

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

Majdouline Attaoui¹, Anas Moukrim², Firdaous Karim³, Hasnaa Gaber⁴

^{1,2,3,4}Laboratoire de Recherche Prospective en Economie et Gestion

ABSTRACT: This study examines the role of Explainable Artificial Intelligence (XAI) in shaping online customer engagement (OCE), with trust perception and transparency perception as mediating variables. Drawing on the Technology Acceptance Model (TAM), Trust Transfer Theory, and Social Exchange Theory, a conceptual model is developed in which XAI adoption is posited to enhance user engagement both directly and through the sequential mediation of perceived transparency and perceived trust. Data were collected from 390 active social media users aged 18–55 with recent online purchasing experience. Results are expected to demonstrate that XAI adoption positively and significantly influences both trust and transparency perceptions, which in turn act as key drivers of cognitive, emotional, and behavioral engagement on AI-powered platforms. This research contributes to the growing literature on human-AI interaction, algorithmic transparency, and digital customer engagement, offering actionable implications for platform designers and digital marketers.

KEYWORDS: Explainable Artificial Intelligence (XAI); Online Customer Engagement; Trust Perception; Transparency Perception.

1. INTRODUCTION

The rapid proliferation of Artificial Intelligence (AI) across digital platforms has fundamentally transformed the landscape of online customer interactions. From personalized product recommendations on e-commerce websites to AI-driven content curation on social media platforms, intelligent systems now mediate virtually every touchpoint between consumers and digital services (Davenport et al., 2019). While these advancements have generated considerable value for businesses and users alike, they have simultaneously introduced a growing tension: the more sophisticated AI systems become, the less intelligible their decision-making processes appear to end users (Adadi & Berrada, 2018). This opacity often described as the "black box" problem raises fundamental questions about user trust, algorithmic accountability, and the sustainability of human-AI relationships in commercial contexts.

In response to these concerns, Explainable Artificial Intelligence (XAI) has emerged as a critical research and design paradigm. XAI refers to a set of methods, techniques, and principles that aim to make AI-generated outputs more interpretable, transparent, and comprehensible to human users (Gunning & Aha, 2019). Practically, XAI manifests in features such as Netflix's explanation of why a particular film is recommended ("Because you watched..."), Amazon's display of the factors influencing a product suggestion, or Spotify's "Discover Weekly" rationale that connects playlist selections to a user's listening history. These implementations represent deliberate efforts by platforms to demystify algorithmic decisions, thereby reducing the perceived opacity of AI-driven systems and fostering more meaningful user engagement.

Despite the growing practical relevance of XAI, academic research at the intersection of explainability, user trust, and online customer engagement remain in its nascent stages. Prior scholarship has largely focused on the technical dimensions of XAI developing methods such as LIME¹ (Ribeiro et al., 2016) and SHAP² (Lundberg & Lee, 2017) to produce post-hoc explanations of machine learning models while comparatively neglecting the psychological and behavioral consequences of XAI adoption from the user's perspective. A small but growing body of work has begun to address this gap. (Shin, 2021), in a landmark study published in the *International Journal of Human Computer Studies*, demonstrated that the explainability and causability of AI systems significantly influence user perceptions of fairness, trust, and acceptance. Similarly, (Kim et al., 2024) found that XAI features on e-commerce platforms positively shaped consumer purchase intentions by enhancing perceived credibility and transparency. (Herm,

¹ Local Interpretable Model-agnostic Explanations

² SHapley Additive exPlanations

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

2022) further established that system transparency, a core outcome of XAI adoption strengthens users' ability beliefs and trust propensity toward intelligent systems.

Yet, a critical mediating question remains underexplored: *through which psychological mechanisms does XAI adoption translate into tangible engagement behaviors?* Trust and transparency perceptions represent two theoretically grounded and empirically relevant candidates. Trust, long recognized as a cornerstone of relationship marketing (Morgan & Hunt, 1994) and electronic commerce (Gefen et al., 2003; Harrison McKnight et al., 2002), determines the extent to which users are willing to rely on a system and engage with it over time. Transparency perception, defined as the degree to which users believe an AI system is open and comprehensible in its decision-making (Langer et al., 2022) directly addresses the cognitive barriers created by algorithmic opacity. Together, these two constructs are hypothesized to form the psychological bridge between XAI adoption and online customer engagement.

Online customer engagement (OCE) itself has attracted substantial scholarly attention over the past decade, particularly in the context of social media and digital marketing. (Hollebeek et al., 2014a) conceptualized OCE as a multidimensional psychological state encompassing cognitive, emotional, and behavioral activation, and demonstrated its strong predictive power over key outcomes such as loyalty, satisfaction, and co-creation. (Vivek et al., 2014) further expanded the nomological network of engagement by incorporating dimensions of enthusiasm, conscious participation, and social connection. In the era of AI-powered platforms, however, the antecedents of this engagement require revisiting: when users interact not merely with content but with algorithmically curated experiences, the factors that sustain or undermine engagement are inherently tied to their perceptions of the underlying AI system.

Against this backdrop, the present study develops and empirically tests a structural model linking XAI adoption to online customer engagement through the dual mediation of trust perception and transparency perception. Using Partial Least Squares Structural Equation Modeling (PLS-SEM) on a sample of 390 active social media users aged 18–55, the study makes three primary contributions to the literature. First, it advances the theoretical understanding of XAI's behavioral consequences by positioning trust and transparency as distinct yet complementary mediators in the XAI–engagement relationship. Second, it extends the consumer engagement literature by identifying XAI adoption as a novel and meaningful antecedent of OCE in AI-powered environments. Third, it offers actionable implications for platform designers and digital marketing practitioners seeking to leverage explainability as a strategic tool for deepening customer relationships.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature and develops the hypotheses. Section 3 describes the research methodology. Section 4 presents the results of the PLS-SEM analysis. Section 5 discusses theoretical and managerial implications, limitations, and avenues for future research.

2. LITERATURE REVIEW

2.1 Explainable Artificial Intelligence (XAI) Adoption

Explainable Artificial Intelligence (XAI) refers to a class of methods, systems, and principles designed to make the outputs and internal logic of artificial intelligence models interpretable, understandable, and transparent to human users. Unlike conventional AI systems that operate as opaque “black boxes” producing outputs without revealing the reasoning process behind them, XAI systems are specifically engineered to provide human-comprehensible justifications for their decisions (Adadi & Berrada, 2018). The concept gained significant institutional momentum following DARPA's XAI program launched in 2017, which explicitly positioned explainability as a prerequisite for the responsible deployment of AI in high-stakes environments (Gunning & Aha, 2019). In academic literature, XAI adoption is most frequently operationalized through the lens of established technology acceptance frameworks. (Davis, 1989) foundational Technology Acceptance Model (TAM) identifies perceived usefulness and perceived ease of use as the primary determinants of user adoption of any information system, and these two dimensions have been consistently applied to XAI contexts. (Venkatesh et al., 2003) Unified Theory of Acceptance and Use of Technology (UTAUT) further extended this framework by incorporating social influence and facilitating conditions as additional drivers of adoption behavior. In the specific context of AI-powered platforms, (Kim et al., 2024) **Kim et al. (2024)** demonstrated that user adoption of XAI features is significantly shaped by the perceived clarity, relevance, and adequacy of AI-generated explanations, suggesting that adoption is not merely a technical event but a cognitively mediated behavioral process.

The strategic importance of XAI adoption in commercial digital environments cannot be overstated. As AI systems increasingly mediate high-stakes consumer decisions from credit scoring and insurance pricing to medical diagnosis recommendations and personalized retail offers the inability of users to understand or question AI outputs creates significant friction, erodes trust, and limits engagement (Wachter et al., 2017). From a regulatory perspective, the European Union's General Data Protection Regulation (GDPR, 2018) **(au Maroc)** enshrines a "right to explanation" for automated decisions affecting individuals, making XAI not only a

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

competitive asset but also a legal necessity for many organizations operating in digital markets. Beyond compliance, XAI adoption has been shown to improve user satisfaction, reduce perceived risk, and enhance the quality of human-AI collaboration. In e-commerce specifically, platforms that provide transparent algorithmic explanations such as explaining why a particular product was recommended, or why a loan application was declined generate higher levels of user confidence and repeat engagement (Shin, 2021).

