

AI Transparency And Consumer Judgments: A Meta-Analysis Of Disclosure And Explainability

Arkashis Das¹; Prof. Bibhas Chandra²; Dr. Koushik Mondal³

¹Research Scholar, Management Studies & Industrial Engineering, Indian Institute of Technology (ISM), Dhanbad, Jharkhand, India, 826004 Email id: 24dr0042@iitism.ac.in, arkashisdas98@gmail.com ORCID ID: 0009-0001-4040-6609

²Department of Management Studies and Industrial Engineering Indian Institute of Technology (ISM), Dhanbad, India Email: bibhaschandra@iitism.ac.in ORCIDID: 0000-0002-7017-5944

³Computer Centre Indian Institute of Technology (ISM), Dhanbad, India Email: gembkousk@gmail.com ORCID ID: 0000-0002-8304-8338

Abstract

This study synthesises the fragmented literature on transparency in artificial intelligence (AI) to examine its relationship with consumer judgments, trust, and behavioural intentions. Drawing on a three-level mixed-effects meta-analysis of 304 effect sizes from 82 empirical articles, the study estimates the overall association between AI transparency and consumer outcomes while examining transparency predictors, outcome domains, and contextual boundary conditions. The findings indicate that AI transparency is positively associated with consumer judgments and behavioural outcomes. Notably, simple disclosure labels are statistically as effective as more complex explainability mechanisms, suggesting that greater explanatory complexity does not necessarily yield stronger consumer responses. Transparency also exhibits stronger associations with perceived system characteristics and trust than with behavioural intentions, while these associations become weaker under more demanding contextual and interface conditions. These findings extend the transparency literature by suggesting that AI transparency may function not only as a mechanism for communicating technical information but also as a heuristic signal of organisational openness. From a practical perspective, simple transparency disclosures can provide an effective foundation for building consumer trust, whereas higher-stakes AI applications may benefit from combining disclosure with human oversight and contextual explanations to encourage favourable consumer responses.

Keywords: AI transparency, AI disclosure, consumer trust, explainability, meta-analysis, EU AI Act.

1. Introduction

Artificial Intelligence (AI) use is no longer veiled behind the drapes of technical systems. Day by day, they are prominently being visible across daily consumer-touchpoints, that includes product recommendations, AI powered chatbots, AI-powered ads, and personalised contents.

Thus, it is no longer enough for businesses to ensure AI is functional; Their task at hand is to ensure that consumers understand and trust how these AI operates.

The marketing literature shows that AI disclosure can potentially impact consumer attitudes, trust, and behavioural intentions significantly, which underscores the growing importance of transparency in the age of generative AI, where consumers, advertisers, and regulators are all demanding the disclosure of AI-generated content (Grigsby et al., 2025; Baek et al., 2026; Zhang and Hur, 2025).

Concurrently, research on AI and consumer behaviour has rapidly proliferated, yet its findings remain fragmented across contexts, transparency mechanisms, and consumer outcomes.

Consumer reactions to AI have always been varied. Early studies on algorithm aversion indicated that people were less likely to use algorithms if they observed errors, even when the performance of these algorithms were evidently more accurate than their human counterparts (Dietvorst et al., 2015).

Subsequent studies documented its opposite phenomenon, termed as algorithm appreciation, demonstrating that, in some contexts, consumers prefer algorithmic judgements over human ones (Logg et al., 2019).

Likewise, a study on the use of AI in the medical field found that consumers are not much willing to accept AI for identity-relevant, personal, and emotionally charged tasks, as they believe human judgment is better at capturing the unique and context-specific aspects of service delivery (Longoni et al., 2019).

In sum, these results indicate that consumers are judging AI not only by its actual performance. Perceptions of meaning, legitimacy and humanity also influence their judgments. An equally crucial cue that influences these perceptions is AI transparency.

In this review, AI transparency is referred broadly to the information consumers receive about how AI systems operate, make decisions, or generate outputs, encompassing disclosure, explainability, and related informational signals (Rai, 2020; Shin, 2021).

As a result, transparency has become one of the centrepieces in AI research today. Recent research has demonstrated that explainability can reduce epistemic uncertainty, as users can connect the AI's output to interpretable causes, which increases the user trust and acceptance of AI recommendations.

Similarly, for AI-powered services, the manner in which the explanations are conveyed, including their level of detail and the extent to which the users can control that information, can crucially impact the trust and acceptance (Nizette et al., 2025).

Thus, the scope of AI transparency goes beyond just merely informing about the use of AI, to also including how it is disclosed and communicated, and whether that disclosure information helps to decrease uncertainty or increase consumer concern.

However, new research evidence also indicates that transparency can have potentially adverse downsides too (Yu & Li, 2022).

Recent findings have shown that AI disclosure does not always increase consumer trust. Disclosing the use of AI, can actually lead to a loss of trust, as it undermines the perceived legitimacy and trustworthiness of the AI user (Schilke and Reimann, 2025).

Similarly, in prosocial advertising, disclosure has been associated with lower ad attitudes, trust, and donation intentions when advertisements lack credibility (Baek et al., 2026). AI disclosure in service advertising can lead to a decline in consumer trust and positive attitudes, especially when combined with non-verifiable cues (Grigsby et al., 2025). Recommendation systems also highlight that consumer reactions vary by the type of product being recommended (Mazzu et al., 2025).

The results contradict the conventional notion that transparency is always good, implying that transparency can be interpreted as trust in some situations, and scepticism in others. Furthermore, previous research has operationalised disclosure, explainability, and transparency in different ways and investigated a range of outcomes, such as trust, credibility, attitudes, authenticity, and behavioural intentions, that warrant a quantitative synthesis.

To overcome this fragmentation, our meta-analysis combines 304 effect sizes from 199 independent studies in 82 empirical papers.

Existing reviews examine only parts of this literature. Atf and Lewis (2025) conducted a meta-analysis of 90 studies on explainability and trust, with a similar correlation to our estimate ($r = 0.194-0.204$), but they examined only a single predictor and outcome and provided limited publication bias checks.

Zehnle et al. (2025) compiled a list of 440 related effects from 76,142 participants regarding AI resistance, with transparency playing a minor role. Gelbrich et al. (2026) analysed 327 papers reporting 943 effects of AI relative to human agents, yet considered transparency a secondary characteristic. Balaskas (2026) similarly reviewed nine FinTech studies, but their scope was restricted to robo-advisors.

In our meta-analysis, we overcome three inter-related limitations. First, disclosure, explainability, trustworthiness, human-like cues, and performance signals are all compared within a single analytical framework.

Second, it reveals that transparency is most strongly correlated with perceived system characteristics and less with attitudes, behavioural intentions, and task performance.

Third, it uses a three-level meta-analytic model to distinguish within-study and between-study heterogeneity, includes tests of moderators across system, user, cultural, and task contexts, and provides comprehensive publication-bias corrections.

These developments collectively offer a holistic quantitative analysis of the impact of AI transparency on consumer judgments, various forms of transparency, consumer outcomes, and contextual circumstances.

Three research questions guide the synthesis.

RQ1: What is the overall effect of AI transparency on consumer judgments and behavior across studies?

RQ2: Do the effects differ across transparency forms (disclosure, explainability, trustworthiness signalling) and across outcome types (trust, credibility, attitude, authenticity, behavioral intention)?

RQ3: Which categories of boundary conditions—spanning system and interface features, individual and cultural differences, and context and task characteristics (including AI task type, product type, explanation richness, consumer control, and consumer AI aversion)—moderate the direction and magnitude of transparency effects?

2. Methods

2.1. Terminology

Based on recent conceptual research (Jia et al., 2026; Wang, 2026; Haresamudram et al., 2023), AI transparency is defined as the visibility or even understandability by consumers of the presence, role, nature or reasoning behind the use of AI in a firm.

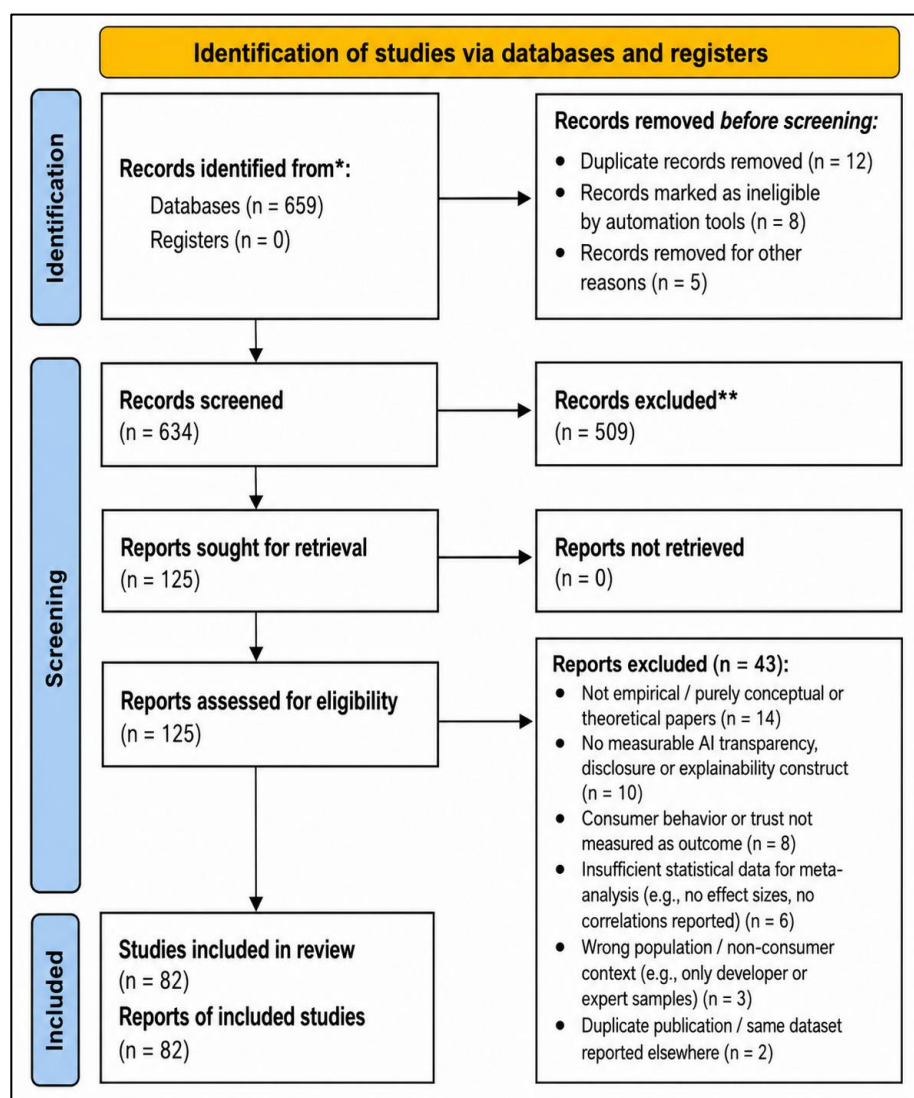
It contains AI disclosure (e.g. content labels or chatbot identity revelation) (Wu et al., 2025; Le et al., 2026; Schilke and Reimann, 2025; Cheng et al., 2022), explainability (intelligible reasons; Hou et al., 2024; Sullivan and Weger, 2025; Ebrahimi et al., 2026), and trustworthiness or anthropomorphic design cues (Cheng et al., 2022; Kim et al., 2022).

