

Privacy and Trust in AI Personalization Strategy: A Systematic Literature Review

Ruzam Almas¹ ✉ and I Made Bayu Dirgantara¹

Universitas Diponegoro, Semarang, Indonesia¹

ABSTRACT

The rapid adoption of artificial intelligence (AI) has enabled increasingly sophisticated personalization strategies while raising concerns about consumer privacy and trust. This study examines managerial strategies for balancing AI personalization effectiveness with privacy protection and consumer trust. A Systematic Literature Review (SLR) was conducted using the PRISMA 2020 framework to analyze 57 peer-reviewed journal articles published between 2014 and 2024 and retrieved from Scopus and IEEE Xplore. The findings reveal that managing AI personalization requires an integrated socio-technical approach involving three interconnected dimensions: Governance and Regulatory Strategies, Transparency and Control Strategies, and Technical Strategies. Governance and regulatory mechanisms establish institutional requirements for privacy, ethics, and accountability, while transparency and control strengthen consumer understanding and agency. Technical strategies, including privacy-preserving mechanisms, support the practical implementation of responsible personalization. The review further identifies an Ideas Evidence Gap, reflecting limited empirical validation of proposed strategies, and a Diversity Gap caused by limited representation of non-Western contexts. These findings highlight the importance of integrating privacy, transparency, user agency, and responsible governance into sustainable AI personalization strategies.

Keywords: *AI Personalization, Algorithmic Trust, Federated Learning, Marketing Ethics.*

CORRESPONDING AUTHOR:

Ruzam Almas

Universitas Diponegoro, Semarang, Indonesia

contact: ruzamalmas@students.undip.ac.id

ARTICLE HISTORY

Received : January 16, 2025

Final Revised : May 28, 2025

Accepted : June 06, 2025

Published : June 30, 2025

1. | INTRODUCTION

In the contemporary digital era, personalization has evolved from a supplementary feature into a core strategy in marketing and customer service. Rapid advances in artificial intelligence (AI), particularly machine learning and generative AI, have enabled firms to move from traditional market segmentation toward increasingly individualized and real-time personalization. These capabilities allow firms to anticipate consumer needs, provide relevant recommendations, and improve customer experiences, potentially contributing to customer loyalty, conversion, and competitive advantage. The effectiveness of digital personalization is also connected to broader customer engagement mechanisms, particularly as organizations increasingly rely on digital service systems to facilitate interaction with consumers. (Aljazzaf et al., 2025). Emerging applications of generative AI also indicate that AI-driven recommendations can extend beyond commercial personalization by influencing consumers' sustainability-related decision-making and behavior (Silalahi, 2025).

For consumers, personalization can reduce information overload and improve the efficiency and relevance of digital services. Highly personalized digital environments may also narrow the range of information presented to users, potentially reinforcing selective exposure and limiting exposure to diverse information (Pariser, 2011). However, the value of personalization depends heavily on the collection and analysis of user data, creating an enduring tension between personalization benefits and privacy concerns (Aguirre et al., 2015; Benlian, 2015; Acquisti et al., 2015). This tension is commonly described as the personalization–privacy paradox, in which consumers express substantial concerns about privacy while continuing to disclose personal information to receive personalized services (Awad & Krishnan, 2006; Barth & de Jong, 2017; Xu et al., 2011).

Privacy Calculus Theory provides an important explanation for this behavior by proposing that individuals evaluate perceived benefits against perceived risks before disclosing information, while procedural fairness and trust can further shape perceptions of information practices (Dinev & Hart, 2006; Culnan & Armstrong, 1999). Recent evidence further demonstrates that consumers evaluate personalization through both privacy risks and personalization benefits, indicating that the paradox is not simply a conflict between personalization and privacy but a decision-making process shaped by contextual factors (Zhu & Chang, 2016; Bol et al., 2018; Wang et al., 2024). Thus, personalization strategies must consider not only the relevance of recommendations but also consumers' perceptions of privacy invasion and control. Recent research on smart-device personalization similarly indicates that consumers evaluate the benefits of personalized experiences alongside privacy considerations, highlighting the need to maintain an appropriate balance between personalization and privacy (Chan-Olmsted et al., 2024).

Beyond privacy, AI-driven personalization creates a second challenge concerning trust. The emergence of generative AI further expands these concerns because the

trustworthiness of AI-generated outputs depends on factors such as reliability, transparency, and the ability to assess system behavior (Fan et al., 2025). The increasing use of algorithmic decision-making has raised concerns about the opacity of AI systems, particularly when users cannot understand why particular recommendations or decisions are generated. Explainable AI (XAI) has therefore emerged as an important mechanism for improving transparency and supporting trust. Research indicates that explanations can influence users' perceptions, trust, and acceptance of AI-based services, although the effectiveness of explanations depends on whether users can meaningfully understand them (Adadi & Berrada, 2018; Shin, 2021). Evidence from AI-based voice assistants further suggests that explanations can shape how users perceive and evaluate AI systems, reinforcing the importance of explanation design in human–AI interaction (Seymour & Such, 2022). Similarly, explainable recommendation research emphasizes that explanations should be designed according to users' needs rather than treated as a universal solution (Zhang & Chen, 2020). Recent empirical research on AI-based recommendations on social media also demonstrates that the effects of explanations depend on how users perceive and process the information provided by the recommendation system (Haque et al., 2024). This suggests that transparency and explainability should be integrated with user-centered design and meaningful control mechanisms.

A further concern involves algorithmic fairness. AI systems trained on historical or potentially biased data may reproduce or amplify existing inequalities, creating risks of discriminatory outcomes and unequal treatment (Barocas & Selbst, 2016). In recommendation environments, fairness has increasingly become a multidimensional issue involving both algorithmic processes and consumers' perceptions of fair treatment (Deldjoo et al., 2023; Vassøy & Langseth, 2024). Consequently, responsible personalization requires a broader socio-technical perspective that simultaneously addresses privacy protection, transparency, trust, user agency, and fairness. In addition, research on algorithmic recommendation systems indicates that social trust and perceptions of algorithmic equity can influence users' willingness to interact with recommendation technologies, suggesting that fairness is closely connected to trust formation (Wu et al., 2024).

Although privacy, trust, personalization, and algorithmic fairness have been extensively investigated, the existing literature remains fragmented in terms of how these challenges can be addressed through integrated managerial and technical strategies. Prior research has frequently focused on identifying privacy concerns, measuring trust, or examining adoption intentions, while fewer studies systematically synthesize actionable solutions. This fragmentation is also evident in research on user profiling, which identifies privacy, data management, and personalization as interconnected challenges in the development of user-centered digital systems (Eke et al., 2019). For example, privacy-enhancing technologies such as federated learning have been proposed to enable personalized services while reducing the need to centralize sensitive user data (Wang et al., 2018; Kaaniche et al., 2020).

At the design level, transparency, explainability, and user control have been investigated as mechanisms for improving trust and acceptance (Shin, 2021). At the governance level, privacy protection and responsible data practices have become increasingly important components of digital marketing strategy (Martin & Murphy, 2016). The broader adoption of AI also creates ethical and policy challenges, making responsible governance increasingly important for organizations deploying AI-enabled personalization (Anshari et al., 2022). Recent work on AI ethics further emphasizes transparency, accountability, and inclusivity as interconnected principles for responsible AI implementation (Tabaghdehi & Arda, 2025).

