans discuss their products — not to manipulate sentiment, but to participate in the conversation that the AI will ultimately synthesize. This means responding to criticism on Reddit, contributing substantive expertise in relevant forums, and maintaining active, authentic profiles on trusted review platforms.

The deeper point is about consensus. LLMs, before generating a response, seek agreement across diverse source types. A brand's owned content (its website, its blog, its press releases) represents one voice. Earned media (press coverage, analyst reports, review aggregator scores) represents a second. UGC represents a third. If the owned and earned narratives claim excellence but the UGC narrative reflects frustration, the LLM will synthesize a response closer to the UGC signal, because the model's architecture is designed to favor multi-source corroboration over single-source claims. The brands that perform best in AI-mediated discovery will be those where the story told on their website, the story told by third parties, and the story told by actual users all point in the same direction.

4. Strategic and Market Implications

In this section, we expand upon the current state of AEO outlined above and develop evidence-based predictions for how the landscape will evolve over the coming years. To do so, we must draw parallels between AEO and SEO, while understanding how the differences between these two paradigms may impact how they unfold in industry. Through this approach, we achieve a reasonable estimation of the future AEO environment while avoiding the compounding errors traditionally found in repeated event-based extrapolation.

4.1. Firm Strategy

AEO adoption is likely to disrupt firm-level marketing strategy as direct advertisements become less effective. Current research indicates that the optimal split in a company's marketing budget is 60% brand and 40% performance, where "brand" covers awareness, positioning, and long-term equity building, and "performance" covers direct-response advertising aimed at immediate conversions (Binet and Field). However, in practice, companies are over-investing in short-term marketing campaigns, pushing their "performance" allocation above the recommended 40% and sacrificing long-term brand development ("Balancing"). This results in companies being top-of-mind for consumers at the cost of being defined as the authoritative company in their space.

Under the AEO paradigm, users increasingly search for solutions rather than specific brand names and rely on the LLM to generate recommendations. Thus, the driving factor for selection evolves from user familiarity with brand names to LLM brand understanding. This challenges the fundamental principles behind short-term marketing advertisements and increases the importance of long-term branding. That is to say, while companies may get a small benefit now by increasing their performance budget, they would likely realize greater long-term benefits by investing more in their brand.

This shift also reframes the role of marketing more broadly. Rather than optimizing for exposure to users, firms must optimize for inclusion in model-generated recommendations. Content strategy becomes more focused on highly specific, bottom-of-funnel queries, where precise alignment with user intent drives retrieval. Brand strategy centers on consistent entity representation across the web, ensuring that models can accurately identify and evaluate the firm. Distribution strategy expands beyond owned channels to

include earned media and user-generated platforms, which provide independent signals of credibility during the grounding process.

Importantly, this line of reasoning predominantly applies to commoditized and substitutable goods. Luxury and specialized goods, where consumers often search for a specific brand or product by name, may still benefit from direct demand that bypasses LLM-mediated discovery. At least for now, this means their advertising marketing, which directly influences consumer behavior, may remain effective. Still, there is a question of whether luxury brands will be able to maintain a consumer presence strong enough to incentivize consumers to purchase directly—especially as LLM search continues to grow in popularity.

