

Explainable Artificial Intelligence (XAI) in Personalized Marketing: A Systematic Literature Review of Algorithms, Interpretability Techniques, and Consumer Trust Implications

Yassir ^{1*} Muhammad Bakri ²

¹ Universitas Wira Bhakti, Makassar, Indonesia. Email: yassirrachman@yahoo.com

² Universitas Wira Bhakti, Makassar, Indonesia. Email: m.bakri933@gmail.com

ARTICLE HISTORY

Submitted : April 27, 2026
Reviewed : May 01, 2026
Revised : May 04, 2026
Accepted : May 08, 2026
Published : May 10, 2026

Conflict of Interest Statement:

The author(s) declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ABSTRACT

Purpose: This study aims to systematically review the role of Explainable Artificial Intelligence (XAI) in personalized marketing by examining AI algorithms, interpretability techniques, and their implications for consumer trust.

Research Method: A systematic literature review was conducted by analyzing peer-reviewed journal articles and conference papers related to AI, XAI, and personalized marketing. The study synthesizes findings across technical and behavioral dimensions to provide an integrated understanding of the research domain.

Results and Discussion: The results indicate that machine learning, deep learning, and recommender systems are the primary algorithms used in personalized marketing. However, increasing model complexity reduces interpretability, creating a need for XAI techniques such as LIME, SHAP, and attention mechanisms. The findings further reveal that XAI enhances consumer trust by improving transparency, understandability, and fairness, although contextual factors, including privacy concerns and user characteristics, influence this relationship.

Implications: This study contributes theoretically by integrating technical and behavioral perspectives into a unified framework. In practice, it provides managers with guidance on designing transparent and trustworthy AI systems and highlights the need for ethical, user-centered AI implementation.

Originality: This study is original in integrating technical perspectives on XAI with behavioral perspectives on consumer trust in personalized marketing. It offers a unified framework explaining how explainability supports transparency, fairness, and trust in AI-driven marketing.

Keywords: artificial intelligence; personalized marketing; consumer trust; machine learning; interpretability.

1. Introduction

The rapid development of artificial intelligence (AI) has significantly reshaped modern marketing, especially in personalized marketing practices. Companies increasingly use machine learning and deep learning to analyze large volumes of consumer data and deliver tailored marketing strategies based on individual preferences (Huang & Rust, 2018). This approach has been proven to improve customer engagement, increase conversion rates, and enhance overall marketing performance (Wedel & Kannan, 2016). For instance, platforms such as Amazon and Netflix generate a substantial share of their revenue, estimated at more than 30 percent, from AI-based recommendation systems. In addition, more than 80



percent of consumers report a higher likelihood of purchasing from brands that offer personalized experiences. These trends indicate that AI-driven personalization is no longer optional but has become a strategic requirement for firms competing in digital markets. As a result, businesses continue to adopt more advanced algorithms to better predict consumer behavior and deliver relevant content in real time.

However, the increasing use of complex AI models also creates important challenges, particularly related to transparency and interpretability. Many advanced models, such as deep neural networks, function as black-box systems, meaning that their decision-making processes are difficult to understand (Adadi & Berrada, 2018). This lack of clarity raises serious concerns, especially in marketing, where consumer trust is critical. Studies show that around 60 to 70 percent of consumers are concerned about how their data is used in algorithmic decisions, and many feel uncomfortable when they cannot understand why certain recommendations are made. This situation can reduce trust, increase perceived risk, and lead to resistance toward AI-based systems (Shin, 2021). At the same time, stricter data privacy regulations further emphasize the need for transparent and accountable AI. Therefore, the gap between high model performance and low interpretability becomes a key issue, highlighting the need to explore how explainable AI can improve trust in personalized marketing.

Recent studies have begun to explore the role of Explainable Artificial Intelligence (XAI) in enhancing transparency and trust in AI-driven systems, although direct applications in personalized marketing remain limited. Existing literature identifies several widely used XAI techniques, including LIME, SHAP, saliency maps, and attention mechanisms, which are designed to provide interpretable insights into complex model decisions (Sewada et al., 2023). In parallel, research on user perception highlights five critical dimensions that influence trust in AI systems, namely trust, transparency, understandability, usability, and fairness (Haque et al., 2023). Furthermore, the effectiveness of explanations is determined by key quality factors such as format, completeness, accuracy, and currency, which shape how users interpret and evaluate algorithmic outputs (Haque et al., 2023). These findings suggest that both technical and perceptual dimensions play an essential role in building trust in AI systems.

