

ARTIFICIAL INTELLIGENCE AND CONSUMER TRUST

İbrahim Halil Efendioğlu*

*Gaziantep University, Faculty of Economics and Administrative Sciences,
Department of Business Administration, Gaziantep, Türkiye.
ORCID Code: 0000-0002-4968-375X

ABSTRACT

This study conducts a systematic literature review (SLR) to synthesize and critically evaluate research on artificial intelligence (AI) and consumer trust in marketing contexts. Following the SPAR-4-SLR protocol and PRISMA 2020 guidelines, we systematically identified and analyzed 17 SSCI/SCI-E articles published between 2021 and 2025. Eligibility criteria covered peer-reviewed empirical and conceptual studies centered on AI-enabled consumer interactions. Data extraction used a structured codebook, and the synthesis applied the TCCM framework (Theory–Context–Characteristics–Method). Given heterogeneity in measures and designs, we used effect-direction vote counting as a descriptive approach rather than a meta-analysis. Five themes emerged: (1) privacy–personalization trade-offs, (2) the ambivalent effects of anthropomorphism, (3) transparency/explainability, (4) fairness and bias in recommenders, and (5) ethical/organizational integration. Anthropomorphic chatbots and GenAI can raise perceived competence and engagement, yet also heighten privacy concerns; however, transparency mechanisms (e.g., algorithmic disclosure, explanations of “why this recommendation?”) reliably strengthen trust. Trust mediates outcomes including adoption, purchase intention, and data disclosure. To our knowledge, this review offers one of the most up-to-date mappings of AI consumer trust in marketing. It outlines a forward agenda on long-term trust dynamics, cross-cultural contexts, and measurement development.

Keywords: Chatbot, Generative AI, Personalization, Privacy Concerns, Anthropomorphism, Transparency, Systematic Literature Review

1. INTRODUCTION

The rapid adoption of artificial intelligence (AI)-based interfaces, such as chatbots, recommender systems, and generative AI, into marketing applications has made the question of how trust is built at consumer touchpoints a strategic priority. In online consumer decision-making, trust is a multidimensional construct shaped by ability, integrity/transparency, and benevolence (Mayer et al., 1995). Within the privacy calculus framework, it is susceptible to the personalization–privacy trade-off (Culnan & Bies, 2003; Dinev & Hart, 2006). As AI becomes more visible in decision-making and recommendation processes, the determinants of trust have shifted beyond classical notions of

web/intermediary trust (Gefen, 2000; Pavlou, 2003) to include AI-specific dimensions such as algorithmic transparency, explainability (XAI), fairness, and anthropomorphism.

Recent empirical evidence suggests that these new dimensions have ambivalent effects. For example, anthropomorphic or social presence cues may enhance perceived competence and closeness but simultaneously trigger privacy concerns through perceived “agent autonomy” and fears of data misuse (Schanke et al., 2021; Song et al., 2024; Kim et al., 2024). In personalized advertising and communication, this tension becomes more pronounced when combined with regulatory focus differences (promotion vs. prevention), producing divergent effects on trust and persuasion (Kim et al., 2023). In recommender systems, “show and tell” explanations reveal that for experience goods, content-based matches are perceived as more persuasive through transparency cues, whereas for search goods, collaborative filtering triggers the bandwagon effect (Liao & Sundar, 2022; see also the Heuristic–Systematic Model). In chatbot contexts, factors such as identity/algorithm disclosure, as well as task complexity, significantly constrain and shape trust relationships (Cheng et al., 2022).

With the advent of generative AI (GenAI), the picture has become even more complex. In retail contexts, GenAI integration enhances familiarity and perceptions of human likeness, thereby strengthening adoption intentions; however, it does not automatically elevate trust and may even heighten privacy concerns (Arce-Urriza et al., 2025). From a relationship science perspective, companion chatbots may create feelings of closeness. However, they can only partially substitute for the deep relational functions and long-term trust dynamics that are essential for genuine connections (Smith et al., 2025). At the ethical and organizational levels, design and governance principles that support security and privacy protection influence employees’ ethical perceptions of AI and their recommendation intentions (Wang et al., 2025). In market communication, firms that provide concrete and actionable AI disclosures generate credibility signals, while speculative or irrelevant narratives fail to create value (Basnet et al., 2025). In recommender systems, fairness dimensions such as gender equity emerge as critical boundary conditions for sustaining trust (Zhang et al., 2025). Neurophysiological comparisons further reveal that humans elicit higher trust than chatbots in specific tasks, particularly those that are subjective (Wang et al., 2023). Emotional and perceptual profiles also diverge between human–human and human–bot interactions: conversational concerns are lower with chatbots, but perceived similarity and liking are higher in human pairings (Drouin et al., 2022). Managerial interviews highlight that in creative processes, GenAI takes over routine tasks while leaving strategic and emotional layers to humans, with transparency and ethical norms acting as key determinants of reputation and trust (Demsar et al., 2025). Within the consumer experience framework, dimensions such as information, entertainment, social presence, and privacy risk influence purchase intention through experiential mechanisms, with user expertise serving as a moderating factor (Puertas et al., 2024). In the education domain, explainable and human-intuitive recommender designs enhance agency and autonomy, contributing to the trust architecture (Bulathwela et al., 2022).

This body of evidence suggests that marketing scholarship has accumulated fragmented and context-sensitive insights on AI and consumer trust. However, findings remain scattered due to methodological, contextual, and measurement heterogeneity. In particular, there is a pressing need for integrative synthesis on (i) the mechanisms through which trust is shaped in the GenAI era (transparency/explainability, fairness, anthropomorphism, human-in-the-loop control), (ii) the role of task type and complexity as boundary conditions, (iii) how cultural/ethical contexts and regulatory focus recalibrate risk–benefit assessments, and (iv) how long-term relational and trust repair processes unfold.

This study systematically reviews 17 SSCI/SCI-E indexed articles published between 2021 and 2025 in the Web of Science Core Collection. Our contribution is threefold:

- *Mapping and integration:* We consolidate the determinants of trust in chatbot, recommender system, and GenAI applications along the axes of personalization–privacy, anthropomorphism–autonomy, and transparency/fairness–persuasion (Schanke et al., 2021; Liao & Sundar, 2022; Cheng et al., 2022; Song et al., 2024; Kim et al., 2023, 2024; Zhang et al., 2025; Arce-Urriza et al., 2025).
- *Boundary conditions and mechanisms:* We identify moderators such as task complexity, regulatory focus, privacy concern, culture, and cognitive need, as well as mediators including competence, fairness, transparency, and experience, to provide an explanatory framework (Cheng et al., 2022; Kim et al., 2023; Liao & Sundar, 2022; Wang et al., 2023; Zhang et al., 2025).
- *Managerial and policy implications:* We develop evidence-based recommendations emphasizing algorithmic disclosure, data minimization, fair design, calibrated anthropomorphism, and human-in-the-loop principles (Wang et al., 2025; Demsar et al., 2025; Basnet et al., 2025).

