Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 474-502
DOI: https://doi.org/10.51903/jtie.v5i1.568

Psychology-Informed Live-Commerce Analytics: Structural Modeling, Predictive Validation, and Governed Script Scoring for TikTok Shopping

Xiaochen Li[1], Yinchen Shi*[2], Jason Zhang[3]
Email: ogyinchen@gmail.com
[1]*Marketing Analytics, Bentley University, MA, USA*
[2]*Computer Science, New York University, NY, USA*
[3]*Computer Science, Cornell Tech, NY, USA*
*Corresponding Author

Abstract

This paper examines the psychological factors associated with impulsive-purchase propensity in TikTok Live Commerce and translates those associations into an auditable analytics workflow. It presents a secondary analysis of a public survey workbook containing 415 Indonesian responses on Social Attraction, fear of missing out, Narrative Involvement, Telepresence, Parasocial Interaction, Social Presence, and Impulsive Purchase. The analysis combines measurement diagnostics, a bootstrapped composite structural model, repeated cross-validated classification and continuous prediction, calibration assessment, a coefficient-based illustrative scenario analysis, and external validation of an ethical microcopy screen. In the full sample, Narrative Involvement has the largest association with Parasocial Interaction (β = .516) and Social Presence (β = .496); Parasocial Interaction and Social Presence jointly account for 67.5% of the in-sample variance in Impulsive Purchase. The pattern remains in the quality-screened TikTok-viewer subset (n = 238; Impulsive Purchase R^2 = .647). For the quality-screened subset, the strongest full-feature classifier is Random Forest (ROC-AUC = .938 ± .031 across 50 held-out folds), while Ridge regression gives the largest mean continuous-outcome R^2 (.622 ± .114). On 2,356 ec-darkpattern strings, a page-grouped hybrid word/character classifier reaches F1 = .959 ± .011 and ROC-AUC = .989 ± .003. The scenario scores are deterministic implications of the estimated coefficients rather than observed purchase effects. The technical contribution is a governed decision-support architecture that keeps structural association, contemporaneous prediction, scenario prioritization, ethical screening, human review, and prospective testing distinct.

Keywords: *TikTok Live Commerce, Impulsive Purchase, Fear of Missing Out, Structural Modeling, Cross-Validation.*

I. INTRODUCTION

Live commerce has shifted digital marketing from static product display to real-time, creator-led persuasion. In TikTok shopping sessions, viewers evaluate product attributes while also responding to storylines, crowd reactions, scarcity cues, host intimacy, and immersive demonstrations. For organizations, this creates both an innovation opportunity and an informatics challenge: teams need to distinguish which psychological states are associated with purchase propensity, which signals can support prediction, and which candidate messages should advance to controlled testing.

This study links social-commerce theory with an auditable marketing-analytics workflow through a secondary analysis of item-level survey data originally collected in Indonesia and released publicly by Hadiyati and Luthfia (2026a, 2026b, 2026c). The behavioral specification follows Stimulus–Organism–Response logic: Social Attraction, fear of missing out (FOMO), Narrative Involvement, and Telepresence are modeled as stimuli; Parasocial Interaction and Social Presence

Received on January, 2026; Revised on February, 2026; Accepted on March, 2026; Published on April, 2026

are parallel organismic states; and Impulsive Purchase is the response construct (Mehrabian & Russell, 1974). The secondary-analysis framing distinguishes the already published behavioral model from the data-quality, predictive, scenario-scoring, and ethical-screening analyses conducted here.

The organizational problem is broader than estimating one propensity score. Live-commerce teams must match products to viewers, decide which script should be tested next, and distinguish incremental persuasion from purchases that would have occurred without the live session. Transaction-sequence representations and language-style personas can support next-purchase prediction and product ranking (Lu & Zhou, 2023; Xin, 2026), while large-language-model (LLM) reranking and product-classification research suggests that behavioral signals are most useful when paired with explainable recommendation logic (Chang et al., 2023; Xu et al., 2024). The present survey contains no transaction log, so its outcome is treated as contemporaneous self-reported purchase propensity rather than conversion.

The objective is methodological and practical. First, the study reassesses the public measurement model and structural associations after explicit eligibility and response-quality checks. Second, it compares classification and continuous-prediction models under repeated cross-validation, alternative target thresholds, calibration diagnostics, and held-out permutation analysis. Third, it converts the estimated indirect associations into a transparent coefficient-based scenario score and evaluates a dark-pattern risk screen on an external e-commerce microcopy benchmark. Candidate scripts may be written by a person or generated by a language model; the evaluation layers are source-agnostic, and no particular generative model is tested here.

Governance is central to this workflow. Generated selling language can be fluent yet unsupported, overly aggressive, or inconsistent with product evidence. Enterprise LLM safety research recommends evidence consistency checks and continual red-teaming before text reaches users (C. Li et al., 2023; D. Zheng et al., 2023). In live commerce, these controls matter because inventory, discounts, delivery promises, and scarcity claims must be accurate and ethically framed. The technical contribution is therefore an auditable sequence connecting data screening, measurement diagnostics, structural association, prediction, scenario prioritization, ethical text screening, and human review. This sequence strengthens the paper's fit with technology informatics and engineering while maintaining distinct interpretations for cross-sectional paths, predictive estimates, and scenario scores.

II. LITERATURE REVIEW

A. Behavioral mechanisms in live commerce

Impulsive purchase is a sudden, affect-laden buying response that occurs with limited prior planning (Rook, 1987). Livestreaming commerce can intensify that response because purchase opportunities unfold while the consumer is emotionally engaged, socially visible, and exposed to time-sensitive persuasion (B. Chen et al., 2022). Compared with static e-commerce, live commerce adds synchronous social proof, host responsiveness, entertainment, and real-time calls to action.

Parasocial Interaction explains why viewers may treat streamers as familiar partners rather than anonymous sellers. Horton and Wohl (1956) described parasocial interaction as an illusion of face-to-face relationship with a media persona; later measurement work emphasizes experienced closeness and reciprocal-seeming attention (Hartmann & Goldhoorn, 2011). Social Presence complements this relationship by capturing the feeling that other people are present in the mediated environment (Short et al., 1976; Biocca et al., 2003). When chat reactions and viewer purchases feel real, a livestream may be experienced as a shared shopping event (L.-R. Chen et al., 2023).

FOMO reflects apprehension that others are gaining rewarding experiences without oneself (Przybylski et al., 2013). In live commerce, countdowns, limited stock, and comment activity can make opportunity costs salient, but urgency language also creates an ethical boundary. Narrative Involvement draws from transportation theory: stories absorb attention and can reduce counter-arguing (Green & Brock, 2000). Prior social-commerce research links narrative involvement, parasocial interaction, and impulse buying (Vazquez et al., 2020).

Telepresence is the subjective sensation of being located within a mediated environment (Kim & Biocca, 1997). Product close-ups, responsive demonstrations, and a stable stream may strengthen that sensation, while poor playback can interrupt it (Zahid et al., 2024). Social Attraction captures perceived likability and similarity rather than physical attractiveness alone; its source measurement tradition treats interpersonal attraction as multidimensional (McCroskey & McCain, 1974). These constructs jointly support the dual-pathway specification assessed in the public survey.

B. Predictive, recommendation, and policy analytics

The recommendation layer extends the psychological model into product and audience decisions. Online-retail studies show that recency, frequency, purchase histories, and learned embeddings remain strong signals for segmentation and next-purchase prediction (Lu & Zhou, 2023; Xin, 2026). LLM-as-reranker work adds a natural-language explanation layer, while review-grounded recommendation emphasizes that explanations should remain faithful to the evidence supporting

a ranking (Chang et al., 2023, 2026). Intelligent product classification likewise shows how structured attributes can support personalized merchandising (Xu et al., 2024).

Propensity is not incrementality. Uplift modeling for email advertising, uncertainty-aware targeting, privacy-robust incrementality estimation, and off-policy policy evaluation all emphasize that a recommendation should be tested for incremental impact and deployment risk rather than judged only by predicted response (Mu et al., 2023; Lu et al., 2024; Bai et al., 2026; Ye et al., 2026). Offline counterfactual evaluation offers a bridge between logged behavior and safer policy selection (Mu et al., 2025). LLM-assisted attribution can help managers summarize heterogeneous pathways, but it does not replace identification assumptions or experimental design (Zhao et al., 2023). Federated topic-preference learning with differential privacy illustrates how personalization can reduce raw-data exposure (Kuo et al., 2025).

Predictive assessment should match the decision target. A median split is convenient but discards information and can be sensitive to ties. Accordingly, the present analysis adds continuous prediction, 40th- and 60th-percentile thresholds, precision-recall area under the curve (PR-AUC), calibration error, Brier score, confusion matrices, and held-out importance stability. These additions follow predictive-assessment guidance that separates explanatory fit from out-of-sample performance (Shmueli et al., 2019) and recognizes the value of PR-AUC when class balance changes (Saito & Rehmsmeier, 2015).

C. Governed script evaluation and ethical persuasion

The content layer should translate behavioral evidence into reviewable message options, not substitute fluency for validation. A script that sounds energetic may overuse urgency while neglecting narrative involvement or telepresence. Evidence-chain retrieval-augmented generation (RAG), scientific-search evidence cards, and evidence-calibrated abstention show how claims can be connected to source records, provenance explanations, and coverage checks (C. Li et al., 2024; Jin, 2025; Z. S. Zhong et al., 2025). Budgeted multi-hop retrieval and narrative-aware claim verification are relevant where product catalogs, promotion rules, reviews, and shipping constraints must be checked together (Su et al., 2025; Su, Chen, & Qian, 2026).

Safety and integrity controls add refusal and review. VerifySafe-style self-verification and behavior-level jailbreak resistance seek to preserve utility while rejecting unsafe or manipulated requests (Zheng & Li, 2024; D. Zheng et al., 2024). Privacy and data-integrity risk cards offer a practical interface for human oversight when prompt injection, private-data leakage, or unsupported claims could affect decisions (Su, Rao, & Ma, 2026).

