Review

Not peer-reviewed version

The Evolution of SEO in the Era of Generative AI: A Business Intelligence Perspective

Konstantinos I. Roumeliotis * , Dionisis Margaris * , Dimitris Spiliotopoulos * , Costas Vassilakis *

Posted Date: 10 July 2026

doi: 10.20944/preprints202607.0744.v1

Keywords: generative AI; Search Engine Optimization (SEO); Business Intelligence (BI); Generative Engine Optimization (GEO); Search Generative Experience (SGE); Large Language Models (LLMs); Explainable AI; digital marketing strategy; integrative literature review; information systems

doi:10.20944/preprints202607.0744.v1

Review

The Evolution of SEO in the Era of Generative AI: A Business Intelligence Perspective

Konstantinos I. Roumeliotis [1,2,*], Dionisis Margaris [3,*], Dimitris Spiliotopoulos [2,*] and Costas Vassilakis [1,*]

1 Department of Informatics and Telecommunications, University of the Peloponnese, 22 131 Tripoli, Greece
2 Department of Management Science and Technology, University of the Peloponnese, Sehi Location (Former 4th Shooting Range), 221 31 Tripoli, Greece
3 Department of Digital Systems, University of the Peloponnese, Valioti's Building, Kladas, 231 00 Sparta, Greece
* Correspondence: k.roumeliotis@uop.gr (K.I.R.); margaris@uop.gr (D.M.); dspiliot@uop.gr (D.S.); costas@uop.gr (C.V.)

Abstract

The rapid paradigm shift from traditional keyword-matching algorithms to AI-driven answer engines has fundamentally disrupted Search Engine Optimization (SEO). As Large Language Models (LLMs) power modern Search Generative Experiences (SGE), organizations must transition from legacy web analytics to sophisticated Business Intelligence (BI) frameworks to capture visibility. Despite the immense strategic implications of this shift, academic literature remains fragmented across computer science, information systems, and digital marketing management. To bridge this gap, this paper adopts an integrative literature review methodology. Rather than utilizing restrictive systematic protocols (e.g., PRISMA) that isolate empirical data within narrow boundaries, the integrative approach enables a holistic synthesis of emerging, multi-disciplinary concepts necessary to decode a rapidly evolving phenomenon. Through this methodological lens, this study introduces the Signal-Structure-Surface-Score (4S) lifecycle framework, illustrating how AI-BI systems capture conversational search intents (Signal), architect machine-readable, entity-based data (Structure), optimize content for LLM retrieval and Generative Engine Optimization (Surface), and define novel attribution metrics for zero-click environments (Score). Furthermore, the paper maps the critical technical and strategic landscape, systematically evaluating prevailing trends (e.g., zero-click searches, AI-generated content velocity), core organizational challenges (e.g., search data attribution loss, algorithmic opacity), and emerging strategic opportunities (e.g., real-time intent mapping, competitor LLM audit trails). Ultimately, this paper bridges the gap between AI search mechanics and strategic BI measurement, providing a robust future research agenda designed to guide scholars and practitioners in navigating data-driven visibility in the age of generative search.

Keywords: generative AI; Search Engine Optimization (SEO); Business Intelligence (BI); Generative Engine Optimization (GEO); Search Generative Experience (SGE); Large Language Models (LLMs); Explainable AI; digital marketing strategy; integrative literature review; information systems

---

1. Introduction

Search Engine Optimization (SEO) has long served as the backbone of digital visibility, enabling organizations to connect with audiences through keyword-driven search algorithms. For over two decades, the fundamental mechanics of SEO—keyword matching, backlink accumulation, and on-page optimization—remained relatively stable, allowing marketers to develop reproducible strategies for organic ranking [1–4]. However, the integration of Large Language Models (LLMs) into search infrastructure has catalyzed a paradigm shift of unprecedented scale. Google's Search Generative Experience (SGE), OpenAI's ChatGPT, Perplexity AI, and similar generative search platforms now synthesize information directly within the search interface, producing conversational answers that

 doi:10.20944/preprints202607.0744.v1

reduce the need for users to click through to external websites [5,6]. This transition from "Retrieve-and-Rank" to "Retrieve-and-Synthesize" architectures [6] fundamentally disrupts the economic model of organic search, where visibility, traffic, and attribution are no longer guaranteed by traditional ranking positions.

The implications for Business Intelligence (BI) are profound. Organizations that once relied on web analytics dashboards to track keyword performance, click-through rates, and conversion funnels now face a "zero-click" environment in which search engines retain users within AI-generated answer panels [7]. The loss of referral traffic data creates attribution gaps that legacy BI frameworks were never designed to handle [8]. Moreover, the rise of Generative Engine Optimization (GEO)—the practice of optimizing content for retrieval and citation by LLM-powered answer engines—introduces an entirely new layer of technical and strategic complexity [6,9]. Content must now be structured not only for human readability and crawler indexing but also for machine-readable entity extraction, semantic clarity, and LLM citation worthiness [10,11]. Furthermore, the opacity of generative search systems creates an urgent need for Explainable AI (XAI) approaches that can make ranking and citation decisions interpretable to practitioners and stakeholders alike.

Despite the immense strategic implications, academic literature on this transformation remains fragmented. Computer science research has advanced rapidly in information retrieval (IR) and retrieval-augmented generation (RAG) [12–15], while marketing and information systems literatures have explored AI-driven personalization and digital marketing automation [16–18]. Yet these streams rarely intersect, leaving a critical gap between the technical mechanics of AI search and the strategic BI frameworks needed to measure and manage visibility in generative search environments.

To bridge this gap, this paper adopts an integrative literature review methodology, synthesizing 70 high-value studies selected from an initial corpus of over 2,900 verified academic papers. The integrative approach, as articulated by Snyder [19], enables a holistic synthesis of emerging, multi-disciplinary concepts that restrictive systematic protocols (e.g., PRISMA) would fragment. Through this methodological lens, this study introduces the **Signal-Structure-Surface-Score (4S) lifecycle framework**, illustrating how AI-BI systems capture conversational search intents (Signal), architect machine-readable, entity-based data (Structure), optimize content for LLM retrieval and Generative Engine Optimization (Surface), and define novel attribution metrics for zero-click environments (Score).

The remainder of this paper is organized as follows. Section 2 details the materials and methods, including the literature retrieval, filtering, and selection criteria. Section 3 presents the 4S lifecycle framework. Section 4 reviews the literature across thematic categories: SEO evolution and GEO, retrieval-augmented generation and information retrieval, knowledge graphs (KG) and semantic search, AI-driven digital marketing and personalization, conversational AI in search, bias and fairness in IR, sentiment analysis and customer behavior, and business intelligence and predictive analytics. Section 5 discusses prevailing trends, organizational challenges, and strategic opportunities, and Section 6 concludes.

2. Materials and Methods

2.1. Research Design

This study employs an integrative literature review methodology, which is particularly suited for emerging, multi-disciplinary research domains where the goal is to synthesize diverse perspectives rather than to isolate specific empirical effects. Unlike systematic reviews governed by restrictive protocols (e.g., PRISMA), the integrative approach enables the inclusion of conceptual, theoretical, and empirical works across computer science, information systems, and marketing management—a breadth necessary to decode the rapidly evolving intersection of SEO, generative AI, and business intelligence.

2.2. Data Source

The primary data source was the Google Scholar. Google Scholar was selected for several reasons: (1) free and unrestricted access; (2) comprehensive cross-disciplinary coverage spanning computer science, information systems, and marketing; (3) persistent identifiers (DOIs) enabling independent verification of every paper; and (4) citation count for each study.

2.3. Literature Retrieval

We used 157 search queries covering the review's thematic scope. The queries spanned the following sub-topics:

- **Core intersection:** SEO + generative AI, AI-driven SEO business intelligence, LLM search engine optimization;
- **GEO/SGE:** generative engine optimization, search generative experience, AI overviews, answer engine optimization;
- **AI search:** AI search engines, conversational search, LLM-powered search, RAG search engines, neural search retrieval;
- **Zero-click:** zero-click search, featured snippets, knowledge panels, search snippet optimization;
- **AI content:** AI-generated content SEO, ChatGPT SEO content, GPT content marketing, automated content SEO;
- **BI & analytics:** business intelligence search analytics, web analytics BI, search analytics dashboards, SEO analytics;
- **Digital marketing:** digital marketing strategy AI search, search marketing generative AI, AI digital marketing transformation;
- **Information retrieval:** information retrieval AI, neural IR, learning to rank, semantic search, knowledge graph search;
- **Predictive analytics:** predictive SEO, search intent prediction, click-through rate prediction, search ranking prediction;
- **Ethics/privacy:** algorithmic transparency search, AI search bias, search algorithm fairness, search privacy.