This review focuses on consumer-facing AI transparency as the predictor of consumer judgments and behaviours.

2.2. Search strategy

We followed PRISMA guidelines (Page et al., 2021) and searched Web of Science, Scopus, and EBSCO Business Source. The Boolean string used was combined ("AI transparency" OR "AI disclosure" OR "explainable AI" OR "XAI" OR "algorithmic transparency" OR "AI-generated disclosure") AND ("trust" OR "consumer" OR "attitude" OR "purchase intention" OR "behavioural intention" OR "adoption" OR "credibility").

The search was limited to peer-reviewed English articles and conference papers, with no lower year limit, and work up to early 2026 (in-press papers with DOIs).



[Figure 1. PRISMA flow diagram of the study identification, screening, eligibility, and inclusion process. Source: Author]

2.3. Eligibility criteria

Studies were included if they investigated the relationships between AI transparency predictors (such as disclosure, explainability, trustworthiness, or anthropomorphism cues) and performance signals (such as accuracy and error rates) and/or consumer judgments or behaviors, and provided sufficient statistical information to enable the calculation of an effect size.

Studies were excluded if they lacked consumer-related outcomes, adequate statistical information, quantitative data, or reported only model-fit indices or technical XAI metrics.

Only individual-level consumer outcomes were retained. Methodological quality was coded based on research design, data type, and sample type, following the PRISMA 2020 guidelines.

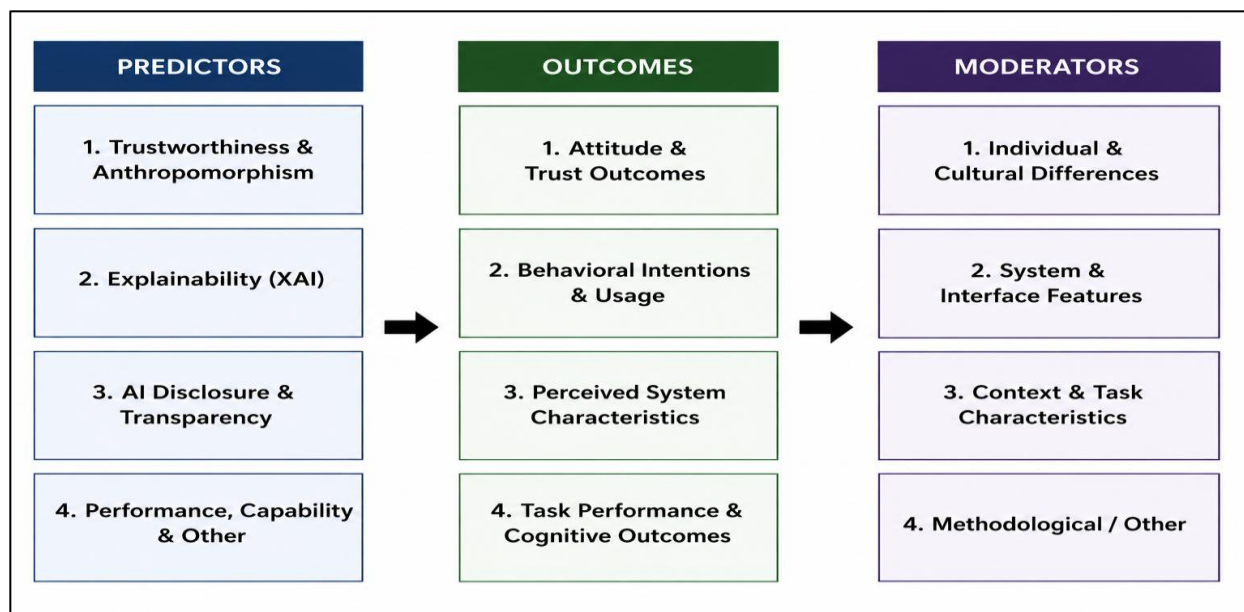
2.4. Additional data sources

The reference lists of eligible articles and recent reviews on trust and disclosure in AI were screened, and supplementary searches were conducted in Google Scholar.

To alleviate the file-drawer problem (Rosenthal, 1979), in-press manuscripts were sought that had been preassigned a DOI. Working papers without peer review or assigned DOIs were excluded.

2.5. Data extraction

From each study, we abstracted the author, year, sample size, country, design, type of AI, transparency predictor, and consumer outcome. Duplicate analyses were excluded, and multiple relevant associations were analysed as separate effect sizes, in line with standard meta-analytic practice (Field and Gillett, 2010). Effects were indicated by the predictor type, the outcome type, and the moderator category, as described in Figure 2. Coding reliability was evaluated with an independent rater by coding 20% of effects, with high levels of agreement found (Cohen's $\kappa = 0.83-0.88$), with disagreements resolved through consensus.



[Figure 2. Conceptual framework: four predictor categories mapping onto four outcome categories, with five moderator categories as boundary conditions. Source: Author]

2.6. Data preparation and effect-size harmonization

All statistics were standardized to Pearson's correlation coefficient (r), consistent with standard practice and supported by the metafor framework, which uses Fisher's z -transformed correlations for meta-analysis (Viechtbauer, 2010). After cleaning, the dataset contained 304 effect sizes from 199 independent studies across 82 papers.

2.6.1. Hierarchical conversion to a common effect size

Direct correlations were retained as reported. For t -statistics, conversion used:

$$r = \text{sign}(t) \times \sqrt{(t^2 / (t^2 + df))}$$

with df typically $N - 2$ for bivariate tests, and the sign of t carried into r . For F -statistics with one numerator df :

$$r = \text{sign}(F) \times \sqrt{(F / (F + df_2))}$$

Because F is unsigned, the sign was assigned from the reported direction. For standardized regression coefficients (β), the Peterson and Brown (2005) approximation was applied:

$$\begin{aligned} r &= 0.98\beta + 0.05 \quad \text{for } \beta \geq 0 \\ r &= 0.98\beta - 0.05 \quad \text{for } \beta < 0 \end{aligned}$$

For p -values, the signed inferential statistic was reconstructed using the reported direction before conversion to r , because two-sided p -values do not contain sign information.

Of the 304 retained effects, roughly two-fifths derived from direct correlations, with the remainder reconstructed via the paths above, and the share derived from standardized betas held under one-sixth.

A sensitivity analysis restricted to the direct- r subset yielded a pooled estimate within 0.01 of the full-sample value, indicating the β -to- r approximation did not materially alter the synthesis.

2.6.2. Fisher's z transformation and standard error

Each correlation was transformed to Fisher's z :

$$z = \frac{1}{2} \times \ln\left(\frac{1 + r}{1 - r}\right)$$

with standard error and sampling variance:

$$SE_z = 1 / \sqrt{N - 3}$$

$$V_z = 1 / (N - 3)$$

Fisher's transformation stabilizes variance and approximates normality, and is implemented in metafor as ZCOR. Rows lacking sufficient information were treated as non-informative and removed rather than imputed, keeping the analytic file auditable.

2.7. Meta-analytic procedures

All syntheses ran in R (R Core Team, 2023) using the metafor package (Viechtbauer, 2010). Effect sizes entered on the Fisher- z scale with inverse-variance weighting via `rma()`. A random-effects framework was used, reflecting the view that included studies are not estimating one fixed true effect but draw from a superpopulation of varying true effects (Hedges, 1992; Field and Gillett, 2010).

Because many papers contributed several effect sizes, a single-level model would understate uncertainty. The primary analysis therefore used a three-level mixed-effects model partitioning variance into three sources (Cheung, 2014; Assink and Wibbelink, 2016; Konstantopoulos, 2011): Level 1 sampling variance ($V = SE^2$), Level 2 within-study variance, and Level 3 between-study variance.

The model was estimated by maximum likelihood (L-BFGS-B). A DerSimonian and Laird (1986) single-level estimator was fitted as a convergent robustness check, and `rma.mv()` confirmed the variance decomposition.

Heterogeneity was assessed with Cochran's Q and I^2 (Higgins et al., 2003):

$$I^2 = \max(0, (Q - df) / Q) \times 100\%$$

I^2 values near 25%, 50%, and 75% indicate low, moderate, and substantial heterogeneity. In the three-level model, the total true heterogeneity was further decomposed into within-study and between-study shares:

$$I^2(\text{Level 2}) = \tau_{22} / (\tau_{22} + \tau_{23} + \tilde{V})$$

$$I^2(\text{Level 3}) = \tau_{23} / (\tau_{22} + \tau_{23} + \tilde{V})$$

where τ_{22} and τ_{23} are within- and between-study variance components and $\tilde{V}$ is the typical sampling variance. Subgroup moderator analyses examined variation by predictor type, outcome target, and analytical context, with between-subgroup differences tested via the omnibus Q_m .

