Latent Identity, Contained Conflict: Cognitive Security In Jakarta's 2017 and 2024 Elections

Septiar Dwi Putra[1*], Pujo Widodo[2], Yuniar[3], Nefra Firdaus[4], Harri Dolli Hutabarat[5]

[1,2,3,4,5]Faculty of National Security, Republic of Indonesia Defense University, IPSC Sentul Area, Sukahati, Citeureup, Bogor, West Java 16810, Indonesia
*Corresponding author:
Email: septiar.putra@kn.idu.ac.id

Abstract.
Indonesia treats digital disinformation as a national security threat, yet election-risk maps list hoaxes without explaining how they become offline conflict. This article compares two Jakarta gubernatorial elections that shared similar information and identity vulnerabilities but ended differently: the 2017 contest saw mass mobilisation and street violence, whereas no comparable mobilisation was identified in 2024. Similar vulnerability does not imply identical political triggers, and the comparison treats that difference as a rival explanation. The article develops a cognitive security framework in which information threats pass through perception shift, collective mobilisation, and clash, with social vulnerability as an amplifier and collective verification as a possible dampener. Using a most-similar comparative case design, it applies theory-testing process tracing to sixteen public documents, comprising survey releases, election-body statements, news reports, and scholarly studies. Identity considerations remained salient in 2024 and internet penetration in Jakarta reached 87.51 percent, yet surveys using different instruments indicate weaker religious alignment of vote choice than in 2017. The election supervisory body named hoaxes and identity-based campaigning as risks and engaged religious and community leaders before the campaign, but the documents cannot separate this pre-emptive verification from the absence of a comparable trigger, nor fix the exact point at which the escalation chain stopped. The article reframes cognitive security as a community and institutional capacity and proposes stage-specific early warning indicators for election and security agencies.
Keywords: *Cognitive Security; Collective Verification; Disinformation; Early Warning and Electoral Conflict.*

I. INTRODUCTION

On 4 November 2016, a march of about 100,000 people protesting alleged blasphemy by Jakarta's governor, Basuki Tjahaja Purnama (Ahok), turned violent after dark. Stores were looted in North Jakarta, twelve people including eight police officers were injured, and an elderly man died of causes that remained unclear (Asmara, 2016). The mobilisation followed weeks of viral circulation of a clip of the governor's remarks and became the defining episode of an election widely read as a turning point in Indonesian identity politics (Hadiz, 2017; Lim, 2017; Mietzner & Muhtadi, 2018). Eight years later, the same city entered another gubernatorial election with a similar warning. On 1 August 2024, the Jakarta Provincial Election Supervisory Body (Badan Pengawas Pemilihan Umum, Bawaslu DKI) rated hoax material on social media and identity-based campaigning on social media, known as SARA (an Indonesian acronym for ethnicity, religion, race, and inter-group relations), as high-vulnerability issues (Bisnis.com, 2024a). Yet the 2024 contest passed without comparable mobilisation, and the runner-up filed no dispute before the Constitutional Court deadline (Adji, 2024). Similar information and identity vulnerabilities in the same city were followed by different escalation outcomes, even though the political triggers were not the same. This puzzle is the starting point of this article.

Existing research establishes that online content can propagate offline harm. Anti-refugee sentiment on Facebook predicted hate crimes in Germany (Müller & Schwarz, 2021), exposure to misinformation raised support for vigilante violence in India and Pakistan (Badrinathan et al., 2025), and social media disinformation is associated with domestic terrorism across countries, partly through polarisation (Piazza, 2022). The same literature also warns that misinformation rarely causes violence on its own; its effect depends on grievance, trust, and emotion (Kurten, 2025; Lewandowsky et al., 2013), and claims of online-to-offline spillover often fail when tested directly (Leverso et al., 2025; Stuart, 2020). In Indonesia, scholars have described algorithmic enclaves and tribal nationalism (Lim, 2017), the securitisation of hoaxes as an existential threat (Lee, 2020), and election-risk mapping as a form of early warning (Nivada & Rassanjani, 2024).

Three gaps follow. First, election-risk mapping lists threats but rarely specifies indicators, thresholds, or the stage at which a threat becomes dangerous (Nivada & Rassanjani, 2024), while computational forecasting predicts persistent hot spots better than new onsets or escalation (Bazzi et al., 2022). Second,

digital literacy is usually tested as an individual's ability to discern false news (Guess et al., 2020; Badrinathan, 2021; Huang et al., 2024), not as a factor that interrupts mobilisation or escalation. Third, cognitive security is framed mostly at the level of artificial intelligence and cyber defence, the individual mind, or strategic and military competition (Casino, 2025; Deppe & Schaal, 2024; Grahn & Taipalus, 2025), and seldom as a capacity of communities and local institutions. Indonesia's legal framework leaves this gap open rather than closing it: Law Number 7 of 2012 on Social Conflict Management (UU 7/2012) requires government to build early warning through communication media, and the Elucidation of Article 10(3) describes these media as covering traditional communication media and print and electronic mass media. The law does not exclude digital platforms, but it gives no guidance on how social media and messaging applications should be monitored, by whom, and under which safeguards.

This article therefore asks two questions. RQ1: Under what conditions, and through which sequence, did digital information threats translate into perception shifts, collective mobilisation, and escalation in Jakarta in 2017 but not in 2024? RQ2: Did collective verification capacity interrupt this sequence, and at which stage? The answers inform a policy question addressed in the Discussion, namely which combination of digital and social indicators an early warning framework should track within Indonesia's national security governance.

The study moves from listing election risks to tracing the mechanism stage by stage, and it treats digital literacy as a conditional and collective protective factor that can fail rather than as an automatic remedy. Theoretically, this study develops cognitive security by incorporating a community-level escalation chain and collective verification. Empirically, it compares two Jakarta elections with similar information and identity vulnerabilities but divergent outcomes, a policy context not yet examined with process tracing. Practically, it derives a stage-specific early warning framework that links the digital indicators already monitored by election bodies to social indicators and pre-agreed responses.

The remainder of this introduction reviews the literature on disinformation, electoral mobilisation, literacy, and early warning, and develops the theoretical framework. Section II outlines the method, Section III presents and discusses the results, and Section IV concludes.

Disinformation and Offline Conflict

The causal evidence linking online content to offline harm is stronger than a decade ago. Müller and Schwarz (2021) show that anti-refugee posts on Facebook predicted hate crimes, and Badrinathan et al. (2025) find experimentally that misinformation increases support for vigilantism. At the level of diffusion, false news spreads farther and faster than true news, driven by people rather than bots (Vosoughi et al., 2018), and much sharing of falsehoods reflects inattention rather than conviction (Pennycook & Rand, 2021). Recent work also treats perception itself as a measurable signal, for example a violence perception index built from YouTube comments (Amarasinghe et al., 2026) or hostile discourse as a predictor of organised collective violence (Mendieta et al., 2026).

