AI Chatbots as Informatics-Enabled Marketing Service Systems: Trust, Transparency, and Youth User Responses

Hanadi Salhab, Munif Zoubi, Huda Istatia, Laith Khrais

Faculty of Business, Middle East University

Hsalhab@meu.edu.jo

Abstract. This manuscript examines how young consumers respond to AI chatbots in social commerce by conceptualizing chatbots as informatics-enabled front-line service systems. Building on a unified model that assigns Stimulus-Organism-Response (SOR) as the system structure, the Persuasion Knowledge Model (PKM) as the ethical-cognition mechanism, and Trust Theory as the service-outcome logic, we test how two service-design choices-identity disclosure (transparency) and conversational tone (personalized vs. generic)-shape trust and perceived manipulation, and ultimately purchase intention. Using a 2x2 between-subjects experiment with UAE youth (18-25), standardized chatbot dialogues were generated and pretested using a large-language-model workflow to ensure consistent stimuli; this design enables controlled comparison but does not fully capture the adaptivity of live chatbots. PLS-SEM results show that transparent AI disclosure and empathetic personalization increase trust and reduce perceived manipulation; trust is the dominant mediator linking design cues to purchase intention, while perceived manipulation imposes a significant negative effect. Digital literacy attenuates the negative influence of manipulation on intention, highlighting a boundary condition relevant for service governance. The results can also be used to inform guidelines for the development and delivery of service systems, which involve the provision of transparency by default, personalization that is explainable, the adaptation of tone to the needs of the user, and the development of an escalation process.

Keywords: Youth, Chatbots, Transparency, Trust, Literacy, Commerce.

1. Introduction

Chatbots are not just neutral tools for younger generations; they are socio-technical actors embedded in the fabric of everyday digital ecosystems (Al-Ma'aitah et al. 2024; Kumar and Shankar 2025). As the use of chatbots is becoming increasingly mainstream, the question is no longer whether chatbots can be used to persuade but, rather, the informatics of the chatbot's use and the associated consequences on the ethics of the interactions that are deemed trustworthy and ethical.

Past research is fragmented in nature, with the majority of the literature considering the issue of trust as the primary motivator of AI interactions and the issue of manipulation as a parallel construct, albeit a competing and complementary organismic response to trust. (Nguyen et al. 2022; Watson et al. 2024). This creates an unresolved conceptual tension: transparency and personalization can increase trust by reducing uncertainty, yet the same mechanisms can heighten persuasion awareness and moral resistance when users infer strategic influence (Singh et al. 2024). A coherent model is therefore needed to explain when AI-mediated persuasion functions as legitimate service support versus ethically problematic manipulation.

Addressing this tension, this manuscript develops a unified SOR-PKM-Trust Theory framework that positions chatbot cues as service-system stimuli, PKM as the ethical-cognition mechanism through which users detect persuasive intent, and Trust Theory as the service-outcome logic linking perceived competence, integrity, and benevolence to behavioural intention. Empirically, we test the model using UAE youth (18-25) as a theory-testing context where high digital adoption and relational norms make trust and moral evaluation particularly salient (Salhab et al. 2025). The contribution is therefore theoretical integration and contextual validation, with implications for informatics-enabled service design and governance rather than new theory creation.

Accordingly, the manuscript addresses the following research questions:

- RQ1: How do chatbot identity disclosure and conversational tone (as transparency and interaction design choices) influence youth users' trust and perceived manipulation in AI-enabled service encounters?

- RQ2: Do trust and perceived manipulation jointly mediate the effects of identity disclosure and tone on purchase intention in social commerce?

- RQ3: Does digital literacy moderate the relationship between perceived manipulation and purchase intention among youth users?

2. Literature Review

2.1. AI chatbots as informatics-enabled service systems

AI chatbots can be understood as informatics-enabled service systems: they operationalize service delivery through data capture, algorithmic personalization, and conversational decision logic, While factors such as identity disclosure, transparency cues, and explainability impact how legitimate and trustworthy this service is perceived to be, this approach expands the scope of such service beyond just persuasion. It includes reliability, consistency, and failure and recovery in information-intensive services (Blümel et al. 2024; Jarrahi et al. 2025; Hui et al. 2025).

2.2. Generative AI and Youth Digital Behaviour

Generative AI is playing an increasingly important role in influencing how young consumers interact with brands, how they consume content, and how they make decisions. Research has revealed that

Gen Z and late millennials interact with generative AI as a tool that can help accomplish tasks and, at times, as a relational presence in the digital space (Kumar & Shankar 2025). Research has primarily focused on the productivity benefits of generative AI in searching, creating, and consuming content in the Western world (Jarrahi et al. 2025).

Research has also revealed that generative AI has different implications in different cultures, with youth in collectivist societies more likely to see generative AI as a "partner" compared with those in individualist societies (Al-Ma'aitah et al. 2024). However, digital nativism and lower privacy concerns can influence how young consumers interact with generative AI, with fairness, persuasion, and manipulation concerns acting as barriers to engagement with generative AI, which can sometimes trigger resistance (Campbell et al. 2025). Therefore, generative AI can be seen as a utility and an ethical actor in digital commerce, which can create value and ethical issues at the same time (Alshemmari, 2023)

.

2.3. Chatbot Identity and Human–AI Interaction in Social Commerce

Chatbot identity not only acts as a cue for communication but also acts as a mechanism to encourage transparency for informatics-enabled service systems. Moreover, the level to which the chatbot discloses information about AI involvement or presents itself as human-like affects the manner in which data usage is perceived, the manner in which the decision logic is understood, and the manner in which accountability is attributed. Empirical findings have been inconsistent regarding this notion. Some findings suggest that disclosing information about AI involvement increases user trust by reducing uncertainty and aligning user expectations. Other findings suggest that presenting the chatbot as human-like increases warmth but can have the unintended consequence of decreasing user trust if the user perceives deception and hidden persuasion intentions (Blömker and Albrecht 2025; Hui et al. 2025).

This indicates that the manner in which identity affects user perception is subject to cultural factors and digital experience. Users who are more inclined towards relational and human-centered interactions may be more accepting of anthropomorphic service behavior but also expect disclosure for data-intensive profiling for personalization purposes (Al-Ma'aitah et al. 2024; Campbell et al. 2025). This indicates that identity disclosure should be viewed as an informatics design choice for chatbots to enhance user trust and also increase user awareness about persuasion intentions, thereby creating the tension between trust and resistance to inform the central argument for this manuscript (Yaseen & Al-Amarneh, 2025)

.

2.4. Conversational Tone and Perceived Personalisation

Conversational tone operationalizes the chatbot's interaction policy, shaping how personalization is delivered in the service encounter. Empathetic or relational language can increase perceived responsiveness and service quality, but it also interacts with transparency: highly personalized tone without clear disclosure may be interpreted as strategic influence rather than genuine support. In informatics terms, tone is an interface-level manifestation of underlying decision rules and personalization algorithms (Blümel et al. 2024; Blömker and Albrecht 2025).