Existing systematic literature has contributed to understanding the broader relationship between digital technologies and marketing innovation, but a focused synthesis of managerial and technical strategies for balancing AI personalization, privacy, and trust remains limited (Athaide et al., 2024). The literature therefore requires a shift from merely identifying the existence of the personalization–privacy–trust paradox toward systematically identifying how organizations can address it. Recent evidence illustrates the importance of this shift: rather than simply measuring the paradox, Wang et al. (2024) demonstrate that recommendation design can influence perceived privacy violation and recommendation acceptance.

Therefore, this Systematic Literature Review (SLR) aims to identify, categorize, and synthesize the managerial and technical strategies proposed or tested in the literature to balance AI-driven personalization with consumer privacy and trust. The study specifically addresses the following research question:

RQ: What managerial strategies does the literature propose to balance AI personalization effectiveness with consumer privacy and trust?

By synthesizing the relevant literature, this study seeks to provide a structured understanding of technology-, design-, and governance-oriented solutions and to clarify how these approaches can collectively support more trustworthy and privacy-conscious AI personalization.

2. | LITERATURE REVIEW

Core Conflict: The Personalization-Privacy Paradox

AI-driven personalization enables firms to provide individualized recommendations and services by collecting and analyzing consumer data. Although personalization can improve relevance and convenience, extensive data collection creates tension between consumers' desire for personalized services and concerns regarding privacy. This phenomenon is widely recognized as the personalization–privacy paradox, in which consumers may express strong privacy concerns while continuing to disclose personal information to obtain personalized benefits (Awad & Krishnan, 2006; Barth & de Jong, 2017). Privacy Calculus Theory explains this behavior by proposing that consumers evaluate perceived benefits against perceived privacy risks before deciding whether to disclose information (Dinev & Hart, 2006). Recent research extends this perspective by demonstrating that consumers' privacy

decisions are influenced by the relative value of personalization benefits and privacy costs (Zhu et al., 2017). The personalization–privacy paradox is also dynamic, as consumers’ willingness to disclose information may change across different stages of the personalization process and according to their perceptions of the benefits and risks of self-disclosure (Zeng et al., 2021). Personalization may therefore be accepted when consumers perceive sufficient utility from the service, while perceived intrusiveness and privacy risks may reduce acceptance.

Research on personalized advertising similarly shows that consumers’ responses depend on the balance between relevance, perceived intrusiveness, and privacy considerations. Trust also plays an important role in personalized advertising, as consumers’ responses to personalization are influenced not only by perceived relevance but also by their level of trust in the advertising process and the organization providing it (Bleier & Eisenbeiss, 2015). The perceived intrusiveness of personalization is also important because consumers may distinguish between advertising that is perceived as appropriately tailored and advertising that is experienced as invasive (Girona & Korgaonkar, 2018). The literature consequently suggests that privacy protection should not be treated as separate from personalization design.

Research on e-commerce privacy further demonstrates that privacy concerns encompass multiple dimensions, indicating that organizations need to consider different forms of consumer privacy concerns when designing digital personalization and data practices (Bandara et al., 2019). When privacy concerns are not adequately addressed, consumers may adopt defensive behaviors in response to digital data practices, making privacy protection relevant not only to compliance but also to consumer relationships and continued engagement (Bandara et al., 2020). Privacy expectations can also vary across informational, temporal, and spatial contexts, suggesting that consumers’ reactions to privacy practices depend partly on whether organizations meet the expectations associated with a particular interaction context (Swani & Milne, 2022). Instead, privacy must be incorporated into the development of personalized services through mechanisms that reduce perceived risks while maintaining service utility. This perspective provides an important foundation for examining privacy-preserving strategies within AI personalization.

Trust, Transparency, and User Control in AI Personalization

Privacy concerns alone do not fully explain consumers’ responses to AI personalization. Trust is also important because users must accept a degree of vulnerability when interacting with algorithmic systems (Mayer et al., 1995; Rousseau et al., 1998). Research on trust in automation further shows that trust development is influenced by multiple characteristics of the system, user, and interaction context, indicating that trust in AI-enabled personalization should not be understood as a simple consequence of system performance alone (Schaefer et al., 2016). The opacity of AI systems can make it difficult for users to understand why particular recommendations are produced, potentially reducing confidence in algorithmic decisions. Consequently,

transparency and explainability have become important approaches for improving users' understanding of AI systems. Explainable AI (XAI) provides mechanisms for communicating how or why algorithmic outputs are generated. Adadi and Berrada (2018) identify explainability as an important response to the black-box nature of AI, while Chen et al. (2022) demonstrate that explainable recommendation can improve transparency, trustworthiness, and user satisfaction by providing understandable reasons for recommendations.

However, explanations should be designed according to users and contexts rather than assumed to have identical effects across all situations. The effectiveness of explainable AI may also depend on users' domain knowledge, as differences in users' understanding of a task can influence how explanations affect trust and task performance (Dikmen & Burns, 2022). The need for explanations may also vary according to users' epistemic uncertainty, suggesting that explanation should be provided when it meaningfully reduces uncertainty rather than being treated as a uniform requirement across all users and situations (Jiang et al., 2022). Evidence from explainable AI recommendations further indicates that consumers may respond differently to post hoc explanations, highlighting the importance of how explanations are presented and interpreted in recommendation contexts (Chen et al., 2023).

This is consistent with evidence in the current SLR showing that transparency can have different effects depending on the information disclosed and the characteristics of users. User control and agency therefore complement transparency. Research on personalized digital news environments similarly identifies agency, trust, and perceived utility as important elements shaping users' perspectives toward personalization (Monzer et al., 2020). Providing meaningful privacy settings, personalization controls, and opt-out mechanisms can give consumers greater control over how their data and personalized experiences are managed. This supports the emerging view that trustworthy personalization requires not only explaining algorithmic processes but also enabling users to exercise meaningful control over them.

Fairness, Privacy-Preserving Technology, and Governance

AI personalization also raises concerns about fairness because algorithms trained on historical or unbalanced data may reproduce existing biases and generate unequal outcomes (Barocas & Selbst, 2016). In recommender systems, fairness has become a multidimensional issue involving algorithmic processes, consumer treatment, and evaluation methods (Deldjoo et al., 2023; Vassøy & Langseth, 2024). These concerns indicate that responsible personalization requires solutions that extend beyond privacy protection alone. From a technical perspective, Privacy-Enhancing Technologies (PETs) provide mechanisms for reducing privacy risks while maintaining personalized services. Kaaniche et al. (2020) classify PETs into several categories and specifically examine their potential to reconcile the conflicting objectives of personalization and privacy.

Federated learning and other privacy-preserving approaches similarly allow personalization-related computation while reducing the need to centralize raw user data (Wang et al., 2018; Ali et al., 2021). Trust-based federated learning has also been investigated as a mechanism for improving recommendation systems while addressing challenges associated with limited user or data participation (Wahab et al., 2022). Federated learning has additionally been applied to recommendation models for financial consumer services, illustrating how decentralized learning can support personalization in data-sensitive consumer environments (Chatterjee et al., 2024). Federated matrix factorization provides another example of how decentralized learning can be applied to recommender systems while reducing the need to directly centralize users' data (Du et al., 2021).

