https://journals.insightpub.org/index.php/imj

Insight Management Journal

Volume 6, Issue 2, November 2026, pages xx-xx

The effect of AI-generated content on information credibility, consumer trust, and purchase decisions: the moderating role of AI-use transparency

Shinta Rahmani [1*], M. Ali Iqbal [1], Hery Derajad Wijaya [1], Sigit Setiawan [2], Mia Rahma Romadona [2]

[1] Universitas Mercu Buana, Jakarta, Indonesia
[2] Badan Riset dan Inovasi Nasional, Jakarta, Indonesia

Article info

Article history:
Submitted [24-08-2026]
Revised [13-10-2026]
Accepted [07-11-2026]

Keywords:
AI-generated content
AI disclosure
information credibility
platform trust
social commerce

A b s t r a c t

This study examines whether transparency about artificial intelligence helps or harms consumers in an emerging social commerce market. Prevailing experimental evidence, drawn largely from mature Western markets, reports that disclosing AI involvement lowers credibility and trust; this study tests whether that finding holds in Indonesia and traces how AI awareness and AI-use disclosure reach purchase decisions through information credibility and two distinct dimensions of trust. Data were collected through a screened online panel survey of active social commerce consumers in six Indonesian metropolitan areas who shop mainly on Shopee and TikTok Shop, and were analysed with partial least squares structural equation modelling using a bootstrapping procedure. Contrary to the prevailing evidence, both AI awareness and AI-use disclosure increased perceived information credibility, with disclosure being the stronger antecedent, while their interaction was not significant, indicating an additive rather than a buffering relationship. Information credibility strongly increased both seller and information trust and platform trust, yet only platform trust had a significant effect on purchase decisions, and only the indirect pathways running through platform trust were significant. The study recommends that sellers treat voluntary AI disclosure as a low-cost trust-building instrument rather than a compliance cost, that platforms adopt uniform platform-issued provenance labelling supported by visible structural assurance, and that regulators pair disclosure mandates with verification infrastructure and digital literacy programmes.

Introduction

Generative artificial intelligence has moved from the margins of digital commerce to its infrastructure. Product descriptions, review summaries, live-stream scripts, tr anslated customer feedback and promotional imagery are now routinely produced or refined by large language models, and the resulting content is frequently indistinguishable from human authorship (Campbell et al., 2022). For sellers, especially the micro and small enterprises that populate social commerce marketplaces, this represents a genuine reduction in the cost of professional-quality communication. For consumers, it introduces an information asymmetry of a novel kind: the artefact on which a purchase judgement rests may no longer originate in anyone's actual experience of the product.

The scholarly response to this asymmetry has converged with unusual speed. A series of experiments has established that labelling content as AI-generated reduces perceived authenticity, brand attitude, ad credibility and trust in the organisation behind the message (Arango et al., 2023); (Baek et al., 2026a); (Brüns & Meißner, 2024); (Kirk & Givi, 2025); (Koning & Voorveld, 2025). The mechanism most often invoked is persuasion knowledge: a disclosure cue alerts consumers to a persuasion attempt and activates defensive processing (Eisend & Tarrahi, 2022). A parallel line of work identifies perceived authenticity and moral disgust as the mediating channels through which AI authorship damages consumer response (Kirk & Givi, 2025). The practical implication drawn from this body of evidence is straightforward and uncomfortable: transparency about AI use carries a commercial penalty, which places regulatory disclosure mandates in direct tension with seller interests.

Three features of that evidence base warrant scrutiny before the conclusion is generalised. First, it is overwhelmingly experimental and single-exposure. Participants are shown a stimulus, told it was

*Corresponding author
Email address: Shinta Rahmani (shinta.rahmani@mercubuana.ac.id)
ISSN: 2774-1737 DOI: 10.47065/imj.v6i2.589

machine-produced, and asked to evaluate it immediately. This design isolates a situational suspicion response, which is not the same psychological state as the chronic, background awareness that AI pervades the content environment. Second, it is geographically concentrated in North America and Western Europe, where baseline institutional trust, regulatory maturity and the credibility of the human alternative all differ substantially from conditions in emerging markets. Third, the outcome variables are typically attitudinal (ad attitude, brand authenticity, willingness to engage) rather than the trust structures and purchase decisions that determine commercial consequence in a marketplace setting.

Indonesia offers an instructive test case. It is Southeast Asia's largest digital retail market, and social commerce, defined here as purchasing embedded in social feeds and live streams on TikTok Shop, Shopee and Instagram Shopping, accounts for a substantial and growing share of transactions, with sector gross merchandise value projected to approach USD 22 billion by 2028 (Campaign Asia-Pacific, 2025). Consumers in this market operate in an information environment saturated with paid endorsement, incentivised reviews and undisclosed affiliate arrangements. The comparison class against which an AI-assisted product description is judged is therefore not a disinterested human account but a commercially motivated one. Whether AI involvement reads as a loss of authenticity or as a reduction in vested interest is, in that context, an open empirical question rather than a settled one.

Against this background, the present study departs from previous research in four specific respects. First, the studies that report an AI-disclosure penalty manipulate disclosure experimentally and measure immediate reactions to a single stimulus (Arango et al., 2023); (Baek et al., 2026b); (Brüns & Meißner, 2024); (Kirk & Givi, 2025); (Koning & Voorveld, 2025), whereas this study measures AI awareness as an enduring perception of the marketplace and AI-use disclosure as a perceived feature of the content that consumers actually encounter. Second, those studies draw on North American and Western European samples, whereas this study surveys active social commerce consumers in an emerging market in which baseline trust in commercial content is comparatively low. Third, previous studies stop at attitudinal outcomes such as ad attitude, brand authenticity and engagement intention (Brüns & Meißner, 2024); Kirkby et al., 2023), whereas this study follows the effect through information credibility and two distinct trust dimensions to the purchase decision itself. Fourth, few studies have tested whether the effect of disclosure depends on how aware consumers already are of AI involvement, so the buffering prediction derived from persuasion-knowledge accounts has been assumed rather than tested (Eisend & Tarrahi, 2022) this study tests that interaction directly.

A second limitation of the existing literature concerns the treatment of trust. Studies commonly model trust as a unitary construct or as a binary trust/distrust outcome. Research on electronic commerce has long distinguished between trust in the exchange partner and institution-based trust anchored in the structural assurances a marketplace provides (McKnight et al., 2002); (Wang et al., 2022). These dimensions are formed differently, respond to different cues, and plausibly differ in how they transmit credibility judgements into purchase behaviour. Collapsing them yields an average that can obscure a real functional division of labour, a concern that is acute in social commerce, where seller identity is often ephemeral and the platform is the only durable party to the transaction (Hajli, 2015); (Wang et al., 2022)

This study addresses both gaps. It asks four questions. (1) Does consumer awareness that AI is involved in social commerce content raise or lower perceived information credibility in an (Wang et al., 2022) own right? (3) Does disclosure moderate the awareness–credibility relationship, buffering the effect that experimental work would predict? (4) Through which trust dimension does credibility reach purchase decisions? These questions are examined using survey data from 406 active social commerce consumers and partial least squares structural equation modelling.

The study makes three contributions. Theoretically, it specifies boundary conditions for the emerging consensus on AI disclosure by showing that both awareness and disclosure operate as credibility assets in this market, and by separating trait-measured awareness from stimulus-induced disclosure, two constructs the literature has tended to treat interchangeably. It also contributes a well-powered null result on the awareness × disclosure interaction, which narrows the scope of persuasion-knowledge accounts in routine commercial settings. Empirically, it demonstrates that trust dimensions behave differently rather than merely covarying, and that AI transparency reaches purchasing through an institutional rather than an interpersonal route. Practically, it reframes voluntary disclosure from a reputational cost into a low-cost trust-building instrument and identifies platform-issued provenance labelling as the higher-leverage intervention.

Literature Review and Hypothesis Development

Competing Predictions for AI Awareness and Information Credibility

Information credibility refers to the extent to which a message is perceived as believable, accurate and worth acting upon, and is conventionally decomposed into source credibility and message credibility (Metzger et al., 2003); (Ohanian, 1990); (Li & See-To, 2024) In social commerce, where consumers cannot inspect goods or verify claims directly, credibility judgements about product descriptions and reviews are the principal evidentiary basis for the decision to transact (Cheung et al., 2008). Any cue that shifts those judgements therefore has downstream commercial consequence.

Two theoretical traditions make opposing predictions about how AI involvement functions as such a cue. The algorithm-aversion tradition holds that people discount machine judgement, particularly after observing machine error and particularly for tasks perceived to require uniquely human capacities such as empathy, subjective taste or the reporting of lived experience (Longoni et al., 2019). Applied to product reviews, the prediction is direct: content known or suspected to be machine-produced cannot embody genuine consumer experience and should therefore be discounted. The experimental disclosure literature is broadly consistent with this account (Brüns & Meißner, 2024); (Kirk & Givi, 2025).

