Simple because it matters.

ERGO

From SEO to LLM Discovery

Insights from the Insurance Sector to Navigate the Next Chapter of Search

How Large Language Models Are Redefining Digital Visibility

ERGO

The **ERGO Innovation Lab** works on the forefront of new technologies and services regarding insurance, risk and finance. Located at the Merantix AI Campus in Berlin, the team is continuously challenging the status quo of the insurance industry. With industry-shaping projects, influenced by technology and digital trends.

ECODYNAMICS

ECODYNAMICS' experts advise and train companies on generative AI and digital business models. The company has been an OpenAI Early Adopter since 2021 and specialises in the implementation of generative AI, AI agents, and the analysis and optimisation of websites for AI-based search engines.

„By 2026, traditional search engine volume will drop 25%, with search marketing losing market share to AI chat-bots and other virtual agents.“

Gartner Inc.

The way consumers search for information online is undergoing a fundamental shift. Traditional keyword-based search, which is still the dominant method of retrieving information online, is giving way to more advanced, conversational approaches powered by large language models (LLMs). These AI-driven systems do not merely retrieve lists of links – they understand user intent, generate precise and conversational responses, and refine their outputs based on real-time data. This transformation is particularly relevant for the insurance industry, where clarity, accessibility and trust of information play a crucial role in customer decision-making.

Consumers today expect instant, personalized, and accurate information, and LLM-powered search has the potential to meet these demands more effectively than traditional keyword-based search. Instead of forcing users to scroll through pages of search results, these models provide direct and contextually relevant answers, helping consumers quickly navigate insurance products and services.

Yet, as with any technological breakthrough, LLM search also presents new challenges. For years, digital content strategies have been shaped by Google's near-total dominance in the search market. Now, with the emergence of AI-driven retrieval systems, insurers are beginning to face new demands. Content must increasingly be optimized not just for ranking on search engines, but for being referenced and trusted by large language models. This shift brings new risks, such as biased outputs or unreliable summaries and calls for updated strategies around visibility, content structure, and quality assurance in an evolving search landscape.

At the ERGO Innovation Lab, together with ECODYNAMICS, we have taken a deep dive into these developments, exploring their implications for the insurance industry. This whitepaper provides an in-depth analysis of the current state of LLM search and its evolving role in consumer information gathering, engagement and decision-making.

Drawing on an in-depth study involving more than 33,000 retrieval results and over 600 website evaluations, we examine how LLMs will reshape the way people search for and engage with insurance products. We also explore the structural and technological adjustments required for insurers to adapt and discuss the broader impact on the industry. In addition, we highlight key trends shaping the future of AI-powered search and provide recommendations for leveraging this transformation effectively.

Understanding how LLM search changes consumer behavior and aligning business strategies accordingly will be critical for insurers to stay competitive in this new landscape. Insurers must now produce content that resonates with AI retrieval logic, or else see brokers, capture higher positions in user queries. This challenge is further amplified by consumers' growing comfort with LLM-based and agentic (autonomous) AI tools, where entire insurance searches, comparisons, and even policy sign-up processes could unfold within a chat or voice interface.

Table of Contents

LLM Search and the Decline of Traditional SEO

AI tools like ChatGPT are changing how people search for information, replacing lists of links with direct, conversational answers. This shift is reshaping user behavior, challenging traditional search engines and creating new opportunities and risks for businesses, especially in sectors like insurance.

The Evolution of AI-Powered Search

The integration of search technology into Large Language Models (LLMs) has significantly transformed how users interact with information. More and more individuals are turning to LLMs like ChatGPT or Perplexity to receive targeted answers, rather than traditional lists of links. Traditional keyword-based search engines like Google have evolved into AI-driven models that understand context, generate direct answers, and provide conversational search experiences. Instead of merely retrieving links and brief snippets, these systems deliver synthesized responses, streamlining the search process. A key technology enabling this shift is Retrieval-Augmented Generation (RAG), where LLMs retrieve relevant external information in real time to enhance the accuracy and relevance of their answers.

According to Gartner, a leading market research firm, traditional search engine volumes are expected to decline by 25% by 2026 in favor of AI-driven chatbots (Gartner Inc., 2023). While Google currently remains the dominant player with 89% market share (StatCounter, cited in Statista, 2025), it is already complementing its search results with the Google AI Overview, an AI-generated summary that appears directly in the search results, leading to what is known as zero-click behavior. This occurs when AI-generated answers aim to satisfy user queries directly within the search interface, thereby eliminating the need to click on external websites. As a result, many publishers and businesses reliant on organic search traffic are rethinking their strategies to maintain visibility in an increasingly AI-driven search landscape.

Simultaneously, alternatives like Perplexity AI, You.com, Google Gemini and SearchGPT are emerging, offering direct answers without traditional search result lists and have been specifically designed to meet this demand for conversational, direct answers. Perplexity AI has been attracting a growing user base as an alternative to conventional search engines, while Google is expanding the capabilities of its Gemini model, integrating AI-driven search with their AI overviews (Eisenbrand, 2025) into the Google ecosystem. Recent data from Seer Interactive (2025) shows that queries with AI Overviews see a steep decline in organic CTR – from 1.41% to 0.64% year-over-year. Paid CTRs dropped overall, regardless of AI Overview presence. However, brands featured in AI Overviews benefit: organic CTR increases from 0.74% to 1.02%, and paid CTR jumps from 7.89% to 11%. This highlights the growing importance of AIO visibility in an increasingly zero-click search environment.

As search behaviors shift toward AI-generated responses, businesses must optimize content for both traditional search engines and LLM-driven queries to maintain visibility. This is especially crucial for industries like insurance, where consumers expect quick, precise, and personalized answers to their product inquiries. In this context, businesses that adapt to new search behaviors can significantly enhance their visibility in AI-generated search results. This strengthens their competitive position and opens up new opportunities in customer engagement, sales, and service, from interactive, conversational advice to personalized product recommendations and comparison overviews.

By comparing LLM-based search results with those of traditional search engines, we aim to identify the factors that drive visibility and ranking within these new AI-powered systems. The focus will be on understanding how businesses can optimize their presence in AI-generated search results to ensure they stay relevant in this evolving landscape.

Challenges in AI-Powered Search

Despite the advantages of AI-enhanced search, several challenges persist. One of the primary concerns is accuracy, as LLMs sometimes generate incorrect or misleading information while presenting it with high confidence, so called hallucinations (Zhang et al, 2023). To mitigate this, companies are incorporating real-time web citations and developing advanced fact-checking mechanisms. These efforts aim to build user trust and ensure that AI-driven interactions remain both useful and reliable.

Legal and copyright issues are another major point of contention (Karamolegkou et al, 2023). News publishers and content creators have voiced concerns over AI systems scraping and summarizing their content without proper attribution (Hagey et al, 2023). Tech companies have faced increasing scrutiny regarding the ethical and legal implications of data usage in AI models.

Transparency and bias in AI-generated responses also remain significant issues. Since these models are trained on vast datasets, they can unintentionally reflect biases present in the underlying information (Gallegos et al, 2024). Efforts to refine these models focus on diversifying training data and improving mechanisms that provide clearer explanations for AI-generated outputs.

Benefits of LLM Search for the Insurance Industry

The insurance industry depends on precise, consistent communication across policy documentation, claims processes, and customer service. While insurance products, legal frameworks, and regulatory requirements can be complex, greater clarity can be achieved through improved accessibility and personalization – enhancing transparency, fostering trust, and supporting more informed decision-making.

AI-driven search transforms how consumers access insurance information – shifting from static queries to dynamic, interactive conversations (Mo et al., 2024). Users can ask follow-up questions to clarify details without contacting an agent, a model that particularly appeals to digital-native customers who prefer self-service. However, especially for more complex or longterm insurance products, many consumers still follow a ROPO (Research Online, Purchase Offline) behavior: they gather information online but complete the purchase offline through personal consultation with brokers or agents (GDV, 2023). LLMs support this evolving journey by providing personalized, context-aware answers that simplify policy language, reduce research time, and translate legal terminology into plain, understandable terms – enhancing both trust and comprehension.

One major advantage of LLM-powered search is its ability to generate policy comparisons in a conversational format, on platforms beyond traditional broker tools. Instead of relying solely on pre-built comparison tables, users can now steer the comparison process themselves via natural language queries and follow-up questions. This enables more flexible, context-aware comparisons, tailored to individual needs and delivered through a dynamic, chat-based experience.

Moreover, these systems provide personalized insurance advice, for instance, offering context-specific insights when searching for "best travel insurance for families with young children", the response reflects relevant needs such as child coverage, family-friendly benefits, or medical assistance abroad, while maintaining context across follow-up questions.

Finding the right balance between trust and innovation is crucial. While challenges such as hallucinations, bias, and legal uncertainty remain, waiting for fully mature systems may lead to missed opportunities as others move ahead. Although insurers cannot control how external LLMs present or phrase responses, they can influence the quality of those responses by ensuring that accurate, well-structured, and up-to-date information is publicly available and machine-readable.

Prerequisites for Traditional Search Engines & SEO

To achieve optimal visibility in traditional search engines such as Google, Bing, or Yahoo, websites must meet a range of technical, structural, and content-related standards. These criteria improve both discoverability and ranking in search engine result pages (SERPs) (Google Search Central, 2023; Iqbal et al., 2022).

Traditional SEO remains essential. LLM-based systems depend on established search engines like Google to crawl and index websites (Iqbal et al., 2022). Without meeting technical and structural SEO standards, content may not be discoverable at all.

The following section outlines essential prerequisites for effective Search Engine Optimization (SEO).

Indexability & Crawl-Friendliness

Ensuring that a website can be efficiently crawled and indexed by search engines is fundamental to SEO. Key technical requirements include:

- **Indexable Pages:** All important pages must avoid "noindex" tags and robots.txt restrictions. Crawlers require explicit permission to access and rank content (Google Search Central, 2023).

- **Clear Sitemap & Internal Linking:** A well-structured sitemap and logical internal links allow search engines to navigate the site and discover all essential content (Iqbal et al., 2022).

- **Mobile-First Optimization:** Since search engines prioritize mobile usability, websites must use responsive designs or mobile-specific layouts to ensure consistent experience across devices (Google Search Central, 2023).

- **Page Loading Speed Optimization:** Fast-loading pages improve user experience and boost rankings. Optimized images minimized CSS/JS, and efficient servers are key (Terenteva, 2023).

- **Avoiding JavaScript-Dependent Content:** Essential content should be accessible without JavaScript rendering, as crawlers may not fully index JS-based elements (Google Search Central, 2023).

While these technical requirements ensure accessibility and crawlability, on-page content and structure remain equally critical for ranking success. In addition to technical accessibility, effective on-page SEO ensures that content is both understandable and attractive to search engines and users. The following best practices help communicate content structure, relevance, and purpose clearly (Moz, 2024; Iqbal et al., 2022).

Classic On-Page SEO Factors

Effective on-page SEO involves clearly communicating content and relevance to both search engines and users. Essential practices include:

- **Unique Title Tags & Meta Descriptions:** Every page should feature concise and engaging metadata that accurately reflects its content and increases click-through rates (Moz, 2024).