These approaches demonstrate that technical solutions can function as important enablers of privacy-conscious personalization. However, technical mechanisms alone cannot resolve the broader socio-technical challenges of AI personalization. Data governance, organizational accountability, ethical principles, and regulatory compliance are also necessary to establish responsible practices (Martin & Murphy, 2016). Thus, the literature increasingly points toward a layered approach combining privacy protection, fairness, technical safeguards, user control, transparency, and governance, providing the conceptual basis for synthesizing managerial and technical strategies in this SLR.

3. | RESEARCH METHOD

This study employed a Systematic Literature Review (SLR) to identify, categorize, and synthesize managerial and technical strategies for balancing AI-driven personalization with consumer privacy and trust. The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 framework to ensure a transparent and systematic selection process (Page et al., 2021). The review focused on peer-reviewed journal articles published between January 1, 2014, and December 31, 2024, and written in English.

Literature was retrieved from Scopus and IEEE Xplore to cover research from management, marketing, artificial intelligence, engineering, and computer science. The search strategy combined four groups of keywords: AI ("artificial intelligence," "AI," "machine learning," "algorithmic," "chatbot," "predictive model"), personalization ("personalization," "customization," "recommendation," "individualization," "targeting"), privacy ("privacy," "data privacy," "information privacy," "data protection"), and trust ("trust," "consumer trust," "customer trust," "user trust," "trustworthiness," "confidence"). Only journal articles that explicitly proposed, tested, or discussed strategies, frameworks, or solutions related to the personalization–privacy–trust paradox were included. Conference papers, book chapters, reviews, editorials, grey literature, non-English publications, and studies that were purely descriptive or unrelated to consumer/marketing contexts were excluded.

The selection process consisted of identification, screening, eligibility assessment, and inclusion. The initial search produced 521 documents. After removing 55 duplicates, 466 unique records remained. During screening, 400 records were excluded because they did not meet the document-type or relevance criteria, leaving 66 articles for full-text assessment. Six articles could not be retrieved, while three were excluded after full-text evaluation. Consequently, 57 journal articles were included in the final qualitative synthesis. The complete selection process is presented in Figure 1.

The article selection process (summarized in the PRISMA Flow Diagram in Figure 1) followed four standard phases:

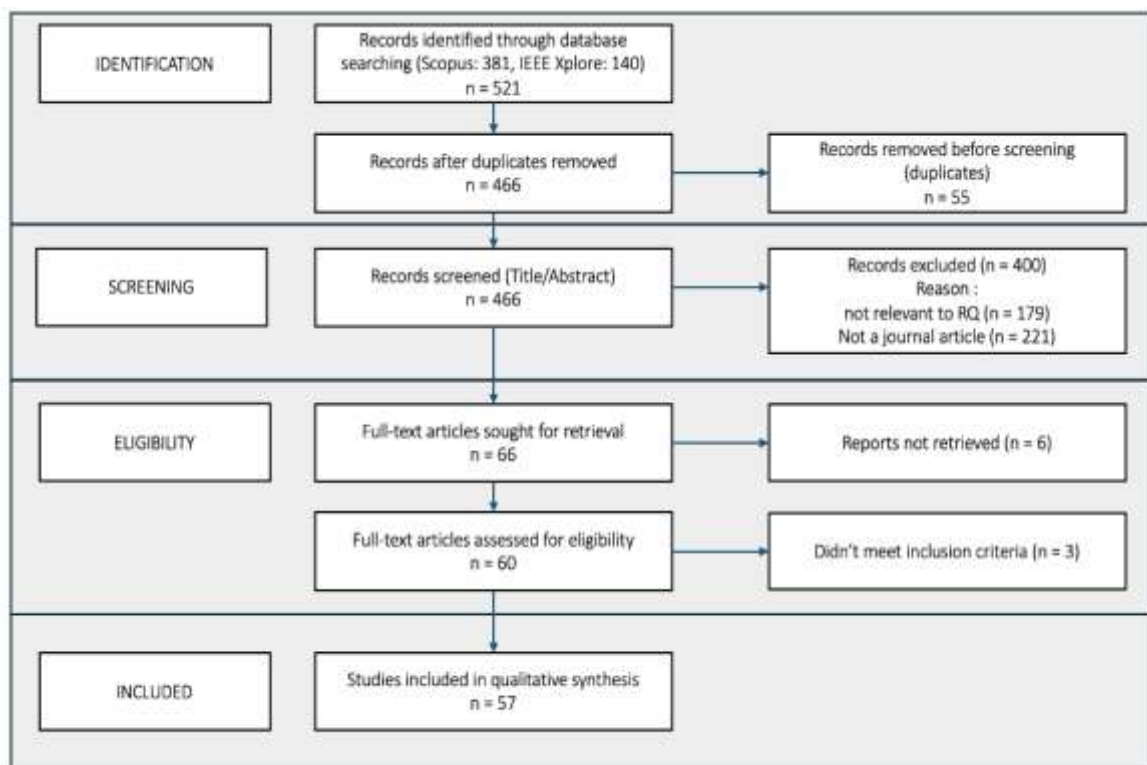


Figure 1. PRISMA Flow Diagram

For the final 57 articles, data were extracted using 11 fields: author and year, title and journal, methodology, context/country, industry or platform, theoretical framework, proposed strategy or solution, solution category, core mechanism, research gaps, and the stated problem or paradox. The extracted data were subsequently analyzed through thematic synthesis to identify recurring patterns across the privacy, trust, personalization, and solution dimensions and to classify the strategies proposed in the literature.

4. | RESULTS

The analysis of the 57 final journal articles produced two interconnected sets of findings: a descriptive synthesis of the research landscape and a thematic synthesis of the problems and solution strategies addressed in the literature. The descriptive analysis

examined publication trends, methodological characteristics, and geographical distribution, while the thematic analysis focused on privacy, trust, personalization, and the strategies proposed to balance AI-driven personalization with consumer privacy and trust.

The publication analysis indicates that research on solutions to the personalization privacy trust paradox has developed rapidly, particularly in recent years. Of the 57 articles included in the synthesis, 84.2% were published during the last three years of the review period. This concentration indicates that research addressing actionable solutions has expanded substantially alongside the rapid development and adoption of AI-based personalization and the increasing importance of data protection and responsible AI practices. The publication distribution is presented in Figure 2.

