

Hate as evaluation: A discourse analysis of language and emotion in YouTube political comments

Khusnul Khotimah^{a, 1,*}, Achamad Fanani^b

^a Universitas Trunojoyo Madura, Madura, Indonesia

^b Universitas Pesantren Tinggi Darul Ulum, Jombang, Indonesia

khusnul.khotimah@trunojoyo.ac.id

* Correspondent author

Received: October 25, 2025

Revised: January 20, 2026

Accepted: February 3, 2026

KEYWORDS

Appraisal Theory

Digital Discourse

Hate Speech

Moral Judgment

ABSTRACT

This study examines the linguistic and psychological dynamics of hate speech in Indonesian political discourse on YouTube by analyzing user comments on a CNN Indonesia talk show discussing Gibran Rakabuming Raka. Using a qualitative approach grounded in appraisal-based discourse analysis, the study analyzes 1,534 comments to explore how evaluative language constructs moral conflict in online political communication. The findings reveal that hate speech is predominantly expressed through negative judgment, which accounts for 45% of the data, followed by contractive engagement strategies that restrict dialogue (44%) and intensified graduation that amplifies emotional force (12%). These linguistic patterns indicate that political disagreement is frequently framed as moral condemnation rather than rational debate. The study further demonstrates that such evaluative strategies are closely linked to psychological processes, including moral conviction, group identity signaling, and emotional contagion, which collectively intensify polarization in digital spaces. By integrating linguistic analysis with psychological perspectives, this research highlights hate speech as a form of moralized discourse that escalates affective conflict. The study concludes by recommending digital literacy interventions that extend beyond content moderation to address underlying evaluative and emotional processes, with the aim of fostering empathy, critical reflection, and more constructive online political engagement.

© 2025 The Author(s). Published by Universitas Ahmad Dahlan.
This is an open-access article under the [CC-BY-SA](#) license.



Introduction

In today's digital era, social media platforms have become major spaces where people express political and ideological opinions (Bobro, 2024; Ariestandy et al., 2024). Among these platforms, YouTube comment sections stand out as interactive arenas where viewers not only respond to video content but also exchange opinions, emotions, and evaluations with others. This is especially evident in politically charged talk shows, where participants, audiences, and online commentators all use language to evaluate public figures, events, and ideas. Unlike traditional media, this digital space allows instant and unfiltered responses, which often intensify polarization, incivility, and the spread of hate speech.

Such interactions illustrate how digital spaces enable strong emotional and evaluative expressions that sometimes go beyond the boundaries of ethical public discourse (Serrano-Puche et al., 2021). Scholars have linked this tendency to rising social polarization and declining digital literacy among Indonesian internet users (Afrina et al., 2024). In linguistic studies, the Appraisal framework has been widely used to analyze

online discourse because it explains how language constructs attitudes, stances, and the strength of interpersonal meaning (Dam & Jensen, 2024).

In the Indonesian context, the growing amount of online hate speech directed at individuals, groups, or ideas has become a serious concern for digital citizenship, content moderation, and democratic culture. A prominent example is the talk show episode “Ijazah Digugat, Legitimasi Diragukan, Jalan Baru Pemakzulan Gibran?” broadcast by CNN Indonesia (October 31, 2025). The episode triggered thousands of YouTube comments, many of which contained derogatory, aggressive, and evaluative language targeting public figures—particularly Gibran Rakabuming Raka—as well as other actors labeled in pejorative ways (e.g., *termul*). These comments reflect typical hate speech patterns such as insults, degrading metaphors, calls for punishment, and rhetorical devices that amplify negativity. They reveal how a seemingly casual comment space can become a site of ideological conflict and interpersonal hostility, shaping the quality of public discourse.

This episode serves as a pertinent case study due to its direct engagement with public legitimacy, a theme that provokes strong evaluative responses. We argue that the episode’s comments provide a concentrated lens to examine how resources of Attitude, Engagement, and Graduation (Martin & White, 2005; Khotimah, 2024) function not merely as isolated insults but as coordinated systems for constructing moral condemnation, enforcing ideological closure, and facilitating emotional contagion. By integrating this linguistic framework with psychological discourse analysis, this research moves beyond description to explain the mechanisms through which digital disagreement is transformed into affectively charged hostility, thereby shaping the very tenor of public discourse.

Previous studies have documented the prevalence and forms of online hate speech in Indonesia. Tahir & Ramadhan (2024) found that during discussions of the 2024 presidential election, the most frequent hate speech types were “early warnings” (43%), followed by dehumanization (21%), incitement or violent language (19%), and general offensiveness (17%). Similarly, Jamilah & Wahyuni (2020) reported that comments during the 2019 election were dominated by sarcasm, profanity, ridicule, and mocking nicknames. Križan & Barbič (2023) used a forensic linguistic approach to show that defamation and insult were the most common forms, while Kazmi et al. (2021) linked Islamophobic hate speech to identity politics and majority-minority tensions.

Despite these valuable contributions, existing studies (e.g., Reichelmann et al., 2020; Walther, 2022) have primarily focused on categorizing hate speech based on thematic content or surface-level linguistic features (e.g., insults, sarcasm, dehumanization). Few have systematically examined *how* evaluative meaning is constructed through nuanced linguistic resources in real-time political discourse. Specifically, there is a lack of research that applies a comprehensive linguistic framework—such as Appraisal Theory—to analyze how attitude, engagement, and graduation interact to shape moral judgment, emotional intensity, and dialogic positioning in Indonesian YouTube political comments. This gap limits our understanding of hate speech not merely as a set of offensive utterances, but as a structured evaluative discourse that reinforces ideological polarization and emotional hostility.