Empirical evidence across domains indicates that XAI can significantly improve user trust and satisfaction, particularly in applications such as recommendation systems and dynamic pricing (Sarkar et al., 2025). However, a "trust-personalisation paradox" has been identified, where increased personalization enhances engagement but often lacks sufficient transparency, thereby undermining user trust (Chingono et al., 2026). Additional studies confirm the utility of SHAP and LIME in detecting algorithmic bias, especially in retail contexts involving hyper-personalization (Agarwal, 2026). Despite these advancements, most XAI research remains concentrated in healthcare, finance, and criminal justice, with limited focus on consumer-oriented marketing applications (Islam et al., 2022). Moreover, challenges such as scalability limitations, bias mitigation, and the lack of standardized evaluation metrics continue to constrain broader implementation (Saarela & Podgorelec, 2024; Sarkar et al., 2025).

Despite the growing body of literature on Explainable Artificial Intelligence (XAI), significant research gaps remain, particularly in the context of personalized marketing. Existing studies predominantly focus on the technical development of XAI methods and their application in high-stakes domains such as healthcare, finance, and criminal justice, where interpretability is critical for risk-sensitive decision-making (Islam et al., 2022). While these studies provide valuable insights into algorithmic transparency, their findings are not directly transferable to marketing contexts, which

involve different user expectations, behavioral dynamics, and decision environments. As a result, there is limited empirical evidence examining how XAI techniques function in consumer-facing applications, especially in personalized marketing systems where user experience and perception play a central role. Furthermore, prior research tends to evaluate XAI performance based on technical metrics, with less emphasis on how explanation quality influences consumer trust, engagement, and decision-making behavior.

In addition, there is a lack of integrative frameworks that connect the technical aspects of XAI with consumer behavior constructs. Although dimensions such as transparency, fairness, and understandability have been identified (Haque et al. (2023), existing studies often examine these variables in isolation rather than as part of a unified model. The identified trust-personalization paradox further highlights inconsistencies in the literature, in which increased personalization does not always lead to higher trust (Chingono et al., 2026). Moreover, methodological limitations persist, including reliance on anecdotal evidence and the absence of standardized evaluation metrics for XAI effectiveness (Saarela & Podgorelec, 2024). These gaps indicate the need for a systematic synthesis of literature that bridges technical innovation with consumer-centric outcomes in personalized marketing.

Building on the identified research gaps, this study makes a novel contribution by systematically integrating the technical and behavioral dimensions of Explainable Artificial Intelligence (XAI) in the context of personalized marketing. Unlike prior studies that tend to separate algorithmic development from user perception, this research develops a comprehensive synthesis that explicitly links artificial intelligence algorithms, interpretability techniques, and their implications for consumer trust within a unified framework. Specifically, this study classifies the types of AI algorithms employed in personalized marketing, identifies the interpretability techniques used to explain AI models, and examines how these techniques influence key dimensions of consumer trust, including transparency, understandability, and perceived fairness. In addition, this study addresses existing limitations by identifying implementation challenges and highlighting unresolved research gaps within the current literature. Therefore, this study aims to answer the following research questions: (1) What types of artificial intelligence algorithms are used in personalized marketing? (2) What interpretability techniques are applied within the XAI framework? (3) How do these techniques influence consumer trust? and (4) What are the key challenges and future research directions in the implementation of XAI in personalized marketing? By addressing these questions, this study seeks to provide both theoretical and practical contributions toward the development of more transparent, accountable, and trust-oriented AI-driven marketing systems.

The remainder of this paper is organized as follows. Section 2 provides a literature review and hypothesis development. Section 3 presents the research method and design. Section 4 provides a discussion. Section 5 is Concluding Remarks and Recommendations.

2. Literature Review and Hypothesis Development

2.1 Artificial Intelligence in Marketing

Artificial Intelligence (AI) has become a central enabler of modern marketing, transforming how organizations analyze data, interact with consumers, and deliver value. The adoption of AI reflects a clear evolution from traditional marketing approaches, which relied on intuition and mass communication, toward data-driven strategies and, more recently, AI-driven systems that emphasize predictive and adaptive capabilities (Huang & Rust, 2021). AI technologies allow firms to process large-

scale and complex datasets, enabling more precise targeting and real-time decision-making. Core applications include predictive analytics, which forecasts customer behavior; marketing automation, which streamlines campaign execution; and recommendation systems, which personalize user experiences based on behavioral data (Davenport et al., 2020). Recent studies further highlight that AI enhances customer journey management and supports hyper-personalization, enabling firms to deliver context-aware, dynamic content (Chintalapati & Pandey, 2022; Mariani et al., 2022). As a result, AI is not only improving operational efficiency but also redefining competitive advantage in digital marketing ecosystems.

However, despite its growing adoption, the literature reveals several critical limitations in the application of AI in marketing. Prior research has predominantly focused on performance outcomes such as engagement, efficiency, and conversion rates, while giving less attention to user-centric issues such as transparency, trust, and ethical concerns. Huang & Rust (2021) emphasize that AI increasingly augments human decision-making, yet this augmentation introduces complexity that may not be easily interpretable by users. Similarly, Davenport et al. (2020) argue that AI-driven systems can reduce managerial visibility into decision-making processes, posing potential risks to accountability. More recent studies also point out that excessive reliance on AI may lead to consumer skepticism, particularly when personalization appears intrusive or lacks transparency (Chintalapati & Pandey, 2022). In addition, concerns related to data privacy and algorithmic bias continue to challenge the widespread acceptance of AI in marketing (Mariani et al., 2022). These limitations suggest that while AI offers substantial benefits, its long-term effectiveness depends on balancing technological performance with consumer trust and ethical considerations.