Warm tone can elevate satisfaction and engagement, yet excessive intimacy or over-personalization can trigger discomfort, privacy concern, and manipulation inference, especially among users with higher persuasion knowledge or stronger privacy norms (Campbell et al. 2025; Watson et al. 2024). This suggests that tone should be treated as a controllable service-design parameter that must be calibrated alongside transparency and explainability.

2.5. Trust Formation in AI-Mediated Marketing

Trust continues to be the foundation upon which technology acceptance is built. However, the determinants of trust evolve when the role of artificial intelligence is central. The earlier model identified competence, integrity, and benevolence as the primary determinants of trust. However, with the role of generative artificial intelligence, new determinants of trust are created. These include transparency, explainability, and perceived control (Luque-Raya et al. 2025). Comparative studies show that trustworthiness is increased by transparency in the West but may decrease comfort when the institutional framework is central. In the case of the United Arab Emirates, institutional trust created by approval and initiatives in digitization and literacy are central in creating trust that aligns with user perceptions (Zoubi et al. 2025).

There is an increasing tendency for researchers to integrate Trust-Commitment Theory and Technology Acceptance Model (TAM) with the objective of addressing the relational and functional dimensions of trust in one model (Almahri and Saleh 2025). However, this model does not consider moral trust, which is central in the case of younger consumers who are exposed to complex and misleading recommendation logics. Therefore, trust needs to be redefined as a multidimensional concept.

2.6. Perceived Manipulation and the Ethics of AI Persuasion

Moreover, the real-time adaptation of messages by AI increases the ethical concern about hidden persuasion. Unlike the assumption of the Persuasion Knowledge Model, which holds that consumers are aware of persuasion intentions and can resist such persuasion, AI makes persuasion intentions unclear by adapting to consumers' responses. Further, comparative studies have also found that younger consumers tend to underestimate AI persuasion compared to older consumer (Watson et al. 2024). A cross-cultural study also indicates that young consumers in Western culture perceive artificial intelligence persuasion as intrusive in terms of privacy, while young consumers in Arab culture perceive artificial intelligence persuasion as acceptable in terms of service and efficiency (Binlibdah 2024).

From an ethical point of view, persuasion is related to perceptual manipulation, where artificial intelligence persuasion leads to a lack of autonomy or a lack of clear consent. Studies related to Nudge Theory and Computational Persuasion state that artificial intelligence persuasion is less likely to influence consumers if its role is transparent, with clear indications of any coercion felt by consumers. However, only a few empirical models exist that consider these factors in a broader context.

2.7. Purchase Intention and Behavioral Outcomes in Social Commerce

In the context of AI marketing, the intention to purchase is a function of the interplay between cognitive trust and the regulation of emotions. The earlier Theories of Planned Behavior and the

Stimulus-Organism-Response model discuss how stimuli induce a reaction, which eventually manifests as a particular behavior (Nguyen et al., 2022). A comparative study across cultures reveals the universal effectiveness of trust as a facilitator of the intention to purchase across cultures, although the mediating effect of manipulation is inconsistent. Western cultures report a negative reaction to manipulation, whereas GCC and Asian cultures report a conditionally positive reaction if the manipulation is aligned with social value or convenience (Cleveland, 2024).

Recent studies on the effects of AI-specific concepts, such as algorithmic transparency and anthropomorphism, on the intention to purchase reveal a mediating effect on emotions, although the link with manipulation is only loosely defined. A comparative synthesis reveals a dual pathway for the intention to purchase in the context of AI marketing. Two routes exist: the ethical route through trust and the psychological route through compliance, with the interaction of these routes, especially among the youth, being the focal point for the extension of the theory.

2.8. Unified theoretical framework, research gap, and hypotheses development

To overcome fragmented thinking, the present manuscript attempts to expand its scope by integrating three basic theories: the Stimulus-Organism-Response (SOR) model, Personal Knowledge Management (PKM), and Trust Theory, into a unified system model of service interaction with the help of informatics. PKM specifies the ethical-cognition mechanism through which users infer persuasive intent and form manipulation judgments, while Trust Theory explains how competence, integrity, and benevolence perceptions translate into service acceptance and continuity (Singh et al. 2024; Watson et al. 2024; Hui et al. 2025).

2.8.1. Stimulus–Organism–Response (SOR) Framework

The SOR framework explains how external stimuli in the digital environment shape internal user states and subsequent behaviour (Mehrabian and Russell 1974). Here, chatbot identity disclosure and conversational tone are treated as controllable service-system stimuli that signal transparency and interaction quality; they are expected to influence two organismic evaluations-trust (service confidence) and perceived manipulation (moral resistance)-which jointly determine purchase intention in social commerce (Singh et al. 2024).

Hypotheses:
H1: Transparent chatbot identity disclosure (AI disclosed vs. human-posed) positively influences trust.
H2: Transparent chatbot identity disclosure negatively influences perceived manipulation.
H3: An empathetic/personalized conversational tone positively influences trust.
H4: An empathetic/personalized conversational tone negatively influences perceived manipulation.

2.8.2. Persuasion Knowledge Model (PKM)

PKM explains how consumers recognize persuasion attempts and evaluate whether influence strategies are legitimate or deceptive (Friestad and Wright 1994). In AI-enabled service encounters, perceived manipulation reflects a user's inference that personalization is designed to steer choices in ways that undermine autonomy, especially when transparency is weak (Watson et al. 2024). PKM therefore clarifies how informatics design cues activate moral resistance in the organism phase of SOR.

Hypotheses:
H5: Perceived manipulation negatively influences purchase intention.

H6: Trust and perceived manipulation jointly mediate the effects of identity disclosure and conversational tone on purchase intention.

2.8.3. Trust Theory

Trust Theory conceptualizes trust as confidence in the competence, integrity, and benevolence of a service provider, and it is central to continued engagement in information-intensive interactions (Hui et al. 2025). In AI chatbot contexts, trust operates as a service-outcome logic: it links perceptions of transparent, reliable, and fair system behaviour to willingness to rely on the chatbot and accept its recommendations in social commerce.

Hypotheses:
H7: Trust positively influences purchase intention in AI-enabled social commerce.
H8: Digital literacy moderates the relationship between perceived manipulation and purchase intention, weakening the negative effect among more digitally literate youth.