Accordingly, the research questions guiding this review are:

- RQ1: What antecedents and mechanisms determine consumer trust in AI-based interfaces (chatbots/recommenders/GenAI) within marketing contexts, and how can this understanding inform AI system design?
- RQ2: How do boundary conditions (task type/complexity, disclosure/transparency, fairness, anthropomorphism, privacy concern) shape trust outcomes (adoption, purchase intention, loyalty)?
- RQ3: How does the rise of GenAI redefine the **long-term** dynamics of consumer trust–AI interactions (closeness, repair, sustainability)?

The remainder of this article is structured as follows: Section 2 provides a detailed review of the conceptual and empirical literature. Section 3 describes the protocol, search and selection criteria, and quality appraisal. Section 4 synthesizes the findings using the TCCM framework, the Theory–Context–Characteristics–Method (TCCM) framework used to systematize constructs, settings, variables, and designs, and effect-direction analysis, a descriptive approach to summarize the direction

of reported associations across heterogeneous designs. Section 5 presents the discussion and integrative model. Section 6 outlines the overall conclusions, theoretical and practical contributions, limitations, and avenues for future research.

2. LITERATURE REVIEW

2.1. Early Approaches to Artificial Intelligence and Consumer Trust

The potential of AI enabled systems to foster trust in consumer interactions has long been a fascinating area of scholarly research. Early studies, particularly in online settings, drew on privacy calculus theory and risk–benefit models to explain consumer trust (Dinev & Hart, 2006; Culnan & Bies, 2003). From this perspective, consumers were thought to weigh the benefits of personal data disclosure (e.g., personalization) against associated risks (e.g., privacy violations). These early theoretical contributions were later applied to AI-based consumer interfaces, such as chatbots and recommender systems, where privacy, transparency, and perceived competence emerged as key determinants of trust (Gefen, 2000; Pavlou, 2003).

2.2. Privacy Concerns and Trust

Recent research has shown that privacy concerns represent one of the most potent inhibitors of trust in AI systems. Kim et al. (2023), drawing on regulatory focus theory, demonstrated how privacy concerns shape responses to personalized chatbot advertising. Their findings revealed that prevention-focused consumers are more susceptible to privacy risks, resulting in lower trust levels. Similarly, Song et al. (2024) found that while anthropomorphic chatbots enhance perceived competence and thus reinforce trust, they simultaneously trigger privacy concerns, which undermine trust. This dual effect has been characterized in the literature as a “double-edged sword.” Furthermore, Wang et al. (2025) emphasized that in ethical AI contexts, privacy and security protection shape employees’ ethical perceptions and recommendation intentions, highlighting that privacy concerns influence trust dynamics not only for consumers but also for employees.

2.3. Anthropomorphism and the Social Response Theory

In AI-based interfaces, anthropomorphism the attribution of human-like features has been considered in the trust literature, with both positive and negative implications. Social Response Theory suggests that individuals apply human–human communication norms when interacting with computers (Nass & Moon, 2000). Within this framework, Schanke et al. (2021) demonstrated in a field experiment that adding anthropomorphic features to customer service chatbots increased transaction conversion rates but also heightened sensitivity to fairness evaluations. Similarly, Kim et al. (2024) demonstrated that anthropomorphic chatbots elicited perceptions of “autonomous intention,” which generated distrust, described as the “uncanny valley of mind.” On the other hand, Song et al. (2024) found that high levels of anthropomorphism boosted trust through perceived competence, yet simultaneously

weakened it due to privacy concerns. These results suggest that the effects of anthropomorphism on trust are ambivalent, depending on context and individual differences.

2.4. Explainability and Transparency

Transparency and explainability (XAI) are critical to building consumer trust. Liao and Sundar (2022) compared content-based versus collaborative filtering algorithms in e-commerce, showing that content-based recommendations created higher transparency perceptions and enhanced trust. In contrast, collaborative filtering relied on social proof to persuade consumers. Similarly, Cheng et al. (2022) found that chatbot disclosure had varying effects on trust: under high task complexity, disclosure weakened the empathy–trust relationship but strengthened the friendliness–trust link. These findings provide valuable insights into the role of transparency in building consumer trust, making the audience more informed and aware of the factors influencing trust.

2.5. Fairness and Perceived Equity

Fairness has recently emerged as a prominent theme in AI-enabled recommender systems. Zhang et al. (2025) examined gender fairness in large language model (LLM)-based recommendation systems and showed that some algorithms created significant disparities between male and female users. Their study underscored that fairness and neutrality are crucial for trust formation, and that LLM-based systems hold promise for improvements in this area, reassuring the audience about the ethical considerations in AI development.

2.6. Generative AI and Consumer Trust

The latest literature has increasingly focused on the implications of Generative AI (GenAI) for consumer trust. Arce-Urriza et al. (2025) found that while GenAI integration in retail chatbots enhanced perceptions of familiarity and human-likeness, thereby strengthening adoption intentions, it did not significantly increase trust levels and even raised privacy concerns. Similarly, Smith et al. (2025), from a relationship science perspective, argued that although GenAI-powered chatbots can create superficial connections, they fall short of replicating deep relational functions and long-term trust mechanisms. This highlights the distinction between “emotional satisfaction” and “deep trust” in consumer-AI relationships.

2.7. Consumer Experience, Purchase Intention, and Trust

Several studies have investigated the indirect effects of trust on consumer experience and purchase intention. Puertas et al. (2024), using the Uses and Gratifications Theory, analyzed chatbot features such as information, entertainment, social presence, and privacy risks, and demonstrated that consumer experience mediates the positive influence of trust on purchase intention. Similarly, Wang et al. (2023), employing an ERP approach, compared human versus chatbot interactions and found that

chatbots generated lower trust than human agents, with the trust gap being more pronounced for subjective tasks.

