INTERNATIONAL REVIEW OF ARTIFICIAL INTELLIGENCE IN MARKETING

NAWALA

Volume 3, Number 2, 2024

Holistic Customer Understanding Through Multimodal Artificial Intelligence in Marketing Analytics

Muhamad Fadyl Frizkia[1*]

[1] Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia

Abstract

Article history:
Received: August 24, 2024
Revised: September 8, 2024
Accepted: October 13, 2024
Published: December 30, 2024

Keywords:
Customer Analytics, Holistic Customer Experience, Marketing Analytics, Multimodal AI, Sentiment Analysis.

Identifier:
Nawala
Page: 90-108
https://nawala.io/index.php/iraim

This article examines how multimodal artificial intelligence (AI) contributes to holistic customer understanding in marketing analytics. It asks to what extent combining text, images, and other data modalities moves firms beyond narrow, touchpoint-specific predictions toward richer views of customer experience and behavior. Using a systematic literature review of peer-reviewed studies published between 2019 and 2023, the study identifies convergences and gaps in current applications of multimodal AI to sentiment analysis, recommendation, and behavioral prediction. Descriptive mapping summarizes dominant contexts, modality combinations, and model families, while thematic synthesis explores how studies conceptualize and operationalize "customer understanding". The review finds that multimodal models consistently outperform unimodal baselines and capture nuanced affective and contextual cues, but remain concentrated on single tasks and public data sources, with limited integration into customer journey or CRM frameworks. Methodological limitations and reporting heterogeneity are highlighted to qualify the strength of existing evidence. Future research directions are outlined to embed multimodal analytics within longitudinal, cross-channel, and ethically governed approaches to customer insight.

*Corresponding author:
(Muhamad Fadyl Frizkia)

1. Introduction

Pervasive digitization has transformed how firms observe and interact with customers. Touchpoints now span websites, mobile apps, social media, in-store sensors, call centers, and third-party platforms, generating vast streams of behavioral, transactional, textual, visual, and sometimes audio data. Artificial intelligence (AI) and machine learning have become central to turning these data into actionable marketing insight, supporting tasks such as personalization, dynamic pricing, churn prediction, and customer journey optimization (Verma et al., 2021; Hossain et al., 2022). Yet most AI-enabled marketing analytics still rely on single data types, typically structured transactions or text, leading to a fragmented view of the customer rather than an integrated understanding of needs, context, and emotions.

Big data and customer relationship management (CRM) research has long emphasized the strategic value of holistic customer knowledge. Structured literature reviews on big data-enabled CRM show how advanced analytics capabilities can enhance segmentation, relationship quality, and competitive advantage when firms consolidate information across channels and life-cycle stages (Del Vecchio et al., 2022). However, these studies predominantly focus on integrating structured and semi-structured data (e.g., purchase histories, demographics, clickstreams), with relatively limited attention to the combined exploitation of heterogeneous unstructured sources such as images, videos, and multi-platform conversational traces. Consequently, many organizations still struggle to operationalize the promise of a "360-degree" customer view in day-to-day marketing decision-making.

In parallel, advances in multimodal AI demonstrate that fusing text, images, audio, and other signals can significantly improve the modelling of human attitudes and behaviors. Affective computing research on multimodal big data shows that combining textual, audio, visual, and physiological signals yields richer affective and sentiment representations than any single modality alone, especially in high-volume customer feedback environments (Shoumy et al., 2020). In the marketing domain, multimodal sentiment analysis of online product information leverages text mining and big-data techniques on reviews and associated media to better capture consumers' evaluative judgments and purchase intentions (Fang et al., 2022). Similarly, deep multimodal architectures such as CRNet integrate textual and image contexts to predict customer revisit behavior more accurately than unimodal baselines (Park, 2023). Taken together, these studies suggest that multimodal AI can move marketing analytics from channel- or task-specific prediction toward more holistic understanding of customers' experiences across touchpoints.

Despite these developments, the emerging literature on multimodal AI in marketing analytics remains fragmented across technical and managerial outlets, uses diverse terminologies, and often focuses on narrow use-cases (e.g., product reviews, tourism platforms) without explicitly theorizing "holistic customer understanding". At the same time, work on AI-enabled customer analytics capability in retailing has started to conceptualize how integrated analytics resources contribute to real-time insight generation and value creation, but typically without a multimodal lens (Hossain et al., 2022). To date, no systematic literature review has synthesized how

multimodal AI has been employed in marketing analytics to construct comprehensive, cross-modal representations of customers.