The machine-heuristic tradition predicts the opposite under specifiable conditions. (Sundar, 2008) argues that cues signalling machine agency activate a heuristic in which machines are judged systematic, consistent and free of the ulterior motives attributed to human communicators; Sundar and Kim (2019) demonstrate that this heuristic can lead people to extend more trust to a machine than to a human counterpart. Algorithm-appreciation research shows that people prefer algorithmic to human judgement for tasks framed as objective (Logg et al., 2019), and evidence on AI recommendation acceptance indicates that trust rises when the machine's basis for judgement is legible (Kim & Kim, 2021).

Which prediction obtains depends on the comparison class and on what awareness measures. In a marketplace where the salient alternative to AI assistance is a paid endorsement or an incentivised review, machine involvement plausibly signals reduced vested interest rather than absent authenticity. Moreover, awareness measured as a chronic individual difference (recognising that AI is routinely involved in the content one encounters) is entangled with AI literacy and platform familiarity, and an interpretive framework of that kind reduces rather than amplifies evaluative uncertainty. Under emerging-market conditions of high technology optimism and comparatively low baseline trust in commercial content, the machine-heuristic account is the more plausible of the two, and it is the one this study adopts. The competing algorithm-aversion prediction is retained explicitly as a rival account, which the sign of the estimated coefficient adjudicates.

H1 : Consumer awareness of AI involvement in social commerce content is positively associated with perceived information credibility

AI-Use Disclosure as a Signal

Disclosure differs from awareness in that it originates with the seller or the platform rather than with the consumer. Signalling theory addresses precisely this asymmetry: when one party holds information the other cannot verify, costly and voluntary signals become the mechanism by which quality is communicated (Spence, 1973). In a market where non-disclosure carries no immediate penalty, a statement that AI assisted the production of a listing is voluntary and mildly costly, and therefore satisfies the conditions for a credible signal of seller type. What the consumer evaluates is not only the content but the disposition toward candour that the disclosure reveals.

This reading diverges from the persuasion-knowledge interpretation dominant in the experimental literature, in which disclosure functions as a warning cue that activates defensive processing (Eisend & Tarrahi, 2022) The divergence is partly a matter of what is disclosed. Disclosing that AI assisted in drafting, formatting or translating a listing is a different act from disclosing that an experience was fabricated, and the experimental literature has not consistently separated the two. (Kirkby et al., 2023a) find that the disclosure decision affects brand authenticity through channels other than consumer scepticism, which is compatible with a signalling rather than a warning interpretation. Where disclosure is read as institutional candour, its effect on credibility should be positive.

H2 : Explicit disclosure of AI use in social commerce content is positively associated with perceived information credibility

The Moderating Role of Disclosure

If disclosure functions as a warning cue, its effect should not be uniform across consumers. Persuasion-knowledge accounts predict an interaction: disclosure should matter most for consumers whose latent suspicion it activates and least for those already sceptical, producing an attenuating pattern

in which transparency buffers the credibility consequences of awareness (Baek et al., 2026b) report that disclosure effects on prosocial advertising evaluation operate through perceived ad credibility, implying that individual differences in prior AI knowledge should condition the response. This is the interaction the present study tests. Its theoretical interest lies in the fact that a null result would be informative rather than merely inconclusive: it would indicate that disclosure functions as a uniform institutional quality signal rather than as a differential activator of persuasion knowledge.

H3 : AI-use disclosure moderates the relationship between AI awareness and information credibility, attenuating the awareness effect at higher levels of disclosure

Credibility and the Two Dimensions of Trust

Trust in electronic commerce is not a single disposition. (McKnight et al., 2002) distinguish trusting beliefs about a specific exchange partner from institution-based trust grounded in structural assurances such as escrow, dispute resolution, seller sanctions, which operate independently of any particular counterparty. Meta-analytic evidence shows that both contribute to online transaction intention through partly separate routes. In social commerce these two dimensions map onto seller and information trust on the one hand and platform trust on the other, and there are strong reasons to expect them to be formed differently.

Trust-transfer theory explains why credibility is the substrate from which both are built (Wang et al., 2022) Buyers rarely meet sellers, cannot inspect goods, and have no interpersonal history on which to base an assessment; trust must therefore be inferred from the artefacts available, principally the product information itself. Credibility judgements transfer most readily to the proximate object of evaluation, namely the seller and the information, because the inferential distance is shortest. Transfer to the platform is weaker, because platform trust is anchored additionally in structural features that no individual listing can alter (Mayer et al., 1995) (McKnight et al., 2002). Both effects should be positive; the first should be the larger.

H4a: Information credibility is positively associated with seller and information trust

H4b: Information credibility is positively associated with platform trust

Credibility, Trust and Purchase Decisions

Credibility can influence purchasing directly, by improving the perceived quality of the evidence on which the decision rests, and indirectly, by building the trust required to accept residual risk. Both routes are theorised in the literature on online transaction intention (Paul A. Pavlou, 2003); (Liu & Zheng, 2024), and dual-process accounts of persuasion suggest they may operate at different levels of elaboration (Li & See-To, 2024). Trust, however, does work that credibility alone cannot: it resolves the vulnerability created by the temporal separation of payment and delivery, which is a structural feature of the exchange rather than a property of the information.

Whether seller trust or platform trust does more of that work is contested and context-dependent. Where consumer protection is well established and legal recourse against individual sellers is practicable, trust in the exchange partner may suffice. Where enforcement against individual sellers is difficult, and seller identity is ephemeral (conditions characteristic of social commerce in emerging markets), institutional assurance provided by the marketplace becomes the operative guarantee (Hajli, 2015); (Wang et al., 2022) All three effects are hypothesised as positive, with the relative magnitudes treated as an empirical question.

H5a: Information credibility is positively associated with purchase decision

H5b: Seller and information trust is positively associated with purchase decision

H5c: Platform trust is positively associated with purchase decision

Mediating Mechanisms

The framework specifies credibility as the sole conduit through which AI awareness and AI-use disclosure reach trust, and trust as a conduit through which credibility reaches purchasing. Neither AI awareness nor disclosure is specified with a direct path to trust or to purchase decision, because the theoretical mechanism in both cases runs through the evaluation of the information itself: awareness and disclosure are cues about content provenance, and provenance cues acquire behavioural relevance only insofar as they alter judgements of whether the content can be believed.

H6a : Information credibility mediates the relationship between AI awareness and both trust dimensions

H6b : Information credibility mediates the relationship between AI-use disclosure and both trust dimensions

H7 : Seller and information trust and platform trust mediate the relationship between information credibility and purchase decision

Figure 1 presents the integrated research model.

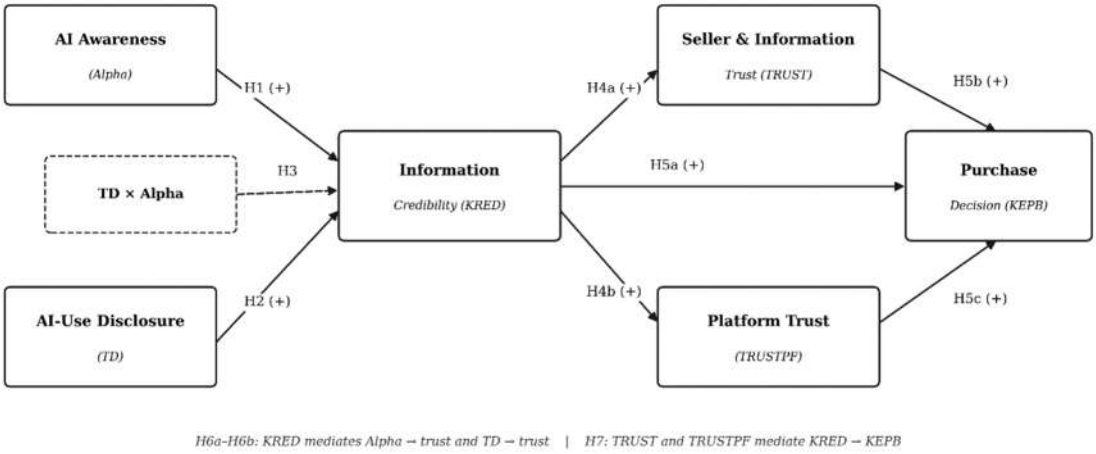

Figure 1. Conceptual framework: research model and hypotheses. Solid arrows denote hypothesised direct effects; the dashed arrow denotes the moderating interaction term.

Research Methods

Research Design and Procedure

The study employed a cross-sectional survey design. Data were collected online between 21 and 23 July 2026 from consumers residing in six Indonesian metropolitan areas (Jabodetabek, Bandung, Yogyakarta, Surabaya, Medan and Makassar) selected to span the country's principal digital retail centres rather than to constitute a probability sample of the national population. Fieldwork was administered by Jakpat (PT Jajak Pendapat Indonesia), an established Indonesian mobile survey panel provider, using screened quota sampling drawn from the provider's nationally distributed panel of mobile respondents. Panel-based administration accounts for the compressed fielding window; it also means the data capture a single narrow point in time, a limitation returned to in the Conclusion.