4.2. Market Structure and Competitive Dynamics

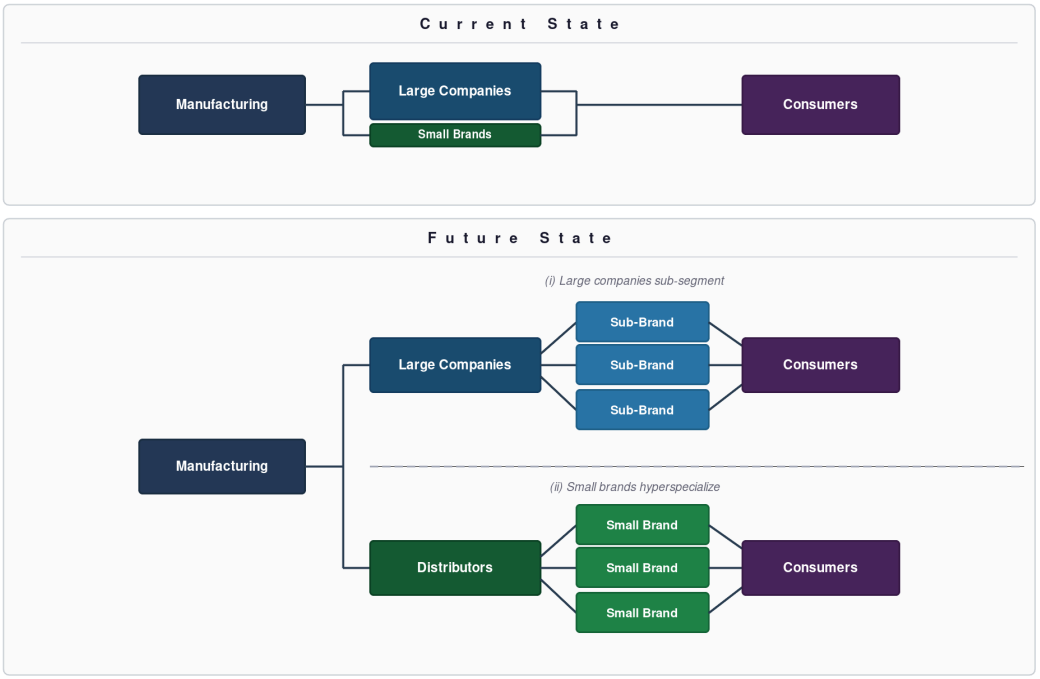

Figure 1: *Impact of AEO on Market Composition.*

If the primary medium of marketing is shifting from advertising to influencing LLM recommendations through branding, a very natural question is then, "Will companies be able to pay to directly influence LLM recommendations?" Just as companies bid for web advertisements, we could imagine a world in which LLMs incorporate bids into recommendations by biasing their output towards companies with bigger bids.

Using SEO as a reference, this outcome appears unlikely. While companies can pay for “sponsored” content slots in traditional search, they are not able to bid for preferential treatment in organic results on major search engines, including Google ("About Measuring"). Although some research has shown a positive correlation between spending on sponsored content and organic results, these effects are indirect and primarily stem from organic content driving sponsorship conversion. This separation of paid and organic content maintains the balance between user trust and advertiser value, which is essential to any two-sided market (Rochet and Tirole).

Given that AEO systems retain a similar two-sided structure, introducing paid bias into organic recommendations would likely erode platform value beyond an acceptable level for users. This is especially true in a competitive environment where multiple frontier models compete for user retention, and customers have shown a high willingness to switch between platforms (Andrews). As a result, firms cannot rely on spend alone to secure visibility within LLM outputs. Instead, they must compete for inclusion through credibility and consensus.

This shift has important implications for competitive dynamics. As previously mentioned, LLM recommendations tend to favor specialized brands over large companies since they more directly address specific user requests. Because retrieval is driven by alignment with specific query attributes, firms that produce highly targeted, domain-specific content are more likely to be surfaced than those with broader but less precise coverage.

Therefore, as brands optimize for LLM recommendations, we expect to see a “hyperspecialization” of smaller brands and segmentation of larger brands. That is to say, small brands will further invest in and emphasize their niches, while larger companies will segment their offerings into different brands so that each can be adequately specialized toward their corresponding products. Larger companies segmenting into sub-brands should have an advantage since their history provides credibility, but as large companies are notoriously slow-moving, it is difficult to determine which group will dominate future markets. In either case, the resulting market composition is similar and involves a new layer of brands operating between consolidated manufacturers and consumers.

The shift from large conglomerates to specialized brands would fundamentally change how we think about branding as a competitive advantage. Instead of a single powerful asset that allows large companies to easily expand into new markets, brand equity becomes localized to specific markets. This process contrasts sharply with current marketing dynamics and raises questions about the sustainability of firms that rely primarily on broad brand equity as their core advantage.