The resulting research gap is not only geographic but conceptual: existing literature does not fully explain how transparency and personalization can simultaneously foster trust while triggering persuasion awareness and moral resistance. By testing this tension within a unified informatics-enabled service system model-and using UAE youth as a theory-testing context-this manuscript offers integrated explanation, design-relevant implications, and a clearer contribution.
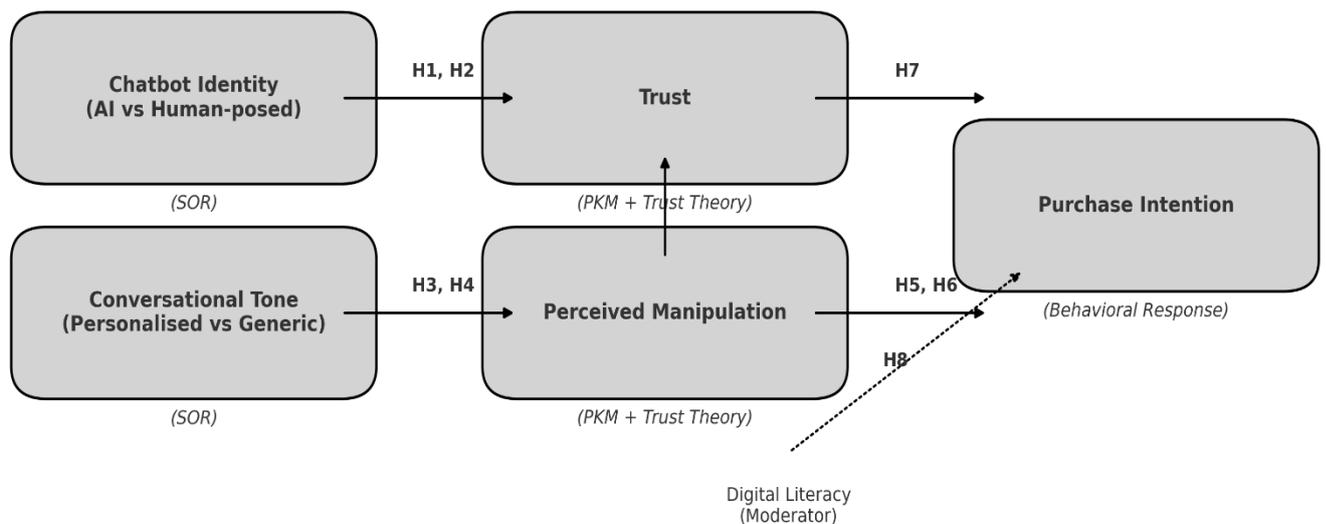

Figure 1. Conceptual framework

3. Methodology

3.1. Research Context and Country Justification

The United Arab Emirates (UAE) provides a relevant theory-testing context because digital services and AI-enabled commerce are widely adopted and youth users are frequent early adopters of conversational technologies. At the same time, the context is used here to examine conceptual tensions (trust versus moral resistance) rather than to claim universal consumer behaviour.

Accordingly, inferences are bounded to digitally active UAE youth (18-25) and similar high-AI-adoption Gulf contexts (Salhab et al. 2025).

3.2. Research Design and Approach

A 2x2 between-subjects experimental survey design was used to test how two service-system design choices influence youth responses: identity disclosure (AI disclosed vs. human-posed) as a transparency cue and conversational tone (personalized/empathetic vs. generic) as an interaction-policy cue. Each participant viewed one standardized chatbot dialogue simulating a social-commerce recommendation scenario, after which they completed measures of trust, perceived manipulation, digital literacy, and purchase intention.

To ensure consistent stimuli, dialogues were generated using a large-language-model workflow and then edited to hold constant product information, length, and structure across conditions. This approach supports internal validity and avoids exposing participants to uncontrolled live-system variation; however, it represents a controlled simulation and does not capture real-time adaptivity, learning, or repeated interaction effects typical of deployed chatbots.

This controlled design isolates the effects of transparency and tone while holding other informatics features constant (e.g., recommendation content, response timing). As a result, the findings should be interpreted as causal evidence about specific design cues rather than as a full ecological account of ongoing human-chatbot service relationships (Taqa, 2025).

3.3. Realism and Manipulation Checks

To assess perceived realism and successful manipulation of the experimental cues, a two-stage verification process was applied. First, a realism pretest (n = 30) evaluated the credibility and naturalness of the dialogues, and items with low realism were revised. Second, manipulation-check items confirmed that participants correctly perceived the intended identity-disclosure and tone conditions. Because the stimuli were standardized rather than adaptive, the experiment is best framed as a controlled simulation; future work should test live chatbots to examine learning effects, service continuity, and failure/recovery dynamics (Jama, 2023).

3.4. Target Population and Sampling

The target population comprised young consumers aged 18-25 residing in the UAE. This population is appropriate for examining youth vulnerability and literacy-based boundary conditions in AI-enabled service encounters, but it does not represent all consumers.

A purposive non-probability sampling strategy recruited participants from universities and early-career workplaces. Findings should therefore be interpreted as applicable to digitally active UAE youth rather than to the broader population. Replication across age groups and cultural contexts is recommended to test robustness and external validity (Oreqat, 2021).

Sample size determination followed Hair et al. (2021), applying both the *10-times rule* and a *G*Power 3.1* analysis to ensure adequate statistical power for Partial Least Squares Structural Equation Modeling (PLS-SEM) (Siregar and Fachrrozi 2023).

3.5. Measurement Instruments

All the latent variables measured by the study employed scales that had been tested for their reliability and had been derived from tested literature on consumer behavior and AI marketing. The study measured the following variables: trust, perceived manipulation, purchase intention, and digital

literacy/AI familiarity, with the latter used as a moderator. The study used a five-point Likert scale for each item, with 1 representing strongly disagree and 5 representing strongly agree (Morshed et al. 2024).

The results showed high reliability with Cronbach's α coefficients above 0.80. The wording of the items was slightly amended to reflect the terminology used in UAE's social commerce (Al-Muntasir, 2022). The reliability and validity criteria used to evaluate the constructs are presented in Table 1. The criteria were used to evaluate the measurement models using PLS-SEM.

Table 1. Measurement Reliability and Validity Thresholds

Assessment Criterion	Purpose	Recommended Threshold	Interpretation / Action	Key References
Cronbach's α	Internal consistency of items within each construct	≥ 0.70	α > 0.80 = good reliability	Hair et al. (2021)
Composite Reliability (CR)	Overall construct reliability	≥ 0.70	Indicates shared variance among indicators	Fornell & Larcker (1981)
Average Variance Extracted (AVE)	Convergent validity	≥ 0.50	Demonstrates sufficient variance captured by indicators	Hair et al. (2021)
Indicator Loadings	Item reliability	≥ 0.70	Retain > 0.70; consider removal < 0.60	Hulland (1999)
Fornell–Larcker Criterion	Discriminant validity	√AVE > inter-construct correlations	Confirms construct distinctiveness	Fornell & Larcker (1981)
HTMT Ratio	Alternative discriminant-validity test	≤ 0.85	HTMT < 0.85 = satisfactory discrimination	Henseler et al. (2015)

3.6. Statistical Method and Analytical Strategy

All statistical procedures were performed using the software tool SmartPLS. Specifically, we relied on version 4.0.9.0. The primary statistical technique we employed is Partial Least Squares Structural Equation Modeling (PLS-SEM). We decided to rely on this technique since we wanted to employ an effective technique that is capable of handling predictive and variance-based models with several latent variables and mediation and moderation effects. Moreover, this technique is very effective when data does not follow a normal distribution and when the sample size is considered medium scale. The use of this technique is consistent with the methodological norms employed by the journal Young Consumers (Morshed 2025b).