This article addresses that gap by conducting a systematic literature review of peer-reviewed studies published between 2019 and 2023 that apply multimodal AI techniques to marketing analytics problems. Drawing on established SLR procedures, the review maps dominant data modalities, modelling approaches, and marketing objectives, critically evaluates how far current work advances holistic customer understanding, and identifies theoretical, methodological, and ethical avenues for future research.

2. Literature Review

The existing literature on AI-enabled marketing analytics has largely evolved along two parallel tracks: work on developing customer analytics capability and work on multimodal machine learning. Customer analytics studies emphasize how firms turn heterogeneous customer data into insight and performance, but typically focus on structured or text data rather than genuinely multimodal representations. For example, Hossain et al. (2020) conceptualize customer analytics capability in retailing as a multidimensional resource spanning value creation, delivery, and management, yet their empirical emphasis is on transactional and tabular data rather than integrated visual, textual, and behavioral signals. This line of research highlights the strategic importance of analytics for customer equity, but still treats "the customer" through relatively narrow data lenses. Broader reviews of AI in marketing similarly foreground personalization, targeting, and predictive modelling while largely

remaining modality-agnostic and centering on structured and textual data (Verma et al., 2021).

In parallel, multimodal AI research has expanded rapidly, but is mostly grounded in computer science or generic social media analytics rather than marketing strategy. Comprehensive surveys such as Deldjoo et al. (2020) and Gandhi et al. (2023) synthesize techniques for combining images, audio, video, and text in recommender systems and sentiment analysis, describing feature extraction, fusion architectures, and evaluation protocols. These reviews demonstrate that multimodal models systematically outperform unimodal baselines by capturing complementary cues across channels, but they rarely connect these gains to constructs like customer experience, journey stages, or relationship value that are central to marketing analytics.

Domain-specific implementations further illustrate both the potential and the fragmentation of current work. In e-commerce, Sales et al. (2021) propose multimodal deep neural networks that fuse product text and images to infer missing catalog attributes, improving search relevance and product organization. In social media analytics, Fan et al. (2023) develop a deep multimodal fusion model that combines BiLSTM-based textual encodings with CNN-based visual features to classify sentiment during public emergencies, showing clear performance gains over text-only models. Mehmet and D'Alessandro (2022) introduce a multimodal social listening analysis framework that qualitatively integrates text and accompanying media to uncover meaning, irony, and emotion in social posts, while Fang et al. (2022) demonstrate how multimodal sentiment analysis of online product marketing

information can better capture evaluative judgments relevant to purchasing decisions.

The emerging gap, therefore, lies not in the absence of multimodal techniques, but in the lack of synthesized knowledge on how multimodal AI can be systematically leveraged within marketing analytics to represent customers as whole persons across channels and contexts. This systematic literature review responds to that gap by mapping and critically evaluating multimodal AI applications through the lens of holistic customer understanding.

Taken together, these strands indicate that multimodal AI can enrich marketing-relevant tasks, such as understanding sentiment, inferring product attributes, and interpreting consumer expressions, yet they remain loosely coupled to the broader agenda of holistic customer understanding. Most studies address a single touchpoint (e.g., reviews, social posts) or a single task (e.g., classification, attribute prediction), with limited attention to how multimodal insights travel across the customer journey, integrate with existing customer analytics capabilities, or support managerial decision-making at segment, relationship, or firm level.

To synthesize these strands of evidence, a summary of the most relevant prior studies is presented in Table 2.1 to highlight their key contributions and limitations in relation to holistic customer understanding. These studies provide insight into how customer analytics capabilities have been conceptualized, how multimodal AI techniques have been developed, and how these approaches have been applied across various domains.

Table 2.1 Prior Research

No	Author(s) & Year	Article Title	Research Findings
1	Hossain et al. (2020)	Revisiting Customer Analytics Capability for Data-Driven Retailing	This study uses a review/theory-building approach in retail to conceptualize customer analytics capability (CAC). Proposes a multidimensional CAC model with 6 capability dimensions and 12 sub-dimensions, grouped into value creation (offering, personalization), value delivery (distribution, communication), and value management (data management, data protection), arguing these capabilities support engagement and customer equity in data-driven retailing.
2	Deldjoo et al. (2020)	Recommender Systems Leveraging Multimedia Content	This study surveys multimedia recommender systems, organizing work by media type, feature representations, and recommendation approaches.