Publication bias was assessed by multi-method triangulation: Egger's regression, a weighted regression variant, Begg and Mazumdar rank correlation, Peters' test, Duval and Tweedie trim-and-fill, and Rosenthal's fail-safe N . Leave-one-out sensitivity analyses at both study and effect levels verified robustness to any single observation.

3. Results

3.1. Overview of included studies

The final dataset included 304 effect sizes from 199 distinct studies in 82 empirical papers (Figure 1), including chatbots, AI-generated and curated ads, virtual influencers, recommender systems, autonomous purchase agents, generative AI assistants, and explainable decision-support tools in East Asia, North America, Europe, and, more recently, South Asia and the Middle East.

The effects were mostly positive, with some negative effects limited to mind-attribution and workload (Wienrich et al., 2022).

Some anthropomorphism manipulations (Bayer et al., 2022), and identity-relevance and human-substitution contexts (Kim et al., 2022). Representative studies are listed in Table 1.

Table 1. Characteristics of representative studies contributing to the meta-analysis.

Study	Region	AI context	Transparency predictor	Primary outcome
Schilke and Reimann (2025)	USA	AI use disclosure	AI Disclosure and Transparency	Attitude and Trust
Wu et al. (2025)	USA	Social media ad label	AI Disclosure and Transparency	Behavioral Intentions
Le et al. (2026)	Multi	AI-generated advertising	AI Disclosure and Transparency	Attitude and Trust
Cheng et al. (2022)	China	Service chatbot identity	Trustworthiness and Anthropol.	Attitude and Trust
Kim (2026)	USA	AI fashion model ad	AI Disclosure and Transparency	Behavioral Intentions
Wang (2026)	Multi	Autonomous retail agent	AI Disclosure and Transparency	Behavioral Intentions
Liang (2026)	China	Live-stream recommender	Trustworthiness and Anthropol.	Attitude and Trust
Ngo (2025)	Vietnam	AI adoption	Explainability (XAI)	Behavioral Intentions
Sullivan and Weger (2025)	USA	XAI decision support	Explainability (XAI)	Task Performance
Ebrahimi et al. (2026)	Multi	Chain-of-thought chatbot	Explainability (XAI)	Behavioral Intentions
Jia et al. (2026)	China	E-commerce AI service	AI Disclosure and Transparency	Attitude and Trust
Hou et al. (2024)	Multi	Anthropomorphic XAI	Explainability (XAI)	Perceived System Char.
Khan and Shehawy (2025)	Saudi Arabia	AI decision integrity	Performance, Capability	Perceived System Char.
Kim et al. (2022)	USA	AI service agent	Trustworthiness and Anthropol.	Behavioral Intentions
Sun et al. (2025)	Multi	Perceived understanding	Explainability (XAI)	Attitude and Trust

Note. Representative subset of the 82 contributing papers. Multi denotes a multi-country sample.

3.2. Overall meta-analysis

The forest plot in Figure 3 displays the overall meta-analysis. The primary estimate came from the three-level mixed-effects model, which accounts for nested effects within studies. On the Fisher-z scale the pooled effect was $z = 0.2074$ ($SE = 0.0080$, 95% CI [0.1918, 0.2231], $Z = 25.87$, $p < .001$), back-transforming to $r = 0.2045$ (95% CI [0.1895, 0.2194]).

The single-level DerSimonian and Laird (1986) estimator produced a closely convergent value ($r = 0.200$, 95% CI [0.187, 0.214], $\tau^2 = 0.008$, $I^2 = 74.98\%$, $H^2 = 3.997$), confirming that the pooled effect is robust to model specification.

Table 2 reports overall results and heterogeneity from both specifications. The 95% prediction interval spanned roughly [0.02, 0.38] on the correlation scale, reflecting wide dispersion of true effects.

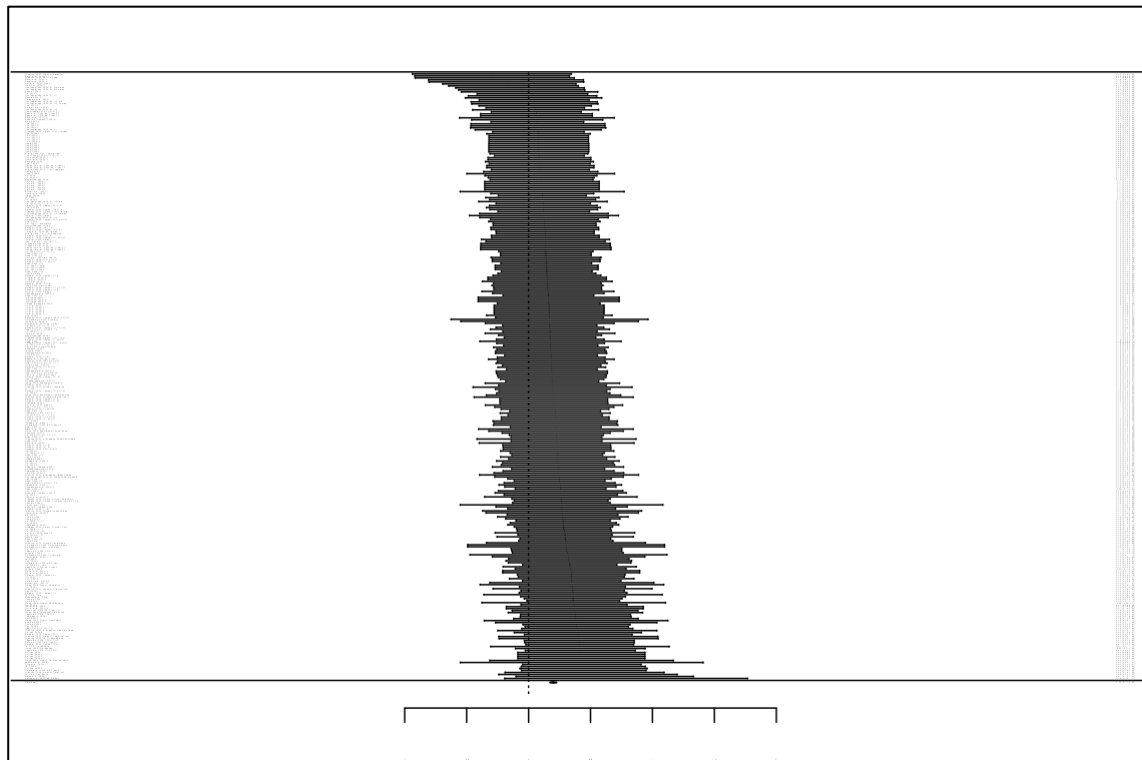
By Cohen's (1988) benchmarks, $r \approx 0.20$ is small to moderate but substantively meaningful: it sits almost exactly on the average effect reported across a century of social psychology ($r = 0.21$; Richard et al., 2003).

Replicated across 82 papers and hundreds of thousands of participants, this is a real signal rather than noise: it is not dramatic enough to convert indifferent browsers into buyers, but it is consistent enough to act as a dependable building block of trust.

Table 2. Overall meta-analytic results and heterogeneity statistics.

Parameter	Three-level mixed-effects model	Single-level random-effects model
Pooled Fisher's z	0.2074	0.203
SE(z)	0.0080	0.007
Z or t statistic	$Z = 25.87$	$t(303) = 29.46$
p value	$< .001$	$< .001$
Back-transformed r	0.205	0.200
95% CI on r	[0.189, 0.219]	[0.187, 0.214]
95% prediction interval on r	[0.00, 0.38]	[0.022, 0.378]
τ^2 (Level 2, within-study)	0.0048	n/a
τ^2 (Level 3, between-study)	0.0059	n/a
τ^2 (total true variance)	0.0107	0.008
τ (total)	0.103	0.090
Q statistic	$Q(303) = 1370.71$	$Q(303) = 1211.08$
Q p value	$< .001$	$< .001$
I^2 (Level 2, within-study)	35.90%	n/a
I^2 (Level 3, between-study)	43.76%	n/a
Total I^2	79.66%	74.98%
H^2	n/a	3.997
Estimation method	ML (L-BFGS-B)	DerSimonian and Laird
Effect sizes (k)	304	304
Studies / Papers	199 / 82	n/a

Note. Three-level model is the primary specification. The single-level model is reported as a robustness check. Both estimators yield substantively convergent pooled correlations.



[Figure 3. Forest plot of the overall random-effects meta-analysis across all 304 effect sizes, showing study-level effect sizes and the pooled summary estimate. Source: Author]

3.3. Heterogeneity and variance decomposition

Variability across the 304 effects was substantial ($Q(303) = 1370.71, p < .001$). The three-level model attributed 79.66% of observed variance to true heterogeneity, and showed where it resides. As Table 3 reports and Figure 4 illustrates, the total true variance $\tau^2 = 0.0107$ split nearly evenly: differences across distinct studies (Level 3) accounted for 43.76% ($\tau^2_3 = 0.0059$); differences among effects within studies (Level 2) accounted for 35.90% ($\tau^2_2 = 0.0048$).

The impact of transparency, therefore, varies not only across studies but also across the specific judgments measured within a study. This is supported by the large prediction interval, which ranges from near zero to approximately 0.38 on the correlation scale; it does not matter where in this interval a particular deployment is positioned, but instead depends on design, audience, and outcome.