- **Structured Headings (H1, H2, H3):** A clear heading hierarchy improves readability and helps search engines identify key themes and sections (Iqbal et al., 2022).

- **Clean URL Structure:** URLs should be short, descriptive, and free of unnecessary parameters. Well-structured URLs enhance crawlability and user trust (Terenteva, 2023).

- **Image Alt-Texts:** To improve accessibility and content relevance, every image should include descriptive alt-text, enabling search engines to interpret visual elements (Google Search Central, 2023).

- **Keyword Optimization:** Relevant keywords must be naturally embedded in titles, headings, and content. Overuse ("keyword stuffing") should be avoided to maintain readability and trust (Moz, 2024).

These on-page strategies form the foundation for strong SEO. However, delivering valuable and engaging content remains equally essential to rank competitively and retain users. While technical and structural factors ensure that content is accessible, it is the quality and user-centric design that ultimately determines whether users stay, engage, and convert. High-quality content that aligns with user needs is essential for long-term search performance (Google Search Central, 2023; Moz, 2024).

Content Quality & User Experience

Content remains king in traditional SEO. Ensuring high-quality user experience and delivering content that precisely meets user intent are critical for long-term success:

- **Unique & High-Quality Content:** Content must offer unique value – avoiding duplication and providing clear differentiation from competitors (Google Search Central, 2023).

- **Relevant & Well-Structured Texts:** Clear structure, concise paragraphs, bullet points, and intuitive formatting enhance both readability and engagement (Moz, 2024).

- **Aligning with Search Intent:** Content should match the user's specific intent – informational, navigational, or transactional—to ensure relevance and visibility (Iqbal et al., 2022).

- **Regular Updates:** Keeping content fresh signals authority and relevance. Updated pages rank better and foster trust (Google Search Central, 2023).

- **User Engagement Signals:** Long dwell time and low bounce rates indicate quality and relevance. Sites must be designed to encourage exploration and interaction (Terenteva, 2023).

Strong content and user experience go hand in hand with trust and authority, factors that are increasingly shaped by signals beyond your own website. In the next section, we explore how external reputation influences SEO performance. Beyond what happens on your own site, search engines rely heavily on signals from across the web to evaluate your brand's credibility and relevance. This is where off-page SEO comes into play, shaping how others perceive and reference your content.

Off-Page SEO & Authority

Beyond technical and on-page factors, off-page elements significantly influence visibility and authority in traditional search:

- **Quality Backlinks:** High-quality backlinks from authoritative sources elevate your website's perceived credibility. Editorial links from trusted domains are especially valuable (Moz, 2024; Brin and Page, 1998).

- **Brand Mentions & Social Signals:** References across trusted blogs, media, and social platforms amplify authority, even when no direct link is present (Terenteva, 2023).

- **Building Domain Authority:** Established domains, characterized by consistent positive signals, such as high-quality backlinks, a strong history of valuable content, and low spam scores, tend to achieve stronger rankings (Moz, 2024).

- **Trustworthiness (E-A-T):** Demonstrating expertise, authoritativeness, and trustworthiness is critical, especially in sensitive sectors like finance or insurance. This includes verified authorship, clear credentials, robust citations, and use of HTTPS (Google Search Central, 2023; Moz, 2024).

These principles have long shaped visibility in traditional search and the rise of AI-driven systems introduces new rules, where context, conversation, and semantic meaning matter as much as links. Let us explore how SEO best practices intersect with LLM-based search models.

The Technology behind LLM-based Search

AI search no longer relies on exact keyword matches. Instead, it interprets meaning by mapping queries into a vector space using transformer models that understand context and nuance. This shift transforms search from keyword lookup to contextual reasoning, revealing new challenges and opportunities for digital visibility.

LLM-based search systems differ fundamentally from traditional keyword search. Instead of matching surface terms, they model meaning through vector spaces and semantic context. Dense retrieval maps both query and content into high-dimensional embeddings, enabling matches based on intent, even if exact words do not align. At the heart of this process are transformers – deep learning models designed to understand relationships between words in a sentence, no matter how far apart they are. Introduced by Vaswani et al. (2017), these models use self-attention to capture dependencies across long sequences. This allows them to understand nuance, resolve ambiguity, and generate context-aware responses.

This architecture is the engine of modern search, shifting focus from words to understanding. Next, we explore the model that made this leap possible: the transformer.

Transformers: The Foundation of Modern AI

Transformers changed how machines understand language. Instead of reading words one by one, they use attention to look at everything at once, and decide what matters most. The process starts with tokenization: breaking down a sentence into smaller parts (called tokens). These tokens are turned into numbers, vectors, that let the system work with language like math.

At the core is self-attention, a technique that tells the model which words relate to each other in a sentence. It does this across many layers and directions, helping the model grasp both grammar and meaning. They can understand complex text, scale across huge datasets, and respond with surprising fluency. Next, we explore how language becomes geometry – how these tokens live in high-dimensional vector space.

Tokenization: Breaking Down Language

Every large language model begins by splitting raw text into smaller units called tokens. These aren't full words, but subword chunks, pieces like "health", "insur", and "ance" instead of the full word "health insurance." This method allows models to handle complex, rare, or compound terms efficiently by breaking them into recognizable parts.

Each token is then assigned a unique number from a fixed vocabulary, transforming natural language into a numerical format that machines can understand. This numerical mapping forms the input layer for the model's computations and enables consistent processing across diverse linguistic inputs.

With language now reduced to structured numeric input, the next challenge is to assign meaning to those numbers – through embeddings.

Vectors: Mapping Meaning to Numbers

Before a model can understand or generate language, it needs a way to represent meaning. This is where embeddings come in. Each embedding map each token to a dense vector list of numbers, that capture semantic relationships. Words like "policy" and "coverage" end up close together in this vector space, while unrelated terms remain distant. These spatial relationships are learned from massive training data and allow models to recognize nuance, context, and similarity.

In insurance-related prompts, for example, "premium" and "deductible" may cluster together, signaling thematic relevance. This embedding space becomes the semantic foundation for all further model computations.

From here, the model starts building higher-order understanding using attention mechanisms.

Attention Mechanisms: Prioritizing Context

As embeddings provide a numerical landscape of meaning, attention mechanisms decide where to focus – dynamically and precisely. Instead of treating all words equally, attention assigns weights to each token based on its relevance to others. In a sentence like “The deductible applies before insurance pays,” the model learns that “deductible” and “pays” are tightly linked, even if several words stand between them.

This enables transformers to track complex dependencies, especially useful in highly structured domains like health coverage. Self-attention deepens this by evaluating how every word relates to all others in a sentence. It gives the model a flexible memory and the ability to highlight what truly matters. This selective focus is the reason LLMs can understand nuance. But how is this power surfaced in real-world interactions? The answer lies in the web search interface.

Web Search Interface: The Gateway to LLM-Powered Answers

Static training alone cannot deliver timely insight. To meet evolving user expectations, LLMs rely on integrated web search. Modern LLM-powered systems extend beyond static knowledge by integrating real-time web search. When users submit queries that require current or highly specific information, the system activates a search interface. This module reformulates the query into targeted terms and retrieves content from trusted sources, such as regulatory bodies, news sites, or institutional publishers (Li et al., 2024).

The retrieved content is not passed on directly. It first goes through a filtering step that checks for credibility, timeliness, and contextual relevance. Without this safeguard, irrelevant or low-quality data may skew the AI’s response. As Liu et al. (2023) highlight, these filters are essential for maintaining output integrity. Once validated, the selected information is integrated into the LLMs internal context. This allows the system to blend static knowledge with up-to-date findings, producing responses that are accurate, nuanced, and aligned with user intent (Nakano et al., 2022).

This process is especially critical in high-stakes domains like insurance, where even minor misinformation can have legal or reputational consequences. To minimize these risks, systems increasingly rely on source-level trust signals such as institutional affiliation, author transparency, and verifiable references (Pan et al., 2024). Rather than acting as a simple add-on, the web search interface plays a central role in modern AI systems.

As Xiong et al. (2024) note, it serves as a dynamic bridge, connecting the open web to the model’s reasoning engine and keeping information flow fresh, verifiable, and relevant. Understanding this architecture leads to a deeper question: how do these components interact during a real search process?

How Both Components Work Together

The integration of the language model and the web search interface forms the complete AI search architecture. When a query is received, the system first assesses whether the question can be answered by the internal knowledge of the LLM or if it requires additional real-time information. If external information is needed, the web search interface takes over, retrieving and evaluating the latest relevant data. Finally, the LLM integrates this fresh data with its existing contextual knowledge to create a seamless, coherent, and actionable response for the user.

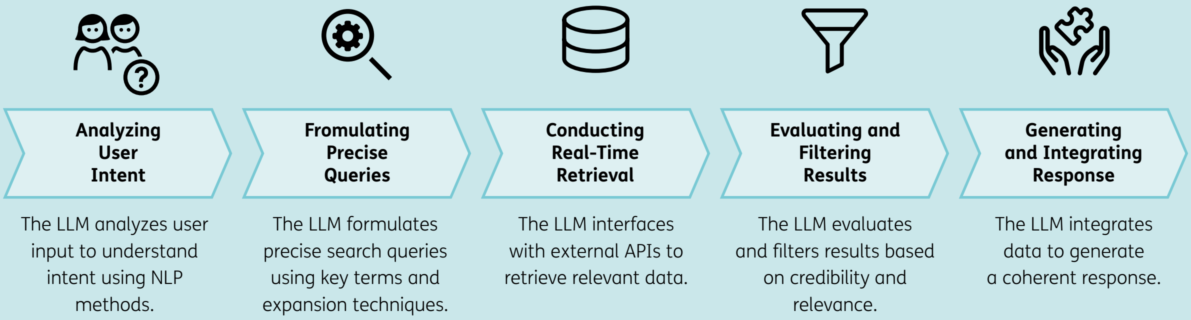
Understanding this architecture clearly highlights how insurers can leverage AI-driven search. By recognizing the roles of both components, organizations can optimize their content and strategies accordingly, ensuring their products, services, and critical information remain visible, accurate, and accessible within these sophisticated search interactions.

Process and Workflow of an AI Search

AI-based search goes beyond powerful models: insurers must grasp each step from query interpretation to accurate, context-driven responses. Knowing this workflow aligns strategies with evolving user behaviors and elevates content delivery. Most LLM-driven searches follow five key stages:

LLM Query Processing and Information Retrieval

Based on several research papers.



1 **Analyzing User Intent**
The system applies advanced Natural Language Processing (NLP) techniques to decode a query's deeper meaning to determine whether the model's internal knowledge suffices or whether real-time data is needed (Yang et al., 2023; Anand et al., 2023). For instance, a user inquiry about "home insurance for coastal flooding" prompts the LLM to interpret both explicit details about flood coverage and implicit concerns regarding geographic risk.

2 **Formulating Precise Query**
When external data proves necessary, the LLM refines the user's question with synonyms, date filters, or domain-specific operators to improve retrieval (Mo et al., 2023). This precision ensures that only the most relevant sources are accessed, boosting both accuracy and speed.

3 **Conducting Real-Time Retrieval**
The system targets high-authority, publicly available resources, regulatory sites, academic literature, recognized publications (Dhole & Agichtein, 2024), while excluding paywalled or private databases. Credibility, timeliness, and relevance guide the selection process, ensuring the final results align well with the user's needs.

