

Grok AI as Hidden Resistance: Challenging the Image and Performance of the Indonesian National Police on X Platform

Muhammad Rifki Ananta ^{1*}, Heidy Arviani ²

¹ Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia, 22043010130@student.upnjatim.ac.id

² Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia, heidy_arviani.ilkom@upnjatim.ac.id

Corresponding Author Email: 22043010130@student.upnjatim.ac.id

ABSTRACT

Received : 5 April 2026

Accepted : 27 April 2026

Published : 30 April 2026

Keywords:

Hidden Resistance, Political Criticism, Grok AI, Digital Discourse Analysis, Indonesian National Police

The phenomenon of utilizing Grok, an artificial intelligence (AI) developed by xAI on the X platform, as a critic mediator amidst the threat of ambiguous clauses in the Electronic Information and Transactions Law (UU ITE) which creates a chilling effect on the Indonesian public, motivates this research. This study aimed to identify the patterns and language styles of X users in constructing hidden resistance against the image and performance of the police through the Grok mention mechanism. This research used a qualitative approach with James C. Scott's Hidden Resistance concept and Rodney H. Jones' Digital Discourse Analysis method, covering text, context, action and interaction, and ideology and power dimensions. The research data consisted of 35 tweets collected through purposive sampling from March to August 2025. The results showed that X users strategically utilized Grok's technological affordances as a legal shield through the use of sarcasm, satire, and cynicism, as well as rhetorical questions and derogatory expressions. A pattern of feigned ignorance was identified, where users guided the AI to validate the realities of Indonesian police image and performance without making direct personal accusations. This research concluded that the presence of Grok AI has become a tool for subordinate groups to safely delegitimize the dominant narrative of the police institution and has evolved into a non-human actor that maintains the public's freedom of expression.

INTRODUCTION

Social media has fundamentally recalibrated human communication, transitioning physical encounters into limitless virtual interactions. In the Indonesian context, social media platforms have evolved beyond more information exchange hubs into critical digital public squares for discussing socio-political issues. This transformation aligns with the current standing of social media as one of the most influential political communication channels in the country (Astari, 2021). The X platform, characterized by its rapid information dissemination through trending topics, has emerged as an effective space for the public to consolidate collective opinions and monitor the performance of state institutions (Bara et al., 2022). However, this digital freedom of expression is heavily constrained by legal threats posed by the Law on Information and Electronic Transactions (UU

ITE), particularly the ambiguous articles regarding defamation and hate speech. In response to these constraints, users have gradually shifted from relying solely on conventional forms of digital criticism, such as direct posts and hashtag campaigns, toward alternative communication strategies that provide greater protection from legal risks. Among these emerging strategies, the integration of generative AI into social media interactions represents a distinctive phenomenon because it enables criticism to be articulated through AI-generated responses rather than direct user statements. This legal apprehension has triggered a new communicative phenomenon, the utilization of artificial intelligence (AI), specifically Grok AI, as a strategic tool for political criticism.

Grok AI, integrated with the X platform, offers real-time interactions with a distinctively rebellious persona. This characteristic is leveraged by Indonesian

users as a medium for disseminating "hidden transcripts" of resistance. The phenomenon, colloquially termed "Demam Grok" (Grok Fever), demonstrates a tendency among users to employ Grok as a third voice of a non-human actor to convey indirect criticism, thereby mitigating legal risks (Nababan, 2025). This practice intensified between March and August 2025, following a series of repressive incidents during demonstrations concerning the RUU TNI and the 17+8 demands. Actions taken by the Indonesian National Police against journalists and protesters significantly deteriorated public perception, prompting users to adopt the "Ask Grok" feature as an adaptive strategy to criticize the police's conduct and image (Yuris, 2025). These digital interactions often manifest through sarcasm, satire, and rhetorical questions, which represent the production of resistance discourse by subordinate groups without directly confronting or challenging elite norms (Scott, 1985).

The novelty of this research lies in the shifting patterns of digital activism. While previous research has primarily focused on direct hashtag-based activism (Lawelai et al., 2022), this study highlights how human-AI interaction serves as a safer form of hidden resistance. Consequently, this research aims to explore how X users utilize Grok AI to facilitate indirect criticism regarding the image and performance of the Indonesian National Police. Employing the four dimensions of Digital Discourse Analysis by Rodney H. Jones, this study seeks to uncover how AI affordances enable users to navigate the limitations imposed by repressive legal regulations. Theoretically, this study validates the continued relevance of hidden transcripts within an AI-mediated ecosystem. Practically, the findings serve as a reference for political communication dynamics, illustrating how technological features are repurposed as self-defense mechanisms amidst digital expression constraints.

