

**DIALOG AKAL SEHAT ROCKY GERUNG – USTADZ ABDUL SOMAD:  
TOXICITY COMMENT ANALYSIS PADA PLATFORM YOUTUBE**



**Disusun Oleh:**  
**Ade Safitri**  
**NIM. 23202011011**

**TESIS**

**Diajukan Kepada Program Magister di Fakultas  
Dakwah dan Komunikasi UIN Sunan Kalijaga untuk Memenuhi Salah Satu  
Syarat Penyusunan Tesis**

**YOGYAKARTA**  
**2025**

## SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Ade Safitri  
NIM : 23202011011  
Prodi : Komunikasi Penyiaran Islam  
Fakultas : Dakwah dan Komunikasi

Menyatakan dengan sesungguhnya, bahwa tesis saya yang berjudul:  
**Dialog Akal Sehat Roky Gerung-Ustadz Abdul Shomad Toxcity Comment Analisis pada Platfom Youtube** adalah hasil karya pribadi yang tidak mengandung plagiarisme, kecuali bagian- bagian tertentu yang penyusun ambil sebagai acuan dengan tata cara yang dibenarkan secara ilmiah.

Apabila terbukti pernyataan ini tidak benar, maka penyusun siap mempertanggungjawabkannya sesuai hukum yang berlaku.

Yogyakarta,  
Yang menyatakan,



Ade Safitri  
23202011011

## SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Ade Safitri  
NIM : 23202011011  
Prodi : Komunikasi Penyiaran Islam  
Fakultas : Dakwah dan Komunikasi

Menyatakan dengan sesungguhnya, bahwa tesis saya yang berjudul: **Dialog Akal Sehat Roky Gerung-Ustadz Abdul Shomad Toxcity Comment Analisis pada Platfom Youtube** adalah hasil karya pribadi yang tidak mengandung plagiarisme, kecuali bagian- bagian tertentu yang penyusun ambil sebagai acuan dengan tata cara yang dibenarkan secara ilmiah.

Apabila terbukti pernyataan ini tidak benar, maka penyusun siap mempertanggungjawabkannya sesuai hukum yang berlaku.

Yogyakarta,  
Yang menyatakan,



Ade Safitri  
23202011011



**PENGESAHAN**  
**KEMENTERIAN AGAMA**  
**UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA**  
**FAKULTAS DAKWAH DAN KOMUNIKASI**  
Jl. Marsda Adisucipto Telp. (0274) 515856 Fax. (0274) 552230 Yogyakarta 55281

**PENGESAHAN TUGAS AKHIR**

Nomor : B-158/Un.02/DD/PP.00.9/01/2026

Tugas Akhir dengan judul : Dialog Akal Sehat Rocky Gerung-Ustadz Abdul Shomad: *Toxicity Comment* Analisis pada Platform Youtube

yang dipersiapkan dan disusun oleh:

Nama : ADE SAFITRI, S.Sos  
Nomor Induk Mahasiswa : 23202011011  
Telah diujikan pada : Selasa, 06 Januari 2026  
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Dakwah dan Komunikasi UIN Sunan Kalijaga Yogyakarta

**TIM UJIAN TUGAS AKHIR**



Valid ID: 69759c823e5dc

Ketua Sidang/Penguji I  
Dr. Hamdan Daulay, M.Si., M.A.  
SIGNED



Valid ID: 6977659b788d3

Penguji II  
Dr. Khadiq, S.Ag., M.Hum  
SIGNED



Valid ID: 6977361d4e1ea

Penguji III  
Prof. Alimatul Qibtiyah, S.Ag., M.Si., Ph.D.  
SIGNED



Valid ID: 69786f6dacc46

Yogyakarta, 06 Januari 2026  
UIN Sunan Kalijaga  
Dekan Fakultas Dakwah dan Komunikasi  
Prof. Dr. Arif Maftuhin, M.Ag., M.A.I.S.  
SIGNED

## HALAMAN PERSETUJUAN

Kepada Yth.  
Dekan Fakultas Dakwah dan Komunikasi  
UIN Sunan Kalijaga Yogyakarta  
Di Yogyakarta

*Assalamualaikum wr.wb.*

Setelah membaca, meneliti, memberikan petunjuk, dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku dosen pembimbing berpendapat bahwa tesis Saudara:

Nama : Ade Safitri  
NIM : 23202011011  
Judul Tesis : Dialog Akal Sehat Roky Gerung – Ustadz Abdul Shomad  
Toxcity Comment Analisis pada Platfom Youtube

tesis tersebut sudah memenuhi syarat

- Bebas dari unsur plagiarisme.
- Hasil pemeriksaan similaritas melalui Turnitin menunjukkan tingkat kemiripan sebesar 15% dengan menggunakan setelan “small match exclusion” sepuluh kata.
- Sistematika penulisan telah sesuai dengan Pedoman Penulisan Tesis yang berlaku. dan sudah dapat diajukan kepada Program Studi Komunikasi Penyiaran Islam Fakultas Dakwah dan Komunikasi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Magister.

Dengan ini kami mengharap agar tesis tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

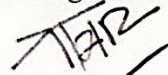
Yogyakarta,



Dr. Hamdan Daulay, M.Si., M.A  
NIP 196612091994031004

Mengetahui:

Ketua Program Studi



Dr. Abdul Rozak, M.Pd.  
NIP 196710061994031003

Silakan beri tanda centang ( ) jika pernyataan telah sesuai.

## **MOTTO**

“You don’t have to be great to start, but you have to start to be great”

## **HALAMAN PERSEMBAHAN**

**Bismillahirrahmanirrahim, Ku persembahkan karya sederhana ini kepada orang-orang yang sangat ku cintai:**

### **MAMA**

Terima kasih atas segala cinta, kesabaran, dan jerih payah dan perjuangan yang telah Mama berikan selama ini. Meskipun Mama tidak pernah merasakan dunia perguruan tinggi, tapi Mama selalu berjuang tanpa lelah untuk memastikan anak sematawayangnya mendapatkan pendidikan yang terbaik.

Mama, terima kasih atas segala do'a, perjuangan, dan kasih sayang yang tiada henti. Semoga sedikit pencapaian ini bisa menjadi kebanggaan untuk Mama.

Mama, Siti Hafsa adalah bukti bahwa kerasnya dunia bukan alasan untuk menyerah, tetap tapak kan kaki pada bumi dan bersujud kepada ilahi , semua kesulitan akan allah mudah kan

Jika lelah, pulang lah ada pelukkan untuk kembali semangat .

Mama, salah satu bukti bahwa ketulusan dan kasih sayang bahkan cinta tulus tanpa pamrih

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb*

*Alhamdulillah* *rabbi'l'alam* puji syukur atas kehadiran Allah subhanahu wata'ala yang telah melimpahkan rahmat serta hidayahnya serta anugrah yang telah diberikan dan tak terhitung jumlahnya, sholawat beriring salam saya curahkan kepada Rasulullah Shalallahu'alaihi Wasallam yang telah mengajarkan suri tauladan, dan yang telah membawa kita dari jaman jahiliyah menuju jaman modern seperti yang kita rasakan saat ini dengan kemudahannya sehingga penulis dapat menyelesaikan tesis dengan judul : **“Dialog Akal Sehat Rocky Gerung – Ustadz Abdul Somad: Toxicity Comment Analysis Pada Platform Youtube”**.

Tesis ini guna disusun melengkapi persyaratan dalam menyelesaikan kelulusan studi pada program Magister (S2) Fakultas Dakwah dan Komunikasi Prodi Komunikasi dan Penyiaran Islam Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Oleh karena itu, hal yang pantas penulis ucapkan adalah kata terimakasih yang teramat dalam kepada semua pihak yang telah mendukung serta membimbing dan yang turut membantu dalam menyelesaikan tesis ini, terutama sekali kepada yang terhormat:

1. Dengan penuh rasa syukur, saya mengucapkan ribuan terima kasih yang sebesar-besarnya kepada Ibu Saya, **Mama Siti Hafsah** yang telah memberikan dukungan sayang, cinta, dan doa'a yang tanpa henti disetiap langkah perjalanan pendidikan ini. Pengorbanan, perjuangan dan do'a-do'a mama adalah sumber inspirasi yang tak ternilai bagi saya, dan tanpa mama pencapaian ini tidak mungkin terwujud.

2. Prof. Noorhaidi, M.A., M.Phil., Ph.D. selaku Rektor UIN Sunan Kalijaga Yogyakarta yang telah memberikan kesempatan untuk menempuh pendidikan di Program Studi Magister Komunikasi dan Penyiaran Islam Fakultas Dakwah dan Komunikasi.
3. Prof. Dr. Arif Maftuhin M.Ag., M.A.I.S. selaku Dekan Fakultas Dakwah dan Komunikasi UIN Sunan Kalijaga Yogyakarta yang telah memberikan kesempatan untuk menempuh pendidikan di Program Studi Magister Komunikasi dan Penyiaran Islam
4. Dr. Hamdan Daulay, M.Si., M.A. selaku dosen pembimbing yang baik dan terbaik serta selalu memberikan motivasi semangat dan ilmu baru serta wawasan tentang materi penulisan ini, semoga Allah berikan kesehatan untuk bapak dan keluarga besar
5. Drs. Abdul Rozak, M.Pd. selaku Ketua Prodi Magister Komunikasi dan Penyiaran Islam Fakultas Dakwah dan Komunikasi UIN Sunan Kalijaga Yogyakarta,
6. Ucapan terimakasih yg tak terhingga untuk SATU NAMA yang tidak bisa disebutkan, semoga senantiasa Allah berikan kemudahan serta kesehatan serta panjang umur karena sudah ikut andil dalam setiap progres penulis dalam menyelesaikan pendidikan ini.
7. Semua pihak yang terlibat dalam penyusunan tesis ini, baik langsung maupun tidak langsung yang tidak bisa disebutkan satu persatu.

Dan akhirnya semoga Allah Subhanahu Wa Ta'ala berkenna membalas segala kebaikan dan amal semua pihak yang telah membantu. Semoga tesis ini bermanfaat bagi pengemban ilmu.

Yogyakarta, 11 Desember 2025

Penulis

Ade Safitri  
NIM. 23202011011

## ABSTRAK

Penelitian ini menganalisis komentar netizen terhadap video “Dialog Akal Sehat” yang menampilkan percakapan antara Ustadz Abdul Somad (UAS) dan Rocky Gerung di platform YouTube. Menggunakan pendekatan Big Data Analytics dan alat bantu Communalytic, sebanyak 2.921 komentar dianalisis berdasarkan enam indikator toksisitas: Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, dan Threat. Hasil penelitian menunjukkan bahwa komentar-komentar tersebut secara umum berada dalam tingkat toksisitas yang rendah hingga sedang, dengan indikator Threat dan Severe Toxicity menempati skor paling rendah secara konsisten. Sebaliknya, indikator Toxicity dan Insult mendominasi, terutama pada momen-momen politis seperti menjelang Pemilu 2024. Visualisasi wordcloud dan emoji cloud memperkuat temuan bahwa wacana dalam kolom komentar didominasi oleh respons positif, dukungan terhadap tokoh, serta ekspresi religius dan politis. Pola waktu interaksi menunjukkan puncak partisipasi dalam tiga minggu pertama setelah unggahan video, diikuti dengan penurunan signifikan. Analisis “Top Ten Posters” mengungkap adanya konsentrasi produksi komentar pada akun-akun tertentu, yang berpotensi mendominasi arah diskursus. Temuan ini memperlihatkan pentingnya mengkaji diskursus digital secara multidimensi untuk memahami dinamika komunikasi publik, sekaligus sebagai masukan bagi kebijakan moderasi konten dan literasi digital di Indonesia.

**Kata Kunci:** Analisis Komentar, Toksisitas, YouTube, Diskursus Digital, UAS, Rocky Gerung

## ABSTRACT

This study analyzes public comments on the YouTube video “Dialog Akal Sehat,” featuring a conversation between Ustadz Abdul Somad (UAS) and Rocky Gerung. Using a Big Data Analytics approach and the CommuAnalytic tool, a total of 2,921 comments were examined across six toxicity indicators: Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat. The findings reveal that the overall toxicity level of comments ranged from low to moderate, with Threat and Severe Toxicity consistently scoring the lowest. In contrast, Toxicity and Insult appeared most frequently, especially around politically sensitive periods such as the lead-up to the 2024 Presidential Election in Indonesia. Wordcloud and emojiCloud visualizations further supported the dominance of positive sentiment, expressions of support for the speakers, as well as religious and political discourse. Temporal interaction patterns show peak engagement in the first three weeks following the video’s publication, followed by a sharp decline. The “Top Ten Posters” analysis indicates a concentration of comment production among specific accounts, potentially influencing the direction of online discussions. These findings underscore the importance of multidimensional analysis of digital discourse to better understand public communication dynamics, while offering valuable insights for content moderation and digital literacy policies in Indonesia.