Thus, the research problem this study addresses is: *How is hate speech linguistically realized and evaluatively constructed in Indonesian digital political discourse, particularly through the interplay of attitude, engagement, and graduation resources?* This problem leads to the following research questions: (1) What attitudinal resources (affect, judgment, appreciation) are employed in YouTube political comments to express evaluation? (2) How do engagement strategies (contractive/expansive) shape the dialogic stance of commenters? (3) In what ways does graduation (force/focus) intensify or mitigate evaluative meaning in hate speech?

Within evaluative discourse analysis, the Appraisal Theory developed by Martin & White (2005) provides a robust and nuanced framework for examining how language expresses interpersonal meaning, evaluative stance, and emotional positioning. Appraisal Theory is grounded in Systemic Functional Linguistics (SFL) and is specifically designed to analyze how speakers and writers adopt attitudes, negotiate viewpoints, and modulate the intensity of their evaluations. The framework consists of three interconnected systems: (1) *Attitude*, which encompasses emotional responses (*Affect*), ethical evaluations of behavior (*Judgment*), and aesthetic assessments of phenomena (*Appreciation*). (2) *Engagement*, which examines how speakers position themselves dialogically—either opening space for alternative viewpoints (*expansive*) or closing it down (*contractive*). (3) *Graduation*, which deals with the scaling of evaluative meaning in terms of intensity (*Force*) or sharpness of categorical boundaries (*Focus*). Together, these systems offer a holistic lens through which to analyze how evaluative language constructs solidarity, conflict, moral alignment, and emotional resonance in discourse.

While many Indonesian studies have examined hate speech in terms of frequency, themes, or surface-level categories, few have employed Appraisal Theory to investigate the detailed linguistic mechanisms that construct evaluative meaning in real-time, politically charged digital environments such as YouTube talk-

show comments. This study adopts Appraisal Theory not only as a theoretical foundation but also as a primary analytical tool to uncover how hate speech is realized through coordinated patterns of judgment, dialogic closure, and emotional intensification. By doing so, it moves beyond descriptive categorizations to explain how hate speech functions as a form of moral and affective discourse that shapes public sentiment and ideological division.

To systematically examine these linguistic and evaluative dynamics, the following section outlines the methodological approach employed in this study. Guided by Appraisal Theory, a qualitative descriptive analysis was conducted on a corpus of 1,534 YouTube comments from the specified episode. The method section details the procedures for data collection, sampling, and the analytical coding framework based on the Attitude, Engagement, and Graduation systems. This methodological foundation enables a fine-grained exploration of how hate speech is constructed, intensified, and positioned within Indonesia's digital political discourse.

Method

This study employed a qualitative descriptive approach based on the principles of Appraisal Theory developed by Martin & White (2005). A qualitative design was chosen because the goal was not to count words statistically, but to explore how linguistic resources express evaluative and emotional meanings in online discourse. Through this approach, the research aims to understand how the three systems of the Appraisal framework—Attitude, Engagement, and Graduation—work together in YouTube comments to construct and intensify hate speech toward individuals and groups. The descriptive framework allows for a deeper interpretation of linguistic patterns, emotional tones, and interpersonal positioning in a politically charged digital environment.

The data were collected from the YouTube comment section of the talk show episode "Ijazah Digugat, Legitimasi Diragukan, Jalan Baru Pemakzulan Gibran? | Head to Head With Elvira", uploaded on the official CNN Indonesia YouTube channel. This episode, streamed on October 31, 2025, discusses a sensitive political issue concerning the legitimacy of a high school diploma allegedly used by Gibran Rakabuming Raka during his vice-presidential candidacy. The video attracted significant public attention, generating over 1,500 comments that reflect intense emotional and ideological reactions.

This channel and episode were chosen because CNN Indonesia is a major mainstream news outlet that often becomes a forum for national political debate. Its talk-show format encourages audience participation through comments, making it a rich site for examining evaluative and emotional language use in real-time political discussions.

From the total dataset, 1,534 comments were identified and classified into two main linguistic categories—sentences and phrases—based on their grammatical structure. Among these, comments containing clear evaluative meaning were analyzed using the Appraisal framework. The selection followed three purposive sampling criteria:

1. The comment explicitly contained evaluative language (positive or negative);
2. The comment reflected personal or group-based attacks, mockery, or emotional aggression;
3. The comment was written in Indonesian and publicly accessible.

Comments that were purely informational or neutral were excluded. This purposive sampling ensured that the dataset captured rich linguistic patterns and meaningful evaluative expressions relevant to hate speech discourse.

To ensure methodological rigor, data analysis was conducted through a systematic, three-stage process guided by the Appraisal Theory framework (Martin & White, 2005). The procedure began with the unitization of 1,534 YouTube comments, which were treated as complete texts. However, for precise coding, the evaluative meaning within each comment was parsed into clauses, allowing for the identification of distinct Attitude, Engagement, and Graduation resources within a single post. This clause-level analysis was crucial, as individual comments often contained multiple and complex evaluative moves.

The first stage involved manual coding using the three core subsystems of Appraisal Theory. The Attitude system was applied to code the core evaluative content, classifying expressions into Affect (emotional reactions), Judgment (evaluations of human behavior and character), and Appreciation (evaluations of objects and phenomena). Simultaneously, the Engagement system was used to analyze the writer's dialogic positioning, categorizing language as either expansive (entertaining or acknowledging alternative views) or contractive (disclaiming, proclaiming, or shutting down dialogue). Finally, the Graduation system scaled the intensity of these evaluations, assessing Force (the amplification or diminishment of meaning) and Focus (the sharpening or softening of categorical boundaries).

