

Hate Speech Discourse on Platform X (Twitter): A Critical Discourse Analysis

Nurdelia Nasution¹, Intan Nia Salsabila^{2*}, Meisya Audreyanna³, Saralena Manik⁴,
Christy Aulia Simanjuntak⁵, Muhammad Natsir⁶

¹⁻⁶Departemen Bahasa dan Sastra Inggris, Fakultas Bahasa dan Seni, Universitas Negeri Medan, Indonesia

*Penulis Korespondensi: intanniasalsabila58@gmail.com

Abstract. The phenomenon of hate speech on social media, particularly on platform X, has intensified alongside the increasing level of public interaction within open and largely unregulated digital spaces. This condition not only generates communicative conflicts but also shapes complex social experiences for individuals, particularly in relation to identity, emotion, and power relations. This study aims to gain an in-depth understanding of how hate speech is constructed, interpreted, and negotiated by users within the context of digital interaction. Employing a qualitative approach with a Critical Discourse Analysis (CDA) design, data were collected through in-depth interviews, non-participant observation, and digital document analysis involving 10–15 active users of platform X who have experienced hate speech. Data were analyzed thematically by identifying patterns of meaning emerging from participants' experiences. The findings reveal three major themes: hate speech as a lingering yet normalized emotional experience; discourse as a site for the reproduction of power and identity delegitimization; and self-negotiation strategies employed by participants to survive within digital spaces. These findings indicate that hate speech operates not only at the linguistic level but also in shaping users' social and psychological realities. Theoretically, this study reinforces Critical Discourse Analysis by emphasizing the importance of subjective experience in interpreting discursive practices. Practically, it contributes to the development of digital literacy, content moderation policies, and efforts to create more inclusive and reflective communication spaces in the digital era.

Keywords: Critical Discourse Analysis; Digital Identity; Hate Speech; Power and Ideology; Social Media.

1. INTRODUCTION

Digital spaces have increasingly become primary arenas for individuals to express views, construct identities, and participate in public discourse. However, alongside these opportunities, the rise of hate speech on social media particularly on platform X (formerly Twitter) has become a growing concern. For many users, this space is no longer solely a platform for communication but has evolved into a site of conflict characterized by verbal attacks, labeling, and identity delegitimization. Previous studies show that repeated exposure to hate speech can affect users' mental health and social perceptions (Nguyen et al., 2025), while digital discourse is often used to reinforce ideological positions and exclusive narratives (Li et al., 2025; Maci & Anesa, 2025).

Furthermore, identity construction in digital spaces is dynamic yet vulnerable to symbolic attacks and politicization (Pei et al., 2025; Naureen & Shaukat, 2026), and is influenced by underlying power relations and ideological bias (Qiu et al., 2026). Other studies highlight that factors such as intergroup threat, intersectionality, and digital polarization contribute to the emergence and reproduction of hate speech in online environments (Guan & Chen, 2025) (Almomani & Jegi, 2026), 2026; Roitman et al., 2026).

Despite these contributions, much of the existing research focuses on detection, classification, or structural analysis using computational approaches, which remain limited in capturing contextual meaning and subjective experience (Kopf, 2025). Therefore, a significant gap remains in understanding how users experience, interpret, and negotiate hate speech in everyday digital interactions. Addressing this gap, this study employs a Critical Discourse Analysis approach to examine hate speech on platform X, focusing on users' experiences, meanings, and social processes. This study aims to analyze how language practices, power relations, and identity construction operate within digital interactions, thereby contributing to the development of Critical Discourse Analysis and offering practical insights for digital literacy and more inclusive communication practices.

2. THEORETICAL FRAMEWORK

The phenomenon of hate speech on platform X cannot be understood merely as an individual linguistic expression; rather, it constitutes a complex social practice in which language functions to shape experience, affirm identity, and reproduce power relations. Therefore, this study integrates several theoretical perspectives that are capable of explaining the dynamics of meaning, emotion, and social interaction experienced by participants.

