International Journal of Social Sciences and English Literature
Vol. 10, No. 6, 1-9, 2026
ISSN(E) 2576-683X
DOI: 10.55220/2576-683x.v10.1348

From Criticism to Digital Aggression: Indonesian Facebook Reactions to the National Data Center Breach

I Dewa Gede Agung Aditya Keramas

Faculty of Humanities, Universitas Udayana, Indonesia.

Abstract

This study examines how digital aggression was linguistically constructed in Indonesian Facebook reactions to the National Data Center (Pusat Data Nasional, PDN) breach and whether such hostility crossed the threshold of hate-oriented expression. Drawing on forensic linguistics, sociolinguistics, and Critical Discourse Analysis, the study applies corpus-assisted qualitative discourse analysis to 404 unique timestamped publicly accessible Facebook comments posted between 29 June and 2 July 2024. The analysis focuses on lexical choice, target construction, blame attribution, illocutionary force, and ideological framing. Rather than treating all hostile expressions as hate speech, the study distinguishes policy criticism, hostile evaluation, personal insult, digital aggression, and hate-oriented expression. The findings show that 193 comments, or 47.8% of the corpus, contained at least one form of digital aggression, yielding 233 coded category instances under multi-label coding. Blame attribution was the dominant category, followed by delegitimation of authority, personal insult, hostile evaluation, and sarcastic ridicule. No comment met the study's strict operational definition of hate-oriented expression. The study argues that hostility in the corpus is better understood as accountability-driven digital aggression rather than hate speech, with implications for forensic linguistics, critical digital literacy, and online discourse education.

Keywords: Critical digital literacy, Critical discourse analysis, Cyber-crisis discourse, Digital aggression, Forensic linguistics.

1. Introduction

The breach of Indonesia's National Data Center (*Pusat Data Nasional*, PDN) triggered one of the country's most visible digital crises in recent years, provoking sustained public reaction across online platforms. Beyond its technical and administrative implications, the incident quickly developed into a discursive event in which public frustration, distrust, and political resentment were articulated through everyday digital communication (Austin et al., 2012; Coombs, 2007; Knight & Nurse, 2020; Kuipers & Schönheit, 2022). Online platforms in Indonesia, and Facebook in particular, became key spaces where citizens transformed institutional failure into vernacular commentary marked by accusation, ridicule, and morally charged evaluation.

In June 2024, Indonesia's National Data Center (*Pusat Data Nasional*, PDN) experienced a major service disruption attributed by the National Cyber and Crypto Agency (BSSN) to a ransomware attack, which affected services across multiple government agencies (Kompas, 2024; Reuters, 2024). In the case of the PDN breach, Facebook comment threads became a particularly visible arena where Facebook users expressed dissatisfaction with state institutions and public officials. The data examined in this study show recurring calls for resignation and dismissal, as reflected in comments such as "wajar *rakyat* minta *Menkominfo* mengundurkan diri" ("it is reasonable for the people to ask the Minister of Communication and Informatics to resign") and "*Menkominfo harus dipecat*" ("*Menkominfo* must be dismissed"), which frame the incident not merely as a technical failure but as a moral and political failure of governance (Coombs, 2007; Coombs & Holladay, 2002; Kuipers & Schönheit, 2022).

The language of reaction, however, extended well beyond policy critique. Many comments shifted into more aggressive and personalized forms of expression, including hostile labeling, sarcastic ridicule, and direct attacks on competence, morality, and legitimacy. Expressions such as "*tidak becus*," "*tidak berkompeten*," "*tak ada otak*," and "*brengsek*" operate as evaluative and aggressive acts that go beyond ordinary disagreement (Hardaker, 2010; Jane, 2015; Runions & Bak, 2015; Suler, 2004). These features make PDN-related Facebook discourse a productive object for forensic linguistics, sociolinguistics, and Critical Discourse Analysis (CDA).

Research on hate speech, online incivility, and aggressive digital communication has expanded considerably in recent years (Davidson et al., 2017; Fortuna & Nunes, 2018; Poletto et al., 2021; Schmidt & Wiegand, 2017). Existing scholarship has examined abusive language, flaming, trolling, cyber-harassment, and hate-oriented discourse across diverse online environments (Hardaker, 2010; Jane, 2015; KhosraviNik & Esposito, 2018; Runions & Bak, 2015). However, much of this work concentrates on identity-based hate (ethnicity, religion, gender) rather than on hostile discourse triggered by state-related cyber incidents.

In the Indonesian context, studies of digital discourse have addressed political communication, polarization, and social media contestation (Ibrohim & Budi, 2023; Lim, 2017; Tahir & Ramadhan, 2024; Benu et al., 2025), yet fewer studies have examined cyber-crisis discourse as a distinct object of linguistic analysis. There remains limited work that approaches Indonesian online hostility through a forensic linguistic lens, particularly in cases where the trigger is institutional failure rather than identity-based conflict. Hostile online discourse is also frequently treated as a single undifferentiated category, despite operating along a continuum from negative evaluation to mockery, personal insult, digital aggression, and hate-oriented expression.

Accordingly, the present study addresses a gap at the intersection of four areas rarely brought together in a single analytical frame: hostile online discourse, digital aggression, blame attribution, and cyber-crisis discourse in Indonesian social media. Rather than assuming that all hostile online expressions constitute hate speech, this study examines how Facebook comments move along a continuum from policy criticism to blame attribution, delegitimation, personal insult, and possible hate-oriented expression. By focusing on Facebook reactions to the PDN breach, it (i) extends forensic and critical discourse research to cyber-crisis contexts, and (ii) offers an integrated framework that distinguishes graded forms of online hostility.