The analysis proceeded through several stages:

1. All comments were read repeatedly to understand their emotional tone and evaluative tendency.

2. Comments were divided into evaluation units (phrases or clauses expressing attitude).
3. Each segment was manually coded according to the three Appraisal subsystems.
4. Codes were cross-checked to maintain consistency and reduce researcher bias.
5. Frequency and co-occurrence patterns of Appraisal resources were observed to identify dominant evaluative tendencies.

Representative comments were presented in tables showing Appraisal categories and interpretive explanations.

As the main research instrument, the researcher acted as the interpreter of linguistic meaning. All data were taken from publicly accessible YouTube comments, and no private or personal information was disclosed. The comments are used solely for academic purposes and are presented in their original Indonesian form to preserve authenticity and cultural nuance.

Ethical protocols were strictly followed in this study. Publicly accessible YouTube comments were collected and analyzed anonymously, with no personally identifiable information retained. Data security was maintained through password-protected storage. Ethical adherence was guided by digital research frameworks such as the AoIR guidelines, respecting platform terms of service as a form of implied consent for public content. Quoted comments are used responsibly and solely for academic analysis.

Results and Discussion

The Appraisal analysis of 1,534 YouTube comments revealed rich evaluative expressions dominated by moral judgment and negative emotional tone. The three main systems of the Appraisal framework—Attitude, Graduation, and Engagement—were identified in varying proportions, providing insights into how online users express ideological stances, personal emotions, and moral evaluations in a politically charged discussion.

Distribution of Attitudinal Resources

The first layer of analysis examined the Attitude system, which captures how users express feelings (Affect), moral evaluations (Judgment), and assessments of events or phenomena (Appreciation). As shown in Table 1, *Judgment* was the most frequent category, representing 45% of all evaluative instances, followed by *Affect* (27%) and *Appreciation* (21%). A small number of comments (7%) were categorized as non-evaluative or neutral expressions.

Table 1. Distribution of Attitudinal Resources in YouTube Comments

Type of Appraisal	Subcategory	Number of Data	Percentage	Example Comment	Interpretation
Affect	Positive/Negative emotions	412	27%	<i>Mual rasanya ngeliat lagu2an si Faldo</i> (I feel sick watching Faldo's songs.) <i>Termul gak punya etika dan moral kebenaran</i> (Termul has no ethics or moral integrity toward the truth.)	Direct emotional reaction toward an object; conveys negative disgust.
Judgment	Moral or behavioral evaluation	693	45%	<i>Bagusnya Gibran langsung yang jelasin bukan relawannya</i> (It's better that Gibran gave the explanation himself rather than through his volunteers.)	Evaluates personal character or morality according to social norms.
Appreciation	Evaluation of actions or events	327	21%	<i>Anak muda Solo.</i> (Young people of Solo.)	Evaluates the quality or appropriateness of an action.
Unclassified (-)	-	102	7%		Nominal phrase without explicit evaluative meaning.

These results indicate that moral judgment is the dominant linguistic resource in the dataset. Many commenters expressed evaluative stances toward individuals—especially political figures such as Gibran

Rakabuming Raka and other actors mentioned in the discussion—rather than evaluating ideas or issues objectively. Comments often focused on moral integrity, honesty, and ethics, suggesting that users framed the controversy around legitimacy and trustworthiness in moral rather than factual terms. This aligns with the social tendency in Indonesian online discourse to moralize political debates, where public figures are frequently judged based on perceived virtue or corruption rather than policy content.

The prevalence of *Judgment* also corresponds with the contractive engagement pattern (see Table 4 below), which shows that commenters typically closed off alternative perspectives, presenting their moral evaluations as absolute truths. This linguistic behavior reflects ideological polarization and moral superiority—both characteristic features of online hate speech.

The *Affect* category (27%) primarily consisted of negative emotional expressions, such as disgust, anger, and frustration. Phrases like “*mual rasanya*” (It feels nauseating), “*kesal banget*” (So annoyed) and “*amit-amit*” (Heaven forbid) reveal an intense emotional involvement with the topic. A smaller number of positive emotions (e.g., “*Hidup peneliti!*” (Long live the researchers!)) expressed support or solidarity, though these were comparatively rare. Such emotional intensification shows that users respond not only cognitively but also affectively to political controversies, framing their comments as emotional performances.

Meanwhile, *Appreciation* (21%) appeared when users evaluated actions, speech acts, or perceived behavior rather than personal character. Comments such as “*Bagusnya Gibran langsung yang jelasin*” (It's better that Gibran gave the explanation himself rather than through his volunteers.) focused on the appropriateness or effectiveness of an act. This shows a limited but important form of rational evaluation—an attempt to discuss the situation rather than the person. However, even these appreciative remarks often contained implicit ideological bias, indicating alignment or disalignment with the speaker's political side.

The remaining 7% of comments were categorized as non-evaluative, consisting mostly of short nominal phrases (e.g., “*Kelompok termul*” (Termul group)) or factual statements lacking emotional or moral stance. Their small proportion indicates that the YouTube comment section largely functions as a space of evaluative and emotional discourse, not neutral discussion.

Valence of Evaluative Meaning

The analysis of valence polarity (positive, negative, or neutral) further revealed the dominant emotional orientation in the dataset. As presented in Table 2, negative valence accounted for 61% of all comments, while positive and neutral stances represented 32% and 7%, respectively.

Table 2. Distribution of Valence

Valence Type	Number	Percentage	Example Comment	Interpretation
Positive	498	32%	<i>Hidup peneliti!</i>	Expression of approval or support; conveys enthusiasm.
Negative	962	61%	<i>Termul tdk punya malu</i>	Criticism, insult, or moral attack directed at individuals.
Neutral	74	7%	<i>Kelompok termul</i>	No clear emotional or evaluative stance.