			Provides a taxonomy and a generic framework showing how multimodal content (audio/visual/text) can enhance recommendation quality across domains such as media, fashion, tourism, food, and e-commerce, while outlining key research challenges.
3	Verma et al. (2021)	Artificial Intelligence in Marketing: Systematic Review and Future Research Direction	This research conducts a systematic review of AI in marketing (with mapping/synthesis of themes) to structure what AI is used for and where the field is heading. Consolidates major research streams and offers a future research agenda by highlighting dominant topics and gaps for subsequent studies.
4	Sales et al. (2021)	Multimodal Deep Neural Networks for Attribute Prediction	Develops multimodal deep learning for e-commerce catalog enhancement by combining

		and Applications to E-Commerce Catalogs Enhancement	product images and unstructured text to predict product attributes. Shows that multimodal representations enable more scalable and accurate attribute prediction than single-modality inputs, supporting practical tasks like catalog structuring and quality improvement.
5	Mehmet & D'Alessandro (2022)	More Than Words Can Say: A Multimodal Approach to Understanding Meaning and Sentiment in Social Media	Introduces a Multimodal Social Listening Analysis (MSLA) framework and validates it on Facebook/Twitter/Instagram posts. Finds MSLA improves interpretation of multimodal meaning by revealing latent structure, capturing multiple sentiment cues within a post, detecting implicit meanings (e.g., irony/sarcasm/humor), and identifying emotions/judgments to inform marketing strategy.

6	Fang et al. (2022)	The Multimodal Sentiment Analysis of Online Product Marketing Information Using Text Mining and Big Data	Builds an attention-based multimodal sentiment approach for online product marketing information (text + images). Reports performance gains over strong baselines (e.g., approx. +4% accuracy improvements vs selected models), supporting the conclusion that multimodal fusion improves sentiment inference in product marketing contexts.
7	Gandhi et al. (2023)	Multimodal Sentiment Analysis: A Systematic Review of History, Datasets, Multimodal Fusion Methods, Applications, Challenges and Future Directions	Systematically reviews multimodal sentiment analysis (MSA), focusing on datasets, fusion methods, applications, and challenges. Synthesizes the evolution of MSA and categorizes fusion approaches, arguing that modern ML/DL improves end-to-end sentiment inference (feature learning → fusion → prediction) while identifying

			persistent challenges and future directions.
8	Fan et al. (2023)	Multimodal Sentiment Analysis for Social Media Contents During Public Emergencies	Proposes a deep multimodal fusion model for sentiment analysis of text-image social media posts during public emergencies, evaluated on platforms/datasets including Weibo and Twitter. Reports the model outperforms baselines and demonstrates that adding visual information improves sentiment classification compared with text-only approaches, supporting potential use for emergency-related decision support.

3. Methods

This study used a systematic literature review design to identify, evaluate, and synthesize peer-reviewed research on multimodal artificial intelligence in marketing analytics. Searches were conducted in major academic databases, including Scopus, Web of Science, ScienceDirect, and IEEE Xplore, supplemented by targeted queries in Google Scholar. A combination of keywords and Boolean operators was used,

such as “multimodal” AND “artificial intelligence” AND “marketing analytics”, “multimodal customer analytics”, and “multimodal sentiment” AND “marketing”. The search was restricted to peer-reviewed journal articles and conference papers published in English between 2019 and 2023. After removing duplicates, titles and abstracts were screened, followed by full-text assessment using predefined inclusion criteria: (1) use of multimodal AI (fusion of at least two data modalities such as text, images, audio, video, or sensor data); (2) a clear focus on marketing or customer-related analytics (e.g., sentiment, recommendation, customer behavior modelling); and (3) empirical implementation or evaluation. Studies that were conceptual only, single-modality, non-peer-reviewed, or unrelated to marketing analytics were excluded.

Data from the included studies were extracted using a structured coding template. For each article, information was recorded on publication details, industry or application context, data modalities, AI and machine learning techniques, type of marketing analytics task, and how customer understanding was operationalized (e.g., sentiment, experience, journey, value), along with performance metrics and any reported managerial, theoretical, ethical, or privacy implications. To enhance reliability, screening and coding were conducted by two reviewers, with disagreements resolved through discussion. The analysis combined descriptive mapping (e.g., by year, sector, modality combinations, and model families) with thematic synthesis to identify recurring patterns and gaps in how multimodal AI is used to construct holistic representations of customers across channels and touchpoints.