Visual and interaction design also shapes how analytics are consumed. Trust-calibrated recommendation cards, text-grounded rationale cards, visual brief cards, and LLM-as-design-critic frameworks translate signals into interfaces that nontechnical teams can inspect (B. Zhang et al., 2025; Q. Wu et al., 2025; H. Tu et al., 2025; Y. Li et al., 2025). Structured visual briefs make creative intentions editable rather than turning them into one-off generated outputs (J. Mu et al., 2026). Dark-pattern detection provides a complementary control because urgency and scarcity language can cross from persuasion into deceptive microcopy (H. Xu et al., 2025; Yada et al., 2022).

D. Transferable implementation and governance patterns

The adjacent-domain studies in this subsection specify implementation controls—evidence provenance, calibration, abstention, service quality, security, and review interfaces—and are not evidence for the live-commerce behavioral paths. Telepresence has an operational analogue in video quality of experience (QoE). Adaptive bitrate (ABR) optimization for HTTP Live Streaming and Dynamic Adaptive Streaming over HTTP, risk-sensitive ABR under mobile traces, content-adaptive AV1/H.264 selection, and Video Multimethod Assessment Fusion proxies demonstrate how bitrate, rebuffering, time to first frame, codec delay, and perceived quality can be monitored (Y. Li, 2023b; Wang & Wang, 2023; Wang & Wu, 2024; B. Zhou et al., 2025). The survey contains no stream telemetry, but the architecture permits later linkage between QoE logs and Telepresence scores.

Conversational analytics should remain executable and verifiable. Execution-feedback text-to-SQL and self-correcting SQL agents show how business questions can be translated into queries while using errors to reduce invalid results (Y. Li, 2023a; S. Chen et al., 2024). Numerical guardrails similarly suggest checking generated inventory, discount, and delivery claims against trusted tables before release (Z. Li et al., 2026).

Calibration, rejection, and cost asymmetry are also transferable. Credit-risk tiering with uncertainty-driven rejection, cost-sensitive fraud modeling, model-risk-oriented probability workflows, and explanation-enhanced gradient boosting show how confidence, error costs, and attribution can be documented for organizational review (Y. Chen et al., 2023; Jin et al., 2024a, 2024b; Zhou & Zhao, 2024). For multi-tool systems, trajectory reliability prediction helps localize failures to retrieval, scoring, query execution, or tool sequencing rather than only the final text (Z. S. Zhong et al., 2026).

E. Analytical rationale

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 474-502
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Partial least squares structural equation modeling (PLS-SEM) is often used for composite-based latent-variable models and mediated paths (Hair et al., 2019;

Hair & Alamer, 2022). The present rerun uses equally weighted construct scores and standardized structural regressions, matching the published path specification while making the score construction explicit. XGBoost provides a complementary nonlinear prediction lens (Chen & Guestrin, 2016). Measurement interpretation is supplemented with the heterotrait-monotrait ratio (HTMT), common-method diagnostics, and quality-screened sensitivity analyses (Henseler et al., 2015; Podsakoff et al., 2003; Kock, 2015; Rönkkö & Cho, 2022).

III. RESEARCH METHOD

A. Data provenance, sampling, and consent

This study is a secondary analysis of the public item-level survey described by Hadiyati and Luthfia (2026a, 2026c). The raw workbook analyzed here is the Excel file associated with Mendeley Data Version 2; the retained Version 4 citation identifies the associated article deposit (Hadiyati & Luthfia, 2026b). The source study reports purposive convenience sampling: active viewers were approached through direct messages following livestream sessions, and electronic consent was obtained before participation. The present authors did not recruit or contact respondents and analyzed the public file under its CC BY 4.0 license.

The workbook contains 416 submissions. One record contains demographics but no psychometric responses and was excluded, leaving 415 records with scale data. The source report describes active TikTok Live Commerce users, whereas the workbook contains 53 records answering that they had never watched e-commerce livestreaming. To preserve transparency, the full 415-record sample is used to reproduce the published pattern and is compared with eligibility- and quality-restricted analyses.

The source ethics statement reports that formal ethical approval was not required for the anonymous, minimal-risk online survey and that electronic consent was obtained. The public age category "<20 years" does not distinguish respondents aged 18–19 from younger respondents. The secondary file therefore cannot independently verify the source report that all participants were adults, and age-bin sensitivity cannot be reconstructed from the released categories.

B. Measures and data-quality screening

The questionnaire uses five-point Likert items for Social Attraction (SA), FOMO, Narrative Involvement (NI), Telepresence (TE), Parasocial Interaction (PI), Social Presence (SP), and Impulsive Purchase (IP). The source instruments draw respectively on McCroskey and McCain (1974); Przybylski et al. (2013); Green and Brock (2000) and Vazquez et al. (2020); Kim and

Biocca (1997) and Zahid et al. (2024); Horton and Wohl (1956) and Hartmann and Goldhoorn (2011); Biocca et al. (2003) and L.-R. Chen et al. (2023); and Rook (1987) and G. Li et al. (2022). Appendix A reports all retained item wording.

Screening included range checks, all-blank psychometric records, item-level missingness, full-vector response invariance, and exact duplicate response patterns. The workbook contains 129 candidate psychometric fields. FOMO2c and IP1b are each missing in 286 of 415 records and do not appear in the published 127-item questionnaire; both were excluded rather than imputed. Construct scores are unweighted means of the remaining 17 SA, 19 FOMO, 17 NI, 18 TE, 20 PI, 19 SP, and 17 IP items.

Two analysis sets were used. The full set includes all 415 records with psychometric data. The strict set requires a prior-livestream response of "yes," selection of TikTok Live, a nonconstant 127-item vector, and exclusion from the four-row nonconstant exact-duplicate pattern, leaving 238 records. Repetition does not itself prove invalid participation; the strict set is used to show how the conclusions change under a conservative screen.

C. Measurement and structural analysis

Internal consistency was summarized by Cronbach's alpha and composite reliability; convergent validity by average variance extracted (AVE) and loading ranges; and discriminant validity by HTMT and the Fornell–Larcker matrix. An unrotated one-factor share and full-collinearity variance inflation factors (VIFs) were reported as descriptive common-method diagnostics. These checks identify risk rather than proving the presence or absence of method bias.

The structural stage standardizes the seven construct scores and estimates three equations: PI from SA, FOMO, and NI; SP from NI and TE; and IP from PI and SP. Direct and indirect associations, standard errors, and percentile confidence intervals were estimated with 5,000 bootstrap resamples. Because all constructs were measured in one cross-sectional survey, the coefficients are interpreted as structural associations, not temporal or causal effects.

D. Predictive modeling and robustness analysis

IP was modeled first as a continuous score and second as a binary high-propensity target. Binary labels were defined as IP strictly above the sample-specific 40th, 50th, or 60th percentile. The median rule produces 207 positive and 208 negative cases in the full sample and 117 positive and 121 negative cases in the strict sample.

Two feature sets were compared: a stimulus-only set (SA, FOMO, NI, and TE) and a full contemporaneous set adding PI and SP. Logistic regression used C = 1, balanced class weights, and standardized inputs. Random Forest used 100 trees, square-root feature sampling, and a

minimum leaf size of 3. XGBoost used 150 trees, learning rate .03, maximum depth 3, minimum child weight 3, 80% row and column subsampling, and L2 regularization. Fixed settings were chosen before comparison to avoid unstable small-sample hyperparameter selection.

Each classifier was evaluated with 10 repeats of stratified five-fold cross-validation using seed 20260718. Preprocessing and fitting occurred within each training fold. Reporting includes accuracy, balanced accuracy, precision, recall, specificity, F1, receiver operating characteristic area under the curve (ROC-AUC), PR-AUC, Brier score, log loss, expected calibration error over ten equal-frequency bins (ECE-10), repeated out-of-fold confusion matrices, and calibration curves. Held-out permutation importance was recomputed with five permutations in each of the 50 test folds. Continuous IP was predicted with Ridge (alpha = 1 after standardization), Random Forest, and XGBoost regressors under repeated five-fold cross-validation; the ensemble settings matched the classifier settings. Continuous prediction was assessed by mean absolute error (MAE), root mean squared error (RMSE), R^2, and Spearman correlation.

E. Coefficient-based illustrative scenario analysis

Four fixed cue profiles represent baseline host appeal, host appeal plus urgency/FOMO, host appeal plus narrative and telepresence, and an integrated profile. Each profile has a binary vector for SA, FOMO, NI, and TE, as reported in Table 5. The full-sample standardized estimates define PI* = .118SA + .230FOMO + .516NI; SP* = .496NI + .337TE; and IP* = .418PI* + .430SP. IP is divided by the integrated profile value and multiplied by 100.

The cue rows are constructed definitions, not ratings of free-form scripts; no subjective coder decision enters the calculation. The resulting scores compare implications of the fitted coefficients and do not estimate a change in purchase behavior. Behavioral validation would require randomized exposure to complete scripts and independently observed outcomes.

F. Ethical microcopy benchmark

The ec-darkpattern benchmark contains 2,356 English e-commerce interface strings: 1,178 dark-pattern and 1,178 acceptable examples across 1,248 page identifiers (Yada et al., 2022). Four classifiers were evaluated: word TF-IDF with logistic regression, character TF-IDF with logistic regression, hybrid word/character TF-IDF with logistic regression, and hybrid TF-IDF with a Linear SVM calibrated by three-fold sigmoid calibration. The primary protocol used five repeats of five-fold StratifiedGroupKFold, keeping every page identifier within one fold; ordinary stratified validation was retained as a comparator. Results are summarized across 25 held-out folds with fold standard deviations and 10,000-resample fold-bootstrap intervals. A 0.50 threshold generated the consensus confusion matrix from each text's mean repeated out-of-fold probability.