At the same time, LLM platforms function as intermediaries between brands and consumers, with their recommendations increasingly determining which products receive attention. This role is structurally analogous to platform intermediaries such as app stores, which gate distribution between producers and users while capturing value from transactions. This intermediation is consequential for several reasons. First, it introduces a dependency on a platform whose ranking logic is largely opaque. Unlike traditional SEO, where ranking factors are partially understood and auditable, LLM recommendation algorithms offer brands limited visibility into why they are or are not being surfaced. Second, the consolidation of consumer discovery into a small number of LLM interfaces concentrates market power in the hands of platform operators. While consumers retain the ability to bypass these systems and navigate directly to brand websites, this constraint may weaken as the convenience of agentic search compounds and habitual usage deepens.

4.3. Measurement, Transactions, and System Constraints

While the preceding analysis outlines the trajectory of AEO, several practical challenges remain unresolved, which will determine how adoption unfolds. Specifically, companies and users face constraints related to measurement, transaction security, and platform dependence.

Possibly the most immediate challenge to AEO adoption among brands is the measurement of its return on investment. In traditional SEO, ROI can be tracked through well-established metrics: keyword rankings, click-through rates, and conversion attribution. AEO, however, disrupts this measurement pipeline. When an LLM recommends a product, there may be no click, no referral URL, and no trackable session (Long, Blyskal).

The primary metric proposed to fill this gap is "share of voice"—the frequency and prominence with which a brand appears in LLM-generated responses relative to its competitors. While several platforms have begun offering share-of-voice tracking tools, the measurement remains imprecise. LLM outputs are non-deterministic; the same prompt may yield different recommendations on successive runs. Additionally, responses vary across models, and there is no standardized methodology for sampling prompts, weighting results, or accounting for model updates that may shift recommendations overnight.

This challenge is further compounded by the increasing role of personalization in LLM outputs. As models incorporate more user-specific information—search history, stated preferences, and behavioral signals—their recommendations will diverge across individuals. This is a natural consequence of the technology’s value proposition. However,

it implies that the concept of a universal brand ranking within LLM search effectively disappears. A brand that ranks first for one user may not appear at all for another.

This does not make measurement impossible, but it does require a shift in methodology. Rather than tracking a brand's absolute rank, companies must instead estimate their expected rank across the distribution of users—formally, E[E[rank | user]]. That is to say, the relevant metric becomes the average of the personalized expected rankings across the user population, rather than the ranking observed in any single query. This approach introduces statistical complexity but more accurately reflects a brand's true visibility in a personalized search environment.

Beyond measurement, the rise of agentic systems introduces challenges related to transaction execution and user control. As LLM-based agents move beyond recommending products to initiating or completing purchases, the degree of autonomy granted to the system becomes critical. We identify three tiers of agent involvement in transactions:

- **AI Recommendation:** The agent suggests products but takes no transactional action.
- **AI Initiation:** The agent populates a checkout form or assembles a cart, but requires a human to confirm the purchase.
- **AI Transaction:** The agent autonomously executes a purchase within a predefined budget.

Of these three options, AI initiation represents the most viable configuration. This is because a one-click confirmation within an LLM interface is functionally equivalent to a one-click purchase on a brand's own website. The user retains final authority over the transaction while still benefiting from the agent's ability to autonomously complete all other aspects of the process. An AI transaction, on the other hand, could result in accidental purchases where no explicit confirmation was given. Not only would this create logistical challenges for businesses, but users may still be liable for the purchase under existing regulatory frameworks ("Electronic Fund Transfers").

Finally, the increasing role of LLM platforms as intermediaries introduces structural constraints for firms. Platform operators control access to users, but their ranking systems are largely opaque and continuously evolving. This creates uncertainty for firms

attempting to optimize their visibility and introduces a dependency on external systems whose incentives may not align with their own.