Against this backdrop, the present study examines how hostile online discourse was linguistically constructed in Indonesian Facebook users' reactions to the PDN breach, guided by three research questions:

1. What linguistic forms of digital aggression appear in Facebook reactions to the PDN breach, and to what extent do they meet the threshold of hate-oriented expression?
2. Who or what becomes the target of hostility, and how is blame discursively constructed?
3. How do these expressions reproduce ideological framings of state (in)competence and public (dis)trust?

Theoretically, the study contributes to forensic linguistics by demonstrating how hostile online discourse can be examined as linguistic evidence of aggression, blame, and target construction in digitally mediated interaction. Practically, it informs digital literacy, critical language awareness in English-language education, content moderation, and the legal-linguistic interpretation of abusive language. Although the data are in Indonesian, the study is relevant to English applied linguistics and English-language education because it addresses critical digital literacy, online discourse evaluation, and the ability to distinguish criticism, aggression, and hate-oriented expression in multilingual digital environments. Such distinctions are increasingly important for English learners and educators who participate in transnational online communication and routinely encounter overlapping Indonesian-English digital registers.

2. Literature Review and Theoretical Framework

2.1. Forensic Linguistics

Forensic linguistics is the study of language in contexts of law, evidence, and conflict, encompassing authorship attribution, threat assessment, defamation, and contested communication (Coulthard et al., 2017; Gibbons & Turell, 2008; Olsson & Luchjenbroers, 2014; Shuy, 1993). In digital environments, its relevance has grown because online communication frequently produces verbal conflict that is public, persistent, and easily circulated (Coulthard et al., 2017; Gibbons & Turell, 2008; KhosraviNik & Esposito, 2018). Within this study, forensic linguistics provides analytical tools for identifying abusive intent, target reference, illocutionary force (insult, threat, accusation, incitement), and the evidential value of utterances as linguistic acts of aggression or delegitimation.

2.2. Hate Speech

Hate speech is broadly understood as language that attacks, degrades, or stigmatizes a person or group on the basis of identity markers such as ethnicity, religion, gender, or political affiliation (Brown, 2017; Fortuna & Nunes, 2018; Waldron, 2012; Waseem & Hovy, 2016). A key analytical distinction must be drawn between general offensive language—profanity, mockery, or harsh criticism aimed at an individual—and hate speech proper, which invokes a socially marked identity category (Davidson et al., 2017; Fortuna & Nunes, 2018; Waseem & Hovy, 2016). In online communication this boundary is often blurred, since rapid, emotionally charged, and intertextual discourse can simultaneously perform mockery, political critique, and symbolic aggression (Jane, 2015; KhosraviNik & Esposito, 2018; Waseem & Hovy, 2016). For the purposes of this study, hate speech is operationally defined as Facebook discourse that goes beyond ordinary criticism by attacking a person or socially identifiable collectivity through degrading, stigmatizing, dehumanizing, or exclusionary expression. In this study, references to political affiliation were not automatically coded as hate-oriented expression unless they involved degrading, stigmatizing, dehumanizing, or exclusionary attacks against a socially identifiable political group; criticism of political appointment, patronage, or competence was coded as delegitimation of authority rather than hate speech.

2.3. Digital Aggression

Digital aggression denotes a broader category of hostile communicative behavior in online environments, including insult, ridicule, flaming, humiliation, and coordinated verbal hostility (Hardaker, 2010; Jane, 2015; Runions & Bak, 2015; Suler, 2004). Unlike hate speech, digital aggression does not require identity-based targeting; it captures hostility directed at individuals, institutions, or perceived opponents on grounds other than protected social identity. This is analytically important for the present study, because PDN-related Facebook comments display extensive verbal hostility—calls for dismissal, ridicule of incompetence, accusations of *balas budi*, pejorative labeling—that is aggressive without necessarily meeting the stricter threshold of hate speech.

2.4. Sociolinguistics

Sociolinguistics provides the second interpretive layer by examining how language use indexes social meaning, stance, and identity in interaction (Bucholtz & Hall, 2005; Eckert, 2008; Jaffe, 2009). In digitally mediated discourse, sociolinguistic analysis is particularly attentive to code choice, register shifting, and stance-taking, all of which carry social meaning beyond propositional content (Androutsopoulos, 2014; Tagg, 2015). Within this study, sociolinguistics interprets how commenters position themselves morally and politically through evaluative lexis,

address terms, and Indonesian-English code-mixing, and how these stance acts construct in-group solidarity and out-group delegitimation in cyber-crisis discourse.

2.5. Critical Discourse Analysis

Critical Discourse Analysis (CDA) treats discourse as a social practice that constructs, legitimizes, and reproduces relations of power and ideology (Fairclough, 1995; van Dijk, 2006; van Leeuwen, 2008; Wodak & Meyer, 2016). Applied to hostile online discourse, CDA frames accusations, insults, and delegitimizing labels as ideological tools through which users construct political identities and moral boundaries. CDA is particularly useful for examining processes of delegitimation—discourse that systematically undermines the credibility, morality, or authority of a social actor or institution—which is highly visible in Facebook reactions to the PDN breach. Building on this perspective, the present study uses the term vernacular accountability to refer to informal, bottom-up, and affectively charged evaluative discourse through which ordinary users demand responsibility from state actors in everyday digital interaction.

