International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

Beyond Algorithmic Exposure: How Perceived Algorithmic Bias Shapes Trust and Engagement on Social Media

Iwan Armawan[1]
Department of Communication Science, Universitas Mpu Tantular, Jakarta, Indonesia
Tri Adi Sarwoko[2]
Department of Communication Science, Universitas Mpu Tantular, Jakarta, Indonesia
Silva Aresta Wulandari[3]
Department of Communication Science, Universitas Mpu Tantular, Jakarta, Indonesia
Wahyu Sulaiman[4]
Department of Communication Science, Universitas Mpu Tantular, Jakarta, Indonesia

Corresponding Author : iwanaradea84@gmail.com

Abstract

This methodological simulation examines how perceived algorithmic bias could shape social media engagement directly and indirectly through platform trust. Existing research establishes that algorithm awareness, equity judgments and explanations may influence users' evaluations, but estimated pathways can vary by context. To demonstrate a reproducible analytical workflow without suggesting new empirical evidence, we generated a fully synthetic dataset of 600 hypothetical users from a documented Gaussian structural process. Perceived algorithmic bias, algorithm awareness, perceived control, platform trust and engagement were standardized and analyzed by ordinary least squares path models, percentile bootstrap mediation (1,200 resamples), and scenario sensitivity checks. In this intentionally specified simulation, the bias–trust path was -0.406, the trust–engagement path was 0.589, and the product indirect effect was -0.239 (illustrative 95% bootstrap interval -0.289 to -0.193); the direct bias–engagement path was -0.220. These values emerge from chosen data-generating assumptions and do not establish behavioral effects in actual populations. The demonstration illustrates how researchers may distinguish exposure metrics from perceived procedural inequity, test mediation transparently, and conduct uncertainty checks before collecting empirical data. A preregistered multi-platform field study is required to assess construct validity, external validity, and causal alternatives.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Keywords: perceived algorithmic bias; procedural fairness; platform trust; social media engagement; algorithm awareness; synthetic data; mediation; simulation.

INTRODUCTION

Algorithmically ranked feeds have become major infrastructures for discovering news, entertainment, commercial messages, and interpersonal expression. Exposure is no longer a neutral sequence of posts: ranking, moderation, prediction and personalization influence which voices receive attention, which communities experience visibility, and how users interpret platform priorities. A crucial distinction follows. Algorithmic exposure refers to the content a system displays, whereas perceived algorithmic bias concerns a user's appraisal that this distribution is systematically unequal, manipulative or insensitive to legitimate interests. Objective differences in ranking and subjective judgments about fairness need not coincide (Huang & Liu, 2025; Lee & Kwon, 2025; Zarouali et al., 2021).

Public debate often treats the performance of recommendation systems as a matter of relevance or accuracy. Yet a system may accurately predict clicks and still prompt distrust if people perceive that it privileges specific viewpoints, commercial interests, creators or social groups. Scholarship on fairness-aware recommendation explicitly warns that trade-offs between accuracy, exposure diversity and equity are context dependent (Jin et al., 2023). Research on social media engagement restrictions shows that the available forms of expression themselves influence procedural fairness perceptions (Lee & Kwon, 2025). Such evidence motivates measuring perceived bias independently from both frequency of exposure and generic attitude toward technology (Geber & Stahel, 2026; Ozanne et al., 2022).

Algorithm awareness provides an important interpretive frame. The Algorithmic Media Content Awareness scale identifies distinct forms of user knowledge concerning filtering, automated decision-making, human–algorithm interplay and ethical considerations (Zarouali et al., 2021). A large survey of recommendation acceptance later associated awareness, usefulness and trust with behavioral intentions, while cautioning that awareness may also foster skepticism (Huang & Liu, 2025). These results do not imply that awareness is identical to perceived unfairness; rather, awareness may alter how users attribute platform experiences to the algorithm (Chen et al., 2024; Dong et al., 2022; Jin et al., 2023).