3. METHODOLOGY

This study was designed in line with the SPAR-4-SLR protocol (Paul et al., 2021) and PRISMA 2020 guidelines (Page et al., 2021) to systematically review research on artificial intelligence (AI) and consumer trust. The methodology consists of four main stages: (i) search strategy, (ii) eligibility criteria, (iii) selection and data extraction, and (iv) quality assessment and synthesis approach.

3.1. Protocol and Transparency

A research protocol was prepared in advance, systematically structured, and reporting was conducted in accordance with the PRISMA checklist (Rethlefsen et al., 2021). The study was preregistered on the Open Science Framework (OSF) with a time-stamp to ensure transparency and reduce reporting bias.

3.2. Search Strategy

We searched the Web of Science Core Collection on 25 September 2025, querying the Topic (TS) fields with proximity operators. A broad query—TS = (("artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR "large language model*" OR LLM* OR "generative AI" OR "foundation model*" OR chatbot* OR "conversational agent*" OR recommender* OR "recommendation system*" OR "dynamic pricing") NEAR/3 (marketing OR advertis* OR "customer service" OR retail* OR "e-commerce" OR "digital market*" OR "customer experience" OR CX)) AND TS = (trust OR "consumer trust" OR "customer trust" OR "trust in AI" OR "algorithmic trust" OR credibility OR transparency OR explainab* OR XAI OR fairness OR bias OR "perceived risk" OR "privacy concern*" OR disclosure)—returned 785 records. We then focused specifically on GenAI/LLMs/chatbots/recommenders in marketing/retail/e-commerce—TS = (("generative AI" OR LLM* OR "large language model*" OR chatbot* OR "conversational agent*" OR recommender*) NEAR/3 (marketing OR advertis* OR retail* OR "e-commerce")) AND TS = ("consumer trust" OR "trust in AI" OR transparency OR explainab* OR fairness OR privacy OR disclosure)—yielding 90 records. Applying sequential filters produced the final corpus: Article (48 records); Publication years 2021–2025 (42); Indexes: SSCI or SCI-EXPANDED (33); and Web of Science Categories: Business, Communication, Management, Psychology Multidisciplinary, Business Finance, Economics, Environmental Studies, Ergonomics, Ethics, Psychology Experimental, Green & Sustainable Science & Technology (17 records). Full query texts, field scope (TS), filters, and counts were logged to ensure reproducibility in line with PRISMA-S guidance for SSCI studies.

3.3. Eligibility Criteria

Inclusion and exclusion criteria were pre-specified as follows:

Inclusion:

- Peer-reviewed articles indexed in SSCI or SCI-E.
- Published between 2021 and 2025.
- Within the contexts of marketing, business, consumer behavior, information systems, or ethics.
- Directly examining the relationship between AI applications (chatbot, recommender, generative AI) and consumer trust.
- Empirical studies (experiments, surveys, mixed-methods, field studies) or conceptual/theoretical contributions.

Exclusion:

- Conference proceedings, book chapters, theses, and gray literature.
- Engineering studies with a purely technical/algorithmic focus.
- Articles not directly addressing the construct of trust.

3.4. Selection Process

The initial search produced 90 records, which were deduplicated using Zotero. After title and abstract screening, 48 articles remained. Following full-text screening, 33 articles were retained, and finally, 17 met all eligibility criteria and were included in the review.

The selection process was conducted independently by two researchers. Discrepancies were resolved through discussion. Inter-rater reliability was measured using Cohen's κ , achieving agreement above 0.85. The PRISMA 2020 flow diagram illustrates the screening and selection process.

3.5. Data Extraction

A pre-developed codebook was employed for data extraction. Extracted information included:

- Author(s), year, and journal,
- Theoretical framework (e.g., privacy calculus, social response theory, heuristic–systematic model),
- Context (chatbot, recommender, generative AI; sector/country),
- Sample characteristics (N, country),
- Research design (experiment, survey, mixed-methods, field study),
- Key variables (trust dimensions: ability, integrity, benevolence; privacy concern; anthropomorphism; transparency/explainability; fairness),
- Mediators and moderators,

- Main findings and effect directions (positive/negative/neutral).

Data coding was performed independently by two researchers, and discrepancies were reconciled through discussion.

3.6. Quality Assessment

The methodological quality of empirical studies was evaluated using the Mixed Methods Appraisal Tool (MMAT 2018). Studies were categorized as low, medium, or high quality, and these ratings were taken into consideration during synthesis. For conceptual contributions, theoretical originality and conceptual rigor were assessed.

3.7. Synthesis Strategy

Due to heterogeneity in data, a meta-analysis was not conducted. Instead, an effect-direction vote-counting approach was applied alongside thematic synthesis guided by the TCCM framework. Findings were mapped along theory, context, characteristics, and method dimensions, and a future research agenda was developed based on the gaps identified in this matrix..

4. FINDINGS

This section presents the findings from the 17 studies included in the systematic literature review, structured through descriptive statistics, thematic synthesis (using the TCCM framework), and effect-direction vote counting. The results are presented in an organized manner, supported by tables and figures.

4.1. Descriptive Mapping

To orient readers, Table 1 summarizes authors, year, context, design, and trust-related dimensions for the 17 studies. This compact view facilitates cross-study comparisons of settings, methods, and focal constructs.

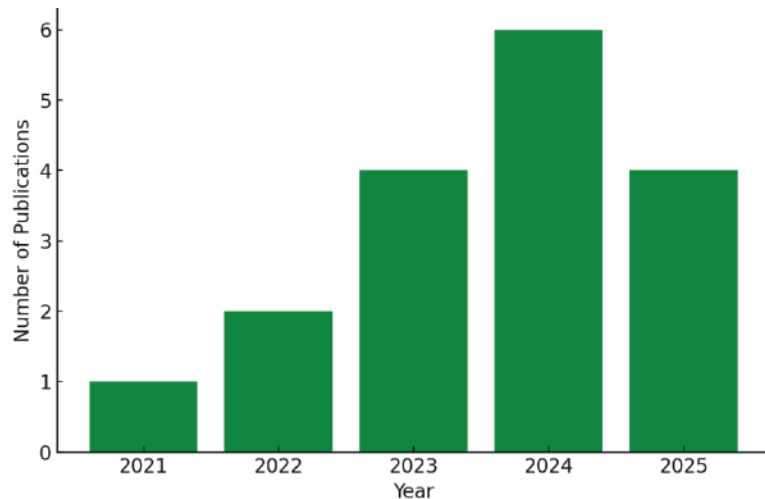
Table 1. Summary of reviewed studies on AI and consumer trust

