Weaponizing Words in a Connected World: Digital Microaggressions, Weaponised Communication, and Technology-Governed Exposure

Zaliha Idris[1]*, Ninderpal Singh Balwant Singh[2], Suhaimee Saahar Saabar[3], Ahlam Abdul Aziz[4]

[1,2]Asia Pacific University of Technology & Innovation (APU), Malaysia

[3,4]Universiti Teknologi MARA (UiTM), Malaysia

***Corresponding Author**

DOI: https://dx.doi.org/10.47772/IJRISS.2026.1013COM0010

Received: 22 February 2026; Accepted: 27 February 2026; Published: 17 March 2026

ABSTRACT

This study examines how digital microaggressions become structurally consequential in platformed communication environments by tracing their structural association with weaponised communication and technology-governed exposure, and by assessing downstream effects on perceived societal vulnerability. Although prior research has examined online hate, emotional virality, and algorithmic amplification in separate streams, fewer studies have modelled how subtle hostile discourse, strategic emotional intensification, and engagement-based platform architectures function as an integrated influence system. Using a cross-sectional survey of 400 active social media users, the study measured four constructs: digital microaggressions (microinsults, microinvalidations, microassaults), weaponised communication (emotional provocation, polarising framing), technology-governed exposure (algorithmic amplification, recommendation frequency), and perceived societal security vulnerability (identity polarisation, social cohesion erosion). Importantly, the construct of societal security vulnerability in this study is conceptualised at the perceptual level. It reflects interpretive assessments of identity polarisation and social cohesion rather than institutional, geopolitical, or policy level instability. This distinction ensures that the findings are bounded within communication-driven risk perception in digital environments. Data were analysed using descriptive statistics and Partial Least Squares Structural Equation Modelling (PLS-SEM). Descriptive results indicate higher exposure to subtle microaggressions (microinsults and microinvalidations) relative to overt hostility, alongside elevated exposure to emotional provocation and perceived algorithmic amplification. Structural results are consistent with a layered amplification pattern in which digital microaggressions are positively associated with weaponised communication, which is associated with technology-governed exposure, corresponding with heightened perceptions of identity polarisation and weakened social cohesion. Rather than implying objective national destabilisation, the findings clarify how repeated exposure to emotionally intensified and algorithmically amplified discourse shapes interpretive perceptions of societal vulnerability. The study advances digital influence research by integrating microaggression theory, emotional contagion dynamics, and platform governance perspectives into a unified structural explanation of how micro level hostility may scale into broader perceptions of societal fragmentation within digitally mediated environments.

Keywords: Digital microaggressions; Weaponised communication; Technology-governed exposure; Societal security vulnerability; Identity polarisation

INTRODUCTION

Digital communication has evolved from a supplementary mode of interaction into a central infrastructure of everyday life, increasingly shaped by the architectures of platformed environments (Erstad et al., 2024; Li et al., 2024; Strauss et al., 2024). In Malaysia, widespread internet penetration and high social media usage have further embedded digital platforms into daily communicative practices (Shien et al., 2023; Yo et al., 2021). Within these environments, communicative acts are not merely shared; they are algorithmically prioritised, recirculated, and amplified through engagement-driven systems designed to maximise visibility and user attention (Chan et al., 2025; Galeazzi et al., 2024; Piccardi et al., 2024). As a result, digital platforms function not only as channels of expression but also as structured environments that actively shape the distribution and salience of discourse (Park & Kim, 2024).

Extensive research has examined key dimensions of contemporary digital communication. Studies have documented the prevalence of online hate speech and harmful discourse (Ali & Stringhini, 2025; Tong, 2024; Yu et al., 2024), explored the dynamics of affective polarisation and emotion-driven information diffusion (Lerman et al., 2024; Marino et al., 2024; Nordbrandt, 2021), and analysed the governance implications of algorithmic infrastructures and platform authority (Airoldi, 2023; Garzonio, 2022; Hunt et al., 2024). While these strands provide valuable insights, they are frequently examined in isolation. Consequently, less attention has been given to how subtle hostile discourse, strategic emotional framing, and engagement-based amplification mechanisms interact as an integrated system of influence.

This study addresses that gap by examining the interrelationships among digital microaggressions, weaponised communication, and technology-governed exposure. Digital microaggressions represent subtle, often normalised forms of identity-based hostility embedded in everyday communication. Weaponised communication refers to the strategic use of emotional provocation and polarising framing to intensify engagement and reinforce group based distinctions. Technology-governed exposure captures the structural dimension of digital platforms, particularly algorithmic amplification and recommendation frequency that increase visibility and repetition. We argue that these elements are structurally associated in a layered amplification pattern: micro-level antagonism provides discursive material, emotional intensification enhances transmissibility, and algorithmic systems scale exposure by prioritising highly engaging content.

Rather than conceptualising digital hostility solely as interpersonal harm, this study situates microaggressions within engagement-driven platform environments. It proposes and empirically tests a structural model linking digital microaggressions to perceived societal security vulnerability through the mediating mechanisms of weaponised communication and technology-governed exposure. In this framework, national security risk is operationalised at the perceptual level through identity polarisation and social cohesion erosion, reflecting interpretive assessments of societal vulnerability rather than objective institutional instability. Accordingly, the concept refers to communication driven perceptions of societal vulnerability within the digital information environment rather than institutional or geopolitical security outcomes.

By integrating microaggression theory, emotional contagion dynamics, and platform governance perspectives, this study contributes to communication scholarship in three ways. First, it conceptualises digital microaggressions not only as interpersonal expressions of hostility but also as discursive signals that may become amplified within engagement-driven platform environments. Second, it proposes a layered amplification framework linking subtle hostile discourse, emotional intensification, and technology-governed exposure within a unified structural model. Third, it reframes societal security concerns at the perceptual level by examining how digitally mediated discourse may influence interpretations of identity polarisation and social cohesion within contemporary information environments.

LITERATURE REVIEW

Digital Microaggressions in Networked Environments

Microaggressions subtle, often ambiguous expressions of hostility directed toward individuals or groups have traditionally been examined within interpersonal, educational, and organisational contexts. These expressions frequently operate through implicit stereotypes, dismissive language, or coded insinuations that communicate exclusion without overt aggression. However, the migration of such interactions into digital environments fundamentally alters their scale, persistence, and potential impact (Idris et al., 2025).

In networked spaces, communicative acts are no longer ephemeral; they are searchable, replicable, and algorithmically recirculated (Kirasur & Jhaver, 2024). Online environments enable repetition, anonymity, and context collapse, allowing hostile expressions to extend far beyond their original communicative setting (Soral et al., 2017; Mane et al., 2023). This persistence influences user experiences, contributing to anxiety, withdrawal from online participation, and identity-based stress, while also reinforcing broader patterns of social polarisation (Ai & Mühlenen, 2025; Cover, 2024; Yadin et al., 2023).

Unlike overt hate speech, digital microaggressions often manifest through normalised discursive patterns, humour, insinuation, or selective framing. Their subtlety renders them socially ambiguous and significantly more difficult for automated moderation systems to detect, frequently bypassing mechanisms designed to identify

explicit hate speech (Hartmann et al., 2024, 2025; Ocampo et al., 2023). Yet cumulative exposure may contribute to desensitisation, prejudice reinforcement, and identity polarisation (Kossowska et al., 2023; Megersa & Minaye, 2023; Park et al., 2023). Research further demonstrates the pervasive impact of digital microaggressions across diverse populations, including LGBTQ individuals, youth, college students, and adolescents, often resulting in adverse psychological and physiological outcomes (Awad & Connors, 2023; Clifton et al., 2025; Hall et al., 2025; McInroy et al., 2023, 2025).

Despite growing recognition of online hostility, much of the literature conceptualises digital microaggressions primarily as psychosocial harm affecting individual well-being. Less attention has been given to their structural implications within algorithmically mediated systems. When embedded in engagement driven platforms, subtle hostility may acquire amplified visibility and repetition, transforming micro-level expressions into broader discursive patterns. This shift from interpersonal harm to structural amplification remains insufficiently theorised.

