

DEEFAKE DISINFORMATION BY POLITICAL BUZZERS: ANALYSIS OF THE IMPLICATIONS FOR CIVIL LIBERTIES IN THE "IBU JILBAB PINK" CASE

Nailah Alifah Mujahidah ^{a*)}, Lia Wulandari ^{a)}

^{a)} Universitas Pembangunan Nasional "Veteran" Jakarta, Jakarta, Indonesia

^{*)}Corresponding Author: nailahalifahmujahidahh@gmail

Article history: received 19 May 2026; revised May 26, 2026; accepted 29 June 2026

DOI: <https://doi.org/10.33751/jhss.v10i2.211>

Abstract. The increasing sophistication of digital media manipulation has intensified the spread of political disinformation, presenting new challenges for civil liberties in digital public spaces. This study examines the implications of deepfake disinformation about "Ibu Jilbab Pink" disseminated by political buzzers on Instagram during the August-September 2025 movement for citizens' civil liberties in Indonesia's digital public sphere. Using a qualitative case study approach that combines netnographic observation of 15 purposively selected Instagram content units with in-depth expert interviews, the analysis identifies three dominant narrative constructions: symbolic delegitimization, identity delegitimization, and public polarization. The findings indicate that political buzzers systematically employed deepfake-based disinformation to undermine both the individual figure and the symbolic meaning of collective protest, thereby narrowing citizens' freedom of expression, weakening deliberative discourse, and suppressing civic participation. The study concludes that deepfake disinformation may function as an instrument of refeudalization, as theorized by Habermas, with implications that extend beyond the individual figure to the broader civil and political citizenship of the public she represented.

Keywords: Deepfake Disinformation; Political Buzzers; Civil Liberties; Digital Public Sphere.

I. INTRODUCTION

Regional and national democratic life in Indonesia has increasingly shifted into the digital sphere, where social media platforms now function as primary arenas for citizens to exercise free expression, voice opinions, and engage in political discourse. Civil liberties, encompassing freedom of expression, freedom of opinion, and the right to accurate information, are expected to be fully realized within this digital space. However, the actual condition of civil liberties in Indonesia remains concerning. The 2024 Democracy Index ranks Indonesia 59th out of 167 countries, with a civil liberties score of only 5.29, one of the lowest among all measured components [1]. The Indonesian Central Bureau of Statistics likewise identifies civil-liberty indicators, freedom of assembly, association, expression, and opinion, as core prerequisites for a functioning democracy, and notes that Indonesia's civil liberties score has continued to decline in recent years [2]. This downward trend forms an important backdrop for understanding how vulnerable citizens' expressive space has become when confronted with organized disinformation in digital environments.

A significant research gap exists in the literature on Indonesian political disinformation: while prior studies have examined deepfakes and buzzer politics primarily in electoral contexts or in relation to elite political targets [3],[4],[5], no study specifically addresses cases in which buzzer-driven deepfake technology targets non-elite citizens who emerge

organically from social movements. Tufekci [6] found that attacking symbolic individuals is among the most effective repression strategies against network-based movements, yet her analysis was written before the rapid development of deepfake technology and does not address AI-based visual manipulation, buzzer activity on visual platforms such as Instagram, or the Indonesian context directly. This study seeks to fill that gap through the case of "Ibu Jilbab Pink."

Instagram is one of the most strategic social media platforms in Indonesia, with approximately 22 million active users dominated by the 18-29 age group [7], a platform that plays a substantial role in shaping opinions, attitudes, and public behavior [8]. Data from Safer Internet Lab indicate that political and election-related disinformation ranks first among issues most detrimental to Indonesian society at 32.80 percent, followed by religious and ideological issues at 30.69 percent [9]. Fake accounts and bots represent the largest category of disinformation actors at 26.76 percent, followed by domestic political actors at 15.18 percent [9], indicating that disinformation in Indonesia's digital space is frequently an organized rather than organic phenomenon.

The figure known as "Ana" or "Ibu Jilbab Pink" first gained public attention during the 17+8 movement, a series of demonstrations spanning labor rights, police accountability, and parliamentary conduct, after footage of her protesting near the House of Representatives went viral [10],[11]. Her presence made her a symbol of civil courage, and the color

pink became a collective emblem of resistance, culminating in the concept of "brave pink" [12].



Fig. 1. Post Ana as the symbol of "Brave Pink" in the 17+8 People's Demands Movement by @menjadimanusia on Instagram.

This recognition was followed by deepfake content portraying her using crude language against President Prabowo Subianto, subsequently confirmed as AI manipulation by AFP Fact Check [13]. The attacks operated through two layers: visual deepfake manipulation and a parallel identity-delegitimization operation through an account later identified as belonging to a police officer [14].



Fig. 2. Deepfake content targeting "Ibu Jilbab Pink" on Instagram.