Author(s)	Year	Context	Design	Trust dimension(s)
Arce-Urriza, M., et al.	2025	Retail chatbots (GenAI vs prior chatbots)	Cross-sectional surveys (Dec 2022; Mar 2024)	Familiarity → Trust; Perceived human-likeness; Privacy risk
Basnet, S., et al.	2025	Capital markets reaction to AI narratives (10-K, etc.)	Event study + text mining	Institutional/corporate trust signals (investor)
Bulathwela, S., et al.	2022	OER recommender (“power to the learner”)	In-the-wild usage analytics + modelling	Transparency, control (structural trust)
Cheng, X., et al.	2022	E-commerce text chatbots	Survey (OLS)	Empathy, friendliness → Trust

Demsar, V., et al.	2025	Advertising creativity & GenAI	Qual interviews (Gioia)	Creative trust in outputs (epistemic)
Drouin, M., et al.	2022	Social chat: chatbot vs online human vs FTF	Lab/online experiments	Social presence / interpersonal trust
Kim, J., & Lee, H.	2023	Chatbot advertising	Experiments (ads via chatbots)	Privacy concern, persuasion knowledge, trust
Kim, J., et al.	2024	Uncanny valley of mind in AI ads	Multi-study experiments	Mind attribution → trust/discomfort
Liao, Y., et al.	2022	E-commerce personalization “show & tell”	Experiments	Transparency/explanations → trust
Puertas, R., et al.	2024	Purchase intention in a chatbot setting	Survey/PLS-SEM (UGT) (journal: <i>Oeconomia Copernicana</i> 15(1) 145–194)	Trust, gratification → intention
Schanke, S., & Burtch, G.	2021	Humanizing service chatbots	Field experiment (humor style)	Warmth/competence → transactional trust
Smith, M. G., & Bradbury, T. N.	2025	Conceptual: Can GenAI emulate human connection?	Theory/review	Relational trust vs pseudo-intimacy
Song, M. M., et al.	2024 (online 2023)	Retail e-commerce chatbots (anthropomorphism)	Scenario experiments (1-factor; 2×2 w/ time pressure)	Competence (ability) & Privacy concerns
Wang, C. C., et al.	2023	Chatbot vs human service (ERP)	Pilot experiment + ERP	Attention/emotion (P2/LPP) → Trust
Wang, X. Q., & Lin, X.	2025 (online 2024)	Security & privacy in ethical AI (marketing employees)	Mixed-methods (JBE)	Organizational privacy/security trust,
Zhang, J. Q., et al.	2025	Fairness: AI-enabled vs LLM-based recsys	Offline eval/simulations	Procedural/distributive fairness (trust proxy)

4.1.1. Distribution by Year

As shown in Figure 1, scholarly interest in AI-enabled consumer trust began to emerge in 2021, grew steadily through 2022, and expanded significantly in 2023 onward. The peak in publications occurred during 2024 and early 2025, reflecting the rising importance of Generative AI and algorithmic trust debates in marketing contexts.

**Figure 1.** Distribution of Publications by Year (2021–2025)

4.1.2. Distribution by Journal

The 17 studies were published across diverse outlets in marketing, information systems, ethics, and behavioral sciences. Leading journals include the Journal of Advertising (3 papers) and the Journal of Retailing and Consumer Services (3 papers), followed by contributions in Behaviour & Information Technology, Information Systems Research, and the Journal of Business Ethics (See Table 2).

Table 2. Journals Publishing Studies on AI and Consumer Trust (2021–2025)

Journal	Number of Articles
Journal of Advertising	3
Journal of Retailing and Consumer Services	3
Behaviour & Information Technology	1
Information Systems Research	1
Journal of Business Ethics	1
Journal of Organizational & End User	1
Perspectives on Psychological Science	1
Sustainability	1
International Review of Financial Analysis	1
Others (various outlets)	4

4.2. Thematic Synthesis

4.2.1. Theoretical Foundations

The reviewed studies applied diverse theoretical perspectives to conceptualize trust in AI interfaces. The most frequently adopted theories include Privacy Calculus Theory, Social Response Theory, the Heuristic–Systematic Model, and the Stimulus–Organism–Response (SOR) framework. Less frequent but notable applications include Cognitive Appraisal Theory and the Service Robot Acceptance Model (SRAM) in the context of GenAI adoption (See Table 3).

Table 3. Theories Applied in the Reviewed Studies

Theory	Representative Studies
Privacy Calculus	Kim et al. (2023, 2024)
Social Response / Anthropomorphism	Song et al. (2024)
Heuristic–Systematic Model	Liao & Sundar (2022)
Stimulus–Organism–Response	Cheng et al. (2022); Wang et al. (2025)
Cognitive Appraisal	Wang et al. (2023)
Service Robot Acceptance Model	Arce-Urriza et al. (2025)

4.2.2. Contexts of Application

As Figure 2 illustrates, the evidence base primarily focuses on three dominant AI-enabled marketing contexts. Chatbots (11 studies; 65%) form the core of the literature, spanning customer service (e.g., service problem resolution and post-purchase support), online retail assistance (product Q&A, order tracking), advertising and persuasion settings (personalized promotions, disclosure of algorithmic targeting), and relational/companionship use cases (para-social connection, social presence). Recommender systems (3 studies; 18%) represent a second cluster focused on e-commerce personalization (content- vs. collaborative-filtering explanations, transparency cues), as well as domain extensions into education (human-intuitive explainable recommenders) and advertising (persuasive effectiveness across product types). Finally, Generative AI (3 studies; 17%) captures the newest wave, covering retail adoption (familiarity and human-likeness effects on acceptance), creative collaboration in advertising (role partitioning between GenAI and human creatives, ethics and disclosure), and human–AI relational dynamics (limits of intimacy and long-term trust). Together, these distributions indicate that trust research is currently chatbot-heavy, while work on GenAI and fairness-aware recommenders is emerging but comparatively underexplored.

