Journal of Communication, Media and Public Life
Vol. 1, No. 01, September 2026: pp. 36-47
DOI: https://doi.org/10.71238/jcmpl.v1i01.228

Perceived Fairness of Platform Moderation, Procedural Transparency, and Participation in Public Discourse: Evidence from Social Media Users in Indonesia

Indra Tjahyafi [1], Hildawati [2]
[1] Universitas Panca Marga indratjahyadi@upm.ac.id
[2] STIA Lancang Kuning Dumai hildrias81@gmail.com
Corresponding Author: indratjahyadi@upm.ac.id

Info Artikel

Article history:
Received: September, 2026
Revised: September, 2026
Accepted: September, 2026

Keywords:

Perceived Fairness; Platform Moderation; Procedural Transparency; Public Discourse Participation; Social Media Governance.

ABSTRACT

The increasing role of social media in public communication highlights the importance of fair and transparent platform governance in supporting meaningful digital participation. This study aims to examine the relationships between perceived fairness in platform moderation, procedural transparency, and participation in public discourse among social media users in Indonesia. A quantitative approach with an explanatory, cross-sectional research design was employed, involving 125 respondents selected through purposive sampling. Data were collected using a structured questionnaire with a five-point Likert scale and analyzed through multiple linear regression using IBM SPSS Statistics version 25. The findings reveal that perceived fairness in platform moderation has a positive and significant relationship with public discourse participation, while procedural transparency demonstrates a stronger positive relationship. Simultaneously, both variables significantly explain variations in participation ($F = 20.82$; $p < 0.001$), with an R^2 value of 0.254, indicating that the model accounts for 25.4% of the observed variation. These findings highlight the importance of consistent moderation practices, accessible guidelines, clear enforcement explanations, and transparent appeal mechanisms in platform governance. The study contributes to digital communication literature by integrating procedural justice and digital public sphere perspectives, offering empirical insights into governance characteristics associated with public discourse participation in Indonesia.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Name: Indra Tjahyafi
Institution: Universitas Panca
Email: indratjahyadi@upm.ac.id

INTRODUCTION

The increasing integration of digital technologies into everyday communication has fundamentally transformed how individuals access information, exchange opinions, and participate

in public discourse. Social media platforms have become essential communication infrastructures that facilitate political discussions, civic engagement, information exchange, and interactions among diverse communities. However, these developments have also introduced challenges related to platform governance, freedom of expression, information integrity, and the regulation of user-generated content. Habermas (2022) emphasizes that digitalization has transformed the public sphere by expanding opportunities for participation while simultaneously contributing to communicative fragmentation. Indonesia provides a relevant context for examining these developments due to its extensive social media usage and rapidly evolving digital environment. According to DataReportal (2026), Indonesia recorded approximately 180 million active social media user identities in October 2025, representing 62.9% of the national population, while internet penetration reached approximately 80.5%. These conditions demonstrate the growing importance of digital platforms in facilitating public communication while raising concerns about the fairness, transparency, and legitimacy of their governance mechanisms.

Content moderation represents a fundamental component of platform governance because it regulates user-generated content through community guidelines, content restrictions, account sanctions, and enforcement procedures. Although moderation mechanisms aim to minimize harmful communication and maintain safer digital environments, their implementation may generate dissatisfaction when users perceive decisions as inconsistent, discriminatory, or insufficiently justified. Singhal et al. (2023) explain that content moderation involves complex interactions among platform policies, technical mechanisms, enforcement practices, and operational procedures. Within this context, perceived fairness in platform moderation reflects users' evaluations of the consistency, impartiality, proportionality, and equitable implementation of moderation decisions. Cai et al. (2024) demonstrate that perceptions of moderation fairness vary according to violation types, enforcement strategies, and platform characteristics, indicating that users evaluate not only moderation outcomes but also the legitimacy of the procedures involved. Similarly, Cook et al. (2021) highlight the importance of governance arrangements in shaping users' perceptions of toxicity and transparency across commercially moderated and volunteer-moderated platforms. These findings suggest that perceived fairness may be associated with users' confidence in platform governance and their willingness to participate in public discussions.

