

Political Identity as a Driver of Inadvertent Disinformation Sharing: Evidence From Thailand and Implications for Civic Participation¹

Sucheewa Sittijinda²

Duangkamol Chartprasert

Faculty of Communication Arts, Chulalongkorn University, Thailand

Abstract

Inadvertent disinformation sharing—the propagation of false political content by users who do not recognize it as such—has become a dominant vector of misinformation in contemporary elections, yet its psychological architecture remains underspecified in non-Western polarized democracies. This study examines how political identity, cognitive reflection, and six additional psychological-communication antecedents shape inadvertent disinformation sharing during Thailand’s 2026 general election, and whether misinformation discernment failure (MDF) mediates these effects. A cross-sectional online survey of 590 Thai social-media users (January–April 2026), recruited via Google Forms distributed through Thai online chat groups and communities, measured ten latent constructs across 59 seven-point Likert items. Confirmatory factor analysis supported the

¹ APA7th Citation: Sittijinda, S. & Chartprasert, C. (2026). Political Identity as a Driver of Inadvertent Disinformation Sharing: Evidence from Thailand and Implications for Civic Participation. In *Proceedings of the ICA Regional Hub Thailand 2026: Official Concurrent Event of the 76th Annual Conference of the International Communication Association (ICA)* (pp. 125–154). Faculty of Communication Arts, Chulalongkorn University.

This research article forms part of the requirements for the Doctor of Communication Arts Program, Faculty of Communication Arts, Chulalongkorn University.

² Correspondence concerning this article should be addressed to Sucheewa Sittijinda at the Faculty of Communication Arts, Chulalongkorn University, Bangkok, Thailand or by email at: sucheewayui@gmail.com

measurement model (CFI = .956, TLI = .953, RMSEA = .033; CR = .87–.96; AVE = .54–.82), and Harman’s test indicated no common-method bias (first factor 18.9%). A structural mediation model with eight antecedents predicting MDF, and MDF together with the antecedents predicting inadvertent sharing, explained 56% of variance in MDF and 66% in inadvertent sharing. All antecedent paths to MDF and the MDF–sharing path ($\beta = .26$) were significant at $p < .001$, with joint-significance tests confirming partial mediation. A balanced polarization–reflection dyad emerged: cognitive reflection produced the largest protective effect (total $\beta = -.35$), slightly exceeding political identity’s positive effect ($\beta = +.30$). Inadvertent sharing showed no significant demographic differences. The findings position MDF as a cognitive bottleneck and identify accuracy-evaluation interventions—calibrated to Thailand and the Mekong region—as high-leverage countermeasures for civic participation.

Keywords: political identity, misinformation discernment failure, inadvertent disinformation sharing, cognitive reflection, Thailand

The global expansion of social media has reshaped how political information moves through electorates and how disinformation reaches voters. A growing body of evidence shows that most users who circulate false or misleading content do so inadvertently rather than maliciously: they forward material that aligns with their identities, emotions, or social ties before they have evaluated its veracity (Pennycook et al., 2020; Vosoughi et al., 2018). In polarized electoral environments, this inadvertent diffusion can outpace deliberate disinformation campaigns and shape public opinion in ways that escape both fact-checking infrastructure and platform moderation.

Thailand's 2026 general election provides a useful setting to study these dynamics. The campaign unfolded across a digital ecosystem in which TikTok has overtaken Facebook as the most-used platform, LINE group chats remain a primary channel for political discussion among older voters, and the partisan divide between progressive and conservative blocs has hardened since 2020. Within this environment, inadvertent sharing is the predominant vector of false content, yet its psychological architecture remains underspecified, particularly in Southeast Asia.

Two limitations characterize prior research. First, antecedents of inadvertent sharing have typically been examined in isolation—single predictors such as social endorsement, altruism, or trust treated in separate studies (e.g., Apuke & Omar, 2021; Messing & Westwood, 2014)—rather than within an integrated model that places multiple psychological and communicative drivers alongside protective factors such as cognitive reflection (Pennycook & Rand, 2019). Second, the cognitive route through which these antecedents reach actual sharing behavior has rarely been modeled; prior work has documented direct associations with sharing intentions or behavior but has seldom specified the proximal evaluative process that intervenes between motivation and action (Pennycook et al., 2020, 2021; Pennycook & Rand, 2021). If antecedents such as political identity and cognitive reflection influence sharing partly by altering users' ability to evaluate the accuracy of political content—what we term misinformation discernment failure (MDF)—then interventions targeting evaluative accuracy may have leverage across multiple psychological pathways simultaneously.

The present study addresses these gaps in three ways. First, it estimates a ten-construct measurement model that includes eight theoretically grounded antecedents (altruism, information overload, social endorsement cues, phatic communication, self-promotion, cognitive reflection, trust in online information, and political identity), the proposed cognitive mediator (MDF), and

the behavioral outcome (inadvertent disinformation sharing, IFNSB). Second, it specifies a structural mediation model in which the eight antecedents predict MDF, and the eight antecedents together with MDF predict IFNSB, allowing the estimation of direct, indirect, and total effects. Third, it tests the model with a survey of 590 Thai social-media users fielded during the active campaign window (January–April 2026) and assesses whether the resulting effects are stable across demographic strata.

The contribution is threefold. Theoretically, the study reframes the polarization–reflection dyad—political identity as primary driver and cognitive reflection as primary protective factor—within a mediated process model in which a substantial portion of each effect travels through the user’s impaired or preserved ability to discern false content. Practically, the identification of MDF as a partial mediator of all eight antecedents indicates that interventions calibrated to restore accuracy evaluation (e.g., accuracy prompts, lateral-reading prompts, friction designs) may produce broad-spectrum reductions in inadvertent sharing even when the underlying psychological drivers cannot be addressed directly. Regionally, the Thai case extends partisan-polarization accounts of misinformation to a non-Western polarized democracy within the wider Mekong-region information environment.