4. Results and Discussion

The studies identified in this review show that multimodal AI in marketing analytics is concentrated in a few dominant application areas rather than evenly distributed across the customer journey. Most included works focus on sentiment or opinion mining from social media posts and product reviews, often combining textual content with associated images to infer attitudes, preferences, or engagement propensity (Fang et al., 2022; Fan et al., 2023). A second cluster addresses recommendation and product discovery, where multimodal representations of items (visual appearance plus textual descriptions or reviews) are used to improve relevance and personalization in e-commerce settings (Deldjoo et al., 2020; Sales et al., 2021). Only a smaller set of studies directly model downstream behavioral outcomes such as revisits or popularity, although those that do highlight the value of multimodal fusion for anticipating future interaction and exposure (Park, 2023; Wang et al., 2023). Overall, the empirical base is rich in task-specific implementations but uneven in coverage of the full lifecycle from awareness and exploration to purchase, usage, and advocacy.

Across these applications, a consistent pattern emerges in how modalities and models are combined. The vast majority of studies integrate two modalities, typically text and images, using deep learning architectures that extract separate latent representations and then fuse them through early, late, or hybrid strategies (Deldjoo et al., 2020; Gandhi et al., 2023). For example, Sales et al. (2021) and Fan et al. (2023) use convolutional and recurrent networks to encode visual and textual information, then merge these representations for attribute prediction or sentiment classification,

while Wang et al. (2023) adopt a hierarchical fusion model to combine image, text, and attribute features when predicting social media popularity. Survey work in computer science reinforces that these fusion choices have important implications for performance, robustness, and interpretability, yet marketing-oriented studies rarely make these design trade-offs explicit or relate them to managerial objectives such as explainability, controllability, or customer fairness (Chandrasekaran et al., 2021; Gandhi et al., 2023). When viewed through the lens of "customer understanding," the corpus shows both progress and blind spots. On the one hand, multimodal models clearly enrich classic marketing analytics tasks by capturing emotion, context, and nuance that are difficult to infer from text or structured data alone. Multimodal sentiment and emotion analytics can surface subtle cues of satisfaction, frustration, or irony in social posts and reviews (Shoumy et al., 2020; Mehmet & D'Alessandro, 2022), while multimodal recommenders implicitly learn aesthetic, semantic, and social signals embedded in images and language (Deldjoo et al., 2020). On the other hand, most studies operationalize "understanding" in terms of predictive accuracy on narrow labels, like sentiment polarity, rating, click, popularity, rather than richer constructs such as experience quality, journey stage, or relationship value. Only a few works, such as CRNet's prediction of customer revisit behavior (Park, 2023), move closer to behavioral indicators that are meaningful for long-term relationship management.

The results also show that integration with broader customer analytics and CRM capabilities is more often implied than realized. Research on AI-enabled customer analytics capability and big-data CRM emphasizes the need to connect

analytics outputs to value creation, delivery, and management processes across channels (Hossain et al., 2020, 2022; Del Vecchio et al., 2022). Yet multimodal marketing studies typically operate at the level of isolated datasets and touchpoints, such as Instagram posts, platform reviews, or single-site logs, without tracing how multimodal insights feed into segmentation, personalization, or journey orchestration over time. Social listening-oriented work offers a partial exception: Mehmet and D'Alessandro (2022) demonstrate how qualitative, multimodal interpretations of meaning and emotion in social posts can inform strategic brand and communication decisions, but such approaches remain rare relative to purely quantitative pipelines.

Taken together, these findings suggest that multimodal AI currently advances holistic customer understanding in depth more than in breadth. At the micro level of individual touchpoints, models are increasingly capable of capturing complex affective, visual, and contextual cues that enrich what marketers know about how customers feel and respond in specific environments (Chandrasekaran et al., 2021; Fang et al., 2022). At the macro level of journeys and relationships, however, there is limited evidence that multimodal representations are systematically linked across channels, time, and organizational functions. Future research can build on these results by designing multimodal analytics that are explicitly aligned with CRM architectures, customer equity metrics, and ethical governance frameworks, and by empirically examining how multimodal insights change managerial decisions about targeting, experience design, and resource allocation. In doing so, multimodal AI

could move from enhancing isolated predictions to genuinely supporting a holistic, person-centered view of customers in marketing analytics.