2.2 Personalized Marketing

Personalized marketing refers to the strategic use of consumer data to deliver tailored products, services, and communication that align with individual preferences and behaviors. In recent years, personalization has evolved into a data-intensive and AI-supported process that goes beyond traditional segmentation, shifting toward individual-level targeting and real-time adaptation (Lemmens et al., 2025). This transformation is driven by advances in digital technologies and machine learning, enabling firms to implement mechanisms such as behavioral targeting, recommender systems, and dynamic content personalization. Contemporary research identifies personalization as a multidimensional construct encompassing personalized recommendations, targeted advertising, and customer relationship management across the entire customer journey (Chandra et al., 2022). These developments position personalized marketing as a core capability in digital ecosystems, where firms continuously adapt their offerings based on user interactions and contextual data.

Empirical evidence consistently highlights the positive impact of personalized marketing on customer engagement, satisfaction, and loyalty. AI-driven personalization strategies, including customized recommendations and dynamic website interfaces, have been shown to significantly enhance user experience and strengthen emotional connections with brands (Zed et al., 2024). Similarly, recent studies confirm that personalized interactions increase engagement and customer satisfaction, particularly when supported by digital innovation and service quality (Hapsari & Asaari, 2025). However, the literature also emphasizes critical challenges, particularly the personalization–privacy paradox, where increased personalization may trigger concerns related to data privacy and perceived intrusiveness (Vishwakarma et al., 2025)). This tension indicates that while personalization improves

marketing effectiveness, its success depends on balancing relevance with transparency and ethical data use. Therefore, personalized marketing serves as a crucial context for examining how AI technologies are applied in consumer-facing systems and how they influence user perceptions and trust.

2.3 AI Algorithms in Personalized Marketing

AI algorithms constitute the technical foundation of personalized marketing by enabling firms to extract patterns from large-scale consumer data and deliver adaptive, data-driven decisions. Traditional machine learning techniques, such as classification and clustering, are widely applied to predict customer behavior and segment users based on shared characteristics. Classification models, including logistic regression and support vector machines, are commonly used to estimate purchase probability and response likelihood, while clustering methods such as k-means facilitate customer grouping based on behavioral similarities (Ali & Oad Rajput, 2024; Hastie, 2009). These approaches are typically categorized into supervised and unsupervised learning, depending on the availability of labeled data. Although these methods are computationally efficient and relatively interpretable, their ability to model complex, nonlinear relationships remains limited, particularly in high-dimensional, dynamic digital environments (Goodfellow, 2016).

To overcome these limitations, recent developments have emphasized the adoption of deep learning techniques, particularly neural networks, which enable more sophisticated representation learning and improved predictive performance. Deep learning models can process large-scale, unstructured data, including text, images, and sequential user interactions, thereby enhancing the accuracy of personalized marketing systems (Goodfellow et al., 2016). In parallel, recommender systems represent a core application of AI algorithms, utilizing collaborative filtering, content-based filtering, and hybrid approaches to deliver personalized recommendations (Ricci et al., 2022). Collaborative filtering leverages user-item interaction data, while content-based methods rely on item attributes, and hybrid models integrate both to improve robustness. Despite their effectiveness, these advanced models often suffer from limited interpretability and high computational complexity, raising concerns regarding transparency and scalability. This trade-off between predictive performance and explainability highlights the need for approaches that balance algorithmic accuracy with interpretability, particularly in consumer-facing personalized marketing systems.

2.4 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) refers to a set of methods and approaches designed to make the outputs of complex AI systems understandable and interpretable to human users. As artificial intelligence models, particularly those based on deep learning, have become increasingly complex, their decision-making processes have often evolved into "black-box" systems that lack transparency and are difficult to interpret (Adadi & Berrada, 2018). This opacity has raised significant concerns regarding accountability, fairness, and trust, especially in applications where decisions directly affect users. In response, XAI has emerged as a critical research area aimed at bridging the gap between high model performance and human interpretability. Recent studies emphasize that XAI not only improves transparency but also supports human-AI interaction better by enabling users to understand, trust, and effectively manage AI-driven systems (Barredo Arrieta et al., 2020) (Gunning & Aha, 2019).

Consequently, XAI is increasingly viewed as a necessary component in the deployment of AI systems in real-world, user-centric environments.