Research Objectives and Hypotheses

The study pursues three objectives. First, to estimate a ten-construct measurement model integrating eight psychological-communication antecedents, the proposed cognitive mediator (MDF), and the behavioral outcome (IFNSB) for active Thai social-media users during the 2026 electoral campaign. Second, to test a structural mediation model in which the eight antecedents predict MDF, and the antecedents together with MDF predict IFNSB, estimating direct, indirect,

and total effects. Third, to assess whether the resulting effects are stable across six demographic strata (gender, age, education, income, daily social-media use, primary platform).

These objectives translate into three formal hypotheses, each derived from the theoretical framework developed in the literature review that follows:

H1: Each of the eight antecedents significantly predicts MDF. Specifically, altruism (H1a), information overload (H1b), social endorsement cues (H1c), phatic communication (H1d), self-promotion (H1e), trust in online information (H1f), and political identity (H1g) each positively predict MDF, while cognitive reflection (H1h) negatively predicts MDF.

H2: MDF positively predicts inadvertent disinformation sharing (IFNSB).

H3: MDF partially mediates the effect of each of the eight antecedents on IFNSB, such that each antecedent retains a significant direct effect on IFNSB alongside a significant indirect effect transmitted through MDF.

Literature Review and Theoretical Framework

Inadvertent Disinformation Sharing and the Information Environment

Inadvertent disinformation sharing (IFNSB) refers to the circulation of false, misleading, or unverified content by users who do not recognize it as such at the moment of sharing (Chadwick & Vaccari, 2019; Pennycook & Rand, 2021). In contrast to organized disinformation campaigns, inadvertent sharing is driven by everyday users responding to identity cues, emotional reactions, social norms, and cognitive shortcuts under conditions of information abundance. Experimental evidence indicates that the dominant pathway is inattention to accuracy rather than malicious intent: users routinely share content they would, on reflection, recognize as inaccurate (Pennycook et al.,

2020, 2021). The Thai context amplifies the salience of this construct because political content travels rapidly across TikTok feeds, Facebook timelines, and closed LINE groups—settings that combine algorithmic amplification with high social trust and limited fact-checking penetration.

Eight Psychological and Communicative Antecedents

Eight constructs have emerged in the literature as theoretically grounded antecedents of inadvertent sharing. Altruism (ALT) captures the prosocial motivation to inform one’s network of perceived threats or opportunities; users high in altruism share political content as a form of public service even when its accuracy is uncertain (Apuke & Omar, 2021). Information overload (IO) reflects the cognitive strain of processing high volumes of incoming political content, which encourages heuristic rather than analytic evaluation (Bawden & Robinson, 2009). Social endorsement cues (SEC) refer to the influence of visible engagement signals—likes, shares, comments—on perceived credibility (Messing & Westwood, 2014). Phatic communication (PHC) describes the relational rather than informational function of sharing, in which content is forwarded primarily to maintain social bonds (Miller, 2017); the closed-group structure of LINE in Thailand makes this pathway particularly salient. Self-promotion (SP) captures impression-management motives, in which users share to signal political sophistication or in-group identity. Trust in online information (TOI) reflects general credulity toward digital sources and weakens deliberate verification. Political identity (PI) operationalizes identity-protective cognition (Kahan, 2017): the tendency to accept and propagate content that affirms one’s partisan in-group while rejecting content that contradicts it. Partisan identification in Thailand does not map cleanly onto the binary structures that dominate much of the Western polarization literature; the construct is therefore operationalized in a way that accommodates the multi-party, multi-camp character of contemporary

Thai electoral politics. Finally, cognitive reflection (CR) is the dual-process protective factor: the disposition to override intuitive responses through deliberative evaluation, which has repeatedly predicted reduced belief in and sharing of false content (Pennycook & Rand, 2019).

Misinformation Discernment Failure as Cognitive Mediator

While each antecedent has documented direct associations with sharing, the cognitive route through which these influences operate has been less systematically theorized. We propose misinformation discernment failure (MDF) as the proximal mediator. MDF refers to the impaired ability to accurately evaluate the veracity of political content encountered on social media—the gap between the cognitive operations required for accurate verification and the operations a user actually performs at the moment of sharing. Conceptually, MDF sits at the convergence of three established literatures: the dual-process account of misinformation belief (Pennycook & Rand, 2019, 2021), which locates susceptibility in the dominance of System 1 over System 2 processing; the inattention-based account of sharing (Pennycook et al., 2020, 2021), which shows that absence of accuracy evaluation—not motivated rejection—is the dominant cause of false-content propagation; and the polarization literature on identity-protective cognition (Kahan, 2017), which identifies partisan motivation as a force that displaces accuracy evaluation.

MDF differs from adjacent constructs in important and specifiable ways. It is not *misinformation susceptibility*, which denotes a stable individual-difference trait in belief acceptance across contexts; it is not *media literacy*, which denotes a dispositional skill set in evaluating media sources and claims; it is not *accuracy discernment*, which is the outcome metric of correctly differentiating true from false items in forced-choice paradigms (Pennycook & Rand, 2019); it is not *perceived credibility*, which is a momentary judgment about a single piece of

content; and it is not *sharing intention*, which is an action-tendency that follows evaluation. MDF is the evaluative *state* itself—the moment-to-moment accuracy-assessment process—and its failure is what we propose mediates the eight antecedents’ effects on actual sharing behavior. This distinction matters because interventions calibrated to a trait (media literacy curricula), to a final judgment (perceived credibility manipulations), or to an outcome metric (accuracy-discernment training) target different points in the cognitive sequence than interventions calibrated to the evaluative state itself (accuracy prompts, friction designs, lateral-reading cues).

Intervention research directly motivates the mediator: the success of accuracy-prompt manipulations (Pennycook et al., 2021) shows that briefly engaging users in accuracy evaluation substantially reduces sharing of false content (Evans & Stanovich, 2013), implying that the absence of such evaluation—the state MDF names—is the active mechanism.