Qualitative studies document that reactions unfold over time. New TikTok users describe training their feeds, recognizing repeated patterns and occasionally rejecting the recommendations they receive (Siles et al., 2024). Ethnographic work further suggests that trust

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

in the feed is experienced as an ongoing human–machine relationship rather than a stable technical judgment (Schellewald, 2025). Accordingly, a pathway from perceived bias through trust to engagement is theoretically plausible, although it cannot be assumed to be causal without longitudinal or experimental identification (Stefanopoulos et al., 2025; Zarouali et al., 2021).

Existing studies examine equity, social trust, moderation accountability, explanatory transparency, algorithmic aversion, and connectedness, but they do not all measure the same psychological mechanism (Ozanne et al., 2022), (Dong et al., 2022), (Chen et al., 2024), (Obreja, 2026), (Shi et al., 2026). Recent short-video research directly tests whether provenance cues and procedural disclosures influence perceived fairness (Liu & Fan, 2026). This suggests a focused research opportunity: isolate a perception-centered mechanism in which an individual's bias appraisal predicts an evaluation of platform trust and then engagement, while accounting for algorithm awareness and perceived control (Obreja, 2026; Siles et al., 2024).

The present paper is a methodological and conceptual demonstration, not a report of new survey evidence. Its contribution is a transparent synthetic example of the proposed mechanism, with fixed data-generating assumptions, explicit disclosure of numerical provenance, regression-based mediation, and sensitivity analysis. Such an exercise can support instrument planning, code verification and power-related reasoning, but cannot substantiate hypotheses about real social media users. The paper asks: What patterns would a specified bias–trust–engagement mechanism generate; how would direct and indirect paths be quantified; and which assumptions must be challenged in future empirical work?

Four objectives organize the study: (i) articulate constructs and an assumed path structure; (ii) generate a reproducible synthetic dataset of hypothetical platform users; (iii) estimate direct and indirect coefficients while checking numerical consistency; and (iv) translate the simulation into a realistic protocol for later validation. We deliberately avoid presenting the simulated coefficients as newly discovered psychological regularities or making unsupported claims of generalizability.

Perceived algorithmic bias and procedural fairness

Perceived algorithmic bias (PAB) is defined here as the subjective expectation that the platform's ranking or moderation systematically advantages some content or actors without sufficient justification. It combines perceptions of distributive imbalance, inconsistent procedures and opacity. In contrast, an algorithmic fairness audit may estimate measurable exposure distributions or performance disparities. These are related but distinct objects of measurement.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

A highly personalized feed could be perceived as fair by one user yet exclusionary by another; thus perceived bias is not evidence of technical discrimination in itself (Liu & Fan, 2026).

Fairness judgments can arise even when engagement options change rather than recommendation code. Experimental research on perceived fairness in social media engagement indicates that restricting opportunities for negative feedback can reduce user fairness evaluations (Lee & Kwon, 2025). Similarly, public disclosures and user-choice affordances in short-video feeds may serve as procedural cues (Liu & Fan, 2026). We therefore specify, strictly as a simulation premise, a negative PAB-to-trust relation and a negative PAB-to-engagement relation (Liu & Fan, 2026).

Trust as a mediating interpretation

Trust refers to the willingness to rely on platform decisions under uncertainty about recommendation logic, organizational incentives and downstream consequences (Shi et al., 2026). Trust in the organization, trust in AI-supported moderation and trust in a particular feed recommendation are analytically separable (Huo et al., 2026). Transparency may supply useful evidence but can sometimes increase perceived instrumental treatment or aversion (Chen et al., 2024). The mediation proposed here assumes that perceived bias first alters users' trust evaluation, which then contributes to engagement decisions (Schellewald, 2025).

Engagement is operationalized as self-reported intention to interact, revisit and voluntarily participate, not as observed clicks or time spent. That distinction is crucial because high dwell time may reflect conflict, curiosity or habitual use instead of satisfaction. Studies questioning engagement-based ranking suggest that revealed behavioral preference and stated user preference can diverge (Obreja, 2024). The outcome in this demonstration is therefore labeled intended engagement to prevent confusion with platform telemetry.