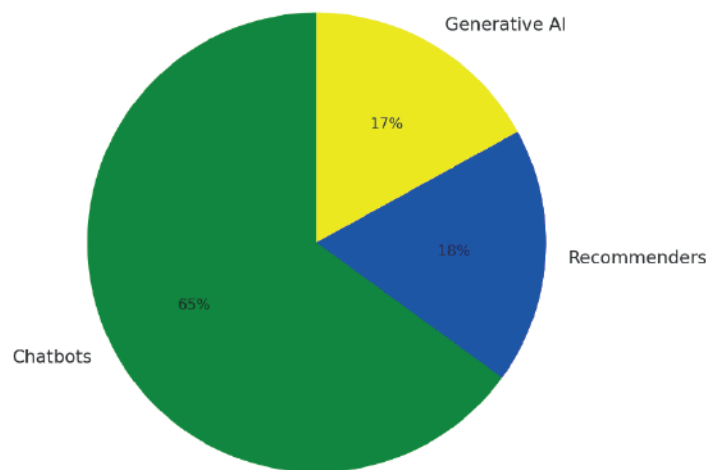


Figure 2. Distribution of Studies by Context

4.3. Effect-Direction Vote Counting

To assess the robustness of relationships, we coded effect directions across all 17 studies and summarized them in Table 4. The pattern is clear: privacy concerns are uniformly detrimental to trust (7/7 negative), corroborating privacy-calculus predictions and highlighting privacy as the most consistent inhibitor in AI-mediated exchanges. In contrast, transparency/explainability (XAI) shows a consistently positive association with trust (4/4 positive), and fairness likewise exhibits positive effects (2/2 positive), suggesting that disclosure, explainability, and equitable treatment operate as reliable enablers of trust. Anthropomorphism, however, yields mixed evidence (3 positive / 2 negative / 1 neutral), underscoring its ambivalence: human-like cues can raise perceived competence and social presence. However, they can also activate autonomy/misuse concerns (the “uncanny” mechanism) that erode trust. Finally, familiarity in GenAI contexts appears beneficial (1 positive), though evidence is still sparse. Patterns were robust when restricting to Medium/High-quality studies under MMAT. We do not interpret directions as effect sizes or causal magnitudes.

Importantly, the mixed anthropomorphism results and the strength of transparency effects align with boundary conditions observed in the primary studies, such as task complexity and the type of disclosure, which can alter the sign or magnitude of effects. Individual differences, including regulatory focus and privacy sensitivity, further moderate outcomes. As detailed in Table 4, these vote counts were derived after quality appraisal; patterns remain when considering only medium- to high-quality studies, increasing confidence that the observed directions are not artifacts of single designs or samples.

Table 4. Direction of Reported Effects on Consumer Trust

Variable	Positive Effect	Negative Effect	Neutral / Mixed
Anthropomorphism	3	2	1
Privacy Concerns	–	7	–
Transparency / XAI	4	–	–
Fairness	2	–	–
Familiarity (GenAI)	1	–	–

4.4. Thematic Findings

Privacy and security: Across contexts, privacy concerns are the most consistent inhibitor of consumer trust. As perceived privacy risk increases, trust in AI-enabled interfaces (chatbots, recommenders, GenAI) declines (Kim et al., 2023; Wang et al., 2025).

Anthropomorphism: Human-like cues (natural language, names/avatars, expressive style) can elevate trust via perceived competence and social presence. However, they can also trigger attributions of agent autonomy and data misuse, lowering trust. Effects are ambivalent and context-dependent (Song et al., 2024; Kim et al., 2024).

Transparency and disclosure: Algorithmic transparency, explainability (XAI), and identity/algorithm disclosure reliably strengthen trust especially in recommender systems and chatbot interactions although the magnitude and direction of effects can vary with task complexity (Liao & Sundar, 2022; Cheng et al., 2022).

Fairness and equity: Perceived fairness particularly in terms of gender and demographic equity in recommendations is a critical condition for sustaining trust. Algorithmic bias undermines legitimacy and erodes trust (Zhang et al., 2025).

Generative AI and user experience: Generative AI (GenAI) increases familiarity and human-likeness, boosting adoption intention; however, it does not automatically raise trust and may heighten privacy concerns, revealing a gap between short-term enjoyment/fluency and durable relational assurance (Arce-Urriza et al., 2025; Smith et al., 2025).

Consumer experience and purchase intention: Trust positively shapes purchase intention through experience dimensions such as information, entertainment, and social presence; human agents tend to elicit higher trust than chatbots in subjective tasks (Puertas et al., 2024; Wang et al., 2023).

Conceptual and managerial perspectives: Concrete, auditable AI disclosures and credible ethical governance signals enhance trust, whereas vague or speculative AI narratives fail to create value (Demsar et al., 2025; Basnet et al., 2025).

The review yields three key insights. First, privacy concerns remain the strongest factor undermining consumer trust. Second, anthropomorphism and transparency generate ambivalent yet potentially positive effects on trust. Third, while GenAI enriches user experiences, it falls short in building sustainable trust. Table 5 (Themes, Key Findings, and Representative Studies on AI and Consumer Trust) synthesizes the evidence base behind these claims.

Table 5. Themes, Key Findings, and Representative Studies on AI and Consumer Trust

Theme	Key Findings	Representative Studies
Early approaches	Trust framed by privacy calculus: consumers weigh personalization benefits against privacy risks.	Dinev & Hart, 2006; Culnan & Bies, 2003; Gefen, 2000; Pavlou, 2003
Privacy concerns	Privacy concerns consistently weaken trust; prevention-focused users are more risk-sensitive.	Kim et al., 2023; Song et al., 2024; Wang et al., 2025
Anthropomorphism	Human-like cues can raise competence/connectedness yet also provoke “uncanny mind” responses and fairness scrutiny.	Schanke et al., 2021; Kim et al., 2024; Song et al., 2024
Transparency & XAI	Transparency/explanations increase trust (e.g., content-based rationales); effects vary with task complexity.	Liao & Sundar, 2022; Cheng et al., 2022
Fairness	Algorithmic fairness (esp. gender equity) is essential for sustaining trust in recommenders.	Zhang et al., 2025
Generative AI (GenAI)	Raises familiarity and adoption intention but does not automatically increase trust; can elevate privacy concerns.	Arce-Urriza et al., 2025; Smith et al., 2025
Experience & intention	Trust boosts purchase intention via user experience; in subjective tasks, humans often inspire higher trust than bots.	Puertas et al., 2024; Wang et al., 2023
Conceptual & managerial	Specific, actionable disclosures and ethical governance build trust; vague AI narratives do not.	Demsar et al., 2025; Basnet et al., 2025

4.5. Overlay Heatmap

Figure 3 (Overlay Heatmap: Themes by Year, 2021–2025) visualizes how these themes have intensified and shifted over time. Together, these displays corroborate that consumer trust in AI-enabled marketing interfaces is multidimensional and must be examined with theoretical, contextual, and methodological diversity, highlighting the need for comprehensive research.