The ethical classifier is a screening layer. Product truthfulness, hallucination, legal compliance, and rare dark-pattern categories still require evidence checks and human review. This division of labor follows the evidence and risk-card literature rather than presenting a text classifier as an ethical guarantee.

G. Decision-support architecture

Figure 1 organizes the empirical components into an engineering workflow. Survey evidence passes through quality and measurement checks before supporting structural and predictive outputs. Candidate wording, whether human-authored or machine-generated, enters the cue-profile and ethical-screening layers. Only text that is supported by product evidence and passes human review advances to prospective randomized evaluation.

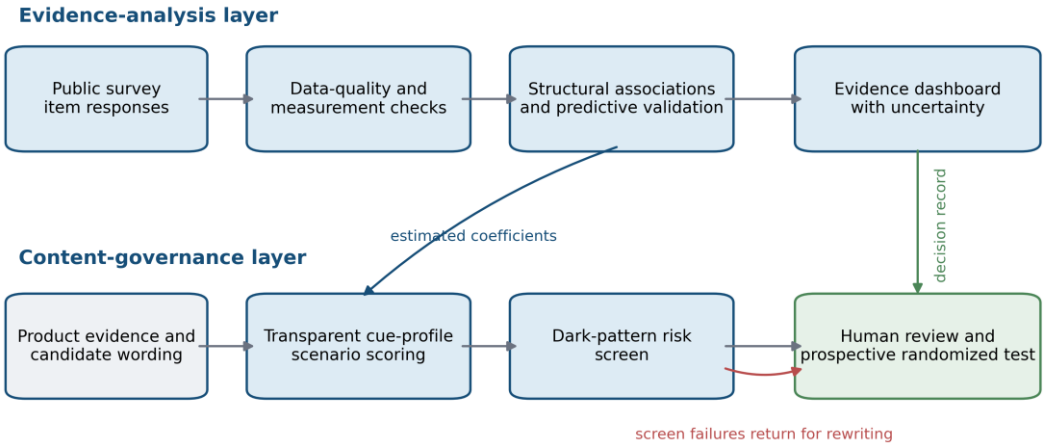

Figure 1. Auditable live-commerce decision-support workflow. Structural association, prediction, scenario scoring, ethical screening, human review, and prospective testing remain separate decision stages.

IV. RESULT AND DISCUSSION

A. Sample profile and response-quality screening

Table 1 summarizes the released sample and the two analysis sets. Women account for 50.8% of the 415 psychometric records and men for 49.2%. The sample is concentrated in the "<20 years" category (65.5%) and among students (54.7%). Although 342 respondents selected TikTok Live among the platforms they watched, only 304 both reported prior livestream viewing and selected TikTok; the strict response-quality and eligibility rules retain 238.

Table 1. Sample profile, eligibility, and response-quality screening.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 474-502
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Domain	Category	n	% of 415
Gender	Women	211	50.8
Gender	Men	204	49.2
Age	<20 years	272	65.5
Age	21–30 years	85	20.5
Age	31–40 years	38	9.2
Age	41–50 years	13	3.1
Age	>51 years	7	1.7
Occupation	Student	227	54.7
Occupation	Private-sector employee	49	11.8
Occupation	Government/state-owned enterprise	40	9.6
Occupation	Entrepreneur	28	6.7
Occupation	Other categories	71	17.1
Residence	Greater Jakarta	185	44.6
Residence	Other Java	53	12.8
Residence	Outside Java	177	42.7
Eligibility	Reported prior livestream viewing	362	87.2
Eligibility	Selected TikTok Live	342	82.4
Eligibility	Both prior viewing and TikTok Live	304	73.3
Quality screen	Straight-line response vector	82	19.8
Quality screen	Nonconstant exact-duplicate rows	4	1.0
Analysis set	Full released analytic sample	415	100.0
Analysis set	Quality-screened TikTok-viewer subset	238	57.3

Two workbook fields, FOMO2c and IP1b, have 286 missing values each (68.9%) and were excluded consistently. All 127 retained items are complete. Eighty-two records (19.8%) use a single response value across all retained items; 85 records occur in exact repeated-vector groups, of which four are nonconstant. Both the full and strict results are reported to show sensitivity to response-pattern screening.

B. Measurement diagnostics

Table 2 reports the full-sample measurement results. Alpha ranges from .981 to .990, composite reliability from .983 to .991, and AVE from .772 to .862. These estimates show high internal consistency, but values close to .99 can also reflect overlapping item wording.

Table 2. Measurement reliability, convergent validity, discriminant validity, and full-collinearity diagnostics in the full sample (n = 415).

Construct	Items	α	CR	AVE	Loading range	Highest HTMT (pair)	Full VIF
SA	17	.981	.983	.772	.851–.897	.821 (TE)	3.176
FOMO	19	.989	.989	.829	.821–.947	.881 (NI)	4.996
NI	17	.989	.990	.847	.889–.937	.909 (TE)	7.141
TE	18	.989	.989	.839	.885–.942	.909 (NI)	7.363
PI	20	.990	.991	.843	.902–.933	.886 (SP)	5.608
SP	19	.990	.991	.848	.886–.943	.886 (PI)	5.281
IP	17	.990	.991	.862	.894–.944	.804 (SP)	3.332

The unrotated first factor accounts for 68.3% of indicator variance in the full sample and 61.7% in the strict sample. The largest full-sample HTMT is .909 for TE–NI, and the largest full-sample collinearity VIF is 7.363 for TE. In the strict sample, the largest HTMT remains .880. These values indicate material construct overlap and common-source risk, especially around NI/TE and PI/SP; the structural paths are interpreted with that limitation.

C. Structural associations

Figure 2 presents the full-sample composite structural model. NI has the largest association with both PI (β = .516, 95% CI [.356, .668]) and SP (β = .496, 95% CI [.323, .653]). FOMO is positively associated with PI, TE with SP, and both mediators with IP. SA–PI has a confidence interval crossing zero.

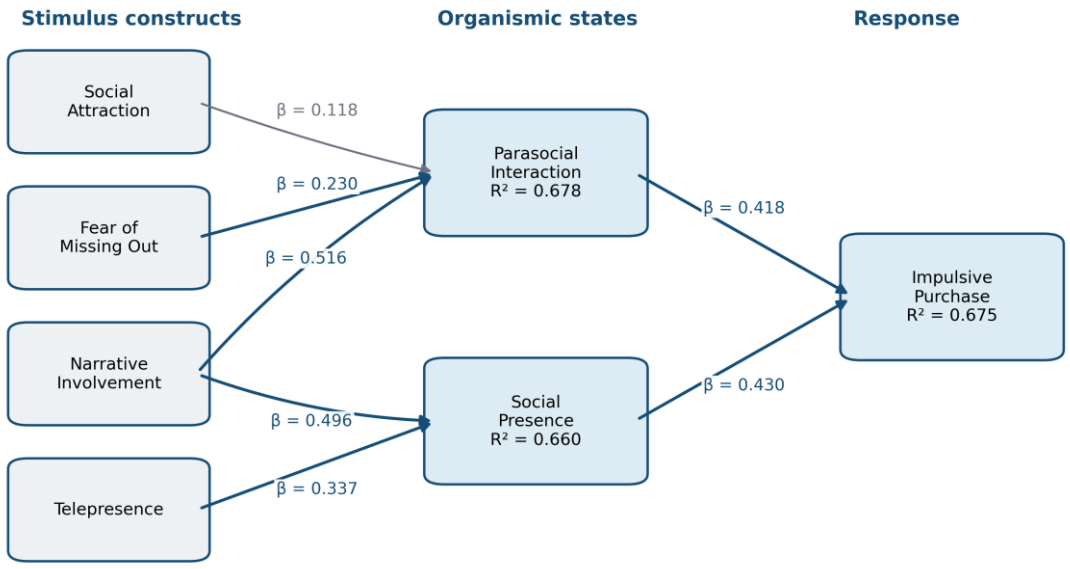

Figure 2. Standardized full-sample structural associations. SA = Social Attraction; FOMO = fear of missing out; NI = Narrative Involvement; TE = Telepresence; PI = Parasocial Interaction; SP = Social Presence; IP = Impulsive Purchase. Gray denotes an interval crossing zero.

Table 3 provides exact bootstrap intervals and the strict-sample sensitivity estimates. Full-sample R^2 is .678 for PI, .660 for SP, and .675 for IP. The corresponding strict-sample values are .556, .559, and .647. The direction and ordering of the six interval-supported direct paths remain stable.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 474-502
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

NI has two positive indirect associations with IP, while the SA → PI → IP interval crosses zero in both samples.

Table 3. Direct and indirect structural associations with 5,000-resample bootstrap intervals.

Association	Full β	Full 95% CI	Full p	Strict β	Strict 95% CI	Strict p
Panel A. Direct structural associations						
SA → PI	.118	[-.007, .237]	= .061	.108	[-.066, .259]	= .192
FOMO → PI	.230	[.106, .367]	< .001	.237	[.085, .408]	= .002
NI → PI	.516	[.356, .668]	< .001	.450	[.249, .644]	< .001
NI → SP	.496	[.323, .653]	< .001	.414	[.221, .594]	< .001
TE → SP	.337	[.175, .508]	< .001	.359	[.166, .552]	< .001
PI → IP	.418	[.223, .624]	< .001	.370	[.127, .643]	= .001
SP → IP	.430	[.223, .624]	< .001	.467	[.187, .711]	< .001
Panel B. Indirect associations						
SA → PI → IP	.049	[-.002, .125]	= .061	.040	[-.017, .137]	= .192
FOMO → PI → IP	.096	[.039, .174]	< .001	.088	[.023, .187]	= .003
NI → PI → IP	.216	[.102, .351]	< .001	.166	[.047, .331]	= .001
NI → SP → IP	.213	[.099, .337]	< .001	.194	[.064, .340]	< .001
NI total indirect → IP	.429	[.309, .543]	< .001	.360	[.213, .505]	< .001
TE → SP → IP	.145	[.056, .264]	< .001	.168	[.049, .323]	< .001

D. Predictive performance, calibration, and stability

Table 4 reports the strict-sample classification and continuous-prediction results, including threshold sensitivity. Under the median rule, the strongest full-feature ROC-AUC is .938 ± .031 for Random Forest. At the 40th, 50th, and 60th percentile cutoffs, that model's ROC-AUC remains within the range shown in Panel B, demonstrating whether the ranking conclusion depends on a single median split. The best strict-sample continuous mean R^2 is .622 ± .114 for Ridge regression.