2.6. Integrated Analytical Framework

The study integrates these three perspectives as complementary interpretive levels rather than parallel add-ons. Forensic linguistics identifies the evidential and pragmatic properties of hostile expressions; sociolinguistics interprets the social meanings, stance, and identity work these utterances perform; CDA situates the resulting patterns within wider structures of ideology and power. Analytically, each utterance is examined sequentially: (i) lexical, pragmatic, and evidential features; (ii) stance and identity positioning; (iii) ideological framing of state, governance, and public trust.

3. Method

3.1. Research Design

This study employs a corpus-assisted qualitative discourse analysis. The qualitative descriptive component is used to describe and interpret linguistic patterns and meaning-making strategies without manipulating the research context, while the corpus-assisted component provides systematic identification of recurring lexical items, collocations, and concordance lines across the dataset (Baker et al., 2008; Gabrielatos & Baker, 2008; Gillings et al., 2023; McEnery & Hardie, 2012; Partington et al., 2013). Quantitative corpus outputs are not treated as final findings but as entry points into qualitative discourse analysis. This design is appropriate for the present study because it is transparent and replicable while allowing context-sensitive interpretation of digitally mediated hostility.

3.2. Data Source and Context

The data were drawn from publicly accessible Facebook comments responding to the breach of Indonesia's National Data Center (*Pusat Data Nasional*, PDN). The raw dataset consisted of 422 comments collected from one publicly accessible Facebook post coded in this study as FB-01. The comments were posted between 29 June 2024 and 2 July 2024. Ten comments were excluded because they lacked complete timestamp or comment-ID metadata, and a further eight entries were removed because they were duplicate records, empty entries, or content consisting solely of links or emojis. After this screening, 404 unique timestamped comments were retained as the primary corpus, comprising approximately 6,100 word tokens. Private messages and inaccessible content were also excluded. Usernames, profile links, and personally identifying details were removed during data preparation.

3.3. Data Selection and Sampling

A purposive sampling strategy was used to select one publicly accessible Facebook comment thread that explicitly responded to the PDN breach and contained substantial public discussion of the incident. Individual comments were included if they were written in Indonesian or in Indonesian-English code-mixing and were publicly visible at the time of collection. The corpus was deliberately not restricted to abusive comments; instead, all relevant comments in the selected thread were retained and subsequently coded to distinguish policy criticism, hostile evaluation, blame attribution, personal insult, delegitimation of authority, and hate-oriented expression. Empty entries, duplicates, and content consisting solely of links or emojis were excluded during data preparation.

3.4. Units of Analysis

The units of analysis comprise lexical choices (e.g., evaluative adjectives, pejorative nouns), phrasal and clausal patterns (insults, accusations, calls for dismissal), address terms and naming practices, hashtags, mocking and dehumanizing expressions, and threatening or accusatory statements. Each unit is examined at the levels of (i) form, (ii) pragmatic function, and (iii) discursive effect.

3.5. Data Analysis Procedures

Analysis proceeded in six steps: (1) data collection and screening against inclusion criteria; (2) preliminary corpus exploration using AntConc to generate frequency lists, keyword lists, and concordance lines; (3) coding of comments into categories of digital aggression and hate-oriented expression using a theory-driven codebook developed from studies on hate speech, offensive language, digital aggression, forensic linguistics, and Critical Discourse Analysis (Davidson et al., 2017; Fortuna & Nunes, 2018; Hardaker, 2010; KhosraviNik & Esposito, 2018; Wodak & Meyer, 2016); (4) forensic-linguistic interpretation of intent, target specificity, and illocutionary force; (5) sociolinguistic interpretation of stance, identity positioning, and code choice; (6) CDA interpretation of ideology, delegitimation, and power relations. Because a single comment may contain more than one aggressive feature, the coding allowed multiple labels for the same comment. Coding was conducted by the researcher; ambiguous cases were revisited several times and compared against the operational definitions adopted in the theoretical framework.

Coding decisions were documented in an audit trail, and ambiguous comments were revisited against the operational definitions before final category assignment.

3.6. Trustworthiness and Ethics

Since the study was conducted by a single researcher-coder, intercoder agreement was not calculated. Instead, analytical trustworthiness was strengthened through systematic codebook application, repeated reading of the data, corpus-assisted checking of recurring patterns, and theoretical triangulation across forensic linguistics, sociolinguistics, and Critical Discourse Analysis. Although the data are publicly accessible, ethical handling followed established guidelines for social media research (Franzke et al., 2020; Markham & Buchanan, 2012): usernames, profile links, and personally identifying details were anonymized; no private content was used; and harmful utterances are reported only as analytical examples, not amplified beyond research necessity. Direct quotations were selected carefully to minimize traceability through search engines, and where necessary minor orthographic normalization or partial masking was applied without altering the analytical meaning of the utterance.

Table 1. Codebook Categories and Operational Definitions.