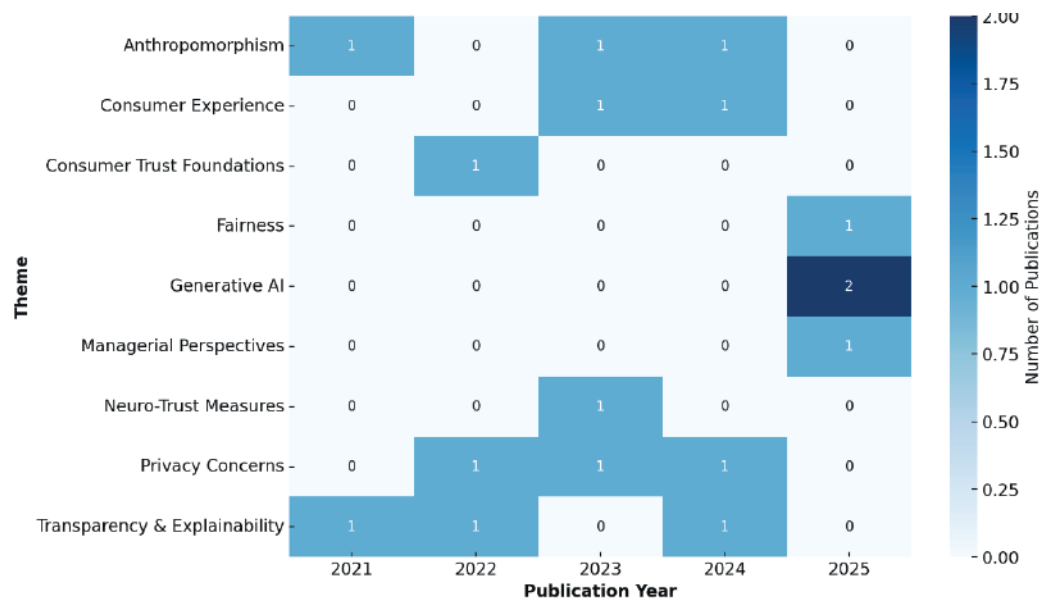


Figure 3. Overlay Heatmap: Themes by Year (2021-2025)

Up to this point, the findings have not only highlighted the central themes in the literature on artificial intelligence and consumer trust but also presented comprehensive conceptual contributions and managerial implications. In particular, the decisive role of transparency, ethical responsibility, and corporate communication in building trust emerges as critical not only at the consumer level but also at the investor and societal levels. The multi-layered nature of these findings provides a solid foundation for advancing theoretical models and rethinking managerial strategies in practice. Therefore, the following section discusses these results in relation to the existing literature and offers a more integrative framework.

5. DISCUSSION

This systematic review analyzed 17 SSCI-indexed articles published between 2021 and 2025 to comprehensively examine the relationship between artificial intelligence (AI) applications and consumer trust. The synthesis of findings demonstrates that trust in AI-enabled interfaces is a multidimensional and context-sensitive construct. The discussion below connects these findings to prior literature and integrates theoretical and managerial implications.

The evidence consistently confirms that privacy concerns remain the most significant barrier to consumer trust in AI applications. Studies such as those by Kim et al. (2023) and Song et al. (2024) demonstrate that privacy concerns consistently erode trust in AI-enabled interactions. These results resonate with the classical privacy calculus model (Culnan & Bies, 2003; Dinev & Hart, 2006), which posits that consumers evaluate the trade-off between risks and benefits when disclosing personal data. Extending this line of reasoning, Wang et al. (2025) demonstrate that security and privacy considerations are not only central to consumers but also shape employees' ethical perceptions and

recommendation intentions, suggesting that privacy is embedded both in customer experience and organizational culture.

Another salient theme is the ambivalent role of anthropomorphism. The literature highlights both its capacity to build trust and its potential to undermine it. For instance, Schanke et al. (2021) observed that anthropomorphic features in customer service chatbots enhanced transaction conversion rates, but simultaneously triggered heightened concerns about fairness and bargaining. Similarly, Kim et al. (2024) reported that high anthropomorphism prompted consumers to attribute “autonomous intentions” to AI agents, evoking the so-called uncanny valley of the mind. Song et al. (2024) likewise found that anthropomorphism strengthens perceptions of competence while amplifying privacy concerns. Together, these findings reveal that anthropomorphism can be an effective but risky trust-building mechanism, whose impact depends heavily on context and consumer perception. Anthropomorphism operates as a double-edged cue: it can enhance perceived competence and social presence yet also activate inferences of agent autonomy and data misuse, thereby dampening trust—especially in high-stakes or subjective tasks.

Transparency and explainability emerge as among the most consistent enablers of trust. Liao and Sundar (2022) demonstrated that content-based recommendations enhanced perceived transparency and thereby trust, while collaborative filtering fostered trust through social proof mechanisms. Cheng et al. (2022) added further nuance by showing that disclosure in chatbot contexts strengthens friendly communication, but under high task complexity, can weaken empathy trust relationships. These insights suggest that transparency is critical but not unconditionally practical; its impact is mediated by context and task characteristics.

Fairness and perceived equity also play an increasingly prominent role. Zhang et al. (2025) demonstrated that gender fairness in recommender systems has a significant influence on the sustainability of consumer trust. This aligns with earlier work emphasizing the importance of distributive and procedural justice in technology acceptance (Lee, 2018; Xu et al. 2005). Consumers expect not only personalized outcomes but also equitable and unbiased treatment, underscoring the theoretical and managerial importance of mitigating algorithmic bias.

The advent of Generative AI (GenAI) has introduced new dynamics into the formation of trust. Arce-Urriza et al. (2025) demonstrated that GenAI increased perceptions of familiarity and human likeness, thereby enhancing adoption intentions without significantly boosting trust. Smith et al. (2025) further argued that GenAI-based conversational agents foster superficial connections but fail to substitute for the deeper trust mechanisms of long-term human relationships. These findings highlight that while GenAI enriches user experiences, it remains limited in fostering sustainable trust.

The behavioral consequences of trust were also evident across the reviewed studies. Puertas et al. (2024), drawing on Uses and Gratifications Theory, found that consumer trust enhances purchase

intentions by affecting user experience. Wang et al. (2023) employed ERP measures to confirm that human–chatbot interactions produce lower trust levels than human–human interactions, particularly in subjective tasks. These results reaffirm trust as a central determinant of adoption and purchase behaviors in AI-mediated environments.

