Retrieval-augmented generation in marketing: an integrative literature review of architectures, applications and strategic implications

Journal of Research in Interactive Marketing

Mayemba Karel Nzita and Marius Wait
Department of Marketing Management,
University of Johannesburg – Auckland Park Kingsway Campus,
Johannesburg, South Africa

Received 23 February 2026
Revised 9 April 2026
1 June 2026
16 June 2026
Accepted 25 June 2026

Abstract

Purpose – This study synthesises and critically interprets the emerging literature on retrieval-augmented generation (RAG) in marketing. It examines how RAG architectures, applications and governance conditions shape interactive marketing, customer engagement, market sensing and responsible AI deployment.

Design/methodology/approach – An integrative literature review guided by the SPIDER framework was conducted across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, Google Scholar and Emerald Insight for publications from 2020 to 2025. A final corpus of 25 marketing-relevant RAG studies was analysed through qualitative coding and thematic synthesis. Adjacent peer-reviewed literature on AI in marketing, recommender systems, customer analytics, data-driven marketing capability, AI-augmented co-creation and immersive digital environments was used to contextualise, rather than expand, the primary corpus.

Findings – Five themes emerge: adaptive and knowledge-graph-enhanced RAG architectures; multilingual, real-time and hyper-personalised customer interaction; intelligent and context-aware retrieval; marketing intelligence and decision support; and robustness, hallucination mitigation and responsible deployment. The review shows that RAG can support retrieval-grounded interaction and market sensing, but evidence is stronger for technical and prototype demonstrations than for longitudinal customer, conversion or strategic outcomes. Responsible deployment requires transparency, privacy safeguards, fairness controls, human oversight and auditable knowledge sources.

Originality/value – The study advances a marketing-specific interpretation of RAG as a socio-technical marketing capability. It contributes by shifting personalisation from predictive targeting to retrieval-grounded interaction, linking market sensing to dynamic knowledge grounding and positioning governance as intrinsic to the capability rather than a downstream safeguard.

Keywords Retrieval-augmented generation, Interactive marketing, Marketing artificial intelligence, Large language models, Customer personalisation, Knowledge graphs, Marketing intelligence, Responsible AI, Integrative literature review

Paper type Literature review

1. Introduction

Generative artificial intelligence has rapidly entered marketing practice, but its adoption has exposed a persistent tension between automation and accountability. Marketing organisations increasingly use large language models (LLMs) to produce customer-facing content, support service interactions, localise messages and analyse market signals. The scale of this shift is considerable: the global market for AI in marketing was valued at approximately USD 47 billion in 2025 and is projected to exceed USD 107 billion by 2028 (Dencheva, 2023). Yet adoption is outpacing safeguards, with more than 70% of marketers reporting an AI-related incident, such as hallucination, bias or off-brand content, and fewer than 35% planning to increase investment in AI governance or brand-integrity oversight (Koch and Pellatt, 2025).

Journal of Research in Interactive Marketing
Emerald Publishing Limited
e-ISSN: 2040-7130
p-ISSN: 2040-7122
DOI 10.1108/JRIM-02-2026-0130

Purely parametric models remain vulnerable to outdated knowledge, unsupported claims, inconsistent brand voice and hallucinated outputs. These weaknesses are consequential in marketing because inaccurate or misleading content can damage trust, distort customer decision-making and create regulatory exposure. RAG has therefore attracted attention because it grounds generation in retrieved external knowledge rather than relying only on a model's internal parameters (Lewis *et al.*, 2020; Fan *et al.*, 2024; Zhao *et al.*, 2026).

RAG is better understood as an architectural logic than as a single tool. A retriever identifies relevant evidence from external sources, such as product catalogues, customer records, policy documents, campaign repositories or knowledge graphs, and a generator uses that evidence to produce a response. In marketing, this architecture is attractive because customer interactions often require current product information, brand-specific phrasing, contextual sensitivity and traceable support for claims. However, retrieval augmentation does not automatically solve the reliability problem. Evidence from high-stakes AI applications shows that retrieval can reduce, but does not eliminate, hallucination and error (Magesh *et al.*, 2025). For marketing scholars, the central question is therefore not whether RAG is technically promising, but how it changes the conditions under which AI-mediated marketing interaction can be trusted, personalised and governed.

This distinction matters because the risks of generative AI in marketing are not limited to fabricated statements. A model may produce an accurate sentence but draw from an unauthorised source, apply a message to the wrong segment or ignore restrictions on promotional claims. RAG creates an opportunity to manage such risks through curated knowledge, but it also makes a firm's knowledge architecture more visible as a determinant of marketing quality. Poorly maintained repositories, fragmented ownership of product information and weak campaign governance can surface directly in generated interactions.

Existing RAG surveys primarily emphasise technical design, retrieval strategies, evaluation benchmarks and robustness (Brown *et al.*, 2025; Fan *et al.*, 2024; Zhao *et al.*, 2026). This work is valuable, but it leaves underdeveloped the specific implications of RAG for marketing. Marketing is a distinctive deployment context because value is created through interaction, interpretation, persuasion and relationship continuity. A response can be technically accurate yet commercially inappropriate if it violates brand tone, retrieves content intended for another segment, ignores privacy expectations or gives advice beyond the firm's authorised claims. Marketing data are also heterogeneous and time-sensitive, combining structured CRM data, transactional histories, product catalogues and campaign metrics with unstructured reviews, social media conversations and competitive intelligence.

The wider marketing literature has examined AI-enabled customer interaction, recommender systems, marketing analytics, customer engagement, data-driven decision-making and immersive digital environments (Davenport *et al.*, 2020; De Bruyn *et al.*, 2020; Huang and Rust, 2021; Wedel and Kannan, 2016). Recent work on AI-augmented co-creation shows how human-AI collaboration in digital touchpoints can shape consumer empowerment, creativity and role identity (Peng and Haq, 2026), while metaverse research highlights the need to understand interaction in immersive, high-velocity and data-rich environments (Yin and Do, 2025). These adjacent streams do not substitute for RAG evidence; they provide theoretical context for interpreting what retrieval grounding adds to interactive marketing.

This study responds by offering an integrative review of RAG in marketing. It makes three linked contributions. First, it reinterprets RAG as a socio-technical marketing capability rather than as a generic technical enhancement to LLMs. Second, it connects RAG to service-dominant logic, customer engagement theory, relationship marketing and market orientation. Third, it clarifies where evidence is empirically grounded, where it is prototype-based and where conclusions remain conceptual. This evidential distinction is necessary because marketing-specific RAG scholarship remains emergent.

The review is aligned with the journal's emphasis on contribution through a coherent scholarly story rather than a descriptive catalogue of technologies (Wang, 2025). The story developed here is that RAG becomes important to marketing when three conditions meet:

customer interaction becomes conversational, organisational knowledge becomes too dispersed for manual retrieval and governance demands make unsupported generative output unacceptable. The study therefore asks: How are RAG architectures being configured for marketing contexts? What marketing applications and strategic implications are reported in the literature? What theoretical, managerial and policy implications follow from the current evidence base?