The spread of deepfake disinformation involves deliberate organized dissemination. A buzzer is an individual or account that systematically spreads particular messages to shape public opinion, a practice that originated from buzz marketing before transforming into political buzzing [4]. The scale of buzzer influence in Indonesia is substantial: then-Minister of Communication and Informatics noted that 92 percent of digital uproar is triggered by buzzer activity [15]. Buzzers typically combine captions and engaging visual content with language styles tailored to appear natural [16], while sending messages continuously to make topics trend [17]. The dominance of account numbers and intensity of repeated postings keep content positioned at the top of engagement [18], and this disinformation is further reinforced by social media algorithms that create filter bubbles susceptible to polarization [19].

This study employs three theoretical frameworks. Jürgen Habermas's Public Sphere Theory [20] serves as the primary lens, particularly his concept of "refeudalization", the capture of rational discourse spaces by organized interested actors [21]. Two supporting frameworks deepen the analysis: Wardle and Derakhshan's Information Disorder framework [22], used to classify the deepfake content as deliberate disinformation; and Marshall's citizenship concept [23], used to identify which civil and political citizenship elements were threatened. Additionally, Nagara [21] applies Habermas's framework

specifically to deepfake phenomena, providing a contextually relevant bridge between the theoretical framework and the Indonesian digital context analyzed in this study.

Based on this analysis, this study addresses the following research question: How does the deepfake disinformation spread by political buzzers on Instagram have implications for citizens' civil liberties in Indonesia's digital public sphere during the August-September 2025 movement? This question matters because Instagram has developed into a primary arena in which non-elite citizens can be turned into targets of organized disinformation, with consequences ranging from individual reputational harm to the collective narrowing of civil liberties for citizens who might otherwise be encouraged to participate in social movements.

II. RESEARCH METHODS

This study employs a qualitative approach with a case study design, appropriate for examining bounded phenomena with clear temporal and spatial limits [24]. The case of "Ibu Jilbab Pink" was selected because it represents an instance of organized deepfake disinformation directed at a non-elite citizen who emerged organically from a social movement rather than through formal political competition.

Data were collected through two primary methods: netnographic observation and semi-structured in-depth interviews. This study deliberately limits its arena of content analysis to a single platform, Instagram, because it was the primary platform on which the "Ibu Jilbab Pink" deepfake content and its accompanying delegitimization narratives first emerged and circulated most intensively during the period under study. Following Kozinets [25], netnography adapts ethnographic techniques to study cultures and communities emerging through computer-mediated communication. Netnography was therefore employed to observe publicly available Instagram content only, including posts, and captions, associated with the disinformation narrative targeting "Ibu Jilbab Pink." Content circulating on other platforms such as TikTok, X, or Facebook lies outside the scope of this analysis.

The 15 content units were selected through purposive netnographic sampling using hashtags and keywords directly associated with the case: #ibujilbabbpink, #bravepink, and #cancelbravepink, supplemented by the figure's name and visual depictions. Inclusion criteria required that content: (1) explicitly referenced or depicted the figure through hashtag, name, or visual description; and (2) contributed to the disinformation narrative through labeling, delegitimization, or polarization. Content that reported the case neutrally or expressed organic support without a delegitimizing function was excluded. Data collection ran from 3 September to 22 October 2025 and continued until thematic saturation was reached, the point at which additional content no longer introduced new narrative patterns. The relatively limited number of units is consistent with the qualitative aim of this study, which prioritizes depth of narrative analysis over statistical representativeness.

The classification of accounts as "political buzzers" is based on observable content-behavior patterns rather than forensic verification of account ownership or organizational affiliation, which lies beyond this study's scope. This analytical limitation must be explicitly acknowledged. Following the buzzer typology of Arianto [26] and characteristics outlined by Camil et al. [27], accounts were classified as exhibiting buzzer-like patterns when activity displayed: (1) repetition of identical labels and hashtags across multiple accounts within short timeframes; (2) synchronized narrative progression from emotional to pseudo-factual framing; and (3) sustained posting intensity consistent with patterns described by Bradshaw and Howard [17] and Kurniawan et al. [18]. The term "political buzzer" in this study therefore denotes accounts demonstrating these coordinated, repetitive, and narrative-driven behaviors, and does not constitute a forensic claim regarding the identity, employer, or formal network behind any individual account.

Relatedly, this study does not undertake any forensic or technical authentication of the deepfake media itself, an analysis that lies outside the disciplinary scope of political communication and political science. The determination that the core video targeting the figure was AI-manipulated is grounded in the independent assessment of AFP Fact Check [13], which this study treats as an adequate evidentiary basis for a socio-political analysis. Accordingly, the analytical focus here is the construction of the disinformation narrative and its implications for civil liberties, not the pixel-level authenticity of any individual artifact; the public contestation over that authenticity is itself treated as part of the phenomenon under study.

Content coding followed the Miles, Huberman, and Saldaña [28] model: data collection, data reduction, data presentation, and conclusion drawing. All 15 content units were coded inductively by the sole researcher to identify recurring themes and narrative patterns, producing three categories with the following operational definitions: (1) Symbolic Delegitimization, content attacking the collective meaning of the movement's symbol through relabeling or reframing of the color pink; (2) Identity Delegitimization, content attacking the personal credibility of the individual figure through claims about her personal identity; (3) Public Polarization, content diverting discourse from the movement's substantive demands toward debate over the individual's suitability as a symbol. To strengthen validity, source, technique, and time triangulation were applied, comparing coding results with interview data and verifying interpretations across different informants.