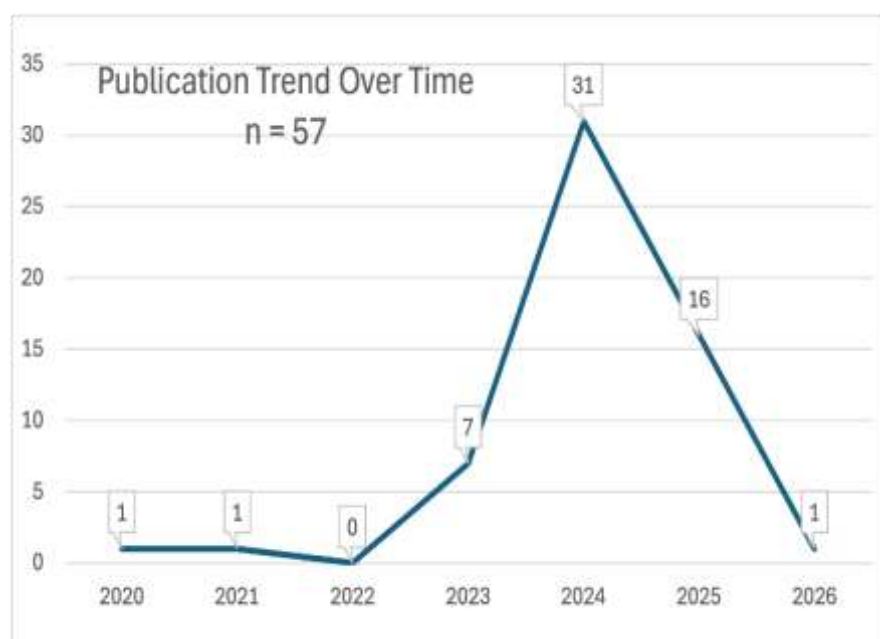


Figure 2. Publication Trend (n = 57)

The methodological distribution shows that the reviewed literature is predominantly evidence-oriented. Quantitative studies, including surveys and experimental approaches, constitute the largest methodological category. This is followed by technical computational studies focusing on system architectures and technical solutions. Conceptual/theoretical studies, other methodological approaches, legal or policy analyses, and qualitative studies constitute the remaining categories. The distribution demonstrates that the literature combines empirical investigation with technical and conceptual development rather than relying exclusively on theoretical discussion. The methodological distribution is presented in Figure 3.

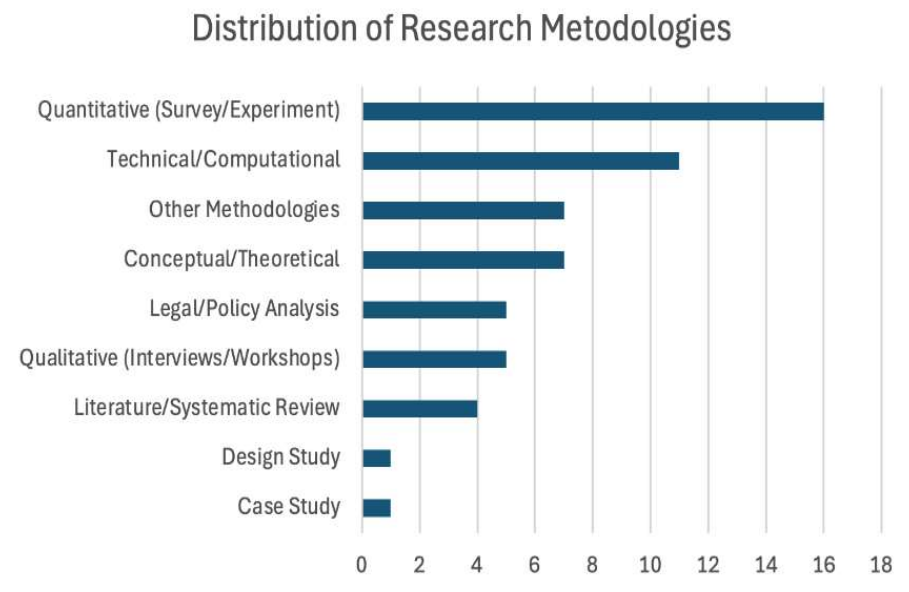


Figure 3. Methodological Distribution (n = 57)

The geographical analysis indicates that the literature is concentrated in broad global or general contexts. Approximately 68% of the studies were classified as Global or General, while location-specific studies were more frequently associated with Western contexts. Only limited representation was identified from Asian and Middle Eastern contexts, including studies involving China, Jordan, and Kuwait. This distribution demonstrates that the empirical evidence represented in the reviewed literature is not geographically balanced. The geographical distribution is presented in Figure 4.

No	Author	Title	Journal	Number of Citations and Scopus Quartile Index	Years of Publication
1	C. I. Eke; A. A. Norman; L. Shuib; H. F. Nweke	Survey of User Profiling: State-of-the-Art, Challenges, Solutions	IEEE Access	208 Q1	2024
2	H. Zhang; B. Wu; X. Yuan; S. Pan; H. Tong; J. Pei	Trustworthy Graph Neural Networks: Aspects, Methods, Trends	Proceedings of the IEEE	194 Q1	2024
3	Virvou, M.	AI and User Experience in Reciprocity	Intelligent Decision Technologies	140 Q3	2024
4	Wahab, O.A., Rioub, G., Bentahar, J., Cohen, R.	Federated Against the Cold: FL for Cold-Start Recommendations	Information Sciences	116 Q1	2024
5	Monzer, C., Moeller, J., Helberger, N., Eskens, S.	User Perspectives on News Personalisation: Agency, Trust, Utility	Digital Journalism	113 Q1	2020

Figure 4. Geographical Distribution

Citation analysis was also conducted to identify influential articles within the reviewed literature. The five most cited articles are presented in Figure 5. This analysis identifies studies that have played an important role as intellectual references within the solution-oriented literature. The citation distribution provides contextual information regarding the development of the research field and identifies studies that have been repeatedly used as foundations for subsequent research.



Figure 5. Citation Analysis of the Most Cited Articles (n = 57)

The thematic analysis subsequently examined how the 57 articles addressed the three central dimensions of the personalization–privacy–trust paradox: privacy, trust, and personalization. The privacy analysis shows that privacy is addressed through multiple dimensions rather than as a single issue. The largest thematic category is Technical Privacy-Preserving, indicating substantial attention to engineered mechanisms for protecting data while supporting personalized services. The second major category is Privacy Concerns & Risks, which addresses users’ perceptions of data collection, misuse, and potential privacy threats. Other identified dimensions include Contextual Privacy and Privacy Governance & Ethics, demonstrating that privacy solutions encompass both technical and socio-organizational considerations. Other identified dimensions include Contextual Privacy and Privacy Governance & Ethics, demonstrating that privacy solutions encompass both technical and socio-organizational considerations. These findings are consistent with prior literature emphasizing that privacy in digital personalization involves both individual perceptions and organizational mechanisms (Martin & Murphy, 2016; Kaaniche et al., 2020). The distribution of privacy themes is presented in Figure 6.



Figure 6. Topics on Trust (n = 57)

The trust analysis demonstrates that trust is also multidimensional. The largest category is Other Trust Aspects, reflecting the conceptual diversity of trust-related issues addressed by the reviewed studies. Among the more clearly identifiable solution-oriented themes, Transparency & XAI represents the most prominent category, followed by Ethical Governance, Trustworthiness Frameworks, and Technical Mechanisms. This finding indicates that trust in AI personalization is addressed through mechanisms that improve system transparency, explainability, governance, and technical reliability. The reviewed studies further demonstrate that trust is influenced by characteristics of users and interaction contexts, including domain knowledge and the conditions under which users interact with automated or explainable systems (Schaefer et al., 2016; Dikmen & Burns, 2022; Jiang et al., 2022). The importance of explainability is consistent with previous research identifying XAI as a mechanism for addressing the opacity of AI systems (Adadi & Berrada, 2018; Zhang & Chen, 2020; Shin, 2021). The trust-related themes are presented in Figure 7.