Other studies counsel caution. Online-to-offline spillover was not significant in a study of Chicago gang violence (Leverso et al., 2025), and overstating the online cause risks criminalising expression (Stuart, 2020). Adolescents describe how platform features escalate conflicts quickly (Elsaesser et al., 2021), but the step from online hostility to street violence still depends on offline conditions. The link is therefore conditional, and the useful question is under which conditions and through which stages it holds.

Electoral Conflict and Mobilisation

Electoral violence is a distinct form of political violence tied to the stakes and timing of electoral contests (Birch et al., 2020). Indonesian elections since 2014 have combined polarisation, identity politics, and hoax-centred campaigns (Lim, 2017; Lee, 2020; Yang Hui, 2020), with much campaigning moving into subversive, underground channels on social media (Tapsell, 2021). The 2016 Islamist mobilisation against Ahok drew on religious intolerance, militant groups, and elite accommodation (Mietzner & Muhtadi, 2018), and it signalled a broader turn toward illiberalism (Hadiz, 2017).

Mobilisation requires framing that converts grievance into collective action (Benford & Snow, 2000). Rumour adoption in conflict zones is driven by worldview, threat perception, and prior exposure rather than by education, income, age, or gender (Greenhill & Oppenheim, 2017), and rumours spread most when leaders hold extreme positions and rivals escalate (Bhavnani et al., 2009). Social structure conditions these

dynamics. Religious and ideological orientation shapes political attitudes in Indonesia (Fossati, 2019), and interethnic civic networks help manage tension, whereas intra-group networks alone raise the risk of violence (Varshney, 2001). Everyday Muslim and Christian coexistence in post-conflict North Maluku shows that such cross-group relations are continually renegotiated under new pressures (Botara, 2022). Large-scale violence in post-Suharto Indonesia subsided largely through state internal security interventions (Barron et al., 2016), a reminder that policing is a rival explanation for any case of containment.

Digital Literacy and Collective Verification

Individual literacy interventions show real but fragile effects. A digital media literacy intervention improved discernment between mainstream and false news in the United States and India (Guess et al., 2020), while an in-person educational intervention in India did not significantly improve respondents' ability to identify misinformation (Badrinathan, 2021). Psychological inoculation improves resilience on social media (Roozenbeek et al., 2022), and a meta-analysis reports a moderate overall effect of media literacy interventions but a weaker effect on belief (Huang et al., 2024). These studies measure discernment, not whether literacy interrupts mobilisation.

A second strand treats verification as a collective practice. Fact-checks reduce false beliefs modestly, and less for campaign statements (Walter et al., 2020), yet they worked across four countries with effects lasting beyond two weeks (Porter & Wood, 2021). Community actors can perform verification in real time, as the Filterinfo Facebook group did during the 2011 Ambon tensions (Rohman et al., 2020), and a community-centred model of response to information harms is now advocated at the global level (Wardle & Scales, 2025). Closed messaging groups matter here as well: a study of an Indonesian alumni WhatsApp group shows how circulating hoaxes shape deliberation within such networks (Soelistyowati et al., 2024). Resilience to disinformation also varies with national media and political environments (Humprecht et al., 2020). In Indonesia, weak hoax literacy has been tied to lower personal resilience (Subarjo & Setianingsih, 2020). Literacy has thus been tested on discernment rather than on escalation, and its collective form remains under-theorised.

Early Warning from Risk Maps to Response

Conflict early warning has advanced from structural forecasting to text and media data. Systems such as ViEWS forecast political violence at fine resolution (Hegre et al., 2019), newspaper text improves prediction (Mueller & Rauh, 2018), and comparative reviews map the strengths of existing systems (Rød et al., 2024). Limits persist. Structural models miss new onsets (Bazzi et al., 2022), social media data skew toward capitals and wealthier areas (Dowd et al., 2020), and atrocity-prevention organisations struggle to turn data into analysis and action (Palihapitiya et al., 2025). The older warning-response problem, in which warnings fail to trigger timely action, remains central (George & Holl, 1997; Wählisch & Kufus, 2026). Indonesian studies map election risks without stage-specific indicators (Nivada & Rassanjani, 2024) or embed hashtag monitoring within police-led collaborative governance (Agus et al., 2026). Proposals for AI-based monitoring of disinformation (Yankoski et al., 2020) likewise focus on detection rather than on the response that should follow.

Theoretical Framework

This article defines cognitive security as the capacity of a community and its institutions to detect, verify, and absorb manipulative information before it changes intergroup perception and collective action. The definition narrows the broader concept reviewed by Grahn and Taipalus (2025) and Casino (2025) to the level at which electoral conflict unfolds. It also connects to Indonesia's doctrine of national resilience, in which the Astagatra framework links the natural Trigatra (geography, natural resources, population) with the social Pancagatra (ideology, politics, economy, socio-culture, defence and security) (Lembaga Ketahanan Nasional Republik Indonesia, 2023; Panca Putra S. et al., 2024; Salsabila et al., 2024). Misinformation in digitalised mass media is framed in this tradition as a threat to social stability (Winarko & Avianto, 2024), and disinformation is listed among the hybrid threats that call for total defence and public literacy (Adhilni et al., 2025).

The framework (Fig. 1) traces an escalation chain with three transitions: an information threat produces a perception shift (T1), the perception shift supports collective mobilisation (T2), and mobilisation becomes

a clash (T3). Four propositions guide the analysis. P1 holds that social vulnerability amplifies each transition. P2 holds that collective verification dampens T1 and T2; it is drawn on these two transitions only, because verification is expected to work before mobilisation rather than after clashes begin. P3 holds that digital indicators predict escalation only when calibrated with social indicators. P4 holds that warnings lead to response only when triggers and responsible actors are agreed in advance; in Indonesia these actors include Bawaslu, the regional national unity and political agencies (Badan Kesatuan Bangsa dan Politik, Kesbangpol), community early warning forums (Forum Kewaspadaan Dini Masyarakat, FKDM), and the police. P1, P2, and P4 are assessed against the case evidence. P3 is a design principle for the early warning framework, derived from the literature, and is not tested empirically in this study.