Semi-structured in-depth interviews were conducted with five informants selected through purposive sampling based on their professional expertise. Interview data were analyzed thematically and integrated with netnographic findings through systematic triangulation. Because this study analyzes a real and identifiable individual who became a target of disinformation, the analysis relies exclusively on content already publicly circulated on Instagram. The figure is referred to only by the publicly known designation "Ana" or "Ibu Jilbab Pink," and claims about her personal condition are treated strictly as objects of analysis rather than factual assertions endorsed by this research

TABLE I
RESEARCH INFORMANTS PROFILE

Code	Categories	Involvement
N-1	Communication Scholar	Researcher on the role of political buzzers in shaping public opinion on social media.
N-2	Political Analyst	Executive Director of a political survey institute, expertise in public opinion and political communication.
N-3	Social Media Analytics Scholar	Specialist in social media analytics and digital platforms.
N-4	Digital Rights Scholar and Activist	Ex Executive Director of a digital rights organization, Expertise in disinformation and digital civil liberties
N-5	Fact-Checking Researcher and Journalist	Researcher on newsroom fact-checking and disinformation

III. RESULTS AND DISCUSSION

The following analysis applies Habermas's Public Sphere Theory [20] as the primary framework, with the Information Disorder framework [22] and Marshall's citizenship concept [23] as complementary lenses for specific analytical dimensions.

A. Narrative Construction by Political Buzzers Against the "Ibu Jilbab Pink" Figure on Instagram

Based on netnographic observation, 15 content units were identified and grouped into three narrative categories through data reduction and coding following the Miles et al. [28] model: Symbolic Delegitimization (8 contents), Identity Delegitimization (4 contents), and Public Polarization (3 contents). The narrative constructions developed chronologically in a mutually reinforcing manner. Symbolic delegitimization emerged first on 3-4 September through repeated labeling and the #CancelBravePink hashtag; identity delegitimization followed on 8 September through a police officer's claim that Ana was a person with special needs, reproduced across three contents within four days; public polarization ran concurrently, diverting discourse from movement substance. The recirculation of content on 22 October, more than one month after the triggering events, indicates that the narrative outlasted the immediate political moment, consistent with findings that buzzer-driven disinformation campaigns are designed to maintain discursive dominance over time [18],[5].

TABLE III
CODING RESULTS OF DEEPPAKE CONTENT "IBU JILBAB PINK" ON
INSTAGRAM

No.	Code	Date	Category	Narrative Indicator & Coding Ratio
1.	SD-01	03/09/25	Symbolic Deleg.	Labels Ana as 'provokator'; questions symbolic legitimacy as movement representative
2.	SD-02	03/09/25	Symbolic Deleg	Emotional distancing from movement symbol; erodes symbolic support without explicit labeling
3.	ID-01	04/09/25	Identity Deleg.	Precursor framing; shifts discourse from collective movement concerns to Ana's personal affairs
4.	SD-03	04/09/25	Symbolic Deleg.	#CancelBravePink campaign; emotionally frames pink = anarchism
5.	SD-04	05/09/25	Symbolic Deleg.	Pseudo-factual strategy: cites specific date/location/event as 'evidence' that pink = anarchism
6.	SD-05	06/09/25	Symbolic Deleg.	Explicit 'provokator' label; links Ana's actions to a specific partisan political agenda
7.	SD-06	06/09/25	Symbolic Deleg.	#CancelBravePink intensification; simultaneous anarchist and racist labeling across accounts
8.	PP-01	07/09/25	Public Polariz.	Use of 'allegedly' as implicit doubt marker; frames Ana's symbolic status as contested
9.	PP-02	07/09/25	Public Polariz.	'Agree or uncomfortable?' binary framing; explicitly divides audience against the symbol
10.	ID-02	08/09/25	Identity Deleg.	Police officer niece claim: Ana has 'special needs'; launches central identity narrative
11.	ID-03	08/09/25	Identity Deleg.	Same claim reproduced across a different media account; creates false-verification effect
12.	ID-04	12/09/25	Identity Deleg.	Third reinforcement of special-needs narrative; sustains delegitimization four days later
13.	SD-07	16/09/25	Symbolic Deleg.	Explicit arrest demand ('TANGKAP IBU INI!!!'); criminalizes the movement's symbol
14.	PP-03	17/09/25	Public Polariz.	Moral-responsibility framing; implicitly discourages civic participation
15.	SD-08	22/10/25	Symbolic Deleg.	Recirculation over one month later; demonstrates narrative durability

Source: Researcher's netnographic observation, 2026.