Table 4. Repeated cross-validated classification, threshold sensitivity, and continuous IP prediction in the quality-screened sample. Entries are fold means ± SD unless otherwise indicated.

Feature set/threshold	Model/classes	Accuracy	F1	ROC-AUC	PR-AUC/MAE	Brier/RMSE	ECE/R^2
Panel A. Median-threshold classification (n = 238)							
Stimulus only	Logistic regression	.829 ± .042	.818 ± .046	.880 ± .043	.901 ± .044	.131	.142

Feature set/threshold	Model/classes	Accuracy	F1	ROC-AUC	PR-AUC/MAE	Brier/RMSE	ECE/R^2
Stimulus only	Random Forest	.852 ± .042	.838 ± .049	.895 ± .038	.915 ± .035	.119	.118
Stimulus only	XGBoost	.857 ± .044	.842 ± .051	.890 ± .038	.911 ± .041	.118	.118
Stimulus + mediators	Logistic regression	.865 ± .043	.860 ± .045	.918 ± .041	.924 ± .044	.104	.125
Stimulus + mediators	Random Forest	.871 ± .038	.865 ± .040	.938 ± .031	.945 ± .031	.096	.090
Stimulus + mediators	XGBoost	.868 ± .037	.861 ± .040	.926 ± .032	.934 ± .037	.100	.104
Panel B. Threshold sensitivity: Random Forest, full features							
40th percentile	Positive 142; negative 96	.847 ± .042	.866 ± .040	.914 ± .027	—	—	—
50th percentile	Positive 117; negative 121	.871 ± .038	.865 ± .040	.938 ± .031	—	—	—
60th percentile	Positive 75; negative 163	.863 ± .039	.788 ± .062	.924 ± .033	—	—	—
Panel C. Continuous IP prediction (n = 238)							
Stimulus only	Ridge regression	—	—	—	MAE .532	RMSE .737	R^2 .494
Stimulus only	Random Forest	—	—	—	MAE .539	RMSE .731	R^2 .501
Stimulus only	XGBoost	—	—	—	MAE .544	RMSE .735	R^2 .495
Stimulus + mediators	Ridge regression	—	—	—	MAE .429	RMSE .636	R^2 .622

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 474-502
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Feature set/threshold	Model/classes	Accuracy	F1	ROC-AUC	PR-AUC/MAE	Brier/RMSE	ECE/R²
Stimulus + mediators	Random Forest	—	—	—	MAE .466	RMSE .647	R² .611
Stimulus + mediators	XGBoost	—	—	—	MAE .472	RMSE .666	R² .586

The full-sample comparison is shown in Figure 3. The strongest full-feature ROC-AUC is .967 ± .017 for Random Forest; the best full-sample continuous R^2 is .677 ± .077 for Ridge regression. Because the full feature set uses PI and SP collected contemporaneously with IP, these values describe same-survey classification rather than a forward forecast available before a livestream.

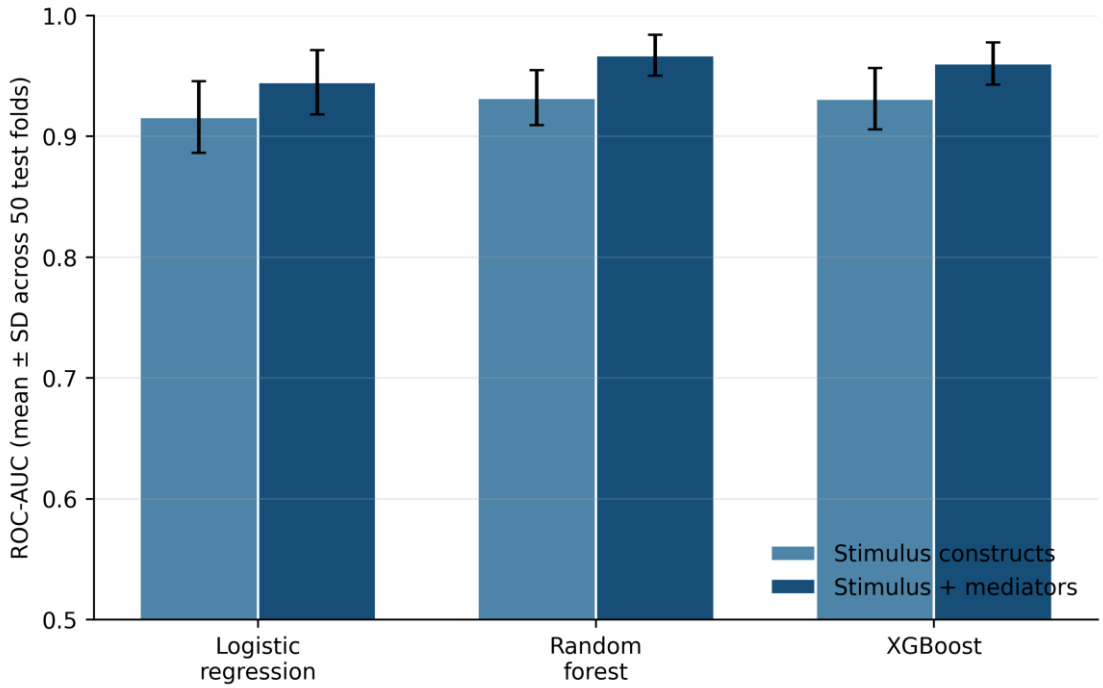

Figure 3. Full-sample ROC-AUC by classifier and feature set under 10 repeats of stratified five-fold cross-validation. Bars show fold means and error bars show fold-to-fold SD.

Figure 4 presents repeated out-of-fold calibration and consensus confusion matrices for the full-feature median target. It makes the operating errors visible rather than relying only on averaged accuracy. Brier, log-loss, and ECE-10 values in the run outputs further distinguish discrimination from probability quality.

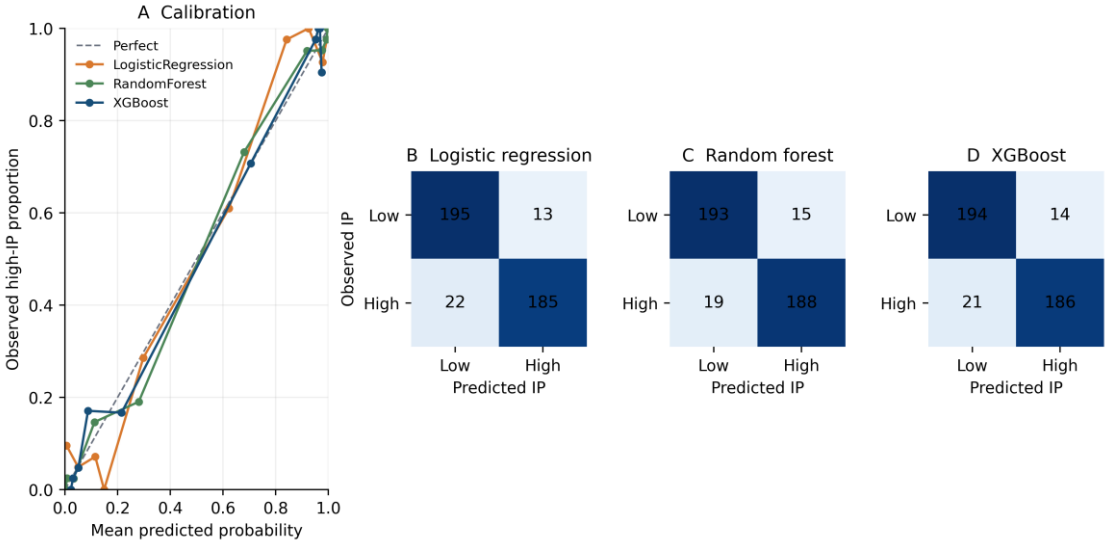

Figure 4. Repeated out-of-fold calibration and consensus confusion matrices for the full-sample, full-feature median target. Each probability is the mean of ten held-out predictions for the same respondent.

Figure 5 shows Random Forest permutation importance recomputed on held-out folds. SP has the largest mean ROC-AUC decrease (.045), while the mean pairwise Spearman correlation among fold rankings is .826. The variability demonstrates why a single fitted-model importance ranking would overstate stability.