A key distinction in the XAI literature lies between inherently interpretable models and post hoc explanation techniques applied to black-box models. Interpretable models, such as decision trees and linear regression, are designed to be transparent by nature but often sacrifice predictive accuracy when dealing with complex data. In contrast, black-box models, including deep neural networks, offer superior performance but require additional explanation methods to make their outputs understandable. Techniques such as feature attribution, rule extraction, and surrogate models are commonly used to provide local or global explanations of model behavior (Doshi-velez & Kim, 2017; Rudin, 2019). However, the literature also highlights ongoing challenges, including a lack of standardization in evaluating explanation quality and a potential trade-off between interpretability and accuracy. These limitations suggest that while XAI offers promising solutions to the black-box problem, further research is needed to develop robust, user-centered approaches that align technical explainability with human understanding and trust.

3. Research Method

This study employs a systematic literature review (SLR) design to comprehensively identify, evaluate, and synthesize existing research on Explainable Artificial Intelligence (XAI) in personalized marketing. The approach is structured to ensure transparency, replicability, and rigor in the selection and analysis of relevant studies. The review focuses on three main dimensions: AI algorithms used in personalized marketing, interpretability techniques within the XAI framework, and their implications for consumer trust. By integrating both technical and behavioral perspectives, the study aims to provide a holistic understanding of the research domain and to identify patterns, trends, and gaps in the existing literature.

The unit of analysis in this study comprises peer-reviewed journal articles and conference papers on AI, XAI, and personalized marketing. The sample includes studies published within the last ten years to ensure the relevance and timeliness of the findings. Articles were selected from reputable academic databases such as Scopus, Web of Science, and Google Scholar using predefined inclusion criteria. These criteria include relevance to the research topic, empirical or conceptual contribution, and availability of full-text access. Studies that do not directly address AI applications, interpretability, or consumer-related outcomes were excluded to maintain focus and consistency.

Data collection was conducted through a structured literature search using specific keywords and Boolean operators, such as "Explainable AI," "personalized marketing," "machine learning," "interpretability," and "consumer trust." The initial search results were screened based on titles and abstracts, followed by a full-text review to ensure alignment with the study objectives. A data extraction form was developed to systematically record key information from each selected article, including author(s), publication year, research context, methodology, AI techniques used, interpretability methods, and main findings. This instrument ensured consistency and minimized bias during data collection.

The collected data were analyzed using a qualitative synthesis approach. Studies were categorized by themes such as AI algorithm types, interpretability techniques, and consumer trust outcomes. A comparative analysis was conducted to identify similarities, differences, and emerging

patterns across studies. Additionally, a descriptive analysis was used to summarize publication trends and research focus areas. This analytical process enabled the study to develop a structured overview of the literature and to highlight key insights, limitations, and directions for future research.

4. Results and Discussion

4.1 Analysis Results

4.1.1 Descriptive Analysis of Selected Studies

The descriptive analysis of the selected studies indicates a significant increase in research on Explainable Artificial Intelligence (XAI) over recent years, reflecting the rapid expansion of artificial intelligence applications across multiple domains. This growth is particularly evident in the proliferation of publications in computer science and data science, which dominate the current body of literature. In contrast, contributions from marketing and consumer behavior fields remain relatively limited. This imbalance aligns with prior findings that most XAI research has been concentrated in high-stakes domains such as healthcare, finance, and legal systems, where interpretability is critical for risk-sensitive decision-making (Islam et al., 2022). The dominance of these domains suggests that the application of XAI in consumer-oriented contexts, particularly in personalized marketing, remains an emerging and underexplored area.

In terms of publication sources, the reviewed studies are primarily drawn from high-impact journals and international conference proceedings, indicating the interdisciplinary nature of XAI research. However, a closer examination reveals that most of these studies adopt conceptual, technical, or methodological approaches, focusing on algorithm development and system performance rather than on user-centric outcomes. Empirical studies that investigate consumer behavior, trust, and perception remain scarce. This methodological imbalance highlights a critical gap between technological advancements and their real-world implications for users. Furthermore, recent trends suggest a gradual shift in research focus from purely optimizing model accuracy toward addressing issues of transparency, interpretability, and trust in AI systems (Sarkar et al., 2025). Despite this shift, the integration of technical and behavioral perspectives remains limited.

4.1.2 RQ1: AI Algorithms in Personalized Marketing

The analysis reveals that AI algorithms in personalized marketing can be broadly categorized into three main groups: traditional machine learning, deep learning, and recommender systems. Traditional machine learning techniques, including classification and clustering, are widely used to segment customers and predict behavioral patterns. These methods enable firms to identify potential customer groups and estimate purchase probabilities based on historical data. However, their effectiveness is often constrained when dealing with complex and high-dimensional datasets. In contrast, deep learning approaches, particularly neural networks, have attracted increasing attention for their ability to capture nonlinear relationships and process unstructured data such as text and images (Goodfellow et al., 2016). This capability makes deep learning particularly well-suited to dynamic, data-intensive marketing environments.