We expect the eight antecedents to predict MDF in theoretically coherent directions: altruism, information overload, social endorsement cues, phatic communication, self-promotion, trust in online information, and political identity should each elevate MDF, while cognitive reflection should reduce MDF. We further expect MDF to predict IFNSB positively, and we expect partial rather than full mediation: each antecedent should retain a direct effect on sharing alongside its indirect effect through MDF, reflecting the fact that psychological drivers shape behavior through both cognitive (MDF) and non-cognitive (e.g., social-bonding, identity-signaling) channels.

The 2026 Thai Electoral Context

Thailand’s 2026 election unfolded under conditions that heighten each pathway. Partisan polarization between progressive and conservative blocs has intensified since the 2020 youth-led

protests and the 2023 election. The information environment is multi-platform: TikTok dominates short-form political content among voters under 35; Facebook remains the principal long-form political space; and LINE group chats serve as the dominant channel for older voters' political discussion. Each platform combines high algorithmic amplification, high social trust within groups, and limited fact-checking penetration—conditions under which inadvertent sharing is expected to be the dominant disinformation vector. These contextual features motivate testing whether the proposed mediation structure replicates outside the predominantly Western samples in which dual-process accounts of misinformation have been developed.

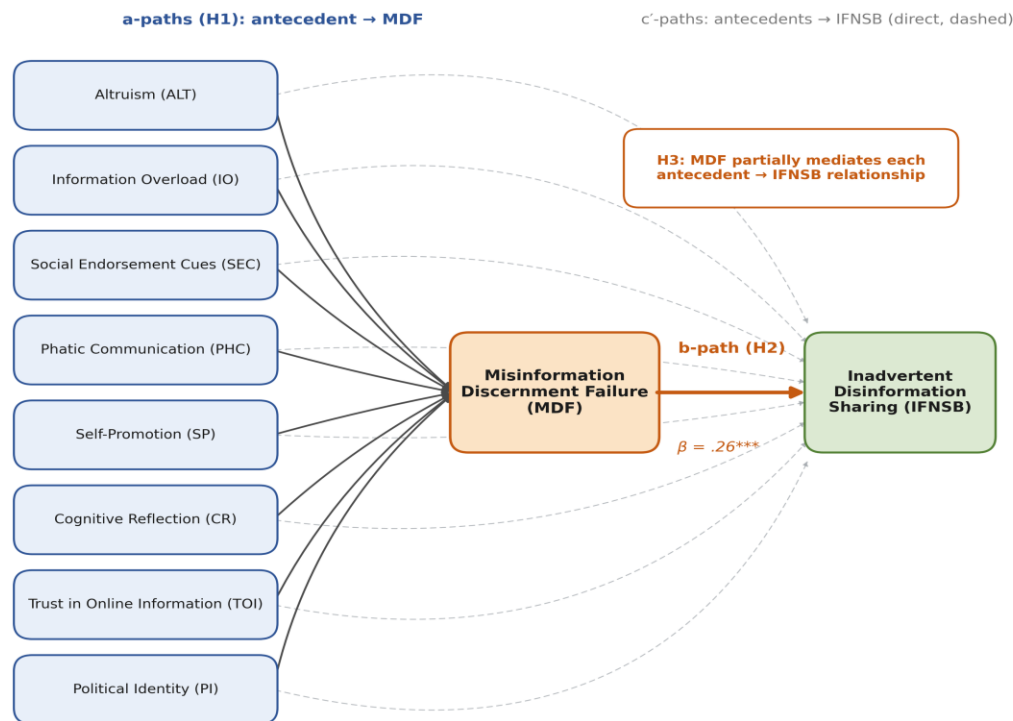
Comparative Context: Thailand Within the Mekong-Region Information Environment

Locating Thailand alongside its Southeast Asian comparators clarifies how far the findings travel. The Philippines combines personalistic patronage networks, Facebook dominance, and a documented record of state-aligned content operations (Ong & Cabañes, 2018). Indonesia's polarization runs along religious-nationalist and pluralist axes intersecting with regional and ethnic cleavages, and its electorate is heavily distributed across WhatsApp groups and TikTok—a closed-group dynamic similar to LINE's role in Thailand (Tapsell, 2021). Vietnam operates a state-controlled information environment with limited partisan pluralism on dominant platforms (Zalo, Facebook), shifting the relevant question from polarization-driven inadvertent sharing to state-curated content diffusion. The Thai case—a competitive multi-party polarized democracy spanning TikTok, Facebook, and LINE—sits between these. The portability of the polarization–reflection dyad and of MDF as a mediating cognitive bottleneck is therefore most plausibly tested in the Philippine and Indonesian cases, with Vietnam serving as a structurally different boundary condition.

Figure 1 summarizes the hypothesized structural relationships.

Figure 1

Hypothesized Structural Mediation Model of Eight Antecedents, Misinformation Discernment Failure (MDF), and Inadvertent Disinformation Sharing (IFNSB)



Note. ALT = altruism; IO = information overload; SEC = social endorsement cues; PHC = phatic communication; SP = self-promotion; CR = cognitive reflection; TOI = trust in online information; PI = political identity; MDF = misinformation discernment failure; IFNSB = inadvertent disinformation sharing behavior. Solid arrows denote the a-paths (antecedents → MDF, H1) and the b-path (MDF → IFNSB, H2). Dashed arrows denote the c'-paths (antecedents → IFNSB direct effects). H3 specifies partial mediation: each antecedent transmits effects to IFNSB through both the indirect route ($a \times b$) and the direct route (c').

Method

Research Design and Sampling Approach

A cross-sectional online survey was administered to 590 Thai social-media users between January and April 2026, spanning the active campaign window of Thailand's 2026 general

election. Given the study’s substantive focus on inadvertent disinformation sharing among active social-media users—the population in which the target behavior actually occurs—the sampling approach was deliberately framed as a community-distributed online survey rather than a probability sample of the general population. This approach prioritizes coverage of users who actively participate in the networked discussion spaces (chat groups, online communities, and social-media forums) in which political content is circulated, forwarded, and inadvertently propagated. The sampling design is therefore best characterized as a non-probability convenience sample of active Thai social-media users, with all generalizability claims framed accordingly.