3.6.1. Stage 1: Measurement-Model Evaluation
- Internal consistency: Cronbach's α and Composite Reliability (CR) ≥ 0.70.

- Convergent validity: Average Variance Extracted (AVE) ≥ 0.50.

- Discriminant validity: Confirmed using the Fornell–Larcker criterion and Heterotrait–Monotrait ratio (HTMT ≤ 0.85)..

3.6.2. Stage 2: Structural-Model Testing

Furthermore, bootstrapping with 5,000 resamples was used for the estimation of the path coefficients (β), the corresponding t-values, and the 95% confidence intervals for direct, mediating, and moderating relationships within the model. In terms of the quality of the model, several key criteria were used: R^2 for the variance explained, f^2 for the effect size, and Q^2 for the predictive relevance of the model. Moreover, the PLSpredict procedure was used for the evaluation of the predictive performance of the model, thus adding further support for the generalization of the model (Ahmad et al., 2023).

3.6.3. Robustness and Sensitivity Checks

Other model specifications were also explored in relation to their stability, such as removing the moderator and reversing the causal directions. A Multi-Group Analysis (MGA) was also performed to examine differences between participants who were high versus low in digital literacy in terms of conditional effects.(Ngah et al. 2023).

4. Findings

4.1. Respondent Profile and Descriptive Statistics

4.1.1. Demographic Profile

The sample comprised 250 UAE youth consumers who were 18-25 years old. As shown in Table 2, males comprised 54% of the sample, whereas females comprised 46%. The majority of the sample were 22-25 years old, representing 63% of the sample. Moreover, most of the participants were undertaking undergraduate courses, which were 72%. In addition, 58% of the participants were categorized as highly digitally literate. The highly digitally literate category identified the upper literacy groups for multi-group analysis (MGA) (Morshed, 2025c).

Table 2. Demographic Profile of Respondents

Variable	Category	Frequency (n = 250)	Percentage (%)
Gender	Male	135	54.0
	Female	115	46.0
Age (years)	18–21	93	37.2

	22–25	157	62.8
Education level	Undergraduate	180	72.0
	Graduate	70	28.0
Digital literacy level	High	145	58.0
	Low	105	42.0

4.1.2. Descriptive Statistics of Latent Constructs

Descriptive analysis was performed for all the latent constructs that were measured using five-point Likert scales ranging from 1 ("strongly disagree") to 5 ("strongly agree").

The results, as presented in Table 3, revealed that Trust and Chatbot Identity had very high means. The means were 4.33 (SD = 0.66) and 4.28 (SD = 0.63) for Trust and Chatbot Identity, respectively. This revealed that authenticity and trustworthiness were associated with AI chatbots. The results also revealed that Perceived Manipulation had a moderate mean of 2.48 (SD = 0.82). The consistently high means for Digital Literacy (M = 4.25, SD = 0.57) validate its role as a potential moderator.

Table 3. Descriptive Statistics of Latent Constructs

Construct	Items	Mean (M)	SD	Min	Max	Interpretation
Chatbot Identity	4	4.28	0.63	1	5	High perceived authenticity
Conversational Tone	5	4.11	0.70	1	5	Personalised interaction
Trust	4	4.33	0.66	1	5	High trust level
Perceived Manipulation	4	2.48	0.82	1	5	Moderate persuasion awareness
Purchase Intention	4	4.02	0.71	1	5	Positive behavioural tendency
Digital Literacy	3	4.25	0.57	1	5	High technological competence

Overall, the descriptive profile suggests that trust levels are relatively high, while levels of manipulation are moderate. In service and informatic terms, the youth users seem to have the willingness to trust the chatbot service interface (as represented by the high levels of trust) but to be aware of the potential for manipulation and the threat to their autonomy (as represented by the moderate levels of manipulation) (Morshed, 2024b).

4.2. Measurement-Model Evaluation

Construct reliability and validity were assessed with SmartPLS 4.0.9.0, where the two-step PLS-SEM approach was employed as suggested by (Hair Jr. et al., 2021).

The outer loadings of all indicators were above 0.70, which confirmed reliability at the indicator level. Internal consistency reliability was also confirmed as Cronbach's α and Composite Reliability (CR) were above 0.70 for all constructs, while Average Variance Extracted (AVE) was above 0.50, thus showing convergent validity (Table 4).

Discriminant validity, as suggested by the Fornell-Larcker criterion and HTMT method, also confirmed satisfactory results as the square root of AVEs was greater than the correlations between the constructs, and HTMT ratios were below 0.85 (Dos Santos and Cirillo 2023).

Table 4. Construct Reliability and Validity Summary

Construct	Cronbach's α	Composite reliability (CR)	Average variance extracted (AVE)	Highest HTMT ratio	Interpretation
Chatbot Identity	0.876	0.912	0.675	0.781	Reliable and valid construct
Conversational Tone	0.884	0.918	0.690	0.802	High internal consistency
Trust	0.893	0.925	0.705	0.793	Satisfactory convergent validity
Perceived Manipulation	0.851	0.897	0.648	0.817	Acceptable discriminant validity
Purchase Intention	0.904	0.934	0.738	0.768	Strong reliability
Digital Literacy	0.865	0.903	0.702	0.759	Measurement adequate

All constructs met recommended thresholds. Substantively, trust indicators were loaded more saliently and cohesively than were indicators of manipulation. This finding supports the view that trust is the primary way in which service encounters with AI are cognitively organized by youth, with manipulation playing a secondary but still important role in moral judgments. Again, this points to the need to consider measurement results in terms of procedure and user representations of service systems and their governance (Morshed, 2025d).

4.3. Structural-Model Results

The second step of the PLS-SEM procedure involved testing the hypothesized structural relationships using the software tool SmartPLS 4.0.9.0 with 5,000 bootstrap resamples (Hair et al., 2021). Since collinearity was not an issue with VIF < 3.3.

The model achieved substantial variance explained for the endogenous constructs: $R^2 = 0.58$ for Trust, 0.46 for Perceived Manipulation, and 0.72 for Purchase Intention (Kalnins and Praitis Hill 2025).

