JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

THE AI TRUST DILEMMA: UNRAVELING THE PSYCHOLOGICAL BARRIERS TO ARTIFICIAL INTELLIGENCE ADOPTION

Mr. Sohail Fazal[1], Dr. Shaukat Ayaz[*2], Mr. Khalid Usman[3], Mr. Amjad Hameed[4], Dr. Maimoona Sadiq[5]

[1]*BBA Marketing Iqra National University Peshawar*
[*2]*Asst. Professor Marketing, Iqra National University Peshawar.*
[3]*Lecturer Human Resource Management Iqra National University Peshawar*
[4]*Lecturer Finance Iqra National University Peshawar*
[5]*Asst. Professor Finance, Iqra National University Peshawar*

[*2]shaukatayaz@inu.edu.pk

DOI: https://doi.org/10.5281/zenodo.23100783

Keywords
Artificial Intelligence, Consumer Trust, Privacy Risk, Algorithmic Bias, AI Hallucination, Job Displacement, Behavioral Intention.

Article History
Received: *25 June 2026*
Accepted: *16 September 2026*
Published: *02 October 2026*

Copyright @Author
Corresponding Author: *
Dr. Shaukat Ayaz

Abstract

The purpose of this study is to empirically examine the impact of perceived artificial intelligence (AI) related risks, specifically privacy risk, algorithmic bias, AI hallucination, and job displacement fear, on consumer trust and subsequent behavioral intentions toward AI technologies. Drawing upon the Technology Acceptance Model (TAM) and Risk-Trust frameworks, this study adopts a quantitative, cross-sectional design. Primary survey data were collected from 250 digitally active consumers in Pakistan with prior experience utilizing AI-enabled services. Direct effects and trust-mediated relationships were evaluated using multiple regression and mediation analysis. The empirical results demonstrate that all four risk dimensions significantly and negatively affect consumer trust in AI. AI hallucination exerts the strongest trust-eroding effect, followed closely by privacy risk, algorithmic bias, and job displacement fear. Furthermore, consumer trust exerts a strong positive influence on behavioral intention and partially mediates the adverse impacts of perceived AI risks on adoption intentions. This research addresses a significant geographic gap by providing evidence from the Pakistani consumer market, offering critical theoretical extensions and practical guidance for practitioners and policymakers navigating emerging digital economies.

INTRODUCTION

Artificial intelligence (AI) has emerged as one of the most revolutionary technological forces of the twenty-first century. Tools such as recommendation engines, voice assistants, automated customer service platforms, and generative AI have become deeply embedded in the everyday consumer experience (Huang & Rust, 2021). Organizations across diverse sectors are rapidly adopting AI-based systems to analyze consumer behavior at scale, deliver personalized services, optimize operational expenses, and secure

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

competitive advantages. However, while the efficiencies generated by AI are indisputable, the accelerated deployment of these systems into consumer-facing environments has precipitated a parallel wave of anxiety regarding potential harms, risks, and ethical challenges (Davenport et al., 2020).

Consumers are not merely passive recipients of AI-mediated services; rather, they construct complex perceptions concerning the trustworthiness, fairness, and security of the technologies with which they interact (Shin, 2021). The literature continually demonstrates that trust serves as a foundational building block for technology adoption, where enhanced trust fosters sustained usage, and diminished trust triggers resistance, disengagement, and avoidance behaviors (Glikson & Woolley, 2020). Consequently, identifying the specific factors that compromise consumer trust in AI is of paramount importance.