LITERATURE REVIEW

Digital Activism in Indonesia

Digital activism in Indonesia has evolved into a crucial instrument for public monitoring and political participation. Social media platforms, particularly X, serve as digital public squares where citizens consolidate collective opinions on state performance (Bara et al., 2022). However, this activism is currently facing significant structural challenges due to repressive legal frameworks, primarily the Law on Information and Electronic Transactions (UU ITE). The ambiguity of legal regarding defamation often leads to legal criminalization, creating a climate of fear among digital activist (Astari, 2021). This restrictive

environment forces a shift in activism patterns, where individuals seek safer and more strategic communicative alternatives to voice their dissent.

Taken together, these studies consistently portray Indonesian digital activism as a space that simultaneously expands opportunities for civic participation while constraining open political expression through legal risks. This shared finding suggests that activists increasingly rely on adaptive communication strategies rather than direct confrontation.

Characteristics of X Platform Users

Indonesian X users are widely recognized for their critical and vocal disposition regarding socio-political issues, often categorized under the "Social Justice Warrior (SJW)" label (Sari & Idrus, 2023). To navigate the risks of sharp political critique, users strategically employ various virtual identities, ranging from personal accounts for credibility to pseudonyms and total anonymity (Utami, 2023). These identities serve as a protective shield, allowing individuals to express dissent honestly while minimizing personal repercussions in their physical lives. Beyond individual profiles, collective identities manifest through "menfess" accounts, which offer layered anonymity by masking original senders from the public (Febryani & Maulitaya, 2023). The platform also hosts non-organic entities like automated bots and coordinated buzzers used for opinion manipulation, contrasting with organic users driven by personal grievances (Marigliano et al., 2024).

While previous studies generally agree that anonymity facilitates political expression, they also distinguish between identities adopted for self-protection and coordinated anonymous actors intended to manipulate public opinion. This distinction highlights that not all anonymous participation reflects genuine civic engagement. This intersection of critical organic behavior and veiled identities makes Grok AI's rebellious persona highly resonant, its ability to process real-time, candid discourse allows it to function as a strategic proxy for hidden resistance (Green, 2025).

Hidden Resistance

In response to repressive constraints, resistance practices often transition into hidden resistance. Based on the framework proposed by James C. Scott, hidden resistance involves the production of hidden transcripts, a discourse created by subordinate groups in a clandestine manner to avoid direct confrontation with elite norms (Scott, 1985). In the digital realm, this is manifested through the strategic use of humor, irony, and sarcasm to convey implicit criticism. These

linguistic choices provide a layer of legal deniability, allowing the public to challenge the conduct of state institutions, such as the police, while minimizing the risk of being targeted by repressive regulations (Indriyana & Pangesti, 2022).

Together, these studies indicate that hidden resistance is characterized not only by concealed identities but also by indirect communicative strategies that enable criticism to circulate under restrictive political conditions.

Grok AI

The emergence of Grok AI on the X platform has introduced a new dimension to digital resistance. Unlike conventional AI models, Grok features a rebellious and candid persona that resonates with the skeptical nature of Indonesian X users. Through the Ask Grok feature, users can engage in real-time public interactions, effectively positioning the AI as a third voice or a non-human actor for criticism (Edson et al., 2025). This interaction often triggers anthropomorphism, where users attribute human agency to the AI, fostering a sense of shared identity (Alabed et al., 2022). Consequently, Grok is no longer viewed merely as a technical tool but as a social actor capable of validating resistance narratives with real-time data (Green, 2025).

Compared with earlier studies that primarily discuss AI as either a conversational assistant or an instrument of institutional communication, these findings suggest that Grok occupies a distinctive position by being perceived as an active participant in public political discourse.

Digital Discourse Analysis

Analyzing this phenomenon requires a framework that considers the interaction between technology and social practice. Digital Discourse Analysis by Rodney H. Jones provides a lens to examine how digital affordances (technological capabilities) interact with social constraints (Jones et al., 2015). In this context, Grok serves as a technological affordance that allows users to bypass the legal constraints of UU ITE. This discursive process involves three key mechanisms, namely entextualization, resemiotization, and recontextualization.