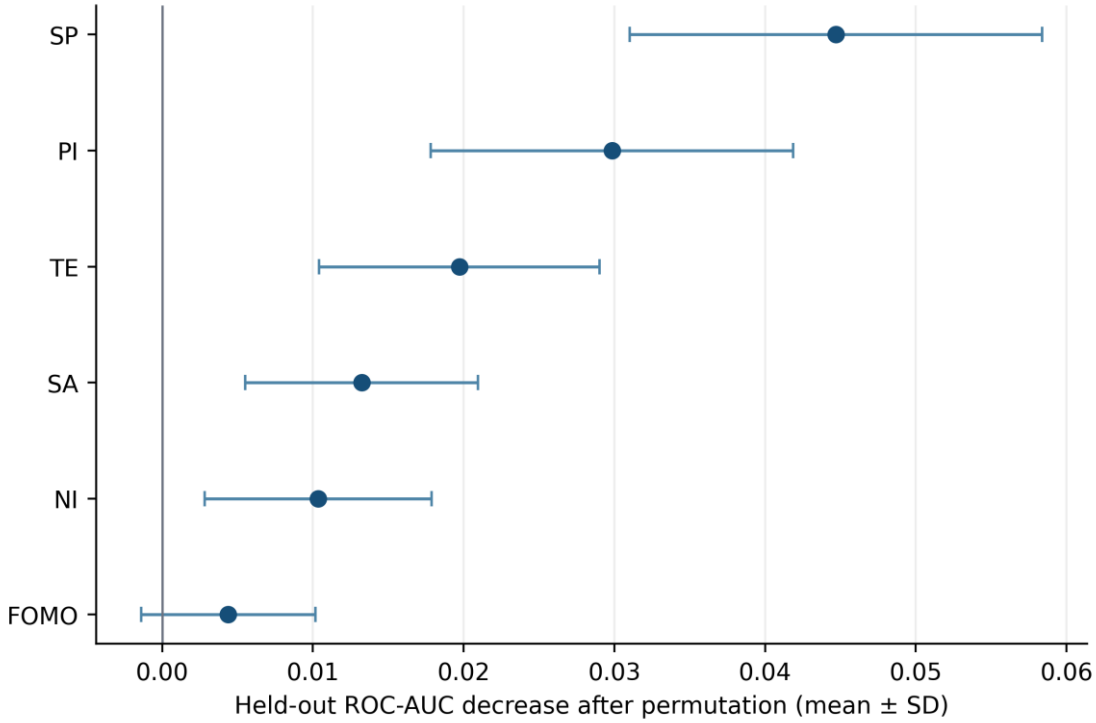

Figure 5. Held-out Random Forest permutation-importance stability for the full-sample median target. Points are mean ROC-AUC decreases, and error bars are SD across 50 test folds.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 474-502
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

E. Coefficient-based scenario scores

Table 5 reports every cue value and intermediate score, making the rubric directly inspectable. The integrated profile defines the common 100-point denominator. The narrative-plus-telepresence profile scores 86.6, compared with 20.2 for urgency/FOMO and 6.9 for baseline host appeal.

Table 5. Fixed cue matrix and coefficient-based illustrative scenario scores.

Scenario	SA	FOMO	NI	TE	PI*	SP*	IP*	Score
Baseline host appeal	1	0	0	0	.118	.000	.049	6.9
Urgency/FOMO	1	1	0	0	.348	.000	.146	20.2
Narrative + telepresence	1	0	1	1	.635	.833	.623	86.6
Integrated	1	1	1	1	.865	.833	.720	100.0

Figure 6 visualizes the same deterministic comparison. The ordering follows from the comparatively large indirect weight assigned to NI through both PI and SP. It is a prioritization device for selecting variants to test, not transaction-level validation.

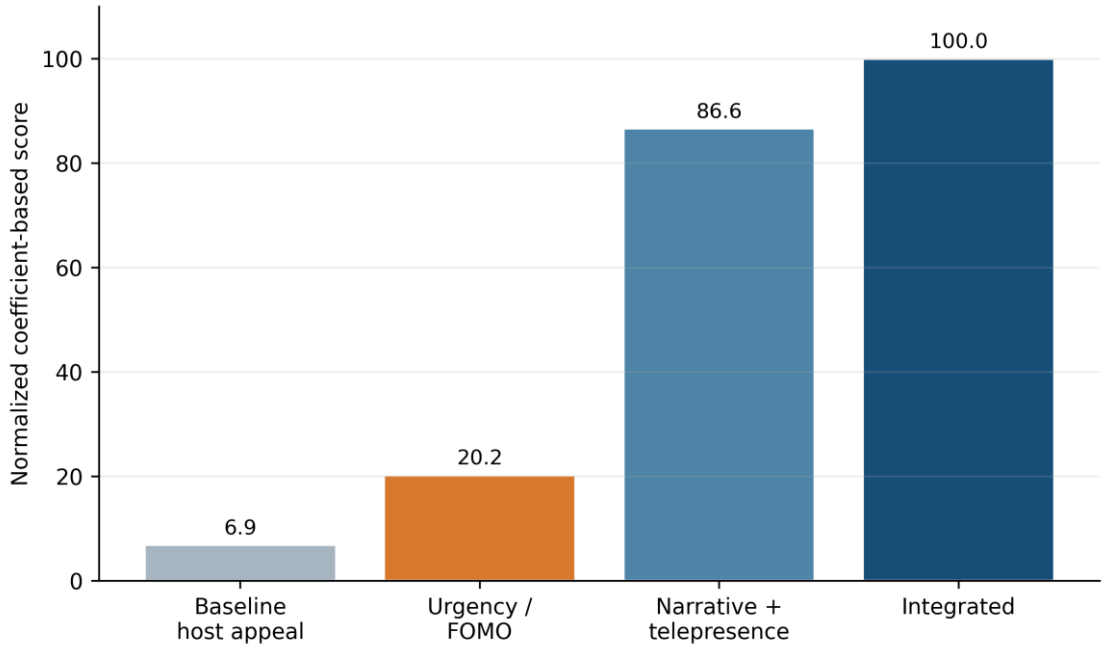

Figure 6. Normalized coefficient-based scenario scores. The values are model-derived comparisons, not observed purchase lift.

F. Ethical-screening performance

Table 6 reports all four ec-darkpattern classifiers under page-identifier-grouped validation. Hybrid word/character TF-IDF with calibrated SVM achieves accuracy .959 ± .011, F1 .959 ± .011, ROC-AUC .989 ± .003, PR-AUC .991 ± .002, Brier score .034 ± .006, and ECE-10 .028 ± .007.

Table 6. ec-darkpattern ethical-screening performance under page-grouped validation.

Model/category	Accuracy/n	F1/recall	ROC-AUC	PR-AUC	Brier	ECE-10
Panel A. Page-grouped 5 × 5 validation						
Word TF-IDF + LR	.953 ± .011	.952 ± .011	.985 ± .004	.988 ± .003	.045 ± .005	.074 ± .008
Character TF-IDF + LR	.953 ± .010	.952 ± .010	.987 ± .004	.989 ± .003	.044 ± .005	.071 ± .007
Hybrid TF-IDF + LR	.958 ± .009	.958 ± .010	.989 ± .004	.991 ± .003	.037 ± .005	.051 ± .006
Hybrid TF-IDF + calibrated SVM	.959 ± .011	.959 ± .011	.989 ± .003	.991 ± .002	.034 ± .006	.028 ± .007
Panel B. Best-model recall by dark-pattern category						
Scarcity	n = 418	Recall .990	—	—	—	—
Social Proof	n = 312	Recall .978	—	—	—	—
Urgency	n = 210	Recall .976	—	—	—	—
Misdirection	n = 195	Recall .826	—	—	—	—
Obstruction	n = 27	Recall .963	—	—	—	—
Sneaking	n = 12	Recall .667	—	—	—	—
Forced Action	n = 4	Recall .500	—	—	—	—

Figure 7 shows the best model's calibration, consensus confusion matrix (1,142 true negatives, 36 false positives, 57 false negatives, and 1,121 true positives), and category recall. Recall is high for Scarcity (.990), Social Proof (.978), and Urgency (.976), but lower for Sneaking and Forced Action, which have only 12 and 4 examples. These rare categories remain human-review priorities.

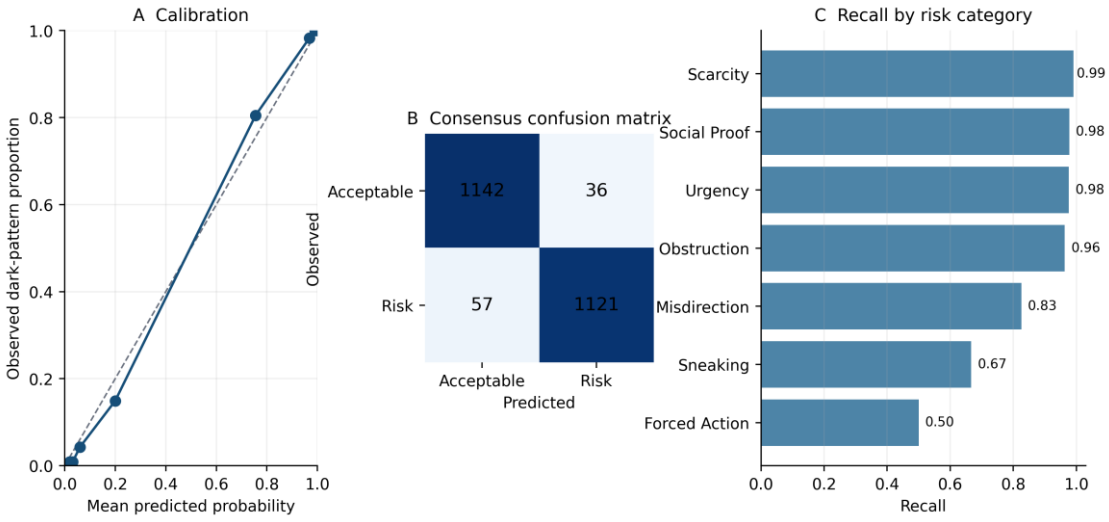

Figure 7. Page-grouped ec-darkpattern validation for the hybrid word/character TF-IDF calibrated SVM: calibration, consensus confusion matrix, and recall by risk category.

B. Discussion

1. Behavioral interpretation

The full and strict analyses support the same descriptive ordering. NI has the largest observed association with both PI and SP, while FOMO is associated with PI and TE with SP. PI and SP are each associated with IP. Static Social Attraction is comparatively weak after the other stimulus constructs enter the model. These patterns are compatible with a dual-pathway account in which

relational connection and co-presence accompany higher impulse-purchase propensity, but the cross-sectional design does not establish that changing a script will change either mediator or purchase behavior.

The measurement diagnostics qualify the magnitude of the paths. Very high reliability estimates coexist with a large first-factor share, high HTMT for NI–TE and PI–SP, and high full-collinearity VIFs. Some shared variance may reflect substantive proximity, but item redundancy and common source can also inflate associations. The lower R^2 values after strict screening reinforce the value of reporting response-quality sensitivity rather than treating the released 415 records as homogeneous.