Academic interest is increasingly converging on four distinct dimensions of AI-related risk. The first, privacy risk, involves concerns regarding unauthorized surveillance, data breaches, and the commercial exploitation of sensitive personal information (Martin & Murphy, 2017). The second, algorithmic bias, denotes systematic unfairness in AI-driven decision-making, where flawed training data or model design places specific consumer groups at a disadvantage (Mehrabi et al., 2021). More recently, two novel risk dimensions have emerged: AI hallucination, which refers to the propensity of generative AI to output highly plausible but factually inaccurate or deceptive information (Ji et al., 2023), and job displacement fear, reflecting rising socioeconomic anxieties that AI-based automation will erode job security across industries (Frey & Osborne, 2017). Despite their individual prominence, existing research largely treats these issues piecemeal (Siau & Wang, 2018; Weidinger et al., 2021). Comprehensive empirical models that simultaneously evaluate how these four disparate risks collectively erode consumer trust and behavioral intentions remain scarce, particularly within developing markets.

Traditional frameworks, such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003), focus predominantly on positive drivers of adoption (e.g., perceived usefulness and ease of use). They do not adequately account for the trust-erosion dynamics triggered by the "dark side" of AI. This theoretical gap is especially evident in the context of the Pakistani consumer market. While the digital economy in Pakistan is booming and exposure to AI-enabled services is expanding rapidly, the socio-technical environment is characterized by unique constraints: variable digital literacy rates, less robust data protection regulatory frameworks compared to Western economies, and heightened sensitivity to unemployment within a labor-intensive economic structure (Arshad et al., 2022; Dwivedi et al., 2021). Consequently, AI-related risk perceptions in Pakistan may operate differently than those reported in international research derived from developed Western or East Asian settings.

To address these critical gaps, this study empirically analyzes the relationships among four AI risk factors (privacy risk, algorithmic bias, AI hallucination, and job displacement fear), consumer trust in AI, and behavioral intention to use AI technologies. By formulating an integrative trust-mediation framework and testing it with primary data from 250 Pakistan-based respondents, this research offers a novel contribution to both AI literature and consumer behavior paradigms. The findings provide evidence-based guidance for technology developers, marketing practitioners, and policymakers seeking to mitigate perceived risks and foster sustainable consumer trust in AI-enhanced service provision.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

2. Theoretical Background and Hypotheses Development

2.1 AI in Consumer Services and the Risk-Trust Framework

Artificial Intelligence encompasses computational methods, including machine learning, natural language processing, computer vision, and deep learning, that enable systems to accomplish tasks traditionally reliant on human cognition (Russell & Norvig, 2021). In consumer services, AI applications such as chatbots, predictive analytics, and generative content systems operate autonomously to analyze data, predict preferences, and customize experiences (Davenport et al., 2020). However, the growing complexity and diminishing interpretability of these systems limit consumers' capacity to critically assess AI outputs. This "black box" dynamic is a fundamental source of risk perception; consumers who do not comprehend how AI arrives at a decision are significantly less likely to perceive the system as accurate, fair, or secure (Doshi-Velez & Kim, 2017; Shin, 2021).

To unpack this dynamic, this study integrates the Technology Acceptance Model (TAM) (Davis, 1989), the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003), and the AI-specific Risk-Trust Framework articulated by Glikson and Woolley (2020). While TAM and UTAUT establish that behavioral intention is traditionally driven by perceived usefulness and ease of use, subsequent extensions highlight that perceived risks act as critical barriers to adoption. Within the AI context, consumer trust, defined as the willingness to rely on a technology system based on optimistic expectations regarding its reliability, competency, and benevolence (McKnight et al., 2002)—is highly fragile. Glikson and Woolley (2020) posit that perceived risks are the primary antecedents of trust erosion when interacting with AI. A single negative experience, such as encountering a fabricated output or an opaque, biased decision, can sustainably damage trust and immediately break the channel to technology acceptance.