Awareness and agency as covariates

Algorithm awareness represents conscious recognition that feeds are curated by computational systems; perceived control captures the extent to which users believe they can adjust recommendations and correct unwanted outputs. Both constructs plausibly correlate with trust and engagement (Zarouali et al., 2021), (Huang & Liu, 2025), (Siles et al., 2024), (Obreja, 2026). They are introduced in the synthetic generator as covariates, without assigning them causal status. A future empirical model should also measure platform, content genre, demographic characteristics, exposure diversity and baseline use intensity (Liao et al., 2026) .

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Three illustrative directional propositions follow: P1—greater PAB corresponds to lower trust; P2—greater trust corresponds to higher engagement while accounting for PAB; P3—the PAB–engagement association contains a negative indirect component through trust. These propositions guide the simulation design only. Hypothesis testing on actual users requires independently collected, appropriate evidence (Shi et al., 2026).

METHOD

Design and disclosure of synthetic data

The article uses a reproducible computational demonstration with N = 600 artificially generated profiles. The records are not anonymized respondents, scraped accounts, recruited participants or platform users. No questionnaire was administered, no consent was collected, and no behavioral traces were accessed. The fixed seed 20261022 defines the random draws, and all reported parameters are properties of this artificial system. Hence conventional significance tests have no population-level inferential meaning in this manuscript.

Five standardized variables were generated: perceived algorithmic bias (B), algorithm awareness (A), perceived control (C), trust (T) and intended engagement (E). Inputs originated from pseudo-random normal draws. Awareness was weakly positively correlated with bias, and perceived control combined bias and awareness with independent noise. Trust was generated from negative bias and positive awareness/control terms. Engagement was then generated from negative bias, positive trust and a small positive awareness term. Independent Gaussian disturbances were included at both structural stages.

Structural assumptions and computational model

The illustrative latent equations before final standardization were specified as: $A^* = 0.28B + \text{sqrt}(1 - 0.28^2)\ \varepsilon_A$; $C^* = -0.23B + 0.18A + 0.95\ \varepsilon_C$; $T^* = -0.48B + 0.25A + 0.20C + 0.77\ \varepsilon_T$; $E^* = -0.20B + 0.56T + 0.13A + 0.72\ \varepsilon_E$. Every ε is an independently sampled standard normal variate. The asterisked variables were standardized to sample mean zero and standard deviation one. These are design choices, not estimates from prior surveys.

The fitted mediation regressions were $T = \alpha_o + aB + \alpha_AA + \alpha_CC + u$, and $E = \beta_o + c'B + bT + \beta_AA + \beta_CC + v$. The indirect coefficient was computed as $a \times b$ and the total as $c' + a \times b$. These expressions are association decompositions in the constructed linear system; causal mediation is not established by this arithmetic alone.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Operational measures for later empirical validation

Table 1. Proposed operationalization (not an administered instrument)

Construct	Illustrative indicators	Measurement caution
PAB	Unjustified favoring; inconsistent reach; biased curation	Distinguish perceived bias from audited disparity
TRU	Reliance on feed; confidence in platform integrity	Separate trust in platform vs. recommender
ENG	Comment/share intention; revisit willingness	Do not equate intentions with observed use
AWA	Filtering awareness; understanding automated decisions	Use validated AMCA construct as anchor
CTL	Ability to hide, correct, or customize feed	Test factor structure and discriminant validity

For a subsequent observational study, use multiple Likert-type items per latent construct, cognitive interviews, translation/back-translation as needed, and confirmatory factor analysis. Reliability (omega/composite reliability), convergent validity, discriminant validity, measurement invariance across platforms, and common-method bias diagnostics should be prespecified. The present synthetic variables are error-contaminated continuous composites created directly; measurement-model reliability cannot legitimately be claimed from them.