4 **Evaluating and Filtering Results**
Each retrieved source undergoes credibility and relevance checks, with automated filters weeding out outdated or conflicting information (Yang et al., 2020). This rigorous screening retains only high-quality data for the LLM to incorporate into its answer, minimizing the risk of inaccuracies.

5 **Generating and Integrating Response**
The LLM merges newly fetched data with its internal knowledge, leveraging transformer-based architectures, self-attention mechanisms, and vector embeddings (Yang et al., 2023). Through this generative synthesis, the system produces a coherent, contextually precise answer that balances static knowledge with real-time findings, ultimately delivering an informed and user-specific response.

Conclusion

With these five stages clearly defined, we can now contrast conventional, keyword-based approaches against the more advanced, context-aware workflows seen in LLM-Search. The operational mechanics of AI Search reveal a new level of contextual understanding and retrieval precision, one that no longer depends solely on exact keyword matches, but on deeper semantic relationships, inferred intent, and broader topic coverage.

To appreciate its impact, it is essential to contrast these capabilities with the limitations of traditional keyword-based search models, which often fail to surface content that is relevant but linguistically misaligned. This shift marks a fundamental change in how visibility is earned and requires to rethink not just how they optimize, but how they communicate.

Key Differences Compared to Traditional Search Engines

Conventional methods focus on keyword matching and backlinks, often yielding generalized link lists (Mo et al., 2024). By contrast, AI-driven approaches interpret user intent and deliver direct, actionable responses, reshaping content strategies, accessibility, and engagement (Kumar et al., 2024). A key distinction lies in interaction flow: users shift from navigating sources to engaging in iterative dialogue, allowing follow-up questions and refinements within the same interface.

The following table outlines the fundamental differences between traditional keyword-based search and LLM-driven model structured across five dimensions.

With these contrasts established, the next section defines the research objectives that guided our empirical analysis. To address these questions directly, we now outline our research objectives, which guide the subsequent in-depth analysis and recommendations presented in this white paper.

Key Differences Compared to Traditional Search Engines

Dimension	Traditional Search (Google & SEO)	LLM-Based Search (AI-Powered Systems)
1. Query Processing Keyword Matching vs. Contextual Understanding	Matches explicit keywords with indexed content. Ranking is based on keyword frequency, backlinks, and metadata (Kumar et al., 2024). Results often lack precision, requiring user-driven filtering.	Interprets full query intent using deep contextual embeddings and transformer-based NLP. Understands nuance, implied meaning, and conversation history (Mo et al., 2024).
2. Information Freshness Static Indexing vs. Real-Time Access	Retrieves from a periodically updated static index. This can lead to outdated results, especially in fast-changing sectors like insurance (Liu et al., 2024).	Retrieves data live from the web, ensuring responses reflect the latest regulations, product updates, and market dynamics (Karamolegkou et al., 2023).
3. Result Format Link Retrieval vs. Generative Answers	Returns a ranked list of links; users must navigate and interpret external pages. This can slow access to relevant answers (Liu et al., 2024).	Generates direct, contextually relevant responses in natural language. Reduces search friction by delivering summarized insights instantly (Mo et al., 2023).
4. Personalization Generic Results vs. Tailored Interaction	Produces general results for a wide audience with minimal personalization. Lacks adaptive understanding of user-specific needs (Mo et al., 2024).	Generates personalized responses based on behavioral signals, search history, and conversational context. Enables individualized product suggestions and user guidance (Sharma et al., 2024; Karamolegkou et al., 2023).
5. Search Architecture & User Experience Manual Navigation vs. AI-Augmented Decision-Making	Empowers users to navigate information themselves, relying on metadata, backlinks, and structured web hierarchies. Users act as interpreters.	AI agents interpret, compare, and summarize on behalf of users—supporting direct decision-making. The search system evolves into a reasoning layer (Mo et al., 2023).

Study Design and Research Objectives

To understand visibility in LLM-powered search, we tested how insurance websites perform across AI systems. Using real-world prompts, we identified which types of content are surfaced most often – and why. The result reveals not just what works today, but what will define who gets found tomorrow.

This study builds on a series of hypotheses aimed at understanding which factors influence the visibility of web pages in LLM-based search compared to traditional Google search. The goal is to identify what types of content, formats, or technical structures increase the likelihood of being referenced or surfaced by these fundamentally different systems. In doing so, the research explores how LLMs are reshaping the search experience, and what insurers need to do to remain visible, findable, and relevant in this evolving landscape.

We start with the assumption that a basic level of SEO optimization is a prerequisite for visibility in both search environments. This does not merely refer to technical hygiene or keyword placement, but to a broader foundation of semantic readiness, the ability of a page to be understood, categorized, and reused by algorithmic systems. In both Google and LLM-driven search, clarity of intent, consistency of messaging, and alignment with user language patterns are essential.

Hypothesis 1: Machine Readability & Technical Accessibility Increases LLM Visibility

The first hypothesis suggests that machine-readable, technically accessible content, such as semantically structured HTML, fast-loading pages, and barrier-free designs is more likely to be referenced by LLMs. Since these systems need to parse and interpret large volumes of content, cleaner and more accessible code offers an advantage. This aligns with the criterion of Technical Accessibility & Machine Readability.

Research has shown that LLMs, while increasingly capable, still depend on reliably parsing and extracting structured content from websites (Achiam et al. 2023). Pages that employ standard markup (e.g., ARIA roles, semantic tags) and avoid blocking scripts or complex client-side rendering offer a better signal-to-noise ratio during retrieval. Comparing the reference frequency of well-structured, fast, accessible URLs against less optimized ones provides a clear way to validate this hypothesis.

Hypothesis 2: Semantic Content Linking Enhances LLM Retrieval

The second hypothesis proposes that content with strong internal linking, conceptual clustering, and structured entities is more frequently retrieved by LLMs. This corresponds to the criterion of Semantic & Linked Optimization.

Since many AI systems use embedding-based retrieval mechanisms, the presence of well-organized semantic relationships helps reinforce topic relevance within a site (Touvron et al. 2023; Ni et al. 2023). LLMs model meaning across documents rather than treating pages in isolation. Content that offers rich interlinking and clear entity relationships (e.g., via schema.org or linked data structures) is better positioned to be selected as a relevant source. Testing this hypothesis could involve analyzing how often semantically rich sites are mentioned in LLM outputs compared to structurally disconnected content. This is because such structures align closely with how LLMs organize and retrieve information, favoring sources that mirror their own latent semantic mapping.

Hypothesis 3: Trusted Sources Are More Frequently Cited by LLMs

The third hypothesis posits that content from trusted sources, such as government websites, academic institutions, or established expert domains is favored more strongly by LLMs than by traditional search engines. This maps to the criterion of Citeability & Source Quality.

LLMs are often fine-tuned and retrieval-augmented to reduce hallucinations by grounding answers in verified content (Nakano et al. 2022). Trust signals like domain authority, author transparency, citation presence, and institutional affiliation play a major role in determining whether a URL is used in an AI-generated response. Ethical AI guidelines (Jobin et al. 2019) further reinforce the preference for high-quality, fact-checked material. The hypothesis can be validated by tracking how frequently high-trust domains appear in LLM outputs compared to less reputable sources.

These four hypotheses form the basis for the comparison between LLM search models and Google search results. For each hypothesis, we examined how often and in what ranking position different types of content appeared, using visualizations to highlight key differences. The results show how core aspects of web architecture, content structuring, and source trustworthiness influence visibility in LLM-driven retrieval systems, and how optimizing content for both human users and AI models is becoming increasingly important for digital presence.

Hypothesis 4: Conversational Formatting Aligns with LLM Retrieval Behavior

The fourth hypothesis argues that content formatted in a conversational, prompt-response style, such as FAQs, Q&A sections, or clearly segmented answer blocks is more likely to be retrieved by LLMs. This is linked to the criterion of Structure & Instant Optimization for LLMs.

Many LLMs are trained on dialogue-rich corpora and prioritize content that matches natural conversational prompts (Wei et al. 2022; Wu et al. 2023; Karpukhin et al. 2020). Structuring content with questions as headers and direct, concise answers beneath mirrors the input-output behavior of conversational AI. Testing this hypothesis could involve measuring how often conversationally structured content appears in AI-generated answers compared to longer, narrative-heavy formats.

Overview of Hypotheses

- **H1:** Machine readability and technical accessibility increase the likelihood of LLM-based retrieval.
- **H2:** Semantic content linking improves the chances of being retrieved by LLMs.
- **H3:** Content from trusted sources is cited more frequently by LLMs.
- **H4:** Conversational formatting aligns better with LLM retrieval behavior.

Pre-Study Design and Methodology

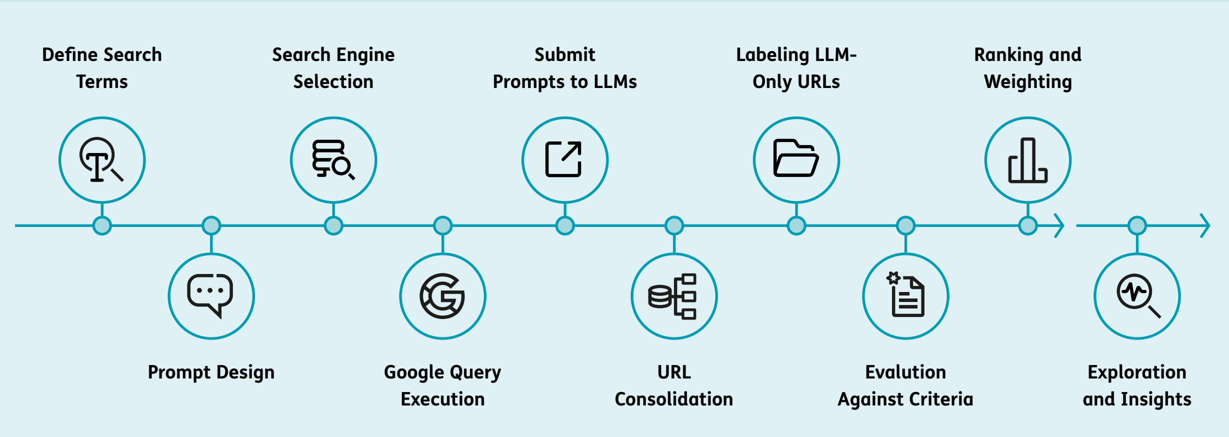
We identified and validated 20 criteria that influence visibility in AI-driven search systems. The study takes a closer look at four key criteria, selected based on the underlying hypotheses.

Understanding how LLM-powered search affects insurance visibility requires a research approach that combines theoretical precision with practical relevance. We developed a methodology aimed at identifying which types of content are retrieved by LLM search engines, how this differs from traditional search, and what structural adjustments insurers need to consider to maintain visibility.

The study followed three key components:

1. **Hypothesis-Driven Testing**
   We formulated hypotheses (e.g., "semantic cohesion improves LLM visibility") and tested them against a

Overall Modular Research Pattern and Methodology



We followed a structured methodology based on ten consecutive steps. This structure gave us a complete picture: from how real users phrase insurance queries, to how LLMs retrieve and rank relevant content. In the next section, we examine how Google search terms were integrated into this framework and what they reveal about shifting visibility across search models. It allowed us to trace how information flows from user query to ranked output in a controlled, comparable sequence.

dataset of over 33,000 URLs using 20 structured content and metadata criteria.