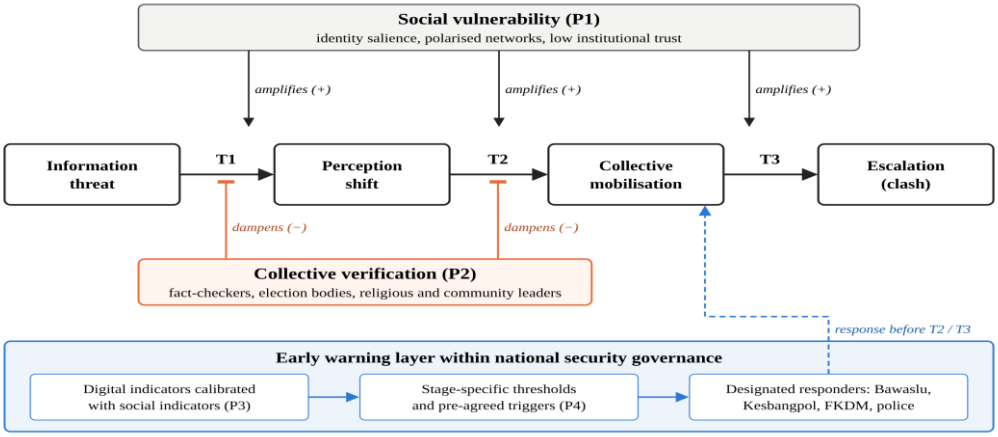


Fig. 1. Conceptual Model of the Escalation Chain and Early Warning Layer
Source: Authors' elaboration

II. METHOD

This study uses a qualitative, mechanism-oriented comparative case study grounded in scientific realism, which treats causal mechanisms as real but context-dependent. Within each case, theory-testing process tracing assesses whether the evidence expected by each transition is present (Beach & Pedersen, 2019; Bennett & Checkel, 2015; Collier, 2011). Across cases, a structured comparison locates where the chain stopped in 2024. Reporting follows the Standards for Reporting Qualitative Research (O'Brien et al., 2014), adapted to a document-based design.

Case selection follows a most-similar logic. Both cases are gubernatorial elections in Jakarta, the national capital and the most closely watched electoral arena, and both show high information and identity vulnerability, yet they differ on the outcome. The design matches cases on vulnerability, not on the political trigger, which differed and is therefore treated as a rival explanation. In 2017, the campaign began on 28 October 2016 (Rahadian, 2016), the first round was held on 15 February 2017, and the run-off on 19 April 2017, with the incumbent on mandatory campaign leave from 7 March to 15 April and provincial recapitulation from 29 April to 6 May 2017 (Sekretariat Kabinet Republik Indonesia, 2017). In 2024, the campaign ran from 25 September to 23 November, voting took place on 27 November, and counting and recapitulation lasted until 16 December 2024 (MC Kabupaten Bengkalis, 2024). The platform environment changed between the cases: internet penetration rose from 51.8 percent nationally in 2016, when 97.4 percent of users accessed social media (Asosiasi Penyelenggara Jasa Internet Indonesia [APJII], 2016), to 79.50 percent nationally and 87.51 percent in Jakarta in 2024 (APJII, 2024).

Documents were selected within two fixed windows: September 2016 to May 2017, covering the viral clip, the late-2016 rallies, both rounds, and recapitulation; and August 2024 to January 2025, covering the Bawaslu DKI vulnerability map, the campaign, the vote, and the post-result period. The analysed corpus consists of sixteen documents, eight per case (Table 1): survey releases on Jakarta voters and national internet use; official documents and statements by Bawaslu and the General Elections Commission (Komisi Pemilihan Umum, KPU); news reports of key events; and peer-reviewed studies of each election. A document was included if it (a) fell within a case window or reported on it, (b) bore directly on a transition or condition in the framework, (c) was publicly accessible and attributable to a named institution, outlet, or

author, and (d) reported figures that could be traced to an original release where one existed. Online sources were located with Indonesian and English keyword searches, including "Bawaslu DKI kerawanan Pilkada 2024", "kampanye SARA media sosial Jakarta", "demo 4 November 2016", "jadwal kampanye Pilkada DKI 2017", "cuti kampanye putaran II", "tahapan Pilkada 2024 PKPU", and "gugatan MK Pilgub Jakarta 2024", and were accessed on 26 October 2026; survey releases were read in their original PDF form. The units of analysis are the dated event, whether a mobilisation, a clash, or a clarification, and the documented statement or survey finding.

Table 1. Documentary Sources by Case

Source type	2017 case (Sep 2016 to May 2017)	2024 case (Aug 2024 to Jan 2025)
Survey releases	Indikator Politik Indonesia (2017); APJII (2016)	Research unit of the Kompas daily (Litbang Kompas), as reported by Gitiyarko (2024); Lembaga Survei Indonesia (2024); APJII (2024)
Official documents and election-body statements	Sekretariat Kabinet Republik Indonesia (2017)	Bawaslu DKI vulnerability map and outreach (Bisnis.com, 2024a, 2024b); KPU schedule (MC Kabupaten Bengkalis, 2024)
News reports of key events	Asmara (2016); Rahadian (2016)	Adji (2024)
Scholarly studies	Hadiz (2017); Lim (2017); Mietzner and Muhtadi (2018)	Nivada and Rassanjani (2024)

Source: Authors' compilation

Each document was read against the framework and classified by transition stage (T1 to T3), actor, narrative type, and verification response, using a codebook derived from the propositions; inductive codes were added when the material required. Indonesian excerpts were translated by the authors. For each transition, evidence was classified by test strength following Mahoney (2012) and Collier (2011). Evidence counted as a hoop test when the proposition requires it, so that its absence would substantially weaken the proposition; as a smoking gun when it would be very unlikely to appear if the proposition were false; and as straw-in-the-wind when it is consistent with the proposition but could also be produced by a rival explanation. Each item was classified by asking two questions: whether the proposition requires the evidence, and whether a rival explanation could produce it. Survey data entered the analysis as contextual and straw-in-the-wind evidence only, because the releases come from different institutes with different instruments. Figures were checked against the original releases where these were available.

Trustworthiness was strengthened by triangulation across source types, a classification matrix linking each claim to its source, and an explicit test of alternative explanations: candidate composition, elite cues, incumbency, national polarisation, and policing posture (Barron et al., 2016). The authors are researchers at a defence university whose institutional perspective emphasises national resilience; to limit the risk that this perspective shapes interpretation, every causal claim is tied to a cited public source and tested against rival explanations. The study involved no human participants; ordinary users named in hoax content are not identified, in line with Law Number 27 of 2022 on Personal Data Protection (UU 27/2022). Because there are no interviews and no systematic coding of fact-check archives, perception and verification are inferred from documents and survey releases, a limitation discussed in Section IV.

III. RESULTS AND DISCUSSION

Information threat profile. Both elections carried recognised information threats. In 2017, a clip of the governor's remarks on a verse of the Qur'an circulated widely before the November rally (Mietzner & Muhtadi, 2018; Lim, 2017). In 2024, Bawaslu DKI listed eight high-vulnerability issues, scored between 75 and 100 on its index, including hoax material on social media, SARA campaigning on social media, SARA

material in public places, and calls by figures or groups to reject particular candidates (Bisnis.com, 2024a). Exposure to such content was structurally higher in 2024: national internet penetration rose from 51.8 percent in 2016 to 79.50 percent in 2024, and reached 87.51 percent in Jakarta (APJII, 2016, 2024).

Transition T1, from information threat to perception shift. A Jakarta survey series on the blasphemy allegation in the 2017 window comes from Indikator Politik Indonesia (2017), whose January 2017 release (fieldwork 12 to 20 January, 697 respondents analysed, margin of error ±3.8 percent) reports a trend among respondents aware of the remark. The share judging the remark blasphemous peaked at 62 percent in November 2016, the month of the first mass rally, and fell to 54 percent in December and 47 percent in January, while the share judging it not blasphemous rose from 18 to 26 and 34 percent (Fig. 2). In the same release, 72 percent said they would accept the court verdict, and same religion was the main reason for vote choice for 6 percent of respondents overall, 12 percent of Agus Harimurti Yudhoyono's voters and 10 percent of Anies Baswedan's voters.