Table II presents the complete coding of the fifteen content units by code, date, category, and narrative indicator, revealing a clear chronological escalation from symbolic labeling to identity delegitimization, and a final recirculation on 22

October. Beyond this temporal pattern, the distribution across the three categories indicates where the disinformation effort was most concentrated, which Table III summarizes below.

TABLE III
NETNOGRAPHY RESEARCH FINDINGS SUMMARY

No.	Categories	Total	Key Coding Criteria
1.	Symbolic Delegitimization	8	Repeated negative labels ('anarchist,' 'racist,' 'provocateur') targeting movement symbol; #CancelBravePink campaign; redefining pink = anarchism through emotional then pseudo-factual framing; explicit criminalization demand
2.	Identity Delegitimization	4	Claims about individual's personal identity targeting credibility; special-needs narrative attributed to police officer; reproduction across multiple accounts creating false-verification effect
3.	Public Polarization	3	Binary framing dividing public opinion; moral-responsibility discourse creating self-censorship conditions; implicit doubt markers questioning symbolic legitimacy

Symbolic Delegitimization

Symbolic Delegitimization is the most dominant category, comprising 8 of the 15 contents analyzed. This dominance suggests that the most intensive attacks were directed not solely at Ana as an individual, but at the collective meaning attached to the movement's symbol, the color pink as a representation of civic resistance. Two symbolic targets may be identified: the "brave pink" symbol as a collective representation, and Ana's status as an individual deemed worthy of embodying it.

A consistent pattern of repeated labeling, comprising the terms "anarchist," "racist," and "provocateur", appeared across multiple accounts within a short timeframe. Following Wardle and Derakhshan [22], disinformation is understood as false information deliberately created and disseminated with the intent to cause harm. This repetitive pattern is consistent with the element of deliberate intent to harm, since the labels did not appear sporadically but were used consistently across multiple accounts within a short period [4],[26]. The most explicit evidence is the #CancelBravePink campaign: two accounts uploaded content with the identical hashtag within a two-day span on 4 and 6 September 2025, while the 5 September content adopted a pseudo-factual strategy by citing a specific date, location, and event as "concrete evidence" that pink was synonymous with anarchism. This escalation from emotional to pseudo-factual framing across coordinated accounts reflects the buzzer working pattern described by Bradshaw and Howard [17] and Kurniawan et al. [18].

"In practice, this figure is used as an intermediate target. The figure of Ibu Jilbab Pink was made into a symbol of courage, translated through the color pink. It is precisely this symbol that was then targeted for destruction. This resembles the pattern in color revolutions elsewhere, yellow umbrellas in Hong Kong, black shirts in Myanmar. What gets attacked is the symbol." (N-4, Interview, 10 June 2026)

The statement above reflects the understanding that the color pink, as a movement symbol, may have functioned as the primary target of delegitimization. Destroying a collective symbol can be as effective as attacking the substance of a movement itself [6]. The peak of symbolic delegitimization is evident in Content SD-07 on 16 September 2025, which demanded repressive action against Ana with the explicit phrase "TANGKAP IBU INI!!!" (ARREST THIS WOMAN!!!), accompanied by the labels "anarchist" and "racist." Chesney and Citron [29] note that deepfakes can function as instruments of reputational sabotage capable of producing real legal and social consequences. This demand appears to constitute part of a delegitimization pattern that could open space for legitimizing repression, directly targeting Ana's exercise of freedom of opinion as protected within Marshall's civil element of citizenship [23]. Content SD-08, appearing on 22 October 2025 more than a month after the main events, indicates the durability of the narrative beyond the immediate political context.

"Labeling narratives, including the label of treason, represent an attempt to manipulate information and delegitimize the symbol of a civil movement. The anarchist and treason narratives function as diversions, so that the public does not respond to the people's genuine grievances but instead becomes trapped in a discussion that can provide justification for repression." (N-1, Interview, 24 May 2026)

Identity Delegitimization

Identity Delegitimization comprises 4 contents and exhibits a clear chronological pattern. Unlike symbolic delegitimization, which targets the collective meaning of the movement's symbol, this category works by attacking Ana's personal credibility through claims about her individual identity. The mechanism is structurally analogous to deepfake manipulation: rather than fabricating visual evidence, it constructs a biographical narrative designed to undermine her symbolic status.

Content ID-01 on 4 September 2025 functioned as a precursor, establishing a frame that shifted attention from the movement's collective concerns to Ana's personal affairs. The central identity narrative emerged on 8 September through Content ID-02, when an online media account disseminated the claim that Ana was a person with special needs, attributed to someone claiming to be her niece and identifying as a police officer [14]. This claim reappeared in Content ID-03 on the same day and was reinforced in Content ID-04 four days later on 12 September. Bradshaw and Howard [17] explain that repetition functions not only to make an issue trend but also to create the impression that information has been verified by multiple independent sources, when in fact what occurred was the reproduction of a single, identical, unverified claim.

Chesney and Citron [29] similarly note that identity exploitation operates through the destruction of an individual's credibility without direct engagement with the substantive claims she raised.