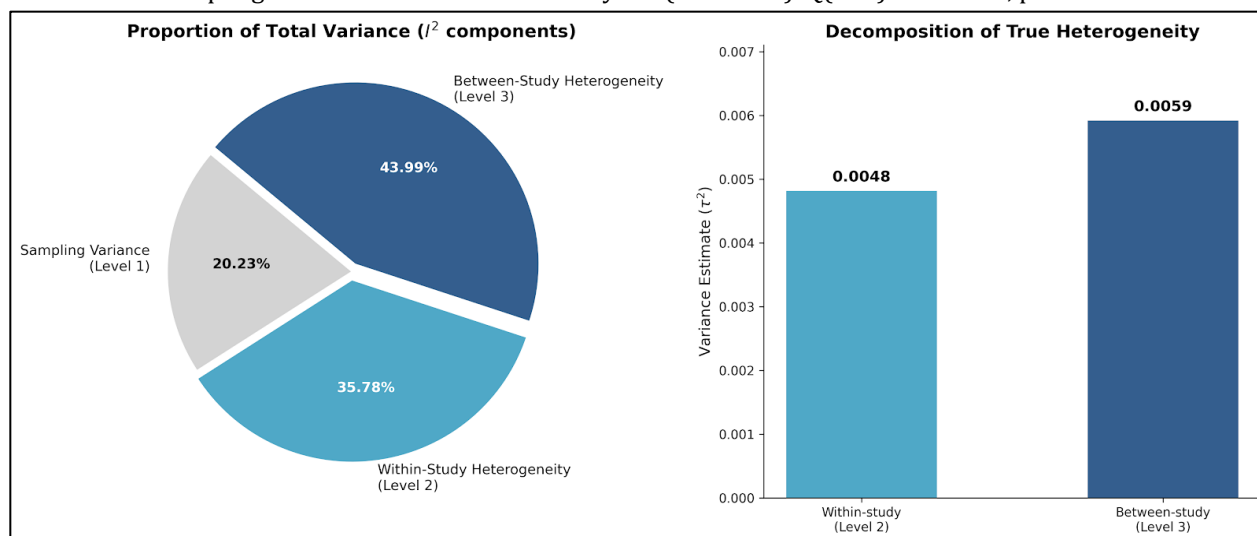
Such heterogeneity is also theoretically expected: as Field and Gillett (2010, p. 670) note, “the effect sizes should be heterogeneous because they come from populations with varying average effect sizes” — under the random-effects framework, studies are drawn from a superpopulation of populations with differing average effect sizes, so effect-size heterogeneity is an inherent feature of the model rather than a deficiency of the literature (Field and Gillett, 2010).

This pattern supports the use of the three-level model rather than a standard random-effects model, which would have combined these two separate sources of variation (Assink and Wibbelink, 2016), and is the reason for conducting the subgroup analyses that follow.

Table 3. Variance components and heterogeneity decomposition from the three-level mixed-effects model.

Variance component	Variance estimate (τ^2)	Proportion of variance (I^2)
Level 2 (within-study)	0.0048	35.90%
Level 3 (between-study)	0.0059	43.76%
Total true heterogeneity	0.0107	79.66%

Note. Level 1 sampling variance $V = SE^2$. Estimated by ML (L-BFGS-B). $Q(303) = 1370.71$, $p < .001$.



[Figure 4. Decomposition of total true heterogeneity into within-study (Level 2) and between-study (Level 3) variance components. Source: Author]

3.4. Subgroup and moderator analyses

To address RQ2 and RQ3, we estimated the pooled effect within subgroups defined by predictor type, outcome type, and moderator category. Table 4 reports the effect-size count, pooled correlation, 95% CI, 95% prediction interval, τ^2 , and I^2 for each subgroup, with the omnibus between-subgroup test.

3.4.1. Predictor type

Differences across the four predictor types were not significant ($Q_m(3) = 2.93$, $p = .403$). The pooled effect was highest for AI Disclosure and Transparency ($r = 0.218$, 95% CI [0.192, 0.245]) and lowest for Explainability ($r = 0.188$, 95% CI [0.159, 0.217]), but confidence intervals overlapped and the difference did not exceed chance. At the level of the whole literature, a simple statement that AI is present produces effects statistically indistinguishable from a rich technical explanation.

The act of being transparent appears to matter more than the depth of transparency, because consumers register that a firm chose to reveal AI rather than studying the technical content of the revelation (Liu, 2021; Le et al., 2026).

3.4.2. Outcome type

Differences across outcomes were significant ($Q_m(3) = 8.64$, $p = .034$). The effect was strongest for Perceived System Characteristics ($r = 0.218$, 95% CI [0.194, 0.243]) and Attitude and Trust Outcomes ($r = 0.214$, 95% CI [0.187, 0.242]), weaker for Behavioral Intentions and Usage ($r = 0.179$, 95% CI [0.156, 0.202]), and weakest for Task Performance and Cognitive Outcomes ($r = 0.176$, 95% CI [0.142, 0.210]).

This gradient runs from perception to attitude to intention to behaviour, with the signal being strongest at the top and decreasing towards behaviour, reflecting the attitude-behaviour gap (Ajzen, 1991).

This is evident through the autonomous retail evidence: Transparency is a strong correlate of cognitive trust, whereas perceived risk is a strong deterrent against the final act of delegating a purchase to AI (Wang, 2026).

3.4.3. Moderator category

Differences across moderator categories were significant ($Q_m(4) = 10.92$, $p = .027$). The effect was greatest in the No Moderator or Direct subgroup ($r = 0.209$, 95% PI [0.035, 0.382]) and consistently reduced when boundary conditions were added: System and Interface Features ($r = 0.159$), Individual and Cultural Differences ($r = 0.158$), Context and Task Characteristics ($r = 0.160$).

The Methodological or Other subgroup produced a higher but unstable estimate ($r = 0.256$) based on only five effects with high heterogeneity ($I^2 = 88.30\%$).

The general pattern shows a drop of about 24% from 0.209 (direct estimations) to 0.159 (with cognitive load). We view this as evidence of cognitive load because when users struggle to navigate through an interface, when tasks are unfamiliar, or when they have culturally specific concerns, some of their attention is diverted from the task to the transparency cue, and the trust-related associations of transparency are attenuated.

Table 4. Subgroup and moderator analyses of the association between AI transparency-related predictors and consumer outcomes.

Subgroup	<i>k</i>	<i>r</i>	95% CI	95% PI	τ^2	I^2 (%)	Between-group test
Panel A. Predictor type							$Q_m(3) = 2.93, p = .403$
AI Disclosure and Transparency	81	0.218	[0.192, 0.245]	[0.023, 0.414]	0.009	76.69	
Performance, Capability and Other	108	0.198	[0.176, 0.220]	[0.035, 0.361]	0.007	72.83	
Trustworthiness and Anthropomorphism	53	0.191	[0.158, 0.224]	[-0.021, 0.403]	0.011	78.06	
Explainability (XAI)	62	0.188	[0.159, 0.217]	[0.011, 0.365]	0.008	72.94	
Panel B. Outcome type							$Q_m(3) = 8.64, p = .034$
Perceived System Characteristics	83	0.218	[0.194, 0.243]	[0.044, 0.393]	0.008	76.55	
Attitude and Trust Outcomes	95	0.214	[0.187, 0.242]	[0.00, 0.429]	0.011	78.47	
Behavioral Intentions and Usage	88	0.179	[0.156, 0.202]	[0.014, 0.344]	0.007	71.41	
Task Performance and Cognitive Outcomes	38	0.176	[0.142, 0.210]	[0.037, 0.315]	0.004	59.37	
Panel C. Moderator category							$Q_m(4) = 10.92, p = .027$
No Moderator / Direct	245	0.209	[0.194, 0.224]	[0.035, 0.382]	0.008	73.08	
Methodological / Other	5	0.256	[0.070, 0.442]	[-0.180, 0.692]	0.020	88.30	
Context and Task Characteristics	14	0.160	[0.082, 0.238]	[-0.105, 0.425]	0.014	81.24	
System and Interface Features	20	0.159	[0.116, 0.201]	[-0.017, 0.335]	0.007	75.89	
Individual and Cultural Differences	20	0.158	[0.112, 0.203]	[-0.018, 0.333]	0.007	73.90	

Note. *k* = effect sizes; *r* = pooled correlation; CI = confidence interval; PI = prediction interval; τ^2 = variance component; I^2 = heterogeneity proportion; Q_m = omnibus test. Estimates back-transformed from Fisher-*z*.