Here's a comprehensive guide to citing sources. The dominance of negative valence highlights how YouTube's political comment space tends to amplify anger and hostility. Negative emotions and moral condemnations create an environment conducive to hate speech. Many comments not only criticized but also delegitimized public figures, using sarcasm, exaggeration, and moral labeling. Positive comments, though present, were often reactive—expressed as support for those attacked or as defense against perceived injustice. Neutral statements were rare and typically limited to information-sharing or clarification.

This pattern reflects a broader discourse culture where digital audiences engage with political content through emotionally charged and polarized evaluations, rather than balanced discussion. It also confirms that negative stance-taking functions as a strategy for identity performance and group alignment, strengthening in-group solidarity by attacking out-groups.

Graduation: Intensity of Evaluative Force

Graduation refers to how speakers amplify or soften their evaluative statements. In this dataset, 85% of comments were neutral in intensity, 12% contained clear intensification, and 3% involved mitigation can be see Table 3.

Table 3. Distribution of Graduation

Graduation Type	Number	Percentage	Example Comment	Interpretation
Intensification	185	12%	<i>Roy semakin sakit amit-amit</i> (Roy is getting sicker—God forbid.)	Emotion strengthened through repetition or hyperbole.
Mitigation	41	3%	<i>Agak lucu ya</i> (It's kind of funny, isn't it?) <i>Beradab itu tidak membela kepaluan</i> (Being civilized means not defending falsehood.)	Evaluation softened by hedging.
Neutral	1308	85%		Normal evaluative expression without intensity markers.

The prevalence of neutral graduation indicates that most users did not explicitly manipulate the intensity of their statements, but those who did tended to intensify rather than mitigate. Words like “*semakin*” (getting more), “*banget*” (really), “*paling*” (the most) or repeated exclamations served to heighten emotional impact. This intensification amplifies the force of moral condemnation, contributing to aggressive tone and persuasive pressure. Conversely, mitigated expressions such as “*agak lucu*” (a bit funny) or “*mungkin berlebihan*” (maybe exaggerated) appeared in attempts to reduce conflict or present irony. However, these were scarce, suggesting that the discourse environment prioritizes confrontation over moderation.

Engagement: Dialogic Positioning

Graduation refers to how speakers amplify or soften their evaluative statements. In this dataset, 85% of comments were neutral in intensity, 12% contained clear intensification, and 3% involved mitigation (see Table 3) Engagement analysis examines how commenters position themselves dialogically toward alternative viewpoints—whether they acknowledge, reject, or ignore them. As shown in Table 4, contractive stance dominated (44%), followed by neutral (42%) and expansive (14%) forms.

Table 4. Distribution of Engagement

Engagement Type	Number	Percentage	Example Comment	Interpretation
Contractive	676	44%	<i>Ya jelas dia salah</i> (Of course he's wrong)	Strong closure of dialogic space; assertion of certainty.
Expansive	212	14%	<i>Mungkin memang begitu</i> (Maybe that's just how it is)	Opens space for alternative perspectives or doubt.
Neutral	646	42%	<i>Serahkan semua kepada hukum</i> (Leave it all to the law)	Minimal dialogic stance; relatively neutral positioning.

The dominance of contractive engagement reinforces the finding that commenters rarely invite discussion or alternative viewpoints. Instead, they use language that asserts certainty, moral authority, or ideological dominance, such as “*jelas*” (obviously), “*pasti*” (definitely), “*memang*” (indeed), and “*harus*” (must). These markers close dialogic space, implying that the speaker's view is unquestionably correct. Such language reflects monoglossic discourse, a key feature of online hostility where participants reject plurality of opinion.

Expansive comments were limited, typically appearing in reflective or ironic tones. For example, “*Mungkin memang begitu*” (Maybe that’s just how it is) or “*Sepertinya sudah biasa di politik kita*” (It seems to be common in our politics) suggest a more dialogically open stance, allowing space for interpretation. Neutral comments, often informational or procedural, made up nearly half of the dataset but were linguistically simple and rarely evaluative.

Summary of Dominant Patterns

Overall, the analysis shows that the comment section operates as a morally and emotionally charged discursive space. The following patterns summarize the results:

1. Moral evaluation (Judgment) dominates, accounting for almost half of all evaluative resources.
2. Negative valence is overwhelmingly prevalent (61%), reflecting public anger and distrust toward political elites.
3. Intensification appears more often than mitigation, amplifying emotional and moral force.
4. Contractive stance indicates strong ideological closure, where users assert their evaluations as absolute truths.

These combined patterns demonstrate that online political discourse in Indonesia often functions as a site of moral contestation, where language is used to judge, accuse, and delegitimize rather than to deliberate. Hate speech thus emerges not only from explicit insults but from the interplay of moral judgment, negative emotion, and dialogic closure, all of which contribute to the construction of hostility in digital communication.

Discussion

Overall, the analysis shows that the comment section operates as a morally and emotionally charged discursive space. The following patterns summarize the results. The analysis of 1,534 YouTube comments on CNN Indonesia’s talk show “*Ijazah Digugat, Legitimasi Diragukan, Jalan Baru Pemakzulan Gibran?*” demonstrates that online hate speech in Indonesia operates not merely through the use of offensive vocabulary but through a systematic interplay of negative judgment, contractive engagement, and emotional intensification (graduation). These linguistic resources work together to construct hostility, moral polarization, and symbolic exclusion in the digital public sphere. Rather than simple insults, hate speech emerges as an evaluative discourse process—one that encodes moral reasoning, emotional arousal, and ideological stance through linguistic choice.