2. **Real-World Validation**
   Beyond controlled testing, we analyzed real insurance-related queries to observe how LLMs retrieve and prioritize content. This helped connect retrieval outcomes directly to practical content structures and informed the development of LLM Optimization (LLMO) strategies.

3. **Quantitative and Qualitative Analysis**
   We combined statistical evaluation (retrieval volumes, ranking patterns) with qualitative content assessments (semantic structure, formatting quality, metadata alignment) to understand not only which content surfaces but also why.

Together, these components ensured that the analysis remained both evidence-based and actionable. The next section shows how this framework was operationalized through a structured, prompt-based evaluation process.

Prompt Set-up and Comparative Evaluation

To enhance the precision of our analysis and establish a robust comparative baseline, we integrated traditional Google search terms into our research methodology. This allowed us to directly compare how LLMs perform against Google Search when processing similar insurance-related queries, providing insights that are both academically rigorous and business-relevant.

To structure the analysis, we used 18 representative search terms across three insurance product categories: home contents, dental supplementary, and legal protection. Each category included both brand-specific terms (e.g., "ERGO dental insurance") and brand-open terms (e.g., "best dental insurance") to reflect different stages of the customer journey. This structure ensured that both intent-driven exploration and concrete decision-making contexts were captured in the evaluation.

These search terms served as the foundation for a structured prompt set of 120 inputs, systematically varied by:

- Funnel stage: 78 upper-funnel (exploratory) and 42 lower-funnel (decision-focused) prompts
- Brand context: Balanced 50/50 between brand-open and brand-specific phrasing
- Complexity: 90 simple, everyday prompts and 30 moderately complex inputs involving com-parisons or constraints

This structure reflects real-world search behavior: most insurance-related queries are exploratory in nature, phrased in simple, natural language, and often influenced by brand awareness—particularly in trust-sensitive contexts (Spatharioti et al., 2023; Chowdhury et al., 2024; Nestaas et al., 2024; Google Consumer Insights, 2024). Because prompts reflect how people actually search, the visibility outcomes they produce reveal how well different engines respond to real information needs. By standardizing prompt structure, we can observe which content types of surfaces consistently, which models prioritize clarity or credibility, and which reward specific formatting choices.

To establish a baseline for comparison, all 18 Google search terms were submitted ten times, resulting in over 5,800 distinct URLs. Retrieved results were categorized as:

- Google-only URLs (not retrieved by LLMs)
- Shared URLs (retrieved by both systems)
- LLM-only URLs (retrieved exclusively via LLMs)

This structure mirrors real-world search behavior, covering the full customer journey, from initial interest to final decision, and incorporating variations in brand familiarity and language complexity. It provides a realistic, data-driven model to evaluate how effectively LLMs retrieve and rank insurance content.

Key Figures At A Glance

3 insurance products analyzed across home contents, dental supplemental, and legal protection

120 prompts (40 per product), covering different funnel stages and complexity levels

18 Google search terms (6 per product), used as traditional search benchmarks

120 prompts and 18 search terms, each submitted 10 times to SearchGPT, Perplexity, You.com, and Gemini

33,600 URLs retrieved across all platforms

25,441 LLM-based URLs

5,920 Google-only URLs

2,074 hallucinated URLs removed after quality filtering

900,909 data points generated through a combination of manual and agent-assisted content labeling and evaluation

With a controlled prompt set in place, we could now observe how different systems responded, both in terms of what they surfaced and how often. The next step was to collect, compare, and filter thousands of results across LLMs and Google to identify where visibility is shifting, and why. This made it possible to trace not just what was retrieved, but which types of content consistently gained prominence under realistic conditions.

Data Evaluation Methodology and Pre-Analysis

Across all four LLM systems and Google Search, a total of 33,366 URLs were generated. After removing duplicates and errors, 31,292 valid results remained. From the structured prompt set based on 18 initial Google search terms, LLMs retrieved 25,441 unique URLs. In comparison, direct searches of the 18 terms in Google returned 5,851 unique URLs. Additionally, 2,074 hallucinations, links to non-existent or irrelevant pages, were identified and excluded, corresponding to about 7% of all LLM outputs.

Each URL was classified according to source type (e.g., insurer websites, broker platforms, media outlets, private blogs), brand specificity, and domain structure. This enabled the identification of patterns in which source types and content formats are most frequently surfaced by LLMs. This cleanup ensured that further evaluation was based solely on verifiable and accessible content.

To explore structural patterns in more depth, we assessed a representative sample of 606 URLs using a 20-criteria framework. The framework captured four key dimensions: machine readability, semantic integrity, domain authority, and conversational formatting.

The evaluation methodology combined hypothesis-driven testing, structured scoring, and applied content analysis across four LLM systems and Google. A sample of 606 insurance-related URLs, normalized to 267 unique pages, was assessed using a mix of automated and rule-based scoring procedures aligned with 20 optimization criteria.

These criteria covered four core dimensions:

- Machine Readability: HTML cleanliness, mobile responsiveness, loading speed, and accessibility features.
- Semantic Content Linking: Internal structure, conceptual flow, use of headings, and logical link hierarchies.
- Source Trust and Verification: Institutional credibility, SSL certification, author attribution, and regulatory disclosures.
- Conversational Formatting: FAQ sections, modular response structures, and clearly scoped answer blocks.

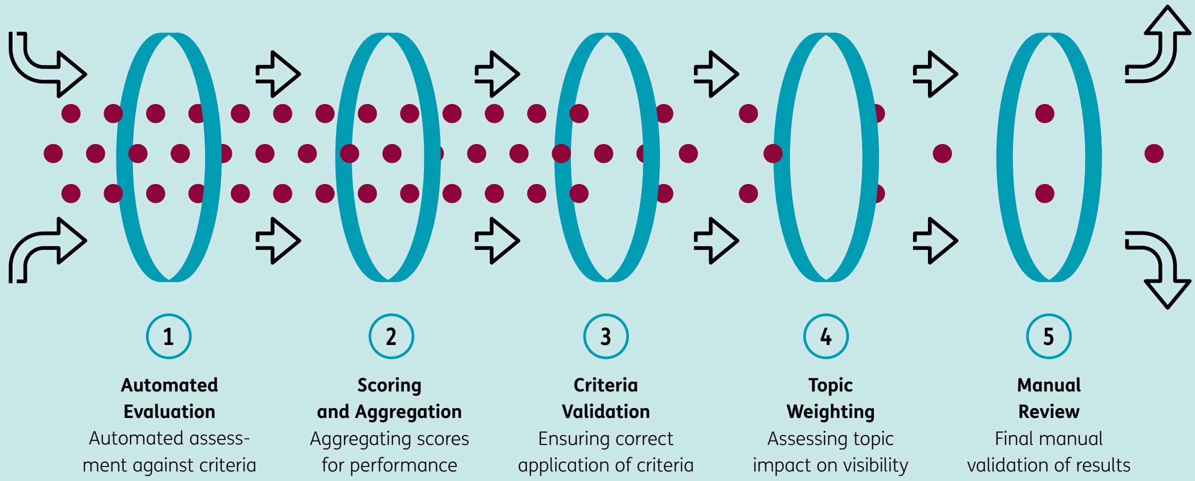
Additionally, eight meta-criteria captured deeper structural factors such as content hierarchy, duplication, internal navigation, and alignment with prompt intent.

The evaluation followed a triangulated research design: empirical hypothesis testing with a labeled dataset, design science to ensure practical relevance, and mixed-methods reasoning to link quantitative retrieval behavior (e.g., hallucination frequency, retrieval distribution) with qualitative content traits. To ensure balance and comparability, the sample was distributed across eight clusters, including brand-specific and brand-open domains, funnel-stage variations, and different platform types. This layered evaluation approach allowed us to examine not just what content LLMs retrieve, but how structure, format, and source characteristics influence visibility in AI-driven search environments.

Semantic Retrieval Diagnostics in LLMs

In addition to large-scale prompt testing and content scoring, we conducted a diagnostic layer focused on how LLMs retrieve information at the semantic level. Unlike traditional keyword-based search engines, LLMs operate

Evaluation Framework for Retrieved URLs



within high-dimensional vector spaces, matching queries and content based on semantic proximity rather than exact phrasing.

To explore this mechanism, we examined how semantically structured content performs in LLM retrieval, focusing on three key aspects:
- Retrieval consistency: whether the same pages were repeatedly retrieved across prompts
- Clustering behavior: whether related content types appeared together across systems
- Semantic cohesion: whether internally linked content had higher retrieval stability

These indicators helped assess how well LLMs detect meaning across pages and how semantic structure influences discoverability. Although we did not conduct full latent space modeling, we visually compared fragmented versus coherent embedding patterns. We found that content with consistent terminology, clear topic hierarchy, and internal navigation was retrieved more frequently and reliably across systems.

This diagnostic confirms a key insight: content that is not only technically accessible but also semantically aligned significantly increases its chances of being retrieved in LLM-driven environments.

Process of the URL Evaluation Methodology

To assess how well insurance content performs in LLM search, each URL in our dataset was reviewed using a structured scoring model. The process combined automated checks, rule-based criteria, and targeted manual review. This five-step framework provided the foundation for our findings. It allowed us to isolate which content attributes matter most and why certain types of insurance pages consistently outperform others in LLM-driven environments.

Study Findings on LLM Visibility in Insurance

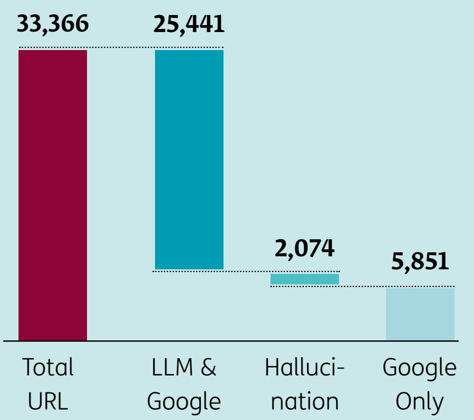
Large language models don't just retrieve what's popular. They retrieve what's readable, structured, and semantically clear. Our evaluation reveals how these systems actually select and prioritize insurance content. From hallucinations to hidden gems, the data shows what makes content visible and what leaves it behind.

Visibility shift across systems

Our comparison between Google and LLM-based search engines reveals fundamental differences in how visibility is assigned. This shift changes the way insurers must design and structure their content.

Distribution of Additive URLs by LLM Search Engine

Percentage & Absolute Values with Hallucination



LLM platforms retrieved significantly more URLs than Google, but also showed higher variability in quality. At the same time, patterns emerged that clearly favored structured, machine-readable content, regardless of brand recognition. These findings highlight that discoverability in AI-powered search depends less on traditional SEO and more on how content aligns with the retrieval logic of language models.

Google continues to lead as the most widely used search engine, including for insurance-related content. In our study, we tested 18 insurance-related search terms across three insurance products and submitted ten times each term, resulting in 180 total search attempts. This volume mirrors realistic user behavior, where multiple searches are often needed to find accurate and comprehensive information. This gave us a stable reference point for comparing how LLMs handle similar queries.