2. Theoretical background and literature review

2.1 RAG and the marketing interaction problem

RAG was formalised as a method for knowledge-intensive natural language processing in which generated outputs are conditioned on retrieved evidence (Lewis *et al.*, 2020). Subsequent work has developed hybrid retrieval, graph-based retrieval, agentic retrieval and evaluation approaches to improve grounding, relevance and robustness (Brown *et al.*, 2025; Fan *et al.*, 2024; Lyu *et al.*, 2025; Zhao *et al.*, 2026). At a basic level, a retriever searches external knowledge sources, a generator integrates the retrieved evidence into a response and a fusion step balances retrieved knowledge with the model's internal representations. The answer is therefore anchored in an external knowledge environment whose quality, currency and permissions shape the output.

For marketing, the significance of RAG lies in the interaction problem. Customers ask questions shaped by context, preference, timing, language, perceived risk and relationship history, while marketers operate under constraints related to brand positioning, product availability, legal claims, privacy rules and channel-specific tone. RAG can retrieve context-specific information before responding, potentially aligning the output with the customer's situation and the firm's authorised knowledge base. However, weak indexing, outdated repositories, biased ranking or poor access controls can still produce outputs that are plausible but unsuitable.

The technical literature shows that RAG quality depends on both retrieval and generation. Retrieval must locate relevant, current and authorised evidence; generation must transform that evidence into a coherent and contextually appropriate output. Evaluation frameworks increasingly recognise this dual requirement by assessing retrieval relevance, faithfulness, answer quality and robustness (Elmitwalli *et al.*, 2025; Lyu *et al.*, 2025; Yu *et al.*, 2025). In marketing, these criteria should be expanded to include brand consistency, consumer comprehension, fairness, privacy sensitivity and channel appropriateness.

2.2 Positioning RAG within marketing theory

The interactive marketing perspective emphasises technology-mediated, bidirectional exchanges between firms and consumers (Wang, 2024). RAG fits this perspective because retrieved knowledge can shape the substance, timing and tone of an interaction in real time. Rather than treating AI as an automated content machine, the interactive lens positions RAG as a mechanism through which firms assemble knowledge during interaction. This interpretation is consistent with the broader argument that AI is changing marketing research and practice in ways that require explicit theoretical development (Wang, 2026). Gao and Liu (2023) locate AI-enabled personalisation across customer-journey touchpoints, while Peltier *et al.* (2024) organise AI in interactive marketing around antecedents to AI use, interactive usage contexts and AI-enabled value co-creation outcomes. This review extends those accounts by identifying retrieval grounding as the mechanism through which interactional knowledge is selected, authorised and made auditable.

Service-dominant logic views value as co-created through resource integration, especially knowledge and competences (Vargo and Lusch, 2004, 2016). RAG can be conceptualised as a resource-integration mechanism because it assembles organisational knowledge, customer context and external market information into a response that supports customer or managerial action. Customer engagement theory clarifies how RAG-enabled interaction may affect cognitive, emotional and behavioural engagement (Brodie *et al.*, 2011; Hollebeek, 2011). Retrieval-grounded responses can support cognitive engagement through relevant information, emotional engagement through tone adaptation and behavioural engagement by assisting decisions along the customer journey.

Taken together, these theories specify what is new about RAG for marketing. The technology does not create value merely because it generates fluent text. Its theoretical significance lies in mediating between customers' demand for immediate relevance and organisations' need to preserve knowledge discipline. In this respect, RAG can be conceptualised as an interface capability: it turns dispersed organisational and market knowledge into interactional resources at the point of customer or managerial need.

Relationship marketing highlights the role of trust, commitment and continuity in firm-customer relationships (Palmatier *et al.*, 2006). RAG can support continuity when systems retrieve prior interactions, preferences or service histories. However, hallucinated claims, inappropriate personalisation or biased retrieval can damage the trust the technology is intended to strengthen. Market orientation is also relevant because it defines marketing capability in terms of generating, disseminating and responding to market intelligence (Kohli and Jaworski, 1990; Narver and Slater, 1990). RAG may operationalise market sensing by retrieving and synthesising signals, but the resulting intelligence is credible only when retrieval quality and source provenance are governed.

2.3 Adjacent literature and boundary-setting

The narrow body of explicitly RAG-labelled marketing research cannot, on its own, support broad claims about the transformation of marketing. This review therefore retains the 25 RAG studies as the primary corpus while using adjacent peer-reviewed literature as a contextual lens. AI-in-marketing and marketing analytics scholarship explains how intelligent systems shape customer interaction, automation, targeting and data-rich decision-making (Davenport *et al.*, 2020; De Bruyn *et al.*, 2020; Huang and Rust, 2021; Wedel and Kannan, 2016). Recommender-system and personalisation research clarifies algorithmic relevance and preference learning, while chatbot and deep-learning studies situate RAG within a longer trajectory of intelligent customer interaction (Pantano and Pizzi, 2020; Yu and Liu, 2025).

This distinction is especially important for recommender systems. Recommenders typically infer relevance from behavioural histories, item attributes and similarity structures, whereas RAG retrieves explicit evidence that can be inspected, updated and governed. The approaches are therefore complementary rather than interchangeable. A recommender may identify which product is likely to be relevant; a RAG system may explain that recommendation using current product specifications, stock information, service policies or customer-specific constraints. This difference strengthens the theoretical claim that RAG introduces retrieval-grounded interaction rather than merely intensifying algorithmic targeting.

Adjacent work also extends the lens to AI-augmented co-creation and immersive digital environments. Peng and Haq (2026) show that human-AI collaboration in co-creation contexts can influence empowerment, creativity and role identity. Yin and Do (2025) show that metaverse consumer behaviour requires attention to immersive interaction and contextual boundary conditions, while Yao and Ren (2025) frame metaverse brand experience as a strategically orchestrated interaction system. These studies indicate where RAG may matter in future immersive and co-creative settings, but they are not treated as evidence that RAG already produces superior marketing outcomes.

This boundary-setting avoids conflating adjacent AI evidence with demonstrated RAG outcomes. Adjacent studies are used to interpret why RAG matters for marketing theory and where future empirical work should be directed; they are not treated as evidence that RAG itself produces superior marketing outcomes.

3. Methodology

An integrative literature review (ILR) was selected because the research area is emergent, interdisciplinary and methodologically heterogeneous. ILRs are appropriate when diverse conceptual, empirical and technical contributions must be synthesised to generate clearer theoretical understanding rather than merely count prior findings (Cronin and George, 2023; Torraco, 2016; Whittemore and Knafl, 2005).