"Research indicates that disinformation spreads far faster than its fact-check. Beyond that, cognitive bias plays a role: even when a fact-check exists, cognitive dissonance tends to occur, people absorb information that aligns with their existing beliefs or prior biases rather than corrections that challenge them." (N-3, Interview, 8 June 2026)

The statement above helps explain why the false-verification mechanism through narrative repetition may have been effective: a public already skeptical of Ana was more likely to accept the special-needs narrative because it reinforced preexisting prejudices, regardless of its factual validity. Miqdad [5] notes that the asymmetry between the speed of disinformation production and the pace of its correction creates a structural advantage for disinformation actors in shaping public perception, particularly among audiences with varying levels of digital literacy [19].

Public Polarization

Public Polarization comprises 3 contents and is qualitatively distinct from the previous two categories. Rather than attacking Ana directly, content in this category opened debate that diverted focus from the substantive demands of the 17+8 movement toward judgment of one individual's suitability as a symbol. This mechanism may constitute the most structurally significant form of the campaign, precisely because it appears neutral while producing divisive effects.

Content PP-01 and PP-02 on 7 September 2025 used different framing mechanisms to divide public opinion. PP-01 deployed the word "allegedly" as an implicit marker of doubt, indirectly questioning whether Ana truly represented an act of resistance. PP-02 openly invited audiences to take a side with the question "agree or uncomfortable?" Habermas's refeudalization concept [20],[21] explains how Instagram, which should have served as an arena for substantive discussion of the 17+8 demands, instead became an arena of debate over one individual's suitability as a symbol.

"This can be read more as a shifting-attention strategy. In every major protest there are two possible counter-strategies: refute the criticism, or divert attention. Diverting attention is the easier path, the way to destroy a symbol is to damage its identity, so that public focus shifts toward matters peripheral to the movement's core demands." (N-3, Interview, 8 June 2026)

Content PP-03 on 17 September 2025 adopted a neutral tone of "respecting differing opinions" while implicitly discouraging citizens from speaking out unless "politically competent." This reflects what Noelle-Neumann [30] describes as the spiral of silence, which operates not through direct threat but through a social norm that leads citizens to self-censor political expression. Nisa [19] notes that in the Indonesian digital context, such normative framing can suppress civic participation precisely because it appears reasonable on the

surface while reinforcing the perception that dissent carries social risk. Together, the three categories form a systematic disinformation structure: Symbolic Delegitimization seizes and reconstructs the movement's collective symbol; Identity Delegitimization undermines the personal credibility of its human embodiment; and Public Polarization ensures that digital discourse remains filled with individual-focused debate rather than movement substance.

B. Implications of Deepfake Disinformation for Civil Liberties in the Digital Public Sphere

The implications of buzzer-driven deepfake disinformation for citizens' civil liberties appear to operate across two interrelated dimensions of Marshall's citizenship framework [23], analyzed within the overarching structural context of Habermas's refeudalization [20],[21].

On the dimension of the civil element, freedom of speech and opinion, and the right to accurate information, Ana's action in demonstrating and criticizing security forces constituted a concrete exercise of this civil element [23]. As N-1 observed, her presence triggered broader collective participation through symbolic actions such as pink-and-green profile filters, suggesting that her civil expression may have catalyzed wider civic engagement. AFP Fact Check confirmed that Ana's video was AI-manipulated [13], yet public debate continued expanding for weeks. Pawelec [31] explains that deepfakes erode trust in information by making it difficult to distinguish fact from manipulation, causing society to fragment around different versions of "truth." Without a shared factual foundation, freedom of opinion loses its substantive meaning. Nagara [21] likewise demonstrates how deepfakes can distort the digital public sphere as theorized by Habermas, replacing rational discourse with manipulated narratives.

"This is an important phenomenon in the post-truth context. Post-truth refers to the condition in which a lie repeated often enough can eventually be believed as truth. Regarding this deepfake phenomenon, if left unaddressed, because deepfakes closely resemble authentic content, people with lower digital literacy are at greater risk of accepting it as genuine." (N-2, Interview, 4 June 2026)

The structural dimension of this problem is further indicated by the finding from N-5: a CSIS study published in 2023 found that 81.4% of respondents had never read a fact-check correction, and 45.4% had never verified information before sharing it. This suggests that the limited effectiveness of AFP Fact Check's confirmation in containing the disinformation about Ana may reflect a systemic gap between the speed of disinformation and the reach of its correction, rather than solely individual negligence. Kusuma [32] and Nasution [33] both note that the right to accurate information, as a component of the civil element, is structurally dependent on the quality and accessibility of information infrastructure in the digital public sphere.

The implications for the civil element also appear to operate through the spiral of silence [30]. Citizens who witnessed labels such as "anarchist" and "racist" attached to Ana across multiple contents, particularly SD-03, SD-06, and SD-07, may have

chosen to restrict their own expression in the digital public sphere for fear of similar treatment. Hampton et al. [34] show that social media can amplify the spiral of silence, as the high visibility of digital attacks creates the perception that a dominant narrative represents majority opinion, even when it originates from a small, organized group.