**Keyword: Comment Analysis, Toxicity, Youtube, Digital Discourse, UAS, Rocky Gerung**

## DAFTAR ISI

|                                                                |             |
|----------------------------------------------------------------|-------------|
| <b>HALAMAN JUDUL .....</b>                                     | <b>i</b>    |
| <b>SURAT PERNYATAAN KEASLIAN.....</b>                          | <b>ii</b>   |
| <b>SURAT PERNYATAAN BEBAS PLASGIASI.....</b>                   | <b>iii</b>  |
| <b>PENGESAHAN .....</b>                                        | <b>iv</b>   |
| <b>HALAMAN PERSETUJUAN .....</b>                               | <b>v</b>    |
| <b>MOTTO .....</b>                                             | <b>vi</b>   |
| <b>HALAMAN PERSEMBAHAN .....</b>                               | <b>vii</b>  |
| <b>KATA PENGANTAR.....</b>                                     | <b>viii</b> |
| <b>ABSTRAK .....</b>                                           | <b>xi</b>   |
| <b>ABSTRACT .....</b>                                          | <b>xii</b>  |
| <b>DAFTAR ISI.....</b>                                         | <b>xiii</b> |
| <b>DAFTAR TABEL .....</b>                                      | <b>xv</b>   |
| <b>DAFTAR GAMBAR.....</b>                                      | <b>xvi</b>  |
| <b>BAB I PENDAHULUAN.....</b>                                  | <b>1</b>    |
| A. Latar Belakang .....                                        | 1           |
| B. Rumusan Masalah .....                                       | 8           |
| C. Tujuan Penelitian.....                                      | 9           |
| D. Manfaat dan Kegunaan .....                                  | 9           |
| E. Kajian Pustaka.....                                         | 11          |
| 1. Media Sosial dan YouTube sebagai Ruang Publik Digital ..... | 11          |
| 2. Komentar Toksik di Media Sosial.....                        | 29          |
| 3. Penelitian Relevan.....                                     | 41          |
| F. Kajian Teori.....                                           | 51          |
| 1. Online Disinhibition Effect Theory.....                     | 51          |
| 2. Uses And Gratifications Theory (UGT).....                   | 54          |
| 3. Kerangka Berpikir .....                                     | 56          |
| G. Metodologi Penelitian .....                                 | 58          |
| 1. Jenis Penelitian.....                                       | 58          |
| 2. Fokus Penelitian .....                                      | 58          |
| 3. Teknik Pengumpulan Data .....                               | 60          |

|                                                                                                                                   |            |
|-----------------------------------------------------------------------------------------------------------------------------------|------------|
| 4. Teknik Validasi Data.....                                                                                                      | 62         |
| 5. Teknik Analisis Data.....                                                                                                      | 63         |
| H. Sistematika Pembahasan .....                                                                                                   | 67         |
| <b>BAB II GAMBARAN UMUM .....</b>                                                                                                 | <b>70</b>  |
| A. Profile Ustadz Abdul Somad dan Rocky Gerung .....                                                                              | 70         |
| 1. Ustadz Abdul Somad.....                                                                                                        | 70         |
| 2. Rocky Gerung .....                                                                                                             | 76         |
| <b>BAB III DIALOG AKAL SEHAT ROCKY GERUNG – USTADZ ABDUL<br/>SOMAD: TOXICITY COMMENT ANALYSIS PADA PLATFORM<br/>YOUTUBE .....</b> | <b>84</b>  |
| A. Distribusi Toksisitas.....                                                                                                     | 84         |
| 1. Deskripsi Umum Datasets.....                                                                                                   | 84         |
| 2. Distribusi Skor Toksisitas .....                                                                                               | 86         |
| 3. Tren Bulanan Rata-Rata Skor Toksisitas pada Komentar Video .....                                                               | 108        |
| B. Distribusi Skor Toksisitas: Variasi Intensitas Komentar Publik.....                                                            | 113        |
| 1. Visualisasi Wordcloud dan Pola Diskursus .....                                                                                 | 116        |
| 2. Pola Waktu dan Puncak Interaksi.....                                                                                           | 120        |
| 3. Kontribusi Interaksi oleh Top Ten Posters.....                                                                                 | 123        |
| <b>BAB IV PENUTUP .....</b>                                                                                                       | <b>126</b> |
| A. Kesimpulan .....                                                                                                               | 126        |
| B. Saran.....                                                                                                                     | 127        |
| <b>DAFTAR PUSTAKA.....</b>                                                                                                        | <b>129</b> |
| <b>LAMPIRAN</b>                                                                                                                   |            |
| <b>CURRICULUM VITAE</b>                                                                                                           |            |

## DAFTAR TABEL

|                                           |    |
|-------------------------------------------|----|
| Tabel 1 Kajian Terdahulu .....            | 42 |
| Tabel 2 Online Disinhibition Effect ..... | 53 |
| Tabel 3 Toxicity Comment Element.....     | 64 |

## DAFTAR GAMBAR

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| Gambar 1 Bagan Kerangka Pemikiran Penelitian.....                  | 57  |
| Gambar 2 UAS Sedang Berceramah di Kabupaten Rokan Hilir, Riau..... | 71  |
| Gambar 3 Rocky Gerung .....                                        | 77  |
| Gambar 4 Tampilan Awal Dashbord App COMmunalytic.org.....          | 85  |
| Gambar 5 Toxicity Scores Table .....                               | 87  |
| Gambar 6 Ten posts with the highest toxicity scores .....          | 91  |
| Gambar 7 Ten posts with the highest severe toxicity scores.....    | 93  |
| Gambar 8 Ten posts with the highest identity attack scores .....   | 96  |
| Gambar 9 Ten posts with the highest insult scores .....            | 100 |
| Gambar 10 Ten posts with the highest profanity scores.....         | 103 |
| Gambar 11 Ten posts with the highest threat scores .....           | 106 |
| Gambar 12 Avarage Toxicity Scores Per Month.....                   | 109 |
| Gambar 13 Toxicity Score Distribution .....                        | 113 |
| Gambar 14 Wordcloud Dialog Akal Sehat UAS – RG .....               | 116 |
| Gambar 15 Emojicloud Dialog Akal Sehat UAS – RG .....              | 118 |
| Gambar 16 Posts Per Day .....                                      | 120 |
| Gambar 17 Top Ten Posters .....                                    | 123 |

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Perkembangan teknologi informasi dan komunikasi dalam dua dekade terakhir telah merevolusi cara manusia berkomunikasi dan berpartisipasi dalam ruang publik digital. Media sosial, khususnya YouTube, telah menjelma menjadi platform distribusi informasi yang tidak hanya bersifat visual dan auditif, tetapi juga interaktif, memungkinkan para pengguna untuk membangun narasi, memengaruhi opini publik, dan menyampaikan aspirasi mereka secara langsung<sup>1</sup>. Dalam konteks komunikasi politik dan keagamaan, platform ini bahkan memainkan peran penting dalam membentuk persepsi publik melalui komentar-komentar warganet yang seringkali mencerminkan respons emosional yang spontan maupun reflektif<sup>2</sup>.

Namun, di balik peran transformatif tersebut, media sosial juga memiliki sisi gelap berupa munculnya ujaran kebencian, disinformasi, dan komentar toksik yang semakin marak seiring meningkatnya penggunaan algoritma dan anonimitas digital (Jane, 2015; Lapidot-Lefler & Barak, 2015; Oboler, 2014; Nahon & Hemsley, 2013; Williams, 2021). Kemudahan akses dan tidak adanya batasan geografis turut mendukung tersebarnya perilaku tidak etis secara masif, menciptakan iklim komunikasi yang berpotensi

---

<sup>1</sup> J. Burgess, J., & Green, *YouTube: Online Video and Participatory Culture*. (Polity Press., 2009).

<sup>2</sup> Z Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. (Yale University Press., 2017).

memecah belah dan menyulut konflik sosial<sup>3</sup>. Dalam hal ini, YouTube tidak hanya menjadi media ekspresi, tetapi juga arena pertarungan narasi yang diwarnai oleh polarisasi politik, ideologi agama, dan perbedaan nilai-nilai sosial budaya<sup>4</sup>.

Fenomena komentar toksik di platform YouTube menjadi perhatian serius, terutama ketika menyangkut tokoh publik yang memiliki pengaruh besar di masyarakat. Salah satu contoh yang menonjol adalah dialog antara Rocky Gerung dan Ustadz Abdul Somad, dua figur publik dengan latar belakang dan pandangan yang berbeda, yang menarik perhatian luas dan memicu berbagai reaksi di ruang komentar YouTube<sup>5</sup>. Komentar-komentar yang muncul tidak hanya mencerminkan perbedaan pendapat, tetapi juga sering kali mengandung unsur toksik yang dapat memengaruhi kualitas diskusi dan kesehatan mental pengguna lainnya<sup>6</sup>.

Penelitian menunjukkan bahwa komentar toksik di YouTube sering kali dipicu oleh konten yang kontroversial atau melibatkan tokoh publik yang polarizing. Komentar-komentar tersebut dapat menciptakan lingkungan yang tidak sehat bagi diskusi publik dan berdampak negatif pada kesehatan mental

---

<sup>3</sup> Alice E. Marwick and Danah Boyd, "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience," *New Media and Society* 13, no. 1 (2011): 114–33, <https://doi.org/10.1177/1461444810365313>.

<sup>4</sup> Howard Kurtz, "Final Edition: Rocky Mountain News to Shut Down Today," *The Washington Post*, 2009, <https://www.washingtonpost.com/archive/business/2009/02/27/final-edition-rocky-mountain-news-to-shut-down-today/25fa709b-f799-425f-98c9-e2656fc713d2/>.

<sup>5</sup> YouTube., "Eksklusif Dialog Akal Sehat Rocky Gerung Ustadz Abdul Somad YouTube.," 2023, <https://www.youtube.com/watch?v=B2WtOAwI158>.

<sup>6</sup> Verywell Mind. (n.d.), "Mental Health Effects of Reading Negative Comments Online.," 2025.

pengguna<sup>7</sup>. Selain itu, algoritma YouTube yang mempromosikan konten berdasarkan tingkat keterlibatan pengguna dapat secara tidak langsung mendorong penyebaran komentar toksik, karena komentar yang memicu emosi cenderung mendapatkan lebih banyak perhatian dan interaksi<sup>8</sup>.

Dampak dari komentar toksik tidak hanya dirasakan oleh individu yang menjadi target, tetapi juga oleh komunitas secara keseluruhan. Lingkungan digital yang dipenuhi oleh komentar negatif dapat menghambat partisipasi aktif pengguna lain, menciptakan efek "silent majority" di mana individu enggan untuk berkontribusi dalam diskusi karena takut menjadi sasaran komentar toksik<sup>9</sup>. Hal ini menunjukkan perlunya pendekatan yang lebih proaktif dalam moderasi konten dan edukasi pengguna untuk menciptakan ruang diskusi yang sehat dan konstruktif di platform digital.

Berbagai penelitian telah dilakukan untuk memahami dan mengklasifikasikan komentar toksik di media sosial. Hanscom mengembangkan pendekatan untuk mendeteksi komentar toksik dengan menggunakan model pembelajaran mesin, yang mencakup analisis siklus ekstremisme internet dan identifikasi kesenjangan dalam literatur yang ada.<sup>10</sup> Mereka menekankan pentingnya memahami konteks sosial dan budaya dalam analisis komentar toksik, serta perlunya pendekatan yang lebih inklusif dalam

---

<sup>7</sup> Griffith University., "The Impact of YouTube on Mental Health.," 2023, <https://news.griffith.edu.au/2023/10/10/the-impact-of-youtube-on-mental-health/>.

<sup>8</sup> A. Bertaglia, T., Goanta, C., & Iamnitchi, "The Monetisation of Toxicity: Analysing YouTube Content Creators and Controversy-Driven Engagement.," no. arXiv preprint arXiv:2408.00534. (2024), <https://arxiv.org/abs/2408.00534>.

<sup>9</sup> G. Juncosa, G., Yasseri, T., Koltai, J., & Iniguez, "Toxic Behavior Silences Online Political Conversations.," *ArXiv Preprint ArXiv:2412.05741.*, 2024, <https://arxiv.org/abs/2412.05741>.

<sup>10</sup> S. Hanscom, R., Lehman, T. S., Lv, Q., & Mishra, "The Toxicity Phenomenon Across Social Media.," *ArXiv Preprint ArXiv:2410.21589.*, 2023, <https://arxiv.org/abs/2410.21589>.

penelitian ini. Namun, sebagian besar penelitian ini berfokus pada konteks global atau negara-negara Barat, sehingga kurang relevan untuk memahami dinamika komentar toksik dalam konteks Indonesia yang memiliki keunikan budaya dan sosial tersendiri <sup>11</sup>.