As shown in Table 5, identity disclosure (transparency cue) and conversational tone (interaction-policy cue) were found to positively influence trust: $\beta = 0.52$, $p < 0.001$ and $\beta = 0.47$, $p < 0.001$, respectively. Moreover, the two cues were also found to decrease perceived manipulation: $\beta = -0.31$, $p = 0.004$ for identity disclosure and $\beta = -0.29$, $p = 0.009$ for conversational tone.

Table 5. Structural Path Results (Direct and Mediating Effects)

Hypothesis	Structural path	B	t-value	p-value	Supported?
H1	Chatbot Identity → Trust	0.52	9.14	< 0.001	Yes
H2	Chatbot Identity → Perceived Manipulation	−0.31	2.88	0.004	Yes
H3	Conversational Tone → Trust	0.47	8.03	< 0.001	Yes
H4	Conversational Tone → Perceived Manipulation	−0.29	2.62	0.009	Yes
H5	Perceived Manipulation → Purchase Intention	−0.33	2.53	0.012	Yes
H7	Trust → Purchase Intention	0.41	3.01	0.003	Yes

Effect-size results (f^2) showed large impacts of *Chatbot Identity* (0.29) and *Conversational Tone* (0.27) on *Trust* and medium effects of *Perceived Manipulation* (0.18) on *Purchase Intention*. Positive Q^2 values for all endogenous constructs confirmed predictive relevance.

Overall, the findings support the SOR framework: *AI identity* and *conversational warmth* act as persuasive stimuli, while *Trust* and *Perceived Manipulation* function as cognitive-affective mechanisms shaping consumer intention (Siregar and Fachrurrozi 2023).

4.4. Mediation Analysis

Mediation (H6) was examined using bootstrapped indirect effect analyses with 5,000 resamples. Significance of the indirect effect was inferred if the CI did not contain zero ($p < 0.05$) (Hair et al., 2021). As shown in Table 6, the trust-based indirect effects are significant and strong: Chatbot Identity → Trust → Purchase Intention ($\beta = 0.21$, 95% CI [0.14, 0.29], $p = 0.002$), and Conversational Tone → Trust → Purchase Intention ($\beta = 0.19$, 95% CI [0.11, 0.27], $p = 0.007$). The manipulation-based mediation is weaker and negative: Chatbot Identity → Manipulation → Purchase Intention is significant and negative ($\beta = −0.08$, $p = 0.031$), while the effect of Conversational Tone on Manipulation and then on Purchase Intention is not significant ($p = 0.087$) (Morshed, 2024a).

Table 6. Indirect (Mediating) Effects

Mediation path	Indirect β	95 % CI (BCa)	p-value	Type	Result
Chatbot Identity → Trust → Purchase Intention	0.21	[0.14, 0.29]	0.002	Partial	Supported
Conversational Tone → Trust → Purchase Intention	0.19	[0.11, 0.27]	0.007	Partial	Supported
Chatbot Identity → Perceived Manipulation → Purchase Intention	–0.08	[–0.16, –0.02]	0.031	Partial (negative)	Supported
Conversational Tone → Perceived Manipulation → Purchase Intention	–0.06	[–0.14, 0.01]	0.087	None	Not supported

This also confirms the primary role of Trust as the mediator between chatbot cues and behavioral intention, as proposed by the SOR model. While the role of Perceived Manipulation, derived from the Persuasion Knowledge Model, is less significant but inverse, the two models validate the mediation process of the dual-path model.

4.5. Moderation and Multi-Group Analysis (MGA)

Digital literacy (H8) was examined as a moderator using the interaction term approach with 5,000 Bootstrap resamples and validated using MGA with the Henseler approach. As shown in Table 7, the result revealed a statistically significant positive interaction effect between the two variables, with $\beta = 0.17$ and $p = 0.018$, indicating that higher levels of digital literacy buffer the effect of the negative relationship between manipulation and purchase intention. As shown in the simple slope plot in Figure 2, the effect of manipulation on purchase intention is strong at lower levels of digital literacy but negligible at higher levels of digital literacy (Wang et al., 2022). MGA also validated the moderating effect of digital literacy, as the relationship between manipulation and purchase intention is significantly different between the high and low literacy groups ($\Delta\beta = 0.21$, $p < 0.05$).

Table 7. Moderation and Multi-Group Analysis Results

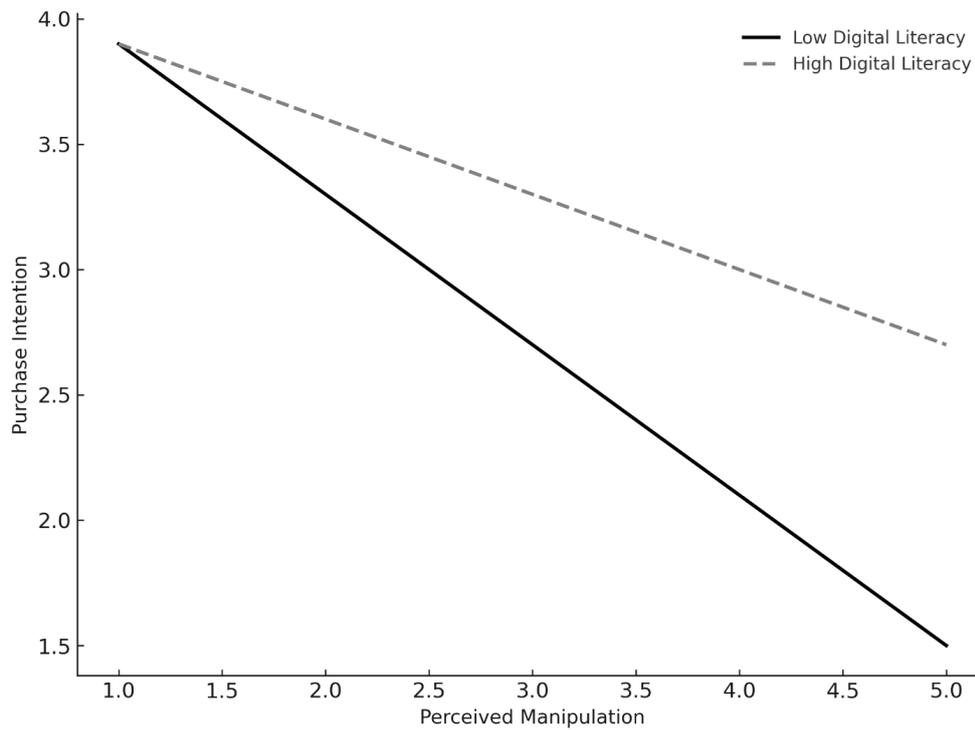| Analysis	Path Tested	β / Δβ	t-value	p-value	Interpretation
Moderation	Manipulation × Digital Literacy → Purchase Intention	0.17	2.38	0.018	High literacy mitigates negative effect
MGA (High vs Low Literacy)	Manipulation → Purchase Intention	Δβ = 0.21	—	< 0.05	Significant group difference

Figure 2. Moderation Effect of Digital Literacy

4.6. Model Fit and Predictive Relevance

The model's adequacy was also checked by using R^2, f^2, Q^2, and PLSpredict, as suggested by Hair Jr. et al. (2021). As presented in Table 8, the results indicate good explanatory and predictive capabilities. The R^2 values indicate moderate-to-substantial variance explained by the model in Trust (0.58), Perceived Manipulation (0.46), and Purchase Intention (0.72). The results of f^2 indicate large effects of Chatbot Identity on Trust (0.29), as well as large effects of Conversational Tone on Trust (0.27). Moreover, f^2 results indicate moderate negative effects of Chatbot Identity on Perceived Manipulation (0.18, 0.16). Regarding Purchase Intention, Trust (0.22) and Manipulation (0.17) have medium effects (Morshed et al., 2024).