Recommender systems represent a critical application of AI algorithms in personalized marketing. Approaches such as collaborative filtering, content-based filtering, and hybrid models are

widely implemented to deliver personalized product and content recommendations (Ricci et al., 2022). Collaborative filtering relies on user interaction data, while content-based methods focus on item attributes, and hybrid approaches combine both to enhance accuracy and robustness. The findings indicate a clear trend toward adopting more advanced models, particularly deep learning-based systems, which offer superior predictive performance. However, this shift also introduces significant challenges related to model interpretability. As models become more complex, they tend to function as black-box systems, making it difficult to understand how decisions are generated (Adadi & Berrada, 2018). This creates a fundamental trade-off between accuracy and transparency. While advanced algorithms improve performance, they also reduce explainability, which is a critical factor in consumer-facing applications such as personalized marketing. Therefore, balancing performance and interpretability remains a central challenge in the use of AI algorithms.

4.1.3 RQ2: Interpretability Techniques in XAI

The findings indicate that interpretability techniques in Explainable Artificial Intelligence (XAI) can be broadly categorized into model-agnostic and model-specific approaches. Model-agnostic techniques, such as LIME and SHAP, are widely used for their flexibility in explaining various machine learning models (Sewada et al., 2023). These methods provide post-hoc explanations by approximating the behavior of complex models, allowing users to understand the contribution of input features to specific predictions. In contrast, model-specific approaches, including decision trees and rule-based systems, offer inherent interpretability by design. However, these models often struggle to maintain high predictive accuracy on complex datasets, limiting their applicability in real-world marketing scenarios.

In addition to these approaches, interpretability techniques for deep learning models, such as saliency maps and attention mechanisms, have gained increasing attention. These techniques aim to provide insights into how neural networks process input data and generate outputs. The literature further classifies interpretability methods into local versus global explanations and post hoc versus intrinsic interpretability. Local explanations focus on individual predictions, while global explanations provide an overall understanding of model behavior. Despite these advancements, the findings reveal a lack of standardized frameworks for evaluating the effectiveness and quality of explanations (Doshi-Velez & Kim, 2017). This limitation makes it challenging to compare different techniques and assess their practical usefulness. Moreover, the effectiveness of interpretability methods is highly context-dependent, varying across domains and user needs. As a result, while numerous techniques have been developed, their application in personalized marketing remains limited and fragmented. This highlights the need for more systematic evaluation and context-specific adaptation of XAI techniques in consumer-oriented environments.

4.1.4 RQ3: Impact of XAI on Consumer Trust

The findings of this study demonstrate that Explainable Artificial Intelligence (XAI) plays a critical role in shaping consumer trust in AI-driven personalized marketing systems. Specifically, XAI enhances trust by improving key dimensions, including transparency, understandability, and perceived fairness (Haque et al., 2023). When users receive clear, meaningful explanations of how recommendations or decisions are generated, they are more likely to perceive the system as reliable and trustworthy. This aligns with prior research suggesting that explainability reduces uncertainty and increases users' perceived control over

AI systems (Shin, 2021). Furthermore, empirical evidence indicates that integrating XAI techniques into recommendation systems can significantly enhance user satisfaction and engagement, particularly when explanations are perceived as relevant and easy to understand (Sarkar et al., 2025).

However, the relationship between XAI and consumer trust is not universally positive. The analysis identifies a “trust–personalisation paradox,” in which increased personalization does not necessarily lead to higher trust (Chingono et al., 2026). In some cases, highly personalized recommendations may raise concerns regarding privacy and data misuse, leading to perceptions of intrusiveness. This suggests that excessive personalization, even when supported by explainability, can negatively affect consumer perceptions. Moreover, the effectiveness of XAI in building trust appears to be context-dependent, influenced by moderating factors such as user characteristics, technological literacy, and data sensitivity. For example, users with higher levels of AI awareness may respond more positively to explanations, while others may experience information overload. These findings indicate that XAI alone is insufficient to guarantee trust; rather, its effectiveness depends on how explanations are designed, communicated, and aligned with user expectations. Therefore, a nuanced and user-centered approach is required to maximize the trust-building potential of XAI in personalized marketing contexts.

4.1.5 RQ4: Challenges and Future Research Directions

The analysis identifies several critical challenges that hinder the effective implementation of XAI in personalized marketing. From a technical perspective, key issues include scalability, computational complexity, and integrating XAI methods into real-time decision-making systems (Saarela & Podgorelec, 2024). As AI models become increasingly complex, generating explanations in a timely and efficient manner becomes more difficult, particularly in large-scale environments such as e-commerce platforms. In addition, many existing XAI techniques are not designed to operate seamlessly in dynamic, high-volume data systems, thereby limiting their practical applicability.