Recruitment and Data Collection Procedure

The Google Forms instrument was distributed through online chat groups and social-media communities in Thailand. Researchers posted the link in topic-relevant discussion spaces, and respondents could forward it within their own networks, producing a network-driven dimension alongside the initial researcher-led distribution. The link was redistributed in waves across the January–April 2026 window to capture both early- and late-campaign respondents.

The Google Forms instrument was anonymous: no personally identifying information (name, email, phone number, IP address, or social-media handle) was collected or stored. The opening page presented an informed-consent statement specifying the study’s purpose, the voluntary nature of participation, the right to withdraw at any point without consequence, the anonymity of responses, and the contact information of the principal investigator. Respondents indicated consent by selecting an explicit consent option before the first item; those declining were routed out of the survey. Quality-control screens—including attention-check items and

consistency review—were applied prior to analysis, and cases failing these screens were excluded from the analytic sample.

Participant Eligibility

Inclusion criteria were: (a) age 18 years or above at the time of participation; (b) current residence in Thailand; (c) self-reported active use of at least one of the major social-media platforms (TikTok, Facebook, LINE, YouTube, Instagram, X/Twitter, or Threads); and (d) sufficient Thai-language reading proficiency to complete the questionnaire without assistance. Exclusion criteria were: (a) failure of quality-control screens; (b) incomplete responses on latent-construct items; and (c) responses identified as duplicate. The achieved analytic sample of $N = 590$ reflects respondents meeting all inclusion criteria and passing all screens, and exceeds the conventional 10:1 ratio of cases to free parameters and the 200-case minimum for SEM with continuous indicators, supporting stable parameter estimation. Because participation was self-initiated within networked online communities, the sample is best interpreted as representative of active, engaged Thai social-media users who participate in chat-group and online-community discussion spaces, with generalizability limits discussed further in the limitations section.

Sample Characteristics

The achieved sample comprised 253 male participants (42.9%), 322 female (54.6%), and 15 gender non-binary or unreported (2.5%). Age was distributed across five bands: 18–25 years ($n = 102$, 17.3%), 26–35 ($n = 125$, 21.2%), 36–45 ($n = 125$, 21.2%), 46–55 ($n = 117$, 19.8%), and 56 and above ($n = 121$, 20.5%), providing broad generational coverage. Educational attainment ranged from below secondary school ($n = 40$, 6.8%) to doctoral level ($n = 14$, 2.4%), with the modal category being bachelor’s degree ($n = 268$, 45.4%). Monthly household income ranged from

below 15,000 THB ($n = 184$, 31.2%) to above 100,001 THB ($n = 24$, 4.1%). Daily social-media use was less than one hour (5.8%), 1–2 hours (17.1%), 3–4 hours (29.2%), 5–6 hours (23.4%), and more than six hours (24.6%). The primary platform was TikTok (31.5%), Facebook (28.6%), LINE (21.2%), YouTube (8.6%), Instagram (6.4%), X/Twitter (2.5%), Threads (0.7%), and other (0.3%).

Measures

Ten latent constructs were measured across 59 items, all rated on 7-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). Eight antecedent constructs each used six items: altruism (ALT; V1–V6), information overload (IO; V7–V12), social endorsement cues (SEC; V13–V18), phatic communication (PHC; V19–V24), self-promotion (SP; V25–V30), cognitive reflection (CR; V31–V36), trust in online information (TOI; V37–V42), and political identity (PI; V43–V48). The PI scale was deliberately constructed for the Thai context, where a binary partisan dichotomy does not apply: items captured (a) self-placement on the progressive–conservative ideological dimension that has structured Thai electoral politics since 2020, (b) identification with one’s preferred political camp (the successor to the ‘red’–‘yellow’ color-coded camps), and (c) attitudes toward salient policy cleavages (constitutional reform, lèse-majesté provisions, decentralization). This multi-component operationalization captures identity-protective cognition without forcing the Thai electorate into the binary partisan structure assumed in much of the cited Western literature. Misinformation discernment failure (MDF; V49–V54) was measured by six items capturing the user’s self-reported difficulty in distinguishing accurate from inaccurate political content encountered during the campaign period. Inadvertent disinformation sharing behavior (IFNSB; V55–V59) was measured by five items capturing the frequency of sharing political content that was later found to be inaccurate or unverified. All scales were translated and

back-translated between English and Thai by bilingual researchers and pilot-tested with 30 respondents prior to fielding.

Representative items are provided to illustrate the operationalization of each construct (full item bank available from the corresponding author):

ALT (V3): “I share political information on social media to help my friends and family make informed decisions.”

IO (V9): “I find it difficult to keep up with all the political content I encounter on social media.”

SEC (V14): “When a political post has many likes, shares, or comments, I tend to view it as more credible.”

PHC (V20): “I forward political messages mainly to keep in touch with people in my group chats.”

SP (V26): “Sharing political content helps me express my views and identity to others.”

CR (V33): “Before sharing political content, I pause to consider whether it is actually accurate.” (reverse-keyed for MDF; positive for CR)

TOI (V39): “Most of the political information I encounter online is generally trustworthy.”

PI (V45): “I identify strongly with my preferred political camp, and I support content that reflects its views.”

MDF (V51): “I find it difficult to tell whether the political content I see online is accurate or inaccurate.”

IFNSB (V57): “I have shared political content on social media that I later discovered was inaccurate or unverified.”