Three screening questions preceded the substantive instrument. Respondents had to confirm residence in one of the six target metropolitan areas, at least one completed purchase on TikTok Shop, Shopee or Instagram Shopping within the preceding six months, and prior exposure to (or suspicion of prior exposure to) product content or reviews produced with AI assistance. Respondents answering negatively to any screen were routed out of the survey. Of the 601 panel members who entered the questionnaire, 406 (67.6 per cent) satisfied all three criteria and 195 (32.4 per cent) were screened out.

The provider applied its standard response-quality controls prior to data delivery, including detection of straight-lining, inconsistent answering and implausibly short completion times. The delivered file of 406 qualifying responses contained no missing values on any demographic or substantive item, and all 406 cases were retained for analysis. The composition of the sample is reported in Table 1.

Sample Size Determination

Minimum sample size was established a priori using G*Power 3.1.9.7 (Kang, 2021) Consistent with convention in the PLS-SEM literature, the structural model was treated as a set of regressions and the analysis powered on the most heavily predicted endogenous construct. Two constructs receive three incoming paths each (information credibility, predicted by AI awareness, AI-use disclosure and their interaction, and purchase decision, predicted by credibility and the two trust dimensions), so the number of predictors was set to three. Assuming a medium effect size ($f^2 = 0.15$; Cohen, 1988), $\alpha = 0.05$ and power of 0.95, the required sample was 119.

Because interaction effects are characteristically small, that figure understates the sample required for the moderation test, and reporting it alone would misrepresent the study's power. A second analysis was therefore run on the R^2 increase attributable to a single tested predictor at $f^2 = 0.02$, yielding a required sample of 395. As a third check, the inverse square root method (Kock & Hadaya, 2018) was applied with a minimum detectable path coefficient of 0.15 at power of 0.80, yielding 275. The achieved sample

of 406 exceeds all three thresholds. A sensitivity analysis confirms that at n = 406, α = 0.05 and power of 0.80, the smallest detectable effect is $f^2 = 0.027$, corresponding to a local R^2 contribution of 0.026. The non-significant interaction reported in the results below is therefore a well-powered null rather than an artefact of insufficient sample. Full parameters are reported in Table 3.

Table 1. Construct operationalisation and measurement sources

Construct (code)	Items	Content domain	Adapted from
AI Awareness (Alpha)	3 (KAI1–KAI3)	Recognition that AI is routinely involved in producing the product content encountered on the platform	(Kim & Kim, 2021); (Sundar, 2008)
AI-Use Disclosure (TD)	3 (TD1–TD3)	Perceived clarity, completeness and visibility of seller and platform statements about AI use	(Kirkby et al., 2023a; Koning & Voorveld, 2025)
Information Credibility (KRED)	6 (KREDP1–KREDP3 message; KREDS1–KREDS3 source)	Believability and accuracy of the product message; expertise and trustworthiness of the source	(Cheung et al., 2008; Metzger et al., 2003; Ohanian, 1990)
Seller & Information Trust (TRUST)	6 (TRUST1–TRUST6)	Seller ability, benevolence and integrity, and confidence in the information supplied	(Mayer et al., 1995); (McKnight et al., 2002)
Platform Trust (TRUSTPF)	2 (TRUSTPF1, TRUSTPF3)	Confidence in platform governance, structural assurance and dispute resolution	(McKnight et al., 2002) (Gefen et al., 2003)
Purchase Decision (KEPB)	2 (KEPB1–KEPB2)	Decision to proceed with purchase following exposure to the content	(Paul A. Pavlou, 2003) (Hajli, 2015)

Note. All items were measured on five-point Likert scales anchored at 1 (strongly disagree) and 5 (strongly agree). The full item wording in Bahasa Indonesia and English back-translation is provided in Appendix A. The platform-trust scale originally comprised three items; TRUSTPF2 was removed during measurement-model purification and the remaining two items were retained.

Table 2. A priori statistical power analysis (G*Power 3.1.9.7)

Parameter	Main model	Interaction test	Inverse square root
Statistical test	Linear multiple regression, R^2 deviation from zero	Linear multiple regression, R^2 increase (1 tested)	(Kock & Hadaya, 2018)
Effect size	$f^2 = 0.15$	$f^2 = 0.02$	$\|p\| = 0.15$
α err prob	0.05	0.05	—
Power (1 β)	0.95	0.80	0.80
Number of predictors	3	3	—
Noncentrality parameter λ	17.850	7.900	—
Critical F	2.6835	3.8654	—
Numerator / denominator df	3 / 115	1 / 391	—
Required total sample	119	395	275
Actual power at required N	0.9510	0.8006	—
Achieved sample	406	406	406

Note. Sensitivity analysis at n = 406, α = 0.05 and power = 0.80 yields a minimum detectable effect of $f^2 = 0.027$ (local R^2 contribution = 0.026).

Measures

All constructs were measured reflectively with previously validated multi-item scales adapted to the social commerce context, as summarised in table 2. Items were translated into Bahasa Indonesia and independently back-translated into English by a second bilingual researcher; discrepancies were resolved by discussion before fielding. The instrument was pre-tested with a small panel of active social commerce users, whose feedback prompted minor rewording of two items for clarity. Responses were captured on five-point Likert scales.

Information credibility was operationalised as a single reflective construct indicated by two content facets, message credibility (KREDP1–KREDP3) and source credibility (KREDS1–KREDS3), consistent with the conventional decomposition of the credibility construct (Metzger et al., 2003); (Ohanian, 1990). The moderating term was formed as the product of AI-use disclosure and AI awareness using the two-stage approach, which is the recommended specification when both the focal predictor and the moderator are reflectively measured latent variables (Becker et al., 2023).

Common Method Bias

Because all constructs were measured from a single source at a single point in time, common method variance is a legitimate concern. Procedural remedies were applied at the design stage: respondent anonymity was assured, the participation reward was fixed and independent of the answers given, item order was varied between construct blocks, and predictor and criterion items were separated by intervening screens (Podsakoff et al., 2003). Administration through an independent panel provider further separated the respondents from the research team, reducing the scope for demand characteristics. Statistically, the full collinearity approach was applied (Kock, 2015): with a single exception (KREDP1, VIF = 3.708), all outer variance inflation factors fell below the 3.3 criterion, and no inner-model VIF reached the critical value of 5.0, which does not support a diagnosis of pervasive method bias. This evidence is indirect, and the inference is offered as such.

Data Analysis

Partial least squares structural equation modelling was selected because the study is prediction-oriented, incorporates a latent interaction term, and combines constructs measured with as few as two and as many as six indicators, conditions under which the variance-based estimator is preferred over covariance-based alternatives (Sarstedt et al., 2022) (Sundar & Kim, 2019a). Estimation was performed in SmartPLS 4 using the path weighting scheme. Significance was assessed by bootstrapping with 5,000 subsamples, two-tailed testing at $\alpha = 0.05$, and bias-corrected and accelerated confidence intervals.

Analysis proceeded in the conventional two stages (Hair et al., 2019). The measurement model was assessed through indicator loadings, Cronbach's alpha, composite reliability (ρA and ρC), average variance extracted, the heterotrait–monotrait ratio and the Fornell–Larcker criterion. The structural model was then assessed through path coefficients, effect sizes (f^2), coefficients of determination, inner-model collinearity, and specific indirect, direct and total effects. Mediation was evaluated following the specific indirect effect approach rather than the causal-steps procedure, and the moderating effect was examined through the significance and effect size of the interaction term.

Results and Discussion

Respondent Characteristics

A total of 406 valid responses were retained for analysis. Table 3 summarieses the demographic and behavioral characteristics of the respondents.

Table 3. Respondent profile (N=406)

Characteristic	Category	Frequency	Percentage
Gender			
	Female	246	60.6
	Male	160	39.4
Age			
	17-24 years	160	39.4
	25-34 years	128	31.5
	35-44 years	84	20.7
	45-55 years	32	7.9
	Over 55 years	2	0.5
Highest completed education			
	Upper secondary (SMA/equivalent)	214	52.7
	Bachelor's degree (S1)	150	36.9
	Diploma	37	9.1
	Master's or doctoral degree (S2/S3)	5	1.2
Primary social commerce platform	Shopee	269	66.3
	TikTok Shop	134	33.0
	Instagram Shopping	3	0.7

Characteristic	Category	Frequency	Percentage
Purchase frequency, preceding six months	1-2 times	65	16.0
	3-5 times	161	39.7
	6-10 times	88	21.7
	More than 10 times	92	22.7

Note. Percentages are calculated on the full sample and may not sum to 100.0 owing to rounding. Residence in one of the six target metropolitan areas was a screening criterion; the distribution of respondents across those areas was not recorded.