2. Informatics and decision-support contribution

The main technical contribution is the separation of four analytic tasks. Structural modeling summarizes a theory-consistent association pattern. Classification and continuous prediction quantify out-of-sample discrimination and error under explicit feature availability. Scenario scoring turns coefficients into a transparent ranking rule. The ec-darkpattern classifier screens candidate wording for manipulative risk. Figure 1 then places those outputs behind evidence and human-review gates before prospective testing.

This separation matters operationally. A same-survey classifier using PI and SP can be a diagnostic dashboard but is not automatically a deployment forecast. A model-derived score can prioritize variants but cannot demonstrate incremental response. Uplift and off-policy methods become relevant only when randomized or suitably logged outcomes exist (Mu et al., 2023, 2025; Ye et al., 2026). The next empirical step should therefore record participant assignment, manipulation and attention checks, clicks, add-to-cart events, purchases, and post-exposure mediator measures.

The architecture also clarifies the role of language models. A future generator can retrieve product evidence and propose variants, but its model name, version, prompt template, decoding settings, raw outputs, repeated-generation variability, factuality checks, and human-review outcomes should be archived for evaluation. Until those artifacts and randomized exposure data exist, the present score remains independent of authoring source. This avoids attributing observed effectiveness to a generative model that has not been tested.

3. Governance implications

Evidence and refusal controls are particularly important for price, inventory, coupons, delivery windows, and product claims. Evidence-chain RAG, abstention, numerical checks, and risk cards offer implementation patterns for linking each statement to a trusted source and escalating low-

coverage cases (C. Li et al., 2024; Z. Li et al., 2026; Z. S. Zhong et al., 2025; Su, Rao, & Ma, 2026). Continual red-teaming and self-verification can test how the system responds to unsafe requests and manipulated prompts (D. Zheng et al., 2023, 2024; Zheng & Li, 2024).

The ec-darkpattern results show that a relatively compact lexical model can identify common scarcity, urgency, and social-proof patterns with high recall. Its weaker performance on rare categories and its inability to verify factual claims argue for a layered review process. A practical interface can combine the risk probability, highlighted terms, product evidence, scenario rationale, and an accept/rewrite/escalate decision on one card (B. Zhang et al., 2025; Q. Wu et al., 2025; J. Mu et al., 2026).

Telepresence also invites a link between consumer psychology and engineering telemetry. Time to first frame, rebuffering, bitrate stability, codec choice, and perceptual-quality proxies can be joined to viewer responses in future studies (Wang & Wang, 2023; Wang & Wu, 2024; B. Zhou et al., 2025). Risk-sensitive ABR suggests optimizing stable experience under uncertainty rather than maximum average quality alone (Y. Li, 2023b). These measures would make the Telepresence construct more actionable without treating service metrics as interchangeable with subjective immersion.

4. Limitations and future research

Several limitations define the scope of the findings. First, all seven constructs were collected in one cross-sectional self-report survey, so temporal ordering and causal effects cannot be established. Common-source and construct-overlap diagnostics indicate that some associations may be inflated by the measurement design. Second, the convenience sample is concentrated among younger Indonesian respondents, includes records inconsistent with the stated active-viewer eligibility, and may not generalize to other platforms or markets. The public age bin also prevents independent confirmation that every respondent was an adult.

Third, the binary outcome is a transformed self-report construct rather than a future transaction. The full feature set additionally uses contemporaneous mediators that may not be available at deployment time. Fourth, the four scenarios are deterministic implications of the fitted model and have no randomized viewer outcomes. Fifth, the ec-darkpattern benchmark is English-language interface microcopy; transfer to Indonesian livestream speech, multimodal cues, and rare risk categories requires new annotation and validation.

Future work should preregister complete script manipulations, randomize exposure, document the generation model and prompts when applicable, and measure attention, manipulation checks, clicks, add-to-cart events, purchases, calibration, and heterogeneous treatment effects.

Transaction logs, streaming telemetry, cross-market replication, and privacy-preserving personalization would then support genuine recommendation and incremental-impact claims.

V. CONCLUSION AND RECOMMENDATION

This secondary analysis identifies a stable dual pattern of structural associations in TikTok Live Commerce. NI has the largest association with both PI and SP in the full and quality-screened samples. FOMO and TE also have positive paths, whereas the SA path is not supported by the bootstrap interval. The measurement diagnostics and lower strict-sample R^2 values place appropriate limits on the magnitude of these findings.

Repeated cross-validation shows how classification conclusions change with feature availability, sample screening, and target threshold, while continuous prediction avoids dependence on a single median split. The coefficient-based scenarios make the scoring rubric inspectable, and the ec-darkpattern benchmark adds an independently validated risk-screening layer for common e-commerce microcopy.

The practical contribution is an auditable sequence: screen data quality, estimate and stress-test structural associations, compare prediction and calibration, translate coefficients into transparent scenarios, screen candidate language for ethical risk, verify product claims, obtain human approval, and test promising variants prospectively. This sequence connects consumer-psychology evidence to live-commerce decision support while preserving the distinction between association, prediction, and experimentally measured lift.

Data Availability

The survey workbook is available from Mendeley Data Version 2 at https://doi.org/10.17632/c5mk3gc473.2 and as a direct Excel download at https://data.mendeley.com/public-files/datasets/c5mk3gc473/files/3831c139-2e24-41eb-ba73-283d14cb6650/file_downloaded. The associated Version 4 record is https://doi.org/10.17632/c5mk3gc473.4. The ec-darkpattern TSV is available under the repository's Apache-2.0 license at https://raw.githubusercontent.com/yamanalab/ec-darkpattern/master/dataset/dataset.tsv. Analysis code and aggregate outputs are provided in Supplementary File S1.

Ethics and Consent

This work reanalyzes a public, anonymized dataset and involved no new participant recruitment or contact. The source study states that institutional ethical approval was not required for its anonymous minimal-risk survey and that participants provided electronic consent before

participation (Hadiyati & Luthfia, 2026a). The public age categories do not permit independent verification of individual adult status.

REFERENCES

Bai, J., Wang, H., Wu, Q., & Zhang, B. (2026). Privacy-robust incrementality estimation in cookieless settings via uplift modeling: Reproducible evidence from the Hillstrom E-Mail Experiment. *Journal of Technology Informatics and Engineering, 5*(1), 17–38. https://doi.org/10.51903/jtie.v5i1.468

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments, 12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Chang, X., Lu, Y., & Zhong, Z. S. (2026). Review-grounded explainable recommendation with faithfulness evaluation on Amazon Reviews. *Journal of Electrical Engineering and Computer Science, 11*(1), 9–22. https://doi.org/10.54732/jeecs.v11i1.2

Chang, X., Ye, T., & Luo, S. (2023). LLM-as-reranker for personalized recommendation: Popularity bias mitigation and faithful natural-language explanations on MovieLens 100K. *Journal of Advanced Computing Systems, 3*(8), 61–78. https://doi.org/10.69987/JACS.2023.30806

Chen, B., Wang, L., Rasool, H., & Wang, J. (2022). Research on the impact of marketing strategy on consumers' impulsive purchase behavior in livestreaming e-commerce. *Frontiers in Psychology, 13*, Article 905531. https://doi.org/10.3389/fpsyg.2022.905531

Chen, L.-R., Chen, F.-S., & Chen, D.-F. (2023). Effect of social presence toward livestream e-commerce on consumers' purchase intention. *Sustainability, 15*(4), Article 3571. https://doi.org/10.3390/su15043571

Chen, S., Su, W., & Ma, J. (2024). Self-correcting text-to-SQL agent with error feedback: A reproducible closed-loop evaluation on compact executable SQLite benchmarks. *Journal of Advanced Computing Systems, 4*(11), 86–104. https://doi.org/10.69987/JACS.2024.41107

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chen, Y., Zhang, Y., Chau, D., & Sherman, M. (2023). Credit card default risk tiering with probability calibration and uncertainty-driven rejection: A reproducible study on the UCI Credit Card Clients Dataset. *Journal of Advanced Computing Systems, 3*(4), 31–47. https://doi.org/10.69987/JACS.2023.30403

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Hadiyati, I. A., & Luthfia, A. (2026a). From stream to purchase: Exploring parasocial interaction and social presence as mediators of impulsive purchase in TikTok live commerce. *Frontiers in Communication, 11*, Article 1786830. https://doi.org/10.3389/fcomm.2026.1786830

Hadiyati, I. A., & Luthfia, A. (2026b). *From stream to purchase: Exploring parasocial interaction and social presence as mediators of impulsive purchase in TikTok Live Commerce* (Version 4) [Data set]. Mendeley Data. https://doi.org/10.17632/c5mk3gc473.4

Hadiyati, I. A., & Luthfia, A. (2026c). *From stream to purchase: Exploring parasocial interaction and social presence as mediators of impulsive purchase in TikTok Live Commerce* (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/c5mk3gc473.2

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

Hair, J., & Alamer, A. (2022). Partial least squares structural equation modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics, 1*(3), Article 100027. https://doi.org/10.1016/j.rmal.2022.100027

Hartmann, T., & Goldhoorn, C. (2011). Horton and Wohl revisited: Exploring viewers' experience of parasocial interaction. *Journal of Communication, 61*(6), 1104–1121. https://doi.org/10.1111/j.1460-2466.2011.01595.x

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*, 115–135. https://doi.org/10.1007/s11747-014-0403-8

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry, 19*(3), 215–229. https://doi.org/10.1080/00332747.1956.11023049

Jin, J. (2025). LLM-style evidence cards for scientific search interfaces: A UI/UX design framework for retrieval transparency, ranking trust, and visual evidence hierarchy. *International Journal of Graphic Design, 3*(2), 397–414. https://doi.org/10.51903/ijgd.v3i2.3698