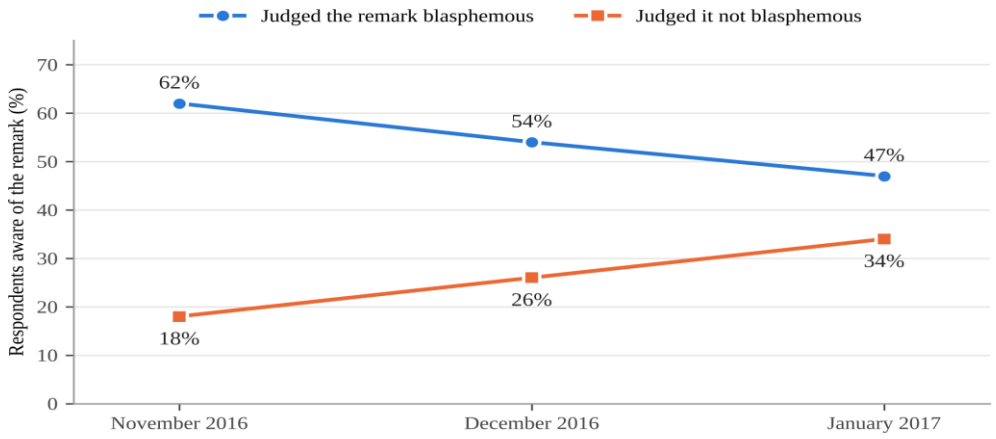


Fig. 2. Perceptions of the Blasphemy Allegation in Jakarta, November 2016 to January 2017
Source: Indikator Politik Indonesia (2017); respondents aware of the remark

For 2024, no survey asked about a comparable allegation, because none arose. Identity considerations nonetheless remained salient. In a Litbang Kompas survey conducted in June 2024, 75.3 percent of respondents said they would weigh whether a candidate shared their religion, 66.3 percent whether the candidate shared their ethnicity, and 56 percent approved of a candidate whose background is that of an Islamic religious figure (Gitiyarko, 2024). The article does not report the sample size or margin of error, so these figures are read as indicative rather than precise. Asked for their main reason for vote choice in September 2024, respondents cited government experience (24.6 percent), a proven track record (14.2 percent), and honesty (12.2 percent); the reported categories include no same-religion option, and being a religious figure was cited by 0.2 percent in a release by Lembaga Survei Indonesia (LSI) (Lembaga Survei Indonesia, 2024).

Vote choice by religion offers a rough indicator of how far identity translated into electoral alignment (Table 2). In January 2017, support for the incumbent pair stood at 31.9 percent among Muslim respondents and about 76.4 percent among non-Muslim respondents, a gap of about 44.5 percentage points. In September 2024, the gap between Muslim and non-Muslim respondents was 7.5 points for Pramono Anung and Rano Karno and 2.7 points for Ridwan Kamil and Suswono. The two surveys come from different institutes, with different samples, question wording, and candidate fields, and the 2017 contest included a non-Muslim incumbent while the 2024 contest did not. The comparison therefore indicates a contextual difference in how identity related to vote choice, not a measured trend in polarisation.

Table 2. Candidate Support by Respondent Religion in the Two Cases

Indicator	January 2017 (Indikator)	September 2024 (LSI)
Fieldwork; sample; margin of error	12 to 20 Jan 2017; 697 analysed; ±3.8%	6 to 12 Sep 2024; 1,200; ±2.9%
Share of Muslim respondents	86.0%	85.3%

Indicator	January 2017 (Indikator)	September 2024 (LSI)
Support among Muslim respondents	Ahok-Djarot 31.9%	Pramono-Rano 27.3%; Ridwan Kamil-Suswono 52.2%
Support among non-Muslim respondents	Protestant and Catholic 78.8%; other 69.8% (combined about 76.4%a)	Pramono-Rano 34.8%; Ridwan Kamil-Suswono 49.5%
Muslim and non-Muslim gap	about 44.5 points	7.5 points; 2.7 points

Source: Indikator Politik Indonesia (2017); Lembaga Survei Indonesia (2024). a Weighted by the authors from subgroup shares (10.3% and 3.7%); LSI reports non-Muslims as one group

Transition T2, from perception shift to mobilisation. In 2017, the blasphemy allegation supported a sequence of mass rallies in late 2016, the first of which drew about 100,000 people on 4 November, the month in which belief that the remark was blasphemous peaked (Asmara, 2016; Mietzner & Muhtadi, 2018). In 2024, Bawaslu DKI announced on the day it released its vulnerability map that it would communicate with political parties, religious leaders, and cross-ethnic community leaders so that the experience of the 2017 election would not be repeated (Bisnis.com, 2024b). The sources examined report no mass mobilisation of comparable scale during the 2024 campaign or after the vote; this absence in the reviewed sources does not rule out smaller local incidents, and the exact point at which the escalation chain stopped cannot be determined from them.

Transition T3, from mobilisation to clash. The 4 November 2016 rally ended in arson, looting, twelve injuries, and one death of unexplained cause (Asmara, 2016). No comparable clash was identified in the sources reviewed for 2024, and the result was not contested: by the deadline of 11 December 2024, the Ridwan Kamil and Suswono pair had filed no dispute (Adji, 2024).

Collective verification. The documentary record allows only a partial assessment of P2. For 2024, it shows pre-emptive verification work by an institutional actor: the election body publicly named hoaxes and SARA content as risks almost two months before the campaign and engaged party, religious, and cross-ethnic leaders (Bisnis.com, 2024a, 2024b). For 2017, the Indikator series shows that belief in the blasphemy allegation declined after November 2016 and that 72 percent of respondents would accept the court verdict, but the sources do not identify which actors drove that decline. The speed of clarification could not be reconstructed from the available documents. Table 3 summarises the evidence for each transition and the tests it supports.