Category	Operational definition	Coding criteria (lexical cues)
Policy criticism	Negative evaluation of policy or institutional performance without abusive wording	Criticizes decision, policy, or responsibility; no evaluative slur
Hostile evaluation	Negative judgment of competence, morality, or capacity	tidak becus, tidak kompeten, gagal, tidak mampu
Blame attribution	Assigning responsibility or demanding consequences	harus mundur, harus dipecat, bertanggung jawab, copot
Personal insult	Direct degradation of an individual target	brengsek, goblok, tolol, tak ada otak, memalukan
Delegitimation of authority	Undermining authority, legitimacy, or basis of appointment	balas budi, titipan, projo, bukan ahli, tidak layak
Sarcastic ridicule	Irony or mockery used to delegitimize a target	hebat sekali, mantap, bravo, salut (in ironic context)
Hate-oriented expression	Degrading or exclusionary attack on a socially identifiable group	Targets identity/group category (ethnicity, religion, etc.), not merely individual incompetence
Other aggressive forms	Aggressive utterances that do not fit the primary categories above	Includes generalized vulgarity, dehumanizing labels, or expressions of disgust not directed at a specific identity group

Note: The codebook was applied via lexicon-assisted coding using regular-expression pattern matching against the lexical cues listed above; ambiguous matches were reviewed manually. Multi-label coding was permitted because a single comment may instantiate more than one category.

4. Result and Discussion

4.1. Forms of Digital Aggression

The analysis identified several recurring forms of digital aggression in Facebook comments responding to the PDN breach. These forms include hostile evaluation, personal insult, sarcastic ridicule, blame attribution, and hate-oriented expression. Although not all hostile utterances can be classified as hate speech in the strict sense, they collectively demonstrate how public anger was linguistically expressed through evaluative, accusatory, and delegitimizing forms of discourse. Because a single comment may contain more than one aggressive feature, multi-label coding was applied. Table 2 presents the distribution of these categories. Of the 404 comments in the primary corpus, 193 (47.8%) contained at least one form of digital aggression, yielding 233 coded category instances under multi-label coding. Blame attribution emerged as by far the most frequent category, present in 113 comments (28.0% of N), followed by delegitimation of authority in 69 comments (17.1%). Personal insult (22 comments; 5.4%) and hostile evaluation (19 comments; 4.7%) appeared less often, while sarcastic ridicule (3 comments; 0.7%) and other aggressive forms (7 comments; 1.7%) were marginal. Notably, no comment in the corpus met the strict definition of hate-oriented expression, which is consistent with the conceptual decision to frame the study around digital aggression rather than hate speech.

Table 2. Forms of Digital Aggression in the Corpus.

Category	Linguistic realization	Frequency	Percentage
Hostile evaluation	evaluative adjectives and phrases such as *tidak becus*, *tidak kompeten*, *gagal*, *tidak mampu*	19	4.7
Personal insult	direct insulting expressions such as *brengsek*, *tak ada otak*, *bodoh*, *goblok*	22	5.4
Sarcastic ridicule	irony, mockery, rhetorical questions, laughter markers, exaggerated praise	3	0.7
Blame attribution	calls for responsibility, resignation, or dismissal such as *harus mundur*, *harus dipecat*, bertanggung jawab	113	28.0
Delegitimation of authority	expressions questioning competence, legitimacy, appointment, or political credibility	69	17.1
Hate-oriented expression	degrading, stigmatizing, or exclusionary labels directed at a person or collectivity	0	0.0
Other aggressive forms	other hostile utterances that do not fit the above categories	7	1.7
Total		233 instances (193 comments)	—

Note: Percentages are calculated against the full corpus size (N = 404). Because multi-label coding was applied, the total number of coded category instances exceeds the number of aggressive comments; a single comment may therefore appear under more than one category.

4.2. Representative Examples of Digital Aggression

The following representative examples are drawn directly from the coded corpus and illustrate the most salient categories of digital aggression observed in the data. Each example is presented with an anonymized comment ID, the original Indonesian text, an English gloss, the analytical category, and a brief analytical note. Comment IDs (C-001 through C-005) replace original Facebook identifiers; minor orthographic normalization (capitalization and spacing) was applied for readability without altering the semantic content of the utterances.

Example 1 (C-001): Blame Attribution

Indonesian: "Bukan cuma *menteri*nya yang layak mundur, presidennya juga *harus mundur*."

English gloss: "Not only does the minister deserve to step down; the president must step down too."

Category: Blame attribution.

Analytical note: The utterance performs blame attribution through the deontic modal harus ('must') and the evaluative predicate layak mundur ('deserves to step down'). Dismissal is framed as a necessary and morally justified consequence, and responsibility is escalated from the individual minister to the head of state.

Example 2 (C-002): Hostile Evaluation

Indonesian: "Mundur aja, *tidak berkompeten*."

English gloss: "Just step down — [you are] not competent."

Category: Hostile evaluation (with co-occurring blame attribution).

Analytical note: The directive mundur aja ('just step down') co-occurs with the evaluative predicate *tidak berkompeten* ('not competent'). The utterance attacks competence rather than only policy: the target is linguistically constructed as incapable of fulfilling institutional responsibility, and dismissal is presented as the natural consequence of that incapacity.

Example 3 (C-003): Personal Insult

Indonesian: "*Modal Ketum Projo… tak ada otak* jadi *menteri*."

English gloss: "Just on the strength of being *Projo* chairman… [he has] no brain to serve as minister."

Category: Personal insult (co-occurring with delegitimation of authority).

Analytical note: The metaphorical insult *tak ada otak* ('has no brain') attacks the target's cognitive capacity, while *modal Ketum Projo* ('only on the strength of being *Projo* chairman') frames the appointment as resting on political affiliation rather than competence. The utterance combines direct personal degradation with delegitimation of the basis of authority.

Example 4 (C-004): Hostile Evaluation through Idiomatic Imagery

Indonesian: "*Pejabat* kita kan hampir semua muka tembok, walaupun *gagal* maju terus."