Table 8. Model Quality and Predictive Indices

Construct	R^2	f^2 (Key Predictors)	Q^2	Predictive Relevance	Interpretation
Trust	0.58	0.29 (Identity), 0.27 (Tone)	0.36	Yes	Strong explanatory power

Perceived Manipulation	0.46	0.18 (Identity), 0.16 (Tone)	0.28	Yes	Moderate predictive strength
Purchase Intention	0.72	0.22 (Trust), 0.17 (Manipulation)	0.43	Yes	High predictive relevance

4.7. Summary of Hypothesis Testing

All eight hypotheses (H1–H8) were evaluated through the integrated PLS-SEM model using bootstrapping and multi-group procedures. As summarised in Table 9, all proposed relationships were statistically significant and directionally consistent with theoretical expectations.

The stimulus variables-identity disclosure and conversational tone-positively influenced trust and reduced perceived manipulation, supporting H1-H4. In turn, trust increased purchase intention (supporting H7), while perceived manipulation reduced purchase intention (supporting H5). Trust and manipulation jointly mediated the effects of the design cues on intention (supporting H6), and digital literacy weakened the negative effect of manipulation on intention (supporting H8).

Table 9. Summary of Hypotheses Testing

Hypothesis	Statement	Path / Test	Result	Supported?
H1	Chatbot Identity → Trust	Direct effect	$\beta = 0.52$, $p < 0.001$	✔ Supported
H2	Chatbot Identity → Perceived Manipulation	Direct effect	$\beta = -0.31$, $p = 0.004$	✔ Supported
H3	Conversational Tone → Trust	Direct effect	$\beta = 0.47$, $p < 0.001$	✔ Supported
H4	Conversational Tone → Perceived Manipulation	Direct effect	$\beta = -0.29$, $p = 0.009$	✔ Supported
H5	Perceived Manipulation → Purchase Intention	Direct effect	$\beta = -0.33$, $p = 0.012$	✔ Supported
H6	Trust & Manipulation as mediators	Indirect effects	$\beta = 0.21 / -0.08$, $p < 0.05$	✔ Supported
H7	Trust → Purchase Intention	Direct effect	$\beta = 0.41$, $p = 0.003$	✔ Supported
H8	Digital Literacy moderates & differentiates Manipulation → Intention	Interaction + MGA	$\beta = 0.17$, $\Delta\beta = 0.21$, $p < 0.05$	✔ Supported

Collectively, these results validate the integrated SOR–PKM–Trust Theory model, demonstrating that chatbot identity and tone influence consumer responses both cognitively (*Trust*) and ethically (*Manipulation*). The moderating role of Digital Literacy reinforces its importance as a boundary condition in AI-driven persuasion dynamics.

5. Discussion and Implications

5.1. Positioning and theoretical implications

The interaction between youth and AI chatbots is influenced by the service system design. Digital literacy further helps to refine the boundary conditions. More digitally literate youth were less likely to translate the awareness of manipulation into behavioral avoidance, suggesting that literacy can help support more sophisticated processing of persuasive intent and possibly provide a buffer for moral resistance.

The contribution of the manuscript should therefore be humbly set: It brings together established theory into a coherent system model, presents empirical results on how specific transparency and interaction policy factors influence trust and manipulation, and tests the underlying mechanisms in the Gulf youth culture. It can therefore extend existing research by emphasizing the co-occurrence of trust and moral resistance, but not by suggesting the development of new theory.

A service science perspective also points to further implications of the results, which go beyond the specific purchase intention of the experiment. In the real-world application of deployed chatbots, the service relationship is continuous, and success, failure, and recovery can all influence long-run service acceptance. Transparency protocols (e.g., persistence of disclosures) and escalation protocols (e.g., handoff to humans when uncertainty or discomfort is detected) become important for reliable and ethical service delivery.

The results support the view of the design and delivery of conversational AI: trust can be fostered not only by success but also by transparent system identity, personalization, and interaction policy.

5.2. Practical Implications

Managerially, what this study's findings suggest is that trustworthy chatbot deployment necessitates specific service system configuration, as opposed to vague recommendations of 'transparency' per se. What this means is that firms need to develop (1) disclosure protocols (persistent AI identifications and purposes), (2) explainable personalization (so users know why they are getting a recommendation and what it's based on), (3) tone calibration (so chatbots avoid too much intimacy or pressure), and (4) escalation and recovery (so users can go to humans or file complaints if they sense manipulation).

5.3. Policy implications

Policy implications of the governance nature of identity disclosure and personalization follow naturally. Regulators and platform operators can improve the overall protection of youth by mandating the disclosure of chatbots in social commerce, specifying the level of transparency in personalization (for example, by specifying the level of explanation of recommendation algorithms), and mandating the availability of reporting and redress systems in cases of user coercion and deception. At the same time, the policy of digital literacy should be seen as an enabling infrastructure for trustworthy AI services, such that certain segments of the population do not suffer systematically in service interactions with AI systems due to their lower levels of literacy in persuasion, privacy, and informed consent.

5.4. Societal and Ethical Implications

Ethical contribution should therefore be operationalized into design and governance principles: transparency-by-default (clear, persistent disclosure and purpose statements), explainable interaction (user-appropriate explanations of recommendation logic and data use), and literacy-adaptive service design (additional prompts, warnings, or opt-outs for lower-literacy users). These principles translate moral concerns into implementable controls that reduce vulnerability while sustaining trustworthy service delivery.

5.5. Future Research Directions

Future research should test these mechanisms using live or adaptive chatbots and behavioural outcomes to capture learning effects, repeated interaction, and service continuity. Longitudinal or field designs can examine how trust and manipulation evolve after service failures and recovery efforts (e.g., human handoff, apology, or correction). Replication across age groups and cultures is also needed to assess whether the literacy boundary condition holds for less digitally experienced users and in different regulatory environments.