The review was guided by the SPIDER framework, which supports qualitative and mixed-evidence synthesis (Cooke *et al.*, 2012). The sample comprised academic publications addressing RAG in enterprise, commercial, e-commerce, customer service or consumer-facing settings. The phenomenon of interest was the adaptation, evaluation and strategic use of RAG for marketing-relevant requirements; evaluation focused on architecture, applications, reported outcomes, governance and evidential strength. Table 1 summarises the SPIDER framework.

Searches were conducted across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, Google Scholar and Emerald Insight. The common search logic combined "retrieval-augmented generation" OR "RAG" with marketing, interactive marketing, digital marketing, e-commerce, customer service, customer support, chatbot, shopping assistant and customer experience. The search covered 2020–2025 because RAG was formalised in 2020 (Lewis *et al.*, 2020), and backward snowballing was used to identify additional relevant sources.

Studies were included when they explicitly addressed RAG and had a clear connection to marketing, e-commerce, customer-facing interaction, customer service, market intelligence or commercial decision support. They also had to provide sufficient information about architecture, application, evaluation or strategic implication. Studies were excluded when they discussed generative AI without retrieval augmentation, focused solely on non-marketing domains without a customer-facing or commercial link, or consisted of non-scholarly commentary.

The search initially produced 204 records. After removing 29 duplicates, 175 records were screened by title, abstract and keywords. One hundred and thirty-two records were excluded, 43 full texts were assessed and 18 were removed because they lacked a genuine retrieval-augmented component, clear marketing relevance or extractable methodological detail. The final primary corpus comprised 25 studies: six journal articles, 14 conference or proceedings papers and five preprints. This composition reflects the field's early stage rather than a narrow search strategy.

Table 1. SPIDER template guiding the research question

SPIDER element	Definition in this study
Sample (S)	Academic publications published between 2020 and 2025 that explicitly address RAG and are linked to marketing or customer-facing contexts, including e-commerce, customer service and chat assistants
Phenomenon of Interest (PI)	How RAG architectures are configured, applied and discussed in marketing-relevant contexts, including use cases, opportunities, constraints and governance implications
Design (D)	Conceptual papers, technical system papers, empirical studies, conference papers and journal articles meeting the inclusion criteria
Evaluation (E)	Architectural features, marketing applications, reported benefits and limitations, strategic implications and governance considerations
Research type (R)	Integrative review of heterogeneous evidence, combining conceptual, empirical and technical contributions

Source(s): Authors' own work

The final corpus was uneven in evidential strength. Some papers reported operational prototypes, some offered benchmark or system-level evaluation, and others presented conceptual or architectural arguments. Peer-reviewed journal articles and conference papers were used to support stronger interpretive claims, while preprints were retained mainly where they reflected emerging technical developments without a peer-reviewed equivalent. This treatment responds to the rapid publication cycle of AI research without allowing unstable sources to carry the main theoretical or managerial contribution.

Data were extracted into a structured matrix capturing author, year, outlet, study type, RAG architecture, knowledge source, marketing context, application task, evaluation approach, reported findings and limitations. Table 2 below presents the data extraction framework.

Coding proceeded through iterative data display, reduction, clustering and pattern identification, following ILR guidance (Whittemore and Knafl, 2005). *In vivo* coding preserved source terms such as hybrid RAG, knowledge graph, multilingual chatbot and hallucination mitigation. Axial coding then grouped these into higher-order categories, including adaptive architecture, customer interaction, context-aware retrieval, marketing intelligence and responsible deployment. Two researchers coded a subset of studies independently and resolved discrepancies through discussion.

The synthesis deliberately avoided descriptive keyword visuals such as word clouds because they add limited analytical value in an integrative review. Instead, the analysis foregrounded conceptual relationships, tensions and evidential strength. Theme development drew on the 4 R logic of recognisability, relevance, responsiveness and reciprocity (Naeem *et al.*, 2023), while evidence classification prevented prototype demonstrations from being treated as established managerial outcomes.

Table 2. Data extraction and evidence-classification framework

Extraction category	Information captured	Purpose for synthesis
Bibliographic profile	Author, year, title, outlet and publication type	Distinguished journal articles, conference papers, system papers and preprints, allowing claims to be weighted according to evidential stability
RAG configuration	Retriever type, generator, knowledge source, graph or hybrid component, and deployment logic where reported	Enabled comparison of architectural choices and their relevance to marketing tasks
Marketing context	Customer-facing setting, e-commerce, customer support, product information, campaign content, market intelligence or decision support	Ensured that included studies had a clear marketing, commercial or customer-interaction connection
Application task	Question answering, chatbot support, product recommendation, translation, content assistance, internal knowledge search or intelligence synthesis	Supported clustering into the five thematic streams reported in the findings
Evaluation and findings	Reported metrics, system comparisons, prototype outcomes, user-facing evidence and stated limitations	Separated demonstrated system performance from broader conceptual or managerial implications
Evidence classification	Empirical evidence, prototype or system-level evidence, conceptual argument, or adjacent contextual literature	Prevented prototype demonstrations from being treated as established organisational outcomes

Source(s): Authors' own work

4. Analysis and findings

4.1 Thematic framework and evidence logic

The thematic synthesis produced five interrelated themes: adaptive RAG architectures for complex marketing environments; multilingual, real-time and hyper-personalised customer interaction; intelligent and context-aware knowledge retrieval; marketing intelligence, forecasting and strategic decision support; and robustness, hallucination mitigation and responsible deployment. These themes should not be read as a list of applications. They form a connected framework in which architecture shapes retrieval quality, retrieval quality shapes interaction and intelligence, and governance conditions determine whether the capability can be deployed responsibly.

The evidence base is uneven across themes. Claims about architecture, retrieval design and system functionality are relatively well supported by technical and prototype studies. Claims about customer engagement, loyalty, conversion and managerial performance are less mature because few studies provide longitudinal, field-based or comparative evidence in live marketing environments. The findings therefore distinguish between what the reviewed studies demonstrate, what they suggest and what remains plausible but unverified. Table 3 presents the thematic framework and marketing implications.

4.2 Adaptive RAG architectures for complex marketing environments

The first theme concerns the movement from generic RAG pipelines towards adaptive architectures for complex marketing environments. The reviewed studies converge on modular, hybrid and dual-phase configurations in which retrieval and generation can be tuned separately. Retrieval modules are optimised for relevance, source quality and domain specificity, while generation modules are tuned for coherence, tone and brand consistency (Brown *et al.*, 2025; Fan *et al.*, 2024). This separation is useful because product catalogues, campaign calendars, promotional claims and consumer preferences change quickly.

The literature also points to a shift from retrieval as document search to retrieval as business-context assembly. In marketing, the system must assemble product rules, customer eligibility, campaign objectives, preferred tone, channel restrictions and service policies. This broader form of retrieval is why architectural choices become strategically relevant for marketing scholars.