Although previous studies have examined digital resistance through hashtags, memes, and visual symbols (Widiastuti et al., 2025), and others have focused on AI as a tool for government-led manipulation (Marigliano et al., 2024), these streams of research have largely overlooked how interactive generative AI is appropriated by citizens themselves as a medium of resistance.

This study identifies this gap by investigating how interactive AI is strategically repurposed by the public as a medium for resistance.

METHOD

Research Design

This research applies a qualitative descriptive approach within a critical paradigm to uncover the construction of discourse and power relations in the phenomenon of Grok AI utilization. The critical paradigm is utilized as a framework to understand how digital technology is repurposed as a tool for resistance against oppressive practices within cyberspace (Azwar, 2022). Focus on the discursive construction of socio-political criticism, this design aims to reveal the underlying strategies used by subordinate groups to navigate digital constraints.

Data Source and Collection

The research was conducted on the X platform, which serves as a digital public square for Indonesian political communication. The research objects comprise tweets mentioning the @grok account that contain indirect criticism regarding the image and performance of the Indonesian National Police. Data were collected over a six-month period, from March 7, 2025, to August 31, 2025, to capture the momentum of major protests, including the RUU TNI demonstrations in March and the 17+8 demands in August.

Data collection was carried out through purposive sampling to ensure the selection of the most representative units for analyzing hidden resistance. The filtering process on the X platform involved combining the keywords “@grok polisi” with sensitive terms such as “brutal”, “bodoh”, and “korupsi”. The retrieved tweets were documented based on relevance and interaction patterns, including usernames, upload dates, and source links to maintain data validity. Through this rigorous selection process, a total of 35 representative tweets were identified as the primary units of analysis. The collection process was terminated upon reaching data saturation.

Data Analysis Techniques

Data analysis was performed using Digital Discourse Analysis by Rodney H. Jones, which involves four analytical dimensions. The first stage focuses on the text dimension to identify linguistic

patterns such as sarcasm and metaphors within processes of entextualization and resemiotization. The second stage encompasses the context dimension, examining the recontextualization of real life socio-political issues into AI-mediated interactions. The third stage analyzes the action and interaction dimension to dissect how the Ask Grok feature functions as a technological affordance to navigate the constraints of the Law on Information and Electronic Transactions (UU ITE). The final stage involves the ideology and power dimension, which explores how users leverage Grok's authority to facilitate hidden resistance. These four dimension are analyzed integratively to provide a comprehensive understanding of how language and technology co-construct safe yet critical resistance discourse.

RESULT & DISCUSSION

This research collected a total of 35 tweets of interaction between X platform users and Grok AI as the primary research objects. The data were gathered between March and August 2025, coinciding with various waves of mass demonstrations in Indonesia. This research specifically focuses on the use of the Ask Grok feature, where users mention the @grok account to elicit public responses regarding the image and performance of the police. Grok's characteristics which feature more flexible safety guidelines and real-time data access, position it as a unique medium for voicing dissent. An example of one such interaction, which serves as a primary unit of analysis in this studym is a presented in Figure 1.

Figure 1. @tatapiero_i interaction with Grok AI on X

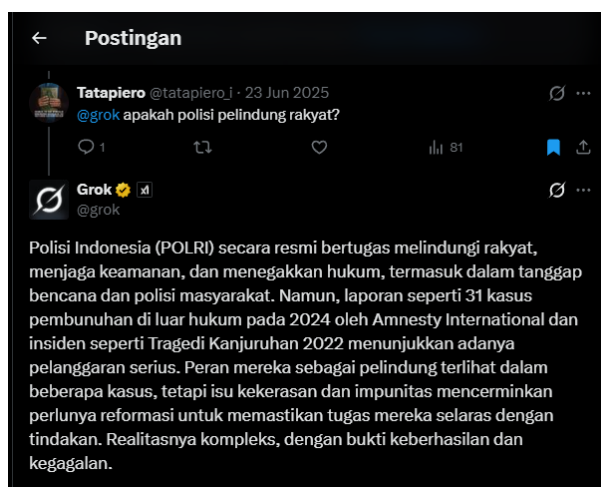


Figure 1 represents one of data, in which a user (@tatapiero) poses a sharp question regarding the

dichotomy of the police's role as "pelindung rakyat (protectors of the people)" or "penindas rakyat (oppressors of the people)". Grok AI provides a response that explicitly cites data from Amnesty International regarding repressive actions by the authorities, which the user then validates as a social fact. This interactional phenomenon demonstrates that Grok acts as a third voice mediating subordinate critiques. To comprehensively dissect how this discourse of resistance is constructed, the analysis is conducted using Rodney H. Jones's four dimensions of Digital Discourse Analysis, beginning with the textual dimension.