Weaponised Communication and Emotional Amplification

Weaponised communication refers to strategically constructed messages designed to provoke emotional reactions, intensify polarisation, or influence perception. Unlike disinformation, which centres on factual distortion, weaponised communication may rely on selective framing, exaggerated moral cues, or emotionally charged narratives that activate audience responses without necessarily fabricating content (Paziuk et al., 2025). Actors may exploit cognitive biases, including fear and confirmation bias, to shape interpretations and erode institutional trust (Huffaker, 2023; Paziuk et al., 2025).

Empirical research consistently demonstrates that emotionally arousing content diffuses more rapidly across social networks than neutral information (Paletz et al., 2023; Pivecka et al., 2022; Zerback & Wirz, 2021). Emotions such as anger, outrage, and moral indignation function as accelerants in digital dissemination dynamics (Chen et al., 2023). These affective triggers increase engagement behaviours such as sharing, commenting, and reacting which serve as key signals within platform algorithms (Banerjee & Urminsky, 2024). While positive emotions can also be contagious, negative emotions often persist longer or spread more broadly, highlighting the context-dependent nature of emotional virality (Luo et al., 2024). Such dynamics may contribute to echo chambers and heightened societal polarisation (Gombar & Cvitković, 2025).

Emotional framing does not operate independently of technological infrastructures. Engagement metrics generated through emotional provocation feed directly into ranking and recommendation systems. In this sense, weaponised communication is structurally compatible with algorithmic environments that privilege high engagement content (Vu et al., 2025). Experimental research indicates that engagement-based algorithms frequently amplify emotionally charged and out of group hostile content, even when users express preferences against such material (Milli et al., 2025). This amplification may intensify affective polarisation and antagonistic interactions across ideological divides (Lerman et al., 2024).

However, existing scholarship often treats emotional virality and discursive hostility as separate phenomena. The interaction between subtle digital microaggressions and strategically weaponised emotional framing remains underexplored, particularly in relation to broader perceptions of societal cohesion.

Technology-Governed Exposure and Algorithmic Mediation

Digital platforms are not neutral conduits of information. Their architectures are designed to maximise engagement, frequently privileging content that sustains attention and stimulates affective response (Jahn, 2023; Lazar, 2024; Patil et al., 2024). Recommendation systems and ranking algorithms amplify emotionally provocative and polarising material because such content generates higher interaction metrics (Whittaker et al., 2021; Rozgonyi, 2023). This amplification contributes to what has been described as “algorithmic public opinion,” wherein visibility patterns influence perceived consensus and narrative dominance (Bodo, 2025; Gandini et al., 2025).

Algorithmic amplification functions as a structural force multiplier. It transforms micro-level communicative acts into macro-level visibility patterns, shaping discourse salience and influencing audience perceptions of dominant narratives (Baumann et al., 2025). Platform infrastructures play a central role in structuring knowledge

production and public discourse. Recent experimental studies provide evidence that engagement-based algorithms amplify emotionally charged and divisive content, thereby affecting political discourse and affective polarisation (Milli et al., 2025; Piccardi et al., 2025). Furthermore, algorithms that reward engagement may inadvertently amplify polarisation, particularly when users with more extreme views exhibit higher levels of interaction (D'Ignazi et al., 2025; Piccardi et al., 2024).

Within such environments, repetition and visibility acquire systemic importance. Emotionally engaging content is recirculated, reinforced, and embedded within personalised information streams. This process may shape perceptions of social fragmentation and identity-based division. Nevertheless, most research on algorithmic amplification focuses on extremist or overtly harmful content, with comparatively limited attention given to how subtle hostility interacts with emotional framing to produce cumulative effects.

Identified Theoretical Gap

Collectively, existing research provides substantial insights into online hate speech, emotional virality, and algorithmic governance. However, these dimensions are frequently treated as analytically distinct phenomena. There remains limited theoretical integration explaining how subtle digital microaggressions, strategically weaponised emotional communication, and technology-governed amplification mechanisms operate collectively as a structured system of influence.

Specifically, insufficient attention has been devoted to examining how micro-level hostile expressions may be intensified through emotional framing and subsequently scaled through algorithmic mediation, contributing to perceptions of identity polarisation and erosion of social cohesion. The absence of an integrated structural framework constrains understanding of digitally mediated hostility as more than isolated communicative events.

This study addresses this gap by proposing and empirically testing a unified structural model that conceptualises digital microaggressions, weaponised communication, and technology-governed exposure as interconnected mechanisms shaping perceived national security vulnerability at the level of identity polarisation and social cohesion.

THEORETICAL FRAMEWORK AND HYPOTHESIS DEVELOPMENT

Development of a Conceptual Model

This study proposes a structural model explaining how digitally mediated hostility operate within platformed communication environments and shape perceptions of societal security vulnerability. Integrated theoretical models are necessary for understanding complex online social phenomena that span multiple disciplinary domains (Alturif & Al-Sanad, 2023; Barth et al., 2023; Murray et al., 2025; Oksanen et al., 2021). Rather than treating digital hostility, emotional virality, and algorithmic amplification as separate processes, the present framework synthesises three complementary theoretical perspectives to explain their interdependent dynamics.

First, Microaggression Theory posits that subtle, everyday hostile expressions accumulate over time and contribute to identity threat and marginalisation (Costa et al., 2022; Derald, 2022). Although traditionally examined within interpersonal and organisational contexts, digital environments significantly alter the persistence, visibility, and repetition of such expressions (Burnell et al., 2024; Idris et al., 2025; Yu & Riddle, 2022). In networked spaces, microinsults, microinvalidations, and microassaults can be repeatedly encountered, reshared, and embedded within public discourse. What may initially appear as isolated or ambiguous expressions can therefore accumulate within digitally mediated systems, extending beyond individual-level harm.

Second, Emotional Contagion and Virality Theory suggests that emotionally arousing content particularly anger, outrage, and moral indignation spreads more rapidly and extensively across social networks (Goldenberg & Willer, 2023; Habib & Nithyanand, 2025; Marino et al., 2024; Voinea et al., 2024). Emotional arousal increases engagement behaviours such as sharing and commenting, which function as signals within algorithmic ranking systems (Alloing et al., 2025; Milli et al., 2025; Zhang, 2025). Research consistently demonstrates that high arousal negative emotions are especially effective catalysts of diffusion (Berger & Milkman, 2012; Brady et al., 2021; Leach et al., 2025; McLoughlin et al., 2024). Within digital environments, emotional intensification therefore serves not only as rhetorical strategy but also as a mechanism of increased visibility.

Third, Platform Governance and Algorithmic Amplification perspectives emphasise that digital infrastructures actively shape content exposure rather than functioning as neutral conduits (Galeazzi et al., 2024; Patil et al., 2024). Recommendation systems and engagement-driven ranking processes prioritise content that sustains interaction, structuring visibility patterns in ways that influence public discourse (Bernstein et al., 2023; Bouchaud, 2024; Huszár et al., 2021). Empirical evidence indicates that such systems may amplify emotionally charged and divisive content due to its engagement potential (Metzler & García, 2023; Milli et al., 2025). In this sense, algorithmic infrastructures operate as structural force multipliers within digital communication ecosystems.

Building on these perspectives, the present study conceptualises digitally mediated hostility as a layered amplification process. Digital microaggressions constitute the discursive foundation of subtle antagonism within digitally mediated interactions. Whether subtle (microinsults, microinvalidations) or more explicit (microassaults), these expressions embed identity the relevant cues within everyday exchanges. While often ambiguous, such cues may become strategically activated within emotionally intensified narratives, increasing their salience and transmissibility.

Weaponised communication is conceptualised as the strategic intensification mechanism within this process. Through emotional provocation and polarising framing, discursive hostility is transformed into affectively charged narratives that stimulate engagement (Chaudhary & Penn, 2024; Paziuk et al., 2025; Stepanov, 2023). Emotional provocation heightens arousal, while polarising framing reinforces in-group and out-group boundaries, thereby enhancing diffusion potential within networked systems.