Di Indonesia, beberapa studi telah dilakukan untuk mengembangkan model deteksi komentar toksik yang sesuai dengan konteks lokal. Salah satunya adalah penelitian oleh Sagama dan Alamsyah pada tahun 2023 yang mengevaluasi efektivitas model bahasa BERT dan RoBERTa dalam mengidentifikasi konten toksik dalam bahasa Indonesia. Mereka menemukan bahwa model IndoBERTweet dan IndoBERT menunjukkan kinerja yang baik dalam tugas klasifikasi multi-label untuk mendeteksi berbagai jenis toksisitas dalam interaksi digital berbahasa Indonesia.<sup>12</sup> Selain itu, penelitian oleh Rumagit dan Girsang pada tahun 2018 juga menunjukkan bahwa pendekatan klasifikasi multi-label dapat digunakan untuk mengidentifikasi komentar toksik dalam bahasa Indonesia, dengan mempertimbangkan berbagai kategori seperti ujaran kebencian, pencemaran nama baik, dan radikalisme<sup>13</sup>. Studi lain oleh Ahmadizzan mengusulkan penggunaan pendekatan pembelajaran mendalam hibrida untuk mengidentifikasi komentar kasar dari data media sosial berbahasa Indonesia, dengan menggabungkan jaringan saraf berulang

---

<sup>11</sup> Hanscom, R., Lehman, T. S., Lv, Q., & Mishra.

<sup>12</sup> A. Sagama, Y., & Alamsyah, "Multi-Label Classification of Indonesian Online Toxicity Using BERT and RoBERTa.," *ResearchGate*, 2023, [https://www.researchgate.net/publication/373036835\\_Multi-Label\\_Classification\\_of\\_Indonesian\\_Online\\_Toxicity\\_using\\_BERT\\_and\\_RoBERTa](https://www.researchgate.net/publication/373036835_Multi-Label_Classification_of_Indonesian_Online_Toxicity_using_BERT_and_RoBERTa).

<sup>13</sup> Reinert Yosua Rumagit, "Multilabel Classification for Toxic Comments in Indonesian," *Engineering, Mathematics and Computer Science (EMACS) Journal* 2, no. 1 (2020): 29–34, <https://doi.org/10.21512/emacsjournal.v2i1.6256>.

(RNN) dan memori jangka pendek panjang (LSTM) untuk meningkatkan akurasi deteksi.<sup>14</sup>

Meskipun telah ada upaya untuk mengembangkan model deteksi komentar toksik yang sesuai dengan konteks Indonesia, tantangan tetap ada. Salah satunya adalah kurangnya data berlabel yang representatif untuk bahasa Indonesia, yang dapat menghambat pelatihan model pembelajaran mesin yang efektif<sup>15</sup> Selain itu, kompleksitas dan nuansa bahasa Indonesia, termasuk penggunaan bahasa gaul dan kode campuran, dapat menyulitkan deteksi toksisitas secara otomatis<sup>16</sup>, Oleh karena itu, diperlukan pendekatan yang lebih holistik dan kontekstual dalam penelitian dan pengembangan sistem deteksi komentar toksik di media sosial Indonesia.

Meskipun telah banyak penelitian tentang komentar toksik di media sosial, masih terdapat kekurangan dalam studi yang secara khusus meneliti fenomena ini dalam konteks Indonesia. Sebagian besar penelitian yang ada berfokus pada konteks global atau negara-negara Barat, sehingga kurang relevan untuk memahami dinamika komentar toksik dalam konteks Indonesia yang memiliki keunikan budaya dan sosial tersendiri<sup>17</sup>.

Penelitian lokal yang membahas komentar toksik di media sosial Indonesia masih terbatas. Beberapa studi telah dilakukan, seperti analisis

---

<sup>14</sup> “Ahmadizzan. (2022). Abusive Comment Identification on Indonesian Social Media Data Using Hybrid Deep Learning. Academia.Edu. [https://www.academia.edu/112446212/Abusive\\_comment\\_identification\\_on\\_indonesian\\_social\\_media\\_data\\_using\\_hybrid\\_deep\\_learning](https://www.academia.edu/112446212/Abusive_comment_identification_on_indonesian_social_media_data_using_hybrid_deep_learning),” n.d.

<sup>15</sup> Rumagit, “Multilabel Classification for Toxic Comments in Indonesian.”

<sup>16</sup> Sagama, Y., & Alamsyah, “Multi-Label Classification of Indonesian Online Toxicity Using BERT and RoBERTa.”

<sup>17</sup> Hanscom, R., Lehman, T. S., Lv, Q., & Mishra, “The Toxicity Phenomenon Across Social Media.”

komentar toksik terhadap video informasi COVID-19 pada akun YouTube Kementerian Kesehatan menggunakan metode Naïve Bayes Classifier <sup>18</sup> Namun, penelitian yang secara khusus mengkaji komentar toksik yang muncul dalam diskusi antara tokoh publik Indonesia dengan latar belakang yang berbeda, seperti Rocky Gerung dan Ustadz Abdul Somad, masih sangat minim.

Kesenjangan ini penting untuk diisi guna memahami bagaimana perbedaan pandangan dan latar belakang dapat mempengaruhi dinamika diskusi dan munculnya komentar toksik di platform seperti YouTube. Studi oleh Mardatrisna dan Banowo menunjukkan bahwa analisis isi komentar warganet pada akun YouTube Channel Metro TV mengenai konten "HOTROOM - Naturalisasi Untuk Prestasi?" dapat memberikan wawasan tentang bagaimana respon publik terhadap topik tertentu. Namun, penelitian serupa yang fokus pada interaksi antara tokoh publik dengan latar belakang berbeda di Indonesia masih sangat diperlukan. <sup>19</sup>

Penelitian ini bertujuan untuk mengisi kesenjangan dalam studi mengenai komentar toksik di media sosial Indonesia dengan menganalisis komentar-komentar yang muncul dalam video dialog antara Rocky Gerung dan Ustadz Abdul Somad di YouTube dengan judul "DIALOG AKAL SEHAT PROF ROCKY GERUNG & USTADZ ABDUL SOMAD".

---

<sup>18</sup> Nuzzulul Ulum, "Pola Komunikasi Kyai Dalam Membentuk Karakter Santri Di Pesantren Ash-Shiddiqi Putri Talangsari Jember," *Fajar: Jurnal Pendidikan Islam* 1, no. 1 (2021): 48–62.

<sup>19</sup> E. Mardatrisna, S. R., & Banowo, "Analisis Isi Komentar Model Analisis Harold Laswell Pada Akun YouTube Channel Metro TV Mengenai Konten 'HOTROOM – Naturalisasi Untuk Prestasi?,'" *Triwikrama: Jurnal Ilmu Sosial* 5 (2024), <https://ejournal.warunayama.org/index.php/triwikrama/article/download/6133/5659/18700>.

Title ‘Prof’ dalam judul konten Dialog akal sehat Prof Rocky Gerung & Ustadz Abdul Somad penting dikritisi mengingat title prof yang disematkan pada Rocky Gerung itu kontroversi di ruang publik. Penting untuk diketahui bahwa peneliti tidak akan membahas kontroversi tersebut selanjutnya. Penelitian ini menepatkan penggunaan title ‘Prof’ sebagai bagian dari konstruksi framing judul konten ‘Dialog akal sehat Rocky Gerung & Ustadz Abdul Somad toxicity comment analysis pada platform youtube. Dengan menggunakan pendekatan analisis teks dan pembelajaran mesin, penelitian ini akan mengidentifikasi pola-pola komentar toksik serta memahami faktor-faktor yang mempengaruhi munculnya komentar tersebut. Pendekatan ini sejalan dengan penelitian sebelumnya yang memanfaatkan algoritma pembelajaran mesin, seperti Naïve Bayes dan Support Vector Machine (SVM), untuk mendeteksi komentar toksik dalam bahasa Indonesia<sup>20</sup>. Selain itu, analisis emosi terhadap komentar You.Tube juga telah dilakukan untuk memahami sentimen pengguna terhadap konten tertentu<sup>21</sup>.

Berangkat dari konteks tersebut, penelitian ini memiliki keterkaitan yang signifikan dengan kajian Komunikasi Penyiaran Islam (KPI), khususnya dalam hal bagaimana pesan-pesan keagamaan dan nilai-nilai Islam diproduksi, disampaikan, dan diterima di ruang publik digital. Dialog antara Rocky Gerung dan Ustadz Abdul Somad tidak hanya menjadi objek diskursus intelektual, tetapi juga representasi dari kontestasi wacana keislaman dalam era digital. Analisis terhadap komentar toksik yang muncul dalam diskusi ini

---

<sup>20</sup> Darmajaya, “Deteksi Kalimat Toxic Dalam Postingan Media Sosial Menggunakan Algoritma Decision Tree.,” 2022, <https://repo.darmajaya.ac.id/15460/>.

<sup>21</sup> Djoko Purwanto, *Komunikasi Bisnis* (Penerbit Erlangga, 2003).

mencerminkan dinamika penerimaan dakwah Islam dan respons publik terhadap tokoh keagamaan di tengah pluralitas ideologi dan gaya komunikasi yang beragam. Dengan demikian, studi ini turut memperkaya ranah KPI dalam memahami bagaimana media digital menjadi medan penyiaran nilai-nilai Islam, tantangan etika komunikasi, serta strategi dakwah yang relevan dalam menghadapi era algoritmik dan polarisasi digital.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan baru tentang dinamika diskusi publik di Indonesia dan kontribusi terhadap upaya menciptakan ruang diskusi yang sehat dan konstruktif di media sosial. Dengan memahami pola komentar toksik dan faktor-faktor yang mempengaruhinya, diharapkan dapat dikembangkan strategi mitigasi yang efektif untuk mengurangi penyebaran komentar negatif dan meningkatkan kualitas diskusi daring. Hal ini penting mengingat peran media sosial sebagai ruang publik digital yang mempengaruhi opini dan perilaku masyarakat <sup>22</sup>.

## **B. Rumusan Masalah**

Berdasarkan latar belakang dan tujuan penelitian, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana profil Ustad Abdul Somad dan Rocky Gerung di platform youtube?

---

<sup>22</sup> Hanscom, R., Lehman, T. S., Lv, Q., & Mishra, "The Toxicity Phenomenon Across Social Media."

2. Bagaimana bentuk dan distribusi komentar toksik yang muncul dalam video dialog antara Rocky Gerung dan Ustadz Abdul Somad di platform YouTube?
3. Apa saja kategori toksisitas yang paling dominan ditemukan dalam komentar warganet, seperti insult, hate speech, identity attack, atau threat?

### **C. Tujuan Penelitian**

Berdasarkan latar belakang tersebut dan rumusan masalah yang telah ditetapkan, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk menganalisis profil Ustad Abdul Somad dan Rocky Gerung di platform youtube
2. Untuk menganalisis bentuk dan distribusi komentar toksik yang muncul dalam video dialog antara Rocky Gerung dan Ustadz Abdul Somad di platform YouTube.
3. Untuk mengidentifikasi kategori toksisitas yang paling dominan ditemukan dalam komentar warganet, seperti insult, hate speech, identity attack, atau threat.

### **D. Manfaat dan Kegunaan**

Penelitian ini diharapkan dapat memberikan kontribusi yang berarti, baik dalam ranah teoretis maupun praktis, terkait dengan kajian komunikasi digital, media sosial, dan fenomena komentar toksik di ruang publik daring. Adapun manfaat dan kegunaan dari penelitian ini dijabarkan sebagai berikut:

## **1. Manfaat Teoritis**

Penelitian ini memberikan kontribusi pada pengembangan kajian komunikasi digital, khususnya dalam memahami dinamika komentar toksik di media sosial YouTube yang melibatkan tokoh publik dengan latar belakang ideologis yang berbeda. Dengan fokus pada konteks Indonesia, penelitian ini memperkaya literatur akademik yang selama ini masih didominasi oleh studi dari negara-negara Barat. Selain itu, penelitian ini turut memberikan kontribusi pada pendekatan interdisipliner antara ilmu komunikasi, sosiolinguistik, dan ilmu komputer dalam memahami pola interaksi digital berbasis teks.

## **2. Manfaat Praktis**

Secara praktis, hasil dari penelitian ini dapat dimanfaatkan oleh berbagai pihak, antara lain:

1. Platform media sosial seperti YouTube, dalam mengembangkan sistem moderasi komentar yang lebih adaptif terhadap dinamika sosial dan budaya lokal.
2. Pemerintah dan pembuat kebijakan, khususnya dalam merancang kebijakan literasi digital dan regulasi konten yang bertujuan menciptakan ruang diskusi yang sehat, terbuka, dan bebas dari ujaran kebencian.
3. Akademisi dan peneliti, sebagai referensi dalam pengembangan riset lanjutan mengenai komunikasi digital, ujaran kebencian, dan perilaku warganet di media sosial Indonesia.

4. Masyarakat umum, sebagai edukasi dan refleksi atas perilaku komunikasi di media sosial, sehingga mendorong partisipasi digital yang lebih bijak, toleran, dan bertanggung jawab.