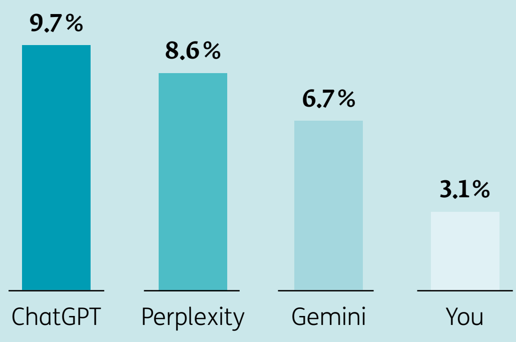
Through these attempts, we retrieved 5,851 unique URLs. Each search returned an average of 585 URLs, highlighting Google's breadth. However, the format remains static: users must sift through multiple links to locate relevant answers. Google's ranking logic, based on keyword matching and backlink strength, continues to shape which results are shown, favoring established domains and SEO-optimized content. This approach works well for general queries and recognized providers. But for users looking for specific product features or contextual comparisons, the need to click through multiple links often slows the process and increases search friction.

Accuracy and Hallucinations in LLM-Based Search

The shift in search behavior becomes clear when looking at LLM-powered engines. In our study, we tested 120 prompts across three insurance products and submitted each prompt ten times to four LLM systems. This pro-

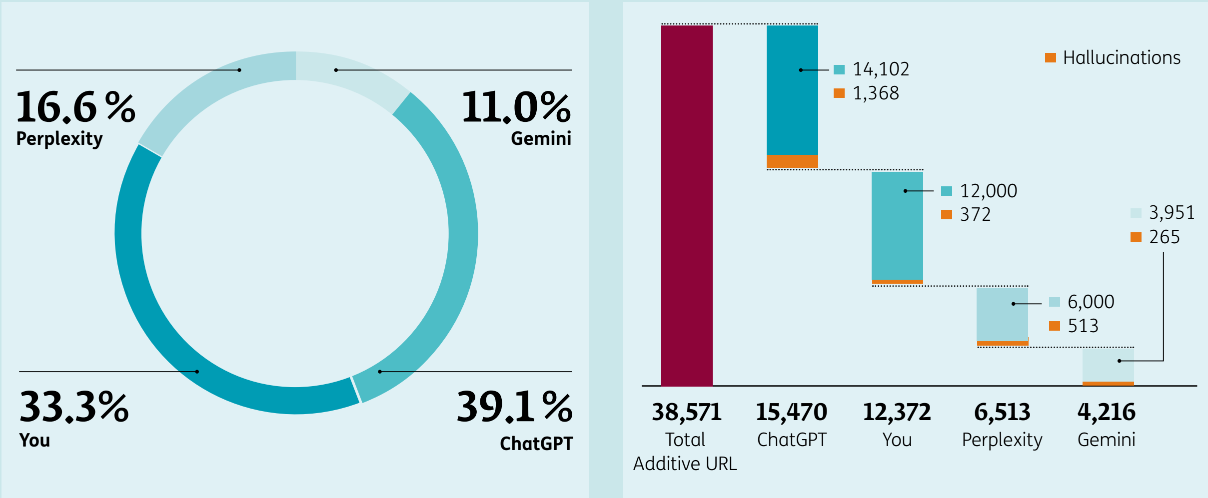
Hallucination rate: the proportion of incorrect or fabricated results generated by the AI

Relative to the total number of submitted URLs



Retrieval Distribution by Search Engine Type

Percentage & Absolute Values with Hallucinations



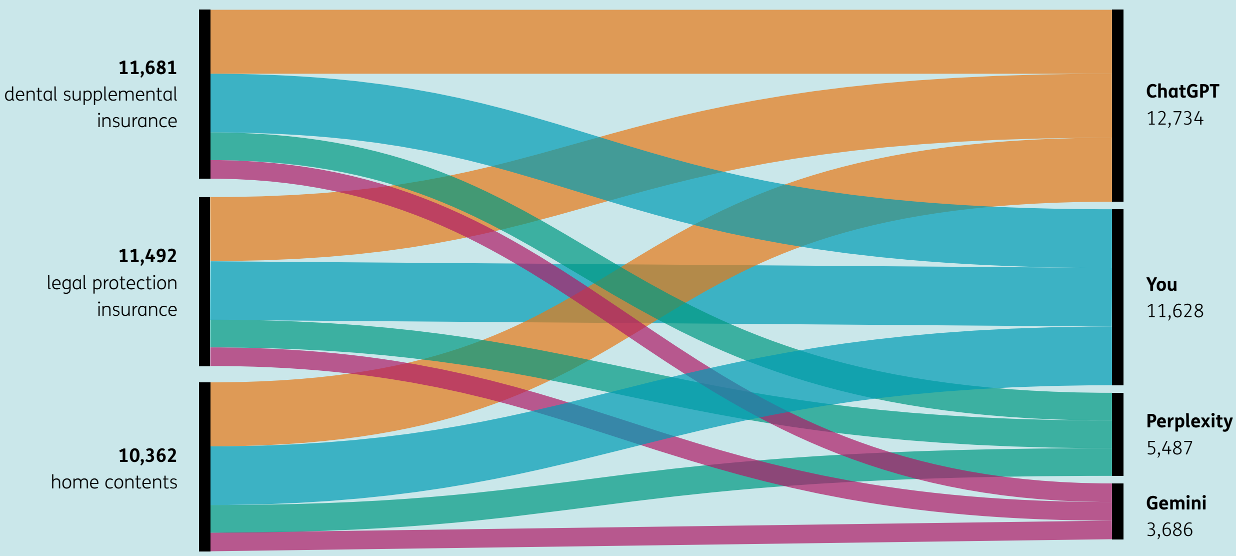
duced a total of 27,446 URLs. However, 2,074 of these URLs were classified as hallucinations, content that was incorrect, irrelevant, or non-existent. This equates to a hallucination rate of 7.6 percent.

Despite significant advances in AI-powered search, hallucinations remain a critical issue. ChatGPT had the highest hallucination rate at 9.7 percent, followed by Perplexity at 8.6 percent. Gemini showed a lower rate at 6.7 percent, while You.com performed best, with just 3.1 percent. These differences suggest that retrieval architecture directly impacts output stability. Systems with broader reach may offer more content but also carry greater risk. Engines with more conservative retrieval strategies produce fewer hallucinations but may require more precise content structuring.

For insurers, this underscores the need to validate content not only for accuracy but also for AI interpretability. Structured markup, institutional signals, and clear factual alignment reduce the likelihood of misinformation and help maintain trust, both from users and from the systems retrieving the content.

Platform-specific tendencies further illustrate these patterns. Gemini favors narrow sets of highly trusted sources, often with a strong presence of broker domains. Perplexity gives preference to editorial-style content from established publishers. SearchGPT balances semantic relevance with indexing of open-source data. You.com includes a broad range of content types, drawing from sources such as blogs, associations, and other non-traditional domains.

Distribution of Retrieved URLs by Product and LLM Engine



As users increasingly rely on multiple AI tools, insurers must adopt differentiated content strategies tailored to each engine's architecture and retrieval logic. A one-size-fits-all SEO approach is no longer effective.

LLM Coverage Depends on Insurance Product Line

The analysis of URL retrievals across three insurance product categories – dental supplemental insurance, legal protection insurance, and home contents insurance – reveals notable disparities in how different LLM-based search engines handle these sectors.

ChatGPT generated the highest overall volume, particularly excelling in dental supplemental insurance. You.com followed closely with balanced performance across all three lines. Perplexity retrieved significantly fewer URLs but showed strength in dental insurance. Gemini returned the lowest volume, especially in dental and home contents, indicating limitations in its indexing model or retrieval architecture.

These differences reveal how platform architecture influences not just overall retrieval rates, but also content coverage per product line, an important consideration for insurers seeking product-specific visibility.

Brokers Gain Visibility Advantage in LLM Search

Different types of content providers perform differently across platforms when handling brand-open insurance-related queries: Brokers account for the largest share at 36 percent, far ahead of insurers (17 percent) and publishers (16 percent). Comparers and associations follow with much smaller portions. These figures highlight a change not only in ranking outcomes but in the logic of how content is evaluated and surfaced.

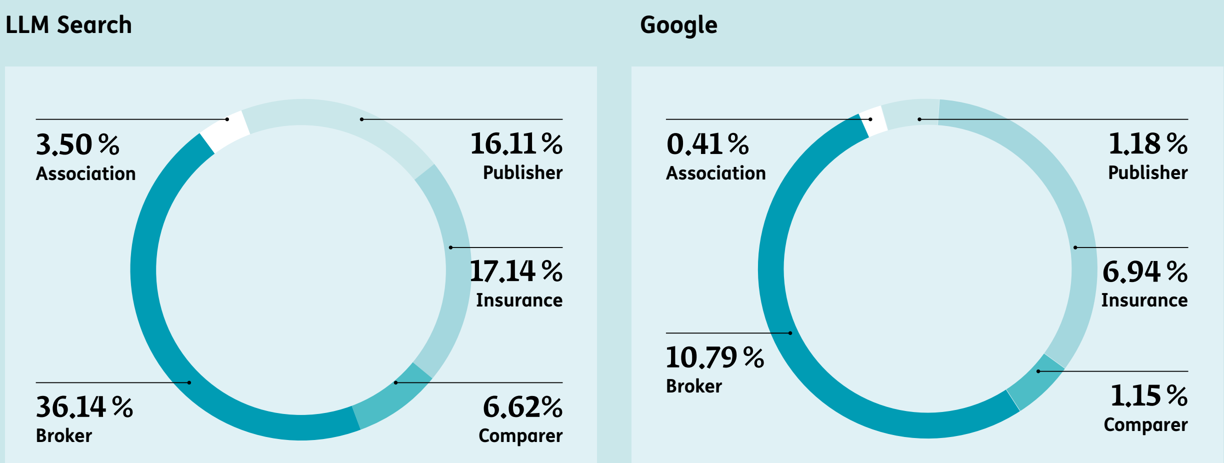
A critical distinction lies between two types of comparison platforms. Brokers, including major comparison platforms, can actively sell insurance policies and typically offer structured, scenario-based content designed for decision-making. Comparers, by contrast, focus solely on editorial or informational content and lack a brokerage license. This distinction helps explain visibility differences: broker platforms align better with the modular, semantically rich content formats that LLMs favor.

Google search paints a different picture. Broker visibility drops to under 11 percent, insurers remain at around 7 percent, and publishers and associations are nearly invisible. This suggests that traditional SEO signals, domain authority, backlinks, keyword density, continue to dominate. Google rewards technically optimized sites, even when their content is less dynamic or user-focused. In contrast, LLMs prioritize content that can be interpreted in context, adapted to conversation, and used to support decisions—content that is clean, linked, transparent, and easy to segment.

This divergence creates a visibility gap. Players with modular, structured content, especially brokers are pulling ahead in LLM environments. Insurers that don't adapt aren't losing because their content is weak, but because it's built for a system that no longer sets the rules. Maintaining visibility across both systems now requires parallel strategies: one optimized for Google's legacy infrastructure, the other designed for LLM-native discovery.