4.4. The Future of AEO

AEO itself is a rapidly developing field that evolves continuously. The infrastructure is already being built. Anthropic's Model Context Protocol, OpenAI's Agentic Commerce Protocol, Google's Universal Commerce Protocol, and the Agent-to-Agent standard are all either live or in active deployment. McKinsey estimates up to $1 trillion in U.S. B2C retail revenue could be orchestrated by AI agents by 2030. Optimizing for today's generative search experience is necessary, but it is not sufficient.

Model developers are routinely updating their LLM search algorithms and agent architecture to more accurately retrieve information and verify sources' trustworthiness. Simultaneously, digital marketing agencies are figuring out how to adjust their content, restructure their websites, and even manipulate the publicly available data about their companies to bias models towards their offerings. This adversarial behavior forces each side to continue innovating and drives the developments we see in AEO.

The competition between marketing specialists and model developers occurs in two categories: reviews and validation. Developers want to find detailed, accurate reviews and be able to validate that their information comes from reputable sources. Marketing specialists are incentivized to bias those reviews to portray their company more favorably and inflate their credibility so they appear reliable. Accordingly, we propose two pathways through which this competition can reach equilibrium.

First, companies could align public web sentiment with their desired brand image. This would eliminate the need for marketing specialists to manipulate their public image, as it would already be in line with their reviews. Notably, research has shown public sentiment toward companies is "sticky" as people's cognitive biases resist change (Schultz et al.). Companies that are willing to invest in changing their image and willing to take the necessary steps, however, can still accomplish this task (Benoit).

Another possibility is that LLM developers may adopt validation systems that are independent or resistant to being manipulated by brands (e.g., a third-party reviewer platform). In this way, they would end the competitive cycle by cutting out the ability for companies to influence model outputs directly. However, this option remains theoretical as

no such system currently exists. The possibility remains that such a system could be built and resolve competition between groups.

Conclusion

The shift from search engine to answer engine is no longer emerging, but underway. AI platforms process billions of queries daily, zero-click sessions dominate a growing share of interactions, and AI-referred traffic converts at rates several multiples higher than traditional organic search ("AI Traffic Converts"). Users arriving via AI recommendation are more qualified, more intent-driven, and more commercially valuable than those arriving through conventional search. Brands that treat AEO as speculative are making a bet against trajectories that have already materialized.

This paper has established the technical and strategic foundations of AEO as it exists today—passage-level retrieval, pairwise ranking, the NEEATT credibility framework, and the emerging agentic commerce layer. However, the more pressing question is what comes next, and here we identify three priorities for companies seeking to act on this analysis.

First, companies should rebalance marketing investment from performance advertising toward brand authority. The retrieval pipeline's reliance on multi-source corroboration during grounding means that credibility cannot be purchased through ad spend alone. The alignment between what a brand claims, what third parties report, and what users experience is now an architectural requirement of the systems that mediate discovery. Investment in earned media, authentic community engagement, and consistent entity representation across the web will compound in value as AI-mediated discovery scales.

Second, companies must build measurement infrastructure suited to a non-deterministic environment. The concept of a universal brand ranking within LLM search is effectively obsolete. As we have argued, the relevant metric is the expected rank across the distribution of users, not the ranking observed in any single query. Companies that develop robust methodologies for sampling across platforms, accounting for personalization, and tracking share of voice against this distributional benchmark will be able to optimize their strategies against a metric that reflects the actual consumer experience. Those that continue to rely on static rank-tracking will be optimizing against a signal that no longer exists.

Third, companies should prepare now for the agentic transaction layer. When AI agents move from recommending products to executing purchases, the brands that are structurally accessible to those agents (through machine-readable product data, structured APIs, and clear fulfillment terms) will capture demand that competitors cannot. The time between the current recommendation-only paradigm and a fully transactional agentic environment is the most strategic window in which companies can act.