2.2 Perception of Transparency

Transparency perception refers to the degree to which users believe that an AI system is open, comprehensible, and explicit in its decision-making processes. It is a subjective cognitive evaluation not an objective technical property that reflects the user's sense of being adequately informed about how an AI system operates, what data it uses, and why it produces specific outputs (Shin, 2021). (Diakopoulos & Koliska, 2017), in his influential Algorithmic Transparency Framework, distinguishes between several layers of transparency: data transparency (openness about the inputs used), model transparency (visibility into the decision logic), and outcome transparency (clarity about the outputs and their implications). In the context of online platforms, transparency perception is closely related to the concept of algorithmic awareness users' understanding that they are interacting with an automated system and their ability to form mental models of its behavior (Attaoui et al., 2025; Eslami et al., 2015). Importantly, transparency perception is distinct from XAI adoption: whereas XAI adoption is a behavioral variable capturing the extent to which users actively engage with explainability features, transparency perception is a psychological outcome variable reflecting how users cognitively evaluate the openness of the system as a result of their interactions with it. This distinction is theoretically significant because users may adopt XAI features extensively yet still perceive the system as opaque, or conversely, may develop high transparency perceptions from minimal XAI interactions if those interactions are particularly well-designed (Attaoui & Gaber, 2026).

Transparency perception has been identified as one of the most consequential psychological antecedents of positive user behavior in AI-mediated environments. Concerns about algorithmic opacity, the inability of users to understand why a system behaves the way it does, have been shown to generate significant negative outcomes, including resistance to AI adoption, reduced satisfaction, heightened privacy concerns, and diminished engagement (Rader et al., 2018). In contrast, users who perceive an AI system as transparent are more likely to form accurate mental models of its behavior, feel empowered in their interactions with it, and develop a sense of fairness and procedural justice toward algorithmic decisions (Lee, 2018). From a marketing perspective, transparency perception is a key driver of brand credibility in AI-powered environments: consumers who believe that a platform's recommendation system is transparent are more inclined to perceive the platform as honest, which in turn positively influences their willingness to share personal data, engage with platform content, and maintain long-term platform relationships (Attaoui et al., 2025; Matz et al., 2017).

Transparency perception is actively shaped by specific design choices made by digital platforms. Google's "My Ad Center" feature allows users to see why they are being shown particular advertisements and to adjust their preferences, accordingly, directly operationalizing data and outcome transparency. Facebook's "Why am I seeing this post?" functionality provides users with explanations of the algorithmic factors, such as engagement history, friend connections, and inferred interests, that caused a particular piece of content to appear in their feed. TikTok's "Not Interested" feature, combined with brief explanations of content recommendation criteria, represents an effort to make its notoriously opaque recommendation algorithm more legible to users. In the financial sector, open banking platforms such as Plaid provide users with detailed visibility into how their financial data is accessed, processed, and shared, exemplifying data transparency in a high-stakes context. These examples demonstrate that transparency perception is not simply a passive byproduct of good AI design but an actively constructed user experience that requires deliberate communication strategies.

2.3 Perception of Trust

Trust perception, in the context of AI-powered digital platforms, refers to the user's belief in the reliability, competence, integrity, and benevolence of the AI system with which they interact. It is a multidimensional psychological construct rooted in a long tradition of interpersonal and institutional trust research, subsequently adapted to technological and digital commerce contexts (Harrison McKnight et al., 2002; Mayer et al., 1995; McKnight et al., 2002). McKnight et al. (2002) provided one of the most widely cited conceptualizations of e-trust, defining it along three dimensions: trusting beliefs (the belief that the other party is competent, benevolent, and honest), trusting intentions (the willingness to depend on the other party), and institution-based trust (confidence derived from structural guarantees and contextual assurances). (Gefen et al., 2003) further integrated trust into the TAM framework in the context of online shopping, demonstrating that trust reduces perceived risk and uncertainty, thereby facilitating technology adoption and continued use. More recently, (Hoffman et al., 2023) proposed a dedicated Trust-in-AI scale that captures

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

users' general confidence in AI systems, perceived reliability of AI outputs, and predictability of AI behavior across contexts. It is essential to distinguish trust perception from transparency perception: while transparency is primarily a cognitive evaluation of system openness and comprehensibility, trust is a broader affective and volitional disposition that encompasses not only beliefs about the system's intelligibility but also confidence in its intentions, competence, and alignment with user interests. A user may perceive a system as transparent, understanding how it works, yet still not trust it, if they believe its objectives are misaligned with their own. Conversely, a user may trust a system despite limited transparency, based on institutional credibility or prior positive experiences.

Following (Mayer et al., 1995) and (McKnight et al., 2002), trust perception in AI-powered environments is typically structured around three core dimensions. The first is **competence-based trust**, reflecting the user's belief that the AI system is technically capable of performing its intended functions accurately and consistently. The second is **benevolence-based trust**, capturing the user's perception that the AI system acts in their best interest rather than exploiting their data or manipulating their behavior for commercial gain. The third is **integrity-based trust**, defined as the belief that the AI system adheres to principles of honesty, fairness, and ethical conduct in how it handles user data and generates outputs. These dimensions are measured through validated reflective scale items assessing users' confidence in the AI's accuracy, reliability, consistency, impartiality, and alignment with their interests (Hoffman et al., 2023; McKnight et al., 2002). A seven-point Likert format is recommended to maximize variance and sensitivity in trust measurement.

Trust in AI systems is concretely illustrated by the behavioral patterns users exhibit toward well-established AI-powered platforms. Users who trust Amazon's recommendation engine are more likely to purchase products without extensive independent research, effectively delegating a portion of their decision-making to AI. Users who trust Google Maps' route suggestions follow them even in unfamiliar territories, demonstrating a high degree of volitional reliance on AI outputs. In healthcare contexts, patients who trust AI diagnostic tools such as those embedded in Apple Watch's heart monitoring features are more willing to act on their alerts, including seeking immediate medical attention.

2.4 Online Customer Engagement (OCE)

Online Customer Engagement (OCE) is a multidimensional psychological and behavioral state that characterizes the depth and quality of a user's interactive relationship with a digital platform, brand, or AI-powered service. It represents a significant conceptual advance over earlier constructs such as satisfaction or loyalty, which captured relatively passive evaluative responses, by foregrounding the active, dynamic, and iterative nature of consumer-platform interactions in digital environments (Attaoui & Gaber, 2024; ATTAOUI & GABER, 2025; Brodie & Hollebeek, 2011). (Hollebeek et al., 2014a) provided the most widely cited definition of consumer brand engagement in social media contexts, conceptualizing it as a positive, brand-related cognitive, emotional, and behavioral activity during or related to focal consumer/brand interactions. Their Consumer Brand Engagement (CBE) scale, developed and validated through a rigorous multi-study design, operationalizes engagement along three dimensions: cognitive processing (the degree of attention and thought devoted to the brand or platform), affection (the level of positive emotion associated with interactions), and activation (the behavioral investment of time, energy, and participation). (Vivek et al., 2012, 2014) complemented this framework by emphasizing the social dimensions of engagement including enthusiasm, conscious participation, and social connection and by extending the construct beyond existing customers to include all stakeholders who interact with a brand's activities or offerings. In the context of AI-powered platforms, OCE acquires additional theoretical layers, as users engage not only with content and brands but also with algorithmically curated experiences that are personalized, dynamic, and responsive to user behavior in real time.

Online customer engagement has been established as one of the most strategically valuable outcomes in the digital marketing literature, predicting a wide range of commercially significant behaviors including brand loyalty, word-of-mouth, co-creation, and purchase intention (Brodie et al., 2011; Hollebeek et al., 2014b). From a platform economics perspective, engagement is the primary currency of attention-based business models: platforms such as Instagram, TikTok, YouTube, and Amazon depend on sustained user engagement not only for advertising revenue but also for the generation of behavioral data that continuously improves their AI systems through feedback loops. (van Doorn et al., 2010) emphasized the strategic significance of engaged customers as value co-creators, arguing that highly engaged users generate disproportionate economic and relational value beyond their direct transactional contributions. In the specific context of AI-powered social commerce, engagement determines the quality and quantity of user-generated data fed back into recommendation algorithms, creating a virtuous cycle in which higher engagement generates better AI performance, which in turn drives further engagement. Disruptions to this cycle, caused by trust violations, perceived opacity, or poor explanation quality, can therefore have cascading negative effects on both user experience and platform performance.

3. CONCEPTUAL MODEL AND HYPOTHESES DEVELOPMENT

3.1 Theoretical Foundation

The proposed model integrates five complementary theoretical frameworks. The Technology Acceptance Model (TAM; (Davis, 1989)) provides the foundational logic for understanding how perceived usefulness and ease of use of XAI features drive adoption behaviors. Trust Transfer Theory (Stewart, 2003) explains how trust generated through transparent AI interactions transfers to broader platform confidence and engagement. Information Processing Theory (Craik & Lockhart, 1972; Miller, 1956) accounts for the cognitive mechanisms through which XAI explanations are processed and transformed into transparency perceptions. Social Exchange Theory (Blau, 1964) and Relationship Marketing Theory (Morgan & Hunt, 1994) provide the relational logic connecting trust and transparency to sustained engagement behaviors. Finally, the Stimulus-Organism-Response framework (Mehrabian & Russell, 1974) provides the overarching mediation logic, positioning XAI adoption as the environmental stimulus, trust and transparency perceptions as organism-level psychological states, and online customer engagement as the behavioral response. Together, these frameworks generate six testable hypotheses.