Papers were deduplicated by Google Scholar URL. This process yielded an initial corpus of **11,382 unique papers**.

2.4. Filtering and Selection

The filtering pipeline applied the following sequential criteria:

1. **Deduplication:** Papers were deduplicated by normalized title, reducing the corpus to 11,181 unique works.
2. **Abstract availability (mandatory):** Papers without a non-empty abstract were excluded, as abstracts are essential for content-level relevance assessment. This reduced the count to 9,868.
3. **Verifiability:** Papers without a DOI or named venue were excluded, ensuring every entry can be independently verified. Count after this step: 9,837.
4. **Non-scholarly type exclusion:** Software releases, datasets, errata, editorials, and retractions were excluded. Count: 9,837.
5. **Topic relevance filtering:** A paper was retained only if its title and abstract contained at least one specific search/SEO/marketing/IR term (e.g., "search engine optimization," "generative AI," "information retrieval," "digital marketing," "RAG," "knowledge graph"). Generic AI/ML terms alone (e.g., "machine learning," "deep learning") were insufficient. Papers whose titles indicated off-topic domains (medical, education, cybersecurity, finance, agriculture, physics, etc.) were excluded. Non-English papers were filtered out using both ASCII title checks and non-English function-word heuristics. This reduced the corpus to **2,962 on-topic papers**—the final `papers.bib` [20].

 doi:10.20944/preprints202607.0744.v1

2.5. Selection of the Top 70 Papers

From the 2,962-paper corpus, the 70 most valuable papers were selected using a composite relevance score combining six criteria, as documented in [20]:

- **R1. Keyword/theme match (40%):** Count of review-relevant keywords in title and abstract, with title hits weighted 3× over abstract hits. Per-theme capping ensured diversity across 10 sub-themes.
- **R2. Citation impact (25%):** Log-scaled citation count from Google Scholar, prioritizing foundational and influential work.
- **R3. Recency (15%):** Papers from 2020–2026 received a recency bonus; a smaller bonus for 2018–2019; foundational pre-2018 papers retained if highly cited.
- **R4. Venue quality (10%):** Papers from recognized publishers (Elsevier, IEEE, Springer, MDPI, ACM, Wiley, AAAI, SIGIR, etc.) received a venue quality bonus.
- **R5. Type diversity (5%):** Preference for a mix of empirical studies, reviews/surveys, and conceptual/theoretical papers.
- **R6. Theme coverage (5%):** Manual adjustment to ensure balanced coverage across all sub-themes, preventing over-representation of any single topic.

Additionally, a title-relevance filter required at least one core search/SEO/marketing/IR term in the paper's title, preventing generic LLM or ChatGPT overview papers from entering the selection. The final 70 papers span publication years 2011–2026, with the majority (over 80%) published in 2023–2026, reflecting the recency of the generative AI search phenomenon. All 70 papers are verifiable via DOI and Google Scholar URL.

2.6. Inclusion and Exclusion Criteria

The full inclusion and exclusion criteria are documented below.

- A paper was **included** if it satisfied all of the following: (I1) topical relevance to at least one of the review's ten core themes; (I2) publication in a recognized academic venue indexed in Google Scholar with a DOI or named venue; (I3) abstract availability (mandatory); (I4) publication year preference 2015–2026, with foundational earlier works included if highly cited; (I5) English language; (I6) preference for reputable sources (Elsevier, IEEE, Springer, MDPI, ACM, Wiley, AAAI, SIGIR, etc.).
- A paper was **excluded** if any of the following applied: (E1) off-topic primary subject (medical, cybersecurity, education, HR, agriculture, manufacturing, physics, biology, etc.); (E2) no abstract; (E3) unverifiable (no DOI and no venue); (E4) duplicates; (E5) editorials, errata, or non-scholarly types; (E6) non-English without English metadata.

3. The 4S Lifecycle Framework

Based on the integrative synthesis of the 70 selected studies, this paper introduces the **Signal-Structure-Surface-Score (4S) lifecycle framework** (Figure 1). The framework captures the end-to-end process by which organizations must adapt their BI and SEO practices to the generative search era, mapping four sequential stages that bridge AI search mechanics with strategic measurement.

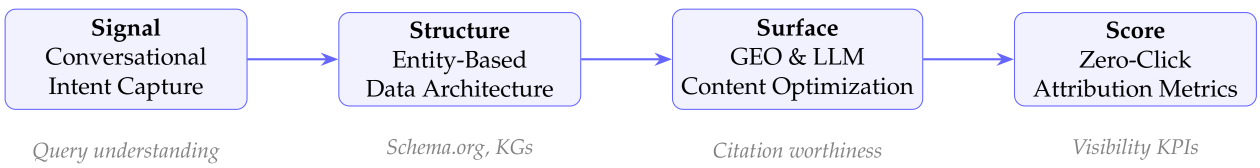

Figure 1. The Signal-Structure-Surface-Score (4S) lifecycle framework for AI-driven search visibility. Each stage represents a critical transition from traditional SEO to generative engine optimization, bridging AI search mechanics with BI measurement.

3.1. Signal: Conversational Intent Capture

The first stage addresses how AI-powered search engines interpret user queries. Unlike traditional keyword-based matching, generative search systems employ LLMs to understand conversational, multi-turn, and context-dependent queries [21,22]. The shift from keyword matching to intent understanding means that organizations must capture not just which keywords users type, but what they mean—a transition that requires NLP-driven intent classification and query rewriting capabilities [23]. Research on conversational search systems demonstrates that user query formulation strategies differ fundamentally between traditional and LLM-based search, with the latter producing longer, more natural-language queries [21]. This stage aligns with the BI "Signal" function: detecting and interpreting the conversational signals that generative engines use to retrieve and synthesize information.

3.2. Structure: Entity-Based Data Architecture

The second stage involves architecting content as machine-readable, entity-based data. Traditional SEO relied on HTML structure and metadata tags; generative search requires structured data in the form of knowledge graphs, Schema.org markup, and linked data [10,11]. Research on knowledge graph augmentation with LLMs demonstrates that microdata embedded in web pages for SEO purposes can serve as a valuable source for enriching knowledge graphs [11], creating a bidirectional relationship between SEO practices and KG construction. The integration of business intelligence with knowledge graph structures enables organizations to represent their products, services, and expertise in formats that LLMs can reliably retrieve and cite [9].

3.3. Surface: GEO and LLM Content Optimization

The third stage focuses on optimizing content for LLM retrieval and citation—the practice increasingly termed Generative Engine Optimization (GEO) or Answer Engine Optimization (AEO) [6,9]. Research shows that visibility in generative search results depends not on traditional ranking signals but on "citation worthiness"—the semantic clarity, authority, and structured format of content [24]. Samet [6] proposes a "GEO-First" framework that unifies search visibility, online reputation management, and digital authority, arguing that narrative inclusion within an AI's retrieval set acts as a catalyst for subsequent high-intent branded searches. Enache [9] introduces a structured approach to transitioning from SEO to AEO, emphasizing Schema.org markup and "extraction-readiness" as prerequisites for LLM citation. AI-generated content pipelines further accelerate this process, with studies demonstrating 98.8% efficiency improvements in SEO content production through generative AI integration [25].

3.4. Score: Zero-Click Attribution Metrics

The final stage addresses the measurement challenge: how organizations define and track visibility in zero-click environments. Traditional SEO metrics—keyword rankings, organic traffic, click-through rates—lose meaning when search engines synthesize answers without directing users to external sites [26]. Research on digital marketing analytics highlights the risk of "analytics myopia"—over-reliance on legacy metrics that fail to capture the full customer decision journey [8]. Samet [6] proposes "Share of Model" (SoM) and citation density as novel KPIs for the AI search era, arguing that organizations must transition from click-oriented to influence-oriented metrics. This stage represents the BI "Score" function: defining the attribution frameworks that capture value in a landscape where clicks no longer serve as the primary visibility signal.

4. Literature Review

The 70 selected papers are organized into eight thematic categories that collectively map the intersection of SEO, generative AI, and business intelligence. Rather than treating these categories in isolation, this review traces the connections between them—how technical advances in retrieval

 doi:10.20944/preprints202607.0744.v1

architecture enable new marketing strategies, how knowledge graph research informs entity-based SEO, and how bias and fairness concerns shape the future of search visibility measurement.