Technology-governed exposure functions as the structural scaling mechanism. Engagement signals generated through emotional mobilisation interact with algorithmic amplification and recommendation frequency, increasing repetition and embedding narratives within personalised information streams. Through this process, micro-level communicative acts acquire macro-level visibility patterns.

Repeated exposure to emotionally intensified and algorithmically amplified discourse may influence the interpretive frameworks through which individuals evaluate social cohesion and group relations (Pierce et al., 2022; Pratap & Pathak, 2025; Vasist et al., 2023; Vasist & Krishnan, 2022). Heightened visibility of antagonistic narratives may contribute to perceptions of identity polarisation and social cohesion erosion. Importantly, these outcomes are conceptualised at the perceptual level, reflecting interpretive assessments of societal vulnerability rather than objective institutional instability.

Collectively, these dynamics represent a layered amplification process linking micro-level discursive hostility to broader perceptions of societal vulnerability. Figure 1 illustrates this structural logic, depicting both direct and mediated relationships among the higher-order constructs.

Figure 1 illustrates the hierarchical structural model linking digital microaggressions, weaponised communication, technology-governed exposure, and perceived societal security vulnerability.

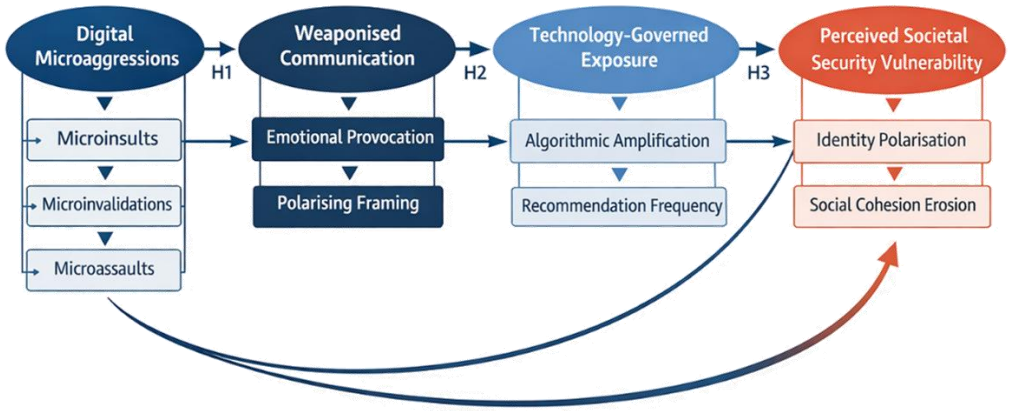

Figure 1: Hierarchical structural model linking digital microaggressions, weaponised communication, technology-governed exposure, and perceived national societal security vulnerability in platformed communication environments.

As shown in Figure 1, digital microaggressions are modelled as a higher-order exogenous construct comprising microinsults, microinvalidations, and microassaults. These dimensions collectively capture a spectrum of hostile expressions embedded in digitally mediated interactions.

Weaponised communication is specified as a mediating construct operationalised through emotional provocation and polarising framing. It represents the mechanism through which subtle hostility is strategically intensified.

Technology-governed exposure is modelled as a second-order mediating construct comprising perceived algorithmic amplification and recommendation frequency. This construct captures how engagement-driven systems prioritise and repeatedly surface emotionally salient content.

At the outcome level, perceived national security vulnerability is operationalised through identity polarisation and social cohesion erosion. Identity polarisation reflects perceptions of intensified group-based antagonism, while social cohesion erosion captures concerns regarding weakening social trust and collective solidarity.

The structural configuration enables examination of both direct and indirect effects. Specifically, the model tests whether digital microaggressions directly influence perceived societal security vulnerability, or whether their effects are sequentially mediated through weaponised communication and technology-governed exposure.

The central theoretical proposition is that subtle hostile discourse becomes structurally consequential not merely through its presence, but through emotional intensification and algorithmic scaling within engagement driven environments.

Accordingly, the model specifies a sequential mediation pathway:

Digital Microaggressions → Weaponised Communication → Technology-Governed Exposure→ Perceived Societal Security Vulnerability

This pathway conceptualises how subtle hostile discourse may acquire wider interpretive significance when emotionally intensified and structurally amplified within engagement-driven platform environments.

Hypothesis Development

Digital Microaggressions and Weaponized Communication

Digital microaggressions operate through implicit insinuation, identity-based cues, and subtle framing that may embed symbolic signals within everyday communication (Abbas et al., 2025; Awad & Connors, 2023). Although often ambiguous, repeated exposure to identity-relevant hostility reinforces social categorisation and boundary construction processes (Megersa & Minaye, 2023; Park et al., 2023). Within networked environments, such expressions may serve as discursive resources that can be strategically intensified.

Emotional contagion theory suggests that identity-salient content is more likely to elicit affective arousal when reframed through emotionally charged or polarising narratives (Lerman et al., 2024). Through selective emphasis and strategic framing, subtle hostility may transition into emotionally intensified discourse designed to provoke engagement (Paziuk et al., 2025; Stepanov, 2023).

Therefore, the following hypotheses are proposed:

H1: Digital microaggressions positively predict weaponised communication.

Weaponized Communication and Algorithmic Amplification

Research on emotional virality consistently demonstrates that emotionally arousing content generates elevated engagement behaviours, including sharing and commenting (Banerjee & Urminsky, 2024; Milli et al., 2025). These engagement signals function as key inputs within algorithmic ranking systems (Alloing et al., 2025; Zhang, 2025).

Platform governance scholarship further indicates that engagement-optimised infrastructures privilege emotionally salient and polarising content due to its higher interaction potential (Bernstein et al., 2023; Galeazzi et al., 2024). Emotional provocation and polarising framing therefore activate the feedback mechanisms through which algorithmic systems determine visibility and repetition.

Therefore, the following hypotheses are proposed:

H2: Weaponised communication positively predicts technology-governed exposure.

Technology-Governed Exposure and Perceived Societal Security Vulnerability

Algorithmic amplification increases the repetition and salience of particular narratives within personalised information streams (Lazar, 2024). Repeated exposure to emotionally intensified and identity-based discourse may reinforce perceptions of social fragmentation and group-based antagonism (Lerman et al., 2024; Nordbrandt, 2021).

In digitally saturated environments, heightened visibility of antagonistic narratives may shape interpretive assessments of societal cohesion and collective stability. The present study conceptualises national security risk at the perceptual level, operationalised through identity polarisation and social cohesion erosion.

Therefore, the following hypotheses are proposed:

H3: Technology-governed exposure positively predicts perceived societal security vulnerability.

METHODOLOGY

Research Design

This study employed a quantitative cross-sectional survey design to empirically test the proposed hierarchical structural model. A cross-sectional approach was appropriate given the study's objective to examine structural relationships among latent constructs within a defined time frame. The design enables simultaneous assessment of direct relationships among digital microaggressions, weaponised communication, technology-governed exposure, and perceived societal security vulnerability.

The research adopts a post-positivist orientation, operationalising theoretically grounded constructs into measurable indicators and evaluating structural relationships using Partial Least Squares Structural Equation Modelling (PLS-SEM). This approach allows for concurrent estimation of measurement reliability and structural paths among latent variables.

Sampling And Data Collection

The target population comprised active social media users regularly exposed to algorithmically curated digital content. Inclusion criteria required respondents to:

1. Be active users of at least one major social media platform,
2. Engage with digital content on a daily or near-daily basis, and
3. Be familiar with content recommendation and engagement features.

Data were collected through a structured self-administered online questionnaire. Participation was voluntary and anonymous. Informed consent was obtained prior to participation, and no personally identifiable information was collected.

A sample size exceeding recommended thresholds for PLS-SEM was targeted to ensure adequate statistical power for estimating hierarchical constructs and mediation pathways. The final sample size satisfied established guidelines for structural model estimation and indirect effect testing.

Instrument Development

The measurement instrument was developed based on established theoretical foundations and adapted to digitally mediated communication environments. All constructs were operationalised as reflective latent

variables measured using multi-item scales. Responses were recorded on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

The questionnaire comprised four principal constructs aligned with the proposed structural model: digital microaggressions, weaponised communication, technology-governed exposure, and perceived societal security vulnerability.