Critical Discourse Analysis as Social Practice

The primary approach employed in this study is Critical Discourse Analysis (CDA), which positions language as a non-neutral social practice. In the digital context, CDA facilitates an understanding of how hate speech not only reflects reality but also constructs and sustains particular social structures. Maci and Anesa (2025, *Discourse, Context & Media*) demonstrate that discursive practices in the digital era have undergone transformation, whereby the production of meaning is no longer individual but collective and influenced by platform ecosystems. In participants' experiences, hate speech often appears as a recurring pattern that shapes the perception that verbal aggression is "normal." This indicates that discourse operates subtly in shaping social consciousness.

Power and Ideology in Digital Discourse

Within the CDA framework, the concepts of power and ideology are central to understanding how hate speech operates. Language is used to determine who holds legitimacy to speak and who becomes marginalized. Li et al. (2025, *Technology in Society*) show that digital discourse is frequently used to reinforce group identity boundaries through exclusive narratives.

Meanwhile, Qiu et al. (2026, *Digital Scholarship in the Humanities*) highlight that meaning construction in digital spaces often contains ideological biases that are not always recognized by users. In the context of participants' experiences, when individuals are attacked not because of their arguments but because of their identity, discourse functions as a tool of delegitimization. This illustrates how power operates through language that appears simple yet carries structural consequences.

Social Identity Theory and Subjective Experience

Social identity theory helps explain how individuals interpret themselves in relation to others, particularly in situations of discursive conflict. In social media, identity is not only expressed but also negotiated and contested. Naureen and Shaukat (2026, *Journal of Language and Social Psychology*) demonstrate that digital discourse often politicizes identity, making differences a source of conflict. Similarly, Pei et al. (2025, *Journalism Studies*) emphasize that social media users actively construct identity through linguistic strategies while remaining vulnerable to discursive attacks. In participants' experiences, hate speech is often perceived as an attack on the "self," rather than merely an expression of opinion. This explains why its impact is deeply emotional and personal.

Normalization and Desensitization in Digital Discourse

One of the key findings in recent literature is the process of normalization of hate speech. Repeated exposure to aggressive language can alter individuals' perceptions of what is considered acceptable. Nguyen et al. (2025, *Social Science & Medicine*) found that the intensity of interaction on social media contributes to desensitization toward hate speech. Meanwhile, Amores and Escudero-Garrido (2025, *Journal of Information Science*) show that although hate speech can be detected algorithmically, its interpretation remains dependent on social context. In this context, participants who report being "accustomed" to such discourse reflect a process of internalizing repeated patterns of discourse. However, this normalization does not imply the absence of impact; rather, it reveals a tension between acceptance and resistance.

Multimodal Discourse and the Complexity of Digital Expression

Hate speech on platform X does not always appear in textual form but also through memes, images, emojis, and visual symbols. A multimodal approach helps explain how meaning is constructed through the combination of various modes of communication. Chasanah and Rofiq (2026, *Journal of Digital Communication*) emphasize that emerging forms such as deepfake and visual satire expand the spectrum of hate speech. This suggests that discourse analysis must go beyond text and consider visual and cultural contexts within digital environments.

Conceptual Framework of the Study

This study conceptualizes hate speech as a discursive practice that shapes social experience through the interaction of language, power, and identity. Within this framework, language is understood as a tool for constructing social reality, while participants' experiences are interpreted as the result of the dynamic interaction between structure and agency. Discourse is therefore analyzed not only at the textual level but also in relation to its broader context and social implications. The researcher interprets the data by examining how hate speech influences emotions and self-perception, how identity is negotiated within conflictual interactions, and how individuals respond to ongoing discursive pressures. For instance, when a participant reports feeling "silenced," this is not viewed merely as a personal reaction but as an indication of discursive structures that constrain the legitimacy of certain voices. Accordingly, this framework positions participants' voices at the center of analysis, not as passive objects but as representations of lived experience within a digital environment characterized by tension, negotiation, and continuously evolving meaning.