English gloss: "Almost all our officials are thick-skinned (lit. 'wall-faced') — even when they fail, they keep going."

Category: Hostile evaluation.

Analytical note: The idiom muka tembok (literally 'wall-faced', i.e., shameless/thick-skinned) and the predicate *gagal* ('fail') jointly construct a generalized evaluation of officials as both incompetent and impervious to accountability. The hostility is collective rather than individual, attaching to the category *pejabat* ('officials') as a whole.

Example 5 (C-005): Delegitimation of Authority

Indonesian: "Karena tim *Projo*, jadinya ya *balas budi*, ya begini, gimana *negara* nggak amburadul."

English gloss: "Because [he is] from the *Projo* team, it's political payback — that's why; how could the state not fall apart?"

Category: Delegitimation of authority (co-occurring with hostile evaluation of the state).

Analytical note: The phrase *balas budi* ('political payback') frames the appointment as patronage rather than merit, while the predicate *negara amburadul* ('the state falls apart') extends the evaluative scope from the individual official to the state itself. The utterance delegitimizes the official by linking the breach to a broader ideology of political patronage and institutional decay.

In this study, hate-oriented expression is distinguished from general offensive language. Profanity or personal insult alone is not automatically classified as hate speech unless it invokes a stigmatized social identity, collective target, or exclusionary framing. The category is therefore reserved for utterances that meet this stricter threshold.

4.3. Target Construction and Blame Attribution

The second pattern concerns how targets of hostility were constructed in the comment threads. The corpus shows that aggression was not randomly distributed; it was directed toward specific social actors and institutions associated with the PDN breach. These targets include individual public officials, government institutions, political groups, and the state apparatus more broadly. Table 3 summarizes the main target categories identified in the data.

Table 3. Target Construction in the Corpus.

Target category	Linguistic reference	Frequency	Percentage	Discursive function
Individual public official	*Menkominfo*, *menteri*, personal name, pronouns such as dia	84	20.8	Personalization of blame
Government institution	*Kominfo*, *pemerintah*, *negara*	46	11.4	Institutional blame
Political actor/group	party labels, volunteer groups, political supporters	53	13.1	Political delegitimation
Bureaucratic/systemic target	*sistem*, *birokrasi*, *pejabat*, *penguasa*	41	10.1	Generalized distrust
Citizens/public	*rakyat*, *warga*, *masyarakat*	31	7.7	Victim positioning
Unspecified target	implicit or context-dependent reference	214	53.0	Diffuse anger

Note: Target categories were coded using a multi-label procedure because a single comment could refer simultaneously to an individual official, an institution, and a broader political group. Percentages are calculated against the total corpus size (N = 404) and therefore do not sum to 100%.

The most prominent identifiable target of hostility was the individual public official, particularly the figure of the Minister of Communication and Informatics, who appeared as the most frequently named referent in 84 comments (20.8% of N). Government institutions (*Kominfo*, *pemerintah*, *negara*) were referenced as targets in 46 comments (11.4%), while political actors and groups (most often Projo and relawan) appeared in 53 comments (13.1%) and bureaucratic or systemic targets (*sistem*, *birokrasi*, *pejabat*) in 41 comments (10.1%). Citizens were positioned as victim-referents in 31 comments (7.7%). A substantial 214 comments (53.0%) had no explicitly named target and relied on contextual deixis, indicating that anger often circulated as diffuse hostility rather than precise accusation. When the target was named, public anger was primarily personalized through references to a single accountable actor, even when responsibility could be more diffusely attributed to institutional or systemic failure. Naming practices such as *Menkominfo* and *menteri* functioned to make responsibility explicit, while pronouns such as dia and demonstrative phrases such as *pejabat* seperti ini allowed commenters to maintain shared reference without repeating the target's full name. Taken together, these patterns suggest that blame attribution operated simultaneously through direct naming, pronominal deixis, and generalized institutional reference, with the individual official functioning as the symbolic focal point even when systemic critique was implied. Having identified the dominant forms of digital aggression and the targets toward which they were directed, the next section interprets these patterns from a forensic-linguistic perspective.

4.4. From Criticism to Aggression: A Forensic-Linguistic Reading

From a forensic-linguistic perspective, the comments can be placed along a continuum of hostility, ranging from legitimate policy criticism to personal insult and hate-oriented expression. This continuum is useful because not every negative utterance constitutes hate speech or legally relevant abuse. Instead, the severity of an utterance depends on its target specificity, lexical intensity, illocutionary force, and potential harm.

Table 4. Continuum of Hostile Online Discourse.

Level	Discourse type	Linguistic indicators	Frequency	Analytical status
1	Policy criticism	criticism of policy, failure, responsibility	211 comments (52.2%)	Legitimate criticism
2	Hostile evaluation	negative competence/morality labels	10 comments (2.5%)	Digital aggression
3	Blame/delegitimation	calls for resignation, dismissal, responsibility, patronage accusations, and delegitimation of authority	162 comments (40.1%)	Accountability-driven digital aggression
4	Personal insult	profanity, cognitive/moral degradation	21 comments (5.2%)	Strong digital aggression
5	Threatening expression	threats, intimidation, harm-oriented utterances	0 comments (0.0%) — none observed	Potentially legally relevant
6	Hate-oriented expression	group-based degrading or exclusionary language	0 comments (0.0%) — none observed	Potential hate speech

Unlike Table 2, which reports multi-label category instances, Table 4 assigns each comment to the highest level of hostility observed in that comment. Level 3 is not identical to the "blame attribution" category in Table 2; rather, it groups accountability-oriented accusations, demands for consequences, delegitimation of authority, and patronage-based criticism when these constitute the highest level of hostility in a comment. The frequencies in Table 4 are therefore not expected to match the category totals reported in Table 2. For example, the 19 comments coded as hostile evaluation in Table 2 are reduced to 10 in Table 4 because 9 of them co-occurred with a higher-level category (personal insult or blame/delegitimation) and were assigned to that higher level under the highest-level rule.