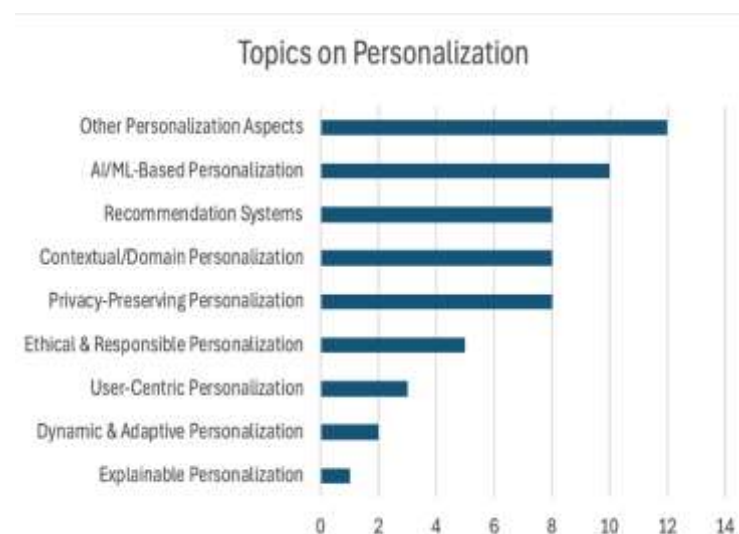


Figure 7. Topics on Personalization (n = 57)

The personalization analysis identifies considerable diversity in the applications and forms of personalization addressed by the reviewed studies. The reviewed studies also demonstrate that personalization outcomes depend on how consumers perceive the relevance, intrusiveness, privacy implications, and trustworthiness of personalized interactions (Bleier & Eisenbeiss, 2015; Girona & Korgaonkar, 2018; Zeng et al., 2021). Other Personalization Aspects represents a substantial category, while the principal identifiable categories are AI/ML-Based Personalization, Personalized Recommendations, Contextual/Domain Personalization, and Privacy-Preserving Personalization. The relatively balanced distribution among these categories demonstrates that personalization is not limited to a single application or technological configuration. Instead, the reviewed studies examine personalization across different AI and recommendation environments, with privacy-preserving personalization representing an important component of the solution landscape. The personalization themes are presented in Figure 8.

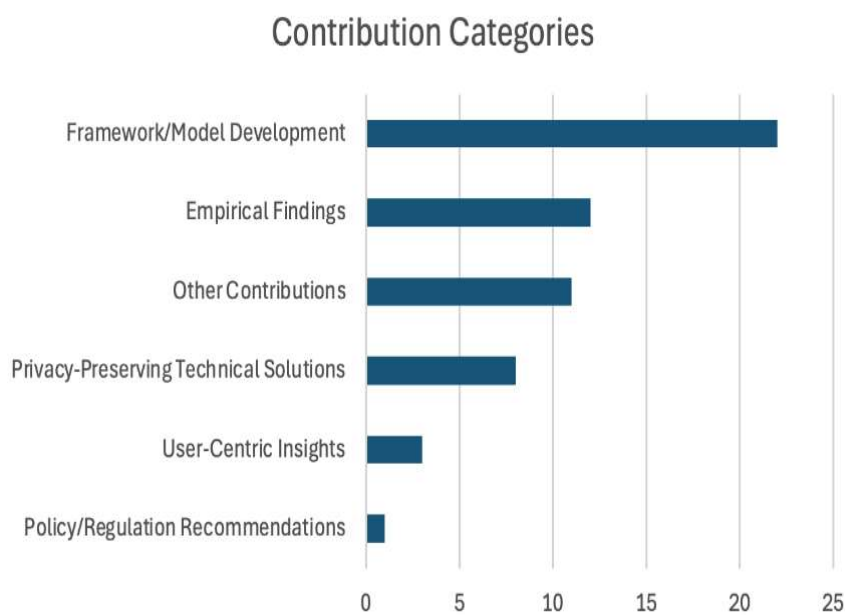


Figure 8. Research Contributions (n = 57)

The final stage of the analysis synthesized the specific strategies proposed or tested across the 57 articles. The analysis identified Framework/Model Development as the most common form of research contribution, followed by Empirical Findings and other forms of contribution. This indicates that the literature provides solutions both through conceptual or technical frameworks and through empirical testing of proposed approaches.

The thematic classification of strategies reveals a fragmented solution landscape. Other Strategies constitutes the largest category, indicating that the reviewed studies propose a broad range of approaches that do not fit within a single standardized

taxonomy. Across the reviewed studies, these strategy categories are represented by approaches including privacy-aware consumer management, explainable AI, user-sensitive explanation design, federated recommendation techniques, and regulatory-oriented data governance (Bandara et al., 2020; Finck & Pallas, 2020; Du et al., 2021; Dikmen & Burns, 2022; Jiang et al., 2022). The distribution of strategy themes is presented in Figure 9.

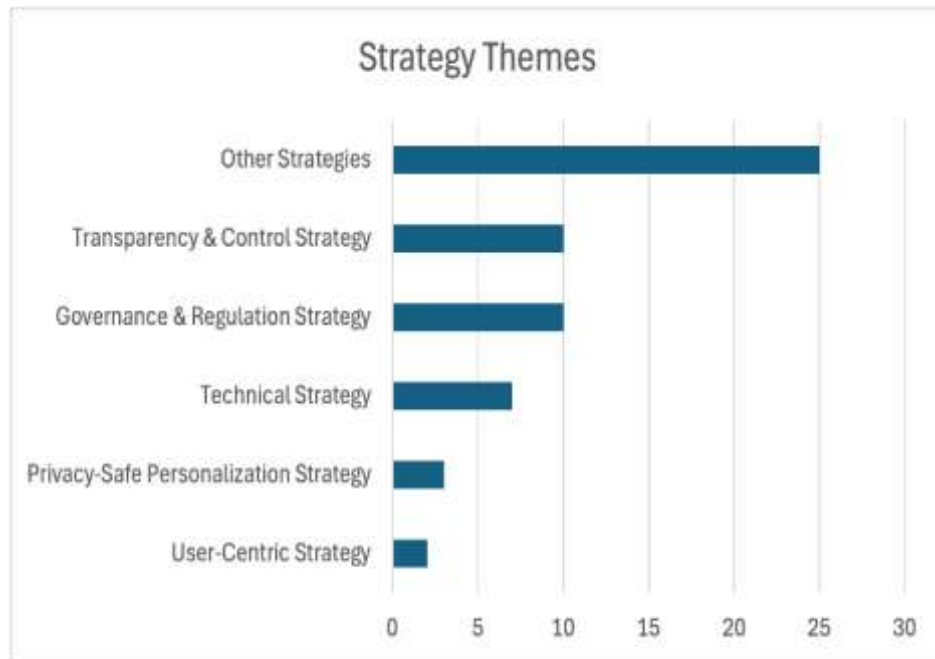


Figure 9. Strategy Themes (n = 57)

The synthesis shows that governance and regulation strategies primarily address the institutional and organizational conditions under which AI personalization is developed and implemented. Recent developments in AI regulation also demonstrate the growing importance of defining institutional responsibilities and regulatory boundaries for the development and use of AI systems (Nikiforov, 2024). Transparency and control strategies address the user-facing dimension of personalization, particularly the provision of understandable information, explainability, privacy controls, and mechanisms through which users can influence personalized experiences. Technical strategies include privacy-preserving mechanisms designed to reduce the exposure or centralization of sensitive data. Examples identified in the reviewed literature include federated learning, differential privacy, and other privacy-enhancing technologies (Wang et al., 2018; Kaaniche et al., 2020; Ali et al., 2021). These technical approaches demonstrate that technological mechanisms form an important enabling layer within broader privacy and trust strategies.