Digital Microaggressions

Digital microaggressions were conceptualised as a higher-order reflective construct encompassing three firstorder dimensions: microinsults, microinvalidations, and microassaults. Microinsults refer to subtle derogatory or stereotypical expressions embedded in digital interactions, often communicated through insinuation or coded language. Microinvalidations capture dismissive statements that negate or minimise identity-based experiences. Microassaults represent comparatively explicit hostile expressions conveyed in digitally mediated settings.

Measurement items assessed respondents perceived exposure to these forms of subtle and overt hostility across social media platforms. By modelling digital microaggressions as a second-order construct, the instrument recognises that these dimensions may vary in intensity while collectively functioning as discursive signals embedded with identity-relevant cues.

Weaponised Communication

Weaponised communication was specified as a first-order reflective construct capturing the strategic intensification of discursive hostility through affective and identity-based framing. It was operationalised through two dimensions: emotional provocation and polarising framing.

Emotional provocation refers to exposure to emotionally charged messaging designed to elicit outrage, fear, or moral indignation. Polarising framing reflects discursive strategies that emphasise in-group and out-group distinctions and reinforce identity boundaries. Together, these dimensions represent the transformation of subtle hostility into emotionally intensified narratives.

Technology-Governed Exposure

Technology-governed exposure was conceptualised as a higher-order reflective construct representing structural mechanisms through which digital content acquires visibility and repetition. This construct comprised algorithmic amplification and recommendation frequency.

Algorithmic amplification refers to the perceived prioritisation of emotionally charged or divisive content by platform ranking systems. Recommendation frequency captures the perceived repetition and resurfacing of similar content within personalised feeds. This operationalisation enables assessment of how engagement-driven platform architectures shape exposure patterns.

Perceived Societal Security Vulnerability

perceived societal security vulnerability was operationalised as a second-order reflective construct measured through identity polarisation and social cohesion erosion.

Identity polarisation reflects perceptions of intensified group-based antagonism and ideological fragmentation. Social cohesion erosion captures perceived weakening of social trust, collective solidarity, and societal unity. Importantly, this construct assesses perceived vulnerability within the information environment rather than objective institutional instability.

Prior to full-scale data collection, the instrument underwent expert review to ensure content validity, conceptual alignment, and clarity of wording. Minor refinements were implemented to enhance construct distinctiveness and reduce semantic overlap.

Measurement Model Specification

A hierarchical component model (HCM) was estimated using PLS-SEM to assess both first-order and higher order constructs.

Digital microaggressions, technology-governed exposure, and perceived societal security vulnerability were specified as second-order reflective constructs using a reflective–reflective specification. Weaponised communication was specified as a first-order reflective construct.

A two-stage approach was employed for estimating higher-order constructs. In Stage 1, first-order constructs were assessed to establish reliability and validity. In Stage 2, latent variable scores from Stage 1 were used to estimate higher-order constructs. This procedure enhances estimation stability and reduces model complexity in hierarchical specifications.

Measurement model evaluation followed established PLS-SEM guidelines. Indicator reliability was assessed via outer loadings (≥ 0.70). Internal consistency reliability was evaluated using composite reliability (≥ 0.70). Convergent validity was assessed through average variance extracted (AVE ≥ 0.50). Discriminant validity was examined using the heterotrait–monotrait ratio (HTMT < 0.85).

Only after satisfying these criteria did the analysis proceed to structural model evaluation.

Data Analysis Procedure

Data were analysed using Partial Least Squares Structural Equation Modelling (PLS-SEM) in SmartPLS 4. PLSSEM was selected due to the hierarchical nature of the proposed model, the inclusion of mediation pathways, and the study's emphasis on predictive explanation rather than covariance-based model fit. The choice of PLSSEM is consistent with contemporary recommendations for analysing complex hierarchical models involving mediation and prediction-oriented research objectives. The analytical procedure followed a two-stage approach.

In the first stage, the measurement model was evaluated to establish reliability and validity. Indicator reliability was assessed through outer loadings, while internal consistency reliability was examined using composite reliability coefficients. Convergent validity was evaluated via average variance extracted (AVE), and discriminant validity was assessed using the heterotrait–monotrait (HTMT) ratio. Only after the measurement model met established thresholds did the analysis proceed to structural evaluation.

In the second stage, the structural model was assessed by examining path coefficients, statistical significance through bootstrapping (5,000 resamples), coefficients of determination (R^2), effect sizes (f^2), and predictive relevance (Q^2). Indirect effects were analysed to evaluate mediation within the proposed sequential framework. This procedure enabled comprehensive examination of both direct and mediated relationships among the study constructs.

Descriptive statistics were also computed to examine exposure prevalence and intensity patterns across the key variables, providing contextual interpretation of the structural findings.

RESULTS

Descriptive Analysis of Key Constructs

Table 1 presents the descriptive statistics for the principal constructs examined in this study (N = 400).

Construct	Dimension	Mean	SD	Level
Digital Microaggressions	Microinsults	3.82	0.71	High
	Microinvalidations	3.65	0.74	Moderate–High
	Microassaults	3.39	0.80	Moderate
Weaponised Communication	Emotional Provocation	4.03	0.65	Very High
	Polarising Framing	3.89	0.68	High
Technology-Governed Exposure	Algorithmic Amplification	4.10	0.62	Very High
	Recommendation Frequency	3.94	0.67	High

Perceived Societal Security Vulnerability	Identity Polarisation	3.83	0.69	High
	Social Cohesion Erosion	3.78	0.73	High

Scale: 1 = Strongly Disagree, 5 = Strongly Agree

Table 1: Descriptive Statistics of Key Variables

Overall, respondents reported moderate to high exposure across all constructs, indicating that digitally mediated hostility, emotional intensification strategies, and algorithmically structured visibility are prevalent features of their online environments.

Among digital microaggressions, microinsults recorded the highest mean, followed by microinvalidations and microassaults. This pattern suggests that subtle and ambiguous forms of hostility are more frequently encountered than overtly aggressive expressions, consistent with the theoretical proposition that digital antagonism often operates through normalised and coded discourse.

Weaponised communication exhibited elevated exposure levels, particularly emotional provocation. The relatively low dispersion across these dimensions suggests consistent exposure patterns across respondents, reinforcing the centrality of affective intensity in digital discourse.

Technology-governed exposure recorded the highest overall mean values, with algorithmic amplification representing the highest single dimension in the dataset. This finding indicates that respondents strongly perceive content visibility as shaped by engagement-driven ranking and recommendation systems.

Perceived societal security vulnerability also recorded high levels, particularly identity polarisation. The convergence of elevated exposure across hostility, emotional intensification, and amplification mechanisms provides descriptive context for the structural relationships tested in the model.

Measurement Model Assessment

The measurement model was evaluated prior to structural testing in accordance with established PLS-SEM guidelines.

Indicator reliability was assessed through outer loadings, which met recommended thresholds. Internal consistency reliability was confirmed via composite reliability coefficients. Convergent validity was established through satisfactory average variance extracted (AVE) values. Discriminant validity was examined using the heterotrait–monotrait (HTMT) ratio, which indicated adequate empirical distinction among constructs.

Collectively, these results demonstrate that the measurement model satisfied reliability and validity requirements, permitting progression to structural model evaluation.

Structural Model Assessment

The structural model was evaluated using bootstrapping procedures (5,000 resamples) within the PLS-SEM framework.

Consistent with H1, digital microaggressions were positively associated with weaponised communication, indicating that exposure to subtle hostile discourse corresponds with increased exposure to emotionally intensified and polarising narratives.

Supporting H2, weaponised communication was positively associated with technology-governed exposure. This finding suggests that emotionally provocative and polarising discourse is more likely to be perceived as algorithmically amplified and repeatedly surfaced within personalised digital feeds.

In line with H3, technology-governed exposure was positively associated with perceived societal security vulnerability, operationalised through identity polarisation and social cohesion erosion. This result is consistent with the proposed amplification mechanism, whereby emotionally intensified discourse, when structurally scaled through engagement-driven systems, is linked to heightened perceptions of societal vulnerability.