Table 3. Evidence by Transition, Case, and Process Tracing Test

Transition	2017 evidence	2024 evidence	Test and assessment
T1 perception shift	Blasphemy judgement peaks at 62% in Nov 2016, the month of the first rally	No comparable allegation; identity considerations salient (75.3%); weaker religious alignment in a survey with a different instrument	Straw-in-the-wind; consistent with P1, causal effect not isolated
T2 mobilisation	Mass rally of about 100,000 on 4 Nov 2016, following the perception peak	No comparable mobilisation identified in the sources reviewed; pre-emptive outreach by Bawaslu DKI	Hoop expectation met for 2017 (temporal order); exact stopping point in 2024 not determinable
T3 clash	Arson, looting, 12 injured, 1 death	No comparable clash identified in the sources reviewed; result uncontested	Outcome observed in 2017; no comparable clash identified in 2024

Transition	2017 evidence	2024 evidence	Test and assessment
Verification and response (P2, P4)	Declining belief in the allegation; actors not identifiable	Election body names risks and engages religious and community leaders before the campaign	Straw-in-the-wind for P4; P2 partially assessed; P3 not tested

Source: Authors' analysis of the sources cited in this section

Analysis of the Escalation Chain

The sources reviewed point to an early interruption of the chain in 2024, but they do not show precisely where it stopped. Two observations support this reading. First, latent vulnerability did not disappear. Three in four respondents in a June 2024 survey still weighed a candidate's religion, and internet penetration in Jakarta was far higher than in 2016. If social vulnerability and digital exposure alone drove escalation, 2024 should have been at least as dangerous as 2017. Second, as far as surveys with different instruments allow comparison, identity appears to have translated less into electoral alignment in 2024 than in 2017, and no mobilisation comparable to that of late 2016 was identified. These observations are consistent with the proposition that social vulnerability amplifies escalation rather than producing it on its own (P1), in line with the conditional view of misinformation effects (Kurten, 2025; Lewandowsky et al., 2013). They do not, however, isolate the effect of vulnerability from political triggers, candidate composition, or security posture.

The 2017 perception series also supports the timing expected at T1. Belief that the remark was blasphemous peaked in the month of the first mass rally and declined as the campaign continued, which is consistent with an information shock whose perceptual effect was strongest at the moment of mobilisation. The evidence is straw-in-the-wind, since the series starts only in November 2016 and cannot show the level before the clip circulated.

Rival explanations remain open. The 2017 cleavage partly reflects candidate composition, because the incumbent was a Christian of Chinese descent accused of blasphemy, whereas the 2024 field offered no comparable trigger. Elite cues and coalition alignments also differed, and policing posture may have changed between the two elections (Barron et al., 2016). The 2024 evidence nonetheless shows an institutional actor naming the 2017 experience as the scenario to prevent and reaching out to religious and community leaders before the campaign began (Bisnis.com, 2024b), which fits P4's expectation that warnings matter when actors and responses are set in advance. On present evidence, the case material is consistent with P1, P4 receives straw-in-the-wind support, P3 remains untested because it concerns the design of future early warning rather than the cases themselves, and P2 is only partially tested: verification activity is visible before mobilisation, where the framework expects it to act, but its effect cannot be separated from the absence of a trigger comparable to the 2016 blasphemy allegation.

Answering the Research Questions

On RQ1, the comparison indicates that in 2017 the escalation chain ran through all three transitions: an identity-laden allegation shifted perceptions, the shift supported mass mobilisation within weeks, and mobilisation ended in a clash, with social vulnerability amplifying each step. Disinformation did not cause the violence on its own; it entered a chain whose later steps depended on mobilising structures and on the street setting. In 2024, comparable vulnerability and greater digital exposure did not produce the same sequence, and the sources reviewed identify no comparable mobilisation, although the exact point at which the chain stopped cannot be determined. On RQ2, collective verification appears in 2024 mainly as an institutional practice that operated before the campaign and drew in religious and community leaders, which is the stage at which P2 expects it to matter. The documentary evidence is consistent with this location but cannot establish that verification, rather than the absence of a comparable trigger, caused the break.

Dialogue with the Literature

The findings support the conditional view of the link between disinformation and violence. They are consistent with evidence that online content can propagate harm (Müller & Schwarz, 2021; Badrinathan et al., 2025), but they also suggest that exposure need not translate into escalation where the intermediate transitions are interrupted, echoing the limits identified by Leverso et al. (2025) and Stuart (2020). The

contrast between persistent identity salience and weaker religious alignment in the available 2024 survey is consistent with the proposition that social vulnerability amplifies escalation (P1), and it is compatible with Varshney's (2001) argument that cross-cutting civic ties shape whether tension becomes violence.

For digital literacy, the study shifts attention from individual discernment (Guess et al., 2020; Badrinathan, 2021; Huang et al., 2024) to collective verification by fact-checkers, election bodies, and community leaders. The 2024 pattern is consistent with a collective form that operates through pre-emption, by naming risks and engaging trusted intermediaries before mobilisation, rather than through correcting individual beliefs after exposure. If confirmed in further cases, this would extend the evidence on fact-checking (Walter et al., 2020; Porter & Wood, 2021) from belief correction toward conflict prevention, while the individual-level findings caution that literacy remains a weak protective factor when it depends on timing and on who verifies. For early warning, the study extends Nivada and Rassanjani (2024) from risk listing to stage-specific indicators. It differs from the community-centred 3E model (Wardle & Scales, 2025), which addresses information harms in general, and from the AGUSTIF model (Agus et al., 2026), which embeds monitoring in police-led governance, by tying each indicator to a transition and a responsible actor.

Theoretical Implications

The study recasts cognitive security as a capacity of communities and local institutions rather than only a property of artificial intelligence systems, individual minds, or strategic competition (Casino, 2025; Deppe & Schaal, 2024; Grahn & Taipalus, 2025). The escalation chain with its conditions offers a testable model: it specifies where information threats should leave observable traces, which conditions should amplify or dampen each step, and what evidence would disconfirm the model. Read through Indonesia's resilience doctrine, the model places information threats at the junction of the ideological, political, and socio-cultural gatra, where perception and mobilisation meet (Lembaga Ketahanan Nasional Republik Indonesia, 2023; Winarko & Avianto, 2024).

Policy Implications for Early Warning

Bawaslu DKI already lists hoaxes and SARA campaigning as issues (Bisnis.com, 2024a), so the gap lies less in detection than in linking issues to stages and responses. Table 4 proposes an indicator set per transition. Because P3 was not tested in this study, the calibration of digital with social indicators is offered as a design principle that requires empirical validation. Digital indicators such as surges of verified hoaxes and the spread of hostile identity narratives are calibrated with social indicators such as reports from community early warning forums (FKDM) and local conflict history, in line with evidence that media data alone are biased toward capitals and wealthier areas (Dowd et al., 2020). Each indicator carries a threshold and a designated responder, addressing the warning-response problem (George & Holl, 1997; Palihapitiya et al., 2025). Because UU 7/2012 does not specify how digital channels fit into early warning, the more useful step is implementing guidance that defines which digital indicators may be monitored, by which agency, and for what purpose, with safeguards for personal data under UU 27/2022, due process for any account or content action, and protection of freedom of expression, so that early warning does not become surveillance of legitimate political speech.