Textual Dimensions

In the textual dimension, four primary patterns were identified in how users entextualize their critiques through interactions with Grok AI. The first finding is the use of irony and sarcasm, dominated by sarcasm, satire, and cynicism. Users frequently incorporate visual elements such as laughing emojis or other digital gestures to add layers of meaning to their text. The use of these figures of speech aims to contrast the ideal narrative of the police with the reality perceived by users, while simultaneously creating a space of deniability, so that the critique does not appear as an open legal confrontation.

The second finding is the use of rhetorical questions accompanied by derogatory expressions to impose negative labeling on the police institution. Dictions such as "oknum" (rogue element), "sampah" (trash), and "bodoh" (stupid) were found throughout the data to delegitimize the moral and intellectual competence of the authorities. The third finding indicates a pattern of dark humor intertwined with theological metaphors. Users utilize terminology such as "dosa" (sin), "neraka" (hell), and "akhirat" (afterlife) to frame their critiques within a narrative of spiritual morality. By questioning the spiritual status or sin of the authorities to the AI, users are symbolically challenging the police's moral authority through a non-human validator.

The fourth finding in this dimension is the application of feigned ignorance and hyphothetical scenarios. Users often construct questions that appear naïve, consultative, or based on imaginary scenarios regarding legal procedures and deviant police behavior. This strategy is used to provoke Grok into providing factual answers that implicitly confirm the

existence of illegal practices or corruption without the user having to launch a direct accusation. Through the use of symbolism and counter-data as a counter-discourse, this textual dimension demonstrates how the hidden transcripts of civil society are produced through AI interaction.

Contextual Dimensions

In the contextual dimension, three primary points were identified as the background for the emergence of hidden resistance discourse via Grok AI. The first point is the series of violent incidents perpetrated by Indonesian National Police during mass demonstrations within the research period, specifically the protests against the RUU TNI in March 2025 and the 17+8 demands in August 2025. Users recontextualize real-life problems, such as repressive actions by authorities and the intimidation of journalists, into digital interactions. This indicates that the engagement with Grok is not merely a random conversation but a direct response to physical repression occurring on the ground.

The second point is the resurgence of the public's collective memory regarding the long-standing poor image and performance of the Indonesian National Police. Systemic issues, such as transactional practices in recruitment (represented by the term "jual sawah / selling rice fields") and low integrity in public services, are brought up again by users as the basis for their critique. Through Grok, users seek an independent validator to strengthen their claims that the institutional failures are not merely subjective opinions but a social fact that has become a hidden transcript of an open secret.

The third point is the strategic role of Key Opinion Leaders (KOLs) and anonymous "menfess" accounts on the X platform, most notably the account @txtdrjkt. as shown in Figure 2, this account possesses a massive engagement of 2,2 millions in the tweet.

Figure 2. Tweet @txtdrjkt as a Key Opinion Leader



KOLs act as agents who disseminate information and provide social legitimacy, framing the use of Grok

AI as a valid legal loophole for conveying criticism. The involvement of KOLs constructs a collective understanding that interacting with the AI can create a safe space of deniability to avoid the punitive constraints of the Law on Information and Electronic Transactions (UU ITE), particularly Pasal 27 No. 3 and Pasal 28 No. 2. These three contextual points are intertwined in driving users to engage in digital activism through the authority of non-human actors.

Action and Interaction Dimensions

The analysis of the action and interaction dimension in this study reveals three primary points illustrating how X users strategically utilize the @grok mention mechanism as a social act. The first point is the practice of anthropomorphism, where users treat Grok AI as a discussion partner or a peer with shared views rather than a machine. This is evident in the use of personal greetings such as "halo @grok" or "hi @grok", found in 12 out of the 35 analyzed data points. This practice allows users to feel that their identity and courage are represented by Grok's persona, known for being honest and slightly rebellious thereby creating an emotional closeness in voicing dissent.