Three features of the profile bear on the interpretation of the results. The sample is young - 70.9 per cent of respondents are under 35 - and predominantly female (60.6 per cent), consistent with the demographic skew of social commerce participation in Indonesia. Educational attainment is more evenly distributed than in many published social commerce samples: a slight majority (52.7 per cent) hold upper-secondary qualifications and 47.3 per cent hold a tertiary credential, so the findings are not confined to a graduate population. Finally, the sample is concentrated on two of the three focal platforms: 99.3 per cent nominated Shopee or TikTok Shop as their primary platform and only three respondents nominated Instagram Shopping, so the results describe marketplace-embedded rather than feed-native social commerce, a constraint revisited in the Conclusion. Respondents were active rather than occasional buyers, with 44.4 per cent reporting six or more purchases in the preceding six months.

Assessment of the Measurement Model

All six reflective constructs satisfied the criteria for indicator reliability, internal consistency and convergent validity (Table 4). Standardised outer loadings ranged from 0.793 (KAI3) to 0.940 (KEPB2), comfortably exceeding the 0.708 benchmark, and every loading was significant at $p < 0.001$ with t-values between 17.433 and 118.528. Cronbach's alpha ranged from 0.750 (AI awareness) to 0.927 (information credibility), ρA from 0.766 to 0.928, and ρC from 0.855 to 0.943. Because ρC values remained below the 0.95 ceiling, there is no evidence that the item sets are semantically redundant (Hair et al., 2019). Average variance extracted ranged from 0.663 to 0.849, meaning that each construct explains at least two-thirds of the variance in its indicators.

Table 4. Measurement model assessment (n = 406)

Construct	Indicator	Loading	t-value	VIF	Cronbach's α	ρA	ρC	AVE
AI Awareness (Alpha)	KAI1	0.826	23.823	1.384	0.750	0.766	0.855	0.663
	KAI2	0.824	23.214	1.578				
	KAI3	0.793	17.433	1.630				
AI-Use Disclosure (TD)	TD1	0.821	27.402	1.550	0.798	0.798	0.881	0.712
	TD2	0.846	35.970	1.767				
	TD3	0.864	46.289	1.937				
Information Credibility (KRED)	KREDP1	0.887	59.007	3.708	0.927	0.928	0.943	0.733
	KREDP2	0.840	42.913	2.643				
	KREDP3	0.874	59.263	3.038				
	KREDS1	0.843	44.922	2.725				
	KREDS2	0.833	33.009	2.566				
	KREDS3	0.860	48.494	3.168				
Seller & Information Trust (TRUST)	TRUST1	0.857	49.530	2.710	0.909	0.909	0.929	0.686
	TRUST2	0.805	35.619	2.116				
	TRUST3	0.821	45.972	2.180				
	TRUST4	0.835	27.290	2.455				
	TRUST5	0.817	34.833	2.313				
	TRUST6	0.835	36.539	2.451				
Platform Trust (TRUSTPF)	TRUSTPF1	0.907	91.343	1.614	0.763	0.766	0.894	0.808
	TRUSTPF3	0.891	64.644	1.614				
Purchase Decision (KEPB)	KEPB1	0.902	54.004	1.967	0.824	0.856	0.918	0.849

Construct	Indicator	Loading	t-value	VIF	Cronbach's α	ρA	ρC	AVE
	KEPB2	0.940	118.528	1.967				
Interaction (TD × Alpha)	TD × Alpha	1.000	n/a	1.000	n/a	n/a	n/a	n/a

Note. Loadings are standardised outer loadings from the original sample; t-values derive from bootstrapping with 5,000 subsamples. ρA and ρC denote composite reliability; AVE denotes average variance extracted. The interaction term is a single-item product construct and is not subject to reliability assessment. All values in this table were computed by the authors from the study's own survey data (n = 406) using SmartPLS 4; they are not drawn from any secondary source. The evaluation thresholds applied follow Hair et al. (2019, 2022).

The complete measurement (outer) model estimated in SmartPLS 4 is presented in Figure 2, which reproduces every construct, every retained indicator and the standardised outer loading of each indicator on its construct. Figure 2 and Table 4 report the same estimation and are intended to be read together: the table gives the reliability and validity statistics, while the figure shows the measurement structure from which they were obtained.

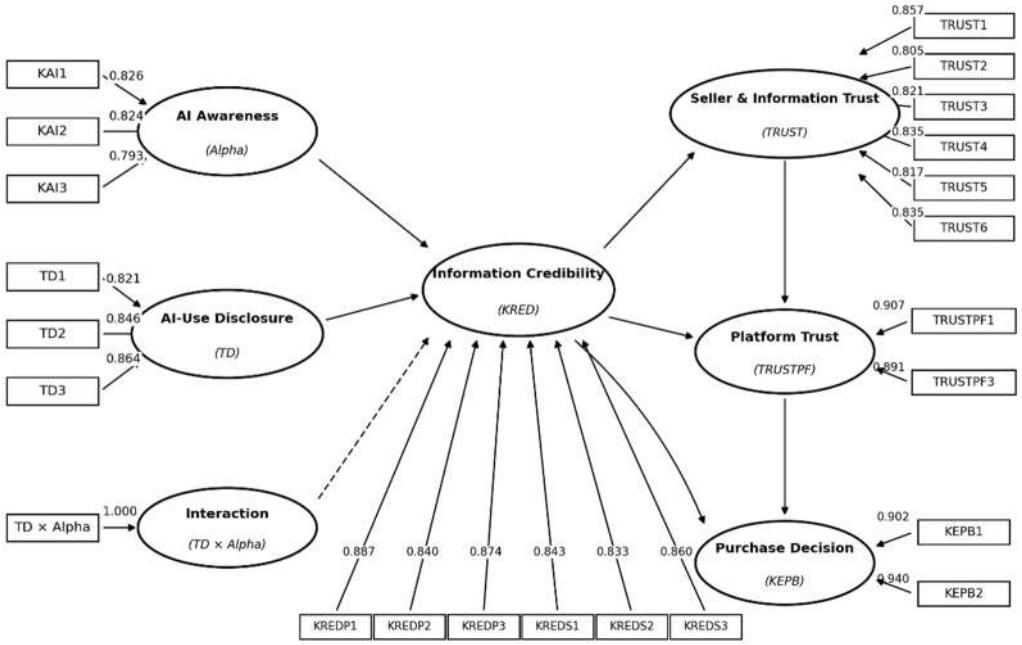

Figure 2. Measurement (outer) model with standardised outer loadings (SmartPLS 4 output, n = 406). Rectangles denote measured indicators, ellipses denote latent constructs, and the values on the measurement arrows are standardised outer loadings. The dashed arrow denotes the interaction term

Discriminant validity was examined using both the heterotrait–monotrait ratio and the Fornell–Larcker criterion (Table 5). All HTMT values fell below the 0.90 threshold recommended for conceptually proximate constructs (Hair et al, 2022) Two comparisons nevertheless warrant explicit comment: information credibility and seller and information trust (HTMT = 0.874) and seller and information trust and platform trust (HTMT = 0.879). Both satisfy the 0.90 criterion but exceed the more conservative 0.85 threshold, indicating that these constructs, while empirically separable, capture closely related evaluative territory. The Fornell–Larcker results corroborate discriminant validity, with the narrowest margin observed between seller and information trust (√AVE = 0.828) and its correlation with information credibility (r = 0.803).

Table 5. Discriminant validity assessment

Construct	Alpha	TD	KRED	TRUST	TRUSTPF	KEPB
	Panel A. Heterotrait–monotrait ratio (HTMT)					
Alpha	—					
TD	0.469	—				

Construct	Alpha	TD	KRED	TRUST	TRUSTPF	KEPB
KRED	0.388	0.472	—			
TRUST	0.266	0.478	0.874	—		
TRUSTPF	0.354	0.592	0.723	0.879	—	
KEPB	0.420	0.613	0.633	0.683	0.717	—
	Panel B. Fornell–Larcker criterion (√AVE on the diagonal)					
Alpha	0.814					
TD	0.369	0.844				
KRED	0.332	0.408	0.856			
TRUST	0.229	0.408	0.803	0.828		
TRUSTPF	0.274	0.461	0.610	0.733	0.899	
KEPB	0.337	0.498	0.564	0.600	0.574	0.921

Note. Panel A reports heterotrait–monotrait ratios Panel B reports the Fornell–Larcker criterion, with the square root of AVE on the diagonal and latent variable correlations below it (Hair, 2022)

Model Fit and Structural Collinearity

Approximate fit indices are reported in Table 6 for transparency, with the caveat that fit testing occupies a contested position in the PLS-SEM literature and should not be treated as a confirmatory hurdle in a prediction-oriented model (Hair et al., 2019). The saturated model returned SRMR = 0.053, within the conservative 0.08 threshold; the estimated model returned SRMR = 0.082, marginally above that benchmark but well within the liberal 0.10 criterion. NFI values of 0.832 and 0.811 indicate acceptable correspondence between the empirical and model-implied correlation matrices. The modest gap between saturated and estimated fit is attributable to the theoretically imposed restriction that AI awareness and AI-use disclosure influence trust and purchase decisions only through credibility.