Table 4. Proposed Stage-Specific Early Warning Indicators

Transition	Digital indicator	Social indicator	Responsible actor
T1	Surge in verified hoaxes on identity themes; clarification lag	Identity salience in local surveys; prior SARA incidents	Bawaslu, fact-checkers, KPU
T2	Calls to rally or reject candidates on social media and messaging apps	FKDM reports of meetings and mobilisation; statements by religious leaders	Kesbangpol, FKDM, religious organisations
T3	Real-time reports of crowd movement and incitement	Police assessment of crowd size and counter-mobilisation	Police, regional government

Source: Authors' elaboration; thresholds to be set by the responsible agencies against local baselines

Comparative Insight and Generalisability

The pattern resonates with other contexts. In Kenya, citizen reporting platforms supported crisis response after the 2007 to 2008 post-election violence (Ajao & Wielenga, 2017), while comparisons of new

and old media show that digital monitoring captures urban areas better than rural ones (Dowd et al., 2020). In India and Pakistan, misinformation raises support for vigilantism (Badrinathan et al., 2025). The Jakarta comparison adds that the same city can move from escalation to containment across two successive electoral cycles, which points to transitions that institutions may be able to influence, although differences in political triggers remain a rival explanation. The propositions should transfer best to high-salience urban elections in Indonesia with dense platform use, and less to rural settings where information travels through different channels. If the pattern holds beyond Jakarta, resilience to election disinformation may be built mainly before the campaign, in the working links between election bodies and community leaders, rather than through reactive content removal once hostile narratives have spread.

IV. CONCLUSION

This article asked why two Jakarta elections with similar information and identity vulnerabilities ended differently. Digital exposure was higher and identity considerations remained salient in the later election, yet the available surveys indicate weaker religious alignment of vote choice, and no mobilisation or clash comparable to that of late 2016 was identified in the sources reviewed. The escalation chain therefore appears to have been interrupted early, although its exact stopping point cannot be determined, and the election body's pre-emptive outreach fits the role that the framework assigns to collective verification and pre-agreed response without being separable from the absence of a comparable political trigger. Theoretically, the study develops cognitive security as a community and institutional capacity and specifies a conditional escalation chain that can be tested in other cases; empirically, it offers a process-tracing comparison of two Jakarta elections; and practically, it proposes stage-specific early warning indicators whose calibration remains to be validated. Its limits are clear: two cases in one city restrict generalisability, fact-check archives were not coded systematically, clarification speed could not be measured, survey instruments differ across the cases, and the absence of reported incidents in the reviewed sources does not prove that none occurred. Future research should code fact-check archives across both windows, extend the comparison to other regional elections, and add interviews with community early warning forums, election officers, and fact-checkers. Ultimately, Indonesia's election security will depend less on detecting more hoaxes than on knowing at which stage a hoax becomes a mobilisation risk and who must act before it does.

DECLARATIONS

Author Contributions. Conceptualisation, SDP, Y, NF, HDH; methodology, Y; formal analysis, PW, NF; investigation and data curation, PW, Y; writing, original draft preparation, SDP; writing, review and editing, Y; supervision, NF, HDH. All authors have read and agreed to the published version of the manuscript.

Data Availability. All sources analysed in this study consist of publicly available documents, listed in Table 1 and the reference list. The document classification matrix and codebook are available from the corresponding author upon reasonable request.

Conflict of Interest. The authors declare no conflict of interest.

Ethical Statement. This study is based exclusively on the analysis of publicly available documents and does not involve human participants or the collection of personal data through direct interaction. Accordingly, formal ethical approval was not sought. Any institutional requirements concerning ethical review or exemption have been addressed in accordance with the applicable regulations of the Republic of Indonesia Defense University.

REFERENCES

[1] Adhilni, R., Gultom, R. A. G., Saragih, H. J. R., Murtiana, S., & Kusuma, A. A. (2025). Total defense systems in national resilience strategy: Assessing Indonesia's readiness to confront hybrid threats. ***Journal of Governance and Social Policy***, 6(2), 300–323. https://doi.org/10.24815/gaspol.v6i2.10

[2] Adji, B. P. (2024, December 12). Tak ajukan gugatan ke MK hingga tenggat berakhir, RK-Suswono akui kemenangan Pramono-Rano?. Republika. https://news.republika.co.id/berita/socb8y409/tak-ajukan-gugatan-ke-mk-hingga-tenggat-berakhir-rksuswono-akui-kemenangan-pramonorano

[3] Agus, M., Niswaty, R., Nur, A. C., & Didin. (2026). The AGUSTIF model: An adaptive governance framework for hybrid electoral conflict management in South Sulawesi. ***International Journal of Safety and Security Engineering,*** 16(1). https://doi.org/10.18280/ijsse.160111

[4] Ajao, T., & Wielenga, C. (2017). Citizen journalism and conflict transformation: The Ushahidi's response to Kenya's 2008 post-election violence. Matatu, 49, 467–486. https://doi.org/10.1163/18757421-04902012

[5] Amarasinghe, A., Nanlohy, S., Morgan, T., Hammond, D., Dahiya, Y., & Bailo, F. (2026). Mapping violence perceptions through YouTube comments: A new approach to real-time violence monitoring. EPJ Data Science, 15, 47. https://doi.org/10.1140/epjds/s13688-026-00649-y

[6] Asmara, T. (2016, November 4). Mass protest turns violent in Jakarta. BenarNews. https://www.benarnews.org/english/news/indonesian/protest-violence-11042016155349.html

[7] Asosiasi Penyelenggara Jasa Internet Indonesia. (2016). Infografis penetrasi dan perilaku pengguna internet Indonesia: Survey 2016. APJII.

[8] Asosiasi Penyelenggara Jasa Internet Indonesia. (2024). Survei penetrasi internet dan perilaku penggunaan internet 2024. APJII.

[9] Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in India. American Political Science Review, 115(4), 1325–1341. https://doi.org/10.1017/S0003055421000459

[10] Badrinathan, S., Chauchard, S., & Siddiqui, N. (2025). Misinformation and support for vigilantism: An experiment in India and Pakistan. American Political Science Review, 119(2), 947–965. https://doi.org/10.1017/S0003055424000790

[11] Barron, P., Jaffrey, S., & Varshney, A. (2016). When large conflicts subside: The ebbs and flows of violence in post-Suharto Indonesia. ***Journal of East Asian Studies***, 16(2), 191–217. https://doi.org/10.1017/jea.2016.6

[12] Bazzi, S., Blair, R. A., Blattman, C., Dube, O., Gudgeon, M., & Peck, R. M. (2022). The promise and pitfalls of conflict prediction: Evidence from Colombia and Indonesia. The Review of Economics and Statistics, 104(4), 764–779. https://doi.org/10.1162/rest_a_01016

[13] Beach, D., & Pedersen, R. B. (2019). Process-tracing methods: Foundations and guidelines (2nd ed.). University of Michigan Press. https://doi.org/10.3998/mpub.10072208

[14] Benford, R. D., & Snow, D. A. (2000). Framing processes and social movements: An overview and assessment. Annual Review of Sociology, 26(1), 611–639. https://doi.org/10.1146/annurev.soc.26.1.611

[15] Bennett, A., & Checkel, J. T. (2015). Process tracing: From metaphor to analytic tool. Cambridge University Press. https://doi.org/10.1017/CBO9781139858472

[16] Bhavnani, R., Findley, M. G., & Kuklinski, J. H. (2009). Rumor dynamics in ethnic violence. ***The Journal of Politics***, 71(3), 876–892. https://doi.org/10.1017/S002238160909077X