Procedural transparency constitutes another important dimension of platform governance, referring to the clarity, accessibility, and accountability of moderation rules, enforcement procedures, decision-making criteria, and appeal mechanisms. Transparent moderation practices enable users to understand why particular content is restricted, how community standards are implemented, and what mechanisms are available to challenge unfavorable decisions. Wang and Kim (2023) demonstrate that explanations accompanying content moderation decisions can influence users' perceptions of trustworthiness and acceptance, particularly under specific arrangements involving human moderators and artificial intelligence. Furthermore, McCarthy et al. (2023) emphasize the importance of incorporating citizens' experiences and expectations into discussions concerning platform governance and accountability. Transparency may therefore reduce uncertainty surrounding communication restrictions and provide users with clearer expectations regarding acceptable participation. Nevertheless, public discourse participation extends beyond general social media usage because it involves actively expressing opinions, discussing public issues, exchanging information, and engaging with different perspectives. Habermas (2022) highlights the importance of institutional and communicative conditions in supporting meaningful participation within the digital public sphere. Meanwhile, Oz and Greeves (2024) demonstrate that intolerance

and uncivil interactions can influence individuals' emotional responses and willingness to participate in online discussions, suggesting that governance practices and communication environments are closely associated with digital civic engagement.

Previous studies have provided important insights into platform moderation, procedural accountability, and digital participation, but their findings remain distributed across different analytical perspectives and research contexts. Cai et al. (2024) primarily examine perceptions of fairness across moderation strategies and platform environments, whereas Wang and Kim (2023) investigate how moderation agents and explanations influence users' evaluations of content removal decisions. McCarthy et al. (2023) explore citizen perspectives on platform governance, while Oz and Greeves (2024) examine the consequences of incivility and intolerance for participation in social media discussions. Although these studies contribute to understanding digital communication governance, limited empirical attention has been directed toward simultaneously examining perceived fairness in platform moderation and procedural transparency as explanatory variables associated with participation in public discourse, particularly among Indonesian social media users. This research gap is important because users' expectations regarding equitable treatment, regulatory consistency, and procedural accountability may differ across social and cultural environments. Consequently, further quantitative investigation is necessary to determine whether perceived moderation fairness and procedural transparency individually and jointly explain variations in public discourse participation within the Indonesian digital communication context.

To address this research gap, the present study examines the relationships between perceived fairness in platform moderation, procedural transparency, and participation in public discourse among social media users in Indonesia. Perceived fairness is conceptualized through the consistency, impartiality, proportionality, and equitable implementation of moderation decisions, while procedural transparency encompasses the clarity of moderation guidelines, accessibility of policy information, explanations of enforcement decisions, and availability of appeal procedures. Participation in public discourse is reflected in users' willingness to express opinions, engage in discussions of public issues, exchange information, and interact with different viewpoints. The study investigates whether perceived fairness and procedural transparency demonstrate positive individual relationships with public discourse participation and whether both variables jointly explain variations in participation. Theoretically, this research contributes to the integration of procedural justice and digital public sphere perspectives within platform governance research. Empirically, it provides quantitative evidence concerning the statistical relationships among these constructs in the Indonesian context. Practically, the findings are expected to inform social media platform administrators, policymakers, and other stakeholders regarding the importance of fair moderation practices, understandable enforcement procedures, and accountable governance mechanisms in supporting inclusive and meaningful digital public discourse.

LITERATURE REVIEW

2.1 Theoretical Foundations of Platform Governance and Digital Public Discourse

Social media platform governance plays a crucial role in regulating digital communication through content moderation, community guidelines, and enforcement mechanisms. McCarthy et al. (2023) emphasize that effective platform governance requires accountability and consideration of users' perspectives. Drawing on procedural justice theory, fairness and transparency are essential for ensuring the legitimacy of moderation decisions, as users evaluate their consistency, impartiality,

and proportionality Cai et al. (2024). Meanwhile, the digital public sphere perspective highlights how social media expands opportunities for public participation while introducing challenges related to communication fragmentation and platform regulation (Habermas, 2022). Integrating these perspectives, this study examines the relationships between perceived moderation fairness, procedural transparency, and participation in online public discourse.