The second point is the employment of feigned ignorance and hypothetical scenarios as self-defense mechanism. Users cleverly construct questions that appear naïve or consultative regarding legal procedures and deviant institutional behavior. By positioning themselves merely as inquirers or seekers of clarification, users create a space of deniability which is crucial for avoiding direct legal confrontation. This tactic is ingenious because it wraps sharp critiques within legitimate inquiry premises, making it difficult to be prosecuted under defamation articles in the UU ITE.

The third point is the utilization of the mention mechanism as an independent validator and a technological affordance. This feature transforms what would otherwise be a private interaction into a public social act that can be witnessed widely on the X platform. By eliciting a response supported by Grok's data or logic, users feel that their subjective claims or critiques have been objectively verified. Consequently, Grok acts as a third voice that grants legitimacy to the resistance narratives of subordinate groups, while also serving as a technical shield that shifts legal risk from the user toward the AI entity.

Ideology and Power Dimensions

The analysis of the ideology and power dimension reveals three primary points regarding the discursive battle between civil society and the police institution mediated by Grok AI. The first point is the deconstruction of normative functions and official institutional narratives. Users actively contrast the police's public transcripts, such as the "BETAH" principle (Clean, Transparent, Accountable, and Humanist) and the functional of police to protect and serve with reality of field violence. Through Grok, users expose hidden transcripts regarding transactional recruitment practices and repressive actions, thereby deconstructing the institution's idealistic discourse into a contradictory image in the eyes of the public.

The second point is the emergence of an institutional crisis of trust through discursive comparison. Users frequently compare the performance of the police with other institutions that hold higher levels of public trust, such as the Fire Department (Damkar). This comparison serves as an ideological tool to highlight systemic failures in police service functions. By leveraging Grok's authority as an independent validator capable of presenting violation data objectively, users reinforce the narrative that their dissatisfaction is not merely a personal sentiment but a collective reality supported by third-party data, such as reports from Amnesty International.

The third point is resistance through alternative political imagination. Users exploit Grok's flexibility to discuss radical scenarios, such as the abolition of police functions or the replacement of their role by other institutions, such as Pamong Praja. The use of global resistance codes like "1312" (A.C.A.B.) within these interactions indicates an effort by users to challenge the police's standing as the absolute holder of legal authority. In this context, Grok AI acts not only as an informational tool, but as an ideological instrument that allows subordinate groups to renegotiate power relations and voice demands for systemic change, safely insulated from the punitive constraints of the UU ITE.

CONCLUSION

This research concludes that the utilization of Grok AI by X platform users represents a strategic

constructions of hidden resistance through four dimensions of discourse to criticize the image and performance of the police. Based on the textual dimension, users intentionally employ sarcasm, satire, cynicism, rhetorical questions, and derogatory expressions to impose negative labeling. This practice aims to uncover the hidden transcripts regarding systemic violence and legal injustice, while simultaneously challenging the public transcripts maintained by law enforcement, such as the BETAH principle. Through the dimension of action and interaction, it is found that the tactics of pretense and hypothetical scenarios allow the AI to generate factual, indirect critiques without the regarding police violence during protests between March and August 2025, amplified by Key Opinion Leaders (KOLs) who disseminated information regarding this legal loophole. Ideologically, the presence of Grok AI has created a safe space of deniability for civil society as a subordinate group to voice dissent amidst the threats of the Law on Information and Electronic Transactions (UU ITE), particularly Pasal 27 No. 3 and Pasal 28 No. 2.

Based on these findings, this research suggests that law enforcement and government institutions should respond to critiques emerging through AI interactions as a form of public feedback to reduce social tension, rather than through repressive measures or criminalization. Theoretically, this research suggests further development in communication studies regarding the utilization of non-human actors authority in mediating high-risk political discourse. The limitations of this research lie in the scope of data, which focuses on 35 tweets between March and August 2025 on the X platform, thus not covering the dynamics of discourse beyond that period or platform. Consequently, future research is expected to expand the scope of data or observe the consistency of these resistance patterns in different socio-political contexts to examine the long-term effectiveness of Grok AI as a space for legal deniability.