[17] Birch, S., Daxecker, U., & Höglund, K. (2020). Electoral violence: An introduction. ***Journal of Peace Research***, 57(1), 3–14. https://doi.org/10.1177/0022343319889657

[18] Bisnis.com. (2024, August 1). Antisipasi isu SARA di Pilkada, Bawaslu dekati parpol hingga tokoh agama. Bisnis.com. https://kabar24.bisnis.com/read/20240801/15/1787479/antisipasi-isu-sara-di-pilkada-bawaslu-dekati-parpol-hingga-tokoh-agama

[19] Bisnis.com. (2024, August 1). Bawaslu rilis tingkat kerawanan Pilgub Jakarta, potensi kampanye SARA tinggi. Bisnis.com. https://kabar24.bisnis.com/read/20240801/15/1787470/bawaslu-rilis-tingkat-kerawanan-pilgub-jakarta-potensi-kampanye-sara-tinggi

[20] Botara, S. J. (2022). Experience of living with Islam-Christianity in North Maluku, Indonesia. ***International Journal of Educational Research & Social Sciences,*** 3(3), 1170–1176. https://doi.org/10.51601/ijersc.v3i3.383

[21] Casino, F. (2025). Unveiling the multifaceted concept of cognitive security: Trends, perspectives, and future challenges. Technology in Society, 83, 102956. https://doi.org/10.1016/j.techsoc.2025.102956

[22] Collier, D. (2011). Understanding process tracing. PS: Political Science & Politics, 44(4), 823–830. https://doi.org/10.1017/S1049096511001429

[23] Deppe, C., & Schaal, G. S. (2024). Cognitive warfare: A conceptual analysis of the NATO ACT cognitive warfare exploratory concept. Frontiers in Big Data, 7, 1452129. https://doi.org/10.3389/fdata.2024.1452129

[24] Dowd, C., Justino, P., Kishi, R., & Marchais, G. (2020). Comparing 'new' and 'old' media for violence monitoring and crisis response: Evidence from Kenya. Research & Politics, 7(3), 2053168020937592. https://doi.org/10.1177/2053168020937592

[25] Elsaesser, C., Patton, D. U., Weinstein, E., Santiago, J., Clarke, A., & Eschmann, R. (2021). Small becomes big, fast: Adolescent perceptions of how social media features escalate online conflict to offline violence. Children and Youth Services Review, 122, 105898. https://doi.org/10.1016/j.childyouth.2020.105898

[26] Fossati, D. (2019). The resurgence of ideology in Indonesia: Political Islam, aliran and political behaviour. ***Journal of Current Southeast Asian Affairs***, 38(2), 119–148. https://doi.org/10.1177/1868103419868400

[27] George, A. L., & Holl, J. E. (1997). The warning-response problem and missed opportunities in preventive diplomacy. Carnegie Corporation of New York. https://carnegie.org/publication/the-warning-response-problem-and-missed-opportunities-in-preventive-diplomacy/

[28] Gitiyarko, V. (2024, July 16). Survei Pilkada 2024: Pertimbangan rasional pemilih Jakarta. Kompas.id. https://www.kompas.id/artikel/survei-pilkada-2024-pertimbangan-rasional-pemilih-jakarta

[29] Grahn, H., & Taipalus, T. (2025). Defining comprehensive cognitive security in the digital era: Literature review and concept analysis. ***Journal of Information Warfare***, 24(2), 39–59. https://urn.fi/URN:NBN:fi:jyu-202505134216

[30] Greenhill, K. M., & Oppenheim, B. (2017). Rumor has it: The adoption of unverified information in conflict zones. International Studies Quarterly, 61(3), 660–676. https://doi.org/10.1093/isq/sqx015

[31] Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences, 117(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

[32] Hadiz, V. R. (2017). Indonesia's year of democratic setbacks: Towards a new phase of deepening illiberalism?. Bulletin of Indonesian Economic Studies, 53(3), 261–278. https://doi.org/10.1080/00074918.2017.1410311

[33] Hegre, H., Allansson, M., Basedau, M., Colaresi, M. P., Croicu, M., Fjelde, H., Hoyles, F., Hultman, L., Högbladh, S., Jansen, R., Mouhleb, N., Muhammad, S. A., Nilsson, D., Nygård, H. M., Olafsdottir, G., Petrova, K., Randahl, D., Rød, E. G., Schneider, G., . . . Vestby, J. (2019). ViEWS: A political violence early-warning system. ***Journal of Peace Research***, 56(2), 155–174. https://doi.org/10.1177/0022343319823860

[34] Huang, G., Jia, W., & Yu, W. (2024). Media literacy interventions improve resilience to misinformation: A meta-analytic investigation of overall effect and moderating factors. Communication Research, 53(7), 899–926. https://doi.org/10.1177/00936502241288103

[35] Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. ***The International Journal of Press/Politics***, 25(3), 493–516. https://doi.org/10.1177/1940161219900126

[36] Indikator Politik Indonesia. (2017). Temuan survei Provinsi DKI Jakarta, 12-20 Januari 2017 [Rilis survei]. Indikator Politik Indonesia.

[37] Kurten, M. (2025). Why we fight for fractured truths: How misinformation fuels political violence in democracies. MediaWell, Social Science Research Council. https://mediawell.ssrc.org/research-reviews/why-we-fight-for-fractured-truths-how-misinformation-fuels-political-violence-in-democracies/

[38] Lee, A. (2020). Online hoaxes, existential threat, and internet shutdown: A case study of securitization dynamics in Indonesia. ***Journal of Indonesian Social Sciences and Humanities***, 10(1), 17–34. https://ejournal.brin.go.id/jissh/article/view/8650

[39] Lembaga Ketahanan Nasional Republik Indonesia. (2023). Buku ajar bidang studi ketahanan nasional (4th ed.). Lemhannas Press.

[40] Lembaga Survei Indonesia. (2024). Rilis temuan survei: Pengaruh Anies di pemilihan gubernur Jakarta [Rilis survei, 18 September 2024]. Lembaga Survei Indonesia.