Analysis and sensitivity procedure

Ordinary least squares models were estimated from the generated matrix, and 1,200 nonparametric resamples of synthetic rows were used to obtain a percentile interval for the product a × b. Because the generator enforces structural relationships, the resulting narrow interval mainly reflects internal Monte Carlo variability; it does not constitute evidence that the mediated association exists in the public. The fitted trust and engagement models include the same bias, awareness and control terms except that trust enters the engagement model.

Two sensitivity exercises were used. First, a direct PAB-to-engagement coefficient was varied over plausible illustrative values, holding the estimated a × b term fixed, to demonstrate how the apparent total association shifts under different assumptions. Second, researchers are advised to introduce an omitted confounder linked to trust and engagement in future simulation

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

runs and report how much such a factor would alter the mediation product. The current exercise is intentionally not a full causal identification sensitivity study.

RESULTS AND DISCUSSION

Synthetic sample and descriptive statistics

Exactly 600 hypothetical observations were generated without missing values. Each modeled variable has mean approximately zero and sample standard deviation one by construction. Accordingly, describing the dataset as having a particular gender, nationality, platform preference, or age distribution would be fictional and is explicitly avoided. The exercise is solely numerical and does not represent a sampling frame.

Table 2. Correlations in synthetic data (N = 600)

Variable	B	A	C	T	E
Bias (B)	1.00	0.31	-0.14	-0.35	-0.39
Awareness (A)	0.31	1.00	0.16	0.20	0.17
Control (C)	-0.14	0.16	1.00	0.37	0.25
Trust (T)	-0.35	0.20	0.37	1.00	0.68
Engagement (E)	-0.39	0.17	0.25	0.68	1.00

The matrix shows the intended pattern of negative correlation between bias and trust, and a positive relation between trust and engagement. Correlations are consequences of programmed coefficients and simulated errors. They should not be compared to values from real-world datasets as independent replication; instead, their purpose is to verify that the generator and estimator behave coherently.

Illustrative regression and mediation decomposition

Table 3. OLS associations estimated from synthetic profiles

Predictor	Trust β (SE)	Engagement β (SE)
PAB	-0.406 (0.037)	-0.220 (0.034)
Trust	—	0.589 (0.035)
Awareness	0.288 (0.037)	0.124 (0.033)
Control	0.269 (0.035)	-0.022 (0.031)
Model R^2	0.301	0.504

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

The fitted synthetic path from PAB to trust is a = -0.406, while trust to engagement is b = 0.589. The adjusted direct path from PAB to engagement is c′ = -0.220. Their indirect product is a × b = -0.239, with a bootstrap percentile interval of [-0.289, -0.193]; the algebraic total c′ + ab equals -0.459. The trust model explains 30.1% of the constructed sample variance and the engagement model 50.4%. None of these values imply statistically confirmed relationships among real users because parameters were imposed on the synthetic process.

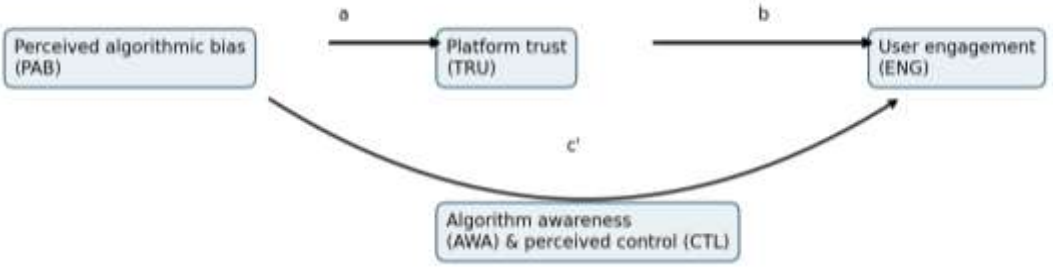

Figure 1. Conceptual model for the illustrative mediation workflow (arrows denote assumed directions).