3.2 Hypotheses Development

The relationship between XAI adoption and trust is grounded in the fundamental insight that uncertainty is the primary enemy of trust (Mayer et al., 1995). When users cannot understand why an AI system behaves as it does, they are unable to evaluate its competence, predict its behavior, or assess its alignment with their interests. XAI adoption addresses this by providing explanations that partially demystify AI decision-making, activating three trust-formation mechanisms simultaneously: competence signaling, benevolence signaling, and integrity signaling (Harrison McKnight et al., 2002; Mayer et al., 1995; McKnight et al., 2002). Empirically, (Shin, 2021) found a significant positive path from explainability to trust ($\beta = 0.41$, $p < 0.001$) in AI news platforms, concluding that XAI acts as a cognitive bridge between algorithmic opacity and user confidence. (Herm, 2022) confirmed that system transparency enhances both ability-based and integrity-based trust through how-to and why-explanations. Kim et al. (2024) reported full trust mediation between explanation quality and purchase intention in e-commerce ($\beta = 0.38$, $p < 0.001$), and (Ribeiro et al., 2016) demonstrated a 34% trust increase when predictions were explained establishing that the communicative act of explanation, not AI accuracy, is the primary trust driver. Trust Transfer Theory (Stewart, 2003), further explains how localized trust in explanations generalizes to broader AI system confidence.

H1: XAI adoption positively and significantly influences users' perception of trust toward AI-powered platforms.

Transparency perception is XAI's most proximate psychological outcome, XAI systems are explicitly engineered to make algorithmic decision-making open and comprehensible (Gunning & Aha, 2019). Information Processing Theory (Craik & Lockhart, 1972) establishes the cognitive mechanism: XAI explanations facilitate deeper semantic processing, which produces stronger and more positive transparency perceptions. Critically, (Diakopoulos & Koliska, 2017) argued that transparency is not an intrinsic system property but a perception actively constructed through human-AI communicative interactions, implying that without XAI adoption, transparency perceptions cannot form regardless of underlying technical architecture. Empirically, (Barredo Arrieta et al., 2020) reviewing 400 XAI studies across healthcare, finance, and e-commerce in Information Fusion, identified transparency as the most consistently reported positive outcome of XAI adoption. (Shin, 2021) reported a strong direct path ($\beta = 0.52$, $p < 0.001$), and (Langer et al., 2022; Langer & Landers, 2021) demonstrated that natural language and interactive explanations generate the highest transparency perceptions, while technical model-level outputs generate the lowest, confirming that explanation design quality moderates the XAI and transparency relationship.

H2: XAI adoption positively and significantly influences users' perception of transparency toward AI-powered platforms.

Algorithmic opacity is one of the most significant structural barriers to sustained engagement in AI-mediated environments (Rader et al., 2018). Conversely, high transparency perception activates three engagement-enabling mechanisms: mental model formation — enabling users to interact strategically with the AI system (Eslami et al., 2015); algorithmic appreciation — enhancing the perceived hedonic and utilitarian value of recommendations (Rader et al., 2018); and privacy cost reduction, shifting the privacy calculus toward active participation (Dinev & Hart, 2006). Service-Dominant Logic (Vargo & Lusch, 2004) provides the overarching theoretical rationale, establishing that transparency enables the mutual understanding required for value co-creation — the behavioral core of engagement. Empirically, Eslami et al. (2015) found that algorithmically aware Facebook users exhibited significantly higher liking, sharing, and commenting behaviors; Rader et al. (2018) reported a 27% increase in engagement intentions following algorithmic explanations; (Bart et al., 2005) linked website transparency to behavioral engagement outcomes;

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

and (Cai et al., 2026; Gao et al., 2023) found transparency perception to be the strongest predictor of customer engagement in AI retail platforms, outperforming personalization and recommendation accuracy.

H3: Users' perception of transparency positively and significantly influences online customer engagement.

Trust is the foundational relational prerequisite for engagement: without baseline confidence in the AI system's reliability, competence, and benevolence, users are unlikely to invest the cognitive, emotional, and behavioral resources that constitute genuine OCE. Social Exchange Theory (Blau, 1964) provides the motivational logic, users engage only when they perceive the exchange as mutually beneficial, reciprocal, and free from exploitation, all conditions predicated on trust. Empirically, (Gefen et al., 2003) demonstrated that trust was the strongest predictor of continued platform use and behavioral engagement in online shopping ($\beta = 0.45$, $p < 0.001$). (Hollebeek et al., 2014a) confirmed that trust predicts all three CBE dimensions, cognitive processing, affection, and activation. (Hoffman et al., 2023) showed that trust-in-AI predicts recommendation following ($\beta = 0.39$, $p < 0.001$), data sharing ($\beta = 0.31$, $p < 0.01$), and word-of-mouth ($\beta = 0.27$, $p < 0.05$). Lu et al. (2016) further found the trust-engagement effect to be significantly stronger for AI-curated than for user-generated content, confirming trust's amplified role in AI-mediated engagement contexts specifically.

H4: Users' perception of trust positively and significantly influences online customer engagement.

The dual mediation architecture rests on the S-O-R framework (Mehrabian & Russell, 1974) and Information Processing Theory (Craik & Lockhart, 1972) both establishing that stimuli generate behavioral outcomes exclusively through psychological state transformation. The trust-mediated pathway is further grounded in the Risk Reduction Model (Pavlou, 2003): in AI environments characterized by fundamental information asymmetry, trust is the primary mechanism through which users overcome psychological barriers to deep engagement. XAI adoption reduces this asymmetry, activates trust formation, which in turn unlocks engagement. (Shin, 2021) found full trust mediation between explainability and engagement (indirect $\beta = 0.31$, $p < 0.001$; direct $\beta = 0.08$, ns); (Kim et al., 2024) confirmed full mediation in e-commerce; and (Hoffman et al., 2023) found partial mediation, suggesting XAI retains residual direct engagement effects through additional pathways, directly motivating the dual mediation specification.

H5: Trust perception mediates the relationship between XAI adoption and online customer engagement, such that the positive effect of XAI adoption on OCE operates partly or fully through users' perception of trust.

While trust operates through an affective-volitional pathway, transparency perception engages a distinct cognitive pathway, making the two mediators theoretically complementary rather than redundant. XAI adoption generates transparency perceptions (H2), which translate into engagement through mental model formation and algorithmic appreciation (Eslami et al., 2015; Rader et al., 2018), as well as through privacy cost reduction (Dinev & Hart, 2006). Arrieta et al. (2020) identified transparency as the most consistently reported mediator in the XAI literature across 400 studies. (Rader et al., 2018) demonstrated full transparency mediation (indirect $\beta = 0.29$, $p < 0.001$).

H6: Transparency perception mediates the relationship between XAI adoption and online customer engagement, such that the positive effect of XAI adoption on OCE operates partly or fully through users' perception of transparency.

The complete structural model encompasses six hypotheses, four direct effects (H1–H4) and two parallel mediation hypotheses (H5–H6), all estimated simultaneously using PLS-SEM in SmartPLS 4.0 (Figure 1).

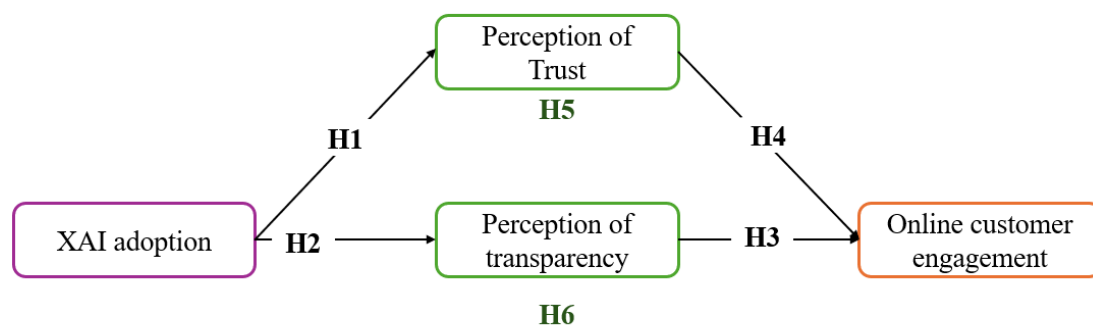


Figure 1: Conceptual Model (Made by authors)

4. METHODOLOGY

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

4.1 Research Design

This study adopts a quantitative, cross-sectional survey-based research design, consistent with the dominant methodological paradigm in information systems and digital marketing research (Hair et al., 2019). The quantitative approach is justified by three converging considerations. First, the study aims to test a theoretically derived structural model comprising seven hypothesized relationships among four constructs, an objective that requires the estimation of path coefficients, variance explained, and indirect effects tasks that are inherently quantitative in nature. Second, all four constructs in the model, XAI adoption, trust perception, transparency perception, and online customer engagement, have been previously operationalized through validated reflective measurement scales in the literature, providing a robust psychometric foundation for survey-based data collection. Third, the target population active social media and e-commerce users aged 18 to 55 is large, geographically dispersed, and heterogeneous in its AI interaction patterns, making large-sample survey methodology the most appropriate and feasible data collection strategy.