4.1. SEO Evolution and Generative Engine Optimization

The most striking finding across this body of literature is the speed and depth of the SEO paradigm shift. What was once a discipline built on keyword matching, backlink profiles, and on-page technical optimization [1,27] has fractured into a multi-layered practice where visibility depends not on ranking position but on whether a generative AI system chooses to cite a source. Liang et al. [5] provide empirical evidence for this transition through a user study comparing Bing Chat (a generative IR system) with traditional search, demonstrating that generative IR returns synthesized text with citations rather than ranked links, thereby reducing user search effort and fundamentally altering the click-based economy of organic search.

This observation is not merely theoretical. Beriozkin [28] provides concrete evidence from a niche luxury transportation firm, where the deployment of Custom GPTs enriched with real-time SEO analytics produced a 23% increase in organic traffic and a 37% uplift in lead conversion rates. What makes this case instructive is not just the positive outcomes but the mechanism: the firm transcended the limitations of generic LLMs by training AI agents on proprietary corporate data and external keyword metrics, effectively creating a localized knowledge base that generative engines could retrieve and cite. This finding suggests that the future of SEO may lie less in optimizing for search engine algorithms and more in optimizing for LLM retrieval—a shift that Samet [6] formalizes as a "GEO-First" framework.

Samet's framework is particularly significant because it reframes the strategic question. Rather than asking "how do we rank?", organizations should ask "how do we become cited?" The framework identifies two operational pillars—Discovery and Sentiment—grounded in a foundational layer of Digital Authority, and introduces "Share of Model" (SoM) as a novel KPI measuring how frequently an AI system references a brand. This represents a fundamental departure from the click-oriented metrics that have dominated digital marketing for two decades [4], aligning instead with what Samet terms "influence-oriented KPIs" suited to a zero-click, agentic landscape.

The transition from SEO to AEO requires not just strategic reframing but technical adaptation. Enache [9] provides a methodological bridge, proposing that content must be repositioned as a "structured, machine-interpretable data object" with Semantic Clarity, Schema.org markup, and "Extraction-Readiness" as prerequisites for LLM citation. This architectural perspective is complemented by Salem et al. [10], who demonstrate that automatically injecting structured data into news articles—validated through Google's Rich Results Test API—can improve discoverability for aggregator platforms. Together, these studies suggest that the technical SEO toolkit is expanding from HTML optimization to semantic web integration, where the ability to produce machine-readable, entity-based content becomes a core competency.

The content production process itself is being transformed. Murdiyanto et al. [25] demonstrate an end-to-end pipeline integrating web scraping, generative AI (OpenAI GPT), and autonomous WordPress publication, achieving a 98.8% reduction in content production time (from 195 minutes to 2.25 minutes per article) while meeting on-page SEO indicators. Vajrobol et al. [29] frame this transformation more broadly, arguing that generative AI is not merely accelerating SEO but fundamentally reshaping it—automating keyword research, enabling chatbot-enhanced user experiences, and providing data-driven insights that were previously inaccessible. However, both studies caution that AI-generated content raises ethical questions about quality, authenticity, and the potential for search engines to penalize machine-generated text.

A critical insight emerges when these findings are read together: the SEO profession is bifurcating. On one branch, technical practitioners must master structured data, knowledge graphs, and semantic web technologies to ensure content is "extraction-ready" for LLM retrieval [9,10]. On the other, content strategists must navigate an AI-accelerated production environment where the line between human and machine authorship is increasingly blurred [25,29]. The traditional on-page optimization

 doi:10.20944/preprints202607.0744.v1

factors identified by Prasad and Chandrika [1]—keyword optimization, site structure, internal linking, usability, mobile versioning, rich snippets, and social media integration—remain necessary but are no longer sufficient. They must be augmented with entity-based SEO, Schema.org markup, and the kind of digital authority signals that Samet [6] identifies as foundational to generative search visibility.

4.2. Retrieval-Augmented Generation and Information Retrieval

The technical foundation of generative search lies in retrieval-augmented generation (RAG), a paradigm that has evolved rapidly from a niche technique to the dominant architecture for LLM-powered search systems. The literature reveals a clear trajectory: from foundational surveys that established the RAG taxonomy [15], through comprehensive frameworks that mapped its architectural variations [12,13], to specialized techniques that address specific retrieval challenges.

Gao et al. [15] provide the foundational survey that subsequent studies build upon, establishing the basic RAG framework of retrieving external knowledge to augment LLM generation. Huang and Huang [12] extend this foundation by organizing the RAG paradigm into four categories—pre-retrieval, retrieval, post-retrieval, and generation—offering a lifecycle perspective that reveals where optimization interventions can be applied. Fan et al. [13] further systematize the field across three dimensions: architectures, training strategies, and applications, demonstrating that RA-LLMs have moved beyond experimental prototypes into production deployment across diverse domains.

What emerges from these surveys is a picture of a maturing field grappling with fundamental reliability challenges. Chen et al. [30] establish the Retrieval-Augmented Generation Benchmark (RGB), evaluating six representative LLMs across four abilities: noise robustness, negative rejection, information integration, and counterfactual robustness. Their findings are sobering: while LLMs exhibit some noise robustness, they struggle significantly with negative rejection (failing to ignore irrelevant retrieved passages), information integration (combining multiple sources coherently), and handling false information. Zhang et al. [31] review hallucination mitigation techniques, confirming that hallucination remains a critical challenge for the reliability of generative search systems—one that RAG mitigates but does not eliminate.

The field's response to these challenges has been a proliferation of specialized retrieval strategies. Sawarkar et al. [32] propose "Blended RAG," combining dense vector indexes with sparse encoder indexes and hybrid query strategies, setting new benchmarks on IR datasets and surpassing fine-tuning performance on generative Q&A tasks. Omolayoa [33] contributes a comparative evaluation of vector embedding frameworks, providing practical guidance for selecting retrieval infrastructure. Su et al. [34] introduce DRAGIN, which dynamically triggers retrieval based on real-time LLM information needs rather than fixed query intervals, while Jeong et al. [35] propose Adaptive-RAG, which adjusts retrieval strategy based on question complexity—simple queries receive lightweight retrieval, complex queries invoke multi-step retrieval. Ma et al. [23] address the critical preprocessing step of query rewriting, showing that reformulating user queries before retrieval significantly improves search relevance in conversational contexts.

A particularly important finding for the SEO community comes from Li et al. [36], who compare RAG with long-context LLMs and propose a hybrid approach. Their study suggests that the choice between retrieving external knowledge and providing longer context windows is not binary—optimal systems will combine both, with implications for how content should be structured and chunked for retrieval. Sun et al. [37] directly investigate whether ChatGPT is effective as a search engine, examining LLMs as re-ranking agents. Their finding that LLM performance varies significantly across query types is a critical insight for SEO practitioners: generative search does not uniformly advantage all content types, and understanding which query classes LLMs handle well or poorly can inform content strategy.

The broader IR perspective is provided by Zhu et al. [38], who survey the confluence of LLMs and IR systems across query rewriters, retrievers, rerankers, and readers, and by Hersh [39], who argues that search still matters in the era of generative AI—information retrieval remains a fundamental capability even as generative systems evolve. Mo et al. [21] map the evolution from keyword-based to dialogue-

 doi:10.20944/preprints202607.0744.v1

based search in their survey of conversational search, while Sekuli et al. [22] analyze utterances in LLM-based user simulation, revealing that users formulate longer, more natural-language queries with LLM-based search than with traditional search engines—a behavioral shift with profound implications for keyword research and content strategy.

4.3. Knowledge Graphs and Semantic Search

If RAG provides the retrieval engine for generative search, knowledge graphs provide the fuel. The literature reveals a symbiotic relationship between SEO-optimized structured data and knowledge graph construction—one that creates new strategic opportunities for organizations willing to invest in entity-based content architecture.

Gonzalez-Garcia et al. [11] demonstrate this symbiosis empirically. By extracting Schema.org microdata from web pages via CommonCrawl and matching it against Wikidata, they show that microdata produced for SEO purposes can serve as a valuable resource for knowledge graph enrichment. The implication is striking: SEO practitioners have been producing structured data for years without realizing its potential value for KG construction. Conversely, KGs enhance search engine understanding of web content, creating a virtuous cycle where SEO investment in structured data improves KG quality, which in turn improves search visibility. Their finding that metadata annotations concentrate in only a few Schema.org attributes—and that authoring tools influence a significant fraction of microdata production—suggests both an opportunity and a quality challenge for the SEO community.