Table 3. Thematic framework, evidence base and marketing implications

Theme	Evidential character	Principal marketing implication
Adaptive RAG architectures for complex marketing environments	Mostly technical and comparative evidence, with some applied demonstrations	Architecture matters for brand coherence, control and scalability
Multilingual, real-time and hyper-personalised customer interaction	Promising prototypes and conceptual arguments; limited field validation	Personalisation value depends on retrieval quality, privacy safeguards and contextual fit
Intelligent and context-aware knowledge retrieval	Relatively stronger technical evidence on coherence and relevance improvements	Retrieval precision is central to trustworthy marketing outputs
Marketing intelligence, forecasting and strategic decision support	Largely conceptual or pilot-stage evidence	Strategic applications are promising but still require real-world validation
Robustness, hallucination mitigation and responsible AI deployment	Supported by evaluation studies and governance-oriented scholarship	Trustworthy deployment is a strategic necessity, not an optional add-on

Source(s): Authors' own work

Knowledge-graph-enhanced RAG is especially relevant because knowledge graphs encode relationships among products, attributes, customer segments, campaign assets, brands and policies. Such relationships are central to meaningful retrieval in tasks involving product compatibility, regional availability, customer history or cross-sell logic. Graph-based retrieval can preserve these relationships more effectively than simple document matching (Peng *et al.*, 2025; Xu *et al.*, 2024b; Zhu *et al.*, 2025), although the stronger claims currently concern retrieval coherence rather than broad commercial performance.

The main tension is between adaptability and control. Modular systems can be updated more easily than monolithic models, but they also increase the number of components requiring monitoring. Retrieval indices, embedding models, source repositories, access permissions and generation rules all become potential points of failure. In a marketing context, architecture cannot be separated from governance: a system designed for rapid content adaptation may also increase the risk of outdated retrieval, off-brand output or unsupported product claims if provenance and validation are weak.

4.3 Multilingual, real-time and hyper-personalised customer interaction

The second theme concerns RAG as a basis for customer-facing interaction. Studies describe RAG-enabled chat assistants, e-commerce question-answering systems and customer support tools that retrieve product, policy or customer-specific information before generating responses (Benita *et al.*, 2024; De Freitas and Lotufo, 2024; Tangarajan *et al.*, 2025). Compared with static FAQs, RAG enables responses more closely tied to the query, the available knowledge base and the firm’s current operating context.

Nevertheless, the evidence does not justify treating RAG-enabled personalisation as automatically superior to established marketing automation. The reviewed studies demonstrate that RAG can produce context-sensitive responses, but they rarely compare those responses with human service agents, rule-based systems or recommender-driven interfaces using behavioural outcomes. Until comparative evidence is available, RAG should be positioned as a promising mechanism for interactive personalisation rather than as an established driver of loyalty or conversion.

RAG contributes to personalisation at three levels. At the query level, it interprets customer intent in relation to context or journey stage. At the retrieval level, it selects evidence relevant to segment, language, location or purchase situation. At the response level, it converts that evidence into an appropriate tone and format. This tri-level view reframes personalisation as retrieval-grounded interaction rather than predictive targeting alone.

This argument is consistent with adjacent work on AI-augmented co-creation. Peng and Haq (2026) show that human-AI collaboration can shape empowerment, creativity and role identity. For RAG, this suggests that customer-facing systems should be evaluated not only for answer accuracy but also for perceived agency, co-creation quality and relationship continuity, although the RAG corpus provides limited longitudinal evidence on these outcomes.

Personalisation quality depends on the interaction between retrieved data and conversational framing. Transaction-only retrieval may be efficient but impersonal, while unfiltered behavioural retrieval may appear intrusive. Curated preference and service data may support continuity without undermining autonomy. Marketing-relevant evaluation should therefore include helpfulness, intrusiveness, trust, explanation quality and relationship continuity.

Multilingual capability expands RAG’s relevance for international marketing but creates risks around tone drift, cultural misalignment and inconsistent claims. A translated response may still be inappropriate if it retrieves content intended for another market or ignores local regulation. Multilingual RAG therefore requires market-specific retrieval boundaries, local content validation and monitoring of brand voice across languages.

4.4 Intelligent and context-aware knowledge retrieval

The third theme identifies retrieval as the decisive mechanism in marketing RAG. Marketing knowledge is fragmented across product information systems, CRM records, social platforms, call-centre transcripts, advertising guidelines and competitor intelligence. RAG adds value only when these sources are indexed, ranked and governed in ways that reflect the marketing task.

This insight changes how marketing teams should think about data integration. RAG does not require all data to be merged into one warehouse, but it does require sources to be discoverable, interpretable and governed. Metadata, taxonomies, ontologies and knowledge graphs therefore become strategically important because they support contextually appropriate retrieval rather than simple text similarity.

Several studies highlight the limits of conventional vector retrieval where marketing meaning depends on relationships. Reviews, product variants, regions, campaigns and service incidents may be interdependent, while campaign messages may be valid only for particular channels or periods. Graph-based and hybrid retrieval can mitigate this problem by combining semantic relevance with structured relationships and metadata (Peng *et al.*, 2025; Xu *et al.*, 2024a, b).

Context-aware retrieval also requires temporal sensitivity. Prices, promotions, delivery windows and competitor claims can change within days or hours. RAG systems that retrieve stale information can create a risk similar to hallucination because the generated response may be fluent and source-supported, while the source itself is outdated. Retrieval pipelines in marketing should therefore include freshness controls, source-ranking rules, audit trails and exclusion of superseded content.

A related boundary condition concerns source authority. Marketing repositories often contain drafts, outdated campaign materials, internal notes and approved claims. RAG systems that do not distinguish among these materials may retrieve text that is semantically relevant but not authorised for public use. Retrieval governance should therefore classify sources by status, audience, validity period and permissible use.

The strongest implication is that retrieval is not a neutral technical step. It encodes organisational priorities, access rights, market boundaries and assumptions about relevance. Retrieval design is therefore a marketing governance issue, requiring clarity about usable sources, source ownership and auditability.

4.5 Marketing intelligence, forecasting and strategic decision support

The fourth theme extends RAG beyond front-office interaction into marketing intelligence and decision support. RAG systems can retrieve and synthesise information from CRM platforms, sales reports, campaign dashboards, social media sentiment, product reviews, competitor websites and external market reports. This potential aligns with market orientation theory because RAG may support the generation and dissemination of market intelligence (Kohli and Jaworski, 1990; Narver and Slater, 1990).

RAG may also support cross-functional coordination because marketing decisions often require evidence from sales, product, customer service, finance and compliance. A retrieval-grounded system can help synthesise these perspectives into a decision brief. However, strategic intelligence is weakened when a system collapses conflicting evidence into a single confident narrative. Decision-support designs should therefore include uncertainty indicators, source comparison and explicit identification of conflicting evidence.

This theme requires careful framing. The reviewed literature offers promising conceptual and prototype evidence, but fewer studies validate strategic decision outcomes in live organisational contexts. It is more accurate to state that RAG may support marketing intelligence than to claim that it transforms decision-making. Adjacent research on data-driven innovation capability in B2B marketing is useful here because it shows that data-driven capability depends on resources, routines and learning mechanisms, not simply on technology adoption (Ravat *et al.*, 2024). RAG can become part of such a capability only when embedded in organisational processes for evidence evaluation, interpretation and action.