## **E. Kajian Pustaka**

### **1. Media Sosial dan YouTube sebagai Ruang Publik Digital**

#### **a. Definisi Media Sosial dan Karakteristiknya**

Media sosial telah menjadi fenomena global yang secara fundamental mengubah cara individu berkomunikasi, berinteraksi, dan berbagi informasi dalam kehidupan sehari-hari. Keberadaan media sosial tidak hanya sekadar memperluas ruang interaksi sosial, tetapi juga membentuk ulang struktur komunikasi masyarakat di era digital. Menurut Kaplan dan Haenlein, media sosial didefinisikan sebagai "sekelompok aplikasi berbasis internet yang dibangun di atas dasar ideologis dan teknologi Web 2.0, yang memungkinkan penciptaan dan pertukaran konten buatan pengguna". Definisi ini menyoroti dua aspek utama yang menjadi dasar keberadaan media sosial: pertama, pemanfaatan teknologi Web 2.0 yang memungkinkan platform menjadi lebih interaktif dan adaptif terhadap pengguna; dan kedua, partisipasi aktif pengguna sebagai produsen dan distributor konten yang terdesentralisasi<sup>23</sup>.

---

<sup>23</sup> M. Kaplan, A. M., & Haenlein, "Users of the World, Unite! The Challenges and Opportunities of Social Media.," *Business Horizons*, no. 53(1), 59–68. (2010), <https://doi.org/10.1016/j.bushor.2009.09.003>.

Aspek Web 2.0 mencerminkan pergeseran dari sistem komunikasi satu arah yang bersifat top-down—seperti yang ditemukan dalam media konvensional—menjadi sistem komunikasi dua arah yang memungkinkan kolaborasi dan dialog antar pengguna<sup>24</sup>. Web 2.0 memberikan kerangka teknis dan ideologis yang memungkinkan lahirnya berbagai platform seperti Facebook, Twitter, Instagram, TikTok, dan YouTube, yang kesemuanya bergantung pada partisipasi pengguna untuk tetap hidup dan berkembang Kaplan & Haenlein, Tidak seperti media tradisional, di mana khalayak hanya berperan sebagai penerima informasi, media sosial menjadikan setiap individu sebagai agen komunikasi aktif yang dapat memproduksi, mengomentari, menyebarluaskan, bahkan memodifikasi informasi yang ada<sup>25</sup>.

Lebih jauh, keterlibatan aktif pengguna ini membentuk budaya partisipatif yang menjadi salah satu ciri utama komunikasi di era digital. Jenkins<sup>26</sup> menyebut fenomena ini sebagai "participatory culture", di mana batas antara produsen dan konsumen konten menjadi kabur, dan tercipta hubungan interaktif antara pengguna dan komunitas digital yang lebih luas. Hal ini sejalan dengan pandangan Rheingold tentang "smart mobs", yaitu kemampuan kolektif masyarakat untuk

---

<sup>24</sup> B. Cormode, G., & Krishnamurthy, "Key Differences between Web 1.0 and Web 2.0.," *First Monday* 13(6). (2008), <https://doi.org/10.5210/fm.v13i6.2125>.

<sup>25</sup> R. Nasrullah, *Media Sosial: Perspektif Komunikasi, Budaya, Dan Sosioteknologi*. (Simbiosis Rekatama Media., 2015).

<sup>26</sup> H. Jenkins, *Convergence Culture: Where Old and New Media Collide*. (New York University Press., 2006).

berkolaborasi secara spontan melalui platform digital dalam menyampaikan aspirasi atau merespons isu-isu sosial.<sup>27</sup>

Transformasi komunikasi melalui media sosial ini juga berdampak pada struktur kekuasaan informasi. Jika sebelumnya informasi dikendalikan oleh lembaga media besar, kini kekuasaan tersebut mulai terdesentralisasi ke tangan individu-individu yang memiliki akses ke internet dan keterampilan digital dasar. Dalam konteks ini, media sosial bukan hanya alat komunikasi, melainkan juga arena kontestasi narasi, pembentukan opini publik, dan pengaruh politik yang signifikan, seperti terlihat dalam berbagai gerakan sosial digital seperti #MeToo atau #ReformasiDikorupsi di Indonesia<sup>28</sup>

Selanjutnya, definisi media sosial sebagaimana dijelaskan oleh Kaplan dan Haenlein tidak hanya menjelaskan aspek teknis, tetapi juga membawa implikasi teoritis yang luas dalam studi komunikasi, budaya, dan masyarakat digital. Pemahaman ini menjadi penting dalam menganalisis berbagai dinamika sosial, termasuk fenomena komentar toksik yang muncul sebagai bagian dari konsekuensi dari keterbukaan dan interaktivitas media sosial sebagai ruang publik virtual.<sup>29</sup>

Karakteristik media sosial dapat diklasifikasikan ke dalam beberapa dimensi utama yang membedakannya secara fundamental

---

<sup>27</sup> H Rheingold, *The Virtual Community: Homesteading on the Electronic Frontier*. (MIT Press., 2000).

<sup>28</sup> Z. Tufekci, "Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency.," *Colorado Technology Law Journal* 13, 203–21 (2015).

<sup>29</sup> Kaplan, A. M., & Haenlein, "Users of the World, Unite! The Challenges and Opportunities of Social Media."

dari media tradisional. Tidak seperti media konvensional yang bersifat satu arah dan terpusat, media sosial bersifat dua arah, interaktif, serta menempatkan pengguna sebagai produsen sekaligus konsumen informasi. Kompas.com menyebutkan lima karakteristik esensial dari media sosial, yaitu: partisipasi, interaktivitas, keterhubungan (connectivity), komunitas (community), dan keterbukaan informasi (accessibility). Karakteristik-karakteristik ini menjadi penopang utama dalam membentuk dinamika komunikasi digital kontemporer <sup>30</sup>.

*Pertama*, partisipasi merujuk pada keterlibatan aktif pengguna dalam menciptakan, memodifikasi, serta menyebarkan informasi dan konten secara luas. Pengguna tidak lagi menjadi entitas pasif seperti dalam model komunikasi massa klasik, melainkan menjadi aktor utama yang memiliki otonomi dalam mengelola dan menyebarkan pesan melalui jejaring digital (Kaplan & Haenlein, 2010; Shirky, 2008; Jenkins, 2006). Hal ini memperkuat munculnya user-generated content (UGC) yang telah mengubah peta ekosistem komunikasi global. Di Indonesia, fenomena ini terlihat dari maraknya konten buatan masyarakat seperti vlog, podcast, dan konten edukatif di platform seperti YouTube, TikTok, maupun Instagram, yang bahkan bisa menyaingi produksi media arus utama <sup>31</sup>

---

<sup>30</sup> Kompas TV., "Rocky Gerung: Pemerintah Produsen Hoaks Terbaik. [Video]. YouTube. Diakses 6 Agustus 2025, Dari," 2017, <https://www.youtube.com>.

<sup>31</sup> Kominfo., "Pedoman Pengendalian Konten Negatif Di Internet.," 2022, <https://kominfo.go.id/>.

*Kedua*, interaktivitas adalah kemampuan media sosial untuk memungkinkan komunikasi dua arah secara simultan antara pengguna, baik dalam bentuk teks, suara, gambar, maupun video. Interaktivitas menciptakan dialog yang bersifat spontan dan responsif, sehingga meningkatkan tingkat keterlibatan emosional dan sosial pengguna<sup>32</sup>. Interaksi ini tidak hanya terbatas antara individu, tetapi juga melibatkan institusi, merek, maupun tokoh publik, yang menjadikan media sosial sebagai ruang strategis untuk membangun relasi dan citra<sup>33</sup>

*Ketiga*, keterhubungan (*connectivity*) merujuk pada kapabilitas media sosial dalam menghubungkan individu lintas batas geografis, budaya, dan kelas sosial. Melalui jaringan digital yang tidak mengenal batas teritorial, media sosial menciptakan ekosistem komunikasi global yang mendukung pembentukan jejaring sosial, kolaborasi lintas negara, dan penyebaran ide secara instan<sup>34</sup>. Dalam konteks ini, media sosial tidak hanya menjadi alat komunikasi, tetapi juga infrastruktur sosial yang memperkuat hubungan interpersonal dan kolektif.

*Keempat*, komunitas (*community*) terbentuk karena media sosial memberikan ruang bagi individu dengan minat, identitas, atau kepentingan serupa untuk saling berinteraksi dan membentuk

---

<sup>32</sup> A. M. Sundar, S. S., & Limperos, "Uses and Grats 2.0: New Gratifications for New Media.," *Journal of Broadcasting & Electronic Media* 57(4), 504 (2023), <https://doi.org/10.1080/08838151.2013.845827>.

<sup>33</sup> D. J. Mangold, W.G., & Faulds, "Social Media: The New Hybrid Element of the Promotion Mix.," *Business Horizons* 52(4), 357 (2009), <https://doi.org/10.1016/j.bushor.2009.03.002>.

<sup>34</sup> M. Castells, *The Rise of the Network Society (2nd Ed.)*. (Wiley-Blackwell., 2011).

kelompok. Komunitas digital ini sering kali melampaui batasan fisik dan temporal, memungkinkan terbentuknya solidaritas virtual dan kohesi sosial yang baru<sup>35</sup>. Misalnya, komunitas *#SaveKPK* atau *#BlackLivesMatter* menunjukkan bagaimana media sosial dapat menjadi alat mobilisasi sosial berbasis komunitas yang memiliki pengaruh nyata di dunia offline.

*Kelima*, keterbukaan informasi (*accessibility*) menunjukkan bahwa media sosial memungkinkan distribusi informasi secara terbuka dan cepat kepada siapa pun yang memiliki akses internet. Prinsip keterbukaan ini menciptakan potensi demokratisasi informasi, di mana setiap orang memiliki peluang yang sama untuk mengakses, memproduksi, dan menyebarkan data<sup>36</sup>. Namun demikian, keterbukaan ini juga menimbulkan tantangan seperti penyebaran disinformasi dan misinformasi, yang dapat berdampak negatif terhadap kualitas ruang publik digital<sup>37</sup>.

Nasrullah dalam bukunya *Media Sosial: Perspektif Komunikasi, Budaya, dan Sosioteknologi*, menambahkan bahwa media sosial memiliki karakteristik penting sebagai ruang simulasi sosial, di mana pengguna tidak hanya berinteraksi, tetapi juga membentuk, merekonstruksi, dan mempresentasikan identitas diri mereka secara

---

<sup>35</sup> Rheingold, *The Virtual Community: Homesteading on the Electronic Frontier*.

<sup>36</sup> C. Shirky, *Here Comes Everybody: The Power of Organizing Without Organizations*. (Penguin Press., 2008).

<sup>37</sup> H. Wardle, C., & Derakhshan, "Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking.," *Council of Europe*, 2017.

virtual.<sup>38</sup> Ruang digital ini menyediakan medium bagi individu untuk menampilkan citra yang dikehendaki, yang tidak selalu mencerminkan identitas asli mereka secara utuh. Representasi diri di media sosial seringkali merupakan hasil dari proses kurasi, seleksi, dan estetika digital yang ditentukan oleh algoritma, ekspektasi sosial, serta keinginan untuk mendapat validasi dalam bentuk like, komentar, dan follower<sup>39</sup>.

Konsep ini beririsan dengan gagasan Jean Baudrillard tentang *simulacra* dan *simulation*, di mana realitas di era media telah digantikan oleh representasi yang direkayasa sehingga realitas sosial menjadi semakin kabur. Baudrillard (1994) menyebut fenomena ini sebagai hyperreality — kondisi di mana representasi menjadi lebih nyata daripada realitas itu sendiri, dan simbol-simbol dalam media membentuk persepsi tentang dunia secara dominan. Dalam konteks media sosial, identitas digital seseorang tidak lagi merupakan cerminan murni dari identitas nyata, melainkan versi yang telah dikonstruksi ulang berdasarkan preferensi personal, budaya populer, dan tuntutan sosial digital<sup>40</sup>.

Media sosial memungkinkan pengguna untuk mengedit, menyeleksi, dan mengatur ulang narasi tentang diri mereka melalui unggahan foto, status, video, hingga bio profil, sehingga menjadi ruang yang penuh dengan simbolisasi identitas. Goffman (1959) dalam teori

---

<sup>38</sup> M. Nasrullah, *Ekonomi Politik Media Dan Demokrasi*. (Prenada Media Group., 2012).

<sup>39</sup> Nasrullah, *Media Sosial: Perspektif Komunikasi, Budaya, Dan Sosioteknologi*.

<sup>40</sup> J. Baudrillard, *Simulacra and Simulation*. (University of Michigan Press., 1995).

dramaturginya menjelaskan bahwa individu dalam kehidupan sosial berperilaku layaknya aktor yang memainkan peran di atas panggung, menampilkan versi terbaik dari diri mereka untuk mendapat pengakuan dari audiens. Prinsip ini semakin diperkuat dalam platform digital, di mana individu memiliki kontrol penuh terhadap bagaimana mereka ingin tampil di hadapan publik. Fenomena personal branding, digital persona, dan aesthetic identity adalah contoh konkret dari bagaimana pengguna media sosial merancang citra diri mereka <sup>41</sup>.