Ibrahim et al. [40] provide the comprehensive survey that maps how LLMs can augment KGs across construction, completion, and question answering, while Wang et al. [41] demonstrate the reverse: using KGs to augment LLM-based recommendation through Knowledge Graph Retrieval-Augmented Generation. This bidirectional enhancement—LLMs improving KGs and KGs improving LLMs—represents a convergence that is reshaping both fields. The practical question for organizations is whether to invest in building proprietary KGs or in optimizing content for inclusion in existing ones.

Sun et al. [42] provide a provocative answer with their "Head-to-Tail" study, which assesses LLM knowledgeability across different frequency bands. Their finding that LLMs excel at common ("head") knowledge but struggle with rare ("tail") knowledge suggests that KGs remain essential for specialized information—and that organizations with deep domain expertise can establish authority by structuring that expertise in KG format. This finding has direct SEO implications: in a generative search environment, being the authoritative source for niche entities that LLMs cannot reliably generate from training data may be more valuable than competing for common knowledge.

The technical foundations for this entity-based approach are well-established. Xie et al. [43] present foundational work on representation learning of KGs with entity descriptions, a technique that underpins modern semantic search. Choi et al. [44] demonstrate how pre-trained language models can enhance KG completion through masked entity prediction, while Wang et al. [45] address explainable reasoning over KGs for recommendation, contributing to the transparency that both search engines and regulators increasingly demand.

The semantic search tradition that predates the generative AI boom remains theoretically relevant. Janowicz et al. [46] examine the semantics of similarity in geographic information retrieval, providing theoretical foundations that inform modern GEO practices. Hollink et al. [47] conduct semantic search log analysis on professional image search, offering early insights into how users interact with semantic search systems—insights that are now being revisited as generative search transforms user behavior [22].

4.4. AI-Driven Digital Marketing and Personalization

The marketing literature reveals a field in rapid transformation, where AI is simultaneously a tool for optimization and a force reshaping the competitive landscape. A cross-cutting insight from this body of work is that AI adoption in digital marketing is no longer a differentiator—it is becoming a

 doi:10.20944/preprints202607.0744.v1

baseline requirement, and the competitive advantage lies in how deeply organizations integrate AI into their strategic decision-making.

Islam et al. [16] provide one of the most comprehensive assessments, using PRISMA methodology to synthesize 150 relevant studies. Their finding that AI enhances digital marketing through predictive analytics, NLP, and chatbots—enabling customer segmentation, content personalization, and campaign optimization—is consistent with the broader literature. What distinguishes their contribution is the identification of persistent challenges: data privacy, algorithmic bias, and high AI adoption costs remain significant barriers, particularly for smaller organizations. These challenges are echoed by Hasan [48], whose PRISMA-guided review of 112 articles confirms that AI-powered SEO tools improve organic reach, technical site health, and semantic keyword alignment, but notes that sustained AI adoption requires compounding strategic advantages including innovation capacity, brand loyalty, and agility.

The personalization dimension is explored by several studies that collectively reveal a shift from segment-based to individual-level marketing. Das [18] emphasizes generative AI's capacity to analyze consumer data and produce tailored text, images, and video, highlighting the growing significance of voice and image search optimization. Raji et al. [49] review AI-powered personalization in e-commerce, showing how advanced algorithms deliver tailored content and recommendations while predicting consumer preferences and streamlining the purchasing journey. Kedi et al. [50] focus on SMEs, finding that machine learning, NLP, and predictive analytics enable personalized content delivery that was previously accessible only to large enterprises with dedicated data science teams.

The e-commerce context receives particular attention. Kathiriya [51] demonstrates LLM-based description and keyword generation from multimodal data for e-commerce listings, showing how generative AI can automate product content creation at scale. Purnomo [52] identifies SEO, content marketing, social media, paid advertising, and user experience optimization as core strategies for sales conversion, while Shaban [53] provides a comprehensive review of digital marketing strategies for e-commerce success. The regional dimension is added by Kamkankaew et al. [54], who examine AI's transformative role in Thai B2C digital marketing, providing evidence that AI-driven marketing transformation is not confined to Western markets.

A particularly interesting finding comes from Abdelkader [55], who investigates ChatGPT's influence on customer experience in digital marketing and finds that familiarity with technology, business type, age, and education level moderate the relationship between ChatGPT-based customer experience and overall satisfaction. This suggests that the effectiveness of AI-driven marketing is not uniform across customer segments—a critical insight for personalization strategies that must account for varying levels of AI acceptance. Arslan et al. [56] extend this line of inquiry with a comprehensive review of LLM applications in business and management, combining bibliometric analysis with thematic synthesis to show a post-2023 shift from general-purpose conversational use toward domain-specific and architecture-aware implementations that bridge the gap between AI capabilities and marketing practice.

The SME perspective is particularly important given the resource constraints these organizations face. Sharabati et al. [57] find that digital marketing strategies—including SEO, social media marketing, and customer engagement through digital channels—are essential drivers of digital transformation and economic results for SMEs. Ali and Gutiérrez [58] emphasize that companies adopting AI and personalized marketing gain clear competitive advantages, while Singh [59] provides a systematic review of digital marketing strategies and consumer engagement, examining emerging trends and measurement frameworks. Hutsaliuk and Mirzoiev [60] highlight the integration of AI, big data analytics, automation, and social media in modern marketing strategies, while Joel [61] examines the broader impact of digital transformation on business development strategies. Sihotang [62] proposes a digital marketing strategy using customer journey analysis, and Fadnavis [63] explores personalized marketing through machine learning—together painting a picture of a marketing discipline that is becoming increasingly data-driven, AI-augmented, and customer-centric.

 doi:10.20944/preprints202607.0744.v1

4.5. Conversational AI and Chatbots in Search

The evolution from intent-based chatbots to LLM-powered conversational agents represents a microcosm of the broader search transformation. Singh and Namin [64] provide a comprehensive survey on chatbots and large language models, offering a taxonomy of testing and evaluation techniques that reveals how LLMs have fundamentally expanded conversational agent capabilities. The significance of this evolution for search lies in the fact that conversational interfaces are becoming the primary mode of interaction with information retrieval systems—a trend that Mo et al. [21] document in their survey of conversational search.

The practical implementation challenges of conversational search are addressed by Pokhrel et al. [65], who demonstrate a RAG-based website chatbot combining web scraping, vectorization, and semantic search. Their system retrieves relevant document segments based on user queries and generates contextually appropriate responses, illustrating how the RAG architecture discussed in Section 4.2 can be operationalized in a conversational interface. Mazumder [66] extends this work to library search systems, showing that RAG-based search can enhance information discovery in specialized contexts beyond commercial search engines—a finding with implications for enterprise search and internal knowledge management.

4.6. Bias, Fairness, Transparency, and Ethics in Information Retrieval

As generative AI becomes embedded in search infrastructure, the question of who gets seen and why takes on new urgency. The literature reveals that bias in IR systems is not a peripheral concern but a structural feature that generative AI may amplify rather than mitigate.

Dai et al. [67] provide the most comprehensive treatment, unifying bias and unfairness in IR systems as distribution mismatch problems. Their framework examines bias across three stages of LLM integration—data collection, model development, and result evaluation—and proposes mitigation strategies through distribution alignment. What makes their contribution particularly significant is the recognition that LLM integration introduces new bias pathways that did not exist in traditional IR systems, requiring fundamentally new fairness frameworks.

Bernard and Balog [68] conduct a systematic review of fairness, accountability, transparency, and ethics in IR, revealing a field struggling with definitional pluralism. Their finding that no standard definitions exist for these concepts—due to their multi-dimensional nature—highlights a governance gap: organizations cannot ensure fairness in generative search if they cannot define what fairness means in this context. They develop taxonomies of requirements and propose practical definitions, but the implementation challenge remains.

The practical consequences of search bias are demonstrated by Feng and Shah [69], who investigate gender bias in image search using adversarial attack queries. Their experiments on Google, Baidu, Naver, and Yandex reveal that adversarial attacks can trigger high levels of gender bias even in systems that have been corrected for such bias—a finding that underscores the fragility of bias mitigation in opaque, proprietary search systems. They design three re-ranking algorithms (epsilon-greedy, relevance-aware swapping, and fairness-greedy) to mitigate bias, providing actionable techniques that could be applied to generative search systems.