Plausible use cases include campaign-performance explanation, customer-feedback comparison, competitor-positioning summaries and product-development support. These applications are credible, but their effectiveness depends on evaluation of decision quality, timeliness, managerial adoption and downstream performance effects.

4.6 Robustness, hallucination mitigation and responsible deployment

The fifth theme concerns responsible deployment. RAG is often presented as a solution to hallucination because it grounds generation in external evidence. The reviewed literature supports a more cautious interpretation: RAG mitigates some hallucination risks but introduces other failure modes, including poor retrieval, conflicting sources, weak fusion, biased ranking, outdated knowledge and overconfident synthesis (Niu *et al.*, 2024; Sun *et al.*, 2025; Zhang and Zhang, 2025).

Mitigation strategies include prompt-level knowledge injection, retrieval filtering, output scoring, post-generation fact checking and human-in-the-loop review (Béchard and Marquez Ayala, 2024; Kumar *et al.*, 2024; Yao and Fujita, 2024). For marketing, oversight is most important where outputs are public-facing, legally sensitive, high value or likely to affect vulnerable consumers.

Responsible deployment also requires accountability across the RAG value chain. Technical teams may maintain retrieval infrastructure, marketing teams may define brand requirements, legal teams may govern claims and privacy teams may set data-use boundaries. If these responsibilities are not coordinated, failures may be attributed to the model even when their causes lie in source governance or organisational process.

Fairness is equally important. Retrieval systems can reproduce bias if certain products, segments, languages or sources are systematically overrepresented. Research on fair ranking in RAG demonstrates that retrieval-stage ranking affects downstream generation and exposure (Kim and Diaz, 2025). In marketing, biased retrieval may influence which products are recommended, which customer groups receive attention and how consumer needs are represented. Responsible RAG deployment must therefore include fairness testing at both retrieval and generation stages.

The responsible deployment theme connects directly to public policy. The EU AI Act includes transparency obligations for systems that interact with natural persons and for AI-generated content, making disclosure and labelling relevant to customer-facing marketing AI (European Parliament and Council, 2024). GDPR provisions on automated decision-making and profiling are relevant when RAG-supported systems use personal data to generate decisions or recommendations with significant effects (European Parliament and Council, 2016). In the United States, FTC enforcement around deceptive AI claims reinforces the need for substantiation when firms market AI capabilities or use AI in ways that may mislead consumers (Federal Trade Commission, 2024). These frameworks do not provide a complete marketing-specific RAG governance model, but they define practical constraints around transparency, evidence, disclosure and accountability.

These regulatory considerations imply that customer-facing interfaces should disclose AI interaction where required, generated content should be traceable to source materials, personal data should be limited to authorised purposes and high-risk outputs should be escalated for approval. The governance question is whether the full retrieval-generation chain can be explained, audited and corrected.

4.7 Cross-theme synthesis: tensions and boundary conditions

Across the five themes, RAG emerges as a socio-technical marketing capability rather than a discrete application. Its value depends on the interaction between architecture, knowledge infrastructure, customer context, organisational routines and governance. Four tensions are central: adaptability versus control, personalisation versus privacy, intelligence versus evidence quality and automation versus accountability.

These tensions define boundary conditions. RAG is most defensible when the knowledge base is curated, current and auditable; retrieval boundaries reflect segment, channel, language and legal constraints; outputs provide source visibility; and high-risk uses include human oversight. It is more likely to fail when repositories are outdated, consumer data are used without clear consent, multilingual retrieval ignores local context or managers treat generated synthesis as verified evidence.

The cross-theme synthesis therefore moves the paper beyond a descriptive inventory of RAG applications. It shows that the same feature can be a source of value and risk depending on context. Dynamic retrieval supports responsiveness but creates freshness and provenance problems. Personalisation supports relevance but intensifies privacy and fairness concerns. Narrative synthesis supports managerial sense-making but may obscure uncertainty. These dual effects are central to theory-building because they explain why RAG outcomes cannot be inferred from technical capability alone.

5. Contributions of the study

5.1 Contribution to theory

The study contributes to theory by reframing RAG as a socio-technical marketing capability. The theoretical advancement is specific rather than general. Recent AI-in-marketing studies have theorised AI either as a predictive and automating force that personalises along the customer journey (Gao and Liu, 2023) or as an infrastructure for value co-creation in interactive buyer–seller relationships (Peltier *et al.*, 2024), and have called for explicit theory-building as the field matures (Wang, 2026). This review extends that body of work by identifying retrieval grounding as a distinct theoretical mechanism that these accounts do not isolate: the marketing value of generative AI depends not only on what a model predicts or generates, but on which external evidence is retrieved, ranked and authorised at the moment of interaction. The new understanding that emerges is therefore that AI-mediated marketing quality becomes contingent on the firm's knowledge architecture rather than on model capability alone. First, it shifts personalisation from prediction to retrieval-grounded interaction. Existing AI-in-marketing work often emphasises targeting, recommendation and automation; RAG adds the ability to ground a response in current organisational and contextual knowledge during interaction. Accordingly, knowledge architecture becomes a front-line marketing capability that shapes brand consistency, service accuracy and consumer trust.

This contribution is deliberately bounded. The review does not claim that RAG replaces existing AI-in-marketing theory. Rather, it specifies retrieval grounding as a distinctive mechanism through which external knowledge may change customer interaction, managerial judgement and organisational accountability. For interactive marketing, this adds a knowledge-governance layer to existing accounts of AI-enabled personalisation and value co-creation. It specifies when the knowledge mobilised in an exchange is current, authorised and auditable. Second, the review extends market orientation and marketing analytics theory by treating retrieval infrastructure as part of market sensing. It shows that data create value through a governed knowledge-grounding layer in which sources are selected, ranked, authorised and traceable. Third, governance is built into the capability: transparency, fairness, privacy and human oversight are enabling conditions, not later safeguards.

5.2 Contribution to practice and policy

For practitioners, the study provides a realistic account of what can be implemented now and what remains speculative. Current evidence supports the use of RAG for bounded, knowledge-intensive tasks such as customer support, product question answering, internal knowledge search, campaign-content assistance and retrieval-grounded product information. These applications are suitable when knowledge bases are curated, access rights are clear and human review is available for high-risk outputs.

A practical implementation roadmap follows from this evidence. Organisations should first identify knowledge-intensive pain points where incorrect or outdated answers create measurable risk, such as product support, policy explanation, campaign compliance or internal knowledge search. They should then curate authorised sources, define retrieval boundaries, test outputs against brand and legal standards, and monitor failure cases. Only after these controls are established should firms expand towards more personalised and agentic use cases.