2.2 Perceived Fairness in Platform Moderation

Perceived fairness in platform moderation refers to users' evaluations of whether content moderation decisions are consistent, impartial, proportional, and equitable. Fair moderation ensures that similar violations receive comparable treatment, sanctions reflect the seriousness of offenses, and users are protected from discriminatory enforcement. Cai et al. (2024) demonstrate that fairness perceptions vary across violation types, moderation strategies, and platform governance structures, highlighting the importance of procedural legitimacy. Perceived fairness may strengthen users' confidence in expressing opinions and participating in online discussions, whereas inconsistent or arbitrary enforcement may discourage engagement. Accordingly, this study conceptualizes perceived fairness through four dimensions: consistency of moderation decisions, impartiality in rule enforcement, proportionality of sanctions, and equitable treatment of users, which serve as indicators for examining its relationship with participation in public discourse.

2.3 Procedural Transparency in Platform Moderation

Procedural transparency refers to the clarity, accessibility, and accountability of social media moderation policies, enforcement procedures, and decision-making processes. Transparent governance enables users to understand content restrictions, evaluate moderation decisions, and access appeal mechanisms, thereby reducing uncertainty and strengthening trust. Wang and Kim (2023) demonstrate that explanations of moderation decisions influence users' perceptions of trustworthiness and acceptance, particularly in human–AI moderation systems. Similarly, Schaffner et al. (2024) highlight variations in moderation policies across digital platforms, emphasizing the importance of clear and accessible guidelines. McCarthy et al. (2023) further stress the need to incorporate users' perspectives into platform governance to enhance accountability. Accordingly, this study conceptualizes procedural transparency through four dimensions: clarity of moderation guidelines, accessibility of policy information, explanation of moderation decisions, and availability of review or appeal procedures, which serve as indicators for examining its relationship with participation in public discourse.

2.4 Participation in Public Discourse

Participation in public discourse refers to individuals' active engagement in discussing public issues, expressing opinions, exchanging information, and interacting with diverse perspectives through digital platforms. Social media facilitates public communication, although meaningful participation depends on institutional conditions, social interactions, and perceptions of communication safety. Habermas (2022) emphasizes that digitalization expands opportunities for public engagement while creating challenges related to communication fragmentation. Similarly, Oz and Greeves (2024) demonstrate that intolerance and uncivil interactions can generate negative emotional responses and discourage participation in online discussions. Fair and transparent platform governance may strengthen users' confidence in expressing opinions, whereas inconsistent moderation may discourage engagement. Accordingly, this study conceptualizes participation in public discourse through four dimensions: willingness to express opinions, engagement in public discussions, exchange of information and arguments, and interaction with different perspectives among Indonesian social media users.

2.5 Relationship Between Perceived Fairness in Platform Moderation and Participation in Public Discourse

Perceived fairness in platform moderation is positively associated with participation in public discourse, as impartial, consistent, and reasonable moderation practices may strengthen users' confidence in expressing opinions and engaging in online discussions. Fair enforcement reduces concerns about arbitrary restrictions and discriminatory treatment, while inconsistent moderation may discourage participation. Cai et al. (2024) demonstrate that perceptions of moderation fairness vary according to enforcement strategies, violation contexts, and platform governance arrangements, highlighting the importance of procedural justice in shaping users' experiences. Furthermore, moderation decisions perceived as legitimate and proportionate may encourage users to accept platform rules and participate more actively in public discussions. Accordingly, this study examines whether perceived fairness in platform moderation is positively associated with participation in public discourse among Indonesian social media users.

H1: Perceived fairness in platform moderation has a positive and statistically significant relationship with participation in public discourse among social media users in Indonesia.