The continuum shows that the corpus contains different degrees of hostility. The utterance "*Menkominfo harus dipecat*" is not hate speech in the strict sense, but it performs a directive and accusatory function: it identifies a specific institutional actor, attributes responsibility, and formulates dismissal as a morally justified consequence. By contrast, expressions such as *tak ada otak* and *brengsek* shift from accountability discourse into personal insult, where the target is delegitimized through moral and cognitive degradation. The forensic relevance of these utterances lies not only in the presence of offensive words but also in their pragmatic function: whether they accuse, insult, threaten, incite, or delegitimize a specific target.

This distinction is crucial because offensive language, digital aggression, and hate speech should not be treated as identical categories. The forensic-linguistic task is to identify what the utterance does, who it targets, and how severe its communicative force is in context. Within the present corpus, the distribution along the continuum is highly skewed toward the lower-to-middle levels: when each comment is assigned its highest level present, 211 comments (52.2%) remain at Level 1 (policy criticism or non-aggressive content), 10 (2.5%) at Level 2 (hostile evaluation), 162 (40.1%) at Level 3 (blame/delegitimation), and 21 (5.2%) at Level 4 (personal insult). No comment in the corpus reached Level 5 (threatening expression) or Level 6 (hate-oriented expression). This distribution empirically supports the conceptual decision to frame the study around digital aggression rather than hate speech: hostility in PDN-related Facebook discourse is dominated by accountability-driven blame and delegitimation (Level 3) and accompanied by a meaningful but minority stream of personal insult, while strict hate speech and explicit threats are absent from the dataset.

4.5. Critical Discourse Analysis: State Competence, Accountability, and Public Distrust

The hostile expressions identified in the corpus are not merely individual emotional reactions. From a Critical Discourse Analysis perspective, they reflect broader ideological framings of state competence, accountability, and public distrust. The PDN breach was discursively constructed not only as a technical failure but also as a moral and political failure of governance.

Table 5. Ideological Framings in the Corpus.

Ideological framing	Linguistic realization	Example	Interpretation
State incompetence	*tidak becus*, *gagal*, *tidak mampu*	"*Pejabat seperti ini tidak becus* mengurus data *negara*."	The state is framed as technically and administratively incapable
Moral failure	*tidak bertanggung jawab*, *memalukan*	"*Memalukan*, sudah dipercaya tapi *gagal* jaga *data rakyat*."	The breach is framed as a failure of responsibility
Political patronage	*balas budi*, *bukan ahli*, *titipan politik*	"Dia diangkat bukan karena kemampuan, tapi karena *balas budi politik*."	Appointment is framed as politically motivated
Public victimhood	*data rakyat*, *warga dirugikan*, *masyarakat jadi korban*	"*Data rakyat* dijual murah, kami jadi korban."	Citizens are positioned as victims of institutional failure
Delegitimation of authority	*harus mundur*, *harus dipecat*	"*Menkominfo harus dipecat*."	Officials are constructed as unfit to govern

The repeated use of expressions such as *tidak becus*, *gagal*, *balas budi politik*, *harus dipecat*, and *data rakyat* constructs the state as an institution that failed to protect citizens' data. This framing transforms the PDN breach from a cybersecurity incident into a symbolic crisis of trust. Through hostile evaluation and blame attribution, commenters positioned themselves as morally entitled to judge state actors. In this sense, digital aggression functions as a form of vernacular accountability. By this term I refer to informal, bottom-up, and often emotionally charged forms of public evaluation through which ordinary users demand responsibility from state actors in everyday digital language, outside the formal channels of audit, parliamentary oversight, or legal procedure.

This vernacular accountability is, however, ambivalent. On one hand, it reflects democratic participation and a public demand for responsibility in the absence of fully effective institutional channels. On the other hand, it can escalate into personal insult, symbolic degradation, and hate-oriented discourse. The data therefore show how online political participation may move from critique to aggression when institutional distrust is linguistically intensified, particularly in the heightened affective context of a national cyber crisis.

4.6. Comparison with Previous Studies and Implications

The findings can be situated against four strands of prior work. First, unlike identity-based hate-speech studies that foreground ethnicity, religion, gender, or partisan identity as primary targets (Davidson et al., 2017; Fortuna & Nunes, 2018; Poletto et al., 2021; Schmidt & Wiegand, 2017), the present corpus contains no comments meeting the strict operational definition of hate-oriented expression; hostility instead clusters around accountability claims directed at named officeholders and institutions. Second, the findings are nonetheless consistent with the broader literature on digital aggression and online incivility (Hardaker, 2010; Jane, 2015; Runions & Bak, 2015; Suler, 2004), in which hostile evaluation and personal insult appear as recurring features even when identity-based hate is absent. Third, compared with research on Indonesian online political discourse, which has tended to centre on electoral polarization and partisan hostility (Ibrohim & Budi, 2023; Lim, 2017; Tahir & Ramadhan, 2024), the present analysis foregrounds cyber-crisis discourse as a distinct discursive trigger: institutional failure generates accountability-driven aggression that is structurally similar to identity-based hostility but is directed at officeholders rather than identity groups. Fourth, the study contributes a continuum-based reading to forensic-linguistic work on online conflict (Coulthard et al., 2017; Gibbons & Turell, 2008; KhosraviNik & Esposito, 2018), showing that the diagnostic question is not simply whether an utterance is hateful but where on the criticism-to-aggression continuum it falls and what target it constructs. Across these comparisons, the continuum framework developed here complements rather than replaces existing hate-speech and aggression frameworks.