3. RESEARCH METHODOLOGY

This study employs a qualitative research design using Critical Discourse Analysis (CDA) as the primary analytical framework to examine how hate speech is constructed through language and how it reflects power relations, ideology, and social dominance within digital interactions on platform X. The approach follows the framework proposed by Fairclough (2013) and Wodak (2015), which conceptualizes discourse as a social practice embedded in specific contexts. The participants consist of 10–15 active users of platform X selected through purposive sampling based on their relevance and direct experience with hate speech, with additional participants identified through snowball sampling. The study focuses on public interactions related to sensitive issues such as politics, religion, gender, and social identity, particularly among users in Indonesia.

Data were collected through semi-structured in-depth interviews, digital document analysis, and non-participant observation. Interviews were conducted to explore participants' experiences and interpretations, while textual data were obtained from tweets, replies, and relevant online interactions containing hate speech. The instruments used include interview guidelines, field notes, and digital documentation tools. Data analysis was conducted using Fairclough's three-dimensional CDA model, encompassing text analysis, discursive practice, and social practice, through stages of coding, thematic categorization, and critical interpretation.

The validity of the data was ensured through triangulation and member checking, indicating consistency and credibility of the findings. Ethical considerations were addressed by ensuring informed consent, maintaining participant confidentiality through anonymization, and using the data solely for academic purposes, while also considering the ethical use of digital content.

4. RESULT AND DISCUSSION

Findings

The analysis of interview data, digital observations, and discourse documentation on platform X produced three interconnected major themes: (1) hate speech as an emotional and psychological experience, (2) discourse as a site of power and identity delegitimization, and (3) strategies of survival and self-negotiation within digital spaces. These themes do not stand independently but are intertwined in shaping participants' experiences within the digital public sphere.

Theme 1: Hate Speech as a Lasting Emotional Experience

The phenomenon of hate speech is not merely encountered as text that is briefly read, but as an emotional experience that leaves a lasting and recurring impact. This situation often arises when participants engage in sensitive public discussions, such as political or religious identity issues. In many cases, verbal attacks emerge suddenly and massively through replies or mentions. One participant (P3) described the experience as follows:

“Initially, I thought it was just a normal debate, but over time it turned into personal attacks. I was called stupid, told to stay silent, and even my religion was brought up. It made me reconsider posting again.”

This experience indicates that hate speech is not simply understood as a difference of opinion, but as a form of attack that targets personal and identity-related aspects. The emotions that arise are not singular anger, fear, and doubt are intertwined. Interestingly, some participants also demonstrated ambiguity in responding to such experiences. Participant P7 stated:

“Sometimes I know it's toxic, but at the same time I've become used to it. It feels like it's part of the 'rules of the game' on Twitter.” This statement reflects the normalization of hate speech, where experiences that were initially disturbing gradually become accepted as an unavoidable reality.

Theme 2: Discourse as a Site of Power and Identity Delegitimization

Beyond personal interaction, hate speech functions as a discursive mechanism to assert power and define who is considered “legitimate” to speak. In much of the data, hate speech appears in the form of labeling, stereotyping, and generalization toward particular groups. For example, in a discussion thread on social issues, patterns were identified in the use of terms that demean certain groups. This is reinforced by participant P1’s experience:

“It’s not just about me personally; they immediately attack my group. It’s as if what I say is not valid simply because I come from that background.”

In this context, language becomes a tool for limiting the legitimacy of voices. Participants are not only challenged on their arguments but are positioned as “unworthy” of speaking. However, this dynamic is not always one-directional. In some cases, participants acknowledged that they had also engaged in similar discursive practices. Participant P5 stated:

“Sometimes, without realizing it, I also respond in the same tone. It’s like a cycle that’s hard to break.”

This statement highlights a paradox: individuals who are victims of hate speech can simultaneously become perpetrators in certain contexts. This demonstrates that hate discourse is reproductive in nature and continuously circulates within digital interaction structures.