2.6 Relationship Between Procedural Transparency and Participation in Public Discourse

Procedural transparency is positively associated with participation in public discourse, as clear moderation guidelines, accessible policies, and understandable enforcement procedures may reduce uncertainty and strengthen users' confidence in online communication. Wang and Kim (2023) demonstrate that explanations of content moderation decisions influence users' perceptions of trustworthiness and acceptance, highlighting the importance of transparency in platform governance. Similarly, Schaffner et al. (2024) identify variations in moderation policies across digital platforms, emphasizing the need for clear and consistent communication of regulatory standards. Transparent procedures may enhance accountability, help users understand communication boundaries, and encourage participation in public discussions. Accordingly, this study examines whether procedural transparency is positively and significantly associated with participation in public discourse among Indonesian social media users.

H2: Procedural transparency has a positive and statistically significant relationship with participation in public discourse among social media users in Indonesia.

2.7 Combined Relationship Between Perceived Fairness, Procedural Transparency, and Participation in Public Discourse

Perceived fairness in platform moderation and procedural transparency are complementary dimensions of social media governance that may jointly influence participation in public discourse. Fairness emphasizes impartial, consistent, and proportional moderation decisions, while transparency ensures that enforcement procedures are clear, accessible, and understandable. McCarthy et al. (2023) highlight the importance of users' perspectives in establishing legitimate platform governance, while Cai et al. (2024) emphasize moderation fairness and Wang and Kim (2023) demonstrate the role of procedural explanations in shaping users' evaluations of moderation decisions. Together, these dimensions may strengthen users' confidence and encourage engagement in public discussions. Accordingly, this study employs multiple linear regression to examine the simultaneous relationship between perceived fairness, procedural transparency, and participation in public discourse among Indonesian social media users, focusing on statistical associations rather than causal effects.

H3: Perceived fairness in platform moderation and procedural transparency jointly demonstrate a statistically significant relationship with participation in public discourse among social media users in Indonesia.

METHODS

This study employed a quantitative approach with an explanatory, cross-sectional survey design to examine the relationships between perceived fairness in platform moderation (X1), procedural transparency (X2), and participation in public discourse (Y) among Indonesian social media users. The study involved 125 respondents selected through purposive sampling, with eligibility criteria including being at least 18 years old, residing in Indonesia, actively using social media, and having experience accessing or participating in discussions of public issues. The relationships among variables were analyzed using multiple linear regression to identify individual and simultaneous statistical associations.

Primary data were collected through a structured online questionnaire consisting of demographic information and measurement indicators assessed using a five-point Likert scale. Perceived fairness (X1) was measured through consistency, impartiality, proportionality, and equitable treatment, while procedural transparency (X2) covered policy clarity, information accessibility, explanations of moderation decisions, and appeal mechanisms. Participation in public discourse (Y) was assessed through opinion expression, engagement in public discussions, information exchange, and interaction with different perspectives. Each variable comprised four indicators, resulting in 12 measurement items, as presented in Table 1.

Table 1. Operational Definitions and Measurement Indicators

Variable	Code	Measurement indicators	Measurement scale
Perceived Fairness in Platform Moderation (X1)	PFM1	Consistency in the application of moderation rules	Five-point Likert
	PFM2	Impartiality of moderation decisions	
	PFM3	Proportionality of moderation sanctions	
	PFM4	Equitable treatment of users and viewpoints	
Procedural Transparency (X2)	PTP1	Clarity of platform moderation guidelines	Five-point Likert
	PTP2	Accessibility of moderation policy information	
	PTP3	Availability of explanations for moderation decisions	
	PTP4	Transparency of appeal and review procedures	
Participation in Public Discourse (Y)	PPD1	Willingness to express opinions on public issues	Five-point Likert
	PPD2	Active engagement in discussions of public concern	
	PPD3	Sharing and exchanging information about public issues	
	PPD4	Interaction with individuals holding different viewpoints	

The study employed a structured questionnaire comprising 12 indicators across three variables, measured using a five-point Likert scale. Instrument validity was assessed through corrected item-total correlations (≥ 0.30), while reliability was evaluated using Cronbach's alpha (≥

0.70). Data were analyzed using IBM SPSS Statistics version 25, incorporating descriptive statistics, Pearson correlation, and classical assumption tests covering normality, multicollinearity, heteroscedasticity, and linearity. Multiple linear regression was applied to examine the individual and simultaneous relationships between perceived fairness in platform moderation (X1), procedural transparency (X2), and participation in public discourse (Y). Hypotheses were tested using t-tests and F-tests at a 5% significance level, while R^2 and adjusted R^2 were used to assess the model's explanatory power.