6. Conclusion

This manuscript examined youth responses to AI chatbots in social commerce by treating identity disclosure and conversational tone as informatics-enabled service design choices. The results demonstrate the effectiveness of making AI transparency a default condition along with a thoughtfully calibrated and empathetic tone for building trust and reducing the experience of manipulation. The results also demonstrate the mediating effect of the two organismic responses on the intention to make a purchase. Finally, the results demonstrate the role of digital literacy as a means of mitigating the behavioral effects of the experience of manipulation, a boundary condition with significant implications for the governance of service for youth.

The study's contribution is integrative and significant for the design of services. The study provides a model for a unified theory of the conditions under which SOR (system structure), PKM (ethical cognition mechanism), and Trust Theory (service outcome logic) can be complementary. The study's integrative contribution is significant because it provides a model for the conditions under which trust and moral counter-resistance can be complementary in the context of AI-mediated service encounters.

From a methodological standpoint, the study's use of standardized dialogues created by a language model provides high internal validity. The study's results could be more complete with the addition of a study using a more realistic chatbot interface. Future studies should extend the study's results with a study using a more realistic chatbot interface.

References

Ahmad, A. K., Nahar, H. M., & Manajreh, M. M. N. (2023). Effect of social media on shaping the agenda of the communicator in the Jordanian TV channels. *Middle East Journal of Communication Studies, 3*(2), Article 3. https://doi.org/10.71220/2585-003-002-003.

Al-Ma'aitah, N., Liao, Y., Soltani, E., *et al.* (2024). Unpacking the cultural impacts on long-term manufacturer–supplier relationships in the Arab Middle East region: The moderating role of trust. *Production Planning & Control, 35*, 2304–2326. https://doi.org/10.1080/09537287.2024.2313516.

Al-Muntasir, M. (2022). The phenomenon of information flow from traditional and new media about the Corona pandemic from the perspective of newly graduated media professionals in Yemen. *Middle East Journal of Communication Studies, 2*(2), Article 1. https://doi.org/10.71220/2585-002-002-005.

Almahri, F. A. A. J., & Saleh, N. I. M. (2024). Insights into technology acceptance: A concise review of key theories and models. In *Innovation and intelligent digital technologies for increased efficiency* (Vol. 2, pp. 797–807). Springer. https://doi.org/10.1007/978-3-031-71649-2_67.

Alshemmari, J. (2023). An empirical study on employee empowerment role in increasing efficiency of employee performance. *Journal of Logistics, Informatics and Service Science, 10*(1), 52–71. https://doi.org/10.33168/LISS.2023.0104.

Binlibdah, S. (2024). Investigating the role of artificial intelligence to measure consumer efficiency: The use of strategic communication and personalized media content. *Journalism and Media, 5*(3), 1142–1161. https://doi.org/10.3390/journalmedia5030073.

Blömker, J., & Albrecht, C.-M. (2025). Reevaluating personalization in AI-powered service chatbots: A study on identity matching via few-shot learning. *Computers in Human Behavior: Artificial Humans, 3*, 100126. https://doi.org/10.1016/j.chbah.2025.100126.

Blümel, J. H., Zaki, M., & Bohné, T. (2024). Personal touch in digital customer service: A conceptual framework of relational personalization for conversational AI. *Journal of Service Theory and Practice, 34*(1), 33–65. https://doi.org/10.1108/JSTP-03-2023-0098.

Campbell, M., Barthwal, A., Joshi, S., & Shrestha, A. (2025). Investigation of the privacy concerns in AI systems for young digital citizens: A comparative stakeholder analysis. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 30–37). IEEE. https://doi.org/10.1109/CCWC62904.2025.10903925.

Cleveland, M. (2024). Within and between two worlds: Conceiving, measuring, and applying mixed-ethnic identity in three countries. *Journal of International Marketing, 32*(2), 65–88. https://doi.org/10.1177/1069031X231212859.

Dos Santos, P. M., & Cirillo, M. Â. (2023). Construction of the average variance extracted index for construct validation in structural equation models with adaptive regressions. *Communications in Statistics—Simulation and Computation, 52*(4), 1639–1650. https://doi.org/10.1080/03610918.2021.1888122.

Griffin, J., Elphinstone, B., Sands, S., & Pallant, J. (2025). Bridging the gap: Aligning academic and practitioner approaches to brand trust measurement. *Australasian Marketing Journal, 33*(4), 430–440. https://doi.org/10.1177/14413582251331608.

Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook.* Springer. https://doi.org/10.1007/978-3-030-80519-7.

Hui, Z., Khan, A. N., & Khan, N. A. (2025). Unveiling the influence of dialogic communication on AI-technology trust in e-commerce. *Aslib Journal of Information Management*. Advance online publication. https://doi.org/10.1108/AJIM-05-2024-0420.

Isaac, M. S., & Calder, B. J. (2025). Thirty years of persuasion knowledge research: From demonstrating effects to building theory to increasing applicability. *Consumer Psychology Review, 8*(1), 3–14. https://doi.org/10.1002/arcp.1107.

Jama, L. A., & Mohamud, N. M. (2024). The impact of procurement practices on organizational performance: A literature review. *Journal of Logistics, Informatics and Service Science, 11*(1), 119–135. https://doi.org/10.33168/JLISS.2024.0108.

Jarrahi, M. H., Li, L., Robinson, A. P., & Meng, S. (2025). Generative AI and the augmentation of information practices in knowledge work. *Behaviour & Information Technology*. Advance online publication. https://doi.org/10.1080/0144929X.2025.2551570.

Jia, Y., Xiao, Y., Chen, H., *et al.* (2025). Effects of group communication norms on daily steps in a team-based financial incentive mobile phone intervention in Shanghai, China. *International Journal of Behavioral Nutrition and Physical Activity, 22*(1), Article 9. https://doi.org/10.1186/s12966-025-01707-w.

Kalnins, A., & Praitis Hill, K. (2025). The VIF score: What is it good for? Absolutely nothing. *Organizational Research Methods, 28*(1), 58–75. https://doi.org/10.1177/10944281231216381.

Kbaier, E., Salminen, J., & Jansen, B. J. (2025). A glimpse into the AI: Exploring the effect of consumers' chatbot interaction experience on brand attachment and advocacy with the moderating role of AI skepticism. *Asia-Pacific Journal of Business Administration.* Advance online publication. https://doi.org/10.1108/APJBA-01-2025-0003.

Kumar, A., & Shankar, A. (2025). Role of generative AI-enabled customer relationship management solutions in achieving agility. *Journal of Business & Industrial Marketing, 40*(8), 1593–1614. https://doi.org/10.1108/JBIM-06-2024-0433.

Li, T., Zhang, C., & Chang, Y. (2025). The impact of prior chatbot identity disclosure on customer tolerance of service failures. *Internet Research.* Advance online publication. https://doi.org/10.1108/INTR-11-2023-1073.