3.5. Publication bias

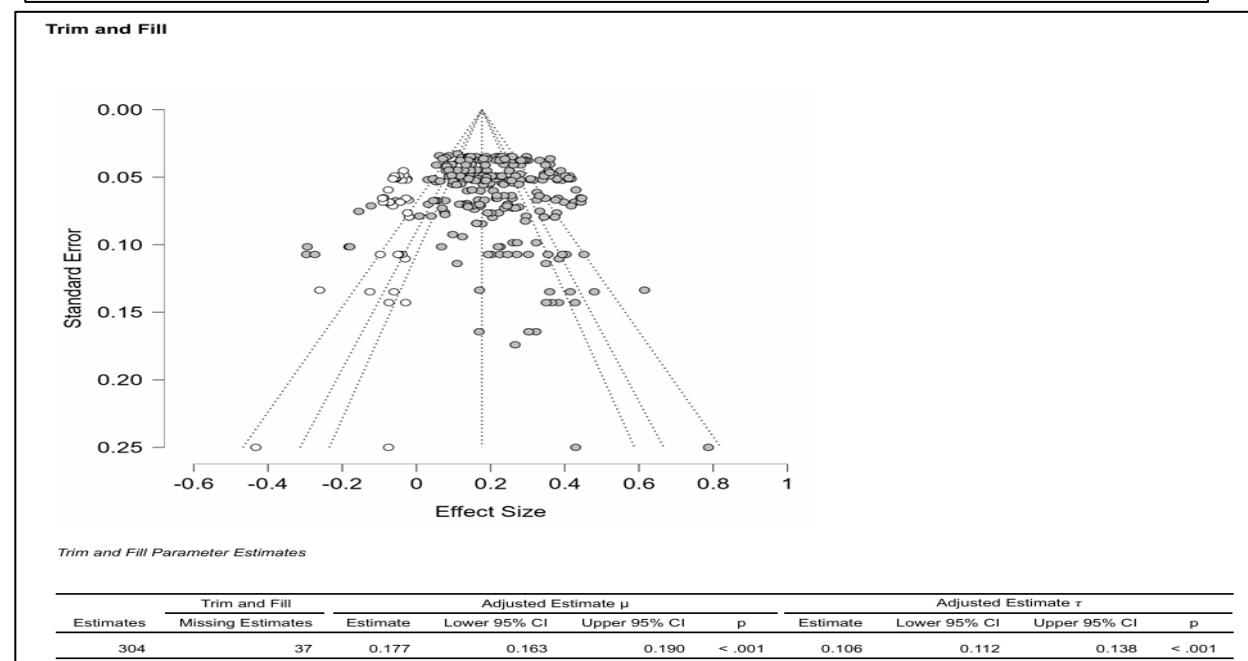
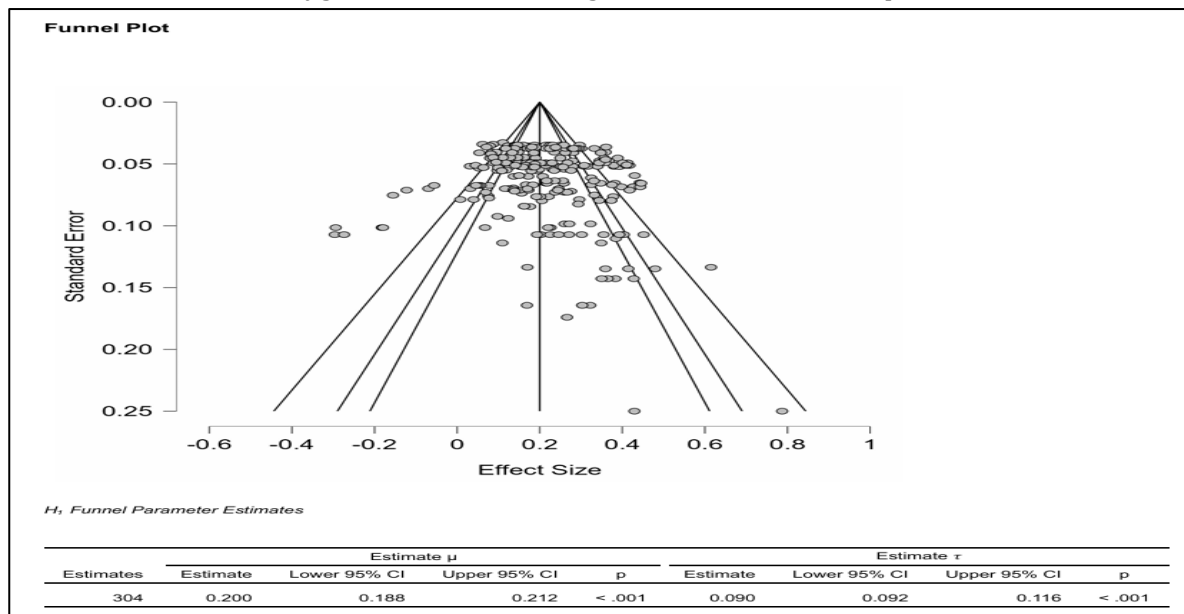
The publication-bias diagnostics are shown in Figures 5A and 5B. Visual examination of the original funnel plot (Figure 5A) indicated asymmetry, and the trim-and-fill analysis (Figure 5B) imputed 37 possible missing studies to the left side of the distribution.

A multi-method triangulation in Table 5 converged on the same conclusion. Egger's regression was significant ($z = 2.619$, $p = .009$, limit $\mu = 0.158$, 95% CI [0.125, 0.192]).

The weighted regression variant was marginal ($t = 1.769$, $df = 302$, $p = .078$), and the Begg and Mazumdar rank correlation was significant ($\tau = 0.106$, $p = .007$). Duval and Tweedie trim-and-fill estimated 37 imputed effects, and after imputation the adjusted estimate remained positive and significant ($r = 0.177$, 95% CI [0.163, 0.190], $p < .001$), with a CI overlapping the primary analysis.

The Rosenthal fail-safe N required 451,246 null studies to negate significance; because this method assumes missing studies have an effect of exactly zero it should be treated as a liberal bound, and the trim-and-fill estimate accordingly provides the more conservative substantive anchor.

The bias-corrected $r = 0.177$ is conservatively interpreted as the purified effect, after accounting for small-sample inflation and selective reporting. AI transparency does not exist as a magic lever, but it is a solid structural basis and has a hygienic function in the digital consumer relationship.



[Figure 5. Publication-bias diagnostics. (A) Funnel plot of the 304 effect sizes. (B) Trim-and-fill funnel plot showing the adjusted distribution after imputing 37 potentially missing studies. Fisher's z effect sizes are plotted against their standard errors. Source: Author]

Table 5. Multi-method publication-bias diagnostic triangulation.

Diagnostic test	Statistic	df	p value	Limit / adjusted estimate	95% CI on estimate	Conclusion
Egger's regression test	$z = 2.619$	n/a	.009	$\mu = 0.158$	[0.125, 0.192]	Significant asymmetry
Weighted regression test	$t = 1.769$	302	.078	$\mu = 0.164$	[0.126, 0.201]	Marginal asymmetry
Begg and Mazumdar rank test	$\tau = 0.106$	n/a	.007	n/a	n/a	Significant asymmetry
Trim-and-fill (Duval and Tweedie)	37 imputed	n/a	< .001	$r = 0.177$	[0.163, 0.190]	Adjusted effect remains positive
Rosenthal fail-safe N	n/a	n/a	n/a	451,246 nulls	n/a	File-drawer threat negligible

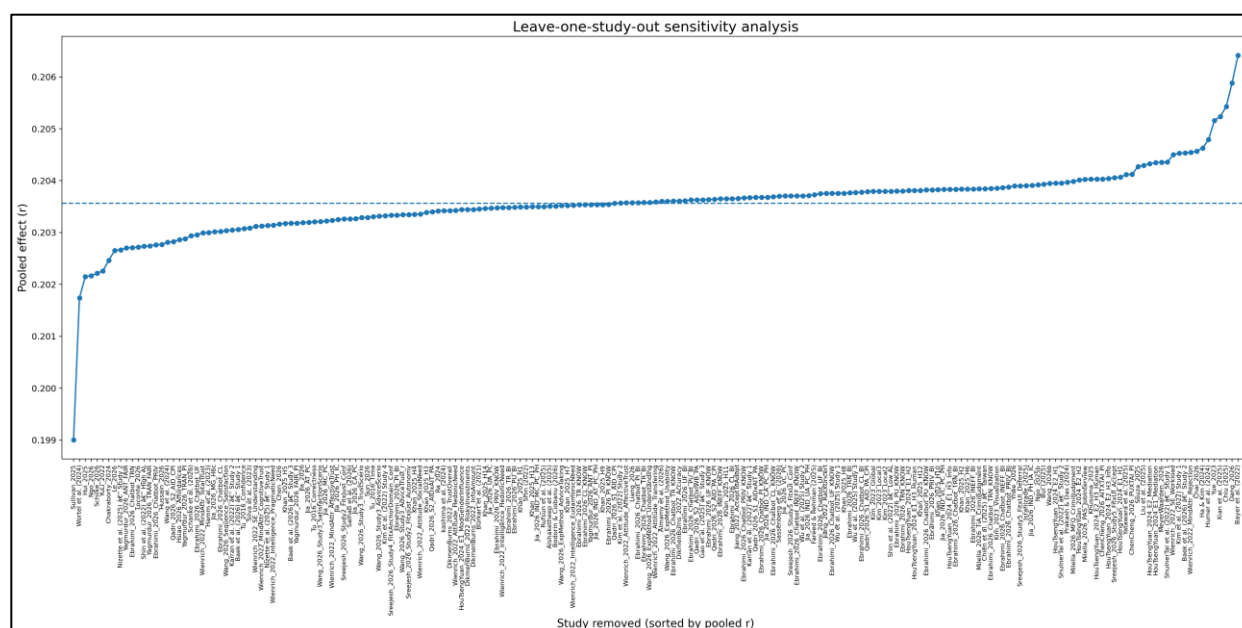
Note. Tests conducted on the Fisher-z metric; adjusted r values on the correlation scale. Convergence of three significant or marginal tests with substantial trim-and-fill adjustment and very large fail-safe N indicates mild bias whose correction does not alter the substantive conclusion.

3.6. Sensitivity and influence analyses

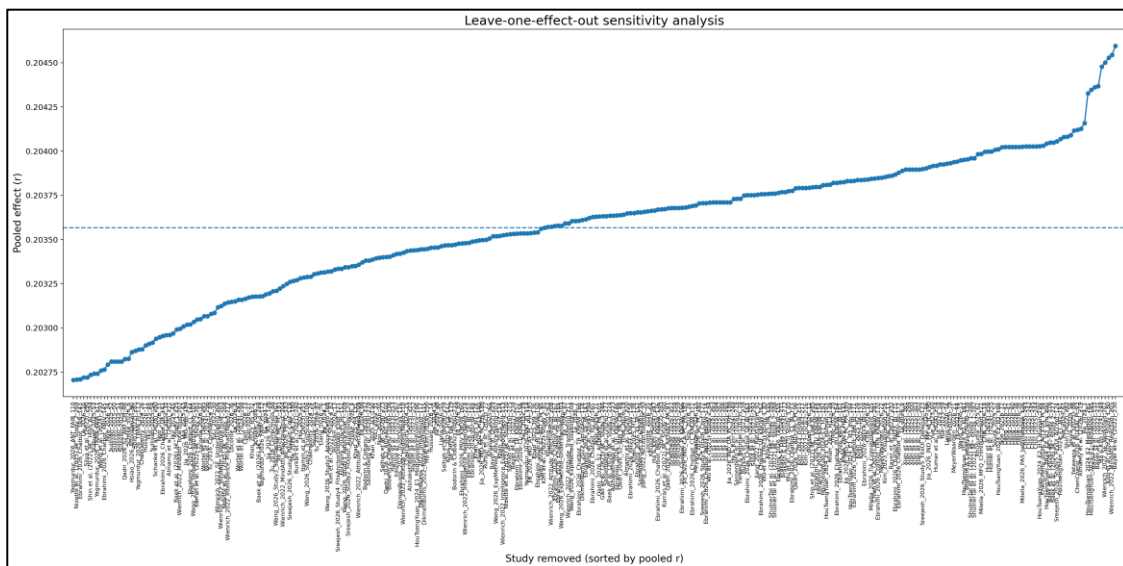
Leave-one-out analyses at both study and effect levels are reported in Table 6 and illustrated in Figures 6 and 7. Removing each study in turn, the pooled estimate ranged narrowly from $r = 0.1995$ to $r = 0.2071$, with no single study shifting it by more than $\Delta r = 0.0047$.