2.2 Hypotheses Development

2.2.1 Privacy Risk

Privacy risk in digital consumer services is the subjective belief that disclosed personal data may be collected, stored, or utilized in a manner that is unauthorized, excessive, or detrimental to the individual (Featherman & Pavlou, 2003). Because AI technologies are inherently data-intensive—requiring vast volumes of personal and behavioral data to optimize system effectiveness—consumers frequently experience acute information privacy concerns (Malhotra et al., 2004). Aguirre et al. (2016) conceptualize this as the "personalization paradox": consumers value the convenience of AI-based personalization but simultaneously exhibit heightened anxiety regarding the requisite data surveillance. As consumers become increasingly sensitized to data monetization strategies, privacy risk emerges as a dominant antecedent of undermined trust in AI (Martin & Murphy, 2017). Therefore, we hypothesize:

H1: Perceived privacy risk has a significant negative effect on consumer trust in AI.

2.2.2 Algorithmic Bias

Algorithmic bias occurs when an AI system generates systematically skewed results that disadvantage specific individuals or groups, often stemming from flawed training data, historical prejudices, or improper model design (Lee, 2018; Mehrabi et al., 2021). This phenomenon has profound societal and psychological impacts. In high-stakes, consumer-facing contexts, such as credit scoring or recruitment, perceptions of unfairness severely undermine the credibility of AI systems. Shin (2021) highlights that a lack of explainability exacerbates these reactions; when consumers lack the criteria to evaluate or challenge an automated decision, they are more likely to perceive the system as inherently discriminatory. Consequently, algorithmic bias directly threatens the benevolence and integrity dimensions of trust. Therefore, we hypothesize:

H2: Perceived algorithmic bias has a significant negative effect on consumer trust in AI.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

2.2.3 AI Hallucination

AI hallucination represents a relatively novel but critical area of inquiry in both technical AI studies and consumer behavior literature. It refers to the generation of outputs, text, images, or code, that are convincing and authoritative in presentation but are factually inaccurate, fake, or deceptive (Ji et al., 2023). Bender et al. (2021) caution that the fluent, authoritative typeface of generative AI systems can deceive users into overestimating the accuracy of the outputs. From a consumer perspective, encountering fabricated information from a chatbot or generative tool constitutes a severe epistemic failure. This failure acts as a direct trust violation, cultivating long-term skepticism regarding the validity of AI outputs (Weidinger et al., 2021). Extending this technical phenomenon to a consumer risk perception framework, we hypothesize:

H3: Perceived AI hallucination has a significant negative effect on consumer trust in AI.

2.2.4 Job Displacement Fear

The macro-economic implications of AI-driven automation have been heavily debated, with scholars warning that large segments of routine and codifiable human labor face computerization (Acemoglu & Restrepo, 2020; Frey & Osborne, 2017). In terms of consumer psychology, this widespread socioeconomic anxiety can translate into an individual-level risk perception. Fear of job displacement induces motivated skepticism and affective distancing toward AI. Consumers may develop defensive or hostile attitudes toward AI technologies out of economic self-interest, viewing the system not as a benevolent tool but as a direct existential threat. Despite its prevalence in macroeconomic discourse, empirical research linking job displacement fear to consumer trust remains scarce. We argue that this economic anxiety directly correlates with a reduction in systemic trust. Therefore, we hypothesize:

H4: Job displacement fear has a significant negative effect on consumer trust in AI.