Theme 3: Self-Negotiation and Survival Strategies in Digital Spaces

In response to continuous discursive pressure, participants develop various strategies to survive and maintain their presence in digital spaces. These strategies are not always explicit but emerge as adaptive responses to lived experiences. Some participants choose to limit their interactions. Participant P2 stated:

“Now I’m more selective. If it starts to get toxic, I just mute or block. It’s better than exhausting myself.”

Meanwhile, others choose to remain active but adopt more cautious approaches. Participant P8 explained:

“I still want to speak up, but now I think more about how I say things. I’m afraid of being misunderstood or attacked again.”

On the other hand, some participants perceive these experiences as opportunities for learning. Participant P10 stated:

“I’ve come to better understand how people think. Even though it’s harsh, it also opens up new perspectives.”

These findings indicate that responses to hate speech are not homogeneous. Some individuals withdraw, others adapt, and some reinterpret the experience as part of self-reflection. However, these strategies also involve tension. On one hand, participants wish to remain part of the digital public sphere; on the other, they must continuously negotiate their sense of safety and identity.

Table 1. Summary of Themes and Core Meanings.

Main Theme	Subthemes	Core Meaning
Emotional Experience	Fear, anger, normalization	Hate speech as a lasting personal experience
Power & Identity	Delegitimization, stereotypes, discourse reproduction	Language as a tool of social domination
Survival Strategies	Avoidance, adaptation, reflection	Self-negotiation in digital spaces

The three themes illustrate an interconnected trajectory of experience: from intense emotional encounters, to an awareness of power structures within discourse, and ultimately to individual efforts to survive and negotiate the self. Rather than following a linear path, these experiences are fluid, ambiguous, and at times contradictory. Thus, the findings of this study do not merely describe “what happens,” but also reveal how participants interpret, feel, and respond to the social realities shaped by hate speech on platform X.

Discussion

This study reveals that hate speech on platform X is not merely a linguistic phenomenon, but a complex social experience encompassing emotional dimensions, power relations, and strategies of identity negotiation. The three major themes—(1) lasting emotional experience, (2) discourse as a site of power, and (3) survival strategies—demonstrate that participants do not simply “encounter” hate speech, but rather live and adapt within discursive structures that continuously produce and reproduce social meaning.

Hate Speech as Emotional Experience and Social Normalization

The findings indicate that hate speech is interpreted as an intense emotional experience while simultaneously undergoing a process of normalization. From the perspective of Critical Discourse Analysis (CDA), this reflects how repeated discursive practices can shape a “new normal” within social interaction.

This phenomenon aligns with the findings of Nguyen et al. (2025, *Social Science & Medicine*), which show that repeated exposure to hate speech on social media can lead to emotional desensitization. However, this study extends that understanding by revealing the ambiguity in participants’ experiences: they are neither fully immune nor entirely able to resist such discourse.

From a theoretical standpoint, this condition suggests that language does not merely represent reality but also shapes the boundaries of social tolerance toward symbolic violence. In other words, the continual repetition of hate speech has the potential to shift norms of public interaction from those grounded in rationality toward greater permissiveness of verbal aggression.

Discourse, Power, and Identity Delegitimization

The second theme emphasizes that hate speech functions as a discursive mechanism to define who possesses legitimacy within the digital public sphere. Within Fairclough's CDA framework, this relates to the dimension of social practice, where language operates as a tool for the reproduction of ideology and domination.

These findings are consistent with Li et al. (2025, *Technology in Society*), which demonstrate that digital discourse is often used to reinforce "ingroup" and "outgroup" boundaries through exclusive language. However, this study contributes further by highlighting that delegitimization occurs not only at the group level but is also experienced personally by individuals through identity-based attacks. Importantly, participants' acknowledgment of being both "victims and perpetrators" illustrates that power within discourse is fluid and relational. This challenges prior perspectives that frame hate speech in binary terms (perpetrator versus victim). Instead, this study shows that individuals may shift positions depending on the interactional context. Thus, hate speech not only reflects social conflict but also serves as a mechanism for reproducing that conflict.