At the organizational and market level, studies also underscore the importance of how firms frame their AI applications. Basnet et al. (2025) found that concrete and actionable AI disclosures enhance market trust, whereas speculative narratives fail to create value. Similarly, Demsar et al. (2025) demonstrated that the ethical and transparent use of GenAI in advertising strengthens brand credibility. These contributions highlight that trust is constructed not only at the consumer level but also in broader investor and societal domains.

Taken together, the synthesis of this literature points to three overarching debates. First, the privacy personalization tension remains the most critical fault line shaping trust. Second, anthropomorphism and transparency act as double-edged mechanisms, holding the potential to strengthen trust but potentially backfiring depending on contextual and individual factors. Third, the contribution of Generative AI is ambiguous, as it enhances experiences but falls short in consolidating durable trust. Overall, consumer trust in AI applications must be theorized and investigated as a multidimensional construct shaped by context, mechanisms, and boundary conditions. This synthesis not only consolidates fragmented findings but also provides a platform for updating theoretical models and guiding future research agendas.

6. CONCLUSION

This systematic literature review examined 17 SSCI-indexed articles published between 2021 and 2025 to provide a comprehensive analysis of how consumer trust is constructed, conditioned, and manifested in AI-enabled marketing applications. The findings highlight that trust extends beyond technical competence or service quality and rests on multi-layered factors such as the privacy personalization trade off, anthropomorphism, transparency and explainability, fairness, and ethical use. Notably, the rise of Generative AI has reshaped the dynamics of trust: while it enriches familiarity and user experience, long-term trust remains fragile and challenging to sustain.

6.1. Theoretical Contributions

This review advances the theory of technology-enabled trust in three integrative ways. First, it articulates an AI-extended trust architecture that layers classic ability–integrity–benevolence with AI-specific levers, including explainability, anthropomorphism, fairness, and privacy governance, and positions these as mechanistic antecedents rather than peripheral design “add-ons.” In this architecture, explainability operates as a diagnostic cue (reducing epistemic opacity), anthropomorphism serves as a social presence cue (shaping inferred agency and intentions), fairness functions as a procedural justice cue (shaping legitimacy), and privacy governance acts as a risk-containment cue (shaping

vulnerability appraisals). The model clarifies why trust can rise or fall across contexts despite similar performance levels.

Second, the review formalizes boundary conditions that reconcile mixed findings in the literature. Trust effects are shown to depend systematically on task properties (objective vs. subjective; simple vs. complex), user states (regulatory focus, privacy sensitivity, need for cognition), and interface choices (degree/type of disclosure; degree/type of anthropomorphism). By specifying these contingencies, the review turns previously fragmented results into testable interaction propositions (e.g., explainability → trust is amplified for experience goods and high-need-for-cognition users; anthropomorphism → trust in low-stakes, objective tasks but can backfire in high-stakes, subjective tasks).

Third, the review reframes Generative AI as a distinct trust context where familiarity and engagement can increase without a commensurate rise in relational assurance. It differentiates experience-proximal outcomes (enjoyment, fluency, human-likeness) from relationship-proximal outcomes (reliance, repairability, accountability), offering a pathway to theorize short-term satisfaction vs. long-term trust. Together, these contributions yield a unified, mechanism- and context-based model of consumer trust in AI that future work can operationalize and test across various domains.

6.2. Practical Implications

Translating these insights into practice requires a trust-by-design mindset that privileges clarity over charm. Explanations should be made first-class elements of the interface: users need concise, human-readable reasons for recommendations and outcomes, with simple recourse options to refine or reject them. The depth of explanation should scale with task stakes, providing brief rationales for routine queries and richer, auditable reasoning for subjective or consequential decisions. Anthropomorphic cues should be right-sized: a warm, polite tone can lower friction in low-stakes service dialogs, but over-humanizing systems risks triggering inferences of autonomous intent over personal data. Capability boundaries should therefore be stated explicitly, and escalation paths to human agents should be kept prominent.

Privacy must be operationalized as part of the value proposition rather than relegated to boilerplate. Just-in-time permissions, data minimization by default, and granular, intelligible controls help users see what is collected, for what purpose, and for how long, alongside visible benefits and explicit constraints on data use. Fairness should be institutionalized through pre-deployment audits, ongoing monitoring for group disparities, and transparent remediation when drift or bias is detected. Where trade-offs arise, documenting them and offering user-facing controls can help preserve legitimacy.

Disclosure should match complexity: identity cues (“AI assistant”) are necessary but insufficient for complex tasks, where process-level disclosures (“generated from your recent orders and preferences”) add credibility. Because trust inevitably fails at times, systems should be engineered for repairability

with rapid human handoff, clear error explanations, and structured apology-and-remedy patterns. Internally, product, legal, and communications functions should align on a single, concrete AI narrative that avoids speculative claims and is consistently reinforced in external messaging. Finally, organizations should measure what matters by instrumenting trust-specific KPIs, perceived transparency, fairness, privacy comfort, reliance, disclosure acceptance, and recovery latency, and segmenting results by task type and user profile. Organizations should instrument trust-specific KPIs and monitor them over time: perceived transparency, perceived fairness (e.g., coverage, Gini, absolute group difference), privacy comfort, reliance, disclosure acceptance, and recovery latency after failures. Segment KPIs by task type (objective vs. subjective) and user profile (privacy sensitivity; regulatory focus). In chatbots, this means emphasizing clarity and safe handoffs; in recommenders, explanation fidelity and de-biasing with tunable controls; and in generative systems, source-grounded outputs, visible limitations, and assured human oversight.

6.3. Limitations

This SLR is limited to 2021–2025 and WoS/SSCI. Therefore, access to early literature and studies in non-WoS indexes is limited. Due to heterogeneous designs and measurements, no meta-analysis was conducted; a qualitative synthesis based on facet counting was preferred. Because this approach fails to capture effect sizes, future meta-analytic estimations with standardized measures and multiple database searches are recommended. We excluded grey literature (reports, theses, non-indexed papers) to focus on peer-reviewed SSCI/SCI-E studies; this choice may introduce coverage bias, as relevant insights outside indexed outlets are omitted. In addition, there is risk of publication bias (positive-results and time-lag bias), which may over-represent significant findings. These limitations should be considered when interpreting the synthesis and motivate future multi-database searches and trim-and-fill-style sensitivity analyses.