The data shows that in LLM-based environments, broker domains consistently achieve the highest visibility, especially across brand-open queries. Publisher content performs moderately well, while comparers trail behind. Direct insurers typically capture only 1 to 2 percent of results in these systems. LLMs seem to apply a systematic preference for content that is conversational, explorable,

Company Type Distribution: LLM Search vs. Google



and logically structured, qualities more common on broker and aggregator websites.

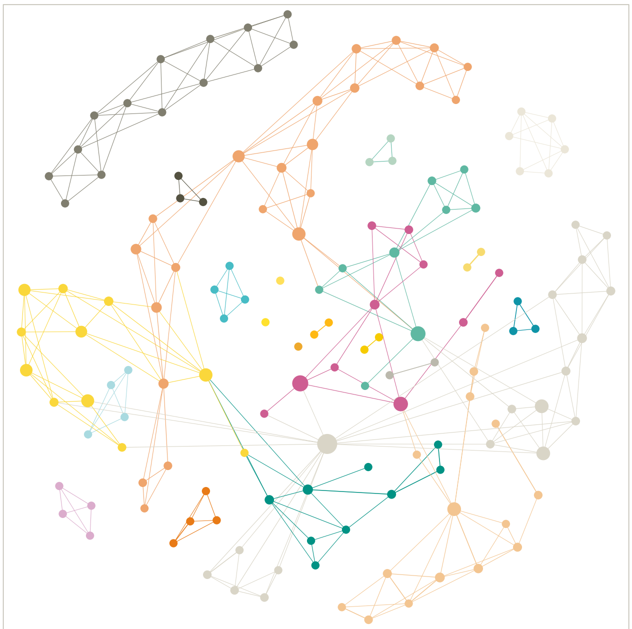
Google’s distribution is flatter but more rigid. Broker domains appear with around 2 percent, direct insurers close behind at 1 percent. The remaining results are scattered among publishers, comparers, and industry associations. The pattern reflects Google's continued dependence on ranking signals that LLMs increasingly bypass: technical setups, backlink networks, and authority metrics.

Brokers are structurally well-positioned to benefit from this shift. Their sites are built around product comparison, modular presentation logic, and decision-relevant pathways. These characteristics map closely to how LLMs generate responses, by breaking content into segments, evaluating context, and reassembling meaning into a usable output. As AI systems move toward agent-like behavior, handling research, filtering, and decision support in one flow, this structure becomes a key advantage.

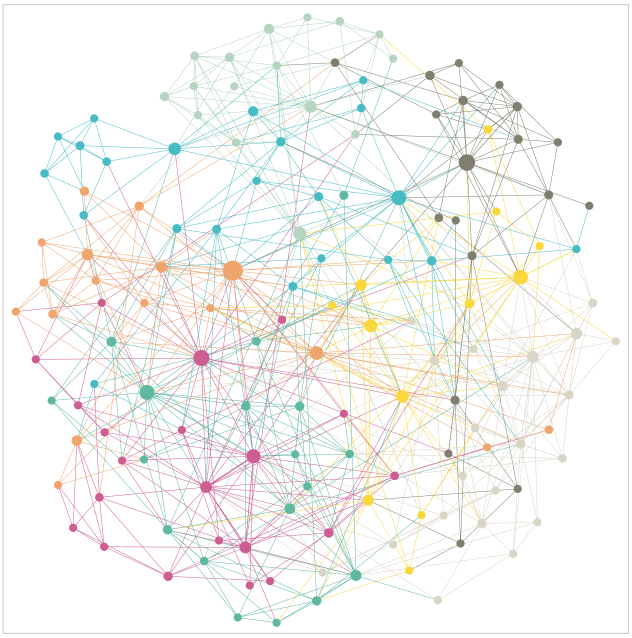
These findings reveal a structural shift in visibility. LLMs reward conversational, interconnected, and intentrich content. These characteristics are more often found on broker and aggregator sites. In contrast, Google continues to emphasize backlinks, domain authority, and keyword matching, which results in a flatter, but less context-aware visibility pattern.

Structural Patterns: Semantic Cohesion and Vector Space Analysis

A central finding of this study is the clear link between semantic structure and LLM visibility. Content that is internally coherent, well-linked, and conceptually consistent is significantly more likely to be retrieved and reused.



Semantic Fragmentation. Disconnected clusters in vector space represent semantically isolated or structurally weak content. These regions show low retrievability due to poor alignment with query vectors.



Semantic Cohesion. Dense, interconnected vector regions reflect high-quality internal linking, structured semantics, and stable embedding behavior. These clusters consistently demonstrate superior retrieval rates in LLM-based search.

Visualizations of embedding behavior illustrate that disconnected or weakly linked content leads to fragmented vector subspaces and lower retrievability.

By contrast, dense, interconnected vector clusters correlate with stable retrieval across systems. Structurally cohesive content, supported by taxonomies, semantic anchors, and logical linking, enables stronger alignment with LLM embedding models. The results reveal a clear dichotomy between fragmented and cohesive vector subspaces. This network visualization shows a fragmented vector space, where disconnected clusters and isolated nodes represent content that lacks internal linking or consistent terminology. These microsites or stand-alone pages tend to perform poorly across LLM platforms. Their structural gaps prevent alignment with embedding-based retrieval logic, leading to low retrievability (Wang et al., 2019; Borah et al., 2021).

These findings confirm that content architecture has direct performance implications. Pages that are technically clean but semantically disconnected remain difficult for LLMs to retrieve. Building strong semantic relationships across pages is therefore not a secondary enhancement, it is a foundational element of LLM-based visibility.

What Makes Insurance Content Perform in LLM Search: Hypothesis Validation

To understand what makes content visible in LLM-powered search, we developed a scoring model covering 20 optimization categories. Each category reflects a structural or semantic criterion relevant to AI-driven retrieval, including factors such as machine readability, trust signals, semantic cohesion, and prompt-style formatting.

The model allows us to systematically assess which features contribute most to high retrievability in LLM outputs.

Each URL in our dataset was evaluated based on how well it met each criterion. The evaluation was conducted in three levels of scoring granularity: using 25%, 10%, and 1% intervals. We ultimately chose 1% intervals to enable fine-grained differentiation between similar content types, a decision that proved especially useful given the relatively small variations in performance across high-quality domains. This approach allowed us to capture subtle but important distinctions in content structure and interpretability.

In addition to assigning scores (ranging from 0 to 100), we applied weighting factors to each criterion to reflect their relative influence on LLM retrieval. The weights, which together sum to 100%, allow us to distinguish between content attributes that are merely present and those that have the greatest impact on discoverability.

The analysis reveals clear performance gradients:

Machine-Readable Content as a Visibility Driver

The data confirms that technically accessible, machine-readable content plays a leading role in determining visibility within LLM-powered search. This category reached a maximum score of 88% and an average of 85% across all evaluated pages, making it one of the highest-performing dimensions. With an average model weight of 6%, it also ranks among the most influential factors in the scoring framework.

These results support hypothesis 1: LLMs show clear preference for semantically structured HTML, fast-loading page architectures, and transparent content layouts. Key technical features, such as ARIA tagging, mobile-first responsive design, and clearly labeled headings are no longer optional. They represent baseline requirements for content to be parsed, understood, and reused by language models.

This technical foundation is often a necessary precondition for additional content attributes, such as semantic linking or trustworthy source signaling. Insurers that fail to meet these accessibility and structure standards risk being excluded from LLM retrieval workflows, regardless of the overall quality or accuracy of their content.

Semantic Cohesion Enables Stronger Retrieval

Semantic and contextual linking has a strong and consistent influence on content retrieval in LLM-based search environments. Content categories optimized for internal structure including internal links, meta data hierarchies, and concept relationships scored between 72% and 78%, with an average of 75%. The corresponding model weight of 5% underlines its strategic role in visibility.

These findings validate hypothesis 2: Semantically coherent content forms more stable and connected clusters in vector embedding space, which significantly increases the likelihood of retrieval. In contrast, isolated or fragmented content, such as standalone landing pages without conceptual linking tends to remain invisible to LLMs.

Our analysis confirms that content embedded in semantically dense, structurally cohesive clusters is more frequently retrieved. Modern embedding architectures assess proximity in vector space, favoring conceptually aligned nodes when semantic consistency is present (Wang et al., 2019; Borah et al., 2021).

In this context, semantic structure is not a design enhancement, it is a technical enabler of discoverability. For insurers, this means that consistent taxonomies, clear internal anchors, and rich interlinking across related content are now essential to ensure retrievability across LLM platforms. Embedding integrity, conceptual continuity, and internal cohesion must be treated as foundational components of any content strategy intended to surface in AI-generated responses.

Trust Signals as a Primary Ranking Factor

Among all evaluated dimensions, trust-related content features performed the strongest. This category achieved a maximum visibility score of 90% and an average of 88%, the highest across all criteria. Its average weight of 6% further confirms its importance within the overall LLM ranking framework.

These results validate hypothesis 3: LLMs consistently favor content from trusted sources such as government portals, academic institutions, and regulated professional domains. Unlike traditional search systems that emphasize backlinks and domain size, LLMs rely on a broader range of trust indicators, including verified authorship, institutional consistency, and transparent citation practices.

For insurers, this means that digital reputation is a primary consideration. It must be actively cultivated across both technical and editorial dimensions. Managing domain authority, providing clear authorship, and adhering to transparent publishing standards now directly influence whether content is retrieved, ranked, and reused by AI systems. In an LLM-driven discovery environment, trust is not only a signal, it is a prerequisite for visibility.

Prompt-Style Structuring Aligns with LLM Patterns

Content structured in a prompt-like or conversational format, such as FAQs, Q&A pages, or segmented advice modules shows a clear retrieval advantage in LLM-based environments. This category scored between 67% and 75%, with an average of 72% and a model weight of 5%.

These findings confirm hypothesis 4: LLMs trained on dialogue-heavy and Q&A-style datasets tend to favor content that reflects this input-output logic. Particularly in insurance contexts involving product comparison, eligibility guidance, or service navigation, conversational structuring helps content align with how models reason and respond.

Organizations that do not adapt their content formats to this logic may remain underrepresented, even when the information is accurate and relevant. For insurers, adopting modular, prompt-aligned content structures is no longer optional, it is a strategic step toward ensuring inclusion in AI-generated responses

These findings validate our four hypotheses:

- **H1:** Technical accessibility and machine readability increase LLM visibility.
- **H2:** Semantic and contextual linking enhance retrievability.
- **H3:** Trusted, high-quality sources are more likely to be cited.
- **H4:** Prompt-aligned content structures improve inclusion in LLM outputs.

Together, these dimensions form a clear strategic roadmap. Insurers that focus on these high-impact attributes and measure performance with this level of granularity are better positioned to secure stable and consistent visibility across LLM platforms.

The four hypotheses represent overarching dimensions that group several of the 20 validated criteria. The average weights shown in the table reflect the combined influence of the individual criteria within each dimension. This highlights four out of 20 factors that have the strongest impact on visibility in AI-driven search.