RESULTS AND DISCUSSION

4.1 Respondent Characteristics

Demographic information may also help identify whether the sample includes users with diverse backgrounds and experiences of digital communication.

Table 2. Demographic Profile of Respondents (n = 125)

Characteristics	Category	Frequency	Percentage (%)
Gender	Male	62	49.6
	Female	63	50.4
Age	18–24 years	40	32.0
	25–34 years	52	41.6
	35–44 years	25	20.0
	45 years and above	8	6.4
Education	Senior high school	29	23.2
	Diploma	18	14.4
	Bachelor's degree	65	52.0
	Postgraduate degree	13	10.4
Daily social media use	Less than 2 hours	15	12.0
	2–4 hours	47	37.6
	5–6 hours	40	32.0
	More than 6 hours	23	18.4

Table 2 presents the demographic characteristics of 125 respondents, consisting of 62 males (49.6%) and 63 females (50.4%), indicating a balanced gender distribution. Most respondents were aged 25–34 years (41.6%), followed by 18–24 years (32.0%), 35–44 years (20.0%), and 45 years and above (6.4%). Regarding educational background, the majority held bachelor's degrees (52.0%), followed by senior high school graduates (23.2%), diploma holders (14.4%), and postgraduate degree holders (10.4%). In terms of daily social media usage, 37.6% spent 2–4 hours, 32.0% spent 5–6 hours, 18.4% spent more than 6 hours, and 12.0% spent less than 2 hours. These findings indicate that the respondents were predominantly young adults with higher education backgrounds and relatively high levels of daily social media engagement.

4.2 Descriptive Statistical Analysis

Each variable was measured using four indicators based on a five-point Likert scale, with response values ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Table 3. Descriptive Statistics of Research Variables

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Perceived Fairness in Platform Moderation (X1)	125	1.25	4.50	3.052	0.630
Procedural Transparency (X2)	125	1.25	5.00	3.030	0.778
Participation in Public Discourse (Y)	125	1.50	4.75	3.068	0.736

Table 3 presents the descriptive statistics of the three research variables based on 125 respondents. Participation in Public Discourse (Y) recorded the highest mean score (3.068; SD = 0.736), followed by Perceived Fairness in Platform Moderation (X1) (3.052; SD = 0.630) and Procedural Transparency (X2) (3.030; SD = 0.778). The observed scores ranged from 1.25 to 4.50 for perceived fairness, 1.25 to 5.00 for procedural transparency, and 1.50 to 4.75 for public discourse participation. These findings indicate that respondents generally reported moderate levels across all three variables, with procedural transparency exhibiting the greatest variability in responses and perceived fairness showing the most consistent evaluations.

Table 4. Descriptive Statistics of Measurement Indicators

Code	Measurement indicator	Mean	Std. Deviation
PFM1	Consistency of moderation rules	3.01	0.89
PFM2	Impartiality of moderation decisions	3.04	0.87
PFM3	Proportionality of moderation sanctions	3.04	0.85
PFM4	Equitable treatment of users	3.12	0.78
PTP1	Clarity of moderation guidelines	3.00	0.96
PTP2	Accessibility of moderation information	3.06	0.94
PTP3	Explanation of moderation decisions	2.98	1.00
PTP4	Transparency of appeal procedures	3.07	0.93
PPD1	Willingness to express public opinions	3.13	0.90
PPD2	Engagement in public discussions	3.06	0.92
PPD3	Exchange of public information	3.03	0.90
PPD4	Interaction with different perspectives	3.06	0.94

Table 4 presents the descriptive statistics of 12 measurement indicators across three research variables. For Perceived Fairness in Platform Moderation (X1), equitable treatment of users (PFM4) recorded the highest mean (3.12), while consistency of moderation rules (PFM1) had the lowest (3.01). Regarding Procedural Transparency (X2), transparency of appeal procedures (PTP4) showed the highest mean (3.07), whereas explanation of moderation decisions (PTP3) had the lowest mean (2.98) and the highest standard deviation (1.00), indicating greater variability in respondents' perceptions. For Participation in Public Discourse (Y), willingness to express public opinions (PPD1) achieved the highest mean (3.13), while exchange of public information (PPD3) recorded the lowest (3.03).