Hal ini berdampak pada cara kita memahami identitas dan otentisitas. Identitas di media sosial menjadi cair, kontekstual, dan sangat bergantung pada narasi yang dibangun pengguna dalam ruang interaksi online. Di sisi lain, hal ini juga menimbulkan ketegangan antara realitas dan ekspektasi, yang dapat berdampak pada psikologi pengguna, seperti perasaan tidak cukup, FOMO (fear of missing out), dan tekanan untuk selalu tampil sempurna<sup>42</sup>.

Media sosial bukan hanya ruang komunikasi, tetapi juga ruang performatif dan simbolik, tempat di mana identitas dinegosiasikan secara terus-menerus. Ini menjadi penting untuk dipahami, terutama dalam studi mengenai dinamika interaksi sosial dan konflik opini seperti yang terjadi di kolom komentar YouTube—di mana pengguna

---

<sup>41</sup> B. Hogan, "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. *Bulletin of Science, Technology & Society*," no. 30(6), 377–386. (2010), <https://doi.org/10.1177/0270467610385893>.

<sup>42</sup> S. Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other*, 2011, Basic Books.

tidak hanya menyampaikan argumen, tetapi juga mempertaruhkan citra, afiliasi, bahkan ideologinya secara simbolik dan terbuka.

Karakteristik lain dari media sosial yang membedakannya dari media tradisional adalah kemampuannya untuk menyebarkan informasi secara viral dalam hitungan detik, tanpa harus melalui proses kurasi atau verifikasi seperti yang biasa diterapkan dalam sistem penyiaran konvensional. Kaplan dan Haenlein (2010) menegaskan bahwa media sosial memungkinkan pesan untuk disebarluaskan secara cepat dan luas melalui jaringan sosial yang terhubung secara global, menciptakan lingkungan komunikasi yang sangat dinamis, desentralistik, dan seringkali tidak terprediksi. Penyebaran ini dipengaruhi oleh struktur jejaring (*network structure*), kekuatan keterhubungan antar pengguna (*tie strength*), serta algoritma yang mendukung visibilitas konten berdasarkan keterlibatan pengguna (*likes, shares, comments*)<sup>43</sup>.

Fenomena *virality* dalam media sosial tidak hanya berlaku pada konten hiburan atau personal, tetapi juga berdampak pada isu-isu sosial, politik, hingga krisis kemanusiaan. Dalam konteks kampanye sosial, kecepatan distribusi informasi ini dapat menjadi alat yang sangat efektif untuk menyebarluaskan kesadaran publik, mobilisasi sosial, hingga penggalangan dukungan, sebagaimana ditunjukkan dalam kampanye global seperti #MeToo, #BlackLivesMatter, maupun

---

<sup>43</sup> Kaplan, A. M., & Haenlein, "Users of the World, Unite! The Challenges and Opportunities of Social Media."

gerakan lokal seperti #GejayanMemanggil di Indonesia<sup>44</sup>. Kekuatan viral juga memungkinkan suara kelompok marginal untuk memperoleh perhatian arus utama yang sebelumnya sulit diakses.

Namun demikian, kecepatan yang sama juga menjadi pedang bermata dua, di mana informasi yang tidak benar atau bersifat manipulatif dapat menyebar dengan mudah dan berdampak luas. Studi oleh Vosoughi, Roy, dan Aral (2018) menunjukkan bahwa informasi palsu atau hoaks memiliki kecenderungan untuk menyebar lebih cepat daripada informasi yang valid, karena sifatnya yang mengejutkan, emosional, dan menarik perhatian. Penyebaran hoaks dan misinformasi ini diperparah oleh fenomena echo chamber dan confirmation bias, di mana pengguna media sosial cenderung berinteraksi dalam lingkaran sosial yang sependapat, sehingga memperkuat opini yang tidak selalu didukung oleh fakta<sup>45</sup>.

Di Indonesia, maraknya penyebaran hoaks melalui media sosial bahkan telah menjadi perhatian pemerintah dan lembaga pengawas informasi. Kementerian Komunikasi dan Informatika (Kominfo) secara aktif melakukan klarifikasi terhadap konten hoaks, terutama selama masa pandemi COVID-19 dan menjelang Pemilu. Namun, kapasitas penyaringan informasi yang begitu cepat menyebar sering kali tertinggal dibandingkan dengan kecepatan viralitas itu sendiri

---

<sup>44</sup> Tufekci, "Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency."

<sup>45</sup> C. R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media*. (Princeton University Press., 2018).

Oleh karena itu, dibutuhkan sinergi antara edukasi literasi digital, peningkatan kesadaran pengguna, serta penguatan algoritma platform dalam menangani penyebaran konten bermasalah.<sup>46</sup>

Karakteristik viralitas dalam media sosial menyimpan potensi transformatif dan risiko disrupsi yang besar. Dalam konteks penelitian ini, pemahaman terhadap karakteristik ini sangat penting untuk menelaah bagaimana komentar-komentar toksik, baik yang bersifat personal, ideologis, maupun emosional dapat menyebar dengan cepat dan membentuk atmosfer diskusi publik yang polaristik di platform seperti YouTube.

b. YouTube sebagai Platform Diskusi Public di Indonesia

YouTube telah berkembang pesat dari sekadar platform berbagi video menjadi salah satu ruang publik digital paling berpengaruh di dunia, termasuk di Indonesia. Keberadaannya telah mengubah lanskap komunikasi publik dengan menjembatani produksi, distribusi, dan konsumsi konten oleh masyarakat luas tanpa harus melalui institusi media arus utama. Menurut Burgess dan Green (2018), YouTube adalah platform partisipatif yang memungkinkan pengguna tidak hanya sebagai konsumen pasif, tetapi juga sebagai produsen aktif dalam ekosistem informasi dan budaya digital. Mereka menekankan bahwa YouTube menyediakan ruang bagi ekspresi diri, aktivisme,

---

<sup>46</sup> Kominfo., “Pedoman Pengendalian Konten Negatif Di Internet.”

kritik sosial, dan bahkan diplomasi digital yang dikelola oleh individu maupun komunitas <sup>47</sup>.

Di Indonesia, YouTube menempati posisi strategis dalam kehidupan digital masyarakat. Berdasarkan laporan DataReportal<sup>48</sup>. YouTube merupakan platform media sosial paling banyak digunakan oleh masyarakat Indonesia, dengan penetrasi lebih dari 94% pengguna internet. Dominasi ini menjadikan YouTube tidak hanya sebagai sarana hiburan, tetapi juga sebagai medium penting dalam menyampaikan opini publik, diskusi politik, dan perdebatan isu-isu sosial yang melibatkan berbagai elemen masyarakat, termasuk jurnalis warga, akademisi, aktivis, hingga tokoh agama dan politik <sup>49</sup>.

Karakteristik YouTube yang berbasis video memungkinkan pendekatan komunikasi yang lebih ekspresif, personal, dan visual, sehingga menciptakan relasi yang lebih kuat antara kreator konten dan audiensnya. Fitur komentar, likes, dan live chat memberi ruang interaktivitas dan partisipasi warganet dalam proses komunikasi publik. Dalam konteks Indonesia, kanal-kanal YouTube seperti Narasi TV, Deddy Corbuzier, Refly Harun, dan Rocky Gerung Official menjadi contoh nyata bagaimana ruang publik digital terbentuk

---

<sup>47</sup> Burgess, J., & Green, *YouTube: Online Video and Participatory Culture*.

<sup>48</sup> “Kemp, S. (2023). Digital 2023: Indonesia. DataReportal. <https://Datareportal.Com/Reports/Digital-2023-Indonesia>,” n.d.

<sup>49</sup> Alim, S., & Dharma, A. F. (2021). YouTube Sebagai Ruang Publik Alternatif Bagi Anak Muda. *Expose: Jurnal Ilmu Komunikasi*, 4(1), 1–10. <https://doi.org/10.33021/Expose.V4i1.1223>,” n.d.

melalui konten berbasis opini, wawancara, dan dialog terbuka mengenai isu-isu nasional<sup>50</sup>.

Konsep YouTube sebagai ruang publik digital dapat dianalisis melalui kerangka ruang publik Habermasian. Habermas (1989) menggambarkan ruang publik sebagai arena diskusi rasional di mana warga negara dapat bertukar gagasan tanpa tekanan institusional. Meskipun YouTube tidak sepenuhnya mencerminkan model ruang publik ideal—karena adanya monetisasi, algoritma, dan polarisasi konten—platform ini tetap memberikan peluang bagi masyarakat untuk berpartisipasi dalam wacana sosial dan politik secara terbuka, terutama ketika media arus utama dinilai kurang independen atau terbatas cakupannya<sup>51</sup>.

Dinamika interaksi di YouTube tidak lepas dari tantangan. Fitur komentar, yang awalnya ditujukan untuk meningkatkan keterlibatan audiens, juga menjadi ladang bagi lahirnya komentar toksik, ujaran kebencian, hingga disinformasi. Studi oleh Santoso dan Widodo (2023) mencatat bahwa dalam berbagai video bertema politik dan agama, komentar-komentar yang muncul sering kali bersifat agresif, provokatif, atau bahkan menyerang identitas pribadi narasumber atau komentator lain. Hal ini menunjukkan bahwa walaupun YouTube menawarkan potensi sebagai ruang publik,

---

<sup>50</sup> R. Purwaningsih, “Aktivisme Digital Dalam Konten Politik YouTube Indonesia.,” *Jurnal Komunikasi Indonesia*, 2022, 11(2), 150–165.

<sup>51</sup> J. Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. (MIT Press., 1989).

kualitas diskursus digital masih dipengaruhi oleh literasi digital, regulasi platform, dan dinamika ideologi warganet<sup>52</sup>.

Posisi YouTube sebagai ruang publik digital di Indonesia mengandung potensi dan paradoks. Di satu sisi, platform ini memberi peluang partisipasi demokratis yang belum pernah terjadi sebelumnya; namun di sisi lain, ia juga berpotensi mereproduksi ketimpangan wacana, konflik sosial, dan penyebaran ujaran toksik yang merusak tatanan diskusi yang sehat. Oleh karena itu, penting untuk mengkaji lebih dalam bagaimana ruang-ruang digital seperti YouTube digunakan, dibentuk, dan dipahami dalam konteks sosiokultural Indonesia.

#### c. Peran Algoritma dan Keterlibatan Pengguna

Secara umum, algoritma dapat dipahami sebagai seperangkat instruksi logis dan terstruktur yang dirancang untuk menyelesaikan suatu permasalahan atau menjalankan tugas tertentu secara otomatis dan sistematis. Dalam konteks teknologi informasi, algoritma merupakan inti dari pemrosesan data dalam sistem komputer, di mana setiap langkah dirancang untuk menghasilkan output yang diinginkan berdasarkan input yang diberikan<sup>53</sup>. Algoritma bekerja melalui proses pengambilan keputusan berbasis aturan logis, statistik, atau pembelajaran mesin, yang memungkinkan perangkat lunak atau sistem

---

<sup>52</sup> Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest*.

<sup>53</sup> C. Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, *Introduction to Algorithms (3rd Ed.)*. (MIT Press., 2009).

digital mengambil tindakan tanpa intervensi manusia secara langsung<sup>54</sup>.

Dalam perkembangan teknologi digital, terutama pada platform media sosial dan layanan berbasis internet seperti Google, Facebook, dan YouTube, algoritma digunakan untuk memfilter, memprioritaskan, dan merekomendasikan informasi yang dianggap paling relevan bagi pengguna. Hal ini dikenal dengan istilah *algorithmic curation*, yakni proses kurasi konten secara otomatis oleh sistem berdasarkan perilaku dan preferensi pengguna<sup>55</sup>. Dengan demikian, algoritma tidak hanya berperan sebagai mesin pengolah data, tetapi juga sebagai aktor tak terlihat yang membentuk pengalaman digital, mempengaruhi opini publik, serta memediasi akses terhadap informasi<sup>56</sup>.

Dalam konteks YouTube, algoritma bekerja dengan menggabungkan berbagai indikator seperti riwayat tontonan, durasi menonton, tingkat keterlibatan (*engagement*), serta metadata konten, untuk membentuk sistem rekomendasi yang bersifat personal<sup>57</sup>. Sistem ini bertujuan untuk meningkatkan waktu tonton pengguna dan menjaga loyalitas terhadap platform, namun di sisi lain juga

---

<sup>54</sup> P. Russell, S. J., & Norvig, "Artificial Intelligence: A Modern Approach," *Pearson Education*. 3rd ed (2010).

<sup>55</sup> T Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. (Yale University Press., 2018).

<sup>56</sup> P. M. Napoli, *Social Media and the Public Interest: Media Regulation in the Disinformation Age*. (Columbia University Press., 2019).

<sup>57</sup> E. Covington, P., Adams, J., & Sargin, "Deep Neural Networks for YouTube Recommendations.," *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016, 191–198., <https://doi.org/10.1145/2959100.2959190>.

menimbulkan kekhawatiran tentang bias algoritmik, efek echo chamber, dan penyebaran informasi ekstrem secara masif<sup>58</sup>.