The results indicate that the solution literature does not converge on a single strategy for balancing AI personalization with privacy and trust. Instead, the 57 studies present a diverse combination of governance, transparency, user control, technical

privacy protection, and related strategies. The prominence of Governance & Regulation and Transparency & Control indicates that solutions are increasingly concerned not only with protecting data through technical mechanisms but also with how organizations govern AI systems and how users understand and control personalized services. This finding is consistent with research showing that privacy-enhancing technologies can address the technical tension between personalization and privacy (Kaaniche et al., 2020), while explainability and transparency can contribute to users' perceptions of AI systems (Zhang & Chen, 2020; Shin, 2021).

The results further identify several characteristics of the current research landscape. First, the strong concentration of publications in recent years indicates a rapidly developing research area. Second, the dominance of quantitative research demonstrates substantial attention to empirical validation, while the presence of technical and conceptual studies indicates continued development of proposed solutions. Third, the geographical concentration of the literature reveals limited representation of non-Western contexts. Fourth, the high proportion of studies classified under Other categories across privacy, trust, personalization, and strategy dimensions indicates that the field remains conceptually diverse and lacks a universally established taxonomy.

Taken together, these findings provide the empirical basis for answering the research question. The literature proposes a multi-layered set of managerial and technical strategies rather than an isolated intervention. Governance and regulation address institutional responsibility, transparency and control address user-facing trust and agency, while technical mechanisms provide privacy-preserving infrastructure. The detailed theoretical, managerial, and policy implications of these findings are discussed in the following section.

5. | DISCUSSION

The synthesis of 57 journal articles indicates that balancing AI personalization with consumer privacy and trust requires a multidimensional approach rather than a single intervention. The results show that the literature is rapidly developing, methodologically diverse, and conceptually fragmented, while Governance & Regulation and Transparency & Control emerge as the most coherent strategy categories. Technical strategies, particularly privacy-preserving technologies, function as important supporting mechanisms. These findings suggest that the personalization–privacy–trust paradox should be understood as a socio-technical challenge involving organizational governance, user-facing design, and technological infrastructure.

The first important implication concerns the relationship between personalization and privacy. The literature confirms that consumers do not simply reject personalization because of privacy concerns; rather, they evaluate perceived benefits and risks when deciding whether to disclose information. This decision is also shaped by the perceived intrusiveness of personalization and the extent to which consumers trust the organization or system delivering personalized services (Bleier & Eisenbeiss, 2015; Gironda & Korgaonkar, 2018). This extends the privacy calculus perspective developed

by Dinev and Hart (2006) and supports subsequent research showing that personalization value, relevance, and perceived privacy risks jointly influence consumer responses (Zhu et al., 2017; Bol et al., 2018). Recent evidence further indicates that recommendation acceptance can be influenced by how personalization and privacy are balanced, reinforcing the need to design personalized services around both consumer value and privacy protection. The dynamic nature of self-disclosure further suggests that consumers' privacy decisions may change across stages of the personalization process rather than remain constant over time (Zeng et al., 2021). This finding supports the results of the present SLR, in which Technical Privacy-Preserving represents a major privacy theme.

The second implication concerns transparency, explainability, and trust. The results identify Transparency & XAI as a prominent trust-related solution. This is consistent with research arguing that explanations can reduce the opacity of AI systems and help users understand algorithmic recommendations (Adadi & Berrada, 2018; Zhang & Chen, 2020; Shin, 2021). However, transparency should not be interpreted as simply providing more information. The reviewed studies further indicate that the usefulness of explanations depends on users' characteristics and informational needs. Domain knowledge can affect how users respond to explanations, while epistemic uncertainty can determine when explanations are most useful (Dikmen & Burns, 2022; Jiang et al., 2022). Naiseh et al. (2021) demonstrate that explanations can produce trust-calibration problems when users ignore, misunderstand, or improperly apply them. Their study emphasizes user-centered explanation design and interaction rather than explanation alone. Similar considerations arise in recommendation settings, where post hoc explanations may generate different consumer responses depending on how the explanation is presented and interpreted (Chen et al., 2023). This supports the present finding that Transparency & Control should be treated as an integrated strategy in which explanations are accompanied by meaningful opportunities for users to understand and influence personalization.

The third implication concerns user agency. The reviewed literature suggests that trust in AI personalization is closely connected to consumers' ability to exercise control over their digital experiences. Privacy dashboards, granular personalization settings, understandable choices, and genuine opt-out mechanisms can transform privacy from a passive concern into an active form of user agency. This interpretation is consistent with the current SLR's finding that user control represents an important component of the transparency and control strategy. It also complements the finding that trust should not be considered only as an outcome of AI performance; consumers may evaluate whether they can understand, influence, or limit the system's behavior.

The fourth implication concerns fairness. The results identify fairness as an important dimension of responsible personalization because personalization systems may reproduce unequal outcomes through their underlying data and recommendation mechanisms. Barocas and Selbst (2016) established the potential for data-driven systems to produce disparate impacts, while more recent research shows that fairness

in recommender systems remains a developing and context-dependent field. Deldjoo et al. (2023) emphasize that fairness involves not only technical measurement but also normative questions concerning what constitutes a fair recommendation in a particular context. Similarly, Vassøy and Langseth (2024) show that consumer-side fairness requires systematic consideration of evaluation methods and fairness objectives. Thus, fairness should be incorporated into responsible personalization alongside privacy and trust rather than treated as a separate technical issue.

The fifth implication concerns privacy-preserving technologies and governance. Federated learning, differential privacy, and other Privacy-Enhancing Technologies (PETs) provide technical mechanisms for reducing privacy risks while maintaining data-driven services. Recent studies further demonstrate the application of federated learning to recommendation systems, including trust-based and consumer-oriented recommendation environments (Wang et al., 2018; Kaaniche et al., 2020; Ali et al., 2021; Wahab et al., 2022; Chatterjee et al., 2024). Kaaniche et al. (2020) specifically position PETs as mechanisms for addressing the personalization–privacy paradox, while Wang et al. (2018) and Ali et al. (2021) demonstrate the relevance of privacy-preserving approaches to personalized recommendation systems.