Lin, H., Zhang, Q., Tian, J., & Liu, N. (2025). Green sparks: Do AI robot or human employees better ignite customer pro-environmental behavioral intention? *Asia Pacific Journal of Marketing and Logistics*. Advance online publication. https://doi.org/10.1108/APJML-02-2025-0207.

Luque-Raya, I., Del Barrio-García, S., & Cordon Garcia, O. (2025). The social dimension of consumer–AI interaction: Alienation, anthropomorphism, and identity in AI-mediated marketing. *Behaviour & Information Technology*. Advance online publication. https://doi.org/10.1080/0144929X.2025.2545313.

Morshed, A. (2024a). Comparative analysis of accounting standards in the Islamic banking industry: A focus on financial leasing. *Journal of Islamic Accounting and Business Research.* https://www.emerald.com/insight/content/doi/10.1108/JIABR-12-2022-0349/full/html

Morshed, A. (2024b). Mathematical analysis of working capital management in MENA SMEs: Panel data insights. *Applied Mathematics & Information Sciences, 18*, 111–124.

Morshed, A. (2025a). Navigating tradition and modernity: Digital accounting and financial integration in family-owned enterprises in the Arab Gulf. *Sustainable Futures*, 100680.

Morshed, A. (2025b). Cultural norms and ethical challenges in MENA accounting: The role of leadership and organizational climate. *International Journal of Ethics Systems, 41*(3), 630–656. https://doi.org/10.1108/IJOES-08-2024-0247.

Morshed, A. (2025c). Ethical challenges in designing sustainable business models for responsible consumption and production: Case studies from Jordan. *Management & Sustainability: An Arab Review*. Advance online publication. https://doi.org/10.1108/MSAR-09-2024-0131.

Morshed, A. (2025d). Sustainable energy revolution: Green finance as the key to the Arab Gulf States' future. *International Journal of Energy Sector Management*. Advance online publication. https://doi.org/10.1108/IJESM-10-2024-0007.

Morshed, A., Ramadan, A., Maali, B., *et al.* (2024). Transforming accounting practices: The impact and challenges of business intelligence integration in invoice processing. *Journal of Infrastructure, Policy and Development, 8*(6), 4241. https://doi.org/10.24294/jipd.v8i6.4241.

Morshed, A., Maali, B., Ramadan, A., Ashal, N., Zoubi, M., & Allahham, M. (2024). The impact of supply chain finance on financial sustainability in Jordanian SMEs. *Uncertain Supply Chain Management, 12*(4), 2767–2776.

Ngah, A. H., Kamalrulzaman, N. I., Mohamad, M. F. H., *et al.* (2023). Do science and social science differ? Multi-group analysis (MGA) of the willingness to continue online learning. *Quality & Quantity, 57*, 2957–2980. https://doi.org/10.1007/s11135-022-01465-y. *(Not web-verified in this check)*

Nguyen, Y. T. H., Tapanainen, T., & Nguyen, H. T. T. (2022). Reputation and its consequences in FinTech services: The case of mobile banking. *International Journal of Bank Marketing, 40*(7), 1364–1397. https://doi.org/10.1108/IJBM-08-2021-0371.

Oreqat, A. (2021). The degree of satisfaction of Facebook users about its features, usage motives and achieved gratifications: An applied study on students of the Faculty of Mass Communication at the Middle East University to attract attention. *Middle East Journal of Communication Studies, 1*(1), Article 1. https://doi.org/10.71220/2585-001-001-001.

Qiu, X., Wang, Y., Zeng, Y., & Cong, R. (2025). Artificial intelligence disclosure in cause-related marketing: A persuasion knowledge perspective. *Journal of Theoretical and Applied Electronic Commerce Research, 20*(3), 193. https://doi.org/10.3390/jtaer20030193.

Salhab, H., Zoubi, M., Khrais, L. T., *et al.* (2025). AI-driven sustainable marketing in Gulf Cooperation Council retail: Advancing SDGs through smart channels. *Administrative Sciences, 15*(1), 20. https://doi.org/10.3390/admsci15010020.

Seo, J. K., & Yoon, H. J. (2025). Promoting mindful consumption through a chatbot with an experiential mind. *Journal of Consumer Marketing, 42*(4), 498–511. https://doi.org/10.1108/JCM-05-2024-6844.

Singh, V., Vishvakarma, N. K., & Kumar, V. (2024). Unveiling digital manipulation and persuasion in e-commerce: A systematic literature review of dark patterns and digital nudging. *Journal of Internet Commerce, 23*(2), 144–171. https://doi.org/10.1080/15332861.2023.2278495.

Siregar, K. R., & Fachrurrozi, I. (2023). The effect of service quality on customer complaints and customer loyalty through customer satisfaction of ZALORA Indonesia e-commerce website users. In *2023 International Conference on Digital Business and Technology Management (ICONDBTM)* (pp. 57–62). IEEE.

Taqa, S. B. A. (2025). The mediating role of remote communication on the relationship between electronic human resource management practices and organizational performance in Iraqi commercial banks. *Middle East Journal of Communication Studies, 5*(1), Article 2. https://doi.org/10.71220/2585-005-001-001.

Wang, X., Ashraf, A. R., Thongpapanl, N., & Wang, K.-Y. (2022). Perceived deception and online repurchase intention: The moderating effect of product type and consumer regulatory orientation. *Journal of Consumer Behaviour, 21*(6), 1522–1539. https://doi.org/10.1002/cb.2109.

Wang, Y., Sauka, K., & Situmeang, F. B. I. (2025). Anthropomorphism and transparency interplay on consumer behaviour in generative AI-driven marketing communication. *Journal of Consumer Marketing, 42*(4), 512–536. https://doi.org/10.1108/JCM-04-2024-6806.

Watson, J., Valsesia, F., & Segal, S. (2024). Assessing AI receptivity through a persuasion knowledge lens. *Current Opinion in Psychology, 58*, 101834. https://doi.org/10.1016/j.copsyc.2024.101834.

Yaseen, H., & Al-Amarneh, A. (2025). Adoption of artificial intelligence-driven fraud detection in banking: The role of trust, transparency, and fairness perception in financial institutions in the United Arab Emirates and Qatar. *Journal of Risk and Financial Management, 18*(4), 217. https://doi.org/10.3390/jrfm18040217.

Zhang, Q., Pan, X., & Fan, J. (2025). Uncovering the mediating effects of AI trust and self-efficacy on engagement in human–GenAI communication for EFL learning. *European Journal of Education, 60*, e70336. https://doi.org/10.1111/ejed.70336.

Zoubi, M., Estaitia, H., Morshed, A., *et al.* (2025). Augmented reality and sustainable luxury: Transforming fashion retail in the UAE. *Technological Sustainability, 4*(3), 275–292. https://doi.org/10.1108/TECHS-03-2025-0068.