From a psychological perspective, these findings confirm that hate speech represents both a linguistic and affective phenomenon. The evaluative acts of judging, condemning, or ridiculing others activate emotional and cognitive mechanisms of in-group favoritism and out-group derogation, as proposed in Social Identity Theory (Kazmi et al., 2021; Houtven et al., 2024). In digital spaces, where anonymity lowers social accountability, these mechanisms become amplified: individuals express aggression not just as personal opinion, but as moral performance, signaling belonging and loyalty to a perceived group identity.

Negative Judgment as the Core of Hate Discourse

The dominance of *Judgment* in the dataset—especially those expressing negative moral propriety and low capacity—reveals that commenters primarily targeted the moral and intellectual integrity of political figures rather than discussing ideas or policy content. Expressions such as “*TERMUL DUNGU DAN MENUTUP MATA TENTANG KEBENARAN*” (TERMUL is stupid and turns a blind eye to the truth) or “*Fix, ijasah bapak dan anak PALSU*” (Definitely, the father’s and son’s diplomas are fake) illustrate a moralized discourse, where the target is portrayed as dishonest, unintelligent, or unworthy of public trust.

This linguistic behavior aligns with previous Indonesian studies (Akbar et al., 2024; Adinugroho et al., 2023; Maalikih et al., 2024), which found that moral degradation and sarcastic humiliation are the most common forms of online aggression. However, the present study extends those findings by showing how *Judgment* is intertwined with collective labeling—a process through which individuals become symbols of an ideological or social group. For instance, the recurrent phrase “*TERMUL itu simbol kedunguan*” transforms a personal insult into a semiotic marker of group identity, turning complex political affiliations into simplified moral categories of “us” versus “them.”

From a psychological standpoint, this shift from personal to collective judgment reflects the operation of moral disengagement (Bandura, 1999; Chen et al., 2025). When individuals perceive their targets not as human beings but as embodiments of immorality or stupidity, aggression becomes morally justified. Such labeling serves cognitive efficiency—reducing ambiguity and reinforcing certainty—while also fulfilling emotional needs for superiority and validation within the in-group.

Thus, negative *Judgment* functions both linguistically and psychologically: linguistically, it encodes condemnation through evaluative lexis; psychologically, it satisfies moral emotions such as anger, contempt, and disgust that sustain collective animosity.

Contractive Engagement and Dialogic Closure

The second major feature of the dataset is the prevalence of contractive engagement, where commenters employ strategies of proclaiming and disclaiming to close off dialogic space (Courtenay & Baraitser, 2021). Statements such as “*Sudah bisa dipastikan Jokowi tidak punya ijazah*” (It can already be confirmed that Jokowi doesn’t have a diploma) or “*Fix, ijasah palsu*” (Definitely, the diploma is fake) exemplify monoglossic certainty—assertions that leave no room for alternative perspectives. Within the Appraisal framework (Martin & White, 2005), such contractive language signals a highly closed dialogic stance, suggesting that the speaker positions their statement as the only acceptable truth.

This rhetorical pattern intensifies ideological polarization. Similar to findings by Wang (2025) and Hackenburg et al. (2023), contractive language in political discourse functions to solidify group boundaries by presenting claims as moral certainties rather than as contestable arguments. Epistemic markers such as “*jelas*,” “*pasti*,” or “*sudah terbukti*” linguistically assert authority and reinforce the speaker’s moral superiority.

In psychological terms, contractive stance reflects cognitive rigidity—a tendency to process information through confirmation bias, selectively attending to cues that support existing beliefs. This rigidity is often accompanied by moral conviction (Yoder & Decety, 2022; Rollwage & Fleming, 2021), where opinions are experienced as moral truths. When moral conviction combines with emotional arousal, individuals become resistant to counterarguments and perceive disagreement as moral transgression rather than intellectual difference.

Therefore, contractive engagement is not only a linguistic mechanism but also a cognitive defense strategy: it protects the self’s moral worldview by denying ambiguity and suppressing dialogic openness. In the context of YouTube discussions, this stance transforms dialog into monologue, turning the comment section into an arena of moral proclamation rather than deliberative exchange.

Intensified Graduation and Emotional Escalation

The *Graduation* subsystem reveals how commenters amplify or soften their evaluative meanings. In this study, 12% of the dataset displayed intensification markers—through repetition, capitalization, and exclamation marks—while only 3% mitigated their tone. Examples such as “*TERMUL DUNGU!!!*” (TERMUL is utterly dumb!!!) and “*benar-benar memalukan*” (Really embarrassing) illustrate upscaled force that dramatizes emotional experience.

This finding echoes global research on hate speech (Gallacher & Bright, 2021; Luo et al., 2020), which shows that intensification is a key rhetorical device for collective emotional contagion. By amplifying force, users externalize their emotions and invite others to share the same emotional intensity. In digital environments, such high-arousal language spreads rapidly, fostering emotional synchrony—a psychological process where group members align their feelings through repeated exposure to similar emotional cues (Doré & Morris, 2018; Houran et al., 2025).

From a linguistic standpoint, intensification functions as performance: it turns private emotions into public acts of outrage (Leuprecht et al., 2025). From a psychological standpoint, it serves as catharsis and social signaling—expressing anger not just to release tension but to affirm one’s identity as part of an emotionally unified collective (Zhang et al., 2024). This dual function explains why highly emotional comments often attract more likes and replies (Kane et al., 2023). The expression of anger becomes socially rewarded, reinforcing the cycle of hostility and group solidarity.

Graduation thus operates as a linguistic amplifier of both moral and emotional meanings. The interaction between intensified graduation and contractive stance produces what can be called rhetorical absolutism: a discourse pattern where emotion and certainty merge to create moral closure and antagonism.