Oleh karena itu, pemahaman tentang algoritma tidak bisa dilepaskan dari dimensi teknis dan sosialnya. Ia bukan hanya sekadar "perhitungan otomatis," tetapi juga bagian dari mekanisme kekuasaan informasi digital, yang menentukan apa yang terlihat dan apa yang tersembunyi dalam ekosistem komunikasi online modern.

Algoritma YouTube memainkan peran krusial dalam membentuk pengalaman pengguna dengan mengatur konten apa yang mereka lihat, kapan mereka melihatnya, dan seberapa sering konten tersebut muncul. Sistem ini dirancang bukan hanya untuk menampilkan video secara acak, tetapi untuk mengoptimalkan waktu tonton, keterlibatan, dan retensi pengguna, dengan cara merekomendasikan konten yang paling sesuai dengan preferensi individu berdasarkan perilaku sebelumnya<sup>59</sup>. Dengan pendekatan berbasis pembelajaran mesin (machine learning), algoritma YouTube menganalisis jutaan titik data secara real-time untuk menyusun daftar rekomendasi yang dipersonalisasi bagi setiap pengguna<sup>60</sup>.

---

<sup>58</sup> E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You.*, 2011 (Penguin Books., n.d.).

<sup>59</sup> Covington, P., Adams, J., & Sargin, "Deep Neural Networks for YouTube Recommendations."

<sup>60</sup> Jim Maddock et al., "Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures," *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 2015, 228–41, <https://doi.org/10.1145/2675133.2675280>.

Beberapa faktor utama yang digunakan algoritma dalam menentukan urutan video yang direkomendasikan antara lain: riwayat pencarian, durasi waktu menonton, frekuensi interaksi (like, komentar, share, dan subscribe), serta frekuensi unggahan dari kanal tertentu <sup>61</sup>. Video dengan performa interaksi yang tinggi cenderung memperoleh "keuntungan algoritmik" karena dianggap relevan dan menarik bagi pengguna lain. Hal ini menciptakan fenomena umpan balik positif, di mana video populer akan semakin populer karena sering direkomendasikan, sementara video dengan engagement rendah berisiko tenggelam dalam sistem distribusi YouTube <sup>62</sup>.

Namun demikian, sistem rekomendasi berbasis engagement ini menyimpan potensi risiko terhadap polarisasi konten dan penyebaran informasi yang salah. Beberapa studi menunjukkan bahwa algoritma YouTube secara tidak langsung dapat memperkuat bias kognitif pengguna dan mendorong mereka ke dalam echo chamber atau filter bubble, di mana mereka hanya terpapar pada informasi yang memperkuat pandangan mereka sendiri <sup>63</sup>. Ini menjadi masalah serius terutama ketika pengguna mengonsumsi konten yang bersifat konspiratif, ekstremis, atau sarat dengan narasi kebencian. Dalam

---

<sup>61</sup> Covington, P., Adams, J., & Sargin, "Deep Neural Networks for YouTube Recommendations."

<sup>62</sup> F. Figueiredo, F., Almeida, J. M., Gonçalves, M. A., & Benevenuto, "On the Dynamics of Social Media Popularity: A YouTube Case Study," *ACM Transactions on Internet Technology*, no. 14(4), 1–23. (2014), <https://doi.org/10.1145/2633121>.

<sup>63</sup> J. Munger, K., & Phillips, "Right-wing YouTube: A Supply and Demand Perspective.," *International Journal of Press/Politics*, no. 25(3), 480–500. (2020), <https://doi.org/10.1177/1940161219892519>.

konteks ini, algoritma dapat secara tidak sengaja memprioritaskan video yang mengandung misinformasi, clickbait, atau bahkan komentar toksik, karena konten semacam itu sering kali memicu respons emosional tinggi dan meningkatkan interaksi<sup>64</sup>

Sebagai contoh, dalam analisis konten video bertema politik dan agama di YouTube Indonesia, komentar dengan nada agresif, ofensif, atau provokatif cenderung memiliki engagement lebih tinggi, baik dalam bentuk like, reply, maupun visibilitas algoritmik. Situasi ini memperlihatkan bahwa keterlibatan pengguna tidak selalu menandakan kualitas atau kedalaman diskusi, tetapi dapat menjadi indikator dari kontroversi dan konflik<sup>65</sup>. Maka dari itu, YouTube menghadapi dilema antara mempertahankan engagement sebagai metrik utama dan menjaga ruang diskusi yang sehat di platformnya.

Sebagai respons terhadap kritik publik dan tekanan dari lembaga pengatur, YouTube telah memperkenalkan berbagai mekanisme content moderation, seperti penghapusan otomatis konten berbahaya, klasifikasi usia, label peringatan, dan algoritma penurunan peringkat (downranking) untuk video yang dianggap borderline harmful<sup>66</sup>. Namun efektivitas pendekatan ini masih diperdebatkan,

---

<sup>64</sup> W. Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira Jr, "Auditing Radicalization Pathways on YouTube. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT)," 2020, 131–141., <https://doi.org/10.1145/3351095.3372879>.

<sup>65</sup> B. Santoso, Y., & Widodo, "Analisis Komentar Toksik Pada Video Politik Di YouTube.," *Jurnal Teknologi Informasi Dan Komunikasi*, no. 10(1), 71–85. (2023).

<sup>66</sup> Google Transparency Report., "YouTube Community Guidelines Enforcement.," 2023, <https://transparencyreport.google.com/youtube-policy/removals>.

terutama ketika konten berbahaya tidak secara eksplisit melanggar kebijakan, tetapi tetap mengandung muatan toksik dalam bentuk komentar, narasi, atau framing yang manipulatif<sup>67</sup>.

Algoritma YouTube tidak hanya menjadi penentu konten apa yang populer, tetapi juga aktor tidak langsung dalam membentuk kualitas wacana publik digital. Dalam konteks penelitian ini, penting untuk memahami bagaimana sistem ini berinteraksi dengan perilaku pengguna dan bagaimana algoritma dapat memperkuat atau meredam kemunculan komentar toksik dalam video-video bertema sensitif seperti yang melibatkan tokoh publik dengan pandangan yang kontras.

## **2. Komentar Toksik di Media Sosial**

### **a. Definisi Komentar Toksik**

Komentar toksik di media sosial merujuk pada pernyataan atau tanggapan yang bersifat kasar, merendahkan, provokatif, atau tidak pantas, yang berpotensi mengganggu kenyamanan pengguna lain dan merusak kualitas diskusi daring. Perspective API (2023) mendefinisikan komentar toksik sebagai “komentar yang kasar, tidak sopan, atau tidak masuk akal, yang kemungkinan besar menyebabkan orang meninggalkan diskusi.” Definisi ini menegaskan bahwa komentar toksik tidak hanya dinilai dari isi verbal yang eksplisit, tetapi juga dari dampaknya terhadap suasana interaksi digital. Komentar toksik sering mencakup ujaran kebencian, penghinaan, pelecehan,

---

<sup>67</sup> Maddock et al., “Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures.”

ancaman, hingga serangan terhadap identitas pribadi seperti ras, agama, jenis kelamin, atau orientasi seksual<sup>68</sup>.

Karakteristik utama dari komentar toksik mencakup penggunaan bahasa yang agresif, provokatif, dan penuh kebencian yang tidak berkontribusi pada diskusi rasional. Komentar ini biasanya disampaikan dengan maksud untuk memicu emosi, konflik, atau memprovokasi reaksi negatif dari pengguna lain<sup>69</sup>. Komentar toksik juga sering kali disertai dengan informasi yang menyesatkan atau manipulatif, sehingga memperburuk kualitas komunikasi publik di ruang digital<sup>70</sup>. Dalam konteks platform seperti YouTube atau Twitter, komentar toksik bisa muncul dalam berbagai bentuk seperti trolling (provokasi sengaja), flaming (serangan verbal langsung), cyberbullying (pelecehan terus-menerus), dan hate speech (ujaran kebencian)<sup>71</sup>.

Selain bentuknya yang bervariasi, komentar toksik juga diperkuat oleh fenomena online disinhibition effect, yaitu kecenderungan pengguna untuk lebih berani, blak-blakan, atau kasar ketika berkomunikasi di internet karena merasa terlindungi oleh

---

<sup>68</sup> Hanscom, R., Lehman, T. S., Lv, Q., & Mishra, "The Toxicity Phenomenon Across Social Media."

<sup>69</sup> H. Zhou, X., Zafarani, R., Shu, K., & Liu, "Characterizing Online Abuse: A Survey," no. arXiv preprint arXiv:2101.12552. (2021), <https://arxiv.org/abs/2101.12552>.

<sup>70</sup> Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*.

<sup>71</sup> J. Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, *Antisocial Behavior in Online Discussion Communities. Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM)*, 2017.

anonimitas dan jarak sosial <sup>72</sup>Pengguna merasa bebas berekspresi tanpa harus mempertanggungjawabkan kata-katanya secara langsung. Hal ini semakin diperparah oleh lemahnya moderasi konten dan algoritma platform yang justru mendorong komentar yang memicu interaksi tinggi tanpa mempertimbangkan substansi atau etika komunikasi <sup>73</sup>.

Komentar toksik bukan hanya sekadar “kata-kata kasar,” melainkan merupakan gejala dari krisis etika komunikasi digital yang dipengaruhi oleh struktur platform, perilaku pengguna, dan insentif sistem algoritmik. Dalam konteks penelitian ini, pemahaman mendalam tentang karakteristik komentar toksik menjadi fondasi penting untuk menganalisis bagaimana dinamika konflik dalam diskusi antara tokoh publik seperti Rocky Gerung dan Ustadz Abdul Somad berkembang di ruang digital.

#### b. Faktor Penyebab Munculnya Komentar Toksik

Munculnya komentar toksik di media sosial bukanlah fenomena yang berdiri sendiri, melainkan hasil dari berbagai faktor struktural, psikologis, dan teknologis yang saling terkait. Pemahaman atas faktor-faktor ini sangat penting agar pendekatan penanggulangan yang diterapkan dapat bersifat preventif, bukan hanya reaktif. Setidaknya terdapat empat faktor utama yang sering disebut dalam literatur sebagai pemicu utama perilaku toksik di ruang digital.

---

<sup>72</sup> J. Suler, “The Online Disinhibition Effect,” *CyberPsychology & Behavior*, no. 7(3), 321–326. (2004), <https://doi.org/10.1089/1094931041291295>.

<sup>73</sup> Maddock et al., “Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures.”

### 1) Anonimitas

Salah satu penyebab utama munculnya komentar toksik adalah anonimitas, yaitu kemampuan pengguna untuk menyembunyikan identitas asli mereka saat berinteraksi secara daring. Menurut Christopherson (2007), anonimitas dapat menurunkan rasa tanggung jawab individu terhadap tindakannya karena tidak adanya konsekuensi sosial langsung. Dalam ruang yang anonim, seseorang lebih mudah mengekspresikan pikiran secara ekstrem, provokatif, bahkan menyakiti orang lain tanpa rasa bersalah<sup>74</sup>. Efek ini diperparah oleh struktur teknis media sosial yang memungkinkan pembuatan akun palsu atau penggunaan identitas samaran, yang menurunkan akuntabilitas dan memperbesar potensi perilaku menyimpang<sup>75</sup>

### 2) Efek Disinhibisi Online (Online Disinhibition Effect)

Fenomena ini menggambarkan bagaimana individu cenderung merasa lebih bebas dan kurang terhambat saat berkomunikasi secara daring dibandingkan dengan interaksi langsung. Suler (2004) menjelaskan bahwa disinhibisi ini dapat bersifat positif (misalnya, pengungkapan diri) maupun negatif (misalnya, penghinaan atau ujaran kebencian). Dalam konteks komentar

---

<sup>74</sup> Noam Lapidot-Lefler and Azy Barak, "The Benign Online Disinhibition Effect: Could Situational Factors Induce Self-Disclosure and Prosocial Behaviors?," *Cyberpsychology* 9, no. 2 (2015), <https://doi.org/10.5817/CP2015-2-3>.

<sup>75</sup> "Barlett, C. P. (2015). Anonymously Hurting Others Online: The Effect of Anonymity on Cyberbullying Frequency. *Psychology of Popular Media Culture*, 4(2), 70–79. <https://doi.org/10.1037/Ppm0000013>," n.d.

toksik, disinhibisi negatif menjadi pemicu agresi verbal, sarkasme, atau bahkan ancaman yang tidak mungkin diucapkan dalam percakapan tatap muka. Faktor-faktor seperti tidak adanya ekspresi wajah lawan bicara, jeda waktu dalam respons, dan persepsi bahwa lawan diskusi bukanlah "manusia nyata," semakin memperbesar efek ini <sup>76</sup>

### 3) Kurangnya Moderasi

Ketiadaan atau kelemahan dalam sistem moderasi komentar turut menjadi pemicu suburnya komentar toksik di media sosial. Menurut Gillespie (2018), sebagian besar platform mengandalkan laporan pengguna atau algoritma untuk menyaring komentar ofensif, namun pendekatan ini sering kali tidak efektif dalam mengidentifikasi ujaran kebencian yang bersifat terselubung, kontekstual, atau sarkastik. Dalam banyak kasus, komentar dengan muatan toksik tetap lolos karena keterbatasan teknologi dalam memahami konteks budaya dan bahasa lokal<sup>77</sup>. Selain itu, kebijakan platform yang longgar terhadap "freedom of speech" sering kali dijadikan pembenaran atas toleransi terhadap perilaku bermasalah yang justru menciptakan iklim diskusi yang tidak sehat.