"Labeling narratives create an atmosphere of fear in the digital public sphere, in which citizens who observe racist labels or treason accusations directed at activists may feel intimidated and reluctant to express their opinions. This labeling can function as an instrument of digital domination aimed at suppressing citizens' communicative autonomy on social media." (N-1, Interview, 24 May 2026)

On the dimension of the political element, the right to participate in the exercise of political power outside formal electoral mechanisms [23], Ana was not attacked solely as an individual but as a symbol of collective political participation. When that symbol was delegitimized through deepfake labeling, the effect extended beyond her personal image to question the legitimacy of the broader movement she represented. Habermas's refeudalization concept [20],[21] explains how Instagram was transformed from an arena for rational discussion of the 17+8 demands into a space dominated by debate over one individual's suitability, a form of discourse capture by organized interested actors. Wijayanto and Berenschot [35] observe that in Indonesia, political buzzer networks have developed precisely as instruments of this kind of discursive control, designed to manage the boundaries of what can be legitimately expressed in the digital public sphere.

"The main goal is not necessarily to find the truth, but to make people doubt the symbol. Creating doubt alone may be sufficient to reduce the militancy of those protesting. In that sense, the debate over whether the video is authentic or fabricated can itself serve the purpose of the disinformation campaign." (N-4, Interview, 10 June 2026)

The long-term implications may extend beyond the August-September 2025 movement. Pratama et al. [36] argue that freedom of opinion is an indicator of the sustainability of democracy and a reflection of the recognition of individual dignity as equal citizens. When a spiral of silence forms, what collapses is not only the freedom of a single individual but the quality of democracy as a whole. Wiratraman [37] asserts that a civil space weakened through digital persecution can result in the paralysis of citizens' political participation. N-2 articulated this risk: when deepfake content associates a moral movement with anarchism and provocation, citizens who had supported it may become doubtful, and those who had not yet participated may become more reluctant, a dynamic consistent with Tufekci's [6] concept of "tactical freeze."

"When a deepfake enters the narrative, it can damage the movement. The public, and even the government, may lose sympathy. People may say the movement has been infiltrated or that its content is fabricated. This is what can endanger future social movements. Governments and security forces may

become increasingly wary, so that even a fundamentally moral movement ends up being stigmatized." (N-2, Interview, 4 June 2026)

It should be noted that not all informants regard deepfake disinformation as the most immediate threat to civil liberties. N-3 observed that the arrest of activists carries more direct and tangible silencing effects than digital delegitimization. This perspective does not diminish the relevance of the digital disinformation dimension analyzed here, but confirms that distinct layers of threat exist: physical repression operates through direct, measurable fear, while deepfake disinformation appears to operate through slower, more systemic erosion of trust, symbolic delegitimization, and collective demobilization. In answer to the research question, the deepfake disinformation spread by political buzzers on Instagram appears to have narrowed civil liberties through two complementary mechanisms: (1) a systematic three-stage narrative construction converting a symbolic figure into an intermediate target; and (2) a consequent narrowing of civil and political citizenship elements, a collapsing factual foundation, a self-censoring spiral of silence, and a stigmatized movement, all within a digital public sphere that had undergone refeudalization.

IV. CONCLUSIONS

This study indicates that the deepfake disinformation spread by political buzzers on Instagram against "Ibu Jilbab Pink" during the August-September 2025 movement systematically narrowed citizens' civil liberties through three mutually reinforcing narrative constructions: Symbolic Delegitimization (8 contents), Identity Delegitimization (4 contents), and Public Polarization (3 contents). These constructions operated chronologically and complementarily to convert a non-elite figure who emerged organically from the 17+8 movement into an intermediate target whose delegitimization served to demoralize the movement's broader civic base. The implications for civil liberties operated across two dimensions of Marshall's citizenship framework. On the civil element, deepfake disinformation appears to have collapsed the shared factual foundation necessary for meaningful freedom of opinion and generated a spiral of silence that may have led citizens to self-censor political expression. On the political element, the delegitimization of the movement's symbol suppressed collective civic participation with potential long-term effects on future social movements. Both dimensions unfolded within a digital public sphere that had undergone refeudalization, consistent with Habermas's theoretical framework. The theoretical contribution of this study lies in demonstrating that non-elite citizens who emerge organically from social movements may be equally, if not more, vulnerable to buzzer-driven deepfake attacks than elite political figures, and that such attacks may constitute a systemic threat to civil and political citizenship rather than merely individual reputational harm. The practical contribution is to raise awareness among citizens, policymakers, and platform regulators of the need for faster fact-checking infrastructure, stronger digital literacy programs, and clearer platform accountability for coordinated inauthentic behavior. This study carries several important limitations. First, the classification of

accounts as political buzzers is based on observable behavioral patterns rather than forensic network verification; conclusions about the organizational structure or affiliation behind individual accounts cannot be drawn. Second, the analysis is limited to publicly available Instagram content and does not cover other platforms such as TikTok, X, or Facebook. Third, the coding was conducted by a single researcher, which may affect interpretive consistency; member checking through expert interviews was applied to partially mitigate this. Fourth, the sample of 15 content units reflects a qualitative priority of depth over breadth and findings are not statistically generalizable. Future research may extend this analysis through comparative case studies across multiple social movements and platforms, forensic digital network analysis to map buzzer organizational structures, and broader platform-based studies spanning Instagram, TikTok, X, and Facebook to capture cross-platform dynamics of deepfake disinformation.