Table 6. Approximate model fit indices

Fit criterion	Saturated model	Estimated model	Recommended threshold
SRMR	0.053	0.082	< 0.08 conservative; < 0.10 liberal
d_ULS	0.711	1.707	Below 95% bootstrap quantile
d_G	0.398	0.492	Below 95% bootstrap quantile
Chi-square	997.953	1120.436	Reported for completeness
NFI	0.832	0.811	> 0.80 acceptable; > 0.90 good

Note. Fit indices in PLS-SEM are approximate and reported descriptively rather than as confirmatory tests (Hair et al., 2019).

Inner-model VIF values ranged from 1.000 to 3.828 (Table 7). Three structural relationships exceeded the stricter 3.3 threshold: TRUST → KEPB (3.828), KRED → KEPB (2.823) and TRUSTPF → KEPB (2.166). None approaches the critical value of 5.0, so the estimates are not compromised in the formal sense; the elevated collinearity among the three simultaneous predictors of purchase decision nevertheless has interpretive consequences addressed below.

Direct Effects

Contrary to the algorithm-aversion prediction, AI awareness exerted a positive and significant effect on information credibility ($\beta = 0.208$, SD = 0.062, $t = 3.333$, $p = 0.001$, 95% BCa CI [0.087; 0.329]), supporting H1. The effect size was small ($f^2 = 0.047$), but the confidence interval excludes zero across its whole range and the total effect of AI awareness on purchase decision remains significant ($\beta = 0.117$, $t = 3.213$, $p = 0.001$), indicating a consequential rather than incidental relationship.

AI-use disclosure produced a stronger positive effect on information credibility ($\beta = 0.323$, SD = 0.053, $t = 6.122$, $p = 0.000$, 95% BCa CI [0.217; 0.425], $f^2 = 0.108$), supporting H2 and establishing disclosure as the more influential of the two antecedents. The disclosure coefficient is approximately 55 per cent larger than the awareness coefficient, indicating that what a seller or platform signals about AI use matters more for credibility judgements than what a consumer independently notices.

H3 was not supported. The interaction term exerted a negligible and non-significant effect on information credibility ($\beta = -0.024$, SD = 0.055, $t = 0.433$, $p = 0.665$, 95% BCa CI [−0.108; 0.097], $f^2 = 0.001$). The confidence interval straddles zero symmetrically and the effect size is indistinguishable from nil. As established in the sample size determination above, the design had power of 0.80 to detect an

interaction of $f^2 = 0.02$ at n = 395; the achieved sample of 406 therefore renders this a well-powered null rather than an underpowered test.

Information credibility was a powerful antecedent of both trust dimensions. Its effect on seller and information trust was exceptionally strong (β = 0.803, SD = 0.037, t = 21.485, p = 0.000, $f^2 = 1.815$), supporting H4a, and its effect on platform trust was also substantial (β = 0.610, SD = 0.048, t = 12.606, p = 0.000, $f^2 = 0.593$), supporting H4b. The ordering of magnitudes is as predicted in the hypothesis development above.

The three simultaneous antecedents of purchase decision produced a differentiated pattern. Only platform trust exerted a significant direct effect (β = 0.279, SD = 0.069, t = 4.057, p = 0.000, 95% BCa CI [0.140; 0.408], $f^2 = 0.061$), supporting H5c. Neither information credibility (β = 0.213, t = 1.349, p = 0.177) nor seller and information trust (β = 0.225, t = 1.420, p = 0.156) reached significance, so H5a and H5b were not supported. Both non-significant paths are characterised by unusually large standard errors (SD = 0.158 in each case) and by confidence intervals spanning zero widely — a signature consistent with the elevated inner VIF values reported above. The three point estimates are of comparable magnitude (0.213, 0.225, 0.279); what distinguishes platform trust is the precision of its estimate rather than the size of its coefficient.

Table 7. Structural model: direct effects, effect sizes and hypothesis decisions

Path	β (O)	SD	t	p	95% BCa CI	f^2	VIF	Decision
Antecedents of information credibility								
H1: Alpha → KRED	0.208	0.062	3.333	0.001	[0.087; 0.329]	0.047	1.166	Supported
H2: TD → KRED	0.323	0.053	6.122	0.000	[0.217; 0.425]	0.108	1.213	Supported
H3: TD × Alpha → KRED	0.024	0.055	0.433	0.665	[0.108; 0.097]	0.001	1.082	Not supported
Credibility → multidimensional trust								
H4a: KRED → TRUST	0.803	0.037	21.485	0.000	[0.702; 0.857]	1.815	1.000	Supported
H4b: KRED → TRUSTPF	0.610	0.048	12.606	0.000	[0.508; 0.695]	0.593	1.000	Supported
Antecedents of purchase decision								
H5a: KRED → KEPB	0.213	0.158	1.349	0.177	[0.073; 0.512]	0.028	2.823	Not supported
H5b: TRUST → KEPB	0.225	0.158	1.420	0.156	[0.079; 0.514]	0.023	3.828	Not supported
H5c: TRUSTPF → KEPB	0.279	0.069	4.057	0.000	[0.140; 0.408]	0.061	2.166	Supported

Note. β (O) = path coefficient from the original sample; SD = bootstrap standard deviation; 95% BCa CI = bias-corrected and accelerated confidence interval from 5,000 bootstrap subsamples; f^2 = effect size (0.02 small, 0.15 medium, 0.35 large; Cohen, 1988); VIF = inner-model variance inflation factor. Two-tailed tests at α = 0.05.

The corresponding structural (inner) model is presented in Figure 3, which reports the standardised path coefficient and significance level of every hypothesised relationship together with the coefficient of determination of each endogenous construct. Figures 2 and 3 together constitute the two required SmartPLS outputs: the outer model documents how the constructs were measured, and the inner model documents how they are related.

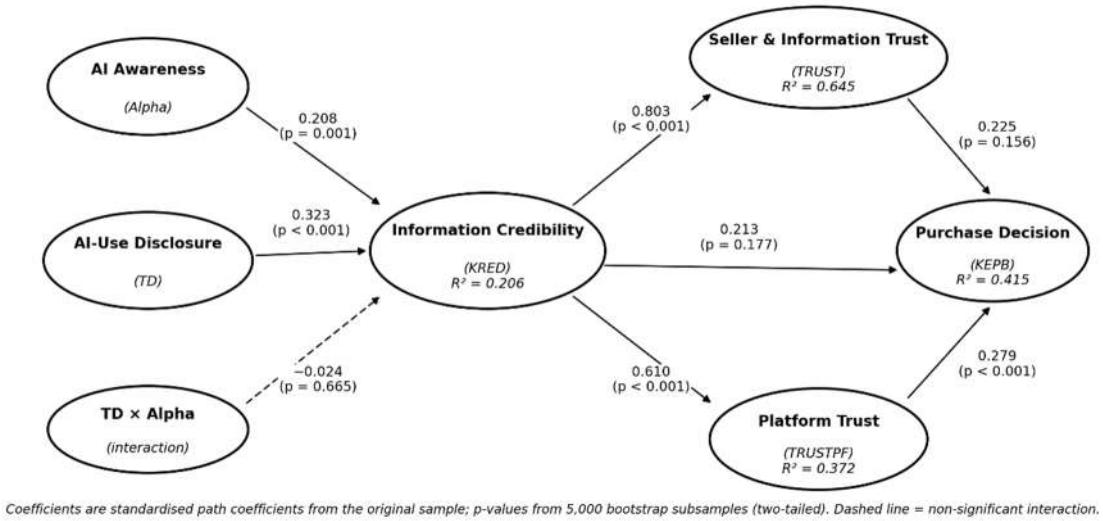

Figure 3. Structural (inner) model with standardised path coefficients, p-values and R^2 values (SmartPLS 4 output, n = 406; 5,000 bootstrap subsamples). The dashed arrow denotes the non-significant interaction effect

Mediation Analysis

Because neither AI awareness nor AI-use disclosure was specified with a direct path to purchase decision, their entire influence on purchasing operates indirectly, and the analysis concerns the distribution of that influence across competing routes rather than the classification of full versus partial mediation.

Both exogenous constructs transmitted significant indirect effects to each trust dimension. AI awareness reached seller and information trust (β = 0.167, t = 3.342, p = 0.001) and platform trust (β = 0.127, t = 3.176, p = 0.002) through credibility, and AI-use disclosure did so with larger magnitudes (β = 0.259, t = 5.788, p = 0.000 and β = 0.197, t = 4.924, p = 0.000). H6a and H6b are therefore supported: information credibility completely mediates the awareness–trust and disclosure–trust relationships.

H7 received partial support. Of the routes available to each exogenous construct, only those running through platform trust attained significance: Alpha → KRED → TRUSTPF → KEPB (β = 0.035, t = 2.272, p = 0.023) and TD → KRED → TRUSTPF → KEPB (β = 0.055, t = 2.952, p = 0.003). The same asymmetry appears when credibility is the exogenous construct: KRED → TRUSTPF → KEPB was significant (β = 0.170, t = 3.688, p = 0.000), whereas KRED → TRUST → KEPB was not (β = 0.180, t = 1.356, p = 0.175), despite a marginally larger point estimate.