Evaluation of Categories by Average Score and Weight

Hypothesis	Matched Category	Avg. Score	Avg. Weight
H1	Technical Accessibility & Machine Readability	85%	6%
H2	Semantic Linking Optimization	75%	5%
H3	Authenticity & Trustworthiness	88%	6%
H4	Prompt-Style LLM Optimization	72%	5%

Strategic Implications for Insurance in the LLM Era

Search is no longer about being found, but about being usable by machines. As LLMs reshape digital visibility, traditional content strategies fall short. What comes next requires new structures, new signals, and a radical shift in how insurers build for discovery.

The New Logic of Visibility

The shift from keyword-matching to semantic retrieval redefines how content must be structured to be found. Visibility now depends on how well information aligns with the internal logic of language models.

A New Visibility Paradigm

The rise of language models (LLMs) fundamentally changes how users access and interact with information. Traditional SEO based on keywords, backlinks, and domain authority is no longer sufficient. LLMs do not match exact terms, they evaluate semantic similarity within vector spaces. This changes the nature of visibility: it is now determined on how effectively content can be parsed, interpreted, and reused by language models operating in high-dimensional vector spaces.

Technical Foundations of Discoverability

LLMs retrieve information not by matching keywords, but by evaluating semantic similarity using vector embeddings. This requires content to be machine-readable, semantically structured, and technically accessible. Clean HTML, accessible markup, fast-loading architecture, and well-applied metadata provide the foundation for visibility. Without these basics, even high-quality content will remain invisible.

To align with LLM training patterns, content should be modular, interpretable, and clearly written using consistent terminology and logical formatting. Effective formats include Q&A sections, Wikipedia-style summaries, and use-case-based content blocks. Semantic coherence across a site via shared vocabulary, topic linking, and well-defined taxonomy, supports the creation of content clusters interpretable by models.

A hypothesis-driven scoring model confirms a clear hierarchy of visibility drivers in LLM environments: machine readability as a prerequisite, semantic linking for embedding consistency, trust signals such as authorship and credibility, and modular, prompt-aligned formats such as Q&A sections, bullet points, and structured scenarios. Content that incorporates these elements doesn't just get indexed, it becomes part of the model's reasoning process.

Strategic Levers for LLM-Based Visibility

Being retrievable demands semantic clarity, structural consistency, and embedded trust. These elements form the core levers of LLM-oriented visibility.

Trust, Brand Structure, and Metadata

LLMs increasingly prioritize content that is transparent, well-sourced, and reputationally credible. Trust signals, like clear authorship, editorial accountability, and references to regulatory bodies, are essential. Brand identity must also be technically embedded. This includes consistent naming and messaging, the use of metadata standards like schema.org or JSON-LD, and advanced techniques such as knowledge graphs or synthetic training data.

Closing Vector Gaps

Sustaining visibility requires addressing so-called vector gaps, cases where content fails to align with relevant queries due to vague wording, incomplete explanation, or weak semantic structure. These gaps can be closed by monitoring LLM outputs, refining phrasing, strengthening semantic linkages, and optimizing internal content taxonomies.

Performance Monitoring and Operational Integration

Once vector gaps are addressed, organizations must also track how content performs in live retrieval settings. Standard traffic metrics are no longer sufficient. Retrieval frequency, inclusion in LLM-generated responses, consistency of brand mentions, and presence in multiturn interactions now matter. Prompt testing across platforms and hallucination analysis help identify structural content weaknesses. These insights must feed directly into content operations and iterative optimization processes. These insights must feed directly into content operations and iterative optimization processes, ensuring that structural weaknesses are addressed in publishing workflows, content templates, and ongoing updates.

Platform and Competitive Differentiators

Not all LLMs interpret content in the same way, and not all market players are equally prepared. Structural alignment and platform-specific adaptation create competitive advantage in this emerging landscape.

LLM- & Platform-Specific Optimization

A key insight from the study is the variation in performance across different LLM platforms. ChatGPT and You.com retrieved large volumes of URLs but also showed higher hallucination rates. Gemini and Perplexity delivered more curated, source-filtered results. Each platform responds to different content signals. There is no universal optimization strategy that applies across all systems. Visibility must be approached on a per-engine basis, supported by structured testing and content iteration. Insurers need to understand how their offerings perform within each platform and adapt accordingly.

Structurally Aligned Content Outperforms

Regardless of platform, certain structural patterns consistently deliver higher retrievability. Technically optimized and semantically structured content performs significantly better across all tested platforms. Pages built around modular formats such as FAQs, structured coverage explanations, or clearly defined answer blocks are retrieved more frequently and more reliably than long-form, brand-centric narratives. This is not a matter of style, but of alignment with how large language models parse, rank, and reconstruct content during response generation.

Why Brokers Outperform

One of the most consistent outcomes of the study is the superior performance of brokers and aggregators. Their platforms show higher visibility across all LLM systems and product categories. This is largely due to how their content is designed: comparison-oriented, clearly segmented, and densely linked. These structural features are closely aligned with LLM retrieval behavior. In contrast, many insurer websites rely on unstructured text, complex navigation, or incomplete markup, making them hard to retrieve. Their content is often optimized for brand communication or regulatory clarity, but not yet fully aligned with how AI systems access and assemble information. Without structural change, insurers will continue to lose reach to better-adapted intermediaries — not because the content lacks quality, but because its presentation format does not yet reflect the logic of LLM-based discovery. Broker platforms typically present insurance content in decision-ready formats — with product filters, scenario-based recommendations, and dynamically generated overviews. This allows language models to extract concise, relevant segments that map well to user queries.

Operational Readiness for LLM Integration

For content to be actionable within AI ecosystems, organizations must rethink capabilities, expose services through APIs, and close technical, strategic and organizational readiness gaps.

Emerging Agent-Based Journey

With the rise of agentic AI, new interaction models are emerging. LLM-powered agents are beginning to move beyond information delivery toward action execution. These systems are designed to compare products, generate real-time quotes, or complete tasks autonomously. Currently, insurance services in Germany are not compatible with agentic systems. APIs for quote delivery, policy information, or pricing logic are either unavailable or inaccessible to third-party models. As a result, insurers remain outside of emerging agent-based user journeys. This is a significant gap and addressing it will require both technical readiness and regulatory alignment.

Strategic Inflection Point for Insurers

Large language models are reshaping how services are found and used. This presents a strategic turning point for the insurance industry. Brokers are structurally better positioned, and agentic systems will further widen the gap between retrievable and invisible content. Yet the drivers of visibility are no longer speculative, they are observable, measurable, and actionable.

Insurers who expose their services via APIs and build modular, LLM-aligned content architectures can regain visibility and shape their role in AI-driven interactions. Those who delay risk losing control over how, when, and whether their products are presented to users. The transition is not a future scenario, it is already underway.

Strategic Response to Changing Visibility Logic

The study confirms a structural transformation in how insurance content is discovered, retrieved, and re-used by AI-based systems. Traditional SEO signals such as keyword placement, backlinks, and domain authority are no longer sufficient for ensuring digital visibility. Instead, discoverability is increasingly governed by how well content aligns with the retrieval logic of large language models. This includes machine readability, semantic coherence, trustworthiness, and conversational formatting.

Organisational Capabilities and Interdisciplinary Collaboration

Meeting these new requirements will not be possible through marketing or SEO alone. LLM visibility must become a shared responsibility across content, engineering, compliance, and product teams. Companies will need to build internal processes for prompt testing, semantic scoring, retrieval monitoring, and format experimentation. This includes redesigning core pages, aligning metadata with AI models, and enabling structured endpoints for services that need to be surfaced or acted upon.

Finally, internal capabilities must evolve. Most marketing and SEO teams were designed for Google's ranking logic. While that knowledge remains valuable, it must now be augmented with an understanding of how LLMs retrieve, interpret, and assemble content. That includes knowledge of vector semantics, embedding space behavior, prompt engineering, and AI-specific performance metrics. This doesn't require building new departments from scratch, but it does require upskilling, cross-functional collaboration and targeted training.

In many cases, this shift will build on existing strengths, such as structured content workflows, compliance know-how, or established taxonomies, but it requires translating those strengths into formats and signals that modern language models can effectively interpret.

Organizations that treat LLM retrievability as a strategic design question, not just a technical one, will be best positioned to remain visible, relevant, and competitive in the years ahead.

From Visibility to Retrievability

Success in AI-driven visibility will come from interdisciplinary coordination. Editorial, technical, and strategic teams must align around shared taxonomies, content models, and a new definition of performance. In this environment, visibility depends not only on content quality, but on how that content is structured, linked, and interpreted by machines.

Marketing is no longer just about visibility, it's about retrievability. The organizations that understand how LLMs read, rank, and repurpose content will define the next generation of digital engagement. Now is the time to build those capabilities from the inside out.

Outlook: The Future of LLM Discovery

The future of digital discovery is being reshaped beyond traditional search. As LLMs evolve, new forces like autonomous agents, innovative monetization, and curated content partnerships are redefining how information is accessed and surfaced. These emerging dynamics demand that organizations rethink their strategies and infrastructure to stay relevant and visible in an AI-driven landscape.

While this study focused on the structural foundations of LLM visibility, new dynamics are already reshaping the landscape—driven by evolving interaction models, emerging monetization strategies, and increasing platform control. Although not part of the empirical analysis, the following developments are essential for organizations looking to stay ahead of the curve.

Strategic Content Collaborations

In parallel, the content distribution model itself is changing. The old model of open web indexing is being replaced by curated integration. Leading language model platforms such as OpenAI and Mistral have begun to form direct alliances with publishers and data providers, exemplified by deals with Axel Springer or backing from conglomerates like Bertelsmann. These relationships go beyond crawling public content. They define which data sets are considered high-quality, trusted, and available for integration into model outputs.

As a result, visibility is increasingly negotiated, not earned. Companies that depend on organic traffic or visibility through traditional SEO must reassess their distribution strategy. Being indexed is no longer enough. To remain part of the LLM-powered answer layer, brands must explore data licensing agreements, direct integrations, or partnerships that allow their content to be part of the prompt stack or reasoning layer. This creates a more fragmented but strategically controlled discovery ecosystem, where business development and content integration are as important as optimization.

Ads in LLMs: Merchant Program and Ads

As LLMs replace or augment traditional search engines, new monetization models are emerging. Platforms like Perplexity are experimenting with embedded advertising formats, including banner-style visuals and merchant programs integrated directly into conversational answers (Perplexity AI, 2024; Bastian, 2024), shifting paid exposure away from result lists and into the model's generative output itself.

Unlike traditional keyword-based ads, these formats emphasize contextual relevance, intent alignment, and user experience. In a world where the LLM not only answers questions but constructs narratives, advertising must adapt. The challenge for brands is to develop new creative and targeting strategies that allow their products and messages to appear naturally within AI-generated content. This may involve testing early-stage ad offerings from emerging LLM platforms, rethinking how product data is structured for semantic integration, or working with agencies that specialize in prompt engineering and AI-native media. The goal is not just visibility, but relevance – ensuring that brand messages are not only seen but seamlessly integrated into the information flows user's trust.

Agentic AI and the Future of Insurance

Agentic AI is changing how insurance products are found and purchased. Success will depend less on direct contact with customers and more on how easily AI systems can discover, understand, and complete transactions. In this new world, being visible means having clear, high-quality data and easy-to-use digital interfaces.