Jin, J., Huang, T., & Lu, S. (2024a). Cost-sensitive learning, simulated PU learning, and one-class autoencoding for extreme-imbalance credit card fraud detection. *Journal of Advanced Computing Systems, 4*(6), 64–73. https://doi.org/10.69987/JACS.2024.40605

Jin, J., Huang, T., & Lu, S. (2024b). A model-risk-friendly probability of default workflow: Calibration, distribution-free uncertainty quantification, and SHAP explanations on the UCI Credit Card Default Dataset. *Journal of Advanced Computing Systems, 4*(6), 74–85. https://doi.org/10.69987/JACS.2024.40606

Kim, T., & Biocca, F. (1997). Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication, 3*(2). https://doi.org/10.1111/j.1083-6101.1997.tb00073.x

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101

Kuo, M.-J., Zheng, D., & Hires, J. (2025). Federated topic-preference learning for knowledge-grounded chat with differential privacy. *Journal of Technology Informatics and Engineering, 4*(2), 385–401. https://doi.org/10.51903/jtie.v4i2.502

Li, C., Bai, J., & Wang, S. (2024). Evidence-chain reliable RAG: Word-level hallucination detection, source attribution, and provenance explanation for LLM applications. *Journal of Advanced Computing Systems, 4*(2), 76–92. https://doi.org/10.69987/JACS.2024.40207

Li, C., Su, W., & Zhang, E. (2023). Lightweight hallucination firewall for enterprise LLM applications: Evidence consistency, self-checking, and small-model detection on TruthfulQA. *Journal of Advanced Computing Systems, 3*(1), 49–65. https://doi.org/10.69987/JACS.2023.30104

Li, G., Jiang, Y., & Chang, L. (2022). The influence mechanism of interaction quality in live streaming shopping on consumers' impulsive purchase intention. *Frontiers in Psychology, 13*, Article 918196. https://doi.org/10.3389/fpsyg.2022.918196

Li, Y. (2023a). Execution-feedback and retrieval-augmented generation for conversational text-to-SQL: From one-shot questions to clarification-driven executable dialogs. *Journal of Advanced Computing Systems, 3*(2), 1–17. https://doi.org/10.69987/JACS.2023.30201

Li, Y. (2023b). Risk-sensitive offline reinforcement learning for stable ABR QoE improvements on real HSDPA and LTE traces. *Journal of Advanced Computing Systems, 3*(4), 1–11. https://doi.org/10.69987/JACS.2023.30401

Li, Y., Lu, S., & Zhao, L. (2025). LLM-as-design-critic: Aligning AI-generated UI feedback with human graphic design judgment. *International Journal of Graphic Design, 3*(1), 196–215. https://doi.org/10.51903/ijgd.v3i1.3661

Li, Z., Zhang, K., & Wong, A. (2026). Numerical-reasoning guardrails for a quant research assistant: A compact reproducible benchmark using SEC and FRED data. *Journal of Technology Informatics and Engineering, 5*(2), 75–90. https://doi.org/10.51903/jtie.v5i2.541

Lu, S., & Zhou, D. (2023). LLM-augmented customer representation learning for next-purchase prediction in online retail. *Journal of Advanced Computing Systems, 3*(3), 50–64. https://doi.org/10.69987/JACS.2023.30305

Lu, Y., Mu, J., & Tran, T. (2024). Uncertainty-aware uplift modeling for safer marketing targeting: Conformal prediction and Bayesian calibration with LCB policies. *Journal of Advanced Computing Systems, 4*(5), 84–101. https://doi.org/10.69987/JACS.2024.40507

McCroskey, J. C., & McCain, T. A. (1974). The measurement of interpersonal attraction. *Speech Monographs, 41*(3), 261–266. https://doi.org/10.1080/03637757409375845

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. MIT Press.

Mu, J., Lu, Y., & Hwang, E. (2026). Structured visual brief interfaces for advertising design: A UI/UX framework for turning creative intentions into designer-editable graphic design cards. *International Journal of Graphic Design, 4*(1), 192–208. https://doi.org/10.51903/ijgd.v4i1.3702

Mu, J., Lu, Y., & Smith, M. (2023). LLM-assisted incrementality (uplift) modeling for email advertising: From feature interactions to interpretable audience-creative-channel policies.

Journal of Advanced Computing Systems, 3(1), 31–48. https://doi.org/10.69987/JACS.2023.30103

Mu, J., Ye, T., & Patel, P. (2025). Offline counterfactual evaluation for advertising and recommendation slot policies: A reproducible study on the Open Bandit Dataset (Small). *Journal of Technology Informatics and Engineering, 4*(3), 521–543. https://doi.org/10.51903/jtie.v4i3.500

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

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior, 29*(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Rook, D. W. (1987). The buying impulse. *Journal of Consumer Research, 14*(2), 189–199. https://doi.org/10.1086/209105

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods, 25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

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

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. John Wiley & Sons.

Su, W., Chen, S., & Qian, E. (2026). Narrative-aware scientific claim verification agent with evidence ranking for ClimateCheck. *Journal of Technology Informatics and Engineering, 5*(1), 327–340. https://doi.org/10.51903/jtie.v5i1.549

Su, W., Chen, S., & Zhao, C. (2025). Budgeted multi-hop retrieval agent for compositional question answering: A retrieval-policy evaluation on the official MultiHop-RAG benchmark. *Journal of Technology Informatics and Engineering, 4*(3), 649–662. https://doi.org/10.51903/jtie.v4i3.543

Su, W., Rao, H., & Ma, E. (2026). Privacy and data-integrity risk cards for LLM agents: A UI/UX design framework for secure human oversight under prompt-injection attacks. *International Journal of Graphic Design, 4*(1), 186–191. https://doi.org/10.51903/ijgd.v4i1.3699

Tu, H., Zhao, S., & Zhou, A. (2025). Visual brief cards for advertising design: A structured UI/UX framework for turning creative intentions into graphic design decisions. *International Journal of Graphic Design, 3*(1), 210–226. https://doi.org/10.51903/ijgd.v3i1.3714

Vazquez, D., Wu, X., Nguyen, B., Kent, A., Gutierrez, A., & Chen, T. (2020). Investigating narrative involvement, parasocial interactions, and impulse buying behaviours within a

second-screen social commerce context. *International Journal of Information Management, 53*, Article 102135. https://doi.org/10.1016/j.ijinfomgt.2020.102135

Wang, E., & Wang, H. (2023). QoE-driven reinforcement learning for joint bitrate, rebuffering, and TTFF optimization in HLS/DASH. *Journal of Advanced Computing Systems, 3*(2), 50–59. https://doi.org/10.69987/JACS.2023.30204

Wang, H., & Wu, P. (2024). Content-adaptive codec selection between AV1 and H.264: A controlled case study of bandwidth savings versus latency at matched quality. *Journal of Advanced Computing Systems, 4*(7), 83–92. https://doi.org/10.69987/JACS.2024.40707

Wu, Q., Meng, S., & Zhao, J. (2025). Text-grounded LLM-assisted design rationale interfaces: Turning advertising layout metadata into explainable UI/UX decision cards. *International Journal of Graphic Design, 3*(1), 216–240. https://doi.org/10.51903/ijgd.v3i1.3713

Xin, Q. (2026). Self-supervised customer representation learning for segmentation and next-purchase prediction on UCI Online Retail. *Journal of Information and Technology, 14*(1). https://doi.org/10.32664/j-intech.v14i01.2229

Xu, H., Chen, Y., & Med, A. (2025). Automatic detection and explanation of dark patterns from interface microcopy: Empirical comparison of BERT-style encoders, RoBERTa-style encoders, and LLM-style decoders on the ec-darkpattern Dataset. *Journal of Technology Informatics and Engineering, 4*(3), 590–612. https://doi.org/10.51903/jtie.v4i3.491

Xu, K., Zhou, H., Zheng, H., Zhu, M., & Xin, Q. (2024). Intelligent classification and personalized recommendation of e-commerce products based on machine learning. In *Proceedings of the 6th International Conference on Computing and Data Science (ICCDS)*.

Yada, Y., Feng, J., Matsumoto, T., Fukushima, N., Kido, F., & Yamana, H. (2022). Dark patterns in e-commerce: A dataset and its baseline evaluations. In *2022 IEEE International Conference on Big Data (Big Data)* (pp. 3015–3022). https://doi.org/10.1109/BigData55660.2022.10020800

Ye, T., Mu, J., & Hunter, J. (2026). Off-policy evaluation and conservative policy selection for slot-level dynamic bidding and ranking on the Open Bandit Dataset (Small). *Journal of Technology Informatics and Engineering, 5*(1), 178–199. https://doi.org/10.51903/jtie.v5i1.503

Zahid, M. N., Kamran, M., Szostak, M., & Awan, T. M. (2024). Telepresence, social presence and involvement in consumer's intention to buy apparels through an interplay of consumer brand engagement. *Foresight, 26*(6), 984–999. https://doi.org/10.1108/FS-10-2023-0197

Zhang, B., Ren, Y., & Zou, J. (2025). LLM-style explainable e-commerce recommendation cards: A UI/UX design framework for trust-calibrated product recommendation. *International Journal of Graphic Design, 3*(2), 381–396. https://doi.org/10.51903/ijgd.v3i2.3697

Zhao, S., Zhou, H., & Martinez, D. (2023). LLM-assisted causal attribution of service performance upgrades on churn and tenure: Full evaluation on the IBM Telco Customer Churn Dataset. *Journal of Advanced Computing Systems, 3*(2), 18–34. https://doi.org/10.69987/JACS.2023.30202

Zheng, D., & Li, C. (2024). Behavior-level jailbreak resistance via multi-stage refusal + utility preservation. *Journal of Advanced Computing Systems, 4*(1), 83–99. https://doi.org/10.69987/JACS.2024.40107

Zheng, D., Li, C., & Davidson, H. (2023). Continual red-teaming for in-the-wild jailbreaks via online guardrail updates and guardrail distillation. *Journal of Advanced Computing Systems, 3*(2), 35–49. https://doi.org/10.69987/JACS.2023.30203