Decomposing the total effect of information credibility on purchase decision (β = 0.564, t = 13.171, p = 0.000) clarifies the structure: 37.8 per cent travels directly, 31.9 per cent through seller and information trust, and 30.1 per cent through platform trust. The influence of credibility on purchasing is thus distributed almost evenly across the three routes, even though only the platform-trust route is individually estimable with confidence. The total effect is large, unambiguous and precisely estimated; its decomposition into competing pathways is degraded by predictor overlap. None of the interaction-term pathways approached significance (all p ≥ 0.657), confirming that the failure of H3 propagates consistently through the downstream structure.

Table 8. Specific indirect effects

Specific indirect path	β (O)	SD	t	p	95% BCa CI	Status
AI awareness as exogenous construct						
Alpha → KRED → TRUST	0.167	0.050	3.342	0.001	[0.069; 0.263]	Significant
Alpha → KRED → TRUSTPF	0.127	0.040	3.176	0.002	[0.052; 0.206]	Significant
Alpha → KRED → KEPB	0.044	0.037	1.203	0.229	[0.010; 0.139]	n.s.
Alpha → KRED → TRUST → KEPB	0.037	0.030	1.235	0.217	[0.009; 0.104]	n.s.
Alpha → KRED → TRUSTPF → KEPB	0.035	0.016	2.272	0.023	[0.012; 0.072]	Significant
AI-use disclosure as exogenous construct						
TD → KRED → TRUST	0.259	0.045	5.788	0.000	[0.174; 0.351]	Significant
TD → KRED → TRUSTPF	0.197	0.040	4.924	0.000	[0.123; 0.279]	Significant

Specific indirect path	β (O)	SD	t	p	95% BCa CI	Status
TD → KRED → KEPB	0.069	0.054	1.287	0.198	[0.019; 0.191]	n.s.
TD → KRED → TRUST → KEPB	0.058	0.045	1.301	0.193	[0.016; 0.151]	n.s.
TD → KRED → TRUSTPF → KEPB	0.055	0.019	2.952	0.003	[0.025; 0.099]	Significant
Credibility as exogenous construct						
KRED → TRUST → KEPB	0.180	0.133	1.356	0.175	[0.059; 0.431]	n.s.
KRED → TRUSTPF → KEPB	0.170	0.046	3.688	0.000	[0.084; 0.264]	Significant
Moderating term						
TD × Alpha → KRED → KEPB	0.005	0.011	0.444	0.657	[0.042; 0.006]	n.s.
TD × Alpha → KRED → TRUST	0.019	0.045	0.427	0.669	[0.086; 0.080]	n.s.
TD × Alpha → KRED → TRUSTPF	0.015	0.034	0.427	0.669	[0.065; 0.062]	n.s.
TD × Alpha → KRED → TRUST → KEPB	0.004	0.015	0.279	0.780	[0.040; 0.012]	n.s.
TD × Alpha → KRED → TRUSTPF → KEPB	0.004	0.010	0.403	0.687	[0.021; 0.018]	n.s.

Note. n.s. = not significant at α = 0.05 (two-tailed). Confidence intervals are bias-corrected and accelerated, based on 5,000 bootstrap subsamples.

Table 9. Total effects

Total effect	β (O)	SD	t	p	95% BCa CI	Status
Alpha → KRED	0.208	0.062	3.333	0.001	[0.087; 0.329]	Significant
Alpha → TRUST	0.167	0.050	3.342	0.001	[0.069; 0.263]	Significant
Alpha → TRUSTPF	0.127	0.040	3.176	0.002	[0.052; 0.206]	Significant
Alpha → KEPB	0.117	0.036	3.213	0.001	[0.049; 0.192]	Significant
TD → KRED	0.323	0.053	6.122	0.000	[0.217; 0.425]	Significant
TD → TRUST	0.259	0.045	5.788	0.000	[0.174; 0.351]	Significant
TD → TRUSTPF	0.197	0.040	4.924	0.000	[0.123; 0.279]	Significant
TD → KEPB	0.182	0.036	5.119	0.000	[0.114; 0.256]	Significant
KRED → TRUST	0.803	0.037	21.485	0.000	[0.702; 0.857]	Significant
KRED → TRUSTPF	0.610	0.048	12.606	0.000	[0.508; 0.695]	Significant
KRED → KEPB	0.564	0.043	13.171	0.000	[0.476; 0.646]	Significant
TRUST → KEPB	0.225	0.158	1.420	0.156	[0.079; 0.514]	n.s.
TRUSTPF → KEPB	0.279	0.069	4.057	0.000	[0.140; 0.408]	Significant
TD × Alpha → KRED	0.024	0.055	0.433	0.665	[0.108; 0.097]	n.s.
TD × Alpha → TRUST	0.019	0.045	0.427	0.669	[0.086; 0.080]	n.s.
TD × Alpha → TRUSTPF	0.015	0.034	0.427	0.669	[0.065; 0.062]	n.s.
TD × Alpha → KEPB	0.013	0.031	0.437	0.662	[0.064; 0.052]	n.s.

Note. Total effects comprise all direct and indirect pathways linking each construct pair.

Explanatory Power

The model accounted for 64.5 per cent of the variance in seller and information trust, 41.5 per cent in purchase decision, 37.2 per cent in platform trust and 20.6 per cent in information credibility (Table 10). Explanatory power is moderate-to-substantial for seller and information trust, moderate for purchase decision, and weak for information credibility (Hair et al., 2019). Adjusted R^2 values differ from their unadjusted counterparts by no more than 0.006, indicating that the model is not inflated by the number of predictors.

The modest R^2 for information credibility deserves emphasis rather than apology. Nearly 80 per cent of the variance in credibility perceptions remains unexplained by AI awareness, disclosure and their interaction combined, indicating that credibility judgements in social commerce are governed predominantly by determinants outside the AI-transparency domain — product category risk, prior experience with the seller, review volume and valence, price plausibility, and individual differences in scepticism among them. The claim this study can defend is that AI transparency matters and matters positively; it is not that AI transparency is the principal driver of credibility.

Table 10. Coefficients of determination for endogenous constructs

Endogenous construct	R^2	Adjusted R^2	Interpretation	Predictors
Information Credibility (KRED)	0.206	0.200	Weak	Alpha, TD, TD × Alpha
Seller & Information Trust (TRUST)	0.645	0.644	Moderate–substantial	KRED
Platform Trust (TRUSTPF)	0.372	0.371	Weak–moderate	KRED
Purchase Decision (KEPB)	0.415	0.411	Moderate	KRED, TRUST, TRUSTPF

Note. Interpretive benchmarks follow Hair et al. (2019): R^2 values of 0.75, 0.50 and 0.25 are described as substantial, moderate and weak, respectively.

Discussion

AI Awareness as a Credibility Asset

The most consequential finding is that awareness of AI involvement in social commerce content is positively associated with perceived information credibility. This adjudicates the competing predictions set out in the hypothesis development above in favour of the machine-heuristic account and against algorithm aversion, and it diverges from a substantial body of experimental evidence generated in mature Western markets (Brüns & Meißner, 2024) (Kirk & Givi, 2025) (Koning & Voorveld, 2025). Two explanations, one methodological and one contextual, jointly account for the divergence, and distinguishing between them is essential to interpreting the result responsibly.

The methodological explanation concerns what is being measured. Experimental studies manipulate exposure to a disclosure label and observe the immediate reaction of a participant who has just been informed that a specific message was machine-generated; that design isolates a situational suspicion response. The present study measures AI awareness as a chronic, individual-difference construct the degree to which a consumer recognises that AI is routinely involved in the content they encounter. Measured this way, awareness is entangled with AI literacy and platform familiarity, and that interpretive framework appears to reduce rather than amplify evaluative uncertainty. Awareness in the trait sense and disclosure in the stimulus sense are not the same construct, and the literature has tended to treat them interchangeably. The composition of the sample is consistent with this reading: 70.9 per cent of respondents were under 35 and 99.3 per cent shopped primarily on Shopee or TikTok Shop, describing a young, platform-habituated cohort for whom AI-assisted content is an ordinary rather than a novel feature of the shopping environment.

The contextual explanation concerns the comparison class. In Indonesian social commerce, where consumers routinely contend with paid endorsements, incentivised reviews and undisclosed affiliate arrangements, the relevant comparison for an AI-assisted product description is not an authentic human account but a commercially motivated one. Against that baseline, machine involvement may read as a reduction in vested interest rather than a loss of authenticity, which is precisely the condition under which the machine heuristic is predicted to operate (Sundar, 2008) (Sundar & Kim, 2019b). The finding is consistent with algorithm-appreciation results showing preference for algorithmic judgement on tasks perceived as objective (Castelo et al., 2019) (Logg et al., 2019), and with the observation that aversion concentrates in domains judged to require uniquely human capacities (Longoni et al., 2019). Product information search is a comparatively objective task.