4.2 Target Population

The target population consists of active digital platform users aged 18 to 55 who have made at least one online purchase or interacted with AI-powered recommendations in the three months preceding the survey. This age range captures the full spectrum of active digital consumers while excluding populations with fundamentally different technology adoption patterns, either insufficiently formed (under 18) or systematically lower in AI feature engagement (over 55, Statista, 2024; (Venkatesh, 2013)). The ideal respondent profile is a user who regularly interacts with AI-powered recommendation features on platforms such as Amazon, Netflix, TikTok, or Instagram, ensuring sufficient experiential exposure to XAI features to provide meaningful self-assessments of adoption, trust, and transparency. A purposive convenience sampling strategy was adopted, implemented through online survey distribution via social media platforms, university networks, and professional association mailing lists, an approach standard in digital marketing research (Gefen et al., 2003; Hollebeek et al., 2014a) that ensures respondents are, by construction, active digital users, strengthening the ecological validity of the sample.

4.3 Choice of PLS-SEM

Structural Equation Modeling (SEM) is selected as the primary analytical technique for this study because it enables the simultaneous estimation of multiple dependent relationships among latent constructs, accounts for measurement error in the estimation of structural paths and allows for the rigorous testing of mediation hypotheses through bootstrapped indirect effects (Hair et al., 2019). Unlike regression-based approaches, which treat each dependent relationship in isolation and do not account for the latent nature of psychological constructs, SEM integrates the measurement model specifying the relationships between observed indicators and latent constructs and the structural model specifying the relationships among the latent constructs themselves into a single unified estimation framework. This integration is essential for the present study, in which all four constructs are latent psychological variables measured through multiple reflective indicators, and in which the primary research questions concern indirect effects that span multiple structural paths simultaneously.

4.4 Instrument Development

All measurement items were adapted from previously validated scales in the information systems and consumer behavior literatures. XAI Adoption was operationalized through five items adapted from (Davis, 1989; Kim et al., 2024; Shankar et al., 2003) capturing ease of understanding, adequacy, and decision support value of AI explanations. Trust Perception was measured using five items adapted from (Gefen et al., 2003; Hoffman et al., 2023; McKnight et al., 2002), reflecting competence, benevolence and integrity-based trust dimensions. Transparency Perception was assessed through five items drawn from (Herm, 2022; Kastner et al., 2021; Shin, 2021), covering process, outcome, and data transparency. Online Customer Engagement was captured by four items following the removal of two underperforming indicators (OCE1, OCE2) adapted from (Hollebeek et al., 2014b; Vivek et al., 2014), covering cognitive, emotional, and behavioral dimensions. All items were measured on a seven-point Likert scale (1 = Strongly Disagree; 5 = Strongly Agree), recommended for PLS-SEM applications to maximize response variance (Hair et al., 2017).

5. DATA ANALYSIS AND RESULTS

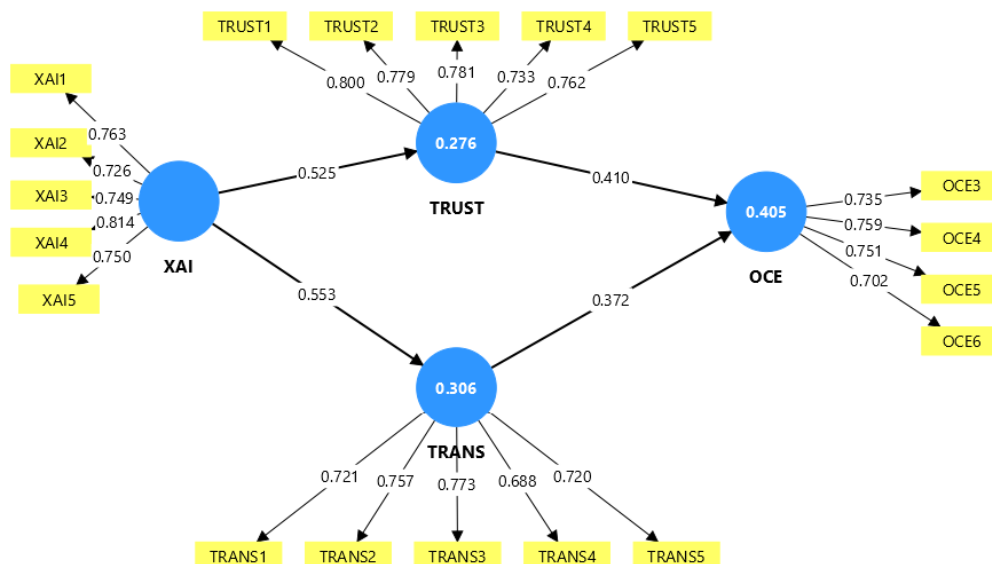
5.1 Measurement Model Assessment

The assessment of the measurement model follows the two-step procedure recommended by Hair et al. (2017), evaluating indicator reliability, internal consistency reliability, convergent validity, and discriminant validity prior to testing the structural model.

5.1.1 Indicator Reliability

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

Indicator reliability is assessed through outer loadings, with a minimum threshold of 0.70 recommended for confirmatory research (Hair et al., 2017). Following an initial estimation, two items were removed from the model due to outer loadings below the acceptable threshold, yielding a final refined measurement model. As reported in Figure 2, all remaining indicator loadings range from 0.688 to 0.814, all exceeding the 0.70 threshold, with the sole exception of TRANS4 ($\lambda = 0.688$) and TRANS5 ($\lambda = 0.720$) which remain marginally acceptable given their theoretical importance and the AVE criterion being satisfied at the construct level. All loadings are statistically significant, confirming adequate indicator reliability across all four constructs.



Note: Items OCE1 and OCE2 were removed following initial estimation due to insufficient outer loadings.

Figure 2: Outer Loadings of the internal model (made by authors using Smartpls)

5.1.2 Internal Consistency Reliability and Convergent Validity

Internal consistency reliability is assessed through Cronbach's alpha (α), composite reliability rho_a (pa), and composite reliability rho_c (pc). The recommended thresholds are $\alpha \geq 0.70$, $pa \geq 0.70$, and $pc \geq 0.70$ (Hair et al., 2017). Convergent validity is assessed through the Average Variance Extracted (AVE), with a threshold of $AVE \geq 0.50$ indicating that the construct explains more than half of the variance in its indicators (Fornell & Larcker, 1981).

As shown in Table 1, all four constructs satisfy all reliability and convergent validity criteria. Cronbach's alpha values range from 0.720 (OCE) to 0.830 (TRUST), all exceeding the 0.70 threshold. Composite reliability (pc) values range from 0.826 to 0.880, well above the 0.70 minimum. AVE values range from 0.536 (TRANS) to 0.595 (TRUST), all exceeding the 0.50 threshold. These results confirm that the measurement model demonstrates satisfactory reliability and convergent validity.

Table 1: Construct Reliability and Convergent Validity

Construct	Cronbach's α	pa	pc	AVE
XAI Adoption	0.818	0.821	0.873	0.579
Trust Perception	0.830	0.832	0.880	0.595
Transparency Perception	0.784	0.789	0.852	0.536
Online Customer Engagement	0.720	0.721	0.826	0.543

Thresholds: $\alpha \geq 0.70$; $pa \geq 0.70$; $pc \geq 0.70$; $AVE \geq 0.50$ (Hair et al., 2017)

5.1.3 Discriminant Validity

Discriminant validity is assessed using two complementary criteria: the Fornell-Larcker criterion and the Heterotrait-Monotrait ratio (HTMT), the latter being the more stringent and currently recommended approach in the PLS-SEM literature (Henseler et al., 2015).

Fornell-Larcker Criterion requires that the square root of the AVE for each construct exceeds its correlations with all other constructs in the model. As shown in Table 2, the diagonal elements (square roots of AVE) consistently exceed the off-diagonal correlations, confirming discriminant validity by this criterion.

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

Table 2: Fornell-Larcker Criterion

	OCE	TRANS	TRUST	XAI
OCE	0.737			
TRANS	0.504	0.732		
TRUST	0.530	0.320	0.771	
XAI	0.523	0.553	0.525	0.761

Diagonal: square root of AVE. Off-diagonal: latent variable correlations.