The largest shifts came from Sullivan and Weger (2025; $\Delta r = -0.0047$), Bayer et al. (2022; $\Delta r = +0.0029$), Dang and Li (2026; $\Delta r = +0.0023$), Wortel et al. (2024; $\Delta r = -0.0019$), and Chiu (2025; $\Delta r = +0.0018$). Removing one effect at a time, the estimate varied even more tightly between $r = 0.2033$ and $r = 0.2053$, with maximum single-effect shifts under $\Delta r = 0.0011$.

Figures 6 and 7 confirm that the overall conclusion does not depend on any single study or observation, and that the corridor of stability around $r = 0.20$ is reassuringly narrow.



[Figure 6. Leave-one-study-out sensitivity plot showing the recalculated pooled estimate as each study is removed. Source: Author]



[Figure 7. Leave-one-effect-out sensitivity plot showing the recalculated pooled estimate as each effect size is removed. Source: Author]

Table 6. Comprehensive sensitivity and influence diagnostics.

Diagnostic	Result
Panel A. Overall stability	
Full random-effects pooled r	0.2042, 95% CI [0.1902, 0.2181]
Leave-one-study-out range of pooled r	0.1995 to 0.2071
Maximum absolute change in pooled r (study level)	$\Delta r = 0.0047$
Leave-one-effect-out range of pooled r	0.2033 to 0.2053
Maximum absolute change in pooled r (effect level)	$\Delta r = 0.0011$
Panel B. Top 5 influential studies by Δr	
Sullivan and Weger (2025)	$\Delta r = -0.0047$; pooled r becomes 0.1995
Bayer et al. (2022)	$\Delta r = +0.0029$; pooled r becomes 0.2071
Dang and Li (2026)	$\Delta r = +0.0023$; pooled r becomes 0.2065
Wortel et al. (2024)	$\Delta r = -0.0019$; pooled r becomes 0.2023
Chiu (2025)	$\Delta r = +0.0018$; pooled r becomes 0.2060
Panel C. Top 5 influential effects by Δr	
Bayer et al. (2022): effect 231	$\Delta r = +0.0010$
Wienrich <i>et al.</i> , (2022) Mind-Attribution Rejection: effect 206	$\Delta r = +0.0010$
Kim et al. (2022) Study 1: effect 81	$\Delta r = +0.0010$
Wienrich <i>et al.</i> , (2022) Workload: effect 208	$\Delta r = +0.0010$
Wang et al. (2024): effect 294	$\Delta r = -0.0009$

Note. Panel A pooled r (0.2042) is the sensitivity-baseline reduced random-effects estimate. Δr is the change in pooled r when the named study or effect is removed. Stability corridor across all leave-one-out resamplings is approximately $r \in [0.20, 0.21]$.

3.7. Robustness summary across analytic specifications

Table 7 collects convergent estimates produced under different analytic choices. Across the three-level model, the single-level DerSimonian and Laird estimator, the trim-and-fill bias-adjusted estimate, the leave-one-out resamplings at both levels, and a restriction to the direct-r subset, the pooled correlation oscillates within a corridor of roughly $r \in [0.177, 0.207]$.

The lower bound ($r = 0.177$) is the bias-corrected floor; the upper ($r \approx 0.207$) is the within-study leave-one-out maximum. The substantive conclusion is invariant: AI transparency is positively and reliably associated with consumer judgments and behavior at a small-to-moderate magnitude by Cohen's benchmarks.

Table 7. Comprehensive robustness summary across analytic specifications.

Specification	Pooled r	95% CI	Notes
Three-level mixed-effects (primary)	0.205	[0.189, 0.219]	$k = 304$; ML, L-BFGS-B; <code>rma.mv()</code>
Single-level random-effects (DL)	0.200	[0.187, 0.214]	$k = 304$; DerSimonian and Laird estimator
Trim-and-fill adjusted (Duval and Tweedie)	0.177	[0.163, 0.190]	37 imputed effects on left side of funnel
Sensitivity-baseline pooled r	0.204	[0.190, 0.218]	Reduced specification used for leave-one-out
Leave-one-study-out minimum	0.1995	n/a	Sullivan and Weger (2025) removed
Leave-one-study-out maximum	0.2071	n/a	Bayer et al. (2022) removed
Leave-one-effect-out minimum	0.2033	n/a	Single observation removed
Leave-one-effect-out maximum	0.2053	n/a	Single observation removed
Direct- r subset only (sensitivity)	≈ 0.20	n/a	Within 0.01 of full-sample value
Egger's regression limit estimate	0.158	[0.125, 0.192]	Asymmetry-adjusted lower bound

Note. Values on the Pearson r scale, back-transformed from Fisher's z . The robustness corridor of $r \in [0.18, 0.21]$ confirms the qualitative conclusion is invariant to model specification, bias correction, and sample restriction.

4. Discussion

4.1. Key findings

Our findings show a clear positive correlation between AI transparency-related factors and consumer judgments and behaviour ($r = 0.205$). The three-level estimate ($r = 0.200$) was very similar to the single-level estimate, suggesting that the result is not sensitive to the modelling of statistical dependency. The finding was also rather robust, with little change after publication-bias adjustment ($r = 0.177$).

Variance decomposition revealed that the true heterogeneity was split between between-study differences (43.76%) and within-study differences (35.90%), indicating that there is variation in transparency effects across studies and specific outcomes. The prediction interval (roughly 0 to 0.38) also suggests that the effect of transparency is context and audience dependent.

These negative effects were not widespread, but rather focused on rejection, workload and some anthropomorphic conditions. Therefore, the answer to RQ1 is that the visibility of AI is generally associated with more positive consumer outcomes, but the strength of the impact is dependent on the situation.

One notable finding is related to RQ2. There was no significant difference in the effects of the four predictor types, and thus, simple disclosure, explainability mechanisms, trustworthiness cues, and performance signals generated statistically similar effects.

Strikingly, in contrast to often-held viewpoints in explainable AI research that providing detailed explanations is superior to simple disclosures, consumers seem mainly interested in the mere presence of a transparency mechanism.

Transparency is a heuristic cue of openness, not a way to evaluate technical aspects. This would align with Liu's (2021) notion of placebo transparency and Le et al. (2026), who discovered that perceptions of process credibility largely mediate disclosure effects.

At the outcome level, there were significant differences. The effects were strongest for perceived system characteristics and weaker for attitudes and behavioural intentions, and weakest for task performance. The pattern is similar to the attitude-behaviour gap (Ajzen, 1991) and research on multi-stage trust formation (Wang, 2026; Liang, 2026; Sun et al., 2025).

Where there is transparency, perceptions and attitudes are reliably improved, but this does not always translate into action without at least some other safeguards, such as risk reduction and human oversight. Moderator analysis was used to answer RQ3. The effect was significantly reduced with the addition of the contextual, individual or interface boundary conditions (from 0.209 to 0.159).

One possible explanation is that increasing contextual and interface complexity elevates cognitive load, reducing consumers' reliance on transparency cues. This explanation is consistent with Ngo's (2025) inverted-U model of transparency and trust, and with Chen and Ahmed's (2026) study results, which show that poorly structured transparency is associated with information overload.

Finally, Schilke and Reimann (2025) found that disclosing AI use can reduce trust by undermining perceived legitimacy, seemingly contradicting our positive overall estimate.

Their impact, however, relates to individuals disclosing their use of personal AI, while our positive results primarily relate to transparency at the brand and system levels. The same term, 'disclosure,' can thus refer to two different psychological processes, a distinction that needs to be tested empirically in the future.

4.2. Limitations and future research directions

It is important for us to acknowledge the several limitations of this study.

First, the meta-analysis estimates associations, not causal effects, because it combines correlational and experimental evidence. And hence, one promising avenue is to investigate the effects of transparency through future experimental and longitudinal studies to determine whether transparency effects are maintained or change over time (Le et al., 2026).

Secondly, the majority of studies used self-report instruments, which are subject to social desirability bias and may suffer from the intention-behaviour gap, which implies that more robust evidence would come from behavioural trace data, such as clicks, completions, and repeat-use records.

Third, although there is coverage of East Asia, North America, and Europe, much of the Global South remains underrepresented. This highlights the need for studies that address the unique dynamics of these cultures and people. Fourth, there is significant heterogeneity in AI transparency, ranging from simple labels to complex explainability interfaces ($I^2 = 79.66\%$).

Finally, moderator analyses were done with broad categories rather than continuous measures. Explanation richness, consumer control, and AI aversion are among the factors that should be directly studied in future meta-regressions. Despite this, the AI transparency aspect of the consumer-AI relationship seems to be a solid, modest and well-regulated element of the interactions (Le et al., 2026).

5. Theoretical Contributions

This study makes four theoretical contributions to the AI transparency literature. First, it advances understanding of the psychological mechanism through which AI transparency shapes consumer responses.

Existing research generally assumes that greater transparency improves consumer evaluations because richer information enables better understanding of AI systems. However, our findings suggest a different mechanism.

The absence of statistically significant differences between simple disclosure labels and detailed explainability suggests that consumers may not systematically evaluate the informational content of transparency.