From Personal Insults to Collective Polarization

One of the most significant findings of this study is that hate speech in Indonesian political discourse transcends personal hostility and evolves into collective polarization. The repeated use of symbolic labels such as “*termul*” marks the construction of an ideological out-group. Through repetition, capitalization, and emotional reinforcement, such terms become cultural symbols of inferiority.

This phenomenon mirrors Civila’s et al. (2020) findings on Islamophobic discourse in Indonesian social media, where group labeling and metaphorical associations construct the “Other” as morally corrupt

or intellectually inferior. In the present dataset, the same mechanism manifests through the recontextualization of political identities. Individual political figures become metonyms for entire ideological blocs, allowing users to express complex political resentment through simple symbolic tokens.

Psychologically, this process involves dehumanization (Haslam, 2006)—the denial of mental or moral qualities to others. Dehumanization reduces empathy and legitimizes aggression, transforming linguistic hostility into normalized social behavior (Bruneau et al., 2017). The combination of *Judgment*, *Contractive stance*, and *Graduation* thus forms a discursive triad that reproduces moral exclusion at both linguistic and psychological levels.

Psychological Interpretation: Emotional Regulation, Identity, and Aggression

The emotional and cognitive mechanisms underlying these linguistic patterns reflect broader psychological dynamics of online behavior. Three dimensions are particularly relevant: emotional regulation, identity signaling, and aggressive motivation.

First, emotional regulation in online settings often fails due to anonymity and depersonalization. The lack of face-to-face feedback diminishes empathy, leading to disinhibition (Syrjämäki et al., 2024; Abdullah et al., 2024). Users vent frustration more freely, using moralized language to justify their hostility. Second, identity signaling occurs when individuals use evaluative and intensified language to demonstrate loyalty to an ideological group. Negative judgment becomes a badge of belonging—a linguistic act of allegiance.

Third, aggression in this context is not merely impulsive but instrumental. It serves to maintain dominance and moral authority within a group. As research in social psychology suggests, individuals often engage in moral aggression to preserve self-esteem and group cohesion (Snowden et al., 2021; Kruglanski et al., 2023). The high frequency of contractive and judgmental language in the dataset suggests that hate speech is not just a breakdown of civility but an organized form of collective emotion management, channeling anger into symbolic attack.

Integrating Appraisal Theory into the study of hate speech broadens our understanding of evaluation as both a linguistic and psychological process. Hate speech is not an anomaly of communication but an extreme manifestation of evaluative discourse. As Martin & White (2005) argued, interpersonal meaning is graduated rather than binary—ranging from mild evaluation to moral condemnation. This study confirms that hate speech occupies the highest end of this gradient, characterized by strong moral judgment, closed engagement, and intensified force.

Moreover, while sentiment analysis captures the polarity of attitudes (positive vs. negative), Appraisal analysis reveals the mechanisms of intensity and stance that construct moral exclusion. By identifying how *Judgment* combines with *Graduation* and *Engagement*, this study shows how linguistic form translates into social function—how evaluation becomes moral boundary-making.

Contextually, the findings reveal a discursive shift in Indonesian digital culture: political discussions increasingly privilege moral certainty and emotional outrage over rational deliberation. This shift aligns with a global trend of affective polarization, where emotions and moral beliefs, rather than facts, drive online engagement.

From a psychological angle, this tendency is underpinned by motivated reasoning and moral identity protection. People process information in ways that affirm their moral worldview, avoiding cognitive dissonance by dismissing counterarguments as immoral. Consequently, hate speech becomes a form of moral performance—a ritualized assertion of virtue and belonging.

These insights carry significant implications for digital literacy and content moderation. Because hate speech operates through subtle evaluative and emotional mechanisms, detection systems based solely on profanity lists or sentiment polarity are insufficient. Integrating Appraisal-based linguistic indicators—especially *Judgment* and *Graduation*—could improve algorithmic sensitivity to implicit hate speech, which often appears in morally coded language rather than direct insults (Calderón et al., 2021; Haq & Rahyono, 2025).

Equally important is the role of psychological education in promoting emotional awareness. Recognizing how moral emotions like anger, contempt, and disgust can escalate into collective hostility can help users regulate responses and resist manipulative rhetoric. Media literacy programs should therefore combine linguistic training with emotional literacy, enabling users to identify when evaluative discourse crosses the line from criticism to dehumanization.

Conclusion

This study concludes that hate speech in Indonesia's digital sphere, particularly in YouTube comments on CNN Indonesia's talk show "*Ijazah Digugat, Legitimasi Diragukan, Jalan Baru Pemakzulan Gibran?*", is a

structured form of evaluative discourse rather than a series of random insults. The Appraisal analysis shows that hate speech operates through three main linguistic mechanisms—negative judgment, contractive stance, and intensified graduation—that together create moral hostility and ideological closure. Most comments expressed moral condemnation instead of rational argument, reflecting a shift from critical dialogue to ethical punishment. Linguistically, commenters attacked personal integrity and intelligence while asserting certainty and amplifying emotional tone through repetition, capitalization, and hyperbole. These linguistic choices reveal how moralized evaluation, closed engagement, and emotional intensification work collectively to transform online spaces into arenas of moral contestation.

Psychologically, the dominance of moral judgment and negative emotion suggests that hate speech is rooted in social identity dynamics, emotional contagion, and moral conviction. Commenters use language to affirm belonging to an ideological group, regulate anger, and display moral superiority. The findings highlight that online aggression is not spontaneous but sustained by cognitive rigidity and affective polarization. Therefore, combating hate speech requires more than censorship—it demands digital literacy and emotional awareness that teach users to recognize evaluative cues, question moral absolutism, and manage emotional responses. By linking linguistic structure with psychological process, this study emphasizes that reducing hate speech in Indonesia's online discourse depends on fostering empathy, reflexivity, and openness in public communication.