Analytical Strategy

Analysis proceeded in four stages. First, descriptive statistics, Cronbach's alpha, and item-level skewness and kurtosis were computed to evaluate distributional properties and internal consistency. Second, a ten-construct confirmatory factor analysis (CFA) estimated standardized factor loadings, composite reliability (CR), and average variance extracted (AVE); the Fornell–Larcker (1981) criterion assessed discriminant validity. Third, common-method bias was evaluated through Harman's single-factor test and inspection of residual inter-item correlations after the CFA. Fourth, a structural mediation model was specified in which the eight antecedents predicted MDF, and the eight antecedents together with MDF predicted IFNSB; the eight antecedents were freely correlated. Standardized direct, indirect (via MDF), and total effects on IFNSB were computed, with the significance of indirect effects assessed through joint-significance tests (MacKinnon et al., 2002). Model estimation used maximum likelihood, and model fit was evaluated using χ^2/df , CFI, TLI, RMSEA, GFI, and AGFI against conventional thresholds. Finally, one-way ANOVAs tested whether IFNSB differed across six demographic strata.

Results

Descriptive Statistics and Internal Reliability

Item-level means ranged from 5.75 to 6.13 on the 7-point scale, indicating moderate-to-high agreement across constructs. Standard deviations ranged from 0.75 to 1.30. Skewness values (−1.04 to −0.20) and kurtosis values (−0.78 to +0.43) fell within conventional bounds for ML estimation ($|\text{skewness}| < 2$, $|\text{kurtosis}| < 7$; Curran et al., 1996). Cronbach's alpha for individual constructs ranged from $\alpha = .87$ (TOI, PI) to $\alpha = .96$ (IFNSB), all exceeding the conventional .80

threshold for established scales. Descriptive statistics, reliability indices, and measurement properties for all ten constructs are reported in Table 1.

Table 1

Descriptive Statistics, Reliability, and Measurement Properties for All Constructs (N = 590)

Construct	Items	<i>M</i>	<i>SD</i>	α	CR	AVE	Loadings
Altruism (ALT)	V1–V6	5.80	0.78	.92	.93	.67	.78–.84
Information Overload (IO)	V7–V12	5.90	0.70	.91	.91	.62	.76–.80
Construct	Items	<i>M</i>	<i>SD</i>	α	CR	AVE	Loadings
Social Endorsement Cues (SEC)	V13–V18	6.01	0.73	.92	.92	.65	.77–.83
Phatic Communication (PHC)	V19–V24	6.03	0.64	.89	.89	.58	.74–.79
Self-Promotion (SP)	V25–V30	6.06	0.65	.90	.90	.60	.76–.79
Cognitive Reflection (CR)	V31–V36	6.08	0.68	.91	.91	.63	.76–.83
Trust in Online Info (TOI)	V37–V42	5.88	0.67	.87	.87	.54	.67–.75
Political Identity (PI)	V43–V48	5.86	0.67	.87	.87	.54	.67–.77
MDF	V49–V54	5.98	0.97	.95	.95	.75	.85–.88
IFNSB	V55–V59	5.87	1.18	.96	.96	.82	.90–.91

Note. *N* = 590. *M* = mean; *SD* = standard deviation; α = Cronbach's alpha; CR = composite reliability; AVE = average variance extracted; Loadings = range of standardized factor loadings from confirmatory factor analysis; MDF = misinformation discernment failure; IFNSB = inadvertent disinformation sharing behavior. All items measured on 7-point Likert scales (1 = *strongly disagree* to 7 = *strongly agree*). All α and CR $\geq$.87, and all AVE $\geq$.54, exceeding conventional thresholds. Discriminant validity was confirmed for all construct pairs via the Fornell–Larcker (1981) criterion.

Confirmatory Factor Analysis

The ten-construct measurement model demonstrated good fit: $\chi^2(1607) = 2658.14, p < .001$, $\chi^2/df = 1.65$, CFI = .956, TLI = .953, RMSEA = .033, GFI = .896, AGFI = .890. Standardized factor loadings ranged from $\lambda = 0.67$ to $\lambda = 0.91$, all exceeding the 0.50 convergent-validity threshold. Composite reliability ranged from .87 to .96. Average variance extracted ranged from .54 (TOI, PI) to .82 (IFNSB), all exceeding the conventional .50 threshold. Discriminant validity was likewise confirmed for all construct pairs using the Fornell–Larcker (1981) criterion: the square root of AVE for each construct exceeded its absolute correlation with every other construct in the model.

Common-Method Bias Assessment

Because all constructs were measured on Likert scales within a single cross-sectional survey, common-method bias was assessed through three complementary procedures (Podsakoff et al., 2003). First, Harman’s single-factor test, conducted via unrotated principal-components analysis on all 59 items, extracted a first factor accounting for 18.9% of total variance, well below the 50% threshold that indicates substantive common-method contamination. Second, inspection of residual inter-item correlations after the ten-factor CFA showed a mean residual of approximately .00 (mean absolute residual = .14), indicating that the measurement model adequately accounted for shared variance across items and leaving no systematic common-method factor unexplained. Third, a marker-variable check was conducted using cognitive reflection (CR), which is theoretically orthogonal to the substantive sharing-related constructs, as a proxy marker. The partial correlations among the substantive constructs after adjusting for CR were within $\pm .02$ of the zero-order correlations, indicating that any common-method variance is small relative to substantive associations. Procedurally, anonymity and response-randomization safeguards (Podsakoff et al., 2003) were also implemented at

the survey-instrument stage. Taken together, these results support the position that observed structural relationships are not artifacts of shared method.