5. Conclusion

This review shows that multimodal AI is beginning to deepen what marketers can infer from individual touchpoints, but has not yet delivered on the promise of truly holistic customer understanding. Most studies concentrate on sentiment analysis, social media content, and e-commerce recommendation, leveraging text-image fusion to predict narrow outcomes such as polarity, ratings, clicks, or popularity. While these approaches clearly improve predictive performance and capture richer emotional and contextual cues than unimodal models, they are only loosely connected to broader constructs such as customer experience, journey stages, or long-term relationship value. Integration with CRM architectures and customer analytics capabilities is more often assumed than demonstrated in empirical work.

At the same time, the corpus reveals important blind spots and limitations that temper the strength of the conclusions. Empirical studies are unevenly distributed across industries, regions, and data sources, with a heavy bias toward public social and review data and relatively little work inside firms' proprietary omnichannel environments. Most models operate on two modalities and rarely address interpretability, fairness, or governance, which constrains their practical use in high-stakes marketing decisions. Methodologically, heterogeneity in tasks, metrics, and reporting standards complicates direct comparison across studies, while the

focus on published, English-language, 2019 to 2023 peer-reviewed research increases the risk of publication and language bias. These limits may affect how comprehensively the current synthesis captures both failures and negative findings.

Future research should therefore move beyond isolated touchpoints and tasks toward multimodal architectures that are explicitly embedded in customer journey and CRM frameworks. This includes longitudinal, cross-channel studies that link multimodal signals to enduring outcomes such as lifetime value, loyalty, and advocacy; the design of explainable fusion strategies that can be understood and challenged by managers; and rigorous evaluation of how multimodal insights change real marketing decisions, not only model accuracy. Researchers should also examine ethical and regulatory dimensions such as privacy, consent, bias, and contestability as multimodal customer profiling becomes more granular and pervasive. By addressing these gaps, subsequent work can transform multimodal AI from a collection of powerful but fragmented techniques into a coherent foundation for holistic, person-centered customer understanding in marketing analytics.

References

Chandrasekaran, G., Nguyen, T. N., & Hemanth D, J. (2021). Multimodal sentimental analysis for social media applications: A comprehensive review. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery, 11*(5), e1415.

Deldjoo, Y., Schedl, M., Cremonesi, P., & Pasi, G. (2020). Recommender systems leveraging multimedia content. *ACM Computing Surveys (CSUR), 53*(5), 1-38.

Del Vecchio, P., Mele, G., Siachou, E., & Schito, G. (2022). A structured literature review on Big Data for customer relationship management (CRM): Toward a future agenda in international marketing. *International Marketing Review, 39*(5), 1069-1092.

Fan, T., Wang, H., Wu, P., Ling, C., & Ahvanooey, M. T. (2023). Multimodal sentiment analysis for social media contents during public emergencies. *Journal of Data and Information Science, 8*(3), 61-87.

Fang, Z., Qian, Y., Su, C., Miao, Y., & Li, Y. (2022). The multimodal sentiment analysis of online product marketing information using text mining and big data. *Journal of Organizational and End User Computing (JOEUC), 34*(1), 1-19.

Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., & Hussain, A. (2023). Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion, 91*, 424-444.

Hossain, M. A., Akter, S., & Yanamandram, V. (2020). Revisiting customer analytics capability for data-driven retailing. *Journal of Retailing and Consumer Services, 56*, 102187.

Hossain, M. A., Akter, S., Yanamandram, V., & Gunasekaran, A. (2022). Operationalizing artificial intelligence-enabled customer analytics capability in retailing. *Journal of Global Information Management (JGIM), 30*(8), 1-23.

Mehmet, M. I., & D'Alessandro, S. (2022). More than words can say: A multimodal approach to understanding meaning and sentiment in social media. *Journal of Marketing Management, 38*(13-14), 1461-1493.

Park, E. (2023). CRNet: A multimodal deep convolutional neural network for customer revisit prediction. *Journal of Big Data, 10*(1), 1.

Sales, L. F., Pereira, A., Vieira, T., & de Barros Costa, E. (2021). Multimodal deep neural networks for attribute prediction and applications to e-commerce catalogs enhancement. *Multimedia Tools and Applications, 80*(17), 25851–25873.

Shoumy, N. J., Ang, L. M., Seng, K. P., Rahaman, D. M. M., & Zia, T. (2020). Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. *Journal of Network and Computer Applications, 149*, 102447.

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. International *Journal of Information Management Data Insights, 1*(1), 100002.

Wang, J., Yang, S., Zhao, H., & Yang, Y. (2023). Social media popularity prediction with multimodal hierarchical fusion model. *Computer Speech & Language, 80*, 101490.