Several limitations should be acknowledged. First, the study focuses exclusively on one YouTube talk show episode, which may limit the generalizability of findings to other digital platforms or political contexts. Second, the qualitative Appraisal analysis, while rich in interpretive depth, relies on manual coding, which introduces potential researcher subjectivity despite systematic cross-checking. Third, the study does not account for demographic variables of commenters (e.g., age, education, political affiliation), which could influence linguistic and evaluative patterns. These limitations suggest that the findings reflect a specific discursive moment rather than universal patterns of online hate speech. Future research could expand the dataset across platforms, employ computational methods to supplement qualitative analysis, and incorporate sociodemographic factors to provide a more comprehensive understanding of evaluative aggression in Indonesian digital discourse.

Declarations

Author contribution : Khusnul Khotimah was responsible for the entire research project. She also led the writing of the manuscript and collaborated with the second author. Achamid Fanani participated in the data collection, transcription, and analysis. He also revised the manuscript. Both authors approved the final manuscript. Khusnul Khotimah contributed to theoretical framework development, data interpretation, and manuscript refinement.

Funding statement : This research did not receive any funding.

Conflict of interest : Authors declare that they have no competing interests.

Ethics Approval : This study used publicly accessible YouTube comments and adhered to ethical guidelines for digital research (AoIR). No personally identifiable information was collected or disclosed.

Additional information : No additional information is available for this paper

References

Abdullah, H., Kunhao, Z., & Ma'ruf, A. A. (2024). The role of moral disengagement, anonymity perception, online disinhibition, and empathy in predicting cyberbullying perpetration among Chinese young adults. *International Journal of Academic Research in Business and Social Sciences*, 14(12), 2178-2195. <https://doi.org/10.6007/IJARBSS/v14-i12/24048>

Adinugroho, I., Kristiani, P., & Nurrachman, N. (2023). Understanding aggression in digital environment: Relationship between shame and guilt and cyber aggression in online social network. *Psikohumaniora: Jurnal Penelitian Psikologi*, 8(2), 215-228. <https://doi.org/10.21580/pjpp.v8i2.14827>

Afrina, C., Zulaikha, S. R., & Jumila, J. (2024). The low level of digital literacy in Indonesia: An analysis of online media content. *Record and Library Journal*, 10(2), 374-387. <https://doi.org/10.20473/rlj.V10-I2.2024.374-387>

Akbar, M. F., Akram, M., & Fiddienika, A. (2024). Fenomena ketidaksopanan berbahasa Indonesia pada saat bermain game online. *Jurnal Loyalitas Sosial: Journal of Community Service in Humanities and Social Sciences*, 6(2), 85-99. <https://doi.org/10.32493/JLS.v6i2.p85-99>

Ariestandy, D., Adidharma, W., & Isdendi, R. R. (2024). Transformation of political participation in the digital age: The role of social media in shaping public opinion and mass mobilization. *Jurnal Ekonomi Teknologi dan Bisnis (JETBIS)*, 3(9), 1738–1745. <https://doi.org/10.57185/jetbis.v3i9.145>

Bandura, A. (1999). Social cognitive theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed.). Guilford Press.

Bobro, N. (2024, August). *Conditions for enhancing students' information and intellectual activity in the digital environment*. In *Theoretical and empirical scientific research: Concept and trends*. <https://doi.org/10.36074/logos-16.08.2024.045>

Bruneau, E. G., Kteily, N. S., & Laustsen, L. (2017). The unique effects of blatant dehumanization on attitudes and behavior towards Muslim refugees during the European 'refugee crisis' across four countries. *European Journal of Social Psychology*, 48(5), 645–662. <https://doi.org/10.1002/ejsp.2357>

Calderón, F. H., Balani, N., Taylor, J., & Peignon, M. (2021). Linguistic patterns for code word resilient hate speech identification. *Sensors*, 21(23), 7859. <https://doi.org/10.3390/s21237859>

Chen, Y., Lu, J., Walker, D. I., Ma, W., Glenn, A. L., & Han, H. (2025). Perceived stress and society-wide moral judgment. *European Journal of Investigation in Health, Psychology and Education*, 15(6), 106. <https://doi.org/10.3390/ejihpe15060106>

Civila, S., Romero-Rodríguez, L. M., & Civila, A. (2020). The demonization of Islam through social media: A case study of #StopIslam on Instagram. *Publications*, 8(4), 1–19. <https://doi.org/10.3390/publications8040052>

Courtenay, T., & Baraitser, P. (2021). Online contraceptive discussion forums: A qualitative study to explore information provision. *BMJ Sexual & Reproductive Health*, 47(3), e5. <https://doi.org/10.1136/bmjsrh-2020-200719>

Dam, L., & Jensen, H. D. (2024). Everyday positivity: An appraisal analysis of online identity in food blogs. *First Monday*, 29(6). <https://doi.org/10.5210/fm.v29i6.13225>

Doré, B. P., & Morris, R. (2018). *Linguistic synchrony predicts the immediate and lasting impact of text-based emotional support*. *Psychological Science*, 29(10), 1716–1729. <https://doi.org/10.1177/0956797618779971>

Gallacher, J. D., & Bright, J. (2021, February). *Hate contagion: Measuring the spread and trajectory of hate on social media*. OSF Preprints. <https://doi.org/10.31234/osf.io/b9qhd>

Hackenburg, K., Brady, W. J., & Tsakiris, M. (2023). Mapping moral language on U.S. presidential primary campaigns reveals rhetorical networks of political division and unity. *PNAS Nexus*, 2(6), 189. <https://doi.org/10.1093/pnasnexus/pgad189>