Beyond their theoretical reach, the findings carry practical weight for digital literacy and critical language awareness in English-language education, where students increasingly operate as bilingual digital citizens (Darvin, 2017; Darvin & Hafner, 2022; Hafner et al., 2015). Distinguishing between criticism, hostile evaluation, insult, and hate-oriented expression equips learners to interpret and produce online discourse responsibly. The findings also speak to platform governance and the legal-linguistic evaluation of online abuse during national cyber crises, where forensic-linguistic categories can support more proportionate moderation than blanket appeals to "hate speech."

5. Conclusion

This study examined how Indonesian Facebook users linguistically constructed digital aggression in response to the PDN breach and whether such hostility crossed the threshold of hate-oriented expression. The findings show that hostile discourse was realized through evaluative adjectives, pejorative labels, insults, sarcastic ridicule, calls for dismissal, and blame-oriented constructions. Public officials and state institutions emerged as the dominant targets, with responsibility constructed through direct naming, deontic expressions, and morally charged evaluations. These expressions did more than vent spontaneous anger; they fed into broader ideological framings of state incompetence, institutional failure, and public distrust.

For forensic linguistics and critical discourse studies, the central contribution is showing that hostile online discourse in cyber-crisis contexts unfolds along a continuum running from legitimate policy criticism through personal insult and digital aggression to hate-oriented expression. This distinction is analytically important because not all offensive or aggressive language qualifies as hate speech in the strict sense. The study also highlights the relevance of corpus-assisted qualitative discourse analysis for identifying recurrent patterns while preserving contextual interpretation, and introduces vernacular accountability as a useful concept for understanding how citizens deploy informal and sometimes abusive language to demand institutional responsibility.

The study is limited by its focus on a single Facebook post, its short sampling window, the modest corpus size, and the difficulty of drawing sharp boundaries between criticism, aggression, and hate speech in rapidly evolving online interaction. The findings should therefore be understood as an in-depth analysis of one selected Facebook

comment thread rather than as a representative account of all Indonesian Facebook reactions to the PDN breach. Coding by a single researcher, rather than two independent coders, also limits the scope of formal reliability assessment, although trustworthiness was supported through repeated coding, codebook consistency, and theoretical triangulation.

Future studies may compare Facebook discourse with X, TikTok, YouTube, or news-comment platforms; examine multimodal forms of hostility that combine image and text; develop comparative forensic-linguistic models of cyber-crisis discourse across Southeast Asia; and integrate machine-assisted corpus analysis with qualitative discourse interpretation. For English-language education research, the framework developed here can also be extended to studies of critical digital literacy and bilingual online behavior among Indonesian learners.

Acknowledgments:

The author thanks colleagues at the Doctoral Program in Linguistics, Faculty of Humanities, Universitas Udayana, for constructive academic discussions during the development of this study. No external funding was received for this research.

References

Androutsopoulos, J. (2014). Languaging when contexts collapse: Audience design in social networking. *Discourse, Context & Media, 4–5*, 62–73. https://doi.org/10.1016/j.dcm.2014.08.006

Austin, L., Fisher Liu, B., & Jin, Y. (2012). How audiences seek out crisis information: Exploring the social-mediated crisis communication model. *Journal of Applied Communication Research, 40*(2), 188–207. https://doi.org/10.1080/00909882.2012.654498

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society, 19*(3), 273–306. https://doi.org/10.1177/0957926508088962

Benu, N. N., Sulasmini, N. M. A., & Supartini, N. L. (2025). Sexist humor in TikTok: Content strategies and the normalization of patriarchy in digital spaces. *International Journal Online of Humanities, 11*(3), 79–106. https://ijohmn.com/index.php/ijohmn/article/view/314/739

Brown, A. (2017). What is hate speech? Part 1: The myth of hate. *Law and Philosophy, 36*(4), 419–468. https://doi.org/10.1007/s10982-017-9297-1

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies, 7*(4–5), 585–614. https://doi.org/10.1177/1461445605054407

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of Situational Crisis Communication Theory. *Corporate Reputation Review, 10*(3), 163–176. https://doi.org/10.1057/palgrave.crr.1550049

Coombs, W. T., & Holladay, S. J. (2002). Helping crisis managers protect reputational assets: Initial tests of the Situational Crisis Communication Theory. *Management Communication Quarterly, 16*(2), 165–186. https://doi.org/10.1177/089331802237233

Coulthard, M., Johnson, A., & Wright, D. (2017). *An introduction to forensic linguistics: Language in evidence* (2nd ed.). Routledge. https://doi.org/10.4324/9781315630311

Darvin, R. (2017). Language, ideology, and critical digital literacy. In S. L. Thorne & S. May (Eds.), *Language, education and technology* (pp. 17–30). Springer. https://doi.org/10.1007/978-3-319-02237-6_35

Darvin, R., & Hafner, C. A. (2022). Digital literacies in TESOL: Mapping out the terrain. *TESOL Quarterly, 56*(3), 865–882. https://doi.org/10.1002/tesq.3161

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media, 11*(1), 512–515. https://doi.org/10.1609/icwsm.v11i1.14955

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics, 12*(4), 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. Longman.