Castillo [70] provides foundational principles for fairness and transparency in ranking that remain relevant as generative search systems increasingly determine information visibility. The ethical dimension extends beyond bias to privacy. Saura et al. [71] investigate the ethics of AI-based digital marketing, identifying a "data privacy paradox" where the most profitable AI-driven marketing actions—cross-device tracking, real-time monitoring, behavioral analytics—are precisely those least connected to privacy and ethical standards. Their finding of a strong proximity between real-time tracking, IoT, and surveillance variables underscores the critical need to ethically understand how user behavior is monitored in AI-driven marketing environments.

4.7. Sentiment Analysis and Customer Behavior

Sentiment analysis serves as a critical bridge between understanding customer behavior and informing search and marketing strategy. The literature reveals that generative AI is transforming this field as profoundly as it is transforming search itself.

Krugmann and Hartmann [72] provide the benchmark study, comparing GPT-3.5, GPT-4, and Llama 2 against established transfer learning models for sentiment analysis. Their finding that LLMs can surpass traditional methods in classification accuracy—despite their zero-shot nature—represents a significant shift for marketing research. Equally important is their finding that linguistic features such as lengthy, content-laden words improve classification performance while single-sentence reviews and unstructured social media text reduce it, providing practical guidance for when LLM-based sentiment analysis is appropriate.

The multilingual dimension is addressed by Miah et al. [73], who achieve over 86% accuracy on cross-lingual sentiment analysis using an ensemble of transformers and LLMs with translation to English. This finding has significant implications for global digital marketing strategies: organizations can leverage LLM-based sentiment analysis across language barriers without requiring language-specific training data. Rane et al. [74] review the broader application of AI, ML, and deep learning for sentiment analysis in CRM, highlighting how NLP-driven sentiment analysis enables real-time understanding of customer emotions across social media, reviews, and support tickets. Paul et al. [75] extend this to digital marketing specifically, finding that transformer-based models (BERT, RoBERTa) excel in multilingual sentiment interpretation and that integrating these models into customer feedback loops enables real-time marketing decision-making and sentiment-driven content optimization.

4.8. Business Intelligence and Predictive Analytics

The BI and predictive analytics literature addresses the measurement dimension of the 4S framework—how organizations capture, analyze, and act on data in an AI-driven search environment. A key insight from this body of work is that the tools and frameworks developed for traditional digital marketing analytics are necessary but insufficient for the generative search era.

Vollrath and Villegas [8] provide the conceptual foundation with their work on avoiding "digital marketing analytics myopia." Their framework places analytics tools in the context of a firm's overall marketing plan, reconnecting measurement with the customer decision journey. This principle is directly applicable to the zero-click attribution challenge: organizations that focus solely on declining click metrics risk missing the broader influence that generative search visibility exerts on brand awareness, consideration, and downstream conversion. The framework's emphasis on theoretically grounded, strategically aligned measurement provides a template for developing the "Score" stage of the 4S lifecycle.

Ajiga et al. [76] review AI-driven predictive analytics in retail, demonstrating how machine learning algorithms, NLP, and computer vision revolutionize the way retailers harness data for strategic decision-making. Their emphasis on data quality and ethical considerations is a recurring theme: the value of predictive analytics is contingent on the quality of input data and the transparency of the models used. Cvitanović [77] provides early evidence of how BI dashboards support marketing decision-making through digital marketing benchmarks, while Chen et al. [78] demonstrate how RAG-based systems can serve as BI tools for organizational knowledge management, bridging the gap between internal knowledge and AI-powered retrieval.

The application of RAG to specialized BI domains is explored by Iaroshev et al. [79], who evaluate RAG models for financial report question answering, and by Toro et al. [80], who present DRAGON-AI for dynamic ontology generation. Together, these studies suggest that the RAG architecture—originally developed for general-purpose question answering—is being adapted for domain-specific BI applications, creating new possibilities for organizations to build internal generative search systems that leverage their proprietary knowledge.

 doi:10.20944/preprints202607.0744.v1

5. Discussion and Future Research Agenda

5.1. Prevailing Trends

The integrative review reveals several prevailing trends that are reshaping the SEO and BI landscape. First, the transition to AI-generated answer panels is reducing organic click-through rates, challenging the economic model of traditional SEO [6,7]. Organizations must adapt to an environment where visibility does not guarantee traffic, and where "Share of Model" (SoM)—the frequency with which an AI system cites a brand—becomes a primary visibility metric [6]. Second, generative AI dramatically accelerates content production, with studies reporting up to 98.8% efficiency improvements [25]; however, this velocity raises quality and authenticity concerns, as AI content detection tools struggle to accurately identify machine-generated text. Third, the literature reveals a convergence between SEO practices and knowledge engineering, as Schema.org markup, knowledge graphs, and linked data become essential for LLM retrieval [9–11]. This convergence requires SEO practitioners to develop new competencies in structured data and entity-based knowledge representation. Fourth, retrieval-augmented generation is emerging as the dominant architecture for generative search systems [12,13,15]. The rapid advancement of RAG techniques—including dynamic retrieval [34], adaptive retrieval [35], and blended retrieval [32]—suggests that the technical foundations of generative search are still maturing, with significant implications for how content must be structured for retrieval.

5.2. Core Organizational Challenges

The review also identifies several core organizational challenges. The zero-click environment creates attribution gaps that legacy BI frameworks cannot address [8]; organizations lack visibility into when and how generative search engines cite their content, making it difficult to measure the ROI of SEO and GEO investments. Compounding this challenge, generative search systems operate as black boxes, with ranking and citation mechanisms that are opaque to content creators [67,68]. This opacity makes it challenging to optimize content for generative retrieval and raises concerns about fairness and accountability in information access. Explainable AI (XAI) approaches—which aim to render model decision processes transparent and interpretable—offer a promising pathway, yet their application to generative search systems remains largely unexplored. Reliability remains a further concern: LLM-based search systems are prone to hallucinations—generating plausible but incorrect information [30,31]. RAG mitigates but does not eliminate this risk, and organizations must develop strategies for ensuring their content is accurately represented in generative search results. Finally, the use of AI in digital marketing raises significant privacy concerns [71], with tensions between personalization and consumer privacy creating what Saura et al. describe as a "data privacy paradox." Organizations must navigate GDPR, cookieless tracking, and evolving regulatory frameworks while maintaining effective search and marketing strategies.

5.3. Emerging Strategic Opportunities

Despite these challenges, the review identifies several emerging strategic opportunities. The conversational nature of generative search enables richer intent signals than traditional keyword data [21,22]; organizations that develop capabilities for capturing and analyzing conversational query patterns can gain earlier and deeper insights into customer needs. As generative search engines cite sources, organizations can audit which competitors are being cited for key queries, creating a new form of competitive intelligence [6]. This "LLM audit trail" provides insights into competitor content strategies and authority signals that generative engines prioritize. The shift toward knowledge graph-based search creates opportunities for organizations to establish entity authority—becoming the definitive source for specific entities that LLMs reference [11,40]. This requires investment in structured data, Schema.org markup, and knowledge graph integration. Finally, research on combining RAG with long-context LLMs [36] suggests that hybrid approaches can deliver superior search and retrieval

 doi:10.20944/preprints202607.0744.v1

performance, offering organizations a path to more reliable and comprehensive generative search integration.

5.4. Future Research Agenda

Based on the identified gaps and opportunities, this review proposes the following research agenda:

1. **Developing zero-click attribution frameworks:** Future research should develop and validate BI frameworks that measure visibility and impact in zero-click environments, moving beyond click-based metrics to citation density, Share of Model, and influence-oriented KPIs [6].
2. **Investigating GEO ranking factors:** Empirical studies are needed to identify the factors that influence LLM citation and retrieval in generative search, analogous to the ranking factor studies that shaped traditional SEO [9,24].
3. **Addressing algorithmic transparency:** Research should develop methods for auditing generative search systems for bias, fairness, and accuracy, building on the frameworks proposed by Dai et al. [67] and Bernard and Balog [68]. Explainable AI (XAI) techniques—such as attention visualization, citation provenance tracking, and counterfactual explanation—represent a particularly promising direction for bridging the gap between algorithmic opacity and practitioner trust.
4. **Exploring RAG optimization for enterprise search:** Studies should investigate how organizations can implement RAG-based internal search systems that leverage their proprietary knowledge graphs and structured data [41,78].
5. **Measuring AI content quality at scale:** As AI-generated content becomes ubiquitous, research must address how to maintain quality, authenticity, and search engine compliance at scale [25,29].
6. **Integrating conversational analytics into BI:** Future work should explore how conversational search data can be integrated into BI dashboards and decision-making processes, bridging the gap between search analytics and strategic planning [8,21].