More ambitious applications, such as autonomous marketing strategy, fully agentic customer management, predictive market forecasting and unsupervised campaign optimisation, should be treated as emerging rather than established. The evidence base does not yet justify broad claims that RAG improves strategic performance at scale. Managers should therefore adopt staged implementation: begin with bounded use cases, define evaluation metrics, audit retrieval behaviour, validate outputs and expand only where performance and governance are demonstrably reliable.

Policy implications are also significant. Customer-facing RAG systems should disclose AI interaction where appropriate, avoid unsupported claims about AI performance and maintain traceability for generated content. Firms operating in or serving EU markets must consider transparency requirements under the EU AI Act and data-protection constraints under the GDPR, particularly where personal data and profiling are involved (European Parliament and Council, 2016, 2024). Firms in markets governed by consumer-protection regimes such as the FTC should ensure that claims about AI capabilities, product performance and automated advice are substantiated (Federal Trade Commission, 2024).

6. Limitations and future research

Several limitations qualify the review. The primary corpus comprises 25 studies, reflecting the early stage of explicitly marketing-focused RAG scholarship. The corpus is heterogeneous and includes journal articles, conference papers, system demonstrations and preprints, which limits strong causal or managerial claims. The review also excludes proprietary enterprise deployments and involves interpretive judgement despite structured coding and evidence classification.

Future research should develop marketing-specific evaluation frameworks that extend beyond generic accuracy to brand consistency, consumer trust, perceived fairness, usefulness, regulatory compliance, localisation quality and relationship effects. Field experiments and longitudinal studies are needed to test whether RAG-enabled interaction improves customer experience, satisfaction, loyalty or service efficiency in real environments.

A second priority is comparative theory-building. Scholars should test whether RAG changes customer responses differently from traditional chatbots, fine-tuned LLMs, recommender systems or human service agents. Such studies would help isolate retrieval grounding as a distinctive mechanism and examine boundary conditions by product category, perceived risk, consumer expertise and cultural context.

Future research should also investigate consumer-level mechanisms, including whether source visibility increases credibility, whether AI disclosure reduces or enhances trust and whether retrieval-grounded explanations improve consumer comprehension. Qualitative work could further explore how consumers interpret AI-assisted service interactions when recommendations are highly personalised.

Finally, the field requires stronger theoretical development. Future studies should examine RAG through service-dominant logic, relationship marketing, market orientation, customer engagement and socio-technical systems theory. Such work can move the field beyond application catalogues and towards explanations of how retrieval-grounded generation changes value co-creation, market sensing and organisational accountability.

To give direction to this research agenda, the review's synthesis can be translated into four testable propositions that connect retrieval grounding to marketing outcomes and specify boundary conditions for future empirical work:

P1. Retrieval grounding improves customer-facing marketing outcomes (trust, perceived helpfulness, and engagement) over non-grounded generation only when the underlying knowledge base is current, authorised, and auditable; where these conditions are absent, retrieval grounding does not improve and may degrade those outcomes.

P2. The effect of source visibility on consumer credibility and trust is moderated by perceived risk and product complexity, such that disclosing retrieved sources strengthens trust more in high-risk or high-complexity interactions than in routine ones.

P3. Retrieval-based market sensing enhances the quality and timeliness of marketing decisions only to the extent that source governance and evidence traceability are embedded in organisational routines; without these routines, retrieval-grounded synthesis increases the risk that managers treat fluent output as verified intelligence.

P4. Governance mechanisms such as fairness testing, freshness controls and human oversight function as enabling conditions rather than downstream safeguards, so that their presence strengthens, and their absence weakens, the relationship between RAG deployment and sustainable marketing performance.

These propositions provide an actionable agenda: P1 and P2 can be tested through controlled experiments and field studies of customer-facing interaction; P3 invites case-based and survey research on marketing decision-making; and P4 calls for longitudinal and organisational-level studies linking governance practices to performance. Together they convert the review's interpretive synthesis into directions that subsequent empirical work can confirm, qualify or refute. Table 4 below presents the future research agenda.

7. Conclusion

This integrative review examined RAG in marketing through a synthesis of 25 marketing-relevant studies and adjacent peer-reviewed literature. The analysis identified five interrelated themes covering adaptive architecture, customer interaction, context-aware retrieval, marketing intelligence and responsible deployment. The central conclusion is that RAG should not be understood merely as a technical extension of LLMs. In marketing, it is better understood as a socio-technical capability that connects customer-facing interaction with organisational knowledge, market sensing and governance.

The review shows that RAG has clear potential to support personalised customer interaction, multilingual service, product question answering and knowledge-grounded decision support. At the same time, the evidence base remains uneven. Many strategic claims are conceptual or prototype-based and require field validation. The paper therefore contributes by sharpening theoretical positioning, clarifying evidential limits and specifying boundary conditions for responsible deployment. As RAG technologies mature, marketing scholarship should focus less on listing applications and more on explaining when retrieval-grounded generation creates value, when it introduces risk and how organisations can govern it responsibly.

Table 4. Future research agenda for RAG in marketing

Priority area	Core issue	Indicative research questions	Likely contribution
Marketing-specific evaluation	Technical benchmarks do not fully capture marketing performance	How should RAG be evaluated for brand consistency, trust, persuasiveness and conversion-related outcomes?	Aligns optimisation with marketing-relevant outcomes
Live deployment and business impact	Most evidence remains at prototype or pilot stage	What happens to customer outcomes, workflow efficiency and organisational learning when RAG is deployed in live marketing settings?	Provides stronger evidence on real organisational value
Retrieval robustness in marketing environments	Queries are often vague, culturally variable and multi-intent	How can retrieval be improved for multilingual, segment-specific and time-sensitive marketing tasks?	Improves contextual precision and output relevance
Governance, fairness and compliance	Technical capability is advancing faster than governance practice	Which governance structures most effectively reduce hallucination, bias and non-compliant outputs in marketing use cases?	Strengthens trustworthy and compliant deployment
Transparency and consumer protection	Consumer-facing implications remain underexplored	How do consumers respond to AI disclosure in marketing content, and what forms of explanation improve trust without reducing usefulness?	Informs disclosure, transparency and consumer autonomy
Theory development	RAG remains weakly integrated into marketing theory	How does RAG reshape value co-creation, relationship continuity, market sensing and interactive marketing processes?	Builds cumulative theorisation of RAG in marketing

Source(s): Authors' own work

References

Béchard, P. and Marquez Ayala, O. (2024), "Reducing hallucination in structured outputs via retrieval-augmented generation", in Yang, Y., Davani, A.M., Sil, A. and Kumar, A. (Eds), *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track)*, Association for Computational Linguistics, pp. 228-238, doi: 10.18653/v1/2024.naacl-industry.19.

Benita, J., Tej, K.V.C., Kumar, E.V., Subbarao, G.V. and Venkatesh, C.H. (2024), "Implementation of retrieval-augmented generation (RAG) in chatbot systems for enhanced real-time customer support in e-commerce", *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, IEEE, pp. 1381-1388, doi: 10.1109/ICACRS62842.2024.10841586.