6.4. Future Research Directions

Future research should build on these findings in several ways. First, there is a clear need for longitudinal studies that can trace the durability of trust over time, particularly in Generative AI contexts where initial engagement may not translate into long-term reliance. Second, multi-contextual studies should expand beyond e-commerce and advertising to sectors such as healthcare, finance, education, and public services, where trust expectations and risks may differ substantially. Third, cross-cultural research is essential to uncover how cultural values and regulatory orientations shape consumer responses to privacy, fairness, and anthropomorphism in AI systems. Such comparative work clarifies why specific trust mechanisms succeed in one market but fail in another. Finally, there is a pressing methodological need for the development and validation of scales. Constructs such as explainability, fairness, and anthropomorphism are still measured inconsistently across studies, and the development of robust, standardized instruments would significantly strengthen cumulative

knowledge. Together, these directions point toward a richer, more nuanced understanding of trust in AI-enabled marketing that is both theoretically grounded and practically actionable.

REFERENCES

- Arce-Urriza, M., Chocarro, R., Cortiñas, M., & Marcos-Matás, G. (2025). From familiarity to acceptance: The impact of generative artificial intelligence on consumer adoption of retail chatbots. *Journal of Retailing and Consumer Services*, 84, 104234. <https://doi.org/10.1016/j.jretconser.2025.104234>
- Basnet, A., Elias, M., Salganik-Shoshan, G., Walker, T., & Zhao, Y. (2025). Analyzing the market's reaction to AI narratives in corporate filings. *International Review of Financial Analysis*, 105, 104378. <https://doi.org/10.1016/j.irfa.2025.104378>
- Bulathwela, S., Pérez-Ortiz, M., Yilmaz, E., & Shawe-Taylor, J. (2022). Power to the learner: Towards human intuitive and integrative recommendations with open educational resources. *Sustainability*, 14(18), 11682. <https://doi.org/10.3390/su141811682>
- Cheng, X. S., Bao, Y., Zarifis, A., Gong, W. K., & Mou, J. (2022). Exploring consumers' response to text-based chatbots in e-commerce: The moderating role of task complexity and chatbot disclosure. *Internet Research*, 32(2), 496–517. <https://doi.org/10.1108/INTR-08-2020-0460>
- Culnan, M. J., & Bies, R. J. (2003). Consumer privacy: Balancing economic and justice considerations. *Journal of Social Issues*, 59(2), 323–342. <https://doi.org/10.1111/1540-4560.00067>
- Demsar, V., Ferraro, C., Sands, S., & Kohn, A. (2025). Harmony or discord? The intersection of generative AI and human creativity in advertising. *Journal of Advertising Research*, 65(2), 150–166. <https://doi.org/10.1080/00218499.2025.2464305>
- 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>
- Drouin, M., Sprecher, S., Nicola, R., & Perkins, T. (2022). Is chatting with a sophisticated chatbot as good as chatting online or FTF with a stranger? *Computers in Human Behavior*, 128, 107100. <https://doi.org/10.1016/j.chb.2021.107100>
- Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega*, 28(6), 725–737. [https://doi.org/10.1016/S0305-0483\(00\)00021-9](https://doi.org/10.1016/S0305-0483(00)00021-9)
- Kim, W., Ryoo, Y., & Choi, Y. K. (2024). That uncanny valley of mind: When anthropomorphic AI agents disrupt personalized advertising. *International Journal of Advertising*. <https://doi.org/10.1080/02650487.2024.2411669>
- Kim, W., Ryoo, Y., Lee, S., & Lee, J. A. (2023). Chatbot advertising as a double-edged sword: The roles of regulatory focus and privacy concerns. *Journal of Advertising*, 52(4), 504–522. <https://doi.org/10.1080/00913367.2022.2043795>
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. <https://doi.org/10.1177/2053951718756684>
- Liao, M. Q., & Sundar, S. S. (2022). When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. *Journal of Advertising*, 51(2), 256–267. <https://doi.org/10.1080/00913367.2021.1887013>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Paul, J., Lim, W. M., & O'Cass, A. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies*, 45(4), O1–O16. <https://doi.org/10.1111/ijcs.12695>

- 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>
- Puertas, S. M., Manzano, M. D. I., López, C. S., & Cardoso, P. R. (2024). Purchase intentions in a chatbot environment: An examination of the effects of customer experience. *Oeconomia Copernicana*, 15(1), 145–194. <https://doi.org/10.24136/oc.2914>
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Journal of the Medical Library Association*, 109(2), 173–200. <https://doi.org/10.5195/jmla.2021.962>
- Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of humanizing customer service chatbots. *Information Systems Research*, 32(3), 736–751. <https://doi.org/10.1287/isre.2021.1015>
- Smith, M. G., Bradbury, T. N., & Karney, B. R. (2025). Can generative AI chatbots emulate human connection? A relationship science perspective. *Perspectives on Psychological Science*. <https://doi.org/10.1177/17456916251351306>
- Song, M. M., Zhu, Y. X., Xing, X. Y., & Du, J. Z. (2024). The double-edged sword effect of chatbot anthropomorphism on customer acceptance intention: The mediating roles of perceived competence and privacy concerns. *Behaviour & Information Technology*, 43(15), 3593–3615. <https://doi.org/10.1080/0144929X.2023.2285943>
- Wang, C. C., Li, Y. Y., Fu, W. Z., & Jin, J. (2023). Whether to trust chatbots: Applying the event-related approach to understand consumers' emotional experiences in interactions with chatbots in e-commerce. *Journal of Retailing and Consumer Services*, 73, 103325. <https://doi.org/10.1016/j.jretconser.2023.103325>
- Wang, X. Q., Lin, X. L., & Shao, B. (2025). Security and privacy protection in developing ethical AI: A mixed methods study from a marketing employee perspective. *Journal of Business Ethics*, 200(2), 373–392. <https://doi.org/10.1007/s10551-024-05894-7>
- Xu, H., Teo, H.-H., & Tan, B. C. Y. (2005). Predicting the adoption of location-based services: The role of trust and privacy risk. In *Proceedings of the 26th International Conference on Information Systems (ICIS)*.
- Zhang, J. Q., Sun, J. S., Xue, Y., & Liu, Y. Z. (2025). A comparative study of fairness in AI-enabled and LLM based recommendation systems. *Journal of Electronic Commerce Research*, 26(3), 237–265.