Haq, Z., & Rahyono, F. X. (2025). Exploring negative judgment language in Indonesian conversational language on Facebook: An appraisal analysis. *Indonesian Journal of EFL and Linguistics*, 10(1). <https://doi.org/10.21462/ijefl.v10i1.879>

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Houran, J., Dagnall, N., Sapkota, R. P., & Lange, R. (2025). The spread of mind: Psychological contagion in theory and critique. *Frontiers in Psychology*, 16, 1631927. <https://doi.org/10.3389/fpsyg.2025.1631927>

Houtven, V. E., Acquah, S. B., Obermaier, M., & Saleem, M. (2024). 'You got my back?' Severity and counter-speech in online hate speech toward minority groups. *Media Psychology*, 27(4), 1–32. <https://doi.org/10.1080/15213269.2023.2298684>

Jamilah, F., & Wahyuni, P. (2020). Ujaran kebencian dalam kolom komentar YouTube pada tahun politik pemilihan presiden 2019. *Silampari Bisa: Jurnal Penelitian Pendidikan Bahasa Indonesia, Daerah, dan Asing*, 3(2), 325–341. <https://doi.org/10.31540/silamparibisa.v3i2.1109>

Kane, A., Van Swol, L. M., & Sarmiento-Lawrence, I. G. (2023). *Emotional contagion in online groups as a function of valence and status*. *Computers in Human Behavior*, 139, 107543. <https://doi.org/10.1016/j.chb.2022.107543>

Kazmi, S., Mohd Noor, S. N. F., Nur Indah, R., & Jaes, L. (2021). Attitude realization in news reports: An interpretation through an appraisal analysis. *Indonesian Journal of Applied Linguistics*, 11(1), 191–200. <https://doi.org/10.17509/ijal.v11i1.34622>

Khotimah, K. (2024). Analisis Sistem Appraisal Pada Pemberitaan Konflik Israel-Palestina Di Media Kompas.Com dan CNN Indonesia. *Jurnal Kandai*, 20(2), 229–245. <https://ojs.badanbahasa.kemendikdasmen.go.id/jurnal/index.php/kandai/article/view/7025>

Križan, A., & Barbič, A. (2023). Appraisal analysis and AI chatbots: Do we even need humans? *ELOPE: English Language Overseas Perspectives and Enquiries*, 22(1), 35–52. <https://doi.org/10.4312/elope.22.1.35-52>

Kruglanski, A. W., Ellenberg, M., Szumowska, E., & Bushman, B. J. (2023). Frustration-aggression hypothesis reconsidered: The role of significance quest. *Aggressive Behavior*, 49(1), 3-13. <https://doi.org/10.1002/ab.22092>

Leuprecht, C., Skillicorn, D. B., & Kernot, D. (2025). Linguistic models of abusive language. *Dynamics of Asymmetric Conflict*, 18(2), 158-174. <https://doi.org/10.1080/17467586.2024.2407922>

Luo, M., Hsu, T. W., Park, J. S., & Hancock, J. T. (2020). *Emotional amplification during live-streaming: Evidence from comments during and after news events*. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1-19. <https://doi.org/10.1145/3392853>

Maalikih, M. D., Sumarwati, & Rakhmawati, A. (2024). Degrading sarcasm in verbal bullying: A speech act study on junior high school students in Surakarta, Indonesia. *Proceeding of The International Conference of Innovation, Science, Technology, Education, Children, and Health*, 4(2), 267-277. <https://icistech.org/index.php/icistech>

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation*. Palgrave Macmillan.

Reichelmann, A., Hawdon, J., Costello, M., & Ryan, J. (2020). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior*, 42(9), 1-12. <https://doi.org/10.1080/01639625.2020.1722337>

Rollwage, M., & Fleming, S. M. (2021). Confirmation bias is adaptive when coupled with efficient metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1822), 20200131. <https://doi.org/10.1098/rstb.2020.0131>

Serrano-Puche, J., Hashom, H., Mustafa, W. A., Mohammed Noor, A., Tambak, S., & Nasir, K. (2021). Digital disinformation and emotions: Exploring the social risks of affective polarization. *International Review of Sociology*, 31(2). <https://doi.org/10.1080/03906701.2021.1947953>

Snowden, R., Amad, S., Morley, E., & Butkute, N. (2021). Explicit and implicit self-esteem and aggression: Differential effects of agency and communion. *Journal of Interpersonal Violence*, 37(11-12), NP10036-NP10059. <https://doi.org/10.1177/0886260520985490>

Syrjämäki, A., Ilves, M., Olsson, T., & Kiskola, J. (2024). Online disinhibition mediates the relationship between emotion regulation difficulties and uncivil communication. *Scientific Reports*, 14(1), 30019. <https://doi.org/10.1038/s41598-024-81086-7>

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), 230-251. <https://doi.org/10.24071/llt.v27i1.8180>

Walther, J. B. (2022). Social media and online hate. *Current Opinion in Psychology*, 45, 101298. <https://doi.org/10.1016/j.copsyc.2021.12.010>

Wang, S. Y. N. (2025). Moral language and political polarisation: An overview. *Sociology Compass*, 19(5), e70063. <https://doi.org/10.1111/soc4.70063>

Yoder, K. J., & Decety, J. (2022). Moral conviction and metacognitive ability shape multiple stages of information processing during social decision-making. *Cortex*, 151, 162-175. <https://doi.org/10.1016/j.cortex.2022.03.008>

Zhang, J., Xiong, Z., Zheng, H., & Ma, X. (2024). The moral psychological justification of anger: An exploration of self-respect and recognition. *Behavioral Sciences*, 15(1), 3. <https://doi.org/10.3390/bs15010003>