However, the results indicate that technology alone does not constitute a complete solution. Governance, organizational accountability, ethical principles, and regulatory compliance are required to determine how these technologies are implemented and communicated to users. Data governance also requires clear identification of personal and non-personal data because such distinctions determine the scope of applicable privacy protections and organizational responsibilities (Finck & Pallas, 2020). Data governance also requires clarity regarding what constitutes personal data, particularly under regulatory frameworks such as the GDPR, because the classification of information determines the scope of applicable data protection obligations (Finck & Pallas, 2020). Comparative developments in data protection regulation also demonstrate the growing importance of aligning AI applications with principles of data protection and privacy governance (Alibeigi, 2025). Data Protection Impact Assessments provide another governance mechanism through which organizations can identify and mitigate privacy risks associated with AI systems before or during deployment (Galandarli, 2025).

The findings support a layered socio-technical perspective in which Governance & Regulation provides the institutional foundation, Transparency & Control provides the user-facing foundation, and technical privacy-preserving mechanisms provide technological support. This interpretation is consistent with the fragmented nature of the literature identified in the SLR. Rather than identifying one universally superior intervention, the evidence indicates that privacy-conscious AI personalization requires complementary strategies whose implementation depends on users, technologies, organizational contexts, and regulatory environments. The remaining gap is therefore not simply the absence of proposed solutions, but the limited empirical validation of how these strategies interact over time and across cultural and organizational contexts.

5. | CONCLUSION

This systematic literature review examined managerial strategies for balancing AI-driven personalization effectiveness with consumer privacy and trust by analyzing 57 peer-reviewed journal articles published between 2014 and 2024. The findings demonstrate that the personalization–privacy challenge cannot be addressed through a single intervention, but requires an integrated socio-technical approach. Three interconnected strategy dimensions emerged from the literature: Governance and Regulatory Strategies, Transparency and Control Strategies, and Technical Strategies.

Governance and regulatory mechanisms provide the institutional foundation for responsible personalization by establishing privacy, ethical, and accountability requirements. Transparency and control strategies strengthen consumer understanding and agency by enabling users to understand how personalization operates and to exercise meaningful control over their data and personalization preferences. Technical strategies, including privacy-preserving technologies and secure personalization mechanisms, function as important enablers for implementing these principles in AI-based systems. Together, these strategies indicate that privacy should be treated not merely as a constraint on personalization, but as an important condition for sustainable consumer trust and long-term personalization effectiveness.

The review also identifies two important research gaps. First, the Ideas–Evidence Gap indicates that many proposed strategies remain conceptually developed but lack sufficient empirical validation in real-world personalization environments. Second, the Diversity Gap reflects the limited representation of non-Western cultural and geographical contexts in the reviewed literature. These gaps suggest the need for future research using longitudinal, experimental, and cross-cultural approaches to examine how governance, transparency, user control, and technical mechanisms influence privacy perceptions, trust, and personalization outcomes across different contexts. This SLR contributes a structured synthesis of the strategies available for managing the tension between personalization, privacy, and trust. It provides a basis for future empirical research and supports managers in developing AI personalization systems that integrate technological effectiveness with transparency, user agency, privacy protection, and responsible governance.

Acknowledgment

We gratefully acknowledge the contributions of individuals who supported the completion of this article.

Funding Information

This research did not receive any funding.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

Ethical Approval and Originality Statement

Ethical approval was obtained for this study. The manuscript represents original work and has not been previously published, nor is it under consideration by another journal.

Data Disclosure Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514. <https://doi.org/10.1126/science.aaa1465>
- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160. <https://doi.org/10.1109/access.2018.2870052>
- Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, 91(1), 34-49. <https://doi.org/10.1016/j.jretai.2014.09.005>
- Ali, W., Kumar, R., Deng, Z., Wang, Y., & Shao, J. (2021). A federated learning approach for privacy protection in context-aware recommender systems. *The Computer Journal*, 64(7), 1016-1027. <https://doi.org/10.1093/comjnl/bxab025>
- Alibeigi, A. (2025). Bridging the gap: Assessing india's digital personal data protection act in light of the EU GDPR. *SN Computer Science*, 6(7). <https://doi.org/10.1007/s42979-025-04269-7>
- Aljazzaf, Z., Rashed, A., Qureshi, K., & Manuel, P. (2025). Understanding barriers to customer engagement center adoption in digital service systems: Evidence from Kuwait. In *Journal of Logistics, Informatics and Service Science*. Hong Kong Success Culture Press. <https://doi.org/10.33168/jliss.2025.0515>
- Anshari, M., Hamdan, M., Ahmad, N., Ali, E., & Haidi, H. (2022). COVID-19, artificial intelligence, ethical challenges and policy implications. *AI & SOCIETY*, 38(2), 707-720. <https://doi.org/10.1007/s00146-022-01471-6>
- Athaide, G. A., Jeon, J., Raj, S. P., Sivakumar, K., & Xiong, G. (2024). Marketing innovations and digital technologies: A systematic review, proposed framework, and future research agenda. *Journal of Product Innovation Management*, 42(1), 144-165. <https://doi.org/10.1111/jpim.12741>
- Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, 30(1), 13-28. <https://doi.org/10.2307/25148715>
- Bandara, R., Fernando, M., & Akter, S. (2019). Privacy concerns in e-commerce: A taxonomy and a future research agenda. *Electronic Markets*, 30(3), 629-647. <https://doi.org/10.1007/s12525-019-00375-6>
- Bandara, R., Fernando, M., & Akter, S. (2020). Managing consumer privacy concerns and defensive behaviours in the digital marketplace. *European Journal of Marketing*, 55(1), 219-246. <https://doi.org/10.1108/ejm-06-2019-0515>
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. <https://doi.org/10.15779/Z38BG31>
- Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic

- literature review. *Telematics and Informatics*, 34(7), 1038-1058. <https://doi.org/10.1016/j.tele.2017.04.013>
- Benlian, A. (2015). Web personalization cues and their differential effects on user assessments of website value. *Journal of Management Information Systems*, 32(1), 225-260. <https://doi.org/10.1080/07421222.2015.1029394>
- Bleier, A., & Eisenbeiss, M. (2015). The importance of trust for personalized online advertising. *Journal of Retailing*, 91(3), 390-409. <https://doi.org/10.1016/j.jretai.2015.04.001>
- Bol, N., Dienlin, T., Kruijkemeier, S., Sax, M., Boerman, S. C., Strycharz, J., ... & De Vreese, C. H. (2018). Understanding the effects of personalization as a privacy calculus: Analyzing self-disclosure across health, news, and commerce contexts. *Journal of Computer-Mediated Communication*, 23(6), 370-388. <https://doi.org/10.1093/jcmc/zmy020>
- Chan-Olmsted, S. M., Chen, H., & Kim, J. (2024). In smartness we trust: consumer experience, smart device personalization and privacy balance. *Journal of Consumer Marketing*, 41(6), 597. <https://doi.org/10.1108/jcm-12-2021-5072>
- Chatterjee, P., Das, D., & Rawat, D. B. (2024). Federated learning empowered recommendation model for financial consumer services. *IEEE Transactions on Consumer Electronics*, 70(1), 2508-2516. <https://doi.org/10.1109/tce.2023.3339702>
- Chen, C., Tian, A. D., & Jiang, R. (2023). When post hoc explanation knocks: consumer responses to explainable AI recommendations. *Journal of Interactive Marketing*, 59(3), 234-250. <https://doi.org/10.1177/10949968231200221>
- Chen, X., Sun, J., & Liu, H. (2022). Balancing web personalization and consumer privacy concerns: Mechanisms of consumer trust and reactance. *Journal of Consumer Behaviour*, 21(3), 572-582. <https://doi.org/10.1002/cb.1947>
- Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10(1), 104-115. <https://doi.org/10.1287/orsc.10.1.104>.
- Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., & Zanzonelli, D. (2023). Fairness in recommender systems: Research landscape and future directions. *User Modeling and User-Adapted Interaction*, 34(1), 59-108. <https://doi.org/10.1007/s11257-023-09364-z>
- Dikmen, M., & Burns, C. (2022). The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human-Computer Studies*, 162, 102792. <https://doi.org/10.1016/j.ijhcs.2022.102792>
- 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>
- Du, Y., Zhou, D., Xie, Y., Shi, J., & Gong, M. (2021). Federated matrix factorization for privacy-preserving recommender systems. *Applied Soft Computing*, 111, 107700. <https://doi.org/10.1016/j.asoc.2021.107700>

- Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access*, 7, 144907. <https://doi.org/10.1109/access.2019.2944243>
- Fan, M., Wang, C., Chen, C., Liu, Y., & Huang, J. (2025). On the trustworthiness landscape of state-of-the-art generative models: A survey and outlook. *International Journal of Computer Vision*, 133(7), 4317-4348. <https://doi.org/10.1007/s11263-025-02375-w>
- Finck, M., & Pallas, F. (2020). They who must not be identified—distinguishing personal from non-personal data under the GDPR. *International Data Privacy Law*, 10(1), 11-36. <https://doi.org/10.1093/idpl/ipz026>
- Galandarli, A. (2025). Mitigating AI risks: A comparative analysis of Data Protection Impact Assessments under GDPR and KVKK. *Journal of Data Protection & Privacy*, 7(3), 252. <https://doi.org/10.69554/att2755>
- Gironda, J. T., & Korgaonkar, P. K. (2018). iSpy? Tailored versus invasive ads and consumers' perceptions of personalized advertising. *Electronic Commerce Research and Applications*, 29, 64-77. <https://doi.org/10.1016/j.elerap.2018.03.007>
- Haque, A. B., Islam, A. K. M. N., & Mikalef, P. (2024). To explain or not to explain: An empirical investigation of AI-based recommendations on social media platforms. *Electronic Markets*, 35(1). <https://doi.org/10.1007/s12525-024-00741-z>
- Jiang, J., Kahai, S., & Yang, M. (2022). Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies*, 165, 102839. <https://doi.org/10.1016/j.ijhcs.2022.102839>
- Kaaniche, N., Laurent, M., & Belguith, S. (2020). Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey. *Journal of Network and Computer Applications*, 171, 102807. <https://doi.org/10.1016/j.jnca.2020.102807>
- Martin, K. D., & Murphy, P. E. (2016). The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, 45(2), 135-155. <https://doi.org/10.1007/s11747-016-0495-4>
- 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/10.2307/258792>
- Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User perspectives on the news personalisation process: Agency, trust and utility as building blocks. *Digital Journalism*, 8(9), 1142. <https://doi.org/10.1080/21670811.2020.1773291>
- Naiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2021). Explainable recommendation: when design meets trust calibration. *World Wide Web*, 24(5), 1857-1884. <https://doi.org/10.1007/s11280-021-00916-0>
- Nikiforov, L. (2024). Groups of persons in the proposed AI Act amendments. *European Journal of Risk Regulation*, 15(3), 757–771. <https://doi.org/10.1017/err.2024.13>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., Shamseer, L., Tetzlaff, J., Akl, E. A., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E., Mayo-Wilson, E.,

- McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York: Penguin Press.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393-404. <https://doi.org/10.2307/259285>
- Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, 58(3), 377-400. <https://doi.org/10.1177/0018720816634228>
- Seymour, W., & Such, J. M. (2022). Ignorance is Bliss? The Effect of Explanations on Perceptions of Voice Assistants. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2211.12900>
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Silalahi, A. D. K. (2025). Can generative artificial intelligence drive sustainable behavior? A consumer-adoption model for AI-driven sustainability recommendations. *Technology in Society*, 83, 102995. <https://doi.org/10.1016/j.techsoc.2025.102995>
- Swani, K., & Milne, G. R. (2022). Impact of unmet privacy expectations across information, time, and space: evidence from four countries. *Journal of Interactive Marketing*, 57(2), 278-295.
- Tabaghdehi, S. A. H., & Arda, Ö. A. (2025). AI ethics in action: a circular model for transparency, accountability and inclusivity. *Journal of Managerial Psychology*. <https://doi.org/10.1108/jmp-03-2024-0177>
- Vassøy, B., & Langseth, H. (2024). Consumer-side fairness in recommender systems: a systematic survey of methods and evaluation: B. Vassøy, H. Langseth. *Artificial Intelligence Review*, 57(4), 101. <https://doi.org/10.1007/s10462-023-10663-5>
- Wahab, O. A., Rjoub, G., Bentahar, J., & Cohen, R. (2022). Federated against the cold: A trust-based federated learning approach to counter the cold start problem in recommendation systems. *Information Sciences*, 601, 189. <https://doi.org/10.1016/j.ins.2022.04.027>
- Wang, C., Zheng, Y., Jiang, J., & Ren, K. (2018). Toward privacy-preserving personalized recommendation services. *Engineering*, 4(1), 21-28. <https://doi.org/10.1016/j.eng.2018.02.005>
- Wang, Y., Zhu, J., Liu, R., & Jiang, Y. (2024). Enhancing recommendation acceptance: Resolving the personalization–privacy paradox in recommender systems: A privacy calculus perspective. *International Journal of Information Management*, 76, 102755. <https://doi.org/10.1016/j.ijinfomgt.2024.102755>

- Wu, W., Huang, Y., & Qian, L. (2024). Social trust and algorithmic equity: The societal perspectives of users' intention to interact with algorithm recommendation systems. *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2023.114115>
- Xu, H., Luo, X. R., Carroll, J. M., & Rosson, M. B. (2011). The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decision Support Systems*, 51(1), 42-52. <https://doi.org/10.1016/j.dss.2010.11.017>
- Zeng, F., Ye, Q., Li, J., & Yang, Z. (2021). Does self-disclosure matter? A dynamic two-stage perspective for the personalization-privacy paradox. *Journal of Business Research*, 124, 667-675. <https://doi.org/10.1016/j.jbusres.2020.02.006>
- Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1), 1-101. <https://doi.org/10.1561/15000000066>
- Zhu, H., Ou, C. X., van den Heuvel, W. J. A., & Liu, H. (2017). Privacy calculus and its utility for personalization services in e-commerce: An analysis of consumer decision-making. *Information & Management*, 54(4), 427-437. <https://doi.org/10.1016/j.im.2016.10.001>
- Zhu, Y. Q., & Chang, J. H. (2016). The key role of relevance in personalized advertisement: Examining its impact on perceptions of privacy invasion, self-awareness, and continuous use intentions. *Computers in Human Behavior*, 65, 442-447. <https://doi.org/10.1016/j.chb.2016.08.048>