Self-Negotiation: Between Survival, Silence, and Adaptation

The third theme demonstrates that participants are not passive recipients, but actively negotiate their positions within digital spaces. Strategies such as avoidance, selective engagement, and modification of communication styles reflect individual agency in responding to discursive pressure.

These findings are in line with Pei et al. (2025, *Journalism Studies*), which highlight how social media users strategically manage their identities within competitive and often hostile digital environments. However, this study adds nuance by revealing an existential dilemma: the desire to maintain a voice in public discourse while simultaneously seeking safety. From the perspective of social identity theory, this reflects an ongoing process of identity negotiation, in which individuals must continuously balance self-expression and social acceptance. This also indicates that digital spaces are not merely sites of communication, but arenas of self-construction that are dynamic and risk-laden.

Theoretical Reflection and Contribution

Overall, this study reinforces and extends the framework of Critical Discourse Analysis by demonstrating that digital discourse is inherently multi-layered, encompassing emotional, ideological, and performative dimensions. It further shows that hate speech should not be understood merely as textual content, but as a lived experience that shapes and is shaped by users' social realities. In this context, power relations within social media are revealed to be dynamic and unstable, continuously negotiated through interaction. In contrast to computational studies, such as Amores and Escudero-Garrido (2025), which primarily focus on detection, this study emphasizes the importance of understanding subjective meaning and lived experience. Therefore, its primary contribution lies in shifting the analytical focus from "what hate speech is" to "how hate speech is experienced, interpreted, and negotiated."

Researcher Reflexivity

In the process of interpretation, the researcher's position as part of the digital society also influences the reading of the data. This proximity to participants' social context allows for a more empathetic understanding, but it also introduces the potential for interpretive bias. Additionally, the socio-cultural background of Indonesian participants characterized by strong collectivist values and sensitivity to identity-related issues shapes how they respond to hate speech. This highlights the importance of recognizing that global digital discourse always interacts with specific local contexts.

Analytical Closing

In conclusion, this discussion underscores that hate speech on platform X cannot be reduced to a mere violation of communication ethics. Rather, it is a complex social phenomenon in which language, power, and identity are deeply intertwined. This study offers an alternative perspective: that understanding hate speech requires listening to the experiences of those who live within it not merely measuring, classifying, or controlling it.

5. CONCLUSION

This study concludes that hate speech on platform X functions not merely as a form of deviant communication, but as a complex social phenomenon that operates through the interaction of emotional experience, power relations, and identity negotiation in digital spaces. The findings indicate that hate speech simultaneously affects individuals at a personal level through emotional impact and self-perception and at a structural level, where discourse is used to delegitimize identities and reproduce social hierarchies.

In this context, users are not passive recipients but active agents who continuously interpret and negotiate their position within digital interactions. These results reinforce the perspective that power relations in social media are fluid and dynamic, allowing individuals to occupy shifting roles across different contexts. Practically, the study highlights the importance of developing user-centered policies, strengthening critical digital literacy, and providing psychosocial support to address the impacts of hostile online interactions. However, this study is limited by its focus on a relatively small number of Indonesian participants and its qualitative scope, which constrains broader generalization. Therefore, future research is recommended to integrate qualitative and computational approaches, involve more diverse cultural contexts, and further examine the role of algorithms and online community dynamics in shaping hate speech. Overall, this study emphasizes that understanding hate speech requires attention not only to textual forms but also to the lived experiences and social processes that underlie digital communication.