2.2.5 The Mediating Role of Consumer Trust and Behavioral Intention

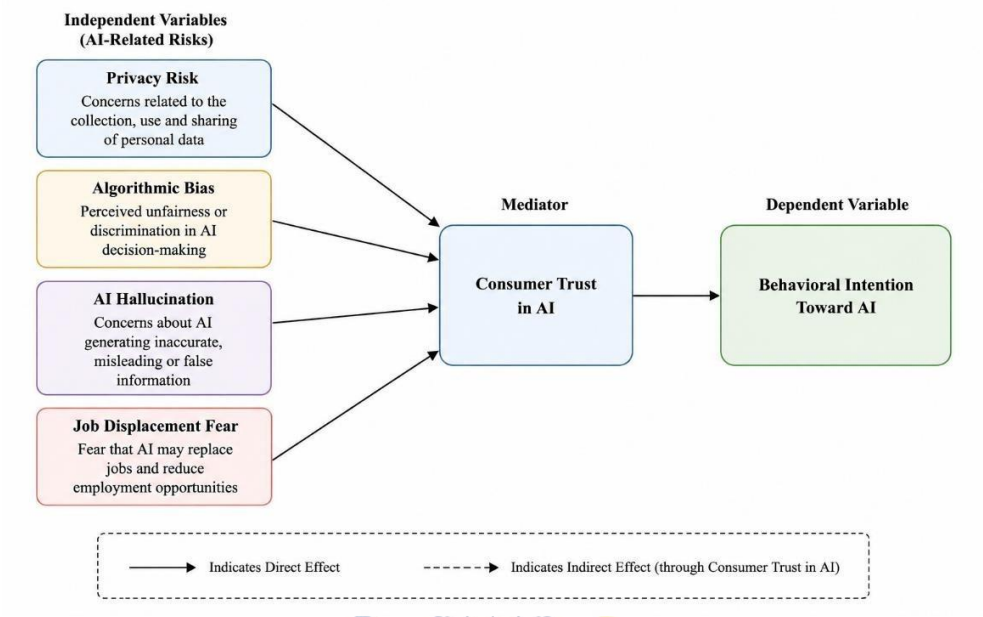
Foundational technology adoption theories indicate that trust acts as the critical gateway translating risk evaluations into final usage decisions (Davis, 1989; Venkatesh et al., 2003). Consumers exhibiting higher levels of calibrated trust in an AI system are more likely to formulate positive behavioral intentions, such as continued usage and recommendations to others, even while acknowledging baseline operational risks. Conversely, the conceptual framework underpinning this study assumes that negative perceptions of privacy, algorithmic bias, hallucination, and job displacement represent distinct theoretical pathways that collectively erode consumer trust. This trust deficit, in turn, extinguishes the desire to engage with AI-enabled services (Glikson & Woolley, 2020). Consequently, consumer trust is posited not merely as an outcome variable, but as the central mediating mechanism between risk perceptions and behavioral intentions. Therefore, we hypothesize:

H5: Consumer trust has a significant positive effect on behavioral intention to use AI technologies.

H6: Consumer trust mediates the relationship between each AI risk factor and behavioral intention toward AI.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026



3. Methodology

3.1 Research Design and Sampling

This study adopted a positivist research philosophy and a deductive approach to empirically test the hypothesized relationships. A quantitative, cross-sectional survey design was utilized, which is well-established in consumer behavior and technology adoption research (Creswell & Creswell, 2018; Venkatesh et al., 2003). The research population consisted of adult consumers (aged 18 years and older) in Pakistan who possessed direct prior experience with AI-enabled digital services, such as AI-powered chatbots (e.g., ChatGPT), voice assistants, algorithmic recommendation systems, and automated customer service tools. Selecting the Pakistani market addresses a significant gap in the international AI literature, which is predominantly skewed toward Western and East Asian contexts.

A purposive convenience sampling technique was employed. The purposive element ensured that all respondents met the strict inclusion criterion of prior AI experience, as individuals without such experience cannot meaningfully assess the nuanced risk dimensions under investigation. To ensure adequate statistical power for both multiple regression and mediation analyses, sample size recommendations by Hair et al. (2014) and Cohen (1992) were consulted. A total of 270 questionnaires were distributed, yielding 258 completed responses. After a rigorous data screening process, which eliminated incomplete responses, those with greater than 10% missing data, failed attention-check items, and straight-line responding, a final usable sample of 250 respondents was retained.

3.2 Data Collection Procedure

Primary data were collected via a structured online questionnaire administered through Google Forms over a four-week period. The survey link was disseminated across various online platforms, including professional networking sites (LinkedIn), academic mailing lists, and social media groups (e.g., WhatsApp). Prior to the main data collection, a pilot test comprising 20 respondents from the target population was conducted to assess item clarity, face validity, and initial internal consistency. Minor wording adjustments were made based on pilot feedback,

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

and all pilot constructs achieved a Cronbach's alpha exceeding 0.70.