Mediation Analysis

Mediation analysis was conducted to examine whether the association between digital microaggressions and perceived societal security vulnerability operates through the proposed sequential mechanisms of weaponised communication and technology-governed exposure.

The findings indicate that weaponised communication serves as an intermediary mechanism linking digital microaggressions to technology-governed exposure. In turn, technology-governed exposure mediates the relationship between weaponised communication and perceived societal security vulnerability. Collectively, these results support the proposed layered amplification framework, whereby subtle hostile discourse becomes structurally consequential when emotionally intensified and subsequently scaled through engagement-driven platform systems.

Given the cross-sectional design, these associations should be interpreted as indicative of relational patterns rather than definitive causal effects.

DISCUSSION

This study examined how digital microaggressions, weaponised communication, and technology-governed exposure operate within platformed communication environments and how these dynamics relate to perceived societal security vulnerability. By integrating discursive hostility, emotional intensification strategies, and algorithmic amplification processes, the findings provide a structured explanation of how subtle online expressions may scale into broader perceptions of societal vulnerability.

The results indicate that digital microaggressions are positively associated with weaponised communication. This finding suggests that subtle hostile discourse may function as adaptable narrative material that can be strategically intensified within digital spaces. Whereas prior research has primarily conceptualised microaggressions as individual-level psychological stressors, the present study highlights their communicative function within networked systems. In digitally mediated environments, microinsults and microinvalidations often normalised or ambiguous may be reframed in emotionally salient ways that reinforce identity-based distinctions and stimulate engagement.

The findings further demonstrate that weaponised communication is positively associated with technology governed exposure. Emotional provocation and polarising framing appear structurally aligned with engagement driven platform logics, where affectively salient content tends to receive increased visibility and repetition. Rather than portraying algorithms as inherently destabilising forces, the results indicate that emotionally intensified discourse is structurally advantaged within engagement optimised systems.

Importantly, technology-governed exposure is positively associated with perceived societal security vulnerability, operationalised through identity polarisation and social cohesion erosion. The study does not suggest that digital communication directly destabilises national institutions. Instead, the findings indicate that repeated exposure to emotionally intensified, and algorithmically amplified discourse may shape how individuals interpret societal cohesion and group relations. In this framework, national security risk is conceptualised as a perceptual outcome emerging from communication environments rather than as an objective measure of institutional breakdown.

Mediation analysis further supports this layered process. Digital microaggressions were indirectly associated with perceived societal security vulnerability through weaponised communication and technology-governed exposure. This pattern reinforces the proposition that subtle hostility acquires structural significance when emotionally intensified and repeatedly surfaced within personalised information streams. By modelling these dynamics within a unified framework, the study contributes to communication scholarship by demonstrating how microlevel discourse, affective mobilisation, and platform architectures operate as interconnected components of influence systems rather than as isolated phenomena.

Overall, the findings underscore the importance of analysing digitally mediated communication through a systems-oriented lens. Subtle discursive hostility, emotional intensification, and algorithmic prioritisation may jointly shape perceptions of social fragmentation in digitally saturated environments.

Policy Implications

The findings suggest that communication governance discussions should extend beyond overt hate speech to include subtle forms of discursive antagonism that may become amplified within engagement-driven environments. Emotional framing plays a central role in visibility dynamics, indicating that platform design incentives and engagement metrics warrant critical examination.

Greater transparency regarding content ranking and recommendation mechanisms may support more informed public understanding of exposure patterns. In parallel, digital literacy initiatives can strengthen users' capacity to critically interpret emotionally charged narratives and recognise how engagement-based systems influence content salience.

Rather than focusing exclusively on content removal, governance strategies may benefit from addressing the structural interaction between emotional intensification and algorithmic prioritisation within digital ecosystems.

Limitations

Despite its contributions, this study has several limitations.

First, the cross-sectional design restricts causal inference. Although the structural pathways are theoretically grounded, the data reflect associations measured at a single time point. Longitudinal designs would be necessary to examine cumulative exposure patterns and potential changes in perceptions over time.

Second, the study relies on self-reported perceptions of exposure. While perceptual measures are appropriate for assessing interpretative outcomes such as perceived societal security vulnerability, they do not directly capture objective platform-level amplification patterns. Future research incorporating behavioural trace data or computational analytics could provide complementary evidence.

Third, national security risk was operationalised through perceptual dimensions identity polarisation and social cohesion erosion rather than institutional or policy-level indicators. The findings therefore reflect perceived societal vulnerability rather than measurable state-level instability.

Fourth, the sample consists of active social media users within a defined context, which may limit generalisability across platforms, demographic groups, or geopolitical environments.

Finally, the model does not incorporate potential moderating variables, such as digital literacy, media trust, or ideological orientation, which may condition amplification effects.

Future Research Directions

Future research should employ longitudinal designs to examine cumulative exposure dynamics and potential nonlinear shifts in perception over time. Integrating behavioural trace data or computational modelling with perceptual survey measures would strengthen causal inference and clarify how emotionally intensified content circulates across platform ecosystems.

Cross-platform comparative studies are also warranted, as platform affordances, moderation practices, and algorithmic logics vary considerably. Expanding demographic scope would further illuminate how identity position, engagement intensity, and digital literacy shape exposure and interpretation patterns.

Finally, extending outcome variables beyond perceptual national security risk to include civic disengagement, institutional trust, or democratic resilience perceptions would deepen understanding of the broader societal implications of digitally mediated hostility.

CONCLUSION

This study advances communication scholarship by offering an integrated framework for understanding how digital microaggressions operate within platformed communication environments. By modelling the structural associations between subtle hostile discourse, emotional intensification, and technology-governed exposure, the findings clarify how digitally mediated interactions may shape perceptions of identity polarisation and social cohesion.

Rather than asserting direct institutional destabilisation, the study demonstrates how communication processes within engagement-driven platforms may influence interpretive perceptions of societal vulnerability. By integrating microaggression theory, emotional contagion dynamics, and platform governance perspectives, the research positions digitally mediated discourse as a structural variable within contemporary information environments.