6. Conclusions

This integrative literature review has synthesized 70 high-value studies from an initial corpus of over 2,900 verified academic papers to map the evolution of SEO in the era of generative AI from a business intelligence perspective. The review makes three primary contributions.

First, it introduces the **Signal-Structure-Surface-Score (4S) lifecycle framework**, which provides a structured approach to understanding how organizations must adapt their SEO and BI practices across four stages: capturing conversational search intents (Signal), architecting entity-based data structures (Structure), optimizing content for LLM retrieval and GEO (Surface), and defining zero-click attribution metrics (Score). The framework bridges the gap between AI search mechanics and strategic BI measurement, offering practitioners a roadmap for navigating the generative search transition.

Second, the review systematically maps the technical and strategic landscape across eight thematic categories, revealing prevailing trends (zero-click normalization, AI content velocity, SEO-knowledge engineering convergence, RAG as search infrastructure), core organizational challenges (attribution loss, algorithmic opacity, hallucination risks, privacy tensions), and emerging strategic opportunities (real-time intent mapping, competitor LLM audit trails, entity-based content architecture, hybrid RAG-long-context strategies).

Third, it provides a robust future research agenda addressing six priority areas: zero-click attribution frameworks, GEO ranking factors, algorithmic transparency, enterprise RAG optimization, AI content quality measurement, and conversational analytics integration.

The findings demonstrate that the transition from traditional SEO to generative engine optimization is not merely a technical evolution but a fundamental paradigm shift that requires new BI frameworks, new metrics, and new organizational competencies. As LLMs increasingly mediate information access, the ability to structure content for machine readability, optimize for LLM citation, and measure visibility in zero-click environments will become critical organizational capabilities. The

4S framework provides a foundation for developing these capabilities, while the research agenda identifies the empirical and theoretical work needed to advance the field.

The review is not without limitations. The integrative methodology, while enabling broad synthesis, does not provide the systematic rigor of PRISMA-based reviews. The reliance on Google Scholar as a single data source may introduce coverage biases, particularly for very recent preprints and conference papers. Future work should complement these findings with domain-specific systematic reviews and empirical studies that test the propositions advanced in this paper.

Author Contributions: Conceptualization, K.I.R., D.M., D.S. and C.V.; methodology, K.I.R., D.M., D.S. and C.V.; software, K.I.R. and D.M.; validation, K.I.R., D.M., D.S. and C.V.; formal analysis, K.I.R., D.M., D.S. and C.V.; investigation, K.I.R., D.M., D.S. and C.V.; resources, K.I.R., D.M., D.S. and C.V.; data curation, K.I.R., D.M., D.S. and C.V.; writing—original draft preparation, K.I.R. and D.M.; writing—review and editing, K.I.R., D.M., D.S. and C.V.; visualization, K.I.R. and D.M.; supervision, C.V.; project administration, K.I.R., D.M.; funding acquisition, K.I.R., D.M., D.S. and C.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The bibliographic data used in this study are available on GitHub https://github.com/Applied-AI-Research-Lab/The-Evolution-of-SEO-in-the-Era-of-Generative-AI.

Acknowledgments: During the preparation of this work the authors used ChatGPT (OpenAI) and Grammarly in order to enhance grammatical accuracy and refine sentence structure. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Prasad, G.; Chandrika, T.S.P. The Role of an On-Page Optimization for an Effective Digital Marketing. *International Journal of Research Publication and Reviews* **2022**. https://doi.org/10.55248/gengpi.2022.3.5.21.
2. Roumeliotis, K.I.; Tselikas, N.D.; Tryfonopoulos, C. Greek Hotels' Web Traffic: A Comparative Study Based on Search Engine Optimization Techniques and Technologies. *Digital* **2022**, *2*, 379–400. https://doi.org/10.3390/digital2030021.
3. Roumeliotis, K.I.; Tselikas, N.D.; Nasiopoulos, D.K. Airlines' Sustainability Study Based on Search Engine Optimization Techniques and Technologies. *Sustainability* **2022**, *14*. https://doi.org/10.3390/su141811225.
4. Mou, A.J.; Hossain, M.S.; Siddiqui, N.A. DIGITAL TRANSFORMATION IN MARKETING: EVALUATING THE IMPACT OF WEB ANALYTICS AND SEO ON SME GROWTH. *American Journal of Interdisciplinary Studies* **2022**. https://doi.org/10.63125/8t10v729.
5. Liang, Y.; Wu, Z.; Zhang, F.; Song, D.; Huang, H. How Users Interact with Generative Information Retrieval Systems: A Study of User Behavior and Search Experience **2025**. https://doi.org/10.1145/3726302.3729998.
6. Samet, U. A GEO-First Framework: Integrating Search Visibility, Sentiment, and Digital Authority for Organic Growth in the AI Era. *World Journal of Advanced Research and Reviews* **2026**. https://doi.org/10.30574/wjarr.2026.29.1.0152.
7. Gleason, J.M.; Hu, D.; Robertson, R.E.; Wilson, C. Google the Gatekeeper: How Search Components Affect Clicks and Attention. *Proceedings of the International AAAI Conference on Web and Social Media* **2023**. https://doi.org/10.1609/icwsm.v17i1.22142.
8. Vollrath, M.; Villegas, S.G. Avoiding digital marketing analytics myopia: revisiting the customer decision journey as a strategic marketing framework. *Journal of Marketing Analytics* **2021**. https://doi.org/10.1057/s41270-020-00098-0.
9. Enache, M.C. From SEO to AEO for Optimizing LLM Output in E-commerce Applications. *Annals of Dunarea de Jos University of Galati Fascicle I Economics and Applied Informatics* **2025**. https://doi.org/10.35219/eai15840409551.
10. Salem, H.; Salloum, H.; Orabi, O.; Sabbagh, K.; Mazzara, M. Enhancing News Articles: Automatic SEO Linked Data Injection for Semantic Web Integration. *Applied Sciences* **2025**. https://doi.org/10.3390/app15031262.