The study strictly adhered to ethical research standards. Participation was entirely voluntary and anonymous. A mandatory informed consent check was embedded at the start of the survey, explicitly stating the academic purpose of the study and assuring participants that no personally identifiable information would be collected or compromised.

3.3 Measurement of Variables

The survey instrument utilized validated measurement scales adapted from prior literature to ensure construct validity and reliability within the AI context. All construct items were measured using a five-point Likert scale, anchored by 1 (Strongly Disagree) and 5 (Strongly Agree). Table 1 summarizes the study constructs, the number of items used to measure them, and their primary source references.

Table 1: Construct Summary and Source References

Construct	Role	Items	Primary Source(s)
Privacy Risk (PR)	Independent Variable	5	Featherman & Pavlou (2003); Malhotra et al. (2004)
Algorithmic Bias (AB)	Independent Variable	5	Lee (2018); Shin (2021)
AI Hallucination (AH)	Independent Variable	5	Ji et al. (2023); Bender et al. (2021)
Job Displacement Fear (JDF)	Independent Variable	5	De Witte (1999); Frey & Osborne (2017)
Consumer Trust in AI (CT)	Mediating Variable	5	McKnight et al. (2002); Glikson & Woolley (2020)
Behavioral Intention (BI)	Dependent Variable	5	Davis (1989); Venkatesh et al. (2003)
Note: All items were measured on a five-point Likert scale (1 = Strongly Disagree; 5 = Strongly Agree).			

4. Data Analysis and Results

Data analysis was conducted using IBM SPSS Statistics Version 26. The analytical procedure proceeded in phases: demographic profiling, measurement validation (reliability and descriptive statistics), correlation analysis, regression analysis for direct effects, and mediation analysis for indirect effects.

4.1 Demographic Profile of Respondents

The final sample consisted of 250 respondents. Reflective of the digital nature of the target population, the sample was predominantly young, with 61.2% aged between 18 and 24 years, and 24.8% aged between 25 and 34 years. In terms of gender, 58.8% identified as male, 38.8% as female, and 2.4% preferred not to say. The respondents demonstrated a high level of digital and educational literacy, with the vast majority holding either undergraduate (56.8%) or postgraduate (30.8%) degrees. Crucially for the study's inclusion criteria, 68.4% of respondents reported using AI-enabled services daily (38.8%) or several times a week (29.6%), ensuring sufficient experiential underpinning for their risk evaluations.

4.2 Measurement Validation and Descriptive Statistics

Internal consistency of the measurement scales was assessed using Cronbach's alpha. As shown in Table 2, all constructs exceeded the widely accepted threshold of 0.70 (Nunnally, 1978), ranging from 0.823 to 0.889, confirming robust scale reliability. Descriptive statistics revealed that respondents held moderate to high perceptions across all four AI risk dimensions. Notably, AI hallucination recorded the highest mean score (M

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

= 3.84, SD = 0.76), indicating that concerns regarding inaccurate or fabricated AI outputs represent the most widespread risk perception in the sample. This was closely followed by privacy risk (M = 3.72, SD = 0.81). Consistent with the theorized negative pressure exerted by these risks, consumer trust in AI was recorded just below the neutral scale midpoint (M = 2.98, SD = 0.82).