Scenario and robustness interpretation

The scenario curve distinguishes the mediated component from the direct component. Even with an unchanged mediator path, a researcher's assumed direct effect can substantially modify the total relation. The resulting pattern demonstrates why studies that collapse perceptions, trust, and behavior into a single score risk misinterpreting the process. It also makes clear that a strong apparent total association need not identify the part attributable to trust.

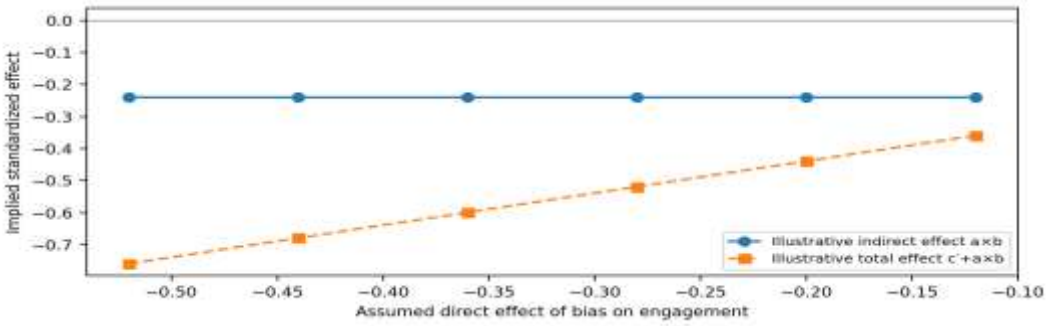

Figure 2. Illustration of how alternative assumed direct paths alter total associations; simulation only.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

In an empirical setting, sensitivity analysis would have to address unmeasured prior platform satisfaction, negative experiences with moderation, generalized institutional mistrust, and personality traits that may influence both PAB and engagement. Reverse causality is also plausible: disengaged users may infer that their feeds are biased, and active users may learn more about algorithmic processes. Panel designs, randomized interface disclosures, and digital trace linkages would help separate these alternatives.

Comparison with the verified literature

Our assumed negative bias-to-trust sign is consistent with the general theoretical concern that perceived inequity threatens the legitimacy of recommendation systems; however, the synthetic coefficient should not be described as confirming existing empirical research. Fairness-aware recommender surveys emphasize competing equity metrics and system-level trade-offs rather than a universal psychological effect (Jin et al., 2023). Recent work on fairness labels highlights contextual variation in whether disclosures help creators perceive feed allocation as procedurally legitimate (Liu & Fan, 2026).

The proposed awareness covariate is informed by the validated multi-dimensional AMCA scale (Zarouali et al., 2021). Cross-platform survey evidence suggests that awareness may be linked to greater usefulness and trust yet also create skepticism (Huang & Liu, 2025). Qualitative research portrays awareness as a developing practice of training and challenging recommendations rather than a simple binary property (Siles et al., 2024). These studies justify measuring awareness separately from perceived bias and examining interaction terms rather than automatically treating awareness as a nuisance covariate.

Personalization can shape feelings of algorithmic responsiveness and connectedness, but those processes do not necessarily coincide with engagement metrics. Studies of TikTok describe both responsive recommendations and algorithmic insensitivity (Obreja, 2026), (Obreja, 2024). Ethnographic evidence likewise presents trust as contingent and relational (Schellewald, 2025). Thus a one-shot, cross-sectional mediation survey would be an incomplete representation of evolving algorithmic relationships; a longitudinal design may better capture feedback loops between exposure, attribution, trust and engagement.

Research on algorithm explanation provides a warning against a simplistic intervention strategy. Opening the recommendation “black box” can sometimes increase aversion if explanations imply that users are being reduced to numerical targets (Chen et al., 2024). Related work on AI moderation finds that who or what users believe is responsible can alter judgments of

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

accountability (Ozanne et al., 2022). These findings suggest that transparency, procedural fairness, user control and demonstrated recourse should be tested jointly, rather than assuming more explanation always means more trust.