Structural Mediation Model

The structural mediation model specified paths from the eight antecedents to MDF, and from the eight antecedents together with MDF to IFNSB; the eight antecedents were allowed to correlate freely. Because all eight antecedent intercorrelations were freely estimated and no structural paths were constrained to zero, the structural model uses the full set of inter-construct covariances available in the measurement model and is therefore statistically equivalent to it; the fit indices reported below are consequently identical to those of the ten-construct CFA. The model demonstrated good fit: $\chi^2(1607) = 2658.14, p < .001, \chi^2/df = 1.65, CFI = .956, TLI = .953, RMSEA = .033, GFI = .896, AGFI = .890$. The model explained 56% of variance in MDF ($R^2 = .56$) and 66% of variance in IFNSB ($R^2 = .66$). All eight antecedent paths to MDF were statistically significant at $p < .001$, with standardized coefficients of $\beta = .23$ (ALT), $.23$ (IO), $.23$ (SEC), $.25$ (PHC), $.25$ (SP), $-.29$ (CR), $.26$ (TOI), and $.25$ (PI). The MDF–IFNSB path was significant ($\beta = .26, p < .001$). Hypothesis 1 (H1a–H1h) was therefore supported for all eight antecedent–MDF paths, and Hypothesis 2 was supported for the MDF–IFNSB path. Standardized direct, indirect (via MDF), and total effects on IFNSB are reported in Table 2, ordered by absolute magnitude of the total effect.

Table 2

Standardized Direct, Indirect, and Total Effects on Inadvertent Disinformation Sharing (N = 590, January–April 2026)

Antecedent	Direct (β)	Indirect (β via MDF)	Total (β)	Direction	Significance
Cognitive Reflection (CR)	−0.27	−0.08	−0.35	−	***
Political Identity (PI)	+0.24	+0.06	+0.30	+	***
Altruism (ALT)	+0.21	+0.06	+0.27	+	***
Self-Promotion (SP)	+0.19	+0.07	+0.26	+	***
Phatic Communication (PHC)	+0.19	+0.06	+0.25	+	***
Information Overload (IO)	+0.17	+0.06	+0.23	+	***
Trust in Online Info (TOI)	+0.16	+0.07	+0.23	+	***
Social Endorsement Cues (SEC)	+0.15	+0.06	+0.21	+	***
MDF → IFNSB (b-path)	+0.26	—	+0.26	+	***

Note. N = 590. β = standardized path coefficient; MDF = misinformation discernment failure; IFNSB = inadvertent disinformation sharing behavior. All coefficients estimated from the full structural model in which the eight antecedents predict MDF, and the eight antecedents together with MDF predict IFNSB. Indirect effects computed as the product of the antecedent → MDF path and the MDF → IFNSB path ($\beta = .26$); because component path coefficients are rounded to two decimals in the Results text, re-multiplying them may not exactly reproduce the indirect effects shown. Total effect = direct + indirect. Significance of indirect effects assessed via joint-significance tests (MacKinnon et al., 2002): both component paths significant at $p < .001$ for all eight antecedents, indicating that MDF partially mediates each antecedent's effect on inadvertent sharing (supporting H3).

*** $p < .001$.

Two findings from the mediation results warrant particular attention. First, the polarization–reflection dyad emerged clearly in the total effects: political identity registered the strongest positive total effect on inadvertent sharing (total $\beta = .30$), and cognitive reflection

produced the strongest protective total effect (total $\beta = -.35$), comparable in absolute magnitude. The protective coefficient slightly exceeds the driving coefficient, indicating that—at least in this Thai electoral sample—deliberative cognitive resources can match or exceed the polarizing pull of partisan identity. Second, all eight indirect effects through MDF were significant by joint-significance criteria, with indirect effects ranging from $\beta = .06$ (ALT, IO, SEC, PI) to $|\beta| = .08$ (CR). The proportion of total effect transmitted through MDF varied from approximately 21% (ALT, PI) to 30% (TOI, SEC), indicating partial rather than full mediation and supporting H3. Substantively, each psychological or communicative antecedent reaches inadvertent sharing through two channels: a non-cognitive channel (e.g., social-bonding, identity-signaling, heuristic processing) that operates as a direct effect, and a cognitive channel through which the antecedent impairs (or, in the case of CR, preserves) the user's ability to discern the veracity of the content at hand.

Among the remaining six antecedents, altruism (total $\beta = .27$) and self-promotion (total $\beta = .26$) produced the largest positive total effects after political identity, indicating that prosocial motivation and identity-signaling are substantial drivers of inadvertent sharing in the Thai electoral context. Phatic communication (total $\beta = .25$) followed closely; this finding is particularly salient given the role of LINE group chats in Thai political communication, where forwarding messages functions as social bonding rather than information transmission. Information overload, trust in online information, and social endorsement cues produced moderate total effects ($\beta = .21-.23$). All eight total effects were statistically significant at $p < .001$.

Demographic Group Comparisons

One-way ANOVAs assessed whether IFNSB differed across six demographic strata. No comparison reached significance at $p < .05$: gender, $F(2, 587) = 0.27$, $p = .766$, $\eta^2 = .001$; age

group, $F(4, 585) = 1.89, p = .111, \eta^2 = .013$; educational attainment, $F(5, 584) = 1.46, p = .202, \eta^2 = .012$; monthly household income, $F(5, 584) = 0.73, p = .599, \eta^2 = .006$; daily social-media use, $F(4, 585) = 1.95, p = .100, \eta^2 = .013$; and primary platform, $F(7, 582) = 0.96, p = .461, \eta^2 = .011$. All effect sizes were negligible (all $\eta^2 \leq .013$), at or below the .01 lower boundary of small effects under Cohen's (1988) conventions. A minimum-detectable-effect analysis at $N = 590, \alpha = .05$, and 80% power indicates a minimum detectable η^2 of approximately .024 for the largest ANOVA design (eight categories for primary platform); the observed η^2 values fall well below this threshold, supporting the interpretation that the null findings reflect substantively small differences rather than statistical under-power. The pattern is consistent with—though does not by itself establish—demographic neutrality of inadvertent sharing in the Thai context, and warrants cross-national replication across other Southeast Asian polarized democracies before being treated as established. The provisional implication is that psychological and communicative variables, rather than demographic characteristics, structure inadvertent sharing in this sample, justifying the omission of demographic moderators from the structural model.