[41] Leverso, J., Diouane, Y., & Mohler, G. (2025). Measuring online-offline spillover of gang violence using bivariate Hawkes processes. ***Journal of Quantitative Criminology***, 41, 103–131. https://doi.org/10.1007/s10940-024-09592-5

[42] Lewandowsky, S., Stritzke, W. G. K., Freund, A. M., Oberauer, K., & Krueger, J. I. (2013). Misinformation, disinformation, and violent conflict: From Iraq and the "War on Terror" to future threats to peace. American Psychologist, 68(7), 487–501. https://doi.org/10.1037/a0034515

[43] Lim, M. (2017). Freedom to hate: Social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia. Critical Asian Studies, 49(3), 411–427. https://doi.org/10.1080/14672715.2017.1341188

[44] Mahoney, J. (2012). The logic of process tracing tests in the social sciences. Sociological Methods & Research, 41(4), 570–597. https://doi.org/10.1177/0049124112437709

[45] MC Kabupaten Bengkalis. (2024, January 31). KPU RI terbitkan PKPU Pilkada 2024, ini jadwal dan tahapannya. InfoPublik. https://infopublik.id/kategori/pemilu/821870/kpu-ri-terbitkan-pkpu-pilkada-2024-ini-jadwal-dan-tahapannya

[46] Mendieta, M. V., Warren, T. C., & Vaca, C. (2026). Predicting organized collective violence through hostile discourse in social media. EPJ Data Science, 15, 3. https://doi.org/10.1140/epjds/s13688-025-00596-0

[47] Mietzner, M., & Muhtadi, B. (2018). Explaining the 2016 Islamist mobilisation in Indonesia: Religious intolerance, militant groups and the politics of accommodation. Asian Studies Review, 42(3), 479–497. https://doi.org/10.1080/10357823.2018.1473335

[48] Mueller, H., & Rauh, C. (2018). Reading between the lines: Prediction of political violence using newspaper text. American Political Science Review, 112(2), 358–375. https://doi.org/10.1017/S0003055417000570

[49] Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. ***Journal of the European Economic Association***, 19(4), 2131–2167. https://doi.org/10.1093/jeea/jvaa045

[50] Nivada, A., & Rassanjani, S. (2024). Pemetaan risiko politik dan keamanan Pemilu 2024: Mekanisme kewaspadaan dini dan implikasinya. ***Jurnal Trias Politika***, 8(1), 33–48. https://doi.org/10.33373/jtp.v8i1.6147

[51] O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. Academic Medicine, 89(9), 1245–1251. https://doi.org/10.1097/ACM.0000000000000388

[52] Palihapitiya, M., Khan, S. A., Ahmed, F., Shenouda, B., & Nelson, E. (2025). Addressing critical gaps in data collection and analysis in atrocities prevention early warning systems. ***Journal of Peacebuilding & Development***, 20(1), 20–37. https://doi.org/10.1177/15423166241293353

[53] Panca Putra S., R. Z., Gusnar, I., Solihin, D., Tantri, S., Andrea, H. P., Trismawan, D., & Santoso, S. (2024). Reactualization of national resilience: A historical and conceptual study. ***Jurnal Lemhannas RI***, 12(1), 65–80. https://doi.org/10.55960/jlri.v12i1.558

[54] 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

[55] Piazza, J. A. (2022). Fake news: The effects of social media disinformation on domestic terrorism. Dynamics of Asymmetric Conflict, 15(1), 55–77. https://doi.org/10.1080/17467586.2021.1895263

[56] Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. Proceedings of the National Academy of Sciences, 118(37), e2104235118. https://doi.org/10.1073/pnas.2104235118

[57] Rahadian, L. (2016, October 25). Satu dua dan tiga untuk Agus, Ahok, dan Anies. CNN Indonesia. https://www.cnnindonesia.com/politik/20161025205250-32-167935/satu-dua-dan-tiga-untuk-agus-ahok-dan-anies/

[58] Rød, E. G., Gåsste, T., & Hegre, H. (2024). A review and comparison of conflict early warning systems. ***International Journal of Forecasting,*** 40(1), 96–112. https://doi.org/10.1016/j.ijforecast.2023.01.001

[59] Rohman, A., Pang, N. L.-S., & Pitaloka, D. (2020). The episodes of a Facebook group for information sharing in the Ambon 2011 conflict prevention movement, Indonesia. Information, Communication & Society, 23(4), 539–554. https://doi.org/10.1080/1369118X.2018.1521456

[60] Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. Science Advances, 8(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254

[61] Salsabila, M., Slam, Z., Alamsyah, M. N., & Hilmi, H. D. (2024). National resilience in historical and geostrategic perspectives: Essence and conception. ***Jurnal Lemhannas RI***, 12(4), 435–452. https://doi.org/10.55960/jlri.v12i4.956

[62] Sekretariat Kabinet Republik Indonesia. (2017, March 6). Ahok wajib cuti kampanye putaran II, Mendagri kembali akan tunjuk Plt Gubernur DKI Jakarta. Sekretariat Kabinet Republik Indonesia. https://setkab.go.id/ahok-wajib-cuti-kampanye-putaran-ii-mendagri-kembali-akan-tunjuk-plt-gubernur-dki-jakarta/

[63] Soelistyowati, R. D., Subhan, A., & Sigit, R. R. (2024). Virtual public sphere on social media (Habermas' critical study of the digital hoax phenomenon WhatsApp group alumni Iluni Menwa UI). ***International Journal of Educational Research & Social Sciences***, 5(2), 284–289. https://doi.org/10.51601/ijersc.v5i2.737

[64] Stuart, F. (2020). Code of the tweet: Urban gang violence in the social media age. Social Problems, 67(2), 191–207. https://doi.org/10.1093/socpro/spz010

[65] Subarjo, A. H., & Setianingsih, W. (2020). Literasi berita hoaxs di internet dan implikasinya terhadap ketahanan pribadi mahasiswa. ***Jurnal Ketahanan Nasional,*** 26(1), 1–23. https://doi.org/10.22146/jkn.51109

[66] 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

[67] Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi. (2022). Lembaran Negara Republik Indonesia.

[68] Undang-Undang Nomor 7 Tahun 2012 tentang Penanganan Konflik Sosial. (2012). Lembaran Negara Republik Indonesia.

[69] Varshney, A. (2001). Ethnic conflict and civil society: India and beyond. World Politics, 53(3), 362–398. https://doi.org/10.1353/wp.2001.0012

[70] 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

[71] Wählisch, M., & Kufus, F. (2026). Structural dilemmas of PeaceTech: AI, power, and peacebuilding in the digital age. Peacebuilding. Advance online publication. https://doi.org/10.1080/21647259.2026.2644694

[72] Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. Political Communication, 37(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894

[73] Wardle, C., & Scales, D. (2025). Advocating for a community-centred model for responding to potential information harms. Nature Human Behaviour, 9(8), 1546–1556. https://doi.org/10.1038/s41562-025-02233-2

[74] Winarko, H., & Avianto, L. (2024). Strengthening digital transformation of mass media for national resilience. ***Jurnal Lemhannas RI***, 12(1), 109–116. https://doi.org/10.55960/jlri.v12i1.559

[75] Yang Hui, J. (2020). Social media and the 2019 Indonesian elections: Hoax takes the centre stage. In Southeast Asian Affairs 2020 (M. Cook & D. Singh, Eds.), ISEAS-Yusof Ishak Institute (pp. 155–172). https://www.cambridge.org/core/books/southeast-asian-affairs-2020/social-media-and-the-2019-indonesian-elections-hoax-takes-the-centre-stage/A8B3D0C5325002FA936F9A60C03C7AA9

[76] Yankoski, M., Weninger, T., & Scheirer, W. (2020). An AI early warning system to monitor online disinformation, stop violence, and protect elections. Bulletin of the Atomic Scientists, 76(2), 85–90. https://doi.org/10.1080/00963402.2020.1728976.