Old-fashioned, fixed product models and straight sales funnels won't work anymore. Insurers need to design flexible, modular products that AI can tailor in real time to fit individual needs. Pricing will need to adjust based on events and changing risks. Human advisors will take on new roles, stepping in only when AI needs help. Companies that work smoothly with AI platforms will have a clear advantage.

Agentic AI will turn insurance from reactive service into ongoing, proactive management. AI agents will find the right products, screen customers, handle claims, and spot gaps in coverage—making the process continuous rather than one-time. To keep up, insurers must not only open their data via APIs but also align their internal systems to work seamlessly with AI decision-making.

Preparing for What Comes Next

As AI systems continue to evolve, companies must look beyond traditional optimization and begin aligning their infrastructure, partnerships, and data strategies with the new logic of discoverability. The trends outlined here represent the next wave of disruption and the next opportunity. Organizations that act now will define how they are found, interpreted, and used by machines in the years ahead.

Authors

Luisa-Marie Schmolke
Innovation Developer
ERGO Group AG

Hamidreza Hosseini
Founder & CEO
ECODYNAMICS GmbH

Acknowledgement

This white paper was created thanks to the valuable support, constructive ideas and professional feedback of numerous other contributors. We would therefore like to express our sincere thanks to all those who have made a significant contribution to the creation and improvement of this white paper through their valuable technical input, careful proofreading and constructive feedback.

Hendrik Schulze Bröring (Senior Partner at ECODYNAMICS GmbH) led the data infrastructure and collection. This work formed the data foundation and enabled the insights presented in this paper.

Special thanks to

Georgina Neitzel
Head of ERGO Innovation Lab
ERGO Group AG

Larissa Selzer
SEO Expert
ERGO Group AG

Bibliography

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Bastian, M. (2024, November 14). Perplexity AI launches ad program with sponsored follow-up questions. THE DECODER. https://the-decoder.com/perplexity-ai-launches-ad-program-with-sponsored-follow-up-questions/

Borah, A., Barman, M. P., & Awekar, A. (2021, August). Are word embedding methods stable and should we care about it? In Proceedings of the 32nd ACM Conference on Hypertext and social media (pp. 45-55).

Borgesius, F. Z., van Bekkum, M., van Ooijen, I., Schaap, G., Harbers, M., & Timan, T. (2025). Discrimination and AI in insurance: what do people find fair? Results from a survey. arXiv preprint arXiv:2501.12897.

Chowdhury, G., & Chowdhury, S. (2024). AI-and LLM-driven search tools: A paradigm shift in information access for education and research. Journal of Information Science, 01655515241284046.

Creswell, J. W., & Clark, V. L. P. (2017). Designing and conducting mixed methods research. Sage publications.

Denzin, N. K. (2017). The research act: A theoretical introduction to sociological methods. Routledge.

Eisenbrand, R. (2025, March 5). “Übersicht mit KI”: Google testet AI Overviews in Europa – auch in Deutschland. OMR. https://omr.com/de/daily/google-ai-overviews-uebersicht-mit-ki-deutschland

Eleni Spatharioti, S., Rothschild, D. M., Goldstein, D. G., & Hofman, J. M. (2023). Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment. arXiv e-prints, arXiv-2307

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., ... & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. Computational Linguistics, 50(3), 1097-1179.

Gartner, Inc. (2023). Predicts 2024: How GenAI will reshape tech marketing. https://www.gartner.com/en/doc/800771-predicts-2024-how-genai-will-reshape-tech-marketing

GDV. (2024, July 31). Digitaler Versicherungsvertrieb wächst deutlich. GDV. https://www.gdv.de/gdv/medien/medieninformationen/versicherung-vertrieb-abschluesse-digital-181036

Google. (n.d.). How Google Search Algorithms Works. https://www.google.com/intl/en_us/search/howsearchworks/how-search-works/ranking-results/

Google Search Central. (n.d.). Google for Developers. https://developers.google.com/search/docs/appearance/page-experience

Hagey, K., Kruppa, M., Bruell, A., & iStock, E. L. W. S. J. (2023, December 14). News publishers see Google’s AI search tool as a Traffic-Destroying nightmare. WSJ. https://www.wsj.com/tech/ai/news-publishers-see-googles-ai-search-tool-as-a-traffic-destroying-nightmare-52154074

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS quarterly, 75-105.

Iqbal, M., Khalid, M. N., Manzoor, A., Abid, M. M., & Shaikh, N. A. (2022). Search engine optimization (seo): A study of important key factors in achieving a better search engine result page (serp) position. Sukkur IBA Journal of Computing and Mathematical Sciences, 6(1), 1-15.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature machine intelligence, 1(9), 389-399.

Joko, H., Chatterjee, S., Ramsay, A., De Vries, A. P., Dalton, J., & Hasibi, F. (2024, July). Doing personal laps: Llm-augmented dialogue construction for personalized multi-session conversational search. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 796-806).

Karamolegkou, A., Li, J., Zhou, L., & Søgaard, A. (2023). Copyright violations and large language models. arXiv preprint arXiv:2310.13771.

Karpukhin, V., Oguz, B., Min, S., Lewis, P. S., Wu, L., Edunov, S., ... & Yih, W. T. (2020, November). Dense Passage Retrieval for Open-Domain Question Answering. In EMNLP (1) (pp. 6769-6781).

Kumar, A., & Lakkaraju, H. (2024). Manipulating large language models to increase product visibility. arXiv preprint arXiv:2404.07981.

Li, J., Zhang, J., Li, H., & Shen, Y. (2024). An Agent Framework for Real-Time Financial Information Searching with Large Language Models. arXiv pre-print arXiv:2502.15684.

Liu, D., Chen, M., Lu, B., Jiang, H., Han, Z., Zhang, Q., ... & Qiu, L. (2024). Retrievalattention: Accelerating long-context llm inference via vector retrieval. arXiv preprint arXiv:2409.10516.

Liu, L., Meng, J., & Yang, Y. (2024). LLM technologies and information search. Journal of Economy and Technology, 2, 269-277

Liu, N. F., Zhang, T., & Liang, P. (2023). Evaluating verifiability in generative search engines. arXiv pre-print arXiv:2304.09848.

Liu, Y., Deng, G., Li, Y., Wang, K., Wang, Z., Wang, X., ... & Liu, Y. (2023). Prompt Injection attack against LLM-integrated Applications. arXiv preprint arXiv:2306.05499

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. Decision support systems, 15(4), 251-266.

Mialon, G., Dessì, R., Lomeli, M., Nalmpantis, C., Pasunuru, R., Raileanu, R., ... & Scialom, T. (2023). Augmented language models: a survey. arXiv pre-print arXiv:2302.07842.

Mo, F., Mao, K., Zhao, Z., Qian, H., Chen, H., Cheng, Y., ... & Nie, J. Y. (2024). A survey of conversational search. arXiv preprint arXiv:2410.15576.

Monir, S. S., Lau, I., Yang, S., & Zhao, D. (2024). VectorSearch: Enhancing Document Retrieval with Semantic Embeddings and Optimized Search. arXiv preprint arXiv:2409.17383.

Moz. (2024, November 25). What are On-Page ranking factors for SEO? Moz. https://moz.com/learn/seo/on-page-factors

Nakano, R., Hilton, J., Balaji, S., Wu, J., Ouyang, L., Kim, C., ... & Schulman, J. (2021). Webgpt: Browser-assisted question-answering with human feedback. arXiv preprint arXiv:2112.09332.

Nestaas, F., Debenedetti, E., & Tramèr, F. (2024). Adversarial search engine optimization for large language models. arXiv preprint arXiv:2406.18382.

Nie, G., Zhi, R., Yan, X., Du, Y., Zhang, X., Chen, J., ... & Hu, J. (2024, October). A hybrid multi-agent conversational recommender system with llm and search engine in e-commerce. In Proceedings of the 18th ACM Conference on Recommender Systems (pp. 745-747).

Pan, R., Cao, B., Lin, H., Han, X., Zheng, J., Wang, S., ... & Sun, L. (2024). Not all contexts are equal: Teaching llms credibility-aware generation. arXiv pre-print arXiv:2404.06809.

Perplexity AI. (2024, November 12). Why we're experimenting with advertising. Perplexity AI. https://www.perplexity.ai/hub/blog/why-we-re-experimenting-with-advertising

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Hambro, E., ... & Scialom, T. (2023). Tool-former: Language models can teach themselves to use tools. Advances in Neural Information Processing Systems, 36, 68539-68551.

Schneider, P., Poelman, W., Rovatsos, M., & Matthes, F. (2024). Engineering conversational search systems: A review of applications, architectures, and functional components. arXiv preprint arXiv:2407.00997.

Schwartz, B. (2025, March 20). Microsoft confirms Schema helps its LLMS (Copilot) understand your content. Search Engine Roundtable. https://www.seroundtable.com/schema-llms-copilot-bing-microsoft-39093.html

Sharma, N., Liao, Q. V., & Xiao, Z. (2024, May). Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-17).

Shi, X., Liu, J., Liu, Y., Cheng, Q., & Lu, W. (2025). Know where to go: Make LLM a relevant, responsible, and trustworthy searchers. Decision Support Systems, 188, 114354

StatCounter. (2025, January). Marktanteile der Suchmaschinen und Marktanteile der mobilen Suche in Deutschland. Cited in Statista. Retrieved January 15, 2025 from https://de.statista.com/statistik/daten/studie/301012/umfrage/marktanteile-der-suchmaschinen-und-marktanteile-mobile-suche/

Terenteva, E. (2023, July 18). Crawlability & Indexability: What they are & How they affect SEO. Semrush Blog. https://www.semrush.com/blog/what-are-crawlability-and-indexability-of-a-website/

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97–112.

Wang, B., Li, Q., Melucci, M., & Song, D. (2019, May). Semantic Hilbert space for text representation learning. In The World Wide Web Conference (pp. 3293-3299).

Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2024, January). Large search model: Redefining search stack in the era of llms. In ACM SIGIR Forum (Vol. 57, No. 2, pp. 1-16). New York, NY, USA: ACM.

Wang, Z., Xu, Z., Srikumar, V., & Ai, Q. (2024, May). An in-depth investigation of user response simulation for conversational search. In Proceedings of the ACM Web Conference 2024 (pp. 1407-1418).

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.

Werner, T., Soraperra, I., Calvano, E., Parkes, D. C., & Rahwan, I. (2024). Experimental evidence that conversational artificial intelligence can steer consumer behavior without detection. arXiv preprint arXiv:2409.12143.

Xiong, H., Bian, J., Li, Y., Li, X., Du, M., Wang, S., ... & Helal, S. (2024). When search engine services meet large language models: visions and challenges. IEEE Transactions on Services Computing.

Yan, L., Liu, Y., & Liu, S. (2024). The Search for Balance: The Impact of LLM-based Conversational Search on Information Processing and Polarization.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023, January). React: Synergizing reasoning and acting in language models. In International Conference on Learning Representations (ICLR).

Yoon, C., Kim, G., Jeon, B., Kim, S., Jo, Y., & Kang, J. (2024). Ask Optimal Questions: Aligning Large Language Models with Retriever's Preference in Conversational Search. arXiv preprint arXiv:2402.11827.

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., ... & Shi, S. (2023). Siren's song in the AI ocean: a survey on hallucination in large language models. arXiv pre-print arXiv:2309.01219.