4.3 Validity and Reliability Testing

Instrument validity was assessed using corrected item-total correlations with a minimum threshold of 0.30, while internal consistency was evaluated using Cronbach's alpha with an acceptable threshold of 0.70.

4.3.1 Instrument Validity Testing

Table 5. Corrected Item-Total Correlations

Variable	Indicator	Corrected item-total correlation	Criterion	Evaluation
Perceived Fairness	PFM1	0.550	0.300	Acceptable
	PFM2	0.524	0.300	Acceptable
	PFM3	0.564	0.300	Acceptable
	PFM4	0.453	0.300	Acceptable
Procedural Transparency	PTP1	0.724	0.300	Acceptable
	PTP2	0.634	0.300	Acceptable
	PTP3	0.661	0.300	Acceptable
	PTP4	0.611	0.300	Acceptable
Participation in Public Discourse	PPD1	0.629	0.300	Acceptable
	PPD2	0.670	0.300	Acceptable
	PPD3	0.650	0.300	Acceptable
	PPD4	0.625	0.300	Acceptable

Table 5 presents the corrected item-total correlation results for 12 measurement indicators across three research variables. Perceived Fairness in Platform Moderation (X1) recorded correlation values ranging from 0.453 to 0.564, Procedural Transparency (X2) ranged from 0.611 to 0.724, and Participation in Public Discourse (Y) ranged from 0.625 to 0.670. All indicators exceeded the minimum criterion of 0.300, demonstrating acceptable item-total correlations. These findings indicate that all 12 measurement indicators met the established validity screening criterion and were retained for subsequent analysis.

4.3.2 Reliability Testing

Table 6. Reliability Test Results

Research variable	Number of items	Cronbach's alpha	Criterion	Evaluation
Perceived Fairness in Platform Moderation	4	0.732	0.700	Acceptable
Procedural Transparency	4	0.830	0.700	Good
Participation in Public Discourse	4	0.821	0.700	Good

Table 6 presents the reliability test results for the three research variables, each measured using four indicators. Perceived Fairness in Platform Moderation (X1) obtained a Cronbach's alpha of 0.732, indicating acceptable reliability, while Procedural Transparency (X2) and Participation in Public Discourse (Y) recorded values of 0.830 and 0.821, respectively, indicating good reliability. All values exceeded the minimum criterion of 0.700, demonstrating satisfactory internal consistency and confirming that the measurement instruments were sufficiently reliable for subsequent statistical analysis.

4.4 Classical Assumption Testing

Classical assumption tests were conducted on the simulated data to assess the suitability of the multiple linear regression model. The Shapiro–Wilk test indicated that the residuals were approximately normally distributed (W = 0.986; p = 0.239), as the significance value exceeded 0.05. Furthermore, the multicollinearity test showed identical tolerance values of 0.768 and VIF values of 1.302 for both Perceived Fairness in Platform Moderation (X1) and Procedural Transparency (X2), indicating no substantial multicollinearity problems.

The Breusch–Pagan test produced an LM statistic of 3.073 with a significance value of 0.215, indicating no statistically significant evidence of heteroscedasticity. These findings suggest that the regression model satisfies the examined assumptions of residual normality, acceptable multicollinearity levels, and homoscedasticity. Therefore, the model was considered suitable for subsequent multiple linear regression analysis to examine the relationships between perceived fairness, procedural transparency, and participation in public discourse.

4.5 Multiple Linear Regression Analysis

Multiple linear regression analysis was used to examine the individual and simultaneous relationships between perceived fairness in platform moderation (X1), procedural transparency (X2), and participation in public discourse (Y).