Table 2: Reliability and Descriptive Statistics

Construct	Mean	SD	Cronbach's Alpha (α)
Privacy Risk (PR)	3.72	0.81	0.847
Algorithmic Bias (AB)	3.58	0.79	0.831
AI Hallucination (AH)	3.84	0.76	0.856
Job Displacement Fear (JDF)	3.66	0.85	0.823
Consumer Trust in AI (CT)	2.98	0.82	0.872
Behavioral Intention (BI)	3.41	0.88	0.889

4.3 Correlation Analysis

Pearson correlation analysis was utilized to examine bivariate associations (see Table 3). All four risk dimensions exhibited statistically significant negative correlations with consumer trust in AI. AI hallucination displayed the strongest negative correlation with trust ($r = -0.481$, $p < .001$), followed by privacy risk ($r = -0.451$, $p < .001$), algorithmic bias ($r = -0.391$, $p < .001$), and job displacement fear ($r = -0.328$, $p < .001$). Additionally, consumer trust was positively and significantly correlated with behavioral intention ($r = 0.581$, $p < .001$). While the risk dimensions showed moderate positive inter-correlations, the magnitudes did not raise significant multicollinearity concerns.

Table 3: Pearson Correlation Matrix

Variable	1	2	3	4	5	6
1. Privacy Risk (PR)	1.000					
2. Algorithmic Bias (AB)	0.412**	1.000				
3. AI Hallucination (AH)	0.388**	0.361**	1.000			
4. Job Displacement Fear (JDF)	0.344**	0.317**	0.299**	1.000		
5. Consumer Trust in AI (CT)	-0.451**	-0.391**	-0.481**	-0.328**	1.000	
6. Behavioral Intention (BI)	-0.283**	-0.241**	-0.309**	-0.198**	0.581**	1.000
*Note: ** $p < 0.01$ (2-tailed). N = 250.*						

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

4.4 Hypothesis Testing: Direct Effects

Multiple regression analysis was conducted to test the impact of the four AI risk factors on consumer trust (H1–H4). The overall model was statistically significant ($F(4, 245) = 29.41$, $p < .001$), explaining 32.4% of the variance in consumer trust ($R^2 = 0.324$, Adjusted $R^2 = 0.313$).

The results (summarized in Table 4) indicate that privacy risk has a significant negative effect on trust ($\beta = -0.31$, $t = -4.88$, $p < .001$), supporting H1. Algorithmic bias also significantly diminishes trust ($\beta = -0.24$, $t = -3.72$, $p < .01$), supporting H2. Notably, AI hallucination exerted the strongest negative effect on consumer trust among the constructs ($\beta = -0.29$, $t = -4.61$, $p < .001$), supporting H3. Finally, job displacement fear exhibited a significant, albeit smaller, negative effect ($\beta = -0.19$, $t = -2.94$, $p < .05$), supporting H4. To test H5, simple regression analysis evaluated the impact of consumer trust on behavioral intention. The model was highly significant ($F(1, 248) = 61.31$, $p < .001$) and explained 33.7% of the variance ($R^2 = 0.337$, Adjusted $R^2 = 0.334$). Consumer trust exerted a strong positive effect on behavioral intention ($\beta = 0.52$, $t = 7.83$, $p < .001$), offering robust support for H5.

Table 4: Summary of Regression Results (Direct Effects)

Hypothesis / Pathway	Standardized β	t-Statistic	Sig. (p-value)	Decision
H1: PR → CT	-0.31	-4.88	< .001	Supported
H2: AB → CT	-0.24	-3.72	< .01	Supported
H3: AH → CT	-0.29	-4.61	< .001	Supported
H4: JDF → CT	-0.19	-2.94	< .05	Supported
H5: CT → BI	0.52	7.83	< .001	Supported

4.5 Mediation Analysis

The Baron and Kenny (1986) causal steps procedure, supplemented by bootstrapping principles inherent in the analytical software, was applied to determine the mediating role of consumer trust (H6) between the four risk factors and behavioral intention. As detailed in Table 5, consumer trust was found to partially mediate the relationship across all four dimensions. In each instance, the indirect effect via consumer trust was statistically significant, and while the direct effect of each risk factor on behavioral intention remained significant, its magnitude was reduced when trust was controlled. These findings provide empirical support for H6, confirming that AI risks erode behavioral intention both directly and indirectly through the destruction of trust.