Brodie, R.J., Hollebeek, L.D., Juric, B. and Ilic, A. (2011), "Customer engagement: conceptual domain, fundamental propositions, and implications for research", *Journal of Service Research*, Vol. 14 No. 3, pp. 252-271, doi: 10.1177/1094670511411703.

Brown, A., Roman, M. and Devereux, B. (2025), "A systematic literature review of retrieval-augmented generation: techniques, metrics, and challenges", *Big Data and Cognitive Computing*, Vol. 9 No. 12, 320, doi: 10.3390/bdcc9120320.

Cooke, A., Smith, D. and Booth, A. (2012), "Beyond PICO: the SPIDER tool for qualitative evidence synthesis", *Qualitative Health Research*, Vol. 22 No. 10, pp. 1435-1443, doi: 10.1177/1049732312452938.

Cronin, M.A. and George, E. (2023), "The why and how of the integrative review", *Organizational Research Methods*, Vol. 26 No. 1, pp. 168-192, doi: 10.1177/1094428120935507.

Davenport, T., Guha, A., Grewal, D. and Bressgott, T. (2020), "How artificial intelligence will change the future of marketing", *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 24-42, doi: 10.1007/s11747-019-00696-0.

De Bruyn, A., Viswanathan, V., Beh, Y.S., Brock, J.K.U. and von Wangenheim, F. (2020), "Artificial intelligence and marketing: pitfalls and opportunities", *Journal of Interactive Marketing*, Vol. 51 No. 1, pp. 91-105, doi: 10.1016/j.intmar.2020.04.007.

De Freitas, B.A.T. and Lotufo, R.D.A. (2024), "Retail-GPT: leveraging retrieval-augmented generation (RAG) for building e-commerce chat assistants", *paper presented at the XXXII Congresso de Iniciação Científica da UNICAMP*, Campinas, Brazil.

Dencheva, V. (2023), "AI in marketing revenue worldwide 2020-2028", Statista, available at: https://www.statista.com/statistics/1293758/ai-marketing-revenue-worldwide/ (accessed 1 April 2026).

Elmitwalli, S., Mehegan, J., Braznell, S. and Gallagher, A. (2025), "Scalable evaluation framework for retrieval augmented generation in tobacco research using large language models", *Scientific Reports*, Vol. 15 No. 1, 22760, doi: 10.1038/s41598-025-05726-2.

European Parliament and Council (2016), "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)", *Official Journal of the European Union*, Vol. L119, pp. 1-88.

European Parliament and Council (2024), "Regulation (EU) 2024/1689 of the European parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (artificial intelligence Act)", *Official Journal of the European Union*, Vols L2024/1689, pp. 1-241.

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.-S. and Li, Q. (2024), "A survey on RAG meeting LLMs: towards retrieval-augmented large language models", *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, pp. 6491-6501, doi: 10.1145/3637528.3671470.

Federal Trade Commission (2024), "FTC announces crackdown on deceptive AI claims and schemes", available at: https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes (accessed 1 June 2026).

Gao, Y. and Liu, H. (2023), "Artificial intelligence-enabled personalization in interactive marketing: a customer journey perspective", *The Journal of Research in Indian Medicine*, Vol. 17 No. 5, pp. 663-680, doi: 10.1108/JRIM-01-2022-0023.

Hollebeek, L.D. (2011), "Demystifying customer brand engagement: exploring the loyalty nexus", *Journal of Marketing Management*, Vol. 27 Nos 7/8, pp. 785-807, doi: 10.1080/0267257X.2010.500132.

Huang, M.-H. and Rust, R.T. (2021), "A strategic framework for artificial intelligence in marketing", *Journal of the Academy of Marketing Science*, Vol. 49 No. 1, pp. 30-50, doi: 10.1007/s11747-020-00749-9.

Kim, T.E. and Diaz, F. (2025), "Towards fair RAG: on the impact of fair ranking in retrieval-augmented generation", *Proceedings of the International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval*, Association for Computing Machinery, pp. 112-121, doi: 10.1145/3731120.3744599.

Koch, J. and Pellatt, C. (2025), "AI adoption is surging in advertising, but is the industry prepared for responsible AI?", *International African Bibliography*, available at: https://www.iab.com/insights/ai-adoption-is-surging-in-advertising-but-is-the-industry-prepared-for-responsible-ai/ (accessed 1 April 2026).

Kohli, A.K. and Jaworski, B.J. (1990), "Market orientation: the construct, research propositions, and managerial implications", *Journal of Marketing*, Vol. 54 No. 2, pp. 1-18, doi: 10.1177/002224299005400201.

Kumar, S., Datta, S., Singh, V., Datta, D., Singh, S.K. and Sharma, R. (2024), "Applications, challenges, and future directions of human-in-the-loop learning", *IEEE Access*, Vol. 12, pp. 75735-75760, doi: 10.1109/ACCESS.2024.3401547.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W.-T., Rocktaschel, T., Riedel, S. and Kiela, D. (2020), "Retrieval-augmented generation for knowledge-intensive NLP tasks", *Advances in Neural Information Processing Systems*, Vol. 33, pp. 9459-9474.

Lyu, Y., Li, Z., Niu, S., Xiong, F., Tang, B., Wang, W., Wu, H., Liu, H., Xu, T. and Chen, E. (2025), "CRUD-RAG: a comprehensive Chinese benchmark for retrieval-augmented generation of large language models", *ACM Transactions on Information Systems*, Vol. 43 No. 2, pp. 41-32, doi: 10.1145/3701228.

Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C.D. and Ho, D.E. (2025), "Hallucination-free? Assessing the reliability of leading AI legal research tools", *Journal of Empirical Legal Studies*, Vol. 22 No. 2, pp. 216-252, doi: 10.1111/jels.12413.

Naeem, M., Ozuem, W., Howell, K. and Ranfagni, S. (2023), "A step-by-step process of thematic analysis to develop a conceptual model in qualitative research", *International Journal of Qualitative Methods*, Vol. 22, pp. 1-18, doi: 10.1177/16094069231205789.

Narver, J.C. and Slater, S.F. (1990), "The effect of a market orientation on business profitability", *Journal of Marketing*, Vol. 54 No. 4, pp. 20-35, doi: 10.1177/002224299005400403.

Niu, C., Wu, Y., Zhu, J., Xu, S., Shum, K., Zhong, R., Song, J. and Zhang, T. (2024), "RAGTruth: a hallucination corpus for developing trustworthy retrieval-augmented language models", in Ku, L.-W., Martins, A.F.T. and Srikumar, V. (Eds), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, pp. 10862-10878, doi: 10.18653/v1/2024.acl-long.585.

Palmatier, R.W., Dant, R.P., Grewal, D. and Evans, K.R. (2006), "Factors influencing the effectiveness of relationship marketing: a meta-analysis", *Journal of Marketing*, Vol. 70 No. 4, pp. 136-153, doi: 10.1509/jmkg.70.4.136.