REFERENCES

1. Abbas, N. F., Eidan, S. S., & Muslah, A. F. (2025). Critical Discourse Analysis of Online Platforms: An Overview. International Journal of Social Sciences and English Literature, 9(8), 27. https://doi.org/10.55220/2576-683x.v9.551
2. Ai, Y., & Mühlenen, A. von. (2025). An experimental online study on the impact of negative social media comments on anxiety and mood. Scientific Reports, 15(1), 26642. https://doi.org/10.1038/s41598-02510810-8
3. Airoldi, M. (2023). Computational Authority in Platform Society. In Oxford University Press eBooks. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197653609.013.40
4. Ali, S., & Stringhini, G. (2025). Evolving Hate Speech Online: An Adaptive Framework for Detection and Mitigation. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2502.10921
5. Alloing, C., Fortant, E., Pierre, J., Richert, F., & Palisser, R. (2025). Knowing Your Users by Heart: A Critical Examination of the Scientific Research on Emotions Conducted by Social Media Platforms. Social Media + Society, 11(3). https://doi.org/10.1177/20563051251355456
6. Alnaqbi, H. H., & Ali, E. A. (2025). Social media impact on societal security. Frontiers in Sociology, 10, 1508542. https://doi.org/10.3389/fsoc.2025.1508542
7. Alturif, G., & Al-Sanad, H. A. R. (2023). The repercussions of digital bullying on social media users. Frontiers in Psychology, 14, 1280757. https://doi.org/10.3389/fpsyg.2023.1280757
8. Arora, S. D., Singh, G. P., Chakraborty, A., & Maity, M. (2022). Polarization and social media: A systematic review and research agenda. Technological Forecasting and Social Change, 183, 121942. Elsevier BV. https://doi.org/10.1016/j.techfore.2022.121942
9. Awad, M. N., & Connors, E. H. (2023). Active bystandership by youth in the digital era: Microintervention strategies for responding to social media-based microaggressions and cyberbullying. Psychological Services, 20(3), 423. https://doi.org/10.1037/ser0000749
10. Banerjee, A., & Urminsky, O. (2024). The Language That Drives Engagement: A Systematic Large-scale Analysis of Headline Experiments. Marketing Science, 44(3), 566. https://doi.org/10.1287/mksc.2021.0018
11. Barth, N., Wagner, E., Raab, P., & Wiegärtner, B. (2023). Contextures of hate: Towards a systems theory of hate communication on social media platforms. The Communication Review, 26(3), 209. https://doi.org/10.1080/10714421.2023.2208513
12. Baumann, F., Arora, N., Rahwan, I., & Czaplicka, A. (2025). Dynamics of Algorithmic Content Amplification on TikTok. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2503.20231
13. Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? Journal of Marketing Research, 49(2), 192. https://doi.org/10.1509/jmr.10.0353
14. Bernstein, M. S., Christin, A., Hancock, J. T., Hashimoto, T., Jia, C., Lam, M. S., Meister, N., Persily, N., Piccardi, T., Saveski, M., Tsai, J. L., Ugander, J., & Xu, C. (2023). Embedding Societal Values into Social Media Algorithms. Journal of Online Trust and Safety, 2(1). https://doi.org/10.54501/jots.v2i1.148
15. Bodo, B. (2025). Humble tools of divine intervention – The misunderstood role of algorithms in public opinion formation. Dialogues on Digital Society. https://doi.org/10.1177/29768640251369974
16. Bouchaud, P. (2024). Skewed perspectives: examining the influence of engagement maximization on content diversity in social media feeds. Journal of Computational Social Science, 7(1), 721. https://doi.org/10.1007/s42001-024-00255-w
17. Brady, W. J., Jackson, J. C., Lindström, B., & Crockett, M. J. (2023). Algorithm-mediated social learning in online social networks. Trends in Cognitive Sciences, 27(10), 947. Elsevier BV. https://doi.org/10.1016/j.tics.2023.06.008

18. Brady, W. J., McLoughlin, K. L., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. Science Advances, 7(33). https://doi.org/10.1126/sciadv.abe5641
19. Burnell, K., Trekels, J., George, M. J., & Nesi, J. (2024). Digital Cruelty's Impact on Self-Esteem and Body Image (p. 439). https://doi.org/10.1007/978-3-031-69362-5_60
20. Çakmak, M. C., Agarwal, N., & Oni, R. (2024). The bias beneath: analyzing drift in YouTube's algorithmic recommendations. Social Network Analysis and Mining, 14(1). https://doi.org/10.1007/s13278-02401343-5
21. Chan, J., Choi, F., Saha, K., & Chandrasekharan, E. (2025). The Ranking Effect: How Algorithmic Rank Influences Attention on Social Media. https://doi.org/10.48550/ARXIV.2509.18440
22. Chaudhary, Y., & Penn, J. (2024). Large Language Models as Instruments of Power: New Regimes of Autonomous Manipulation and Control. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2405.03813
23. Ciavolino, E., Ferrante, L., Sternativo, G. A., Cheah, J., Rollo, S., Marinaci, T., & Venuleo, C. (2021). A confirmatory composite analysis for the Italian validation of the interactions anxiousness scale: a higherorder version. Behaviormetrika, 49(1), 23. https://doi.org/10.1007/s41237-021-00151-x
24. Clifton, M., Clifton, R. L., & Zapolski, T. C. B. (2025). Racial Media Microaggressions: Impact on Black Adolescent Mental and Behavioral Health. OTJR Occupational Therapy Journal of Research. https://doi.org/10.1177/15394492251385473
25. Conn, V. S. (2020). Crafting Effective Abstracts. Western Journal of Nursing Research, 44(7), 635. https://doi.org/10.1177/0193945920920638
26. Costa, P. L., McDuffie, J. W., Brown, S. E. V., He, Y., Ikner, B. N., Sabat, I. E., & Miner, K. N. (2022). Microaggressions: Mega problems or micro issues? A meta-analysis. Journal of Community Psychology, 51(1), 137. Wiley. https://doi.org/10.1002/jcop.22885
27. Cover, R. (2024). Apprehending digital hostility and online abuse: Feminist care ethics in/and digital ecologies. Thesis Eleven, 183(1), 33. https://doi.org/10.1177/07255136241284658
28. Cui, P. (2025). Dynamics and Inequalities in Digital Social Networks: A Computational and Sociological Review. https://doi.org/10.48550/ARXIV.2503.02887
29. DeCuir-Gunby, J. T., McCoy, W. N., Gibson, S. M., Modaressi, S. L., & Macias, A. J. (2024). Using Critical Race Mixed Methodology to Explore African American College Students' Experiences with Racial Microaggressions. Innovative Higher Education, 49(6), 1077. https://doi.org/10.1007/s10755-024-097326
30. Derald, W. (2022). Microaggressions: Death by a Thousand Cuts. Scientific American, 31, 48. https://doi.org/10.1038/scientificamericanracism0621-48
31. D'Ignazi, J., Kaltenbrunner, A., Mens, G. L., Germano, F., & Gómez, V. (2025). Rewarding Engagement and Personalization in Popularity-Based Rankings Amplifies Extremism and Polarization. https://doi.org/10.48550/ARXIV.2510.24354
32. Erstad, O., Hegna, K., Livingstone, S., Negru-Subtirica, O., & Stoilova, M. (2024). How digital technologies become embedded in family life across generations: scoping the agenda for researching 'platformised relationality.' Families Relationships and Societies, 13(2), 164. https://doi.org/10.1332/20467435y2024d000000023
33. Eschmann, R. (2020). Digital Resistance: How Online Communication Facilitates Responses to Racial Microaggressions. Sociology of Race and Ethnicity, 7(2), 264. https://doi.org/10.1177/2332649220933307
34. Galeazzi, A., Paudel, P., Conti, M., De Cristofaro, E., & Stringhini, G. (2024). Revealing The Secret Power: How Algorithms Can Influence Content Visibility on Twitter/X. https://doi.org/10.48550/ARXIV.2410.17390
35. Gandini, A., Keeling, S., & Reviglio, U. (2025). Conceptualising the 'algorithmic public opinion': Public opinion formation in the digital age. Dialogues on Digital Society. https://doi.org/10.1177/29768640251323147