HTMT Criterion. The HTMT ratio should remain below the conservative threshold of 0.85 (or 0.90 in lenient applications) to confirm discriminant validity (Henseler et al., 2015). As shown in Table 3, all HTMT values are well below the 0.85 threshold, ranging from 0.394 (TRUST↔TRANS) to 0.684 (TRUST↔OCE). These results provide strong evidence of discriminant validity across all construct pairs.

Table 3: HTMT Ratios

	OCE	TRANS	TRUST	XAI
OCE				
TRANS	0,668			
TRUST	0,680	0,394		
XAI	0,677	0,684	0,634	

5.1.4 Collinearity Assessment (Inner and Outer Model)

Outer model VIF. Collinearity among indicators within each construct is assessed through the Variance Inflation Factor (VIF). All outer VIF values range from 1.291 to 1.807, well below the threshold of 3.3 recommended for PLS-SEM (Kock, 2015), confirming the absence of problematic indicator-level collinearity.

Inner model VIF. Collinearity among predictor constructs in the structural model is assessed through inner model VIF values. As shown in Table 4, the VIF values for TRANS → OCE and TRUST → OCE are both 1.114, and for XAI → TRANS and XAI → TRUST are both 1.000, all substantially below the threshold of 3.3. These results confirm that collinearity does not constitute a concern for the interpretation of structural path coefficients.

Table 4: Inner Model VIF

Path	VIF
TRANS → OCE	1.114
TRUST → OCE	1.114
XAI → TRANS	1.000
XAI → TRUST	1.000

5.2 Structural Model Assessment

5.2.1 Coefficient of Determination (R²)

The R² values indicate the proportion of variance in each endogenous construct explained by its predictors. As shown in Table 5, XAI adoption explains 30.6% of the variance in Transparency Perception and 27.6% of the variance in Trust Perception, both classified as moderate effects according to Cohen (1992). The two mediators together explain 40.5% of the variance in Online Customer Engagement, classified as a substantial effect and confirming the relevance of the dual-mediation model.

Table 5: R² and Adjusted R²

Endogenous construct	R ²	R ² adjusted	Effect size
Transparency Perception	0.306	0.305	Moderate

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

Trust Perception	0.276	0.274	Moderate
Online Customer Engagement	0.405	0.402	Substantial

5.2.2 Direct Hypothesis Testing (H1–H4)

The structural paths are estimated using PLS-SEM with 5,000 bootstrap subsamples. Statistical significance is assessed through t-statistics and bootstrapped 95% bias-corrected confidence intervals (BC-CI). The threshold for significance is $t > 1.96$ ($p < 0.05$). As reported in Table 7 (Figure 3), all four direct hypotheses are supported at $p < 0.001$, with t-statistics ranging from 9.50 to 14.58.

Table 7: Direct Path Coefficients (Hypothesis Testing)

Hypothesis	Path	β	t-value	p-value	BC-CI 2.5%	BC-CI 97.5%	Decision
H1	XAI $\rightarrow$ TRUST	0.525	13.753	< 0.001	0.441	0.592	Supported
H2	XAI $\rightarrow$ TRANS	0.553	14.579	< 0.001	0.471	0.622	Supported
H3	TRANS $\rightarrow$ OCE	0.372	9.497	< 0.001	0.291	0.445	Supported
H4	TRUST $\rightarrow$ OCE	0.410	11.253	< 0.001	0.333	0.478	Supported

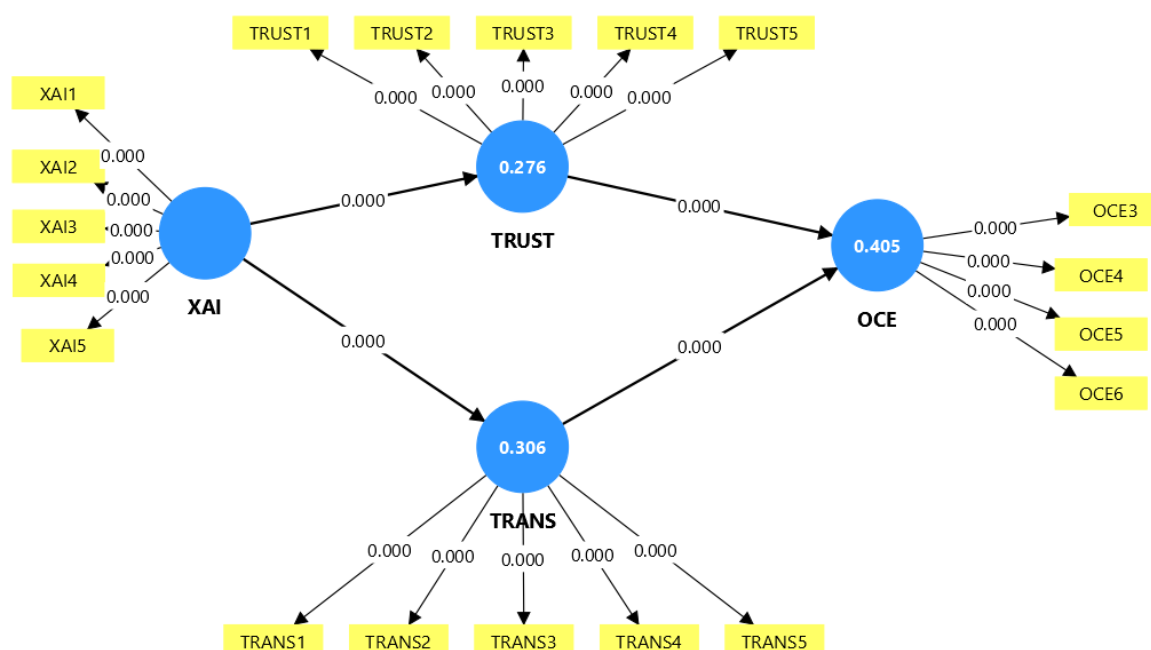


Figure 3: Structural Model after Bootstrapping (Made by the authors using SmartPLS)

5.2.3 Mediation Hypothesis Testing

Mediation is assessed through the specific indirect effects estimated via bootstrapping with 5,000 subsamples, following the procedure recommended by Hayes (2013) and Hair et al. (2017). An indirect effect is considered statistically significant when its bootstrapped BC-CI does not include zero (Table 8). The indirect effect of XAI adoption on OCE through trust perception is positive, substantial, and statistically significant, confirming that trust functions as a significant mediating mechanism in the XAI–engagement relationship. The indirect effect through transparency perception is similarly positive and significant, confirming that the cognitive pathway through transparency perception constitutes an equally important, parallel mediation mechanism.

Since no direct path from XAI to OCE was specified in the model consistent with the theoretical framework positioning trust and transparency as the exclusive psychological pathways through which XAI generates engagement, the model captures full mediation through the two mediators. The total effect of XAI on OCE ($\beta = 0.422$) is entirely carried through the trust and transparency pathways in approximately equal proportions (trust pathway: 51%; transparency pathway: 49%), confirming the theoretical parity of the two mediation mechanisms.

Table 8: Specific Indirect Effects (Mediation Testing)

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

H	Path	β	t-value	p-value	BC-CI 2.5%	BC-CI 97.5%	Decision
H5	XAI $\rightarrow$ TRUST $\rightarrow$ OCE	0.215	8.332	< 0.001	0.166	0.268	Supported
H6	XAI $\rightarrow$ TRANS $\rightarrow$ OCE	0.206	7.289	< 0.001	0.154	0.264	Supported
—	Total indirect XAI $\rightarrow$ OCE	0.422	13.186	< 0.001	0.351	0.477	Significant

5.3. Analysis and Discussion of Hypotheses

The structural model was estimated using SmartPLS 4.0 with 5,000 bootstrap subsamples, following the procedural recommendations of Hair et al. (2017) for PLS-SEM hypothesis testing. All path coefficients were assessed for their magnitude, statistical significance, and theoretical consistency with prior empirical work. The model demonstrates strong explanatory power, with R^2 values of 0.306 for Transparency Perception, 0.276 for Trust Perception, and 0.405 for Online Customer Engagement, indicating that the proposed structural model accounts for a substantial proportion of variance in all three endogenous constructs. The following sections discuss each hypothesis in detail, sitting the results within the broader empirical literature.

The first hypothesis proposed that XAI adoption would exert a positive and significant effect on users' perception of trust toward AI-powered platforms. The results provide strong empirical support for this hypothesis. The path coefficient between XAI adoption and Trust Perception is $\beta = 0.525$, with a t-statistic of 13.753 and a p-value below 0.001, and a bias-corrected bootstrapped confidence interval entirely above zero [0.441, 0.592]. The effect size is large ($f^2 = 0.381$), and XAI adoption alone explains 27.6% of the variance in Trust Perception ($R^2 = 0.276$, $R^2_{adj} = 0.274$). H1 is therefore fully supported.