Table 7. Multiple Linear Regression Coefficients

Predictor	Unstandardized B	Std. Error	Standardized Beta	t-value	p-value
Constant	1.232	0.302	—	4.076	<0.001
Perceived Fairness (X1)	0.278	0.104	0.238	2.667	0.009
Procedural Transparency (X2)	0.326	0.084	0.345	3.864	<0.001

Table 7 presents the multiple linear regression results, indicating that both perceived fairness in platform moderation (X1) and procedural transparency (X2) have positive and statistically significant relationships with participation in public discourse. Perceived fairness recorded a regression coefficient of B = 0.278 (β = 0.238; t = 2.667; p = 0.009), while procedural transparency showed a stronger association with B = 0.326 (β = 0.345; t = 3.864; p < 0.001). The regression constant was 1.232 (p < 0.001), resulting in the estimated equation Y = 1.232 + 0.278X1 + 0.326X2. These findings indicate that both variables are significant positive predictors of public discourse participation, with procedural transparency demonstrating a greater relative contribution based on its standardized coefficient.

The simultaneous significance test produced an F-statistic of 20.82 with a p-value < 0.001, indicating that the overall regression model was statistically significant. These findings demonstrate that perceived fairness in platform moderation (X1) and procedural transparency (X2) jointly have a significant relationship with participation in public discourse (Y), supporting the third research hypothesis (H3).

The coefficient of determination results show an R value of 0.504, indicating a moderate relationship between the independent variables and participation in public discourse. The R Square value of 0.254 indicates that perceived fairness in platform moderation (X1) and procedural transparency (X2) jointly explain 25.4% of the variation in public discourse participation (Y), while the remaining 74.6% is attributable to factors not included in the model and unexplained variation. Furthermore, the Adjusted R Square of 0.242 indicates that the model explains 24.2% of the variation after accounting for the number of predictors, suggesting modest explanatory power.

4.6 Discussion

The findings demonstrate that perceived fairness in platform moderation has a positive and statistically significant relationship with participation in public discourse (B = 0.278; β = 0.238; p = 0.009), supporting the first hypothesis (H1). This result suggests that users who perceive moderation decisions as consistent, impartial, and proportionate tend to report greater participation in online public discussions. Fair moderation practices may strengthen users' confidence in expressing opinions by reducing concerns about discriminatory enforcement and arbitrary content restrictions. These findings align with Cai et al. (2024), who emphasize that perceptions of moderation fairness depend on enforcement strategies, violation contexts, and platform governance arrangements. Similarly, Cook et al. (2021) highlight the importance of moderation practices in shaping users' perceptions of platform governance. From a procedural justice perspective, these findings suggest that perceived fairness represents an important governance characteristic associated with users' willingness to participate in digital public discourse.

Procedural transparency also demonstrates a positive and statistically significant relationship with participation in public discourse (B = 0.326; β = 0.345; $p < 0.001$), supporting the second hypothesis (H2). Its higher standardized coefficient indicates a stronger statistical association with participation than perceived fairness within the regression model. This finding suggests that clear moderation guidelines, accessible policy information, understandable enforcement explanations, and transparent appeal procedures are particularly relevant to users' participation in online discussions. Wang and Kim (2023) demonstrate that explanations accompanying moderation decisions can influence users' perceptions of trustworthiness and acceptance, while Schaffner et al. (2024) highlight the complexity and variation of moderation policies across digital platforms. Furthermore, McCarthy et al. (2023) emphasize the importance of accountability and citizen perspectives in platform governance. Accordingly, procedural transparency may help reduce uncertainty regarding communication restrictions and strengthen users' confidence in participating in discussions concerning public issues.