---

<sup>76</sup> D. L. Buckels, E. E., Trapnell, P. D., & Paulhus, "Trolls Just Want to Have Fun. *Personality and Individual Differences*" 67, 97–102 (2014).

<sup>77</sup> E. Jhaver, S., Bruckman, A., & Gilbert, "Does Transparency in Moderation Really Matter? *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW)," 2019, 1–27.

#### 4) Algoritma Platform

Algoritma media sosial—termasuk YouTube, Facebook, dan TikTok—umumnya dirancang untuk memaksimalkan engagement, seperti klik, komentar, dan waktu tonton. Sayangnya, komentar toksik sering kali menghasilkan tingkat keterlibatan yang tinggi karena memicu emosi dan kontroversi<sup>78</sup>. Akibatnya, algoritma dapat secara tidak langsung memperkuat visibilitas komentar semacam ini, memperbesar peluangnya untuk dilihat dan direspons oleh pengguna lain<sup>79</sup>. Hal ini dikenal sebagai amplifikasi algoritmik, di mana sistem justru memberi insentif terhadap konten yang kontroversial atau problematik karena dianggap “menarik.” Dalam jangka panjang, algoritma semacam ini dapat menciptakan efek feedback loop yang memperkuat budaya debat yang kasar dan penuh konflik di ruang digital.

#### c. Dimensi dan Kategori Toksisitas

Untuk memahami lebih dalam fenomena komentar toksik di media sosial, penting untuk mengidentifikasi dimensi-dimensi utama yang menyusun sifat toksisitas itu sendiri. Dalam konteks penelitian komunikasi digital dan pengolahan bahasa alami (Natural Language Processing/NLP), berbagai model klasifikasi telah mengelompokkan komentar toksik ke dalam sejumlah kategori semantik dan pragmatis.

---

<sup>78</sup> Maddock et al., “Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures.”

<sup>79</sup> Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira Jr, “Auditing Radicalization Pathways on YouTube. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT).”

---

Salah satu model yang paling sering digunakan, seperti yang dikembangkan oleh Jigsaw melalui Perspective API dan didukung oleh penelitian Wulczyn pada tahun 2017, membagi toksisitas ke dalam beberapa dimensi penting:

1) *Toxicity Score*

Salah satu metrik utama adalah Toxicity Score, yang mewakili sejauh mana komentar tersebut dianggap kasar, tidak sopan, atau tidak masuk akal<sup>80</sup>. Skor ini membantu peneliti mengidentifikasi komentar yang berpotensi merusak suasana diskusi dan mempengaruhi dinamika sosial dalam komunitas online.

2) *Severe Toxicity Score*

Selain *Toxicity Score*, *Communalytic* juga menyediakan *Severe Toxicity Score*, yang mengukur seberapa besar komentar tersebut bersifat penuh kebencian, agresif, dan tidak hormat. Skor ini berguna untuk mengidentifikasi komentar yang bukan hanya kasar, tetapi juga berbahaya karena mengandung elemen-elemen kebencian yang kuat<sup>81</sup>. Metrik ini sangat penting dalam konteks memantau dan mengelola komentar yang dapat memicu konflik atau merusak kesejahteraan mental individu.

---

<sup>80</sup> Lusy Novitasari, "Penyimpangan Perilaku Seks Waria Dalam Novel Taman Api Karya Yonathan Rahardjo," *Deiksis* 10, no. 02 (2018): 125, <https://doi.org/10.30998/deiksis.v10i02.2339>.

<sup>81</sup> Irwan Nugroho and Irwansyah Irwansyah, "Konvergensi Konten Audio Di Media Online (Studi Kasus Podcast Detik.Com)," *Jurnal Komunikasi* 15, no. 1 (2021): 55–70, <https://doi.org/10.21107/ilkom.v15i1.9753>.

### 3) *Identity Attack Score*

Metrik lain yang disediakan oleh *Communalytic* adalah *Identity Attack Score*, yang menunjukkan apakah suatu postingan mengandung bahasa yang penuh kebencian yang menargetkan seseorang karena identitas mereka, seperti ras, agama, atau orientasi seksual. Pengukuran ini penting untuk memahami bagaimana komentar yang bersifat diskriminatif dapat menyebar di media sosial dan dampaknya terhadap kelompok yang rentan <sup>82</sup>. Dengan menggunakan metrik ini, peneliti dapat mengidentifikasi dan mengatasi masalah diskriminasi dan marginalisasi di dunia digital.

### 4) *Insult Score*

*Insult Score* adalah metrik lain yang membantu mengidentifikasi postingan yang bersifat menghina atau bersifat inflamasi. Komentar yang termasuk dalam kategori ini sering kali dirancang untuk memprovokasi atau merendahkan individu atau kelompok tertentu <sup>83</sup>. Mengukur *Insult Score* dapat memberikan wawasan tentang seberapa sering penghinaan terjadi dalam diskusi online dan bagaimana hal ini mempengaruhi interaksi sosial di platform digital.

---

<sup>82</sup> Santoso, Y., & Widodo, “Analisis Komentar Toksik Pada Video Politik Di YouTube.”

<sup>83</sup> Adel Andila Putri and Jurnalis Data, “Tantangan Baru Industri Media Di Indonesia” (Jakarta, 2023), <https://goodstats.id/article/tantangan-baru-industri-media-di-indonesia-XIPEa>.

### 5) *Profanity Score*

Selain itu, *Communalytic* juga mengukur *Profanity Score*, yang menunjukkan apakah dalam komentar tersebut terdapat kata-kata kasar atau bahasa yang tidak senonoh. Metrik ini penting untuk menjaga standar komunikasi di media sosial, terutama di platform yang digunakan oleh berbagai kelompok umur. Mengidentifikasi dan memoderasi komentar dengan tingkat Profanity yang tinggi dapat membantu menciptakan lingkungan diskusi yang lebih sehat dan konstruktif.

### 6) *Threat Score*

Terakhir, *Threat Score* di *Communalytic* mengukur sejauh mana suatu postingan mengandung ancaman kekerasan atau niat untuk menyakiti individu atau kelompok. Metrik ini sangat penting dalam upaya menjaga keamanan di komunitas online, karena ancaman kekerasan tidak hanya merusak komunikasi tetapi juga dapat menimbulkan risiko nyata bagi keselamatan fisik seseorang<sup>84</sup>. Dengan mengidentifikasi dan menindaklanjuti komentar-komentar yang mengandung ancaman, platform media sosial dapat melindungi penggunanya dari bahaya.

---

<sup>84</sup> Dina Nisa Aulia and Dunan Amri, "Konvergensi Media Dan New Media Dalam Transformasi" 6, no. 1 (2025): 48–63.

Secara keseluruhan, *Communalytic.org* menawarkan alat yang komprehensif dan canggih untuk mengukur dan menganalisis toxic comments di media sosial. Dengan menggunakan metrik yang beragam, aplikasi ini memungkinkan peneliti untuk mendapatkan gambaran yang lebih mendalam tentang dinamika komunikasi di dunia digital dan membantu mengembangkan strategi untuk mengatasi masalah toxic comments secara efektif<sup>85</sup>. Ini membuat *Communalytic* menjadi alat yang sangat berharga dalam upaya menciptakan lingkungan media sosial yang lebih aman dan inklusif.

d. Dampak komentar toksik terhadap komunikasi publik dan psikologi pengguna

Komentar toksik di media sosial tidak hanya menimbulkan gangguan sementara dalam percakapan daring, tetapi juga membawa dampak jangka panjang yang signifikan terhadap kualitas komunikasi publik serta kondisi psikologis pengguna. Secara umum, komentar toksik dapat melemahkan ekosistem diskusi digital, menciptakan lingkungan komunikasi yang penuh ketegangan, serta menghambat partisipasi konstruktif dari warganet yang berniat berdialog secara sehat<sup>86</sup>

---

<sup>85</sup> Nugroho and Irwansyah, "Konvergensi Konten Audio Di Media Online (Studi Kasus Podcast Detik.Com)."

<sup>86</sup> Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*.

Dalam konteks komunikasi publik, kehadiran komentar toksik menyebabkan apa yang disebut sebagai efek "silencing", yaitu kecenderungan pengguna untuk menahan diri dari berkomentar atau mengekspresikan pendapat karena takut menjadi sasaran pelecehan atau agresi verbal (Matias, 2019; Nadim & Fladmoe, 2021). Hal ini terutama dialami oleh perempuan, kelompok minoritas, dan individu dengan pandangan politik atau agama yang dianggap kontroversial. Ketika sebagian kelompok merasa tidak aman untuk berbicara di ruang publik digital, maka yang terjadi adalah distorsi wacana, di mana hanya kelompok dominan atau paling vokal yang menguasai ruang dialog<sup>87</sup>. Akibatnya, komunikasi publik kehilangan sifat deliberatif dan partisipatif yang menjadi fondasi demokrasi digital.

Selain mengganggu struktur komunikasi, komentar toksik juga berdampak serius terhadap psikologi individu. Studi menunjukkan bahwa paparan terus-menerus terhadap ujaran kebencian, penghinaan, atau ancaman dapat menyebabkan gangguan psikologis seperti stres, kecemasan, depresi, bahkan trauma psikososial, terutama pada pengguna muda dan kelompok rentan<sup>88</sup>. Dalam banyak kasus, korban komentar toksik mengalami apa yang disebut sebagai online

---

<sup>87</sup> Tubbs. Stewart L dan Moss Sylvia, *Human Communication : Prinsip - Prinsip Dasar* (PT. Remaja Rosdakarya, 2008).

<sup>88</sup> N. Chou, H. T. G., & Edge, "“They Are Happier and Having Better Lives than I Am”: The Impact of Using Facebook on Perceptions of Others’ Lives. *Cyberpsychology, Behavior, and Social Networking*” 15(2), 117 (2012), <https://doi.org/10.1089/cyber.2011.0324>.

harassment fatigue, yaitu kelelahan emosional akibat interaksi beracun secara berulang yang tidak bisa dikendalikan<sup>89</sup>.

Komentar toksik juga merusak kepercayaan sosial antar pengguna dan memperkuat polarisasi masyarakat. Konten komentar yang mengandung penghinaan atau ujaran kebencian cenderung memicu respons emosional negatif dari pengguna lain, menciptakan efek bola salju dalam bentuk perdebatan yang semakin panas dan destruktif<sup>90</sup>. Ini berkontribusi terhadap terbentuknya ekosistem informasi yang penuh dengan konflik, bukan kolaborasi. Dalam jangka panjang, efek ini dapat memperdalam jurang perbedaan ideologi, memperlemah kohesi sosial, dan mengurangi kemampuan masyarakat untuk berdialog dalam keberagaman<sup>91</sup>.

Dari sisi platform, komentar toksik berdampak pada legitimasi dan kredibilitas media sosial sebagai ruang publik modern. Jika tidak ditangani dengan tepat, tingginya tingkat toksisitas dapat menurunkan reputasi platform, menyebabkan eksodus pengguna, dan memicu tekanan dari regulator maupun pengiklan<sup>92</sup>. Oleh karena itu, penanggulangan komentar toksik tidak hanya menjadi isu teknis atau

---

<sup>89</sup> E. A. Jane, "Flaming? What Flaming? The Pitfalls and Potentials of Researching Online Hostility. *Ethics and Information Technology*, 17(1)," 2015, 65–87., <https://doi.org/10.1007/s10676-015-9362-0> <https://doi.org/10.1007/s10676-015-9362-0>.

<sup>90</sup> Maddock et al., "Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures."

<sup>91</sup> Sunstein, *#Republic: Divided Democracy in the Age of Social Media*.

<sup>92</sup> Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*.

moderasi konten, tetapi juga menyangkut etika, kesehatan mental, dan keberlangsungan ekosistem digital yang demokratis.

### **3. Penelitian Relevan**

Kajian terdahulu memberikan landasan penting bagi penelitian ini untuk memahami perkembangan keilmuan seputar fenomena komentar toksik di media sosial. Dalam beberapa tahun terakhir, topik ini mendapat perhatian yang semakin besar seiring meningkatnya penggunaan platform digital seperti YouTube sebagai ruang diskusi publik. Para peneliti dari berbagai disiplin telah mengkaji bagaimana komentar beracun atau toksik memengaruhi dinamika komunikasi daring, membentuk persepsi publik, dan berdampak terhadap kesejahteraan psikologis pengguna.