Zheng, D., Zhang, B., & Geibel, J. (2024). VerifySafe: Toxicity-safe agent responses under adversarial prompts with evidence-based self-verification. *Journal of Advanced Computing Systems, 4*(1), 67–82. https://doi.org/10.69987/JACS.2024.40106

Zhong, Z. S., Chen, J., Zhong, E., & Sun, X. (2025). Evidence-calibrated RAG for unanswerable question answering: Retrieval coverage, abstention calibration, and hallucination-proxy analysis on SQuAD 2.0. *Journal of Technology Informatics and Engineering, 4*(2), 502–520. https://doi.org/10.51903/jtie.v4i2.536

Zhong, Z. S., Li, C., & Rao, H. (2026). Trajectory reliability prediction for generalist AI agents: Tool-use failure analysis and success forecasting on ZClawBench. *Journal of Technology Informatics and Engineering, 5*(1), 341–360. https://doi.org/10.51903/jtie.v5i1.539

Zhou, B., Wang, H., & Chang, X. (2025). Distilling VMAF into an edge-deployable quality predictor: A pilot shot-level proxy with LLM-ready quality tokens. *Journal of Technology Informatics and Engineering, 4*(2), 447–463. https://doi.org/10.51903/jtie.v4i2.522

Zhou, H., & Zhao, S. (2024). LLM-explanation-enhanced retail credit default prediction with gradient boosting on the UCI Default of Credit Card Clients Dataset. *Journal of Advanced Computing Systems, 4*(5), 102–118. https://doi.org/10.69987/JACS.2024.40508

APPENDIX

Appendix A. Measurement items and construct sources

The retained wording below follows the public questionnaire reported by Hadiyati and Luthfia (2026a, 2026c). All items used a five-point agreement scale. FOMO2c and IP1b are absent from the published 127-item instrument and were not scored.

Social Attraction (SA)

SA1	: I feel that the host has a personality similar to mine.
SA2	: I feel compatible and comfortable when watching the content shared by the host.
SA2a	: I feel comfortable with the host's way of speaking.
SA3	: The host feels like a friend with whom I can share thoughts and feelings.
SA3a	: The host has an appearance that I find appealing.
SA4	: I want to keep following or interacting with the host because I feel socially connected.
SA4a	: I feel that the host behaves in a friendly manner during live streaming.
SA5	: I find the host easy to interact with.
SA5a	: I feel that the host is open to receiving questions from viewers.
SA6	: I feel that the host cares about their audience.
SA6a	: I feel that the host responds sincerely to viewers' comments.
SA7	: I feel that the host appears genuine during the live session.
SA7a	: I feel that the host does not pretend or act fake in their content.
SA8	: The host's experiences are similar to mine.
SA8a	: The host's stories are relevant to my life.
SA9	: The host is relatable, like an ordinary person.
SA9a	: The host does not create distance from the audience.

Fear of Missing Out (FOMO)

FOMO1	: I am afraid that I will regret not buying the product being promoted.
FOMO1a	: I feel regretful when I miss a live promotion opportunity.
FOMO2	: I worry about missing out on the product because the stock is limited.
FOMO2a	: I fear that the product will sell out quickly if I do not buy it.
FOMO2b	: I feel pressured to make an immediate purchase due to the time-limited offer.
FOMO3	: I worry that others are getting something more enjoyable than I am.
FOMO3a	: I feel envious when others get to enjoy the product before I do.
FOMO4	: I am concerned that others are having more fun with the promoted product while I am not.
FOMO4a	: I feel left out of the social excitement if I do not join the purchase.
FOMO4b	: I compare myself with other viewers who are enjoying the product.
FOMO4c	: I feel left behind when others buy the product before I do.
FOMO5	: I feel out of trend if I do not have the product promoted by the host.
FOMO5a	: I feel the need to buy in order not to miss out on the trend.
FOMO6	: I regret not trying the product promoted by the host.
FOMO6a	: I feel a sense of loss if I do not try the product shown in the live session.
FOMO6b	: I feel the need to buy to avoid being excluded from the community.
FOMO6c	: I worry that I will be seen as outdated if I do not make a purchase.
FOMO7	: I feel anxious because I do not have the product promoted by the host.
FOMO8	: I feel bothered for missing the chance to try the product promoted by the host.

Narrative Involvement (NI)

NI1	: I am truly absorbed in the storyline presented by the live host.
NI1a	: While watching, I feel completely immersed in the host's story.
NI2	: While watching, I find myself drawn into the host's narrative world.
NI2a	: I feel as if I have entered the story being presented.
NI2b	: I experience an emotional connection when the host delivers the story.
NI2c	: I feel emotionally close to the product narrative.
NI3	: I become curious about how the story will end after watching the live session.
NI3a	: I feel intrigued to know what happens next in the host's story.
NI3b	: I feel motivated to watch until the live session ends.
NI3c	: I find it difficult to stop watching before the story finishes.
NI3d	: I feel curious about how the product will be featured in the live story.
NI3e	: I want to know more product details after hearing the host's narrative.
NI4	: I pay full attention to the story told by the host.
NI4a	: I focus on listening to the details of the narrative.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 474-502
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

NI4c : I feel that time passes quickly while watching the live session.
NI4d : My thoughts become absorbed in the story details.
NI4e : My mind feels fully concentrated on the host's narrative content.

Telepresence (TE)

TE1 : While watching the live-streaming session, I feel as if the experience is similar to shopping in a physical store.
TE1a : I feel as though I am physically present in the live-streaming room.
TE1b : The products displayed appear realistic in front of me.
TE1c : It feels as if I can touch the products directly, even though I only see them on screen.
TE1d : The live-streaming session creates an atmosphere similar to being in a real store.
TE1e : During the live session, it feels like an authentic in-store shopping experience.
TE2 : I feel deeply immersed in the atmosphere of the live broadcast.
TE2a : I am carried away as if I were physically present at the event location.
TE2b : I focus completely on observing the host's interactions.
TE2c : I ignore other things around me because I am focused on the host.
TE2d : When watching the host's live-streaming session, I feel as if I am in a "virtual world" rather than the "real world" around me.
TE2e : I feel that the virtual environment dominates my attention.
TE3 : While watching the live stream, my mind becomes fully immersed in the host's world and the products being presented.
TE3a : The host's storytelling captures all my attention.
TE3b : My attention is fully directed toward the live content.
TE3c : My thoughts are entirely focused on the live session.
TE3d : I lose track of time while watching the live session.
TE3e : Time seems to pass quickly while I am watching.

Parasocial Interaction (PI)

PI1 : When watching the TikTok host's live session, the host feels like an old friend of mine.
PI1a : I feel that the host is as close to me as a personal friend.
PI1b : I feel that the host provides emotional support even though it is only through the screen.
PI1c : The host is able to uplift my mood while I am watching.
PI1d : I feel comfortable when watching the host's live sessions.
PI1e : I feel calmer knowing that the host is accompanying me.
PI2 : While watching the live session, the host's presence makes me feel like I am part of the event.
PI2a : I feel that the host makes me feel less alone.
PI2b : I feel accompanied even when watching alone at home.
PI2c : I feel that the host is genuinely present to accompany the audience.
PI2d : The host talks directly to me.
PI2e : The host responds to me personally.
PI3 : While watching the live-streaming session, I feel as though I am part of it.
PI3a : I feel that I play a role in the live event.
PI3b : I feel accepted by other viewers.
PI3c : I feel warmly welcomed by the audience.
PI3d : I feel a shared identity with other viewers.
PI3e : I feel that I belong to the same community as other viewers.
PI4 : I look forward to watching the host's next live-streaming session.
PI5 : I would like to meet the streamer in person someday.

Social Presence (SP)

SP1 : While watching the host's live-streaming session, other users can sense my presence when I ask questions.
SP1a : I am aware of the presence of other viewers during the live session.
SP1b : The presence of other viewers feels real.
SP1c : I feel as if I am in the same room as the audience.
SP1d : I feel noticed when my comments are acknowledged by other viewers.
SP1e : I feel that my presence is recognized by the audience.
SP2 : When watching the host's live-streaming session, my emotions are influenced by other viewers.
SP2a : I feel happy when other viewers are excited.
SP2b : I share the same joy when the audience shows enthusiasm.
SP2c : I experience the excitement together with other viewers.
SP2d : I feel anxious when other viewers react quickly to make a purchase.
SP2e : I feel encouraged to buy after seeing other viewers' reactions.

SP3 : While watching the host's live-streaming session, I feel close to other viewers.
SP3a : I feel a friendly bond with other viewers.
SP3b : I feel familiar with other viewers even though we have never met in person.
SP3c : I feel that other viewers could become my new friends.
SP3d : I feel that I can build new social relationships through the live session.
SP3e : I feel like part of a social network with other viewers.
SP4 : While watching the host's live-streaming session, the distance between me and other viewers feels smaller.

Impulsive Purchase (IP)

IP1 : While watching live-streaming sessions, I buy items that I did not originally intend to purchase.
IP1a : I purchase products even though they are not on my shopping list.
IP1c : I buy products without much prior thought.
IP1d : I purchase products even when they are not part of my budget.
IP1e : I buy products without considering my financial situation.
IP2 : I often make unplanned purchases.
IP2a : I buy products without thinking for too long.
IP2b : I decide to purchase products quickly without much deliberation.
IP2c : I buy products on impulse.
IP2d : I make purchases because I get carried away by the live-streaming atmosphere.
IP2e : I purchase products due to sudden promotional offers.
IP3 : Making spontaneous purchases feels enjoyable.
IP3a : I feel happy after making an impulsive purchase.
IP3b : I feel satisfied with my spontaneous buying decisions.
IP3c : I find that sudden shopping makes me feel happy.
IP3d : I consider spontaneous shopping as a form of entertainment.
IP3e : Impulsive buying gives me a pleasant experience.