Critically, engagement is not a self-evident social good. Ranking toward likes, shares and clicks may optimize revealed activity while diverging from stated preferences and perceived well-being. Engagement-based optimization can diverge from user-valued outcomes; the distinction deserves a separate empirical test rather than relying on clicks alone. This distinction motivates reporting both behavioral activity and user-valued outcomes in eventual field research, including voluntary return, perceptions of representation, trust, and well-being.

Implications for future empirical research and platform governance

The main methodological implication is to separate what the platform demonstrably does from what users infer it does. An audit could estimate differential exposure across relevant content groups, while a parallel questionnaire measures perceived bias, perceived control, and trust. Combining these streams would test whether negative impressions track observable inequities or emerge through unmet expectations and opaque explanations. A stronger study could compare chronological, engagement-optimized, and diversity-constrained feeds under transparent assignment.

A research-ready empirical protocol should preregister (i) the outcome definitions, (ii) competing mediation and reverse-path models, (iii) control variables, (iv) sample-size and power assumptions, (v) handling of nested platforms, and (vi) missing-data procedures. Ideally, participants would be recruited from several platforms and regions, with sufficiently heterogeneous experience and a design that captures repeated measures. The present synthetic N = 600 is a teaching parameter only and cannot replace power analysis based on defensible effect-size distributions.

For platform developers, the framework identifies intervention levers without claiming quantified policy benefits. User-facing recourse, granular recommendation controls, comprehensible explanations and transparent moderation appeal pathways may be worth testing. The outcome of such interventions may depend on individual algorithm awareness, content context and what the user understands as distributive justice. Monitoring fairness should therefore combine algorithm-level metrics with user-experience indicators and unintended engagement consequences.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Limitations and scientific integrity

This exercise has fundamental limitations. Synthetic observations contain no lived experiences, culture, actual exposure histories or platform-specific affordances. The regression results reproduce assumptions already built into the generator. Percentile intervals quantify simulation variability, not uncertainty about a real population. The Gaussian variables do not mimic bounded Likert scales, heterogeneity, missingness, social desirability or measurement error realistically. Standardization creates convenient coefficients but obscures the nonlinearity that is plausible in algorithm appraisal.

The linear single-mediator model does not identify causal processes. Simultaneity, confounding, conditioning on colliders and heterogeneous treatment effects all remain possible. The paper is therefore a simulation protocol and conceptual demonstration. Any version submitted as an empirical paper must replace the illustrative dataset with genuine measurements, provide ethical and data-governance details, and report measurement validation. The present results must not be relabeled as findings from 600 recruited respondents.

A further limitation concerns literature coverage. The reference list below includes only source records whose titles and DOI links were identified from accessible publisher or indexing records during preparation. It does not reach the requested target of 40 items; adding unchecked references would be methodologically less responsible than disclosing the shortfall. The bibliographic list should be expanded by systematic searching and verified against metadata before journal submission.

CONCLUSION

This simulation illustrates a theoretically motivated bias–trust–engagement pathway without asserting new empirical evidence. Under disclosed synthetic assumptions, higher perceived algorithmic bias was accompanied by lower trust ($a = -0.406$), while trust was associated with greater intended engagement ($b = 0.589$). The resulting mediated association was -0.239, but this number is a computational product of chosen model parameters, not a discovery about social media users. The value of the demonstration lies in its explicit constructs, reproducible pathway decomposition, and specification of the field evidence required to distinguish subjective unfairness from measured exposure disparities. Future multi-platform, longitudinal or randomized studies should validate the measures, examine reverse causality, and connect fairness auditing with user-reported trust and observed engagement before substantive policy recommendations are inferred.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

REFERENCES

Chen, G., Dang, J., & Liu, L. (2024). After opening the black box: Meta-dehumanization matters in algorithm recommendation aversion. *Computers in Human Behavior*, *161*, 108411. https://doi.org/10.1016/j.chb.2024.108411