From a methodological standpoint, the literature is dominated by conceptual and technical studies, with little rigorous empirical validation. Many studies rely on theoretical assumptions or small-scale experiments, which limit the generalizability of their findings. Furthermore, the lack of standardized evaluation metrics for XAI effectiveness hinders comparisons across approaches and assessments of their real-world impact. This lack of standardization has been highlighted as a major barrier to advancing XAI research (Saarela & Podgorelec, 2024).

From a consumer behavior perspective, additional challenges include inconsistent levels of trust, increasing privacy concerns, and perceptions of intrusiveness associated with personalized systems. These issues are particularly relevant in marketing contexts, where user acceptance is a critical success factor. Moreover, the literature indicates a significant gap in research on XAI in consumer-oriented domains, with most studies focusing on high-risk sectors (Islam et al., 2022). This fragmentation underscores the need for future research to adopt more integrative and interdisciplinary approaches. In particular, future studies should focus on empirical validation, the development of standardized evaluation frameworks, and the exploration of user-centered design principles to enhance both transparency and trust in AI systems.

4.1.6 Synthesis and Integrative Framework

Based on a comprehensive literature analysis, this study develops an integrative framework that links AI algorithms, XAI techniques, and consumer trust within a unified conceptual structure. In this framework, AI algorithms serve as the technical foundation for personalized decision-making, leveraging data-driven models such as machine learning, deep learning, and recommender systems. However, as these algorithms become more complex, they increasingly operate as black-box systems, creating challenges related to transparency and interpretability (Adadi & Berrada, 2018). To address this issue, XAI techniques serve as an intermediary layer that translates complex model outputs into understandable explanations for users.

The framework further illustrates that the quality of explanations provided by XAI directly influences key dimensions of consumer trust, including transparency, fairness, and understandability (Haque et al., 2023). These dimensions, in turn, shape user perceptions, acceptance, and engagement with personalized marketing systems. Importantly, the relationship between XAI and consumer trust is not linear but is moderated by contextual factors such as privacy concerns, user characteristics, and the level of personalization. This aligns with the observed trust–personalization paradox, in which increased personalization may simultaneously enhance engagement and reduce trust (Chingono et al., 2026).

The primary contribution of this framework lies in its ability to integrate technical and behavioral perspectives, which have traditionally been studied in isolation. By linking algorithmic performance with user perception, the model provides a more comprehensive understanding of how AI systems operate in consumer-facing environments. This integrative perspective offers a foundation for future empirical research and supports the development of AI-driven marketing systems that are not only accurate but also transparent, accountable, and trustworthy.

4.2 Discussion

The findings of this study provide a comprehensive, integrative perspective on the role of Explainable Artificial Intelligence (XAI) in personalized marketing by linking technical advancements to consumer behavioral outcomes. From a descriptive standpoint, the results indicate that research on XAI has experienced rapid growth in recent years, yet remains heavily concentrated in technical domains such as computer science, with limited contributions from marketing and consumer behavior research. This observation is consistent with previous studies showing that XAI research has predominantly focused on high-stakes domains such as healthcare and finance, where interpretability is essential for risk-sensitive decisions (Islam et al., 2022). The dominance of conceptual and technical studies further highlights a gap between algorithm development and real-world consumer implications. This imbalance suggests that while technological innovation has advanced rapidly, its application in consumer-facing contexts, such as personalized marketing, remains underexplored, thereby reinforcing the relevance and contribution of the present study.

From the perspective of AI algorithms (RQ1), the findings confirm that personalized marketing is primarily driven by machine learning, deep learning, and recommender systems. Traditional machine learning techniques, such as classification and clustering, continue to play an important role in customer segmentation and behavior prediction. However, the increasing adoption of deep learning models reflects a shift toward more advanced approaches capable of handling large-scale and unstructured data (Goodfellow et al., 2016). This trend is consistent with prior literature highlighting the growing importance of deep learning in marketing analytics and personalization (Chintalapati & Pandey, 2022;

Mariani et al., 2022). Furthermore, recommender systems, particularly those based on collaborative filtering and hybrid approaches, remain central to delivering personalized experiences (Ricci et al., 2022). Despite these advancements, the findings reveal a critical trade-off between predictive accuracy and interpretability. As models become more complex, they increasingly function as black-box systems, limiting transparency and user understanding (Adadi & Berrada, 2018). This supports earlier arguments that the lack of interpretability in high-performing models poses significant challenges for their adoption in consumer-facing applications (Rudin, 2019).

In response to this limitation, the findings related to interpretability techniques (RQ2) demonstrate that XAI methods, including LIME, SHAP, saliency maps, and attention mechanisms, play a crucial role in improving model transparency. Model-agnostic approaches such as LIME and SHAP are widely adopted for their flexibility in explaining diverse model types. In contrast, model-specific techniques offer inherent interpretability but often at the cost of reduced predictive performance (Sewada et al., 2023). These findings align with prior studies emphasizing the importance of both post hoc and intrinsic interpretability for understanding AI systems (Barredo Arrieta et al., 2020). However, the results also reveal a lack of standardized evaluation frameworks for measuring explanation quality, which limits the comparability and practical applicability of different techniques (Doshi-Velez & Kim, 2017). This inconsistency suggests that although significant progress has been made in developing XAI methods, their effectiveness remains context-dependent and requires further empirical validation, particularly in marketing environments.