Drawing on the Heuristic Systematic Model (Chaiken, 1980), the Elaboration Likelihood Model (Petty and Cacioppo, 1986), signalling theory (Spence, 1973), and cue utilisation theory (Olson and Jacoby, 1972), we argue that consumers primarily interpret transparency as a heuristic cue signalling organisational openness, accountability, and trustworthiness rather than as detailed technical information.

This extends prior conceptualisations of AI transparency and is consistent with Liu's (2021) notion of placebo transparency, suggesting that the perception of openness often outweighs the informational richness of the explanation.

Second, our findings identify the stage at which transparency exerts its greatest influence. Transparency exhibits its strongest association with perceived system characteristics and trust, followed by progressively weaker associations with behavioural intentions and task performance.

This pattern suggests that transparency primarily shapes early cognitive evaluations before consumers increasingly integrate other considerations, including perceived risk, convenience, economic value, and habit, into behavioural decisions. Transparency should therefore be viewed as a foundational trust building mechanism that shapes early consumer judgments rather than as a direct driver of behavioural outcomes.

Third, our results demonstrate that transparency is contingent rather than universally beneficial. Its influence weakens under more demanding contextual, individual, and interface conditions, indicating that transparency is one of several cues consumers integrate when evaluating AI systems. This identifies important theoretical boundary conditions that explain when transparency is likely to be amplified or attenuated.

Finally, our findings introduce the concept of diminishing informational returns in AI transparency. The absence of meaningful differences between simple disclosure and detailed explainability suggests that additional explanatory detail does not necessarily produce proportionally greater consumer trust.

Instead, consumers appear to derive sufficient reassurance from relatively simple disclosure cues, indicating that transparency is effective because it signals organisational openness, accountability, and responsible AI governance rather than because it provides increasingly detailed technical information.

Taken together, these contributions reconceptualise AI transparency as a contingent psychological cue that primarily shapes early consumer judgments through heuristic cue processing rather than through increasingly detailed information provision.

6. Practical Implications

While these underlying primary studies are cross-sectional, the overall robustness of the relationship across 82 studies provides actionable heuristics for practitioners. The synthesis is about when brands become visible to consumers via AI, and this has implications in everyday disclosure moments, whether it be a label, an opening line, a campaign acknowledgement, or a verified badge.

6.1. Disclosure labels on social media posts

The most common form of AI transparency is a label indicating that content was generated or curated by AI. The type of predictor does not make a significant difference in the effect, so the label itself carries the psychological weight.

But direction is determined by what the label suggests the AI did.

In the study conducted by Wu et al. (2025), they discovered that labelling a post as chosen by the AI for the consumer led to higher word-of-mouth intent as it aligned with the machine heuristic of objective optimisation.

In contrast, brand attitude was negatively affected when it was attributed to AI ($M = 4.80$ vs 5.38). Brands should therefore focus on the functional aspects of AI, emphasising data-based curation language instead of highlighting machine authorship, in accordance with the EU AI Act.

6.2. AI-powered virtual influencers

Disclosure is particularly important in the context of virtual and AI influencers, because consumers develop relationships with the persona itself. As shown by Kim (2026), advertising scepticism is higher in ads that include AI models ($M = 4.18$) than in ads that include human models ($M = 3.34$), which is associated with significantly lower brand advocacy, word-of-mouth, and purchase intention.

Our outcome analysis suggests that the primary harm lies in attitudes and authenticity, not in immediate transactions, which is why this is a long-term brand-equity risk. Brands, hence, need to align disclosure with social responsibility signals and human oversight.

6.3. Chatbots and identity disclosure

Disclosing the bot's machine identity is when consumers decide to trust advice or reveal information. Cheng et al. (2022) demonstrate that this reveal enhances friendliness-trust but reduces empathy-trust, as empathy is a human-coded quality that is revealed as simulated. If the chatbot begins warmly and reveals that it is an AI, the user may lose trust, but if it reveals its AI identity first and focuses on accuracy, it will better cater to the user's expectations.

The results of this study also align with those of Khusuma et al. (2025), who found that although warmth is a crucial element of trust, it is not sufficient on its own.

6.4. Proactive versus reactive framing

Le et al. (2026) demonstrate that there are different effects when the transparency signals a firm's values and when it is perceived as simply required (proactive versus reactive framing).

Proactive framing encourages benevolent motive inferences, whereas reactive framing raises scepticism and makes disclosure quality more significant. Pairing disclosure with EU AI Act compliance signalling is associated with higher perceived credibility of the creation process. Table 8 synthesises these dimensions into a practitioner decision matrix for AI transparency design.

Table 8. A practitioner decision matrix for AI transparency design, derived from the meta-analytic findings.

Dimension	Lower-stakes / simpler contexts	Higher-stakes / complex contexts
Disclosure depth	Simple visible label is sufficient; heuristic signalling dominates	Layered transparency with process detail and human oversight
Functional framing	Curation or optimization framing fits the machine heuristic	Creative or relational framing needs authenticity and oversight cues
User expertise	Process-level transparency suffices for novice users	Decision-logic transparency suits experts; avoid overload for novices
Compliance framing	Proactive, values-based framing builds a trust buffer	Reactive framing demands high disclosure quality and verification badges
Associated effect	Associated effect near upper bound, around $r = 0.21$	Associated effect near lower bound, around $r = 0.16$, potentially due to cognitive load

Note. Matrix translates the pooled and subgroup estimates into design guidance. The effect-size range reflects the observed oscillation between roughly 0.158 and 0.218 across subgroups.

7. Conclusions

The meta-analysis results indicate a consistent positive association between AI transparency and consumer judgments and behaviours. While further experimental and longitudinal studies are needed to establish causal directionality and move beyond cross-sectional surveys, the evidence underscores that transparency is a robust correlate of consumer trust rather than an optional add-on.

The observed effect is small but very consistent, and also resistant to the exclusion of any single study or effect. Transparency is not a panacea, because its effects are diluted when the task is complex or the interface is unfamiliar. But when used along with manned control, contextual explanations, and credible compliance signalling, it is associated with higher trust.

With disclosure requirements growing worldwide, especially under Article 50 of the EU AI Act, transparency could be a key factor in restoring trust structures crucial for sustainable human-AI consumer relations.

Data Availability

The dataset generated and analyzed during this study, along with the corresponding analytical outputs, can be accessed at <https://doi.org/10.6084/m9.figshare.32838230>.

Funding Statement

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

References

- Ajzen, I. (1991) 'The theory of planned behavior', *Organizational Behavior and Human Decision Processes*, 50(2), pp. 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Assink, M. and Wibbelink, C.J. (2016) 'Fitting three-level meta-analytic models in R: A step-by-step tutorial', *The Quantitative Methods for Psychology*, 12(3), pp. 154–174. <https://doi.org/10.20982/tqmp.12.3.p154>
- Atf, Z. and Lewis, P.R. (2025) 'Is trust correlated with explainability in AI? A meta-analysis', *IEEE Transactions on Technology and Society*. <https://doi.org/10.1109/TTS.2025.3558448>
- Baek, T.H., Kim, J. and Kim, J.H. (2026) 'Effect of disclosing AI-generated content on prosocial advertising evaluation', *International Journal of Advertising*, 45(1), pp. 171–192. <https://doi.org/10.1080/02650487.2024.2401319>
- Balaskas, S. (2026) 'Transparency by design: A narrative synthesis of AI disclosure, explainability, and trust in consumer-facing FinTech', *FinTech*, 5(2), p. 41. <https://doi.org/10.3390/fintech5020041>
- Bayer, S., Gimpel, H. and Markgraf, M. (2022) 'The role of domain expertise in trusting and following explainable AI decision support systems', *Journal of Decision Systems*, 32(1), pp. 110–138. <https://doi.org/10.1080/12460125.2021.1958505>
- Begg, C.B. and Mazumdar, M. (1994) 'Operating characteristics of a rank correlation test for publication bias', *Biometrics*, 50(4), pp. 1088–1101. <https://doi.org/10.2307/2533446>
- Chen, B. and Ahmed, S. (2026) 'Is transparency always good? Dual cognitive responses to different transparency dimensions and their effects on users' intention to use', *International Journal of Human-Computer Interaction*, pp. 1–17. <https://doi.org/10.1080/10447318.2026.2623213>
- Cheng, X., Bao, Y., Zarifis, A., Gong, W. and Mou, J. (2022) 'Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure', *Internet Research*, 32(2), pp. 496–517. <https://doi.org/10.1108/INTR-08-2020-0460>
- Cheung, M.W.-L. (2014) 'Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach', *Psychological Methods*, 19(2), pp. 211–229. <https://doi.org/10.1037/a0032968>
- Chiu, M.L. (2025) 'Exploring user awareness and perceived usefulness of generative AI in higher education: The moderating role of trust', *Education and Information Technologies*, 30(15), pp. 21317–21351. <https://doi.org/10.1007/s10639-025-13612-7>
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. 2nd edn. Lawrence Erlbaum.