 doi:10.20944/preprints202607.0744.v1

11. Gonzalez-Garcia, L.; González-Carreño, G.; Machota, A.M.R.; Fernández-Vega, J.P. Enhancing knowledge graphs with microdata and LLMs: the case of Schema.org and Wikidata in touristic information. *The Electronic Library* **2024**. https://doi.org/10.1108/el-06-2023-0160.
12. Huang, Y.; Huang, J.X. A Survey on Retrieval-Augmented Text Generation for Large Language Models. *ACM Computing Surveys* **2026**. https://doi.org/10.1145/3805774.
13. Fan, W.; Ding, Y.; Ning, L.; Wang, S.; Li, H.; Yin, D.; Chua, T.S.; Li, Q. A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. In Proceedings of the Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2024; KDD '24, p. 6491–6501. https://doi.org/10.1145/3637528.3671470.
14. Roumeliotis, K.I.; Zestas, O.N.; Kyriakou, K.D.; Soumis, D.N.; Kapellaki, S.H.; Seklou, K.; Tselikas, N.D. ORAM.AI: A multi-agent chatbot with RAG-based reasoning and MCP integration for adaptive fitness recommendation. *Array* **2026**, *30*, 100933. https://doi.org/10.1016/j.array.2026.100933.
15. Gao, Y.; Xiong, Y.; Gao, X.; Jia, K.; Pan, J.; Bi, Y.; Dai, Y.; Sun, J.; Wang, M.; Wang, H. Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv (Cornell University)* **2023**. https://doi.org/10.48550/arxiv.2312.10997.
16. Islam, M.A.; Fakir, S.I.; Masud, S.B.; Hossen, M.D.; Islam, M.; Siddiky, M.R. Artificial intelligence in digital marketing automation: Enhancing personalization, predictive analytics, and ethical integration. *Edelweiss Applied Science and Technology* **2024**. https://doi.org/10.55214/25768484.v8i6.3404.
17. Roumeliotis, K.I.; Tselikas, N.D. A Machine Learning Python-Based Search Engine Optimization Audit Software. *Informatics* **2023**, *10*. https://doi.org/10.3390/informatics10030068.
18. Das, S. Investigating Generative AI innovative Strategies for Customer Engagement in Marketing Automations in the Digital Era. *International Journal of Applied Science and Engineering* **2024**. https://doi.org/10.30954/2322-0465.1.2024.5.
19. Snyder, H. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research* **2019**, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039.
20. Applied AI Research Lab. The Evolution of SEO in the Era of Generative AI. https://github.com/Applied-AI-Research-Lab/The-Evolution-of-SEO-in-the-Era-of-Generative-AI, n.d.
21. Mo, F.; Mao, K.; Zhao, Z.; Qian, H.; Chen, H.; Cheng, Y.; Li, X.; Zhu, Y.; Dou, Z.; Nie, J. A Survey of Conversational Search. *ACM Transactions on Information Systems* **2025**. https://doi.org/10.1145/3759453.
22. Sekulić, I.; Alinannejadi, M.; Crestani, F. Analysing Utterances in LLM-Based User Simulation for Conversational Search. *ACM Transactions on Intelligent Systems and Technology* **2024**. https://doi.org/10.1145/3650041.
23. Ma, X.; Gong, Y.; He, P.; Zhao, H.; Duan, N. Query Rewriting in Retrieval-Augmented Large Language Models **2023**. https://doi.org/10.18653/v1/2023.emnlp-main.322.
24. Font-Julián, C.I.; na Malea, E.O.; Codina, L. ChatGPT Search as a tool for scholarly tasks: evolution or devolution? *Infonomy* **2024**. https://doi.org/10.3145/infonomy.24.059.
25. Murdiyanto, A.W.; Sulistiyantoro, D.; Kurniawati, M.W. SEO-Based Blog Content Pipeline Automation: Integrating Web Scraping and Generative AI for Digital Marketing Efficiency. *APPLIED SCIENCE AND TECHNOLOGY REASERCH JOURNAL* **2026**. https://doi.org/10.31316/astro.v4i2.9454.
26. Agrawal, K. The Future of AI in Digital Search: Towards a Fully Conversational Experience. *Journal of Computer Science and Technology Studies* **2025**. https://doi.org/10.32996/jcsts.2025.7.4.34.
27. Roumeliotis, K.I.; Tselikas, N.D. Search Engine Optimization Techniques: The Story of an Old-Fashioned Website. In Proceedings of the Business Intelligence and Modelling; Sakas, D.P.; Nasiopoulos, D.K.; Taratuhina, Y., Eds., Cham, 2021; pp. 47–55.
28. Beriozkin, N. AI SEO Case Study: Driving Exponential Digital Growth through Custom GPTs and SEO Analytics in Niche Markets. *Open MIND* **2026**. https://doi.org/10.5281/zenodo.20406494.
29. Vajrobol, V.; Aggarwal, N.; Saxena, G.J.; Singh, S.; Pundir, A. Transforming SEO in the Era of Generative AI: Challenges, Opportunities, and Future Prospects; 2024. https://doi.org/10.4324/9781032688305-6.
30. Chen, J.; Lin, H.; Han, X.; Sun, L. Benchmarking Large Language Models in Retrieval-Augmented Generation. *Proceedings of the AAAI Conference on Artificial Intelligence* **2024**. https://doi.org/10.1609/aaai.v38i16.29728.
31. Zhang, W.; Zhang, J. Hallucination Mitigation for Retrieval-Augmented Large Language Models: A Review. *Mathematics* **2025**. https://doi.org/10.3390/math13050856.
32. Sawarkar, K.; Mangal, A.; Solanki, S.R. Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Semantic Search and Hybrid Query-Based Retrievers **2024**. https://doi.org/10.1109/mipr62202.2024.00031.

 doi:10.20944/preprints202607.0744.v1

33. Omolayoa, O.; Babatope, O. Comparative Evaluation of Vector Embedding Frameworks for Scalable Sematic Retrieval in PDF-Based RAG Systems. *International Journal of Research Publication and Reviews* **2025**. https://doi.org/10.55248/gengpi.6.0625.23100.
34. Su, W.; Tang, Y.; Ai, Q.; Wu, Z.; Liu, Y. DRAGIN: Dynamic Retrieval Augmented Generation based on the Real-time Information Needs of Large Language Models **2024**. https://doi.org/10.18653/v1/2024.acl-long.702.
35. Jeong, S.; Baek, J.; Cho, S.; Hwang, S.J.; Park, J.M. Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity **2024**. https://doi.org/10.18653/v1/2024.naacl-long.389.
36. Li, Z.; Li, C.; Zhang, M.; Mei, Q.; Bendersky, M. Retrieval Augmented Generation or Long-Context LLMs? A Comprehensive Study and Hybrid Approach **2024**. https://doi.org/10.18653/v1/2024.emnlp-industry.66.
37. Sun, W.; Yan, L.; Ma, X.; Wang, S.; Ren, P.; Chen, Z.; Yin, D.; Ren, Z. Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agents **2023**. https://doi.org/10.18653/v1/2023.emnlp-main.923.
38. Zhu, Y.; Yuan, H.; Wang, S.; Liu, J.; Liu, W.; Deng, C.; Chen, H.; Liu, Z.; Dou, Z.; Wen, J.R. Large Language Models for Information Retrieval: A Survey. *ACM Trans. Inf. Syst.* **2025**, *44*. https://doi.org/10.1145/3748304.
39. Hersh, W. Search still matters: information retrieval in the era of generative AI. *Journal of the American Medical Informatics Association* **2024**. https://doi.org/10.1093/jamia/ocae014.
40. Ibrahim, N.; AboulEla, S.; Ibrahim, A.; Kashef, R. A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): models, evaluation metrics, benchmarks, and challenges. *Discover Artificial Intelligence* **2024**. https://doi.org/10.1007/s44163-024-00175-8.
41. Wang, S.; Fan, W.; Feng, Y.; Shanru, L.; Ma, X.; Wang, S.; Yin, D. Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation **2025**. https://doi.org/10.18653/v1/2025.acl-long.1317.
42. Sun, K.; Xu, Y.; Zha, H.; Liu, Y.; Dong, X.L. Head-to-Tail: How Knowledgeable are Large Language Models (LLMs)? A.K.A. Will LLMs Replace Knowledge Graphs? **2024**. https://doi.org/10.18653/v1/2024.naacl-long.18.
43. Xie, R.; Liu, Z.; Jia, J.; Luan, H.; Sun, M. Representation Learning of Knowledge Graphs with Entity Descriptions. *Proceedings of the AAAI Conference on Artificial Intelligence* **2016**. https://doi.org/10.1609/aaai.v30i1.10329.
44. Choi, B.; Jang, D.; Ko, Y. MEM-KGC: Masked Entity Model for Knowledge Graph Completion With Pre-Trained Language Model. *IEEE Access* **2021**. https://doi.org/10.1109/access.2021.3113329.
45. Wang, X.; Wang, D.; Xu, C.; He, X.; Cao, Y.; Chua, T. Explainable Reasoning over Knowledge Graphs for Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence* **2019**. https://doi.org/10.1609/aaai.v33i01.33015329.
46. Janowicz, K.; Raubal, M.; Kühn, W. The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science* **2011**. https://doi.org/10.5311/josis.2011.2.3.
47. Hollink, V.; Tsikrika, T.; de Vries, A.P. Semantic search log analysis: A method and a study on professional image search. *Journal of the American Society for Information Science and Technology* **2011**. https://doi.org/10.1002/asi.21484.
48. Hasan, R. ENHANCING MARKET COMPETITIVENESS THROUGH AI-POWERED SEO AND DIGITAL MARKETING STRATEGIES IN E-COMMERCE **2025**. https://doi.org/10.63125/31tpjc54.
49. Raji, M.A.; Olodo, H.B.; Oke, T.T.; Addy, W.A.; Ofodile, O.C.; Oyewole, A.T. E-commerce and consumer behavior: A review of AI-powered personalization and market trends. *GSC Advanced Research and Reviews* **2024**. https://doi.org/10.30574/gscarr.2024.18.3.0090.
50. Kedi, W.E.; Ejimuda, C.; Idemudia, C.; Ijomah, T.I. AI software for personalized marketing automation in SMEs: Enhancing customer experience and sales. *World Journal of Advanced Research and Reviews* **2024**. https://doi.org/10.30574/wjarr.2024.23.1.2159.
51. Kathiriya, S.; Mullapudi, M.; Karangara, R. Optimizing ECommerce Listing: LLM Based Description and Keyword Generation from Multimodal Data. *International Journal of Science and Research (IJSR)* **2023**. https://doi.org/10.21275/sr24304113521.
52. Purnomo, Y.J. Digital Marketing Strategy to Increase Sales Conversion on E-commerce Platforms. *Journal of Contemporary Administration and Management (ADMAN)* **2023**. https://doi.org/10.61100/adman.v1i2.23.
53. Shaban, A.A.; Zeebaree, S.R.M. A comprehensive review of digital marketing strategies for e-commerce success. *International Journal of Scientific World* **2025**. https://doi.org/10.14419/th8ey858.