Table 5: Mediation Analysis Results — Indirect Effects via Consumer Trust

Pathway	Direct Effect (β)	Indirect Effect (β)	Total Effect (β)	Mediation Type
PR → CT → BI	-0.19*	-0.16**	-0.35***	Partial
AB → CT → BI	-0.14*	-0.13**	-0.27***	Partial
AH → CT → BI	-0.18*	-0.15**	-0.33***	Partial
JDF → CT → BI	-0.11*	-0.10*	-0.21***	Partial
*Note: * $p < .05$; ** $p < .01$; *** $p < .001$.*				

5. Discussion and Implications

5.1 Discussion of Findings

The primary objective of this study was to empirically examine how four dimensions of perceived AI risk, privacy risk, algorithmic bias, AI hallucination, and job displacement fear, affect consumer trust and, subsequently, behavioral intention toward AI adoption. The empirical results support all hypothesized relationships,

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

providing critical insights into the "dark side" of AI within a developing digital economy.

Consistent with established literature (Featherman & Pavlou, 2003; Malhotra et al., 2004), privacy risk emerged as a dominant negative predictor of consumer trust. This finding carries specific contextual weight in Pakistan, where comprehensive data protection regulations are still in nascent stages compared to frameworks like the European Union's GDPR. Consequently, Pakistani consumers possess limited regulatory redress concerning data misuse, amplifying their baseline privacy anxieties. Similarly, algorithmic bias significantly diminished consumer trust, aligning with Shin (2021) and Lee (2018). As consumers become increasingly cognizant of AI's role in high-stakes areas like credit assessment and content filtering, the inability to verify procedural fairness directly undermines their confidence in the integrity of these systems.

A standout finding of this study is the potent trust-eroding effect of AI hallucination. While primarily discussed as a technical limitation in prior literature (Ji et al., 2023), this study validates it as a profound psychological barrier. The immediacy of encountering fabricated or misleading information generated by an authoritative-sounding AI system constitutes a severe epistemic failure, causing lasting damage to consumer trust beliefs.

Furthermore, the fear of job displacement exhibited a significant negative impact on trust. While traditionally framed as a macroeconomic concern (Acemoglu & Restrepo, 2020), our findings demonstrate that this anxiety cascades down to the individual level. In a labor-intensive economy like Pakistan, the perceived existential threat of automation fosters a motivated skepticism that reduces interpersonal trust in AI technologies. Finally, the mediation analysis confirmed that consumer trust serves as the vital conduit linking risk perceptions to behavioral intentions, illustrating that AI adoption is a holistic process where mitigating risk is just as crucial as highlighting technological utility.

5.2 Theoretical Contributions

This research offers substantial theoretical contributions to the technology adoption and consumer behavior literature. Foremost, it provides an initial, integrated empirical test of four disparate AI risk dimensions within a single trust-mediation framework. By operationalizing AI hallucination and job displacement fear alongside conventional technological risks (privacy and bias), this study expands the nomological network of the Risk-Trust Framework (Glikson & Woolley, 2020). It successfully translates macroeconomic (job displacement) and technical (hallucination) phenomena into measurable consumer risk perceptions. Additionally, the study addresses a critical geographic gap by extending theoretical models predominantly validated in Western or East Asian contexts to the rapidly digitizing, yet uniquely constrained, Pakistani consumer market.

5.3 Practical Implications

The findings yield highly actionable recommendations for industry stakeholders.