- Dang, Q. and Li, G. (2026) 'Unveiling trust in AI: The interplay of antecedents, consequences, and cultural dynamics', *AI and Society*, 41(1), pp. 669–692. <https://doi.org/10.1007/s00146-025-02477-6>
- DerSimonian, R. and Laird, N. (1986) 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, 7(3), pp. 177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
- Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015) 'Algorithm aversion: people erroneously avoid algorithms after seeing them err', *Journal of Experimental Psychology: General*, 144(1), p. 114. <https://doi.org/10.1037/xge0000033>
- Duval, S. and Tweedie, R. (2000) 'Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis', *Biometrics*, 56(2), pp. 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Ebrahimi, P., Hoffmann, S. and Schneider, J. (2026) 'Opening the black box: How reasoning-enabled AI agents influence user perceptions and behavior in sustainable consumption', *International Journal of Information Management*, 90, p. 103075. <https://doi.org/10.1016/j.ijinfomgt.2026.103075>
- Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) 'Bias in meta-analysis detected by a simple, graphical test', *BMJ*, 315(7109), pp. 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Field, A.P. and Gillett, R. (2010) 'How to do a meta-analysis', *British Journal of Mathematical and Statistical Psychology*, 63(3), pp. 665–694. <https://doi.org/10.1348/000711010X502733>
- Gelbrich, K., Roschk, H., Miederer, S. and Kerath, A. (2026) 'Automated versus human agents: A meta-analysis of customer responses to robots, chatbots, and algorithms and their contingencies', *Journal of Marketing*, 90(2), pp. 1–26. <https://doi.org/10.1177/00222429251344139>
- Grigsby, J.L., Michelsen, M. and Zamudio, C. (2025) 'Service ads in the era of generative AI: Disclosures, trust, and intangibility', *Journal of Retailing and Consumer Services*, 84, p. 104231. <https://doi.org/10.1016/j.jretconser.2025.104231>
- Haresamudram, K., Larsson, S. and Heintz, F. (2023) 'Three levels of AI transparency', *Computer*, 56(2), pp. 93–100. <https://doi.org/10.1109/MC.2022.3213181>
- Hedges, L.V. (1992) 'Meta-analysis', *Journal of Educational Statistics*, 17(4), pp. 279–296. <https://doi.org/10.2307/1165125>
- Higgins, J.P., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003) 'Measuring inconsistency in meta-analyses', *BMJ*, 327(7414), pp. 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Hou, T.Y., Tseng, Y.C. and Yuan, C.W.T. (2024) 'Is this AI sexist? The effects of a biased AI's anthropomorphic appearance and explainability on users' bias perceptions and trust', *International Journal of Information Management*, 76, p. 102775. <https://doi.org/10.1016/j.ijinfomgt.2024.102775>
- Jia, W., Mehta, A.M., Han, C., Asif, M. and Shahzad, M.F. (2026) 'The dark side of AI transparency: investigating AI-induced anxiety, technostress, and consumer satisfaction through a hybrid SEM-ANN approach', *BMC Psychology*. <https://doi.org/10.1186/s40359-026-04147-8>
- Khan, S.M.F.A. and Shehawy, Y.M. (2025) 'Perceived AI Consumer-Driven Decision Integrity: Assessing Mediating Effect of Cognitive Load and Response Bias', *Technologies*, 13(8), p. 374. <https://doi.org/10.3390/technologies13080374>
- Khusuma, M.F., Wilbert, D. and Kristian, W. (2025) 'The effect of information quality, human-like language style, privacy concerns, and AI-based chatbot interaction on consumer trust in the e-commerce world', *ICIMCIS 2025*, p. 514. <https://doi.org/10.1109/ICIMCIS68501.2025.11327439>
- Kim, S. (2026) 'AI vs. human models in fashion brand advertising: A schema theoretical perspective on consumer responses', *Journal of Retailing and Consumer Services*, 92, p. 104820. <https://doi.org/10.1016/j.jretconser.2025.104820>
- Kim, T.W., Jiang, L., Duhachek, A., Lee, H. and Garvey, A. (2022) 'Do you mind if I ask you a personal question? How AI service agents alter consumer self-disclosure', *Journal of Service Research*, 25(4), pp. 649–666. <https://doi.org/10.1177/10946705221120232>
- Konstantopoulos, S. (2011) 'Fixed effects and variance components estimation in three-level meta-analysis', *Research Synthesis Methods*, 2(1), pp. 61–76. <https://doi.org/10.1002/jrsm.35>
- Le, K.B.Q., Khan, H., Li, F. and Kunz, W.H. (2026) 'More Than Saying "It's AI": How Role Disclosure Transparency in AI-Generated Ads Influences Persuasion', *Psychology and Marketing*. <https://doi.org/10.1002/mar.70175>
- Liang, M. (2026) 'How marketing mechanisms build trust in AI recommendation systems: A serial mediation model in live-streaming commerce', *Cogent Business and Management*, 13(1), p. 2658868.

<https://doi.org/10.1080/23311975.2026.2658868>

Liu, B. (2021) 'In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human-AI interaction', *Journal of Computer-Mediated Communication*, 26(6), pp. 384–402. <https://doi.org/10.1093/jcmc/zmab013>

Logg, J.M., Minson, J.A. and Moore, D.A. (2019) 'Algorithm appreciation: People prefer algorithmic to human judgment', *Organizational Behavior and Human Decision Processes*, 151, pp. 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>

Longoni, C., Bonezzi, A. and Morewedge, C.K. (2019) 'Resistance to medical artificial intelligence', *Journal of Consumer Research*, 46(4), pp. 629–650. <https://doi.org/10.1093/jcr/ucz013>

Mazzù, M.F., Andria, A., Baccelloni, A. and De Angelis, M. (2025) 'The impact of humans vs. AI recommendation on consumer reactions to products exposure', *Electronic Commerce Research*, pp. 1–41. <https://doi.org/10.1007/s10660-025-10048-6>

Ngo, V.M. (2025) 'The AI transparency dilemma: when more is less for trust and adoption', *Information Discovery and Delivery*. <https://doi.org/10.1108/IDD-03-2025-0056>

Nizette, F., Hammedi, W., van Riel, A.C. and Steils, N. (2025) 'Why should I trust you? Influence of explanation design on consumer behavior in AI-based services', *Journal of Service Management*, 36(1), pp. 50–74. <https://doi.org/10.1108/JOSM-05-2024-0223>

Olson, J.C. and Jacoby, J. (1972) 'Cue utilization in the quality perception process', *ACR Special Volumes*.

Page, M.J. et al. (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>

Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R. and Rushton, L. (2006) 'Comparison of two methods to detect publication bias in meta-analysis', *JAMA*, 295(6), pp. 676–680. <https://doi.org/10.1001/jama.295.6.676>

Peterson, R.A. and Brown, S.P. (2005) 'On the use of beta coefficients in meta-analysis', *Journal of Applied Psychology*, 90(1), pp. 175–181. <https://doi.org/10.1037/0021-9010.90.1.175>

R Core Team (2023) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141. <https://doi.org/10.1007/s11747-019-00710-5>

Richard, F.D., Bond Jr, C.F. and Stokes-Zoota, J.J. (2003) 'One hundred years of social psychology quantitatively described', *Review of General Psychology*, 7(4), pp. 331–363. <https://doi.org/10.1037/1089-2680.7.4.331>

Rosenthal, R. (1979) 'The file drawer problem and tolerance for null results', *Psychological Bulletin*, 86(3), pp. 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>

Schilke, O. and Reimann, M. (2025) 'The transparency dilemma: How AI disclosure erodes trust', *Organizational Behavior and Human Decision Processes*, 188, p. 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>

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>

Spence, M. (1973) 'Job market signaling', *The Quarterly Journal of Economics*, 87(3), pp. 355–374. <https://doi.org/10.2307/1882010>

Sullivan, V. and Weger, K. (2025) 'Transparency and explainability in AI-assisted decision making: Effects on trust, perceived reliability, confidence, and ease of understanding', in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 69(1). Los Angeles, CA: SAGE Publications, pp. 1827–1833. <https://doi.org/10.1177/10711813251369473>

Sun, Y., Liao, M.M., Sundar, S.S. and Walther, J.B. (2025) 'Does Transparency Matter When an AI System Meets Performance Expectations? An Experiment with an Online Dating Site', *Computers in Human Behavior*, p. 108875. <https://doi.org/10.1016/j.chb.2025.108875>

Sweller, J. (1988) 'Cognitive load during problem solving: Effects on learning', *Cognitive Science*, 12(2), pp. 257–285. https://doi.org/10.1207/s15516709cog1202_4

Viechtbauer, W. (2010) 'Conducting meta-analyses in R with the metafor package', *Journal of Statistical Software*, 36, pp. 1–48. <https://doi.org/10.18637/jss.v036.i03>

Wang, D., Bian, C. and Chen, G. (2024) 'Using explainable AI to unravel classroom dialogue analysis: Effects of explanations on teachers' trust, technology acceptance and cognitive load', *British Journal of Educational Technology*, 55(6), pp. 2530–2556. <https://doi.org/10.1111/bjet.13466>

Wang, J. (2026) 'Letting the bot decide: Trust and risk perceptions in autonomous retail purchase decisions', *Journal of Retailing and Consumer Services*, 93, p. 104867. <https://doi.org/10.1016/j.jretconser.2026.104867>

Wienrich, C., Carolus, A., Roth-Isigkeit, D. and Hotho, A. (2022) 'Inhibitors and enablers to explainable AI success: a systematic examination of explanation complexity and individual characteristics', *Multimodal Technologies and Interaction*, 6(12), p. 106. <https://doi.org/10.3390/mti6120106>

Wortel, C., Vanwesenbeeck, I. and Tomas, F. (2024) 'Made with artificial intelligence: The effect of artificial intelligence disclosures in Instagram advertisements on consumer attitudes', *Emerging Media*, 2(3), pp. 547–570. <https://doi.org/10.1177/2752354324129209>

Wu, L., Doodoo, N.A. and Wen, T.J. (2025) 'Disclosing AI's involvement in advertising to consumers: A task-dependent perspective', *Journal of Advertising*, 54(1), pp. 20–38. <https://doi.org/10.1080/00913367.2024.2309929>

Yu, L. and Li, Y. (2022) 'Artificial intelligence decision-making transparency and employees' trust: The parallel multiple mediating effect of effectiveness and discomfort', *Behavioral Sciences*, 12(5), p. 127. <https://doi.org/10.3390/bs12050127>

Zehnle, M., Hildebrand, C. and Valenzuela, A. (2025) 'Not all AI is created equal: A meta-analysis revealing drivers of AI resistance across markets, methods, and time', *International Journal of Research in Marketing*, 42(3), pp. 729–751. <https://doi.org/10.1016/j.ijresmar.2025.02.005>

Zhang, L. and Hur, C. (2025) 'The impact of generative AI images on consumer attitudes in advertising', *Administrative Sciences*, 15(10), p. 395. <https://doi.org/10.3390/admsci15010014>