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys, 51*(4), Article 85. https://doi.org/10.1145/3232676

Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C. M., & Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Association of Internet Researchers. https://aoir.org/reports/ethics3.pdf

Gabrielatos, C., & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996–2005. *Journal of English Linguistics, 36*(1), 5–38. https://doi.org/10.1177/0075424207311247

Gibbons, J., & Turell, M. T. (Eds.). (2008). *Dimensions of forensic linguistics*. John Benjamins. https://doi.org/10.1075/aals.5

Gillings, M., Mautner, G., & Baker, P. (2023). *Corpus-assisted discourse studies*. Cambridge University Press. https://doi.org/10.1017/9781009168144

Hafner, C. A., Chik, A., & Jones, R. H. (2015). Digital literacies and language learning. *Language Learning & Technology, 19*(3), 1–7. https://www.lltjournal.org/item/10125-44426/

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research, 6*(2), 215–242. https://doi.org/10.1515/jplr.2010.011

Ibrohim, M. O., & Budi, I. (2023). Hate speech and abusive language detection in Indonesian social media: Progress and challenges. *Heliyon, 9*(8), e18647. https://doi.org/10.1016/j.heliyon.2023.e18647

Jaffe, A. (Ed.). (2009). *Stance: Sociolinguistic perspectives*. Oxford University Press.

Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethics and Information Technology, 17*(1), 65–87. https://doi.org/10.1007/s10676-015-9362-0

KhosraviNik, M., & Esposito, E. (2018). Online hate, digital discourse and critique: Exploring digitally mediated discursive practices of gender-based hostility. *Lodz Papers in Pragmatics, 14*(1), 45–68. https://doi.org/10.1515/lpp-2018-0003

Kompas. (2024, June 24). *BSSN: Gangguan Pusat Data Nasional disebabkan serangan siber "ransomware" [BSSN: National Data Center disruption caused by "ransomware" cyberattack]*. *Kompas.com*. https://nasional.kompas.com/read/2024/06/24/15413051/bssn-gangguan-pusat-data-nasional-disebabkan-serangan-siber-ransomware

Knight, R., & Nurse, J. R. C. (2020). A framework for effective corporate communication after cyber security incidents. *Computers & Security, 99*, Article 102036. https://doi.org/10.1016/j.cose.2020.102036

Kuipers, S., & Schönheit, M. (2022). Data breaches and effective crisis communication: A comparative analysis of corporate reputational crises. *Corporate Reputation Review, 25*(3), 176–197. https://doi.org/10.1057/s41299-021-00121-9

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

Markham, A., & Buchanan, E. (2012). *Ethical decision-making and internet research: Recommendations from the AoIR Ethics Working Committee* (Version 2.0). Association of Internet Researchers. https://aoir.org/reports/ethics2.pdf

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Olsson, J., & Luchjenbroers, J. (2014). *Forensic linguistics* (3rd ed.). Bloomsbury Academic.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies*. John Benjamins. https://doi.org/10.1075/scl.55

Poletto, F., Basile, V., Sanguinetti, M., Bosco, C., & Patti, V. (2021). Resources and benchmark corpora for hate speech detection: A systematic review. *Language Resources and Evaluation, 55*(2), 477–523. https://doi.org/10.1007/s10579-020-09502-8

Reuters. (2024, June 24). *Cyber attack compromised Indonesia data centre, ransom sought*. Reuters. https://www.reuters.com/technology/cybersecurity/cyber-attack-compromised-indonesia-data-centre-ransom-sought-2024-06-24/

Runions, K. C., & Bak, M. (2015). Online moral disengagement, cyberbullying, and cyber-aggression. *Cyberpsychology, Behavior, and Social Networking, 18*(7), 400–405. https://doi.org/10.1089/cyber.2014.0670

Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. https://doi.org/10.18653/v1/W17-1101

Shuy, R. W. (1993). *Language crimes: The use and abuse of language evidence in the courtroom*. Blackwell.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321–326. https://doi.org/10.1089/1094931041291295

Tagg, C. (2015). *Exploring digital communication: Language in action*. Routledge. https://doi.org/10.4324/9781315727165

Tahir, I., & Ramadhan, M. G. F. (2024). Hate speech on social media: Indonesian netizens' hate comments of presidential talk shows on YouTube. *LLT Journal: A Journal on Language and Language Teaching, 27*(1), Article 8180. https://doi.org/10.24071/llt.v27i1.8180

van Dijk, T. A. (2006). Discourse and manipulation. *Discourse & Society, 17*(3), 359–383. https://doi.org/10.1177/0957926506060250

van Leeuwen, T. (2008). *Discourse and practice: New tools for critical discourse analysis*. Oxford University Press.

Waldron, J. (2012). *The harm in hate speech*. Harvard University Press.

Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. *Proceedings of the NAACL Student Research Workshop*, 88–93. https://doi.org/10.18653/v1/N16-2013

Wodak, R., & Meyer, M. (Eds.). (2016). *Methods of critical discourse studies* (3rd ed.). SAGE.