- **For AI Developers and Technology Firms:** Trust-building must center on explicit signaling of uncertainty and epistemic humility. Because AI hallucination severely violates trust, developers should implement mechanisms that indicate knowledge limitations or low AI confidence when generating outputs. Furthermore, privacy assurance must be treated as a primary design parameter rather than an afterthought, featuring transparent data usage policies, opt-in consent, and explainable algorithmic decision-making.
- **For Marketing Practitioners:** Marketers must adopt proactive and truthful communication strategies that recognize AI's limitations rather than over-promising performance. In labor-sensitive markets, communication should carefully position AI as an augmentative tool that complements human judgment rather than an outright substitute, thereby easing job displacement anxieties.
- **For Policymakers and Regulators:** The strong negative effect of privacy risk underscores an

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

urgent need to accelerate the implementation of comprehensive, legally enforceable data protection and AI responsibility frameworks in Pakistan. Furthermore, state investment in digital literacy initiatives will assist consumers in developing calibrated, balanced perceptions of AI's strengths and asymmetric harms.

5.4 Limitations and Future Research

Several limitations present opportunities for future research. First, the cross-sectional design precludes definitive causal inferences; future studies should employ longitudinal data to track how risk perceptions evolve as consumers gain more experience with specific AI systems. Second, the reliance on purposive convenience sampling skewed the demographic profile toward younger, more educated consumers, somewhat limiting the generalizability of the findings across the broader Pakistani population. Third, the study relied on self-reported data, which remains vulnerable to social desirability biases. Methodologically, while the Baron and Kenny (1986) procedure validated the mediation effects, future researchers are encouraged to utilize modern bootstrapping techniques, such as the PROCESS Macro, for more robust statistical inferences. Finally, the current model did not differentiate between specific types of AI services (e.g., healthcare triage vs. entertainment recommendation); subsequent research should isolate risk perceptions across distinct AI application domains.

6. Conclusion

As artificial intelligence becomes inextricably woven into the fabric of daily consumer services, understanding the mechanisms that threaten user acceptance is imperative. This study empirically confirms that the "dark side" of AI—characterized by privacy risks, algorithmic bias, AI hallucination, and the fear of job displacement—exerts a profound, trust-destroying impact on consumers. The findings reveal that AI hallucination and privacy risk are the most powerful forces eroding consumer confidence. Ultimately, this research underscores that sustained behavioral intention toward AI cannot be achieved through technological convenience alone; it necessitates a foundational bedrock of trust. Fostering this trust requires a concerted, ethical effort from developers, marketers, and policymakers to prioritize accuracy, transparency, and consumer protection in the unfolding digital era.

REFERENCES

Acemoglu, D., & Restrepo, P. (2020). Automation and New Tasks: How Technology Displaces and Reinstates Labor, *Journal of Economic Perspectives*, *33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2016). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing, 92*(1), 34–49. https://doi.org/10.1016/j.jretai.2015.11.005

Arshad, S., Malik, O. A., & Hameed, A. (2022). Digital adoption and consumer trust in AI-powered services in Pakistan: An empirical examination. *Journal of Computer Information Systems*, *63*(4), 872–886. https://doi.org/10.1080/08874417.2022.2073186

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 610–623. https://doi.org/10.1145/3442188.3445922

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

De Witte, H. (1999). Job insecurity and psychological well-being: Review of the literature and exploration of some unresolved issues. *European Journal of Work and Organizational Psychology, 8*(2), 155–177. https://doi.org/10.1080/135943299398521

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. https://arxiv.org/abs/1702.08608

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human-Computer Studies, 59*(4), 451–474. https://doi.org/10.1016/S1071-5819(03)00111-3

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson.

Huang, M. H., & Rust, R. T. (2021). Artificial intelligence in service. *Journal of Service Research, 24*(1), 3–15. https://doi.org/10.1177/1094670520902266

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 1–38. https://doi.org/10.1145/3571730

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society, 5*(1), 1–16. https://doi.org/10.1177/2053951718756684

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research, 15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Martin, K., & Murphy, P. (2017). 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

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance of AI systems. *Computers in Human Behavior, 115*, 106606. https://doi.org/10.1016/j.chb.2020.106606

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal, 31*(2), 47–53.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P., & Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*. https://arxiv.org/abs/2112.04359