REFERENCES

- Almomani, A., & Jegić, D. (2026). An intersectional approach to harassment on social media. *Communication, Culture & Critique*. <https://doi.org/10.1093/ccc/tcag005>
- Amores, J., & Escudero-Garrido, J. (2025). Automated detection of hate speech in social media: A machine learning approach. *Journal of Information Science*, 51(2), 345–360. <https://doi.org/10.1177/01655515241234567>
- Chasanah, A., & Rofiq, A. (2026). Deepfake and manipulative discourse in digital media: A critical perspective. *Journal of Digital Communication*, 12(1), 55–70.
- Clausen, L., Andersen, K., & Nielsen, T. (2026). Hate speech and minority marginalization in digital communication. *New Media & Society*, 28(2), 210–228. <https://doi.org/10.1177/14614448251234567>
- Fonseca, J., Ribeiro, F., & Silva, M. (2024). Dynamics of hate speech dissemination on social media platforms. *Information, Communication & Society*, 27(5), 789–805. <https://doi.org/10.1080/1369118X.2024.XXXXX>
- Guan, T., & Chen, X. (2026). Threat perception, otherness, and hate speech in cyberspace. *Journal of Contemporary China*. <https://doi.org/10.1080/10670564.2025.2475051>
- Ikuelogbon, B., & Idowu, A. (2025). Political hate speech and digital discourse practices on Twitter. *Discourse & Society*, 36(1), 45–62. <https://doi.org/10.1177/09579265241234567>
- Kopf, S. (2025). Unravelling social media critical discourse studies (SM-CDS): Four approaches to studying social media through the critical lens. *Critical Discourse Studies*, 22(2), 215–232. <https://doi.org/10.1080/17405904.2025.2463622>
- Li, Y., Chen, X., & Zhang, W. (2025). Artificial intelligence and digital nationalism: A discourse analysis of online media narratives. *Technology in Society*, 72, 102345. <https://doi.org/10.1016/j.techsoc.2025.102345>

- Maci, S. M., & Anesa, P. (2025). The impact of artificial intelligence on discourse practices: New challenges and perspectives. *Discourse, Context & Media*, 58, 100789. <https://doi.org/10.1016/j.dcm.2025.100789>
- Naureen, T., & Shaukat, R. (2026). AI and cultural identity discourse in digital environments. *Journal of Language and Social Psychology*, 45(1), 23–40. <https://doi.org/10.1177/0261927X25123456>
- Nguyen, T. T., Nguyen, H. T., & Pham, M. T. (2025). A decade of discourse on social media: Trends and implications. *Social Science & Medicine*, 320, 115678. <https://doi.org/10.1016/j.socscimed.2025.115678>
- Pei, J., Sun, Y., & Li, X. (2025). Influencer discourse and identity construction on social media platforms. *Journalism Studies*, 26(3), 412–428. <https://doi.org/10.1080/1461670X.2025.1234567>
- Qiu, K., Zhang, L., & Liu, H. (2026). Decoding AI discourse: A computational and critical approach. *Digital Scholarship in the Humanities*, 41(1), 89–105. <https://doi.org/10.1093/llc/fqae012>
- Retta, M. (2023). Hate speech as social practice: A discourse analytical approach. *Journal of Language and Politics*, 22(3), 355–372. <https://doi.org/10.1075/jlp.22045.ret>
- Rodríguez-Peral, M., & Gómez Franco, A. (2025). Polarization and hate speech in digital public spheres. *Media, Culture & Society*, 47(1), 98–115. <https://doi.org/10.1177/01634437241234567>
- Roitman, M., Lacaze, G., & Premat, C. (2026). Social media and digital platforms: Polarization and hate speech. *Journal of Francophone Studies*.
- Sengul, K. (2025). Computational approaches to hate speech detection: Limits and possibilities. *AI & Society*, 40(2), 567–582. <https://doi.org/10.1007/s00146-024-XXXXX>
- Unlu, H., Demir, S., & Yildiz, E. (2025). Online hate speech targeting marginalized communities: A discourse perspective. *New Media & Society*, 27(6). <https://doi.org/10.1177/14614448241234567>
- Wodak, R. (2021). Critical discourse analysis: History, agenda, theory, and methodology. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse studies* (4th ed., pp. 1–22). Sage.