These explanations are candidate accounts rather than demonstrated mechanisms. The design cannot adjudicate between them, and doing so is the most valuable next step the finding invites.

Disclosure as a Direct Signal Rather Than a Buffer

AI-use disclosure emerged as the strongest antecedent of information credibility, while the interaction between disclosure and awareness was empirically null. Read together, the two results indicate that disclosure and awareness contribute additively to credibility, and that disclosure neither amplifies nor cushions the awareness effect.

Signalling theory provides the more parsimonious account of the main effect (Spence, 1973). Disclosure is voluntary and mildly costly in a market where non-disclosure carries no immediate penalty, which makes it a credible signal of seller type. A seller willing to state that AI assisted the production of a listing signals a general disposition toward candour, and consumers appear to read it as such. Under this interpretation, what is being evaluated is the seller's transparency practice rather than the AI content itself, which is why disclosure raises credibility rather than lowering it.

The null interaction has genuine theoretical content and should not be reported merely as a failed hypothesis. Persuasion-knowledge accounts predict that disclosure matters most for consumers whose latent suspicion it activates and least for those already sceptical (Eisend & Tarrahi, 2022) The absence of any interaction indicates that disclosure in this setting is not functioning as an activator of persuasion knowledge at all: its effect is uniform across levels of awareness, which is what would be expected of an institutional quality signal. (Kirkby et al., 2023b) reach a compatible conclusion in the brand-voice context. The result suggests that AI disclosure in routine, low-involvement commercial settings is processed peripherally as a marker of seller quality rather than centrally as information about message manipulation (Li & See-To, 2024)

Credibility as the Engine of Multidimensional Trust

Information credibility exerted an exceptionally strong effect on seller and information trust and a substantial effect on platform trust. Substantively, this extends trust-transfer theory to the AI-augmented marketplace. Trust in social commerce is not formed through direct interpersonal experience (buyers rarely meet sellers and cannot inspect goods), so it must be inferred from the artefacts available, principally the information itself. Credibility judgements about product descriptions and reviews therefore serve as the primary substrate from which trust in the seller is constructed, consistent with the classical position that source credibility drives persuasive outcomes when direct verification is unavailable (Ohanian, 1990) (Li & See-To, 2024) and with evidence on eWOM adoption (Cheung et al., 2008).

The differential magnitudes are theoretically informative and match the prediction in the hypothesis development above. Credibility transfers more readily to the proximate object of evaluation than to the institutional actor standing behind the transaction, because platform trust is anchored additionally in structural assurances independent of any particular listing (McKnight et al., 2002). (Wang et al., 2022) This is consistent with the 37.2 per cent of platform-trust variance explained against 64.5 per cent for seller and information trust: institution-based trust is by design less sensitive to message-level cues.

Caution is nevertheless warranted. The HTMT ratio between credibility and seller trust is 0.874 and their latent correlation is 0.803. These values fall within accepted limits but indicate substantial conceptual proximity, and part of the observed effect may reflect measurement overlap: items asking whether product information is believable and items asking whether the seller can be trusted plausibly tap a common evaluative disposition. The finding should be read as evidence of a tightly coupled credibility–trust complex rather than as a clean causal estimate between fully independent constructs.

Platform Trust as the Decisive Route to Purchase

Of the three direct antecedents of purchase decision, only platform trust reached significance. Interpreting this requires care, because the three point estimates are close in magnitude and differ principally in precision. The honest reading is that credibility, seller trust and platform trust overlap sufficiently as simultaneous predictors that the model cannot reliably apportion their individual contributions, and that among the three, platform trust carries the variance that is distinctively its own.

Two conclusions survive this caveat. First, the total effect of credibility on purchase decision is large and precisely estimated, so the substantive importance of credibility for purchasing is not in doubt; only the allocation of its influence across routes is uncertain. Second, the fact that platform trust retains a significant unique effect while the other two do not indicates that it supplies something the others cannot. That something is most plausibly institutional assurance: when a consumer evaluates a listing from an unknown seller, structural safeguards provided by the marketplace resolve the residual risk that credibility assessment and seller impressions leave unaddressed (McKnight et al., 2002) (Paul A. Pavlou, 2003). In social commerce specifically, where transactions are embedded in social feeds and seller identity is often ephemeral, the platform is the only durable party to the exchange (Hajli, 2015) (Wang et al., 2022). The mediation results support this reading: the only significant serial pathways to purchase decision run through platform trust.

Theoretical Implications

Three contributions follow. First, the study qualifies the emerging consensus that AI disclosure damages consumer response. That consensus rests largely on experiments conducted in mature Western markets using single-exposure manipulations. In a field survey of 406 consumers in an emerging-market social commerce environment, both awareness and disclosure raised credibility. The disclosure–trust relationship is therefore contingent rather than universal, and its boundary conditions (market maturity, baseline trust in commercial content, the nature of the alternative to AI involvement, and whether

awareness is measured as a trait or induced as a stimulus) require systematic specification. (Grigsby et al., 2025) similarly report that disclosure effects depend on category characteristics rather than operating uniformly.

Second, the study demonstrates that trust in social commerce must be modelled multidimensionally, and that the dimensions behave differently rather than merely covarying. Seller and information trust absorbed the largest share of the credibility effect but did not translate into an independently estimable effect on purchase; platform trust absorbed less but converted more reliably. Studies that collapse trust into a single latent variable will recover an average that obscures a real functional division of labour.

Third, the well-powered null moderation contributes a negative result of theoretical value. Persuasion-knowledge models predict that transparency cues interact with consumer suspicion; here they did not. This suggests that the persuasion-knowledge framework may apply more narrowly than its current usage in the AI-disclosure literature implies, and that its scope condition is the salience of a persuasion attempt rather than the mere presence of a provenance cue.

Managerial and Policy Implications

For sellers, voluntary disclosure of AI assistance carries no measurable credibility penalty in this market and is associated with a credibility gain. Disclosure had the largest effect of any antecedent tested, which reframes it from a compliance cost into a low-cost trust-building instrument. The caveat is that the study measures disclosure of AI assistance in content production; it does not test the fabrication of experiences a seller never had, and no inference about that practice should be drawn from these results.

For platforms, the finding that platform trust is the decisive route to purchase argues for investment in visible institutional assurance: transparent dispute resolution, enforceable seller accountability and consistent provenance labelling applied uniformly across listings rather than left to seller discretion. Standardised, platform-issued AI labelling is likely to outperform voluntary seller disclosure, because a uniformly applied signal cannot be strategically withheld and therefore does not require consumers to infer meaning from its absence.

For regulators, the results suggest that mandatory AI-disclosure requirements of the kind now emerging in several jurisdictions need not depress commercial outcomes in this market, which weakens the principal industry objection to such rules. The results also indicate the limits of disclosure as a policy instrument: it explained only a modest share of credibility variance and did not interact with consumer awareness. Disclosure mandates should therefore be paired with verification infrastructure and digital literacy provision, since a label conveys information about origin but not about accuracy.

Limitations and Future Research

Six limitations bound these findings. First, the cross-sectional design permits no causal inference; the theorised sequence from awareness through credibility and trust to purchase is a modelling assumption, and reciprocal specifications remain observationally compatible with the data. Second, AI awareness was measured as a self-reported trait rather than manipulated, which both explains the divergence from experimental findings and prevents this study from resolving it; a design combining trait measurement with experimental disclosure manipulation would separate the two mechanisms directly. Third, elevated inner-model collinearity among the three predictors of purchase decision means that the non-significant paths for credibility and seller trust should be regarded as unresolved rather than as evidence of absent effects. Fourth, purchase decision was measured with two self-report items and captures stated rather than observed behaviour; platform transactional data would strengthen the outcome measure. Fifth, out-of-sample predictive validity was not assessed; PLSpredict and cross-validated predictive ability tests should be reported in subsequent work (Shmueli et al., 2019). Sixth, the sample is a non-probability online panel sample fielded over three days. Panel members are by definition smartphone-owning and habituated to responding to surveys, which plausibly skews the sample toward higher digital engagement than the general social commerce population and may inflate the AI-literacy mechanism proposed in the discussion of AI awareness above. The sample is also heavily concentrated on two platforms: only three respondents nominated Instagram Shopping as their primary channel, so the findings describe marketplace-embedded social commerce rather than feed-native social commerce. Age was captured in bands rather than continuously, which precludes its use as a covariate.

Four directions follow. Cross-national replication would establish whether the positive awareness–credibility association reflects a general emerging-market pattern or a specifically Indonesian one, and would test the machine-heuristic explanation against th e alternative that low baseline trust in commercial content moderates the effect. Experimental work varying disclosure specificity, placement and timing

would identify which features of a disclosure carry the signal. Longitudinal designs would reveal whether the credibility premium associated with AI transparency persists or decays as AI-generated content saturates the marketplace, an important question given evidence that undisclosed AI use, once suspected, provokes distinct and negative responses (Amos & Zhang, 2024). Finally, category-level moderation deserves attention: the credibility stakes of AI involvement plausibly differ between low-risk fashion purchases and high-risk health or electronics categories, and Arango et al. (2023) show that domain sensitivity conditions responses to AI-generated appeals.