From a behavioral perspective (RQ3), the study demonstrates that XAI has a significant positive impact on consumer trust by enhancing transparency, understandability, and perceived fairness (Haque et al., 2023). These findings are consistent with previous research indicating that explainability reduces uncertainty and increases users' perceived control over AI systems, which are critical antecedents of trust (Shin, 2021). In addition, empirical evidence shows that XAI improves user satisfaction and engagement in recommendation systems, particularly when explanations are perceived as meaningful and relevant (Sarkar et al., 2025; Agarwal, 2026). However, the findings also reveal important complexities, particularly the trust–personalization paradox, in which increased personalization does not always lead to greater trust (Chingono et al., 2026). This paradox reflects earlier findings in personalized marketing research that highlight the tension between the benefits of personalization and privacy concerns (Chandra et al., 2022; Vishwakarma et al., 2025). While personalization enhances relevance and engagement, it may simultaneously trigger perceptions of intrusiveness and data misuse, thereby undermining trust. These results indicate that the relationship between XAI and consumer trust is not linear but is moderated by factors such as user characteristics, technological literacy, and data sensitivity.

Furthermore, the study identifies several critical challenges and research gaps (RQ4) that limit the effective implementation of XAI in personalized marketing. From a technical perspective, issues such as scalability, computational complexity, and integration with real-time systems remain significant barriers (Saarela & Podgorelec, 2024). These findings are consistent with prior literature emphasizing the practical limitations of deploying XAI in large-scale environments (Islam et al., 2022). From a methodological standpoint, the dominance of conceptual studies and the lack of robust empirical validation limit the generalizability of existing findings. This is further compounded by the lack of standardized metrics for evaluating XAI effectiveness, a key challenge widely recognized in the field (Doshi-Velez & Kim, 2017). From a consumer perspective, persistent concerns regarding privacy, trust instability, and perceived intrusiveness continue to hinder user acceptance of AI-driven personalization

systems. These challenges highlight the need for more interdisciplinary research that integrates technical innovation with behavioral insights.

5. Concluding Remarks and Recommendation

This study provides a systematic and integrative examination of Explainable Artificial Intelligence (XAI) within the context of personalized marketing by synthesizing insights across technical and behavioral domains. The review addresses key research questions related to the types of AI algorithms employed, the interpretability techniques used to explain these systems, their implications for consumer trust, and the challenges associated with their implementation. Overall, the study consolidates existing knowledge by mapping how machine learning, deep learning, and recommender systems underpin personalized marketing, while XAI techniques, such as model-agnostic and model-specific approaches, enhance transparency and interpretability. Furthermore, the study highlights the multifaceted nature of consumer trust, emphasizing its dependence on transparency, understandability, and perceived fairness. In doing so, this research provides a structured understanding of how technical mechanisms and user perceptions interact within AI-driven marketing environments.

This study contributes to both theory and practice by offering an original integrative framework that links AI algorithms, XAI techniques, and consumer trust in a unified model. From a theoretical perspective, it advances the literature by bridging the gap between technical AI research and consumer behavior studies, which have traditionally been examined in isolation. From a practical and managerial standpoint, the findings underscore the importance of organizations not only focusing on algorithmic performance but also prioritizing transparency and user-centric design in AI-driven systems. Managers are encouraged to implement explainability features that enhance user understanding and trust, while carefully balancing personalization with privacy considerations to avoid negative consumer responses. The study also provides actionable insights for policymakers, emphasizing the need for guidelines and standards that promote transparency, accountability, and ethical AI deployment in marketing practices.

This study has several limitations that open avenues for future research. First, the review relies primarily on existing literature, which is dominated by conceptual and technical studies, thereby limiting the availability of empirical evidence, particularly in marketing contexts. Second, the lack of standardized evaluation frameworks for XAI restricts the ability to compare findings across studies. Third, the scope of the review may not fully capture emerging developments in rapidly evolving AI technologies. Future research should focus on conducting empirical studies to validate the proposed relationships between XAI and consumer trust, particularly through experimental and longitudinal designs. Additionally, further investigation is needed to develop standardized metrics for evaluating explainability and to explore the role of moderating factors such as cultural differences, user experience, and industry context. Expanding research into real-world applications of XAI in marketing will also be critical in advancing both academic understanding and practical implementation.

Statement of Use of Generative AI

During the preparation of this work, the author used ChatGPT to assist in improving the clarity and readability of the text. The author reviewed and edited the output and takes full responsibility for the content of the publication.