Finally, we reiterate that AEO is not a marginal update to SEO. It is the ability for search to directly answer users' queries and, increasingly, transact on their behalf. The brands that internalize this shift earliest will hold a durable advantage as AI-mediated discovery becomes the dominant mode of consumer search.

Works Cited

"About Measuring Paid and Organic Search Results." *Google Ads Help*, Google, support.google.com/google-ads/answer/3097241?hl=en. Accessed 16 Mar. 2026.

Aggarwal, Pranjal, et al. "GEO: Generative Engine Optimization." Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, 2024, pp. 5–16.

"AI Platform Citation Patterns: How ChatGPT, Google AI Overviews, and Perplexity Source Information." Profound, 2025, www.tryprofound.com/blog/ai-platform-citation-patterns.

"AI Traffic Converts at 3x the Rate of Other Channels (Study)." Microsoft Clarity, 6 Nov. 2025, clarity.microsoft.com/blog/ai-traffic-converts-at-3x-the-rate-of-other-channels-study/.

Andrews, Debra. "Why Are Millions of Users Leaving ChatGPT for Claude?" *Built In*, 6 Mar. 2026, builtin.com/articles/chatgpt-claude-switching-analysis.

Apptopia. "Gen AI Chatbots: February 2026 Apptopia Data Brief." Apptopia, Feb. 2026, apptopia.com/en/insights/gen-ai-chatbots-february-2026-apptopia-data-brief/.

"Balancing Your Budget between Performance Marketing and Brand Building Has Never Been More Important." *iO*, 23 Sept. 2024, www.iodigital.com/en/insights/blogs/balance-brand-building-vs-performance-marketing.

Barnard, Jason. "N.E.E.A.T.T in SEO: Things You Need to Know." Kalicube, 22 June 2025.

Barnard, Jason. Personal interview. 3 Feb. 2026.

Benoit, William L. *Accounts, Excuses, and Apologies: A Theory of Image Restoration Strategies*. State University of New York Press, 1995.

Binet, Les, and Peter Field. *Effectiveness in Context: A Manual for Brand Building*. IPA, 2018, downloads.ctfassets.net/ptzdhtf6t0jg/7hlC0oHJIdCqATrkrKINKT/246b2755a78a7872b8e75389aab5203f/Effectiveness_In_Context.pdf.

Blyskal, Josh. Personal interview. 6 Feb. 2026.

Chatterji, et al. "How People Use ChatGPT." *NBER*, 12 Sept. 2025, https://www.nber.org/papers/w34255.

Coggins, Becca, et al. "The Agentic Commerce Opportunity: How AI Agents Are Ushering in a New Era for Consumers and Merchants." McKinsey & Company, Oct. 2025, www.mckinsey.com/capabilities/quantumblack/our-insights/the-agentic-commerce-opportunity-how-ai-agents-are-ushering-in-a-new-era-for-consumers-and-merchants.

"Digital Ad Spend Worldwide to Pass $600 Billion This Year." eMarketer, 2023, www.emarketer.com/content/digital-ad-spend-worldwide-pass-600-billion-this-year.

"Electronic Fund Transfers (Regulation E)." *Code of Federal Regulations*, title 12, sec. 1005.6, 2011. *Electronic Code of Federal Regulations*, www.ecfr.gov/current/title-12/chapter-X/part-1005/subpart-A.

Feather, Nathan, and Brian Nowak. "Agentic Commerce Impact Could Reach $385 Billion by 2030." Morgan Stanley Research, 2025, www.morganstanley.com/insights/articles/agentic-commerce-market-impact-outlook.

First Page Sage. "Top Generative AI Chatbots by Market Share – March 2026." First Page Sage, 10 Mar. 2026.

Fortune. "ChatGPT's Market Share Is Slipping as Google and Rivals Close the Gap, New App-Tracker Data Shows." Fortune, 5 Feb. 2026.

Gao, Yunfan, et al. "Retrieval-Augmented Generation for Large Language Models: A Survey." arXiv, 18 Dec. 2023, arxiv.org/abs/2312.10997.