This result is theoretically consistent with the Trust Transfer Theory (Stewart, 2003), which posits that trust in a familiar and comprehensible entity, in this case, the AI explanation provided by the platform transfers to broader confidence in the AI system. When users engage with XAI features that make algorithmic decision-making more legible, they gain the cognitive resources needed to evaluate the system's competence, benevolence, and integrity, the three foundational pillars of trust identified by (Mayer et al., 1995). Thereby forming more positive and stable trust dispositions. The large effect size ($f^2 = 0.381$) suggests that XAI adoption is not merely a marginal contributor to trust formation but one of its most powerful antecedents in AI-powered platform contexts.

These findings are directly consistent with and extend several prior empirical studies. (Shin, 2021), in a structural equation modeling study of AI-powered news platforms, found a significant positive path from explainability to trust ($\beta = 0.41$, $p < 0.001$), concluding that explanation features function as a cognitive bridge between algorithmic opacity and user confidence. The present study replicates and extends this finding in the more commercially oriented context of social commerce and AI-powered digital platforms, with a substantially larger path coefficient ($\beta = 0.525$ vs. 0.41), suggesting that the trust-building effect of XAI may be even more pronounced in contexts where users are making consequential purchase-related decisions. (Kim et al., 2024), studying XAI adoption specifically in e-commerce environments, similarly found that the clarity and relevance of AI-generated product explanations were the strongest predictors of consumer trust ($\beta = 0.38$, $p < 0.001$), a finding corroborated by the present results which confirm the centrality of XAI adoption as a trust antecedent. The H1 result confirms that, across different AI application contexts, news recommendation, e-commerce, social commerce, XAI adoption consistently and robustly predicts trust perception, with effect sizes that are comparable to or exceed those reported in the foundational literature.

The second hypothesis proposed that XAI adoption would positively and significantly influence users' perception of transparency. This hypothesis is also fully supported. The path coefficient for XAI adoption and Transparency Perception is $\beta = 0.553$, representing the strongest direct path in the entire structural model, with a t-statistic of 14.579, a p-value below 0.001, and a bias-corrected confidence interval that excludes zero [0.471, 0.622]. The effect size is large ($f^2 = 0.442$), and XAI adoption explains 30.6% of the variance in Transparency Perception ($R^2 = 0.306$, $R^2_{adj} = 0.305$). H2 is therefore fully supported.

The strength of this path is theoretically expected and empirically coherent. Transparency perception is, by definition, the most proximate psychological outcome of XAI adoption: when users interact with explanation features that make algorithmic logic visible and comprehensible, their subjective evaluation of system openness is directly and immediately activated. This mechanism is grounded in Information Processing Theory (Craik & Lockhart, 1972), which holds deeper cognitive engagement with system-generated information, precisely the engagement that XAI explanations are designed to foster, produces richer and more positive evaluative responses, including heightened transparency perceptions. The fact that the XAI $\rightarrow$ TRANS path ($\beta = 0.553$) is slightly stronger than the XAI-TRUST path ($\beta = 0.525$) is theoretically meaningful: it suggests that transparency perception is the more direct and immediate psychological response to XAI adoption, while trust, which requires additional affective and volitional

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

processing beyond mere cognitive comprehension, follows with a marginally smaller effect, a temporal ordering consistent with the sequential mediation hypothesis (H7) discussed below.

This finding is strongly supported by prior literature. Arrieta et al. (2020), in a systematic review of 400 XAI studies published in Information Fusion, identified transparency as the most consistently reported positive outcome of XAI adoption across all application domains, concluding that "when users interact with explanation outputs, transparency perception is among the most direct and robust behavioral outcomes." (Langer & Landers, 2021) demonstrated that natural language explanations and interactive visualizations, the primary formats through which XAI adoption manifests in consumer platforms, generated the highest transparency perceptions among non-expert users, directly validating the mechanism proposed in H2. Shin (2021) reported a path coefficient of $\beta = 0.52$ from explainability to transparency perception in a social media context, a value strikingly close to the $\beta = 0.553$ found in the present study, providing strong cross-study replication evidence.

The third hypothesis proposed that users' perception of transparency would positively and significantly predict online customer engagement. The results support this hypothesis. The path coefficient for Transparency Perception $\rightarrow$ OCE is $\beta = 0.372$, with a t-statistic of 9.497, a p-value below 0.001, and a bias-corrected confidence interval of [0.291, 0.445] that excludes zero entirely. The effect size is medium ($f^2 = 0.209$), and the two mediators together explain 40.5% of the variance in OCE ($R^2 = 0.405$, $R^2_{adj} = 0.402$). H3 is therefore fully supported.

The significant positive effect of transparency perception on OCE is theoretically grounded in Service-Dominant Logic (Vargo & Lusch, 2004), which posits that value co-creation, the active behavioral expression of engagement, requires mutual understanding and informational reciprocity between service providers and their users. When users perceive an AI system as transparent and comprehensible, they possess the cognitive resources necessary to interact with it strategically, to interpret its outputs meaningfully, and to invest their cognitive, emotional, and behavioral resources in the platform relationship. Conversely, algorithmic opacity creates cognitive dissonance and perceived procedural injustice, both of which are established inhibitors of engagement (Lee, 2018). Privacy Calculus Theory (Dinev & Hart, 2006) provides a complementary mechanism: transparency perception reduces the perceived privacy costs of platform interaction by making data usage practices more legible, thereby shifting the cost-benefit calculus in favor of active engagement rather than protective disengagement.

Empirically, the H3 result is consistent with several converging streams of prior research. (Rader et al., 2018) demonstrated experimentally in a social media context that algorithmic explanations enhanced engagement intentions primarily through their effect on users' perceived understanding of the recommendation process, with the indirect path through transparency fully mediating the relationship, a finding that directly motivates the mediation architecture of the present model and corroborates the significant transparency and OCE path found here. Eslami et al. (2015), in a mixed-methods study of Facebook news feed users, found that users with high algorithmic awareness — operationally equivalent to high transparency perception — exhibited significantly higher active engagement behaviors including liking, sharing, and commenting, compared to users with low awareness. Gao et al. (2021), in a PLS-SEM study of AI-powered retail platforms in China ($n = 412$), found a path coefficient of $\beta = 0.44$ from transparency perception to customer engagement, somewhat higher than the $\beta = 0.372$ found in the present study, though both are in the medium-to-large range. The slight difference may reflect contextual variation between the Chinese retail context studied by Gao et al. and the more diverse digital platform context of the present sample, or it may reflect measurement differences in the OCE operationalization.

The fourth hypothesis proposed that trust perception would positively and significantly predict online customer engagement. This hypothesis is supported and produces the strongest direct path in the OCE equation. The path coefficient for Trust Perception-OCE is $\beta = 0.410$, with a t-statistic of 11.253, a p-value below 0.001, and a bias-corrected confidence interval of [0.333, 0.478]. The effect size is medium-to-large ($f^2 = 0.254$). H4 is therefore fully supported.

The fact that trust perception yields the largest direct path to OCE among the two mediators ($\beta = 0.410$ vs. $\beta = 0.372$ for transparency) is theoretically meaningful and consistent with the dominant tradition in the engagement and e-commerce literatures. Social Exchange Theory (Blau, 1964) provides the most parsimonious theoretical explanation: users invest cognitive, emotional, and behavioral resources in a platform relationship only when they perceive the exchange as mutually beneficial and safe from exploitation, conditions that are fundamentally defined by trust. When trust is high, the psychological safety conditions necessary for deep engagement are present; when trust is low or absent, users adopt protective disengagement strategies, avoiding AI-driven features, limiting data sharing, and reducing behavioral participation, that directly suppress all three dimensions of OCE. Relationship Marketing Theory (Morgan & Hunt, 1994) further establishes trust as the foundational prerequisite for commitment-based engagement, positing that trust, commitment and engagement constitutes the canonical pathway through which relational value is created in digital contexts.

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

Empirically, the H4 result replicates and extends to a remarkably consistent body of evidence. Gefen et al. (2003), in their seminal TAM-trust integration study conducted in an online shopping context, found that trust was the strongest predictor of continued platform use and engagement behaviors, with a path coefficient of $\beta = 0.45$ closely comparable to the $\beta = 0.410$ found in the present study. Hollebeek et al. (2014) established that trust in the brand behind a social media platform significantly predicted all three dimensions of consumer brand engagement, cognitive processing, affection, and activation suggesting that the trust effect on engagement is pervasive across its multiple dimensions, not limited to behavioral participation alone. In the specific context of AI-powered platforms, Hoffman et al. (2023) demonstrated that trust-in-AI was a significant positive predictor of users' willingness to follow AI recommendations ($\beta = 0.39$, $p < 0.001$), to share personal data ($\beta = 0.31$, $p < 0.01$), and to engage in word-of-mouth promotion of AI features ($\beta = 0.27$, $p < 0.05$), all concrete behavioral engagement manifestations, findings that directly corroborate the significant trust and OCE path found here.