Discussion

The mediation analysis yields three findings of theoretical and practical consequence. First, the data support a balanced polarization–reflection dyad rather than a political-identity-first account of inadvertent sharing. Cognitive reflection produced the largest protective total effect ($\beta = -.35$), comparable to and slightly exceeding political identity's positive total effect ($\beta = +.30$). The headline reading is therefore 'polarization drives, reflection protects,' with the two forces in approximate balance and the reflective pole marginally dominant. Substantively, this confirms

identity-protective cognition (Kahan, 2017) as a major engine of false-content diffusion in Thailand's 2026 election while simultaneously extending dual-process accounts of misinformation belief (Pennycook & Rand, 2019) to inadvertent sharing behavior in a Southeast Asian electoral context. That cognitive reflection's protective coefficient slightly exceeds political identity's driving coefficient is more optimistic than identity-protective accounts have typically offered: deliberation, at least in this Thai sample, can match—and modestly outpace—the pull of partisan identity.

Second, MDF emerges as a partial mediator of all eight antecedents, transmitting between roughly one-fifth and one-third of each antecedent's total effect. Theoretically, this shows that psychological and communicative drivers of inadvertent sharing reach behavior through both cognitive and non-cognitive channels: a portion of each effect travels through impaired or preserved discernment, while a substantial residual portion operates through direct routes such as social bonding, identity signaling, prosocial motivation, and heuristic processing under load. The partial-mediation pattern is theoretically more informative than a fully mediated specification would be, because it indicates that interventions targeting cognitive evaluation alone are necessary but not sufficient. Practically, the centrality of MDF across all eight pathways identifies it as a high-leverage intervention target: any procedure that restores accuracy evaluation at the moment of sharing—accuracy prompts, lateral-reading cues, friction designs, source-credibility nudges—should reduce inadvertent sharing across multiple psychological drivers simultaneously, even when those drivers themselves cannot be addressed.

Third, the absence of demographic differentiation in inadvertent sharing has provisional implications for intervention design. Because gender, age, education, income, daily use, and

primary platform produced negligible group differences in IFNSB (all $\eta^2 \leq .013$, well below the minimum detectable η^2 of $\approx .024$), the present data are consistent with the hypothesis that inadvertent sharing in Thailand is structured by psychological rather than demographic variables. If this pattern replicates cross-nationally, it would support broadly universal—rather than demographically segmented—intervention designs; if it does not, demographically tailored designs would become essential. We frame this finding as a hypothesis warranting comparative replication rather than as established demographic neutrality.

Contributions Beyond the Thai Context

Beyond its specific Thai application, the study makes three contributions to the broader misinformation and political communication literature. First, the eight-antecedent mediation model addresses a recurring gap in the field: most prior research has examined single antecedents of inadvertent sharing in isolation (e.g., social endorsement in Messing & Westwood, 2014; altruism in Apuke & Omar, 2021; cognitive reflection in Pennycook & Rand, 2019), leaving the relative magnitudes of multiple drivers and their joint cognitive route unaddressed. By estimating direct, indirect, and total effects within a single mediation framework, the study provides a template for cross-construct comparison that other national contexts can replicate. Second, the MDF construct itself contributes to the conceptual vocabulary of the misinformation literature by naming the moment-to-moment evaluative state that the dual-process, inattention, and identity-protective traditions all imply but rarely operationalize. Distinguishing MDF from misinformation susceptibility (trait), media literacy (skill), accuracy discernment (outcome metric), perceived credibility (single-content judgment), and sharing intention (action-tendency) clarifies which point in the cognitive sequence different interventions target. Third, the polarization–reflection dyad

finding—cognitive reflection’s protective effect marginally exceeding political identity’s driving effect—offers a slightly more optimistic empirical anchor than identity-protective accounts have typically supplied. If this balance holds in comparator settings, then deliberation-supporting interventions are a credible policy lever in polarized democracies generally, not only in Thailand. Cross-national replication in the Philippines, Indonesia, and beyond Southeast Asia will determine the generality of these contributions.

Intervention Implications

Within the framing developed above, four concrete intervention paths follow. (a) Platform-level accuracy nudges adapted to Thailand’s three dominant channels: a TikTok ‘pause-before-share’ prompt before forwarding political content, Facebook newsfeed friction calibrated for long-form posts, and a LINE group-chat soft-warning when forwarded political messages exceed a within-group share threshold. (b) Secondary and tertiary digital-literacy curricula focused not on detection-as-skill but on the cognitive habit of accuracy evaluation, drawing on lateral-reading exercises (Wineburg & McGrew, 2019) and accuracy-prompt research (Pennycook et al., 2021) adapted to Thai examples. (c) Civic-society fact-checking partnerships, building on the Thai Fact-Check Center and Cofact ecosystem, that prioritize MDF-targeted interventions over post-hoc debunking. (d) Mekong-region inter-governmental literacy frameworks: ASEAN-aligned coordination among Thailand, the Philippines, Indonesia, and Cambodia sharing intervention designs and election-period response procedures. Translating this into the comparator polities, MDF-targeted intervention implies accuracy prompts on Facebook in the Philippines, group-chat friction designs in Indonesia (where WhatsApp/TikTok dynamics most closely resemble Thailand’s LINE ecosystem), and re-theorization of MDF for Vietnam’s state-curated

environment—a graded portability pattern that should structure any Mekong-region coordination effort.

Limitations and Future Research

Five boundary conditions warrant explicit acknowledgment. First, the cross-sectional design supports a mediational specification consistent with theory but does not establish temporal precedence; longitudinal or experimental designs (e.g., randomized accuracy-prompt interventions) are needed to confirm the causal status of MDF as a mediator. Second, the use of self-report measures introduces the possibility of socially desirable responding in the IFNSB items, although the high reported levels of sharing argue against substantial under-reporting. Third, the convenience-sampling design—distributing a Google Forms questionnaire through Thai chat groups and online communities with onward sharing by respondents—generalizes to active Thai social-media users in these networked discussion spaces and does not constitute a probability sample of the Thai population. Non-users, occasional users outside these networks, and users without smartphone or stable-internet access are systematically under-represented; findings should not be extended uncritically to these segments. Fourth, the use of self-reported behavioral measures (rather than behavioral-trace data such as platform-logged shares) limits inferences about actual sharing magnitudes. Fifth, the demographic-neutrality result rests on six non-significant tests in a single convenience sample; cross-national replication with probability-based recruitment where feasible is needed before it can support universal-intervention claims.