"Google's Gemini App Has Surpassed 750M Monthly Active Users." TechCrunch, 4 Feb. 2026, techcrunch.com/2026/02/04/googles-gemini-app-has-surpassed-750m-monthly-active-users/.

Google. "Generative Summaries for Search Results." US Patent No. US11769017B1, 26 Sept. 2023.

Google. "Method for Text Ranking with Pairwise Ranking Prompting." US Patent Application No. US20250124067A1, 17 Apr. 2025.

Google. "Response Generation Using a Retrieval Augmented AI Model." US Patent Application No. US20240346256A1, 17 Oct. 2024.

Goralewicz, Bart. Personal interview. 30 Jan. 2026.

Guaglione, Sara. "Many GEO Tactics Are Not That Different from Search Optimization." Digiday, 10 Mar. 2026.

iPullRank. "AI Search Architecture Deep Dive: Teardowns of Leading Platforms." iPullRank AI Search Manual, 13 Aug. 2025.

Kahn, Jeremy. "As 'Agentic Commerce' Gains, Brands Shouldn't Put Too Much Faith in 'GEO,' an Industry Insider Says." Fortune, 13 Jan. 2026, fortune.com/2026/01/13/agentic-commerce-generative-engine-optimization-geo-unreliable-aivo-standard/.

King, Michael. Personal interview. 29 Jan. 2026.

Kopp, Olaf. "How AI Mode and AI Overviews Work Based on Patents." Search Engine Land, 2 June 2025.

Kopp, Olaf. Personal interview. 3 Feb. 2026.

Kopp, Olaf. "What Is Brand Context Optimization for GEO?" Kopp Online Marketing, 21 Feb. 2026.

Kopp, Olaf. "What We Can Learn about Google's AI Search from the Official Vertex & Cloud Documentation." Kopp Online Marketing, 31 Dec. 2025.

Lewis, Patrick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 9459–9474, proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Long, Chris. Personal interview. 3 Feb. 2026.

Microsoft Advertising. "From Discovery to Influence: A Guide to AEO and GEO." Microsoft Advertising Blog, 6 Jan. 2026, about.ads.microsoft.com/en/blog/post/january-2026/from-discovery-to-influence-a-guide-to-aeo-and-geo.

Petrovic, Dan. Personal interview. 2 Mar. 2026.

Qin, Zhen, et al. "Large Language Models Are Effective Text Rankers with Pairwise Ranking Prompting." Findings of the Association for Computational Linguistics: NAACL 2024, 2024, pp. 1504–1518, aclanthology.org/2024.findings-naacl.97.

Rochet, Jean-Charles, and Jean Tirole. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association*, vol. 1, no. 4, 2003, pp. 990–1029.

"Sam Altman Says ChatGPT Has Hit 800M Weekly Active Users." TechCrunch, 6 Oct. 2025, techcrunch.com/2025/10/06/sam-altman-says-chatgpt-has-hit-800m-weekly-active-users/.

SAP News Center. "For Retailers, Agentic Commerce Is Here." SAP News Center, 22 Jan. 2026, news.sap.com/2026/01/for-retailers-agentic-commerce-is-here/.

Schultz, Majken, et al. "Sticky Reputation: Analyzing a Ranking System." *Corporate Reputation Review*, vol. 4, no. 1, 2001, pp. 24–41.

Stec, Jeff. Personal interview. 4 Mar. 2026.

theCUBE Research. "AI Engine Optimization (AEO): How to Get Cited in AI Answers." theCUBE Research, 17 Dec. 2025, thecuberesearch.com/ai-engine-optimization/.

World Economic Forum. "New Era of Performance Marketing: How Brands Are Repositioning for Agentic Engine Optimization (AEO)." World Economic Forum, Jan. 2026, www.weforum.org/stories/2026/01/new-era-of-performance-marketing-how-brands-are-repositioning-for-agentic-engine-optimization/.

Yesilyurt, Metehan. Personal interview. 10 Feb. 2026.