Pantano, E. and Pizzi, G. (2020), "Forecasting artificial intelligence on online customer assistance: evidence from chatbot patents analysis", *Journal of Retailing and Consumer Services*, Vol. 55, 102096, doi: 10.1016/j.jretconser.2020.102096.

Peltier, J.W., Dahl, A.J. and Schibrowsky, J.A. (2024), "Artificial intelligence in interactive marketing: a conceptual framework and research agenda", *The Journal of Research in Indian Medicine*, Vol. 18 No. 1, pp. 54-90, doi: 10.1108/JRIM-01-2023-0030.

Peng, C.-Y. and Haq, M.D.U. (2026), "AI-augmented co-creation via social media posts and consumer identity formation: the interplay of interaction orientation, empowerment, creativity, and willingness", *Journal of Retailing and Consumer Services*, Vol. 90, 104718, doi: 10.1016/j.jretconser.2025.104718.

Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., Zhang, Y. and Tang, S. (2025), "Graph retrieval-augmented generation: a survey", *ACM Transactions on Information Systems*, Vol. 44 No. 2, pp. 1-52, doi: 10.1145/3777378.

Ravat, L., Hemonnet-Goujot, A. and Hollet-Haudebert, S. (2024), "Exploring how to develop data-driven innovation capability of marketing within B2B firms: toward a capability model and process-oriented approach", *Industrial Marketing Management*, Vol. 118, pp. 110-125, doi: 10.1016/j.indmarman.2023.12.015.

Sun, Z., Zang, X., Zheng, K., Song, Y., Xu, J., Zhang, X., Yu, W., Song, Y. and Li, H. (2025), "ReDeEP: detecting hallucination in retrieval-augmented generation via mechanistic interpretability", *paper presented at the International Conference on Learning Representations (ICLR 2025)*, available at: https://openreview.net/forum?id=ztzZDzgfrh (accessed 1 June 2026).

Tangarajan, P., Rajasekar, A.A., Rathi, M., Dandin, V.R. and Ersoy, O. (2025), "Contextually aware e-commerce product question answering using RAG", arXiv preprint, available at: https://arxiv.org/abs/2508.01990 (accessed 1 June 2026).

Torraco, R.J. (2016), "Writing integrative literature reviews: using the past and present to explore the future", *Human Resource Development Review*, Vol. 15 No. 4, pp. 404-428, doi: 10.1177/1534484316671606.

Vargo, S.L. and Lusch, R.F. (2004), "Evolving to a new dominant logic for marketing", *Journal of Marketing*, Vol. 68 No. 1, pp. 1-17, doi: 10.1509/jmkg.68.1.1.24036.

Vargo, S.L. and Lusch, R.F. (2016), "Institutions and axioms: an extension and update of service-dominant logic", *Journal of the Academy of Marketing Science*, Vol. 44 No. 1, pp. 5-23, doi: 10.1007/s11747-015-0456-3.

Wang, C.L. (2024), “Editorial - what is an interactive marketing perspective and what are emerging research areas?”, *The Journal of Research in Indian Medicine*, Vol. 18 No. 2, pp. 161-165, doi: 10.1108/JRIM-03-2024-371.

Wang, C.L. (2025), “Editorial: demonstrating contributions through storytelling”, *The Journal of Research in Indian Medicine*, Vol. 19 No. 1, pp. 1-4, doi: 10.1108/JRIM-01-2025-455.

Wang, C.L. (2026), “Editorial: the changing landscape of marketing research in the AI era: prospects and challenges”, *The Journal of Research in Indian Medicine*, Vol. 20 No. 1, pp. 1-10, doi: 10.1108/JRIM-02-2026-766.

Wedel, M. and Kannan, P.K. (2016), “Marketing analytics for data-rich environments”, *Journal of Marketing*, Vol. 80 No. 6, pp. 97-121, doi: 10.1509/jm.15.0413.

Whittemore, R. and Knafl, K. (2005), “The integrative review: updated methodology”, *Journal of Advanced Nursing*, Vol. 52 No. 5, pp. 546-553, doi: 10.1111/j.1365-2648.2005.03621.x.

Xu, Z., Cruz, M.J., Guevara, M., Wang, T., Deshpande, M., Wang, X. and Li, Z. (2024a), “Retrieval-augmented generation with knowledge graphs for customer service question answering”, *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, pp. 2905-2909, doi: 10.1145/3626772.3661370.

Xu, Z., Shen, Z., Zhou, Q. and Ristoski, P. (2024b), “Knowledge graph-enhanced retrieval augmented generation for e-commerce”, *Proceedings of the International Workshop on Retrieval-Augmented Generation Enabled by Knowledge Graphs (RAGE-KG 2024), CEUR Workshop Proceedings*, Vol. 3950, pp. 31-39.

Yao, C. and Fujita, S. (2024), “Adaptive control of retrieval-augmented generation for large language models through reflective tags”, *Electronics*, Vol. 13 No. 23, 4643, doi: 10.3390/electronics13234643.

Yao, X. and Ren, W. (2025), “Enhancing customer experiences: a holistic brand strategy framework in the metaverse”, *Journal of Applied Business and Behavioral Sciences*, Vol. 1 No. 1, pp. 115-135, doi: 10.63522/jabbs.101006.

Yin, C.-C. and Do, K.-M. (2025), “Metaverse and consumer behavior: a systematic literature review and future research agenda”, *International Journal of Consumer Studies*, Vol. 49 No. 4, e70097, doi: 10.1111/ijcs.70097.

Yu, B.T.W. and Liu, S.T.X. (2025), “Deep learning application for marketing engagement - its thematic evolution”, *The Journal of Research in Indian Medicine*, Vol. 19 No. 5, pp. 861-879, doi: 10.1108/JRIM-08-2024-0371.

Yu, H., Gan, A., Zhang, K., Tong, S., Liu, Q. and Liu, Z. (2025), “Evaluation of retrieval-augmented generation: a survey”, *Big Data: 12th CCF Conference, BigData 2024, Qingdao, China, August 9-11, 2024, Proceedings*, Springer, pp. 102-120, doi: 10.1007/978-981-96-1024-2_8.

Zhang, W. and Zhang, J. (2025), “Hallucination mitigation for retrieval-augmented large language models: a review”, *Mathematics*, Vol. 13 No. 5, 856, doi: 10.3390/math13050856.

Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J. and Cui, B. (2026), “Retrieval-augmented generation for AI-generated content: a survey”, *Data Science and Engineering*, Vol. 11, pp. 1-29, doi: 10.1007/s41019-025-00335-5.

Zhu, X., Xie, Y., Liu, Y., Li, Y. and Hu, W. (2025), “Knowledge graph-guided retrieval augmented generation”, *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Association for Computational Linguistics, pp. 8912-8924, doi: 10.18653/v1/2025.naacl-long.449.

Corresponding author
Mayemba Karel Nzita can be contacted at: mayemban@uj.ac.za