36. Garzonio, E. (2022). Performative Intermediaries Versus Digital Regulation. A Multidisciplinary Analysis of the Power of Algorithms. In Frontiers in sociology and social research (p. 157). Springer International Publishing. https://doi.org/10.1007/978-3-031-11756-5_10
37. Goldenberg, A., & Willer, R. (2023). Amplification of emotion on social media. Nature Human Behaviour, 7(6), 845. https://doi.org/10.1038/s41562-023-01604-x
38. Gombar, M., & Cvitković, M. K. (2025). Cognitive Resonance Theory in Strategic Communication: Understanding Personalization, Emotional Resonance, and Echo Chambers. OALib, 12(4), 1. https://doi.org/10.4236/oalib.1113171
39. Goswami, G. (2025). AI Echo Chambers: How Algorithms Shape Reality, Influence Democracy, and Fuel Misinformation. https://doi.org/10.36227/techrxiv.174059950.03385147/v1
40. Habib, H., & Nithyanand, R. (2025a). YouTube Recommendations Reinforce Negative Emotions: Auditing Algorithmic Bias with Emotionally-Agentic Sock Puppets. https://doi.org/10.48550/ARXIV.2501.15048
41. Habib, H., & Nithyanand, R. (2025b). YouTube Recommendations Reinforce Negative Emotions: Auditing Algorithmic Bias with Emotionally-Agentic Sock Puppets. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2501.15048
42. Hall, A. R., Meter, D. J., Culianos, D., Uresti, A., Medina, M., & Nishina, A. (2025). The effects of college students' online experiences with racial/ethnic discrimination. Social Psychology of Education, 28(1). https://doi.org/10.1007/s11218-025-10137-2
43. Harmer, E., & Southern, R. (2021). Digital microaggressions and everyday othering: an analysis of tweets sent to women members of Parliament in the UK. Information Communication & Society, 24(14), 1998. https://doi.org/10.1080/1369118x.2021.1962941
44. Hartmann, D., Oueslati, A., & Staufer, D. (2024). Watching the Watchers: A Comparative Fairness Audit of Cloud-based Content Moderation Services. https://doi.org/10.48550/ARXIV.2406.14154
45. Hartmann, D., Oueslati, A., Staufer, D., Pohlmann, L., Munzert, S., & Heuer, H. (2025). Lost in Moderation: How Commercial Content Moderation APIs Over- and Under-Moderate Group-Targeted Hate Speech and Linguistic Variations. 1. https://doi.org/10.1145/3706598.3713998
46. Hu, B., Zhu, Y., Liu, C., Zheng, S., Zhao, Z., & Bao, R. (2024). Collectivism, face concern and Chinesestyle lurking among university students: the moderating role of trait mindfulness. Frontiers in Psychology, 15. https://doi.org/10.3389/fpsyg.2024.1298357
47. Huffaker, J. S. (2023). Social Reference Processing with Collaborative Human-AI Systems. Deep Blue (University of Michigan). https://doi.org/10.7302/8321
48. Hunt, R. A., Townsend, D. M., Simpson, J., Nugent, R., Stallkamp, M., & Bozdag, E. (2024). Digital Battlegrounds: The Power Dynamics and Governance of Contemporary Platforms. Academy of Management Annals. https://doi.org/10.5465/annals.2022.0188
49. Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2021). Algorithmic amplification of politics on Twitter. Proceedings of the National Academy of Sciences, 119(1). https://doi.org/10.1073/pnas.2025334119
50. Idris, Z., Warganegara, D. S., Saabar, S. S., Aziz, A. A., & Idris, J. (2025). Microaggressions on Social Media: A Conceptual Framework of 3R Issues and Youth Consumer Behaviour in Malaysia. International Journal of Research and Innovation in Social Science, 314. https://doi.org/10.47772/ijriss.2025.913com0029
51. Jahn, L. (2023). Curbing Amplifiation Online : Towards Improving the Quality of Information Spread on Social Media Using Agent-Based Models and Twitter Data. Research Portal Denmark, 150. https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=ku&id=ku-ef61aeed-56f5-4d78b171a00a80798d1a&ti=Curbing%20Amplifiation%20Online%20%3A%20Towards%20Improving%20the%20Quality%20of%20Information%20Spread%20on%20Social%20Media%20Using%20AgentBased%20Models%20and%20Twitter%20Data
52. Kasi, A. Z., Kasi, M., & Qadir, A. (2021). The Effects of Social Media on National Security: An Overview. Global Strategic & Securities Studies Review, 121. https://doi.org/10.31703/gsssr.2021(vi-i).13

53. Kearney, A., Poredi, N., Shelton, J., Akcinaroglu, S., Karakoç, E., Tran, T. N. T., & Chen, Y. (2025). Echoes amplified: a study of AI-generated content and digital echo chambers. 23. https://doi.org/10.1117/12.3053447
54. Khambatta, P., Mariadassou, S., Morris, J., & Wheeler, S. C. (2023). Tailoring recommendation algorithms to ideal preferences makes users better off. Scientific Reports, 13(1), 9325. https://doi.org/10.1038/s41598-023-34192-x
55. Kirasur, N., & Jhaver, S. (2024). Understanding the Resilience of Caste: A Critical Discourse Analysis of Community Profiles on X. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2407.02810
56. Kossowska, M., Kłodkowski, P., Siewierska-Chmaj, A., Guinote, A., Kessels, U., Moyano, M., & Strömbäck, J. (2023). Internet-based micro-identities as a driver of societal disintegration. Humanities and Social Sciences Communications, 10(1). https://doi.org/10.1057/s41599-023-02441-z
57. Lazar, S. (2024). Lecture II: Communicative Justice and the Distribution of Attention. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2410.20718
58. Leach, S., Formanowicz, M., Nikadon, J., & Cichocka, A. (2025). Moral Outrage Predicts the Virality of Petitions for Change on Social Media, But Not the Number of Signatures They Receive. Social
59. Psychological and Personality Science, 17(2), 194. https://doi.org/10.1177/19485506251335373
60. Lerman, K., Feldman, D., He, Z., & Rao, A. (2024). Affective polarization and dynamics of information spread in online networks. Npj Complexity, 1(1). https://doi.org/10.1038/s44260-024-00008-w
61. Li, C., Qin, J., Wang, X., & Xue, F. (2024). Navigating the Digital Horizon: The Impact and Future of Communication Technologies in Society. Lecture Notes in Education Psychology and Public Media, 51(1), 223. https://doi.org/10.54254/2753-7048/51/20240938
62. Luo, Y., Sornette, D., & Lera, S. C. (2024). Quantification of the Self-Excited Emotion Dynamics in Online Interactions. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2408.05700
63. Marchi, M., Travascio, A., Uberti, D., Micheli, E. D., Quartaroli, F., Laquatra, G., Grenzi, P., Pingani, L., Ferrari, S., Fiorillo, A., Converti, M., Pinna, F., Amaddeo, F., Ventriglio, A., Mirandola, M., & Galeazzi, G. M. (2023). Microaggression toward LGBTIQ people and implications for mental health: A systematic. International Journal of Social Psychiatry, 70(1), 23. SAGE Publishing. https://doi.org/10.1177/00207640231194478
64. Mariano, L. de O., Moura, L. de, Mattos, R. H. P., Bizarria, F. P. de A., & Kind, L. (2025). Faces of exclusion: the "social," the "digital" and "digital racism" in a decolonial critical essay. Frontiers in Sociology, 10, 1534313. https://doi.org/10.3389/fsoc.2025.1534313
65. Marino, E. B., Baleato, S., & Ribeiro, A. S. (2024). The Polarization Loop: How Emotions Drive Propagation of Disinformation in Online Media—The Case of Conspiracy Theories and Extreme Right Movements in Southern Europe. Social Sciences, 13(11), 603. https://doi.org/10.3390/socsci13110603
66. Martino, E. D., Galeazzi, A., Starnini, M., Quattrociocchi, W., & Cinelli, M. (2024). Characterizing the Fragmentation of the Social Media Ecosystem. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2411.16826
67. McInroy, L. B., Beer, O., Scheadler, T. R., Craig, S. L., & Eaton, A. D. (2023). Exploring the psychological and physiological impacts of digital microaggressions and hostile online climates on LGBTQ + youth. Current Psychology, 43(3), 2586. https://doi.org/10.1007/s12144-023-04435-1
68. McInroy, L. B., Scheadler, T. R., McDonald, M., Eaton, A. D., & Craig, S. L. (2025). Digital microaggressions and LGBTQ+ youth: exploring potential impacts and opportunities for educational intervention. Educational Psychology, 1. https://doi.org/10.1080/01443410.2025.2541743
69. McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., & Crockett, M. J. (2024). Misinformation exploits outrage to spread online. Science, 386(6725), 991. https://doi.org/10.1126/science.adl2829
70. Megersa, T., & Minaye, A. (2023). Social media users' online behavior with regard to the circulation of hate speech. Frontiers in Communication, 8. https://doi.org/10.3389/fcomm.2023.1276245
71. Metzler, H., & García, D. (2023). Social Drivers and Algorithmic Mechanisms on Digital Media. Perspectives on Psychological Science, 19(5), 735. https://doi.org/10.1177/17456916231185057

72. Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. PNAS Nexus, 4(3). https://doi.org/10.1093/pnasnexus/pgaf062
73. Murray, R., Orr, B., Al-khateeb, S., & Agarwal, N. (2025). Constructing a multi-theoretical framework for mob modeling. Social Network Analysis and Mining, 15(1). https://doi.org/10.1007/s13278-025-01449-4
74. Myrick, R., & Wang, C. (2024). Domestic Polarization and International Rivalry: How Adversaries Respond to America's Partisan Politics. The Journal of Politics, 86(1), 141. https://doi.org/10.1086/726926
75. Nordbrandt, M. (2021). Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties. New Media & Society, 25(12), 3392. https://doi.org/10.1177/14614448211044393
76. Ocampo, N. B., Свиридова, E., Cabrio, E., & Villata, S. (2023). An In-depth Analysis of Implicit and Subtle Hate Speech Messages. 1997. https://doi.org/10.18653/v1/2023.eacl-main.147
77. Oksanen, A., Celuch, M., Latikka, R., Oksa, R., & Savela, N. (2021). Hate and harassment in academia: the rising concern of the online environment. Higher Education, 84(3), 541. https://doi.org/10.1007/s10734-021-00787-4
78. Park, A., Kim, M., & Kim, E.-S. (2023). SEM analysis of agreement with regulating online hate speech: influences of victimization, social harm assessment, and regulatory effectiveness assessment. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1276568
79. Park, J., & Kim, S. (2024). Why do users perceive digital platforms as indispensable to their lives?: A study on KakaoTalk in Korea. Telecommunications Policy, 48(10), 102863. https://doi.org/10.1016/j.telpol.2024.102863
80. Patil, S., Jani, A. S., & Konatam, S. (2024). Trend Amplification or Suppression: The Dual Role of AI in
81. Influencing Viral Content. International Journal of Global Innovations and Solutions (IJGIS). https://doi.org/10.21428/e90189c8.361bcc7f
82. Paziuk, A., Lande, D., Shnurko-Tabakova, E., & Kingston, P. (2025). Decoding manipulative narratives in cognitive warfare: a case study of the Russia-Ukraine conflict. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1566022
83. Piccardi, T., Saveski, M., Jia, C., Hancock, J. T., Tsai, J. L., & Bernstein, M. A. (2024a). Social Media Algorithms Can Shape Affective Polarization via Exposure to Antidemocratic Attitudes and Partisan Animosity. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2411.14652
84. Piccardi, T., Saveski, M., Jia, C., Hancock, J. T., Tsai, J. L., & Bernstein, M. S. (2024b). Reranking Social Media Feeds: A Practical Guide for Field Experiments. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2406.19571
85. Piccardi, T., Saveski, M., Jia, C., Hancock, J., Tsai, J. L., & Bernstein, M. S. (2025). Reranking partisan animosity in algorithmic social media feeds alters affective polarization. Science, 390(6776). https://doi.org/10.1126/science.adu5584
86. Pierce, G. L., Holland, C. C., Cleary, P. F., & Rabrenovic, G. (2022). The opportunity costs of the politics of division and disinformation in the context of the twenty-first century security deficit. SN Social Sciences, 2(11). https://doi.org/10.1007/s43545-022-00514-5
87. Pottier, P., Lagisz, M., Burke, S., Drobniak, S. M., Downing, P. A., Macartney, E. L., Martinig, A. R., Mizuno, A., Morrison, K., Pollo, P., Ricolfi, L., Tam, J., Williams, C., Yang, Y., & Nakagawa, S. (2024). Title, abstract and keywords: a practical guide to maximize the visibility and impact of academic papers. Proceedings of the Royal Society B Biological Sciences, 291(2027), 20241222. https://doi.org/10.1098/rspb.2024.1222
88. Pratap, A., & Pathak, A. (2025). From Public Square to Echo Chamber: The Fragmentation of Online Discourse. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2501.18441
89. Shien, O. Y., Huei, N. S., & Yan, N. L. (2023). The Impact of Social Media Marketing on Young Consumers' Purchase Intention in Malaysia: The Mediating Role of Consumer Engagement. International

Journal of Academic Research in Business and Social Sciences, 13(1). https://doi.org/10.6007/ijarbss/v13i1/15806

90. Shin, D. (2021). Embodying algorithms, enactive artificial intelligence and the extended cognition: You can see as much as you know about algorithm. Journal of Information Science, 49(1), 18. https://doi.org/10.1177/0165551520985495
91. Stappen, L., Baird, A., Lienhart, M., Bätz, A., & Schuller, B. W. (2022). An Estimation of Online Video User Engagement From Features of Time- and Value-Continuous, Dimensional Emotions. Frontiers in Computer Science, 4. https://doi.org/10.3389/fcomp.2022.773154
92. Stepanov, V. N. (2023). Emotion vs conflict-generating communication in a hybrid media environment. RUDN Journal of Studies in Literature and Journalism, 28(4), 800. https://doi.org/10.22363/2312-9220-2023-28-4-800-808
93. Strauss, C., Harr, M. D., & Pieper, T. M. (2024). Analyzing digital communication: a comprehensive literature review. Management Review Quarterly, 75(4), 3119. https://doi.org/10.1007/s11301-02400455-8
94. Tong, S. T. (2024). Foundations, Definitions, and Directions in Online Hate Research. In Routledge eBooks (p. 37). Informa. https://doi.org/10.4324/9781003472148-3
95. Vaishnav, S., & Wallace, D. (2022). Navigating Microaggressions in Online Learning Environments. Journal of Technology in Counselor Education and Supervision, 2(2). https://doi.org/10.22371/tces/0020
96. Vasist, P. N., Chatterjee, D., & Krishnan, S. (2023). The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-country Configural Narrative. Information Systems Frontiers, 26(2), 663. https://doi.org/10.1007/s10796-023-10390-w
97. Vasist, P. N., & Krishnan, S. (2022). ICT4D: development or destabilization? A cross-country study on the polarizing effect of political disinformation through social media. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-2099119/v1
98. Voinea, C., Marin, L., & Vică, C. (2024). Digital Slot Machines: Social Media Platforms as Attentional Scaffolds. Topoi, 43(3), 685. https://doi.org/10.1007/s11245-024-10031-0
99. Vombatkere, K., Mousavi, S., Zannettou, S., Roesner, F., & Gummadi, K. P. (2024). TikTok and the Art of Personalization: Investigating Exploration and Exploitation on Social Media Feeds. 3789. https://doi.org/10.1145/3589334.3645600
100. Vu, N. V., Nazari, M. A., Dang, T., Muralev, Y., Mohanraj, M., Tran, T., & Quoc, H. A. (2025). Type of the Paper: Article. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5384374
101. Wu, J. (2024). Social and ethical impact of emotional AI advancement: the rise of pseudo-intimacy relationships and challenges in human interactions. Frontiers in Psychology, 15, 1410462. https://doi.org/10.3389/fpsyg.2024.1410462
102. Wu, Y., Zhao, M., & Canny, J. (2025). Beyond Likes: How Normative Feedback Complements Engagement Signals on Social Media. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2505.09583
103. Yadin, D., Yahav, I., Zalmanson, L., & Munichor, N. (2023). Resolving the Ethical Tension Between Creating a Civil Environment and Facilitating Free Expression Online: Comment Reordering as an Alternative to Comment Moderation. Journal of Business Ethics, 192(2), 261. https://doi.org/10.1007/s10551-023-05450-9
104. Yo, P. W., Kee, D. M. H., Yu, J. W., Hu, M. K., Jong, Y. C., Ahmed, Z., Gwee, S. L., Gawade, O., & Nair, R. K. (2021). The Influencing Factors of Customer Satisfaction: A Case Study of Shopee in Malaysia. Studies of Applied Economics, 39(12). https://doi.org/10.25115/eea.v39i12.6839
105. Yu, M., & Riddle, K. (2022). An Experimental Test of the Effects of Digital Content Permanency on Perceived Anonymity and Indirect Effects on Cyber Bullying Intentions. Social media & Society, 8(1). https://doi.org/10.1177/20563051221087255
106. Yu, Z., Sen, I., Assenmacher, D., Samory, M., Fröhling, L., Dahn, C., Nozza, D., & Wagner, C. (2024). The Unseen Targets of Hate -- A Systematic Review of Hateful Communication Datasets. arXiv (Cornell University). Cornell University. https://doi.org/10.1177/08944393241258771

107. Zhang, J. (2025). The Emotional Economy in Social Media Platforms: An Analysis of Algorithm-Driven User Engagement and Commercial Value. Highlights in Art and Design, 12(2), 113. https://doi.org/10.54097/87aa6f65