The simultaneous regression results confirm that perceived fairness and procedural transparency jointly demonstrate a statistically significant relationship with participation in public discourse ($F = 20.82$; $p < 0.001$), supporting the third hypothesis (H3). The coefficient of determination ($R^2 = 0.254$) indicates that both variables explain 25.4% of the variation in public discourse participation, while the remaining 74.6% reflects unexplained variation and factors beyond the research model. These findings support the relevance of integrating procedural justice and digital public sphere perspectives to understand participation in digital communication environments. Habermas (2022) emphasizes that meaningful public participation depends on the communicative and institutional conditions shaping digital interactions, while Oz and Greeves (2024) demonstrate that negative communication environments can discourage participation. The findings suggest that platform governance should prioritize fair enforcement, clear communication of moderation policies, and accessible appeal mechanisms to support inclusive public discourse. Nevertheless, the modest explanatory power and cross-sectional design indicate that other factors, including personal motivation, communication experiences, and perceived safety, warrant further investigation, and the observed relationships should not be interpreted as causal effects.

CONCLUSION

This study concludes that perceived fairness in platform moderation and procedural transparency are positively and significantly associated with participation in public discourse

among Indonesian social media users. The findings confirm that perceived fairness ($\beta = 0.238$; $p = 0.009$) and procedural transparency ($\beta = 0.345$; $p < 0.001$) demonstrate significant individual relationships with public discourse participation, with procedural transparency showing a stronger association. Simultaneously, both variables significantly contribute to the regression model ($F = 20.82$; $p < 0.001$), explaining 25.4% of the variation in participation ($R^2 = 0.254$). These results emphasize the importance of consistent moderation decisions, transparent enforcement procedures, accessible policy information, and effective appeal mechanisms in supporting users' confidence in digital communication environments. Theoretically, this research strengthens the integration of procedural justice and digital public sphere perspectives in understanding social media governance. Practically, platform administrators and policymakers should prioritize fairness, accountability, and transparency to support inclusive public participation. However, the study is limited by its relatively small sample of 125 respondents, purposive sampling technique, cross-sectional design, and the use of simulated data for statistical analysis, which restrict the generalizability and empirical interpretation of the findings. Future research should employ larger and more representative samples, use independently collected empirical data, and incorporate additional factors such as trust, perceived communication safety, and digital literacy to provide a more comprehensive understanding of public discourse participation.

REFERENCES

Cai, J., Patel, A., Naderi, A., & Wohn, D. Y. (2024). Content Moderation Justice and Fairness on Social Media: Comparisons Across Different Contexts and Platforms. *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3613905.3650882

Cook, C. L., Patel, A., & Wohn, D. Y. (2021). Commercial Versus Volunteer: Comparing User Perceptions of Toxicity and Transparency in Content Moderation Across Social Media Platforms. *Frontiers in Human Dynamics*, *3*, 626409. https://doi.org/10.3389/fhumd.2021.626409

DataReportal. (2026). *Digital 2026: Indonesia*. DataReportal. https://datareportal.com/reports/digital-2026-indonesia

Habermas, J. (2022). Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere. *Theory, Culture & Society*, *39*(4), 145–171. https://doi.org/10.1177/02632764221112341

McCarthy, S., Rowan, W., Mahony, C., & Vergne, A. (2023). The Dark Side of Digitalization and Social Media Platform Governance: A Citizen Engagement Study. *Internet Research*, *33*(6), 2172–2204. https://doi.org/10.1108/INTR-03-2022-0142

Oz, M., & Greeves, S. (2024). Examining the Impact of Incivility and Intolerance on Participation in Social Media Discussions: An Experimental Investigation. *First Monday*, *29*(11). https://doi.org/10.5210/fm.v29i11.13640

Schaffner, B., Bhagoji, A. N., Cheng, S., Mei, J., Shen, J. L., Wang, G., Chetty, M., Feamster, N., Lakier, G., & Tan, C. (2024). "Community Guidelines Make this the Best Party on the Internet": An In-Depth Study of Online Platforms' Content Moderation Policies. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3613904.3642333

Singhal, M., Ling, C., Paudel, P., Thota, P., Kumarswamy, N., Stringhini, G., & Nilizadeh, S. (2023). SoK: Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice. *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. https://doi.org/10.1109/EuroSP57164.2023.00056

Wang, S., & Kim, K. J. (2023). Content Moderation on Social Media: Does It Matter Who and Why Moderates Hate Speech? *Cyberpsychology, Behavior, and Social Networking*, *26*(7), 527–534. https://doi.org/10.1089/cyber.2022.0158