REFERENCES

- [1] The Economist Intelligence Unit, "Democracy Index 2024: Democracy in an Age of Conflict," Economist Intelligence Unit, 2025. [Online]. Available: <https://www.eiu.com/n/campaigns/democracy-index-2024/>. Accessed: Mar. 1, 2025.
- [2] Badan Pusat Statistik, "Statistik 80 Tahun Indonesia Merdeka," BPS RI, 2025. [Online]. Available: <https://www.bps.go.id/id/publication/2025/09/26/89069c5f3f244aae3b7fd913/statistik-80-tahun-indonesia-merdeka.html>. Accessed: Oct. 5, 2025.
- [3] C. T. Noerman and A. L. Ibrahim, "Kriminalisasi deepfake di Indonesia sebagai bentuk perlindungan negara," *Jurnal USM Law Review*, vol. 7, no. 2, pp. 603-621, 2024.
- [4] F. I. M. Daeni, F. A. Rachmarani, and I. Rhiza, "Pengaruh buzzer politik dalam pemilu: Tantangan terhadap electoral justice dalam mempertahankan prinsip demokrasi," *Padjadjaran Law Review*, vol. 11, no. 2, pp. 183-200, 2023.
- [5] M. Miqdad, "Literature review: Buzzer politik dan pengembangan opini di media sosial di Indonesia," *NeoRespublica: Jurnal Ilmu Pemerintahan*, vol. 5, no. 2, pp. 689-698, 2024.
- [6] Z. Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven, CT, USA: Yale University Press, 2017.
- [7] D. R. Indika and C. Jovita, "Media sosial Instagram sebagai sarana promosi untuk meningkatkan minat beli konsumen," *Jurnal Bisnis Terapan*, vol. 1, no. 01, pp. 25-32, 2017.
- [8] A. Pujiono, "Media sosial sebagai media pembelajaran bagi generasi Z," *Didache: Journal of Christian Education*, vol. 2, no. 1, pp. 1-19, 2021.
- [9] H. J. Salim, "Survei Safer Internet Lab: Hoaks politik dan pemilu paling banyak rugikan masyarakat," *Liputan6.com*, Jan. 18, 2025. [Online]. Available: <https://www.liputan6.com/cek-fakta/read/5885876/survei-safer-internet-lab-hoaks-politik-dan-pemilu-paling-banyak-rugikan-masyarakat?page=2>. Accessed: Mar. 10, 2025.
- [10] N. K. Ranawati, "Apa itu 17+8 tuntutan rakyat yang ramai di medsos dan maknanya," *DetikJabar*, Sep. 2, 2025. [Online]. Available: <https://www.detik.com/jabar/berita/d->