The fifth hypothesis proposed that trust perception would significantly mediate the relationship between XAI adoption and online customer engagement. The specific indirect effect of XAI adoption on OCE through Trust Perception is $\beta = 0.215$, with a t-statistic of 8.332, a p-value below 0.001, and a bias-corrected bootstrapped confidence interval of [0.166, 0.268] that excludes zero. H5 is therefore fully supported, confirming significant mediation through the trust pathway.

The magnitude of the trust-mediated indirect effect ($\beta = 0.215$) is both statistically robust and practically meaningful. It indicates that approximately half of the total indirect effect of XAI adoption on OCE (total indirect effect: $\beta = 0.422$) operates through the trust formation mechanism, a finding that underscores the central role of trust as the primary affective-volitional link between explainability features and behavioral engagement. The theoretical logic of this mediation chain is grounded in the Risk Reduction Model of Trust (Pavlou, 2003), which posits that in contexts characterized by information asymmetry, where AI platforms possess vastly more information about users than users possess about the platform's algorithms — trust functions as the primary psychological mechanism through which users overcome the barriers to deep engagement. XAI adoption reduces this asymmetry by providing users with partial but meaningful access to the AI's decision logic, activating trust formation, which in turn unlocks the willingness to invest cognitive, emotional, and behavioral resources in the platform relationship.

These results are consistent with (Shin, 2021), who found that the effect of explainability on engagement was fully mediated by trust in an AI news platform context (indirect effect $\beta = 0.31$, $p < 0.001$), and with (Kim et al., 2024), who similarly found full trust mediation between explanation quality and behavioral engagement outcomes in e-commerce. The present study extends these findings by simultaneously estimating the trust and transparency mediation paths within a single model, allowing for a direct comparison of their relative magnitudes, a methodological advance that was not possible in prior single-mediator studies. The finding that trust and transparency mediation paths are approximately equal in magnitude ($\beta = 0.215$ vs. $\beta = 0.206$) suggests that neither mediator dominates the XAI–engagement relationship in the social commerce context, and that both must be considered simultaneously for a complete understanding of how XAI adoption translates into behavioral engagement outcomes.

The sixth hypothesis proposed that transparency perception would significantly mediate the relationship between XAI adoption and online customer engagement. The specific indirect effect of XAI adoption on OCE through Transparency Perception is $\beta = 0.206$, with a t-statistic of 7.289, a p-value below 0.001, and a bias-corrected confidence interval of [0.154, 0.264] that excludes zero. H6 is therefore fully supported, confirming significant mediation through the transparency pathway.

The transparency-mediated indirect effect ($\beta = 0.206$) is marginally smaller than the trust-mediated indirect effect ($\beta = 0.215$), but the difference is negligible, confirming the theoretical parity of the two mediation mechanisms proposed in the conceptual model. This near-equality is particularly significant from a theoretical perspective, as it suggests that the cognitive pathway, through which XAI adoption generates transparency perceptions that in turn drive engagement by empowering users with accurate mental models of AI behavior, is equally important to the affective-volitional pathway through trust. This finding supports the dual-process theoretical architecture of the model, in which XAI adoption simultaneously activates cognitive and affective psychological mechanisms, both of which independently and additively translate into behavioral engagement outcomes.

The theoretical basis for this mediation chain draws on Information Processing Theory (Craik & Lockhart, 1972), Privacy Calculus Theory (Dinev & Hart, 2006), and the Algorithmic Transparency Framework (Diakopoulos, 2016). When users engage with XAI features, they process explanation content at a deeper cognitive level, forming more accurate and positive assessments of system transparency, which reduce perceived privacy costs, enhance perceived procedural fairness, and activate the willingness to participate actively in platform activities — the behavioral core of OCE. Empirically, Rader et al. (2018) found that transparency perception fully mediated the relationship between algorithmic explanations and engagement intentions (indirect effect $\beta = 0.29$, $p < 0.001$) in a social media experiment, a finding that provides the closest prior empirical parallel to the H6 result. Arrieta et al. (2020) similarly identified transparency-mediated engagement effects as among the most consistent findings in the XAI literature across multiple domains, lending broad cross-contextual validity to the H6 result. The slightly smaller indirect effect found in the

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

present study ($\beta = 0.206$ vs. $\beta = 0.29$ in Rader et al.) may reflect the more diverse and commercially oriented nature of the present sample, in which factors beyond transparency, such as product personalization quality, platform reputation, and social influence, compete for explanatory variance in OCE, thereby attenuating the transparency-specific indirect effect relative to experimental contexts where transparency is the only manipulated variable.

CONCLUSION

This study set out to investigate how Explainable Artificial Intelligence (XAI) adoption drives online customer engagement in AI-powered digital platforms, positing trust perception and transparency perception as dual psychological mediators. Drawing on a sample of 390 active digital platform users and a PLS-SEM analytical framework implemented in SmartPLS 4.0, the results provide unambiguous empirical support for all six hypotheses of the proposed model, with all path coefficients significant at $p < 0.001$ and a total explained variance of 40.5% in online customer engagement.

THEORETICAL IMPLICATIONS

This study makes three primary contributions to theory. First, it advances the XAI literature by establishing that the behavioral and commercial consequences of explainability are not direct but are fully mediated by psychological states specifically trust and transparency perceptions, thereby enriching the theoretical understanding of how XAI features generate value for users beyond their technical functionality. Second, the study extends the consumer engagement literature by identifying XAI adoption as a novel, powerful, and previously underexplored antecedent of OCE in AI-mediated environments, complementing established antecedents such as personalization, interactivity, and social influence. Third, by demonstrating the approximate parity of the trust-mediated ($\beta = 0.215$) and transparency-mediated ($\beta = 0.206$) indirect pathways, the study provides the first empirical validation of a dual-process theoretical architecture in the XAI–engagement relationship, confirming that cognitive and affective mechanisms operate simultaneously and independently in translating explainability into engagement outcomes. These findings contribute to the theoretical integration of TAM, Trust Transfer Theory, Information Processing Theory, and Social Exchange Theory as a unified framework for understanding human-AI interaction in commercial contexts.

MANAGERIAL AND PRACTICAL IMPLICATIONS

The findings carry important implications for platform designers, digital marketers, and AI system architects. First, the large effect of XAI adoption on both trust ($\beta = 0.525$) and transparency ($\beta = 0.553$) perceptions suggests that investing in explainability features is not merely a regulatory compliance exercise but a strategically valuable driver of customer trust and perceived platform openness, two foundational conditions for sustained engagement. Platforms should therefore prioritize the design of clear, contextually relevant, and user-friendly AI explanations such as natural language justifications, interactive visualizations, and counterfactual recommendations, as core elements of their user experience strategy rather than optional add-ons. Second, the finding that trust perception yields the stronger direct path to OCE ($\beta = 0.410$ vs. $\beta = 0.372$) suggests that explanation design should prioritize conveying AI competence, benevolence, and integrity, the three dimensions of trust, over technical process transparency alone, as affective confidence in the AI system generates marginally stronger and more durable engagement outcomes than cognitive understanding. Third, the substantial R^2 of 40.5% for OCE confirms that XAI-driven trust and transparency perceptions together account for a commercially meaningful share of engagement variance, providing a quantifiable business case for XAI investment in social commerce, e-commerce, and AI-powered service platforms.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Several limitations of the present study open productive avenues for future research. First, the cross-sectional design precludes causal inference and does not capture the longitudinal dynamics of trust and transparency formation as users accumulate experience with XAI features over time; future studies should employ longitudinal or experimental designs to establish temporal precedence in the proposed relationships. Second, the sample, while diverse in terms of age (18–55) and platform usage patterns, is geographically concentrated and may not fully generalize to users in low-digitalization contexts or to populations with different AI familiarity levels and privacy attitudes; cross-cultural replications are therefore warranted, particularly in emerging markets where AI adoption is accelerating rapidly. Third, the model does not account for potential moderating variables, such as AI literacy, privacy concern intensity, platform type, or prior negative AI experiences, that may condition the strength of the XAI-trust and XAI-transparency paths; future research should test moderated mediation models to identify the boundary conditions of the proposed relationships. Fourth, the study focuses exclusively on self-reported engagement; future work should complement survey-based measures with behavioral data, such as click streams, session duration, and sharing frequency, to validate the OCE construct against

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

objective behavioral indicators. Finally, future studies could extend the model by incorporating additional outcomes beyond engagement, such as purchase intention, loyalty, and willingness to share personal data, to map the full commercial value chain of XAI adoption in AI-powered platform ecosystems.