54. Kamkankaew, P.; Thanitbenjasith, P.; Sribenjachot, S.; Sanpatanon, N.; Phattarowas, V.; Thanin, P. How Artificial Intelligence is Helping Businesses Grow and Thrive: The Transformative Role of Artificial Intelligence in Thai B2C Digital Marketing. *International Journal of Sociologies and Anthropologies Science Reviews* **2024**. https://doi.org/10.60027/ijsasr.2024.3651.
55. Abdelkader, O.A. ChatGPT's influence on customer experience in digital marketing: Investigating the moderating roles. *Heliyon* **2023**. https://doi.org/10.1016/j.heliyon.2023.e18770.
56. Arslan, M.; Munawar, S.; Riaz, Z.; Cruz, C. Large language models for business and management applications: A review. *Information Processing & Management* **2026**, *63*, 104864. https://doi.org/10.1016/j.ipm.2026.104864.
57. Sharabati, A.A.; Ali, A.A.A.; Allahham, M.; Hussein, A.M.A.; Alheet, A.F.; Mohammad, A.S. The Impact of Digital Marketing on the Performance of SMEs: An Analytical Study in Light of Modern Digital Transformations. *Sustainability* **2024**. https://doi.org/10.3390/su16198667.
58. Ali, M.M.W.; Gutiérrez, J.O. Digital Marketing: Strategies, Challenges, and Opportunities in the Digital Technology. *Global Journal of Economic and Business* **2025**. https://doi.org/10.31559/gjeb2025.15.1.3.
59. Singh, D.N.K.; Peters, M.S.; Bhagat, P. Digital Marketing Strategies and Consumer Engagement: A Systematic Review of Emerging Trends, Technologies, and Measurement Frameworks. *Milky Way Scientific Journal* **2026**. https://doi.org/10.65021/mwsj.v1.i2.26.
60. Hutsaliuk, O.; Mirzoiev, D. Formation of digital marketing tools in modern conditions. *Actual problems of innovative economy and law* **2025**. https://doi.org/10.36887/2524-0455-2025-1-12.
61. Joel, O.S.; Oyewole, A.T.; Odunaiya, O.G.; Soyombo, O.T. The impact of digital transformation on business development strategies: Trends, challenges, and opportunities analyzed. *World Journal of Advanced Research and Reviews* **2024**. https://doi.org/10.30574/wjarr.2024.21.3.0706.
62. Sihotang, H.K.N.; Hudrasyah, H. Proposed Digital Marketing Strategy Using Existing Customer Journey Analysis Case Study: PT SS. *International Journal of Current Science Research and Review* **2023**. https://doi.org/10.47191/ijcsrr/v6-i3-07.
63. Patil, G.B.; Padyana, U.K.; Rai, H.P.; Ogeti, P.; Fadnavis, N.S. Personalized Marketing Strategies Through Machine Learning: Enhancing Customer Engagement. *Journal of Informatics Education and Research* **2021**. https://doi.org/10.52783/jier.v1i1.1345.
64. Singh, S.U.; Namin, A.S. A survey on chatbots and large language models: Testing and evaluation techniques. *Natural Language Processing Journal* **2025**, *10*, 100128. https://doi.org/10.1016/j.nlp.2025.100128.
65. Pokhrel, S.; Bina, K.C.; Shah, P. A Practical Application of Retrieval-Augmented Generation for Website-Based Chatbots: Combining Web Scraping, Vectorization, and Semantic Search. *Journal of Trends in Computer Science and Smart Technology* **2025**. https://doi.org/10.36548/jtcsst.2024.4.007.
66. Mazumder, J.; Mukhopadhyay, P. Designing Question-Answer Based Search System in Libraries: Application of Open Source Retrieval Augmented Generation (RAG) Pipeline. *Journal of Information and Knowledge* **2024**. https://doi.org/10.17821/srels/2024/v61i5/171583.
67. Dai, S.; Chen, X.; Xu, S.; Pang, L.; Dong, Z.; Xu, J. Bias and Unfairness in Information Retrieval Systems: New Challenges in the LLM Era **2024**. https://doi.org/10.1145/3637528.3671458.
68. Bernard, N.; Balog, K. A Systematic Review of Fairness, Accountability, Transparency, and Ethics in Information Retrieval. *ACM Computing Surveys* **2023**. https://doi.org/10.1145/3637211.
69. Feng, Y.; Shah, C. Has CEO Gender Bias Really Been Fixed? Adversarial Attacking and Improving Gender Fairness in Image Search. *Proceedings of the AAAI Conference on Artificial Intelligence* **2022**. https://doi.org/10.1609/aaai.v36i11.21445.
70. Castillo, C. Fairness and Transparency in Ranking. *ACM SIGIR Forum* **2019**. https://doi.org/10.1145/3308774.3308783.
71. Saura, J.R.; Škare, V.; Đurđana Ozretić Došen. Is AI-based digital marketing ethical? Assessing a new data privacy paradox. *Journal of Innovation & Knowledge* **2024**. https://doi.org/10.1016/j.jik.2024.100597.
72. Krugmann, J.O.; Hartmann, J. Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions* **2024**. https://doi.org/10.1007/s40547-024-00143-4.
73. Miah, M.S.U.; Kabir, M.M.; Sarwar, T.B.; Safran, M.; Alfarhood, S.; Mridha, M.F. A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. *Scientific Reports* **2024**. https://doi.org/10.1038/s41598-024-60210-7.
74. Rane, N.L.; Desai, P.; Rane, J.; Mallick, S.K. Using artificial intelligence, machine learning, and deep learning for sentiment analysis in customer relationship management to improve customer experience, loyalty, and satisfaction; 2024. https://doi.org/10.70593/978-81-981367-4-9_7.

doi:10.20944/preprints202607.0744.v1

75. Paul, R.; Imam, M.H.; Mou, A.J. AI-POWERED SENTIMENT ANALYSIS IN DIGITAL MARKETING: A REVIEW OF CUSTOMER FEEDBACK LOOPS IN IT SERVICES. *American Journal of Scholarly Research and Innovation* **2023**. https://doi.org/10.63125/61pqqq54.
76. Ajiga, D.I.; Ndubuisi, N.L.; Asuzu, O.F.; Owolabi, O.R.; Tubokirifuruar, T.S.; Adeleye, R.A. AI-DRIVEN PREDICTIVE ANALYTICS IN RETAIL: A REVIEW OF EMERGING TRENDS AND CUSTOMER ENGAGEMENT STRATEGIES. *International Journal of Management & Entrepreneurship Research* **2024**. https://doi.org/10.51594/ijmer.v6i2.772.
77. Cvitanović, P.L. Digital Marketing Benchmarks Leveraged by Marketing Analytics Tools. *Hrčak Portal of scientific journals of Croatia (University Computing Centre)* **2019**. https://doi.org/10.13140/rg.2.2.13289.65129.
78. Chen, L.C.; Pardeshi, M.S.; Liao, Y.; Pai, K.C. Application of retrieval-augmented generation for interactive industrial knowledge management via a large language model. *Computer Standards & Interfaces* **2025**. https://doi.org/10.1016/j.csi.2025.103995.
79. Iaroshev, I.; Pillai, R.; Vaglietti, L.; Hanne, T. Evaluating Retrieval-Augmented Generation Models for Financial Report Question and Answering. *Applied Sciences* **2024**. https://doi.org/10.3390/app14209318.
80. Toro, S.; Anagnostopoulos, A.V.; Bello, S.M.; Blumberg, K.; Cameron, R.; Carmody, L.; Diehl, A.D.; Dooley, D.; Duncan, W.D.; Fey, P.; et al. Dynamic Retrieval Augmented Generation of Ontologies using Artificial Intelligence (DRAGON-AI). *Journal of Biomedical Semantics* **2024**. https://doi.org/10.1186/s13326-024-00320-3.