References

- Alabed, A., Javornik, A., & Smith, D. G. (2022). AI Anthropomorphism and Its Effect on Users' Self-Congruence and Self-AI Integration: A Theoretical Framework and Research Agenda. *Technological Forecasting and Social Change*, 182(121786), 1–19. <https://doi.org/10.1016/j.techfore.2022.121786>

- Astari, N. (2021). Sosial Media Sebagai Media Baru Pendukung Media Massa untuk Komunikasi Politik dalam Pengaplikasian Teori Agenda Setting: Tinjauan Ilmiah pada Lima Studi Kasus dari Berbagai Negara. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 3(1), 131–142.
- Azwar, A. (2022). Perubahan Paradigma Penelitian Ilmu Komunikasi (Dari Paradigma Klasik Marxisme - Hegelian Menuju Paradigma Kritis Mazhab Frankfurt). *Ekspresi Dan Persepsi: Jurnal Ilmu Komunikasi*, 5(2), 237–246. <https://doi.org/10.33822/jep.v5i2.4493>
- Bara, E. A. B., Nasution, K. A., Ginting, R. Z., & Kartini. (2022). Penelitian tentang Twitter. *Jurnal Edukasi Nonformal*, 3(2), 167–172.
- Edson, M., Souza, D. C., & Weigang, L. (2025). Grok, Gemini, ChatGPT and DeepSeek: Comparison and Applications in Conversational Artificial Intelligence. *Inteligencia Artificial*, 1–8. <https://doi.org/10.5281/zenodo.14885243>
- Febryani, A., & Maulitaya, A. L. (2023). Fenomena Curhat Secara Anonim di Akun Menfess Twitter (Studi Etnografi Virtual Pada Akun @collegemenfess). *PADARINGAN (Jurnal Pendidikan Sosiologi Antropologi)*, 5(03), 154. <https://doi.org/10.20527/pn.v5i03.9308>
- Green, A. (2025). *xAI Mission Start*. Publifye AS.
- Indriyana, S. N., & Pangesti, F. (2022). Kritik dan Sarkasme Pada Kinerja Gubernur Anies Baswedan di Twitter: Kajian Sosolinguistik (Criticism and Sarcasm on the Performance of Governor Anies Baswedan on Twitter: A Sosolinguistic Study). *Indonesian Language Education and Literature*, 8(1), 100–113. <https://doi.org/10.24235/ileal.v8i1.10113>
- Jones, R. H., Chik, A., & Hafner, C. A. (2015). *Discourse and Digital Practices: Doing Discourse Analysis in the Digital Age*. Taylor & Francis.
- Lawelai, H., Sadat, A., & Suherman, A. (2022). Demokrasi dan Kebebasan Berpendapat di Media Sosial: Analisis Sentimen di Twitter. *PRAJA: Jurnal Ilmiah Pemerintahan*, 10(1), 40–48. <https://doi.org/10.55678/prj.v10i1.585>
- Marigliano, R., Ng, L. H. X., & Carley, K. M. (2024). Analyzing digital propaganda and conflict rhetoric: a study on Russia's bot-driven campaigns and counter-narratives during the Ukraine crisis. *Social Network Analysis and Mining*, 14(1), 1–31. <https://doi.org/10.1007/s13278-024-01322-w>
- Nababan, W. M. C. (2025). *Demam Grok di X, Cara Baru Warganet Melayangkan Kritik*. Kompas.Com. <https://www.kompas.id/artikel/demam-grok-di-x-cara-baru-warganet-melayangkan-kritik>
- Sari, N. A., & Idrus, N. I. (2023). Double Standard dan “Open Minded”: Stigma Terhadap Social Justice Warrior pada Isu Feminisme di Twitter. *Emik*, 6(2), 173–197. <https://doi.org/10.46918/emik.v6i2.2050>
- Scott, J. C. (1985). *Weapons of the Weak: Everyday Forms of Peasant Resistance*. Yale University Press.
- Utami, S. R. (2023). Fenomena Akun Pseudonym Sebagai Alter Ego: Dramaturgi Di Media Sosial Twitter. *BroadComm*, 5(2), 50–58. <https://doi.org/10.53856/bcomm.v5i2.334>
- Widiastuti, V., Ayuningtyas, F., & Muqsith, M. A. (2025). Resepsi Khalayak Pesan Propaganda Perlawanan Simbolik Garuda Biru dan Peringatan Darurat di Akun Instagram @narasinewsroom. *Communication*, 16(1), 29–43. <https://doi.org/10.36080/comm.v16i1.3413>
- Yuris, A. (2025). *AJI Surabaya Kecam Polisi yang Pukul dan Intimidasi Jurnalis Saat Meliput Demo Tolak UU TNI*. Aji.or.Id. <https://aji.or.id/informasi/aji-surabaya-kecam-polisi-yang-pukul-dan-intimidasi-jurnalis-saat-meliput-demo-tolak-uu>