Future research should extend the present model along four lines. First, replication in the 2027 Indonesian, 2028 Philippine, and 2029 Thai election cycles would test whether the polarization–reflection dyad and the MDF mediation structure are stable across electoral periods

and across Southeast Asian comparator polities. Second, multi-group SEM by primary platform would test whether the magnitude (rather than the presence) of mediation differs across TikTok, Facebook, and LINE environments. Third, experimental tests of accuracy-prompt interventions in the Thai context would directly evaluate the causal status of MDF as a manipulable mediator. Fourth, the inclusion of behavioral-trace measures (actual shares logged via platform APIs or device telemetry) would address the self-report limitation directly.

Conclusion

This study examined the psychological and communicative architecture of inadvertent disinformation sharing among Thai social-media users during a polarized national election, with particular attention to whether misinformation discernment failure mediates the effects of eight antecedents on sharing behavior. Three principal contributions emerge.

Theoretically, the study reframes the polarization–reflection dyad—political identity as principal driver and cognitive reflection as principal protective factor—within a mediated process model in which a substantial portion of each effect travels through users’ impaired or preserved ability to discern false content. By integrating eight psychological and communicative antecedents within a single structural mediation framework, the study moves beyond fragmented single-predictor accounts that have characterized much prior research, offering a more integrative theoretical foundation for studying inadvertent disinformation sharing in polarized electoral contexts.

Conceptually, the study contributes the construct of misinformation discernment failure (MDF) as a theoretically distinct cognitive bottleneck. By naming the situational evaluative state

that the dual-process, inattention-based, and identity-protective traditions all imply but rarely operationalize, MDF provides a unifying proximal mediator amenable to systematic intervention design. The construct is distinguished from misinformation susceptibility, media literacy, accuracy discernment, perceived credibility, and sharing intention, clarifying which point in the cognitive sequence interventions can target.

Practically, the centrality of MDF across all eight antecedent pathways identifies accuracy-evaluation interventions—platform-level nudges adapted to TikTok, Facebook, and LINE; digital-literacy curricula focused on accuracy-evaluation habits as cognitive practice rather than detection-as-skill; and Mekong-region coordination linking Thailand to its Philippine, Indonesian, and Cambodian comparators—as high-leverage countermeasures. These interventions are particularly significant because they can produce broad-spectrum effects across multiple psychological drivers simultaneously, even when those drivers themselves cannot be addressed directly.

Regionally and globally, the Thai case extends partisan-polarization accounts of misinformation to a non-Western polarized democracy and offers a template for comparative investigation across Southeast Asian electoral contexts and beyond. The implications for civic participation are direct: equipping citizens to discern political information accurately is a precondition for informed electoral engagement, and the polarization–reflection dyad offers a tractable framework for civic-education and platform-governance design that respects both the realities of polarized identity politics and the empirical evidence that deliberation can meaningfully counter them. Future research is invited to test the model longitudinally, replicate it cross-nationally, and evaluate accuracy-prompt interventions experimentally to establish the causal status of MDF as a manipulable target for sustaining democratic participation in a digital age.

Ethics Statement

This research was conducted in accordance with internationally recognized ethical principles for research involving human participants. Because the study involved an anonymous, minimal-risk online survey that collected no personally identifying information from consenting adult participants, it did not require review by a formal research-ethics committee under the host institution's policy governing human-subjects research. Participation was voluntary, and all participants provided informed consent electronically before completing the questionnaire. The consent procedure disclosed the study's purpose, the voluntary nature of participation, the right to withdraw at any point without consequence, and the anonymity of responses. All participants were 18 years of age or older. No personally identifying information (name, email address, phone number, IP address, or social-media handle) was collected, stored, or analyzed; data are reported only in aggregate form and individual responses cannot be linked to identifiable persons. No vulnerable populations were specifically targeted in this research, and no incentive payments were offered that could constitute undue inducement. De-identified, aggregated data are available from the corresponding author upon reasonable request.

AI Disclosure Statement

During the preparation of this work, the author(s) used Claude Opus 4.7 (developed by Anthropic) to improve the readability and language editing of the manuscript. After using this tool, the author(s) thoroughly reviewed, verified, and edited the content as needed, and take(s) full responsibility for the final text of the publication.

References

- Apuke, O. D., & Omar, B. (2021). Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 56, 101475. <https://doi.org/10.1016/j.tele.2020.101475>
- Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191. <https://doi.org/10.1177/0165551508095781>
- Chadwick, A., & Vaccari, C. (2019). *News sharing on UK social media: Misinformation, disinformation, and correction*. Loughborough University Online Civic Culture Centre.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.1177/002224378101800104>
- Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2973067>
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83–104. <https://doi.org/10.1037/1082-989X.7.1.83>
- Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042–1063. <https://doi.org/10.1177/0093650212466406>
- Miller, V. (2017). Phatic culture and the status quo: Reconsidering the purpose of social media activism. *Convergence*, 23(3), 251–269. <https://doi.org/10.1177/1354856515592512>
- Ong, J. C., & Cabañes, J. V. A. (2018). *Architects of networked disinformation: Behind the scenes of troll accounts and fake news production in the Philippines*. The Newton Tech4Dev Network.

- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7), 770–780. <https://doi.org/10.1177/0956797620939054>
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Tapsell, R. (2021). Social media and elections in Southeast Asia: The emergence of subversive, underground campaigning. *Asian Studies Review*, 45(1), 117–134. <https://doi.org/10.1080/10357823.2020.1841093>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record*, 121(11), 1–40. <https://doi.org/10.1177/016146811912101102>