References

- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Agarwal, U. (2026). Explainable AI Toolkits for Trustworthy Retail Decision-Making Systems. *IJAIDR-Journal of Advances in Developmental Research*, 17(1). <https://doi.org/10.71097/IJAIDR.v17.i1.1680>
- Ali, A., & Oad Rajput, S. K. (2024). Cracking the Code: Hidden Choices and Visible Impacts Pattern Recognition in Corporate Finance. *Asia-Pacific Financial Markets*. <https://doi.org/10.1007/s10690-024-09487-2>
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/https://doi.org/10.1016/j.inffus.2019.12.012>
- Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. *Psychology & Marketing*, 39(8), 1529–1562. <https://doi.org/10.1002/mar.21670>
- Chingono, A. T. A., Chakweza, C. T., Kanyongo, R. S., & Tlou, R. (2026). Ai-Driven Personalisation In Digital Marketing: Balancing Innovation and Consumer Privacy. *International Journal of Computer Applications*, 187(75), 68–85. <https://doi.org/10.5120/ijca2026926297>
- Chintalapati, S., & Pandey, S. K. (2022). Artificial intelligence in marketing: A systematic literature review. *International Journal of Market Research*, 64(1), 38–68. <https://doi.org/10.1177/14707853211018428>
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42. <https://doi.org/10.1007/s11747-019-00696-0>
- Doshi-velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *ML*, 1–13. <https://doi.org/10.48550/arXiv.1702.08608>
- Goodfellow, I. (2016). *Deep Learning*/Goodfellow I., Bengio Y. and Courville A. Cambridge, MA, USA: MIT Press.
- Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44–58. <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2850>.
- Hapsari, A. Y., & Asaari, A. H. (2025). The Role of Personalized Marketing and Service Quality in Enhancing Patient Engagement in Health Care Ayuningtyas. *JKBM (JURNAL KONSEP BISNIS DAN MANAJEMEN)*, 11(2), 167–180. <https://doi.org/10.31289/jkbm.v11i2.14597>
- Haque, A. K. M. B., Islam, A. K. M. N., & Mikalef, P. (2023). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change*, 186, 122120. <https://doi.org/https://doi.org/10.1016/j.techfore.2022.122120>
- Hastie, T. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. <https://doi.org/10.1007/s11747-020-00749-9>
- Islam, M. R., Ahmed, M. U., Barua, S., & Begum, S. (2022). A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Applied Sciences*, 12(3), 1353. <https://doi.org/10.3390/app12031353>
- Lemmens, A., Roos, J. M. T., Gabel, S., Ascarza, E., Bruno, H. A., Gordon, B. R., Israeli, A., McDonnell Feit, E., Mela, C. F., & Netzer, O. (2025). Personalization and targeting: how to experiment, learn & optimize. *International Journal of Research in Marketing*. <https://doi.org/https://doi.org/10.1016/j.ijresmar.2025.07.004>
- Mariani, M. M., Perez-Vega, R., & Wirtz, J. (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, 39(4), 755–776. <https://doi.org/10.1002/mar.21619>



- Ricci, F., Rokach, L., & Shapira, B. (2022). Fairness in recommender systems. *Recommender Systems Handbook*. Springer US, New York, NY. <https://doi.org/10.1007/978-1-0716-2197-4>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Saarela, M., & Podgorelec, V. (2024). Recent applications of explainable AI (XAI): A systematic literature review. *Applied Sciences*, 14(19), 8884. <https://doi.org/10.3390/app14198884>
- Sarkar, M., Rashid, M. H. O., Hoque, M. R., & Mahmud, M. R. (2025). Explainable AI in e-commerce: Enhancing trust and transparency in AI-driven decisions. *Innovatech Engineering Journal*, 2(01), 12–39. <https://doi.org/10.70937/itej.v2i01.53>
- Sewada, R., Jangid, A., Kumar, P., & Mishra, N. (2023). Explainable artificial intelligence (xai). *Journal of Nonlinear Analysis and Optimization*, 13(1), 41–47. <https://doi.org/10.36893/JNAO.2022.V13I02.041-047>
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2020.102551>
- Vishwakarma, R. K., Pandey, A., Kundnani, M. P., Yadav, A. K., Singh, M. N., & Yadav, M. S. (2025). Personalization vs. privacy: Marketing strategies in the digital age. *Journal of Marketing & Social Research*, 2, 177–191.
- Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121. <https://doi.org/10.1509/jm.15.0413>
- Zed, E. Z., Tri Mulyani Kartini, & Pupung Purnamasari. (2024). The Power Of Personalization: Exploring The Impact Of Ai-Driven Marketing Strategies On Consumer Loyalty In E-Commerce. *Jurnal Ekonomi*, 13(04 SE-Articles), 1303–1314. <https://www.ejournal.seaninstitute.or.id/index.php/Ekonomi/article/view/5694>

Corresponding author

Yassir can be contacted at: yassirrachman@yahoo.com