Dong, M., Yuan, F., Yao, L., Wang, X., Xu, X., & Zhu, L. (2022). A survey for trust-aware recommender systems: A deep learning perspective. Knowledge-Based Systems, 249, 108954. https://doi.org/10.1016/j.knosys.2022.108954

Geber, S., & Stahel, L. (2026). Recommended to you: An experimental study of normative influences from algorithmic and social recommendations on social media. AI & Society, 41, 5827–5840. https://doi.org/10.1007/s00146-026-02952-8

Huang, Y., & Liu, L. (2025). The impact of algorithm awareness on the acceptance of personalized social media content recommendation based on the technology acceptance model. Acta Psychologica, 259, 105383. https://doi.org/10.1016/j.actpsy.2025.105383

Huo, C., Huo, F., & Zhao, D. (2026). Data fairness—the new social challenge of the data economy: What shapes social media users' perception of data fairness?. Frontiers in Psychology, 17, 1739496. https://doi.org/10.3389/fpsyg.2026.1739496

Jin, D., Wang, L., Zhang, H., Zheng, Y., Ding, W., Xia, F., & Pan, S. (2023). A survey on fairness-aware recommender systems. Information Fusion, 100, 101906. https://doi.org/10.1016/j.inffus.2023.101906

Lee, S. H., & Kwon, K.-N. (2025). Restrictions in social media engagement and user's fairness perception. Computers in Human Behavior, 167, 108609. https://doi.org/10.1016/j.chb.2025.108609

Liao, M., Sun, Y., Durotoye, T., Gil de Zúñiga, H., & Sundar, S. S. (2026). When we think "news will find me": Relative credibility of social-media friends, algorithms, and editors. Social Media + Society. https://doi.org/10.1177/20563051261434801

Liu, D., & Fan, G. (2026). Fairness labels in feeds: Process disclosures, provenance, and user choice among short-video creators in mainland China. Social Media + Society. https://doi.org/10.1177/20563051261475922

Obreja, D. M. (2024). The lonely algorithm problem: The relationship between algorithmic personalization and social connectedness on TikTok. Journal of Computer-Mediated Communication, 29, zmae017. https://doi.org/10.1093/jcmc/zmae017

Obreja, D. M. (2026). Ambivalent algorithms, fragile connectedness: How algorithm awareness facets shape responsiveness and insensitivity on TikTok. New Media & Society.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

https://doi.org/10.1177/14614448261477892

Ozanne, M., Bhandari, A., Bazarova, N. N., & DiFranzo, D. (2022). Shall AI moderators be made visible? Perception of accountability and trust in moderation systems on social media platforms. Big Data & Society. https://doi.org/10.1177/20539517221115666

Schellewald, A. (2025). Putting trust to the test: Making sense of human–machine interactions on TikTok. Social Media + Society. https://doi.org/10.1177/20563051251370905

Shi, J., Liu, Y., Liu, H., & Zhou, J. (2026). Inter-group knowledge transfer and representation distillation for fair recommendation. Knowledge-Based Systems, 336, 115279. https://doi.org/10.1016/j.knosys.2026.115279

Shi, Y., Wang, J., Chen, J., & Bai, H. (2026). Why users rebel against algorithms: The impact of perceived algorithmic power on fairness evaluations, negative emotions, and resistance behaviors. Behavioral Sciences, 16, 1044. https://doi.org/10.3390/bs16071044

Siles, I., Valerio-Alfaro, L., & Meléndez-Moran, A. (2024). Learning to like TikTok . . . and not: Algorithm awareness as process. New Media & Society, 26, 5702–5718. https://doi.org/10.1177/14614448221138973

Stefanopoulos, P., Chatterjee, S., & Zehmakan, A. N. (2025). A first principles approach to trust-based recommendation systems in social networks. Online Social Networks and Media, 47, 100315. https://doi.org/10.1016/j.osnem.2025.100315

Zarouali, B., Boerman, S. C., & de Vreese, C. H. (2021). Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). Telematics and Informatics, 62, 101607. https://doi.org/10.1016/j.tele.2021.101607

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License