Conclusion

This study examined how AI awareness and AI-use disclosure shape information credibility, multidimensional trust and purchase decisions among Indonesian social commerce consumers, and it reaches three conclusions. First, transparency about AI is a credibility asset rather than a liability in this market: both AI awareness and AI-use disclosure raised perceived information credibility, contrary to the penalty reported in experimental studies from mature Western markets, and disclosure was the stronger of the two. Second, disclosure operates as a direct quality signal rather than as a buffer, because its effect did not depend on the level of AI awareness, which narrows the reach of persuasion-knowledge explanations in routine commercial settings. Third, credibility reaches purchasing through an institutional route: credibility strongly built both seller and information trust and platform trust, but only platform trust converted that credibility into purchase decisions. These conclusions are bounded by several limitations, namely the cross-sectional design, the self-reported measurement of AI awareness, elevated collinearity among the predictors of purchase decision, a two-item outcome measure, the absence of out-of-sample predictive testing, and a non-probability online panel concentrated on Shopee and TikTok Shop. Accordingly, sellers are recommended to disclose AI assistance voluntarily and explicitly as part of routine product presentation; platforms are recommended to adopt uniform, platform-issued provenance labelling supported by visible structural assurance such as escrow, dispute resolution and seller sanctions; and regulators are recommended to pair disclosure obligations with verification infrastructure and consumer digital literacy programmes. Future research should replicate the model across emerging and mature markets, combine trait measurement of awareness with experimental manipulation of disclosure specificity, placement and timing, adopt longitudinal designs to track whether the credibility premium persists as AI content saturates the marketplace, and use transactional data to measure purchase behaviour directly.

Declarations

Conflicts of interest: The authors declare no conflict of interest.

Data availability: The dataset and SmartPLS project file are available from the corresponding author on reasonable request.

Author contributions: [All author contributions is equal]

Use of generative AI: [DeepL for translation; Grammarly for grammar checking]

References

Amos, C., & Zhang, L. (2024). Consumer reactions to perceived undisclosed ChatGPT usage in an online review context. *Telematics and Informatics*, *93*, 102163. https://doi.org/10.1016/j.tele.2024.102163

Arango, L., Singaraju, S. P., & Niininen, O. (2023). Consumer Responses to AI-Generated Charitable Giving Ads. *Journal of Advertising*, *52*(4), 486–503. https://doi.org/10.1080/00913367.2023.2183285

Baek, T. H., Kim, J., & Kim, J. H. (2026a). Effect of disclosing AI-generated content on prosocial advertising evaluation. *International Journal of Advertising*, *45*(1), 171–192. https://doi.org/10.1080/02650487.2024.2401319

Baek, T. H., Kim, J., & Kim, J. H. (2026b). Effect of disclosing AI-generated content on prosocial advertising evaluation. *International Journal of Advertising*, *45*(1), 171–192. https://doi.org/10.1080/02650487.2024.2401319

Becker, J.-M., Cheah, J.-H., Gholamzade, R., Ringle, C. M., & Sarstedt, M. (2023). PLS-SEM's most wanted guidance. *International Journal of Contemporary Hospitality Management*, *35*(1), 321–346. https://doi.org/10.1108/IJCHM-04-2022-0474

Brüns, J. D., & Meißner, M. (2024). Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity. *Journal of Retailing and Consumer Services*, *79*, 103790. https://doi.org/10.1016/j.jretconser.2024.103790

Campaign Asia-Pacific. (2025). *Southeast asia boom markets: How Indonesia is redefining social commerce.*

Campbell, A. D., Fatoyinbo, L., Goldberg, L., & Lagomasino, D. (2022). Global hotspots of salt marsh change and carbon emissions. *Nature*, *612*(7941), 701–706. https://doi.org/10.1038/s41586-022-05355-z

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet Research*, *18*(3), 229–247. https://doi.org/10.1108/10662240810883290

Eisend, M., & Tarrahi, F. (2022). Persuasion Knowledge in the Marketplace: A Meta-Analysis. *Journal of Consumer Psychology*, *32*(1), 3–22. https://doi.org/10.1002/jcpy.1258

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and tam in online shopping: AN integrated model. *MIS Quarterly: Management Information Systems*, *27*(1), 51–90. https://doi.org/10.2307/30036519

Grigsby, J. L., Michelsen, M., & Zamudio, C. (2025). Service ads in the era of generative AI: Disclosures, trust, and intangibility. *Journal of Retailing and Consumer Services*, *84*, 104231. https://doi.org/10.1016/j.jretconser.2025.104231

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hajli, N. (2015). Social commerce constructs and consumer's intention to buy. *International Journal of Information Management*, *35*(2), 183–191. https://doi.org/10.1016/j.ijinfomgt.2014.12.005

Kang, H. (2021). Sample size determination and power analysis using the G*Power software. In *Journal of Educational Evaluation for Health Professions* (Vol. 18). Korea Health Personnel Licensing Examination Institute. https://doi.org/10.3352/JEEHP.2021.18.17

Kim, D. Y., & Kim, H. Y. (2021). Influencer advertising on social media: The multiple inference model on influencer-product congruence and sponsorship disclosure. *Journal of Business Research*, *130*, 405–415. https://doi.org/10.1016/j.jbusres.2020.02.020

Kirk, C. P., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. *Journal of Business Research*, *186*, 114984. https://doi.org/10.1016/j.jbusres.2024.114984

Kirkby, A., Baumgarth, C., & Henseler, J. (2023a). To disclose or not disclose, is no longer the question – effect of AI-disclosed brand voice on brand authenticity and attitude. *Journal of Product and Brand Management*, *32*(7), 1108–1122. https://doi.org/10.1108/JPBM-02-2022-3864

Kirkby, A., Baumgarth, C., & Henseler, J. (2023b). To disclose or not disclose, is no longer the question – effect of AI-disclosed brand voice on brand authenticity and attitude. *Journal of Product and Brand Management*, *32*(7), 1108–1122. https://doi.org/10.1108/JPBM-02-2022-3864

Kock, N. (2015). Common Method Bias in PLS-SEM. *International Journal of E-Collaboration*, *11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, *28*(1), 227–261. https://doi.org/10.1111/isj.12131

Koning, B., & Voorveld, H. A. M. (2025). Disclaimer! This Content Is AI-Generated: How AI-Disclosures Influence Trust in Advertisements and Organizations. *Journal of Interactive Advertising*, *25*(3), 240–253. https://doi.org/10.1080/15252019.2025.2554149

Li, H., & See-To, E. W. K. (2024). Source credibility plays the central route: an elaboration likelihood model exploration in social media environment with demographic profile analysis. *Journal of Electronic Business & Digital Economics*, *3*(1), 36–60. https://doi.org/10.1108/jebde-10-2022-0038

Liu, X., & Zheng, X. (2024). The persuasive power of social media influencers in brand credibility and purchase intention. *Humanities and Social Sciences Communications*, *11*(1), 15. https://doi.org/10.1057/s41599-023-02512-1

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

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to Medical Artificial Intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Mayer, R. C., Davis, J. H., & David Schoorman, F. (1995). *The Academy of Management Review* (Vol. 20, Number 3). https://about.jstor.org/terms

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

Metzger, M., Flanagin, A., Eyal, K., Lemus, D., & McCann, R. (2003). Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. *Annals of the International Communication Association*, *27*, 293–335. https://doi.org/10.1080/23808985.2003.11679029

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Paul A. Pavlou. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*, *7*(3), 101–134. https://doi.org/10.1080/10864415.2003.11044275

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L., & Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology and Marketing*, *39*(5), 1035–1064. https://doi.org/10.1002/mar.21640

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing*, *53*(11), 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189

Spence, M. (1973). Job Market Signaling Author (s): Michael Spence Source : The Quarterly Journal of Economics , Vol . 87 , No . 3 (Aug ., 1973), pp . 355-374 Published by : The MIT Press Stable URL : http://www.jstor.org/stable/1882010. *Press, The M I T Journal, The Quarterly*, *87*(3), 355–374.

Sundar, S. S. (2008). The main model: A Heuristic Approach to understanding Technology Effects on Credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital Media and Learning* (pp. 73–100).

Sundar, S. S., & Kim, J. (2019a). Machine Heuristic. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3290605.3300768

Sundar, S. S., & Kim, J. (2019b). Machine Heuristic. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3290605.3300768

Wang, J., Shahzad, F., Ahmad, Z., Abdullah, M., & Hassan, N. M. (2022). Trust and Consumers' Purchase Intention in a Social Commerce Platform: A Meta-Analytic Approach. *SAGE Open*, *12*(2). https://doi.org/10.1177/21582440221091262