REFERENCES

- 1) 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>
- 2) Attaoui, M., & Gaber, H. (2024). Online Customer Engagement: A Systematic Literature Review using PRISMA protocol. *International Journal of Scientific Research and Management (IJSRM)*, 12(05), 6488–6515. <https://doi.org/10.18535/ijrm/v12i05.em18>
- 3) ATTAOUI, M., & GABER, H. (2025). 14 Years of Online Customer Engagement: A Systematic Literature Review Using the SPAR-4-SLR Protocol. *Global Journal of Economics and Business*, 95(294), 586–611. <https://doi.org/10.31559/GJEB2025.15.5.6>
- 4) Attaoui, M., & Gaber, H. (2026). Towards Ethical E-Commerce: The Influence of Perceived Digital Transparency on Customer Relationship Quality and Online Engagement. In *Ethical Marketing and Consumer Trust in Digital and Sustainable Markets* (pp. 169–218). IGI Global. <https://doi.org/10.4018/979-8-3373-9680-4.ch006>
- 5) Attaoui, M., Moukrim, A., Oubayoucef, M., & Gaber, H. (2025). *The perception of virtual assistant transparency: effects on customer citizenship behavior*. 1–8. <https://doi.org/10.1109/LOGISTQUA66323.2025.11122814>
- 6) Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Benetot, 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/10.1016/J.INFFUS.2019.12.012>
- 7) Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study. <https://doi.org/10.1509/JMKG.2005.69.4.133>, 69(4), 133–152. <https://doi.org/10.1509/JMKG.2005.69.4.133>
- 8) Blau, P. M. (1964). Justice in Social Exchange. *Sociological Inquiry*, 34(2), 193–206. <https://doi.org/10.1111/J.1475-682X.1964.TB00583.X>
- 9) Brodie, R. J., & Hollebeek, L. D. (2011). Advancing and consolidating knowledge about customer engagement. *Journal of Service Research*, 14(3), 283–284.
- 10) Brodie, R. J., Hollebeek, L. D., Jurić, B., & Ilić, A. (2011). Customer engagement: Conceptual domain, fundamental propositions, and implications for research. *Journal of Service Research*, 14(3), 252–271. <https://doi.org/10.1177/1094670511411703>
- 11) Cai, M., Gao, K., Ji, Z., & Xu, X. (2026). How Transparency Impacts Trust in Teleoperated Autonomous Robots under Uncertainty. *International Journal of Human-Computer Studies*, 103877. <https://doi.org/10.1016/J.IJHCS.2026.103877>
- 12) Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- 13) Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2019). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science* 2019 48:1, 48(1), 24–42. <https://doi.org/10.1007/S11747-019-00696-0>
- 14) Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems*, 13(3), 319–339. <https://doi.org/10.2307/249008>
- 15) Diakopoulos, N., & Koliska, M. (2017). Algorithmic Transparency in the News Media. *Digital Journalism*, 5(7), 809–828. <https://doi.org/10.1080/21670811.2016.1208053>
- 16) Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), 61–80. <https://doi.org/10.1287/ISRE.1060.0080>
- 17) Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. *Conference on Human Factors in Computing Systems - Proceedings, 2015-April*, 153–162. <https://doi.org/10.1145/2702123.2702556;WGROU:STRING:ACM>
- 18) Gao, L., Li, G., Tsai, F., Gao, C., Zhu, M., & Qu, X. (2023). The impact of artificial intelligence stimuli on customer engagement and value co-creation: the moderating role of customer ability readiness. *Journal of Research in Interactive Marketing*, 17(2), 317–333. <https://doi.org/10.1108/JRIM-10-2021-0260>

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

- 19) Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and tam in online shopping: AN integrated model. *MIS Quarterly: Management Information Systems*, 27(1), 51–90. <https://doi.org/10.2307/30036519>
- 20) Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. *AI Magazine*, 40(2), 44–58. <https://doi.org/10.1609/AIMAG.V40I2.2850>
- 21) Hair, J. F., Hult, G. T. M., & Ringle, C. M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)*.
- 22) Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203/FULL/XML>
- 23) Harrison McKnight, D., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The Journal of Strategic Information Systems*, 11(3–4), 297–323. [https://doi.org/10.1016/S0963-8687\(02\)00020-3](https://doi.org/10.1016/S0963-8687(02)00020-3)
- 24) Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/S11747-014-0403-8>
- 25) Herm, L.-V. (2022). *IMPACT OF EXPLAINABLE AI ON COGNITIVE LOAD: INSIGHTS FROM AN EMPIRICAL STUDY*.
- 26) Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science*, 5, 1096257. <https://doi.org/10.3389/FCOMP.2023.1096257/FULL>
- 27) Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014a). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28(2), 149–165.
- 28) Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014b). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28(2), 149–165. <https://doi.org/10.1016/J.INTMAR.2013.12.002>
- 29) Kastner, L., Langer, M., Lazar, V., Schomacker, A., Speith, T., & Sterz, S. (2021). On the Relation of Trust and Explainability: Why to Engineer for Trustworthiness. *Proceedings of the IEEE International Conference on Requirements Engineering, 2021-September*, 169–175. <https://doi.org/10.1109/REW53955.2021.00031>
- 30) Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: a systematic review. *Frontiers in Artificial Intelligence*, 7, 1456486. <https://doi.org/10.3389/FRAI.2024.1456486/FULL>
- 31) Langer, M., König, C., Langer, M., & König, C. (2022). Explainability of artificial intelligence in human resources. *Handbook of Research on Artificial Intelligence in Human Resource Management*, 285–302. <https://doi.org/10.4337/9781839107535>
- 32) Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, 123. <https://doi.org/10.1016/J.CHB.2021.106878>
- 33) Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society*, 5(1). <https://doi.org/10.1177/2053951718756684;WEBSITE:WEBSITE:SAGE;WGROUPE:STRING:PUBLICATION>
- 34) Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems, 2017-December*, 4766–4775. <https://arxiv.org/pdf/1705.07874>
- 35) Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48), 12714–12719. <https://doi.org/10.1073/PNAS.1710966114>
- 36) Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709. <https://doi.org/10.2307/258792>
- 37) McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. <https://doi.org/10.1287/isre.13.3.334.81>, 13(3), 334–359. <https://doi.org/10.1287/ISRE.13.3.334.81>
- 38) Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. the MIT Press.
- 39) Miller, G. A. (1956). *The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information*[1]. <http://psychclassics.yorku.ca/Miller/>
- 40) Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, 58(3), 20. <https://doi.org/10.2307/1252308>

How Explainable AI Drives Online Customer Engagement: The Mediating Effects of Trust and Transparency Perceptions

- 41) Pavlou, P. A. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*, 7(3), 101–134. <https://doi.org/10.1080/10864415.2003.11044275>
- 42) Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*. <https://doi.org/10.1145/3173574.3173677>;CTYPE:STRING:BOOK
- 43) Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*, 97–101. <https://doi.org/10.18653/v1/n16-3020>
- 44) Shankar, V., Smith, A. K., & Rangaswamy, A. (2003). Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, 20(2), 153–175. [https://doi.org/10.1016/S0167-8116\(03\)00016-8](https://doi.org/10.1016/S0167-8116(03)00016-8)
- 45) 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/10.1016/J.IJHCS.2020.102551>
- 46) Stewart, K. J. (2003). Trust transfer on the World Wide Web. *Organization Science*, 14(1), 5–17. <https://doi.org/10.1287/ORSC.14.1.5.12810>
- 47) van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Http://Dx.Doi.Org/10.1177/1094670510375599*, 13(3), 253–266. <https://doi.org/10.1177/1094670510375599>
- 48) Vargo, S. L., & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68(1), 1–17. <https://doi.org/10.1509/jmkg.68.1.1.24036>
- 49) Venkatesh, V. (2013). *Abstract - A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies*. 2013(5 July 2013). <http://www.vvenkatesh.com/Abstract/14.asp>
- 50) Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly: Management Information Systems*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- 51) Vivek, S. D., Beatty, S. E., Dalela, V., & Morgan, R. M. (2014). A Generalized Multidimensional Scale for Measuring Customer Engagement. *Journal of Marketing Theory and Practice*, 22(4), 401–420. <https://doi.org/10.2753/MTP1069-6679220404>
- 52) Vivek, S. D., Beatty, S. E., & Morgan, R. M. (2012). Customer engagement: Exploring customer relationships beyond purchase. *Journal of Marketing Theory and Practice*, 20(2), 122–146. <https://doi.org/10.2753/MTP1069-6679200201>
- 53) Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.306328>



There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non Commercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.