- 8092490/apa-itu-17-8-tuntutan-rakyat-yang-ramai-di-medsos-dan-maknanya. Accessed: Oct. 1, 2025.
- [11] Detik.com, "Potret heroik ibu jilbab pink di garis depan demo DPR, viral bawa merah putih," Aug. 29, 2025. [Online]. Available: <https://wolipop.detik.com/worklife/d-8085918/potret-heroik-ibu-jilbab-pink-di-garis-depan-demo-dpr-viral-bawa-merah-putih>. Accessed: Sep. 5, 2025.
- [12] IDN Times, "Fakta-fakta ibu jilbab pink saat demo DPR yang kembali viral," Sep. 4, 2025. [Online]. Available: <https://www.idntimes.com/news/indonesia/fakta-fakta-ibu-jilbab-pink-saat-demo-dpr-yang-kembali-viral-00-f48mr-klhnj3>. Accessed: Sep. 10, 2025.
- [13] M. Nurhadi, "Video ibu jilbab pink maki-maki Prabowo dan minta Anies jadi presiden deepfake," Suara.com, Sep. 4, 2025. [Online]. Available: <https://www.suara.com/teknologi/2025/09/04/212326/video-ibu-jilbab-pink-maki-maki-prabowo-dan-minta-anies-jadi-presiden-deepfake>. Accessed: Sep. 12, 2025.
- [14] Tribunnewsmaker.com, "Sosok keponakan ibu jilbab pink di demo DPR, ternyata polisi, kini minta maaf ke teman seprofesi," Sep. 8, 2025. [Online]. Available: <https://newsmaker.tribunnews.com/viral/164765/sosok-keponakan-ibu-jilbab-pink-di-demo-dpr-ternyata-polisi-kini-minta-maaf-ke-teman-seprofesi>. Accessed: Sep. 14, 2025.
- [15] T. Rizal, "Budi Arie: 92 persen kegaduhan ruang digital ulah buzzer," BeritaSatu, Mar. 20, 2024. [Online]. Available: <https://www.beritasatu.com/nasional/2806057/budi-arie-92-persen-kegaduhan-ruang-digital-ulah-buzzer>. Accessed: Apr. 1, 2025.
- [16] D. Yulianti and Zulfebriges, "Aktivitas buzzer di Instagram dalam promosi MD Clinic," in *Prosiding Manajemen Komunikasi*, vol. 4, no. 2, 2018.
- [17] S. Bradshaw and P. N. Howard, "The global disinformation order: 2019 global inventory of organised social media manipulation," Oxford Internet Institute, Oxford, UK, Comprop Research Report, no. 24, 2019. [Online]. Available: <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>. Accessed: Feb. 10, 2025.
- [18] R. Kurniawan, R. Y. Muliana, F. Maesaroh, M. I. Nurcahyo, and A. J. Kusuma, "Buzzer media sosial dan pembentukan perspektif pemilih millennial dalam pemilu 2019," *Jurnal Politik Walisongo*, vol. 3, no. 1, pp. 54-72, 2021.
- [19] K. Nisa, "Peran literasi di era digital dalam menghadapi hoaks dan disinformasi di media sosial," *Impressive: Journal of Education*, vol. 2, no. 1, pp. 1-11, 2024.
- [20] J. Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Great Britain: Polity Press, 1989.
- [21] M. A. Nagara, "Deepfake dan distorsi ruang publik digital: Analisis teori public sphere Jurgen Habermas," *Jurnal Komunikasi & Media Digital*, vol. 3, no. 1, pp. 44-52, 2025.
- [22] C. Wardle and H. Derakhshan, "Information disorder: Toward an interdisciplinary framework for research and policymaking," *Council of Europe*, vol. 27, pp. 1-107, 2017.
- [23] T. H. Marshall, *Citizenship and Social Class, and Other Essays*. Cambridge, UK: Cambridge University Press, 1950.
- [24] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA, USA: SAGE Publications, 2017.
- [25] R. V. Kozinets, *Netnography: The Essential Guide to Qualitative Social Media Research*, 3rd ed. London, UK: SAGE Publications, 2020.
- [26] B. Arianto, "Salah kaprah ihwal buzzer: Analisis percakapan warganet di media sosial," *JiIP: Jurnal Ilmiah Ilmu Pemerintahan*, vol. 5, no. 1, pp. 1-20, 2020.
- [27] R. Camil, N. H. Attamimi, and K. Esti, "Dibalik fenomena buzzer: Memahami lanskap industri dan pengaruh buzzer di Indonesia," *Centre for Innovation Policy and Governance*, vol. 1, no. 1, pp. 1-30, 2017.
- [28] M. B. Miles, A. M. Huberman, and J. Saldaña, *Qualitative Data Analysis: A Methods Sourcebook*, 3rd ed. Thousand Oaks, CA, USA: SAGE Publications, 2014.
- [29] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *California Law Review*, vol. 107, no. 6, pp. 1753-1819, 2019.
- [30] E. Noelle-Neumann, "The spiral of silence: A theory of public opinion," *Journal of Communication*, vol. 24, no. 2, pp. 43-51, 1974.
- [31] M. Pawelec, "Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions," *Digital Society: Ethics, Socio-Legal and Governance of Digital Technology*, vol. 1, no. 2, p. 19, 2022.
- [32] E. Kusuma, "Kebebasan berpendapat dan kaitannya dengan hak asasi manusia (HAM)," *Sanskara Hukum Dan HAM*, vol. 1, no. 3, pp. 97-101, 2023.
- [33] L. Nasution, "Hak kebebasan berpendapat dan berekspresi dalam ruang publik di era digital," *Adalah*, vol. 4, no. 3, pp. 37-48, 2020.
- [34] K. N. Hampton, L. Rainie, W. Lu, M. Dwyer, I. Shin, and K. Purcell, "Social media and the spiral of silence," *Pew Research Center*, Aug. 2014. [Online]. Available: <https://www.pewresearch.org/internet/2014/08/26/social-media-and-the-spiral-of-silence/>. Accessed: Feb. 20, 2025.
- [35] Wijayanto and W. Berenschot, Eds., *Pasukan Siber: Politik Buzzer di Indonesia*. Jakarta, Indonesia: Yayasan Obor Indonesia, 2021.
- [36] M. I. Pratama, A. Rahman, and F. Bachmid, "Kebebasan berpendapat dan berekspresi di media sosial dalam perspektif hak asasi manusia," *Qawanin Jurnal Ilmu Hukum*, vol. 3, no. 1, pp. 1-16, 2022.
- [37] H. P. Wiratraman, "Kebebasan sipil dan politik hukum di era pemanipulasian," *Pusat Studi Hukum dan Kebijakan Indonesia*, Jul. 26, 2023. [Online]. Available: <https://www.pshk.or.id/aktivitas/kebebasan-sipil-dan-politik-hukum-di-era-pemanipulasian/>. Accessed: Mar. 5, 2025.