Penelitian-penelitian terdahulu juga telah mengeksplorasi dimensi linguistik komentar toksik, peran algoritma dalam amplifikasi konten bermasalah, serta strategi moderasi dan intervensi untuk mengurangi ujaran kebencian secara daring.

Studi-studi tersebut tidak hanya memperkuat urgensi penelitian ini, tetapi juga menggarisbawahi adanya kesenjangan kontekstual, terutama dalam ruang diskusi publik di Indonesia yang kerap kali diwarnai perdebatan tajam, polarisasi ideologis, dan ekspresi kebencian yang terlegitimasi oleh fitur komentar terbuka di platform seperti YouTube. Oleh karena itu, pemetaan dan analisis atas penelitian-penelitian terdahulu menjadi langkah strategis untuk

memperjelas posisi penelitian ini dalam lanskap keilmuan yang relevan.

Adapun kanjian terdahulunya sebagaimana dijabarkan dalam table berikut ini:

**Tabel 1 Kajian Terdahulu**

| No | Judul                                                                           | Tahun | Jurnal                         | Temuan                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|----|---------------------------------------------------------------------------------|-------|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | <i>Analyzing Toxicity in YouTube Comments with the Help of Machine Learning</i> | 2023  | (Sasan & Gunica, thesis)       | Penelitian ini mengumpulkan sampel acak 79.577 komentar dari saluran YouTube Swedia, menggunakan Hatescan untuk memberi skor toksisitas. Hasil menunjukkan bahwa hanya 0,643 % dari komentar yang tergolong toxic; sebagian besar berupa hinaan personal dan komentar merendahkan kecerdasan atau kompetensi seseorang. Komentar politik cenderung umum atau tidak menyasar pribadi. Para penulis menekankan pentingnya memahami distribusi jenis toxicity agar pembuat konten siap menghadapi ancaman digital secara spesifik <sup>93</sup> . |
| 2  | <i>Multilabel Classification for Toxic Comments in Indonesian</i>               | 2020  | EMACS Journal (Rumagit et al.) | Studi membangun model deteksi komentar toxic dalam bahasa Indonesia menggunakan machine learning (multilabel). Metode seperti CNN, FastText, logistic regression diuji untuk mendeteksi ancaman, ujaran kebencian, dan hinaan. Hasilnya                                                                                                                                                                                                                                                                                                        |

<sup>93</sup> J. Sasan, D., & Gunica, "Analyzing Toxicity in YouTube Comments with the Help of Machine Learning [Undergraduate Thesis, Stockholm University].," *DiVA Portal.*, 2023, <https://www.diva-portal.org/smash/get/diva2:1784332/FULLTEXT01.pdf>.

|   |                                                                                                                                  |      |                              |                                                                                                                                                                                                                                                                                                                                                                                                      |
|---|----------------------------------------------------------------------------------------------------------------------------------|------|------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |                                                                                                                                  |      |                              | menunjukkan model efektif dalam memetakan berbagai tingkat toxicity dan identitas target, serta mampu membantu menjaga diskusi daring yang inklusif <sup>94</sup> .                                                                                                                                                                                                                                  |
| 3 | <i>Abusive comment identification on Indonesian social media data using hybrid deep learning</i>                                 | 2022 | IJ-AI Journal                | Tiara Intana Sari dkk. mengembangkan model hybrid (RNN + LSTM) untuk mendeteksi komentar abusif berbahasa Indonesia. Dengan dataset Twitter, model mencapai f-measure 94 % dan peningkatan akurasi 23 %. Hasil mengonfirmasi bahwa teknik deep learning hybrid lebih unggul dibanding metode tunggal, serta relevan untuk memproses komentar toxic di platform berbasis bahasa lokal <sup>95</sup> . |
| 4 | <i>YouTube, Public Discourse, and the 'Makan Siang Gratis' Program: An Analysis of Toxicity Comments on the Liputan6 Channel</i> | 2025 | JCC-Indonesia (Jayus et al.) | Fokus pada program "Makan Siang Gratis" di kanal Liputan6 YouTube, studi menggunakan Communalytic untuk scraping dan scoring komentar berdasarkan severity, hate speech, identity attack, dan threat. Ditemukan sebagian besar komentar mempunyai skor toxicity rendah, namun sejumlah kecil mencakup severe toxicity dan profanity yang dapat memicu konflik emosional                              |

<sup>94</sup> Reinert Yosua Rumagit, "Multilabel Classification for Toxic Comments in Indonesian," *Engineering Mathematics and Computer Science (EMACS) Journal*, no. (1):29-34 (2020), <https://doi.org/DOI:10.21512/emacsjournal.v2i1.6256>.

<sup>95</sup> Tiara Intana Sari et al., "Abusive Comment Identification on Indonesian Social Media Data Using Hybrid Deep Learning," *IAES International Journal of Artificial Intelligence* 11, no. 3 (2022): 895–904, <https://doi.org/10.11591/ijai.v11.i3.pp895-904>.

|   |                                                                                                                                                                                                        |      |                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |                                                                                                                                                                                                        |      |                                   | dan mengurangi kualitas interaksi. Studi menyarankan moderasi proaktif untuk menjaga diskursus yang inklusif dan konstruktif <sup>96</sup> .                                                                                                                                                                                                                                                                                                                                                          |
| 5 | <i>Clicks, comments, consequences: Are content creators' socio-structural And platform characteristics shaping the exposure to negative sentiment, offensive language, and hate speech on YouTube?</i> | 2025 | ArXiv pre-print (Weißmann et al.) | Penelitian pada 3.695 channel YouTube Jerman-bahasa melibatkan ~40 juta komentar. Temuan menunjukkan bahwa channel dengan topik politik/conspirasi menerima lebih banyak negative sentiment dan hate speech; perempuan terkena lebih sedikit komentar negatif; sementara BIPoC lebih sering menerima negative sentiment tetapi tidak lebih sering mendapat hate speech ofensif. Hal ini menunjukkan hubungan antara karakteristik pembuat konten dengan tingkat eksposur komentar toxic <sup>97</sup> |
| 6 | <i>Commenter Behavior Characterization on YouTube Channels</i>                                                                                                                                         | 2023 | ArXiv pre-print (Shajari et al.)  | Studi menggunakan analisis jaringan sosial (SNA) untuk mendeteksi commenter mobs di YouTube. Data mencakup 7.782 video, 294.199 komentator, dan 596.982 komentar. Hasil menunjukkan adanya perilaku koordinasi                                                                                                                                                                                                                                                                                        |

<sup>96</sup> S Jayus, J., Abdullah, A., Mustafa, A., & Sumaiyah, "YouTube, Public Discourse, and the 'Makan Siang Gratis' Program: An Analysis of Toxicity Comments on the Liputan6 Channel.," *Proceeding Jogja Communication Conference (JCC-Indonesia)*, 8(1), 1–20., 2015, <https://jcc-indonesia.id/proceeding/index.php/jcc/article/view/383>.

<sup>97</sup> C. Weißmann, S., Philipp, A., Verwiebe, R., Krauter, C. O., Fritsch, N. S., & Buder, "Clicks, Comments, Consequences: Are Content Creators' Socio-Structural and Platform Characteristics Shaping Exposure to Negative Sentiment, Offensive Language, and Hate Speech on YouTube?," *ArXiv Preprint.*, 2025, <https://arxiv.org/abs/2403.00745>.

|   |                                                                                                              |      |       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|---|--------------------------------------------------------------------------------------------------------------|------|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |                                                                                                              |      |       | terstruktur antar komunitas komentator dalam mendukung narasi tertentu—fenomena yang relevan dengan strategi buzzer dan pembentukan opini publik secara terorganisir <sup>98</sup> .                                                                                                                                                                                                                                                                                                  |
| 7 | <i>Hiding in Plain Sight: A Measurement and Analysis of Kids' Exposure to Malicious URLs on YouTube</i>      | 2020 | arXiv | Studi ini mengumpulkan dataset besar (~4 juta komentar) pada video anak-anak dan menemukan banyak URL jahat atau tidak pantas yang berpotensi membahayakan anak. Meskipun bukan secara eksplisit toxic speech, kehadiran URL ini menjadi medium penyebaran konten berbahaya, menunjukkan bahwa moderasi komentar juga perlu mempertimbangkan tautan yang disisipkan. Rekomendasi menyarankan kontrol URL dalam komentar sebagai bagian penting dari moderasi platform <sup>99</sup> . |
| 8 | <i>Hate, Obscenity, and Insults: Measuring the Exposure of Children to Inappropriate Comments in YouTube</i> | 2021 | arXiv | Analisis ~4 juta komentar di channel anak-anak menunjukkan ~11% komentar tergolong toxic, termasuk hate speech, obscenity, dan pelecehan. Temuan memperlihatkan anak-anak sangat                                                                                                                                                                                                                                                                                                      |

<sup>98</sup> Adewale Obadimu et al., “Identifying Toxicity within Youtube Video Comment,” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11549 LNCS (2019): 214–23, [https://doi.org/10.1007/978-3-030-21741-9\\_22](https://doi.org/10.1007/978-3-030-21741-9_22).

<sup>99</sup> “Alshamrani, S., Abusnaina, A., Abuhamad, M., Nyang, D., & Mohaisen, D. (2021). Hate, Obscenity, and Insults: Measuring the Exposure of Children to Inappropriate Comments in YouTube. ArXiv Preprint. <https://arxiv.org/abs/2103.09050>,” n.d.

|    |                                                                                                                        |      |              |                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----|------------------------------------------------------------------------------------------------------------------------|------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                                                        |      |              | terekspose konten tidak layak, bahkan setelah penggunaan classifier NLP dan machine learning. Penelitian menekankan perlunya moderasi otomatis untuk meminimalisir dampak psikologis bagi pengguna muda <sup>100</sup> .                                                                                                                                                                                                                 |
| 9  | <i>Deradicalizing YouTube: Characterization, Detection and Personalization of Religiously Intolerant Arabic Videos</i> | 2022 | ResearchGate | Fokus pada komentar berbahasa Arab, studi mengidentifikasi lima bentuk toxic yang sering muncul dalam video pro/anti-NATO di YouTube. Pengguna di channel anti-NATO lebih cenderung menghasilkan komentar beracun daripada pro-NATO. Word-cloud dan LDA menunjukkan penggunaan kata profan dan topik fake news lebih banyak di channel anti. Studi menyoroti perbedaan pola toxicity tergantung konteks ideologis kanal <sup>101</sup> . |
| 10 | <i>Analyzing Right-wing YouTube Channels: Hate, Violence and Discrimination</i>                                        | 2018 | arXiv        | Dataset mencakup >17 juta komentar dari >7.000 video kanal sayap kanan. Peneliti menemukan frekuensi tinggi kata-kata negatif, bias diskriminasi terhadap LGBT, dan ujaran kebencian terhadap Muslim. Pendekatan tiga-lapis (lexicon, topik, bias                                                                                                                                                                                        |

<sup>100</sup> “Alshamrani, S., Abusnaina, A., Mohaisen, D. (2020). Hiding in Plain Sight: A Measurement and Analysis of Kids’ Exposure to Malicious URLs on YouTube. ArXiv Preprint. <https://arxiv.org/abs/2009.07923>,” n.d.

<sup>101</sup> V. Ottoni, R., Cunha, E., Magno, G., Bernadina, P., Meira Jr., W., & Almeida, “Analyzing Right-Wing YouTube Channels: Hate, Violence and Discrimination,” *ArXiv Preprint*, 2018, <https://arxiv.org/abs/1804.04096>.

|    |                                                                             |      |                           |                                                                                                                                                                                                                                                                                                                                                                                  |
|----|-----------------------------------------------------------------------------|------|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                             |      |                           | implisit) mendukung kesimpulan bahwa kanal politik ideologis intensif mendatangkan toxic comment kategori berat <sup>102</sup> .                                                                                                                                                                                                                                                 |
| 11 | <i>Identifying Toxicity Within YouTube Video Comment</i>                    | 2020 | Springer                  | Studi membandingkan channel pro- dan anti-NATO. Komentator anti-NATO menunjukkan lebih banyak toxic score, sementara pro-NATO juga menyertakan komentar positif. Penggunaan LDA dan word cloud mendetailkan bahwa komentar anti sering diarahkan pada “Fake News” dan “Profanity”. Temuan memperkuat pentingnya memahami konteks ideologis dalam riset toxicity <sup>103</sup> . |
| 12 | <i>An approach to counter online toxicity and enhance digital discourse</i> | 2024 | Procedia Computer Science | Alamsyah menyoroti prevalensi komentar toxic pada media sosial Indonesia tahun-tahun terakhir. Studi menekankan bahwa strategi counter-speech dan digital literacy sangat penting untuk meningkatkan kualitas diskusi online serta menekan toxic discourse di platform publik <sup>104</sup> .                                                                                   |
| 13 | <i>Social-mediated</i>                                                      | 2025 | Frontiers in              | Analisis SNA terhadap                                                                                                                                                                                                                                                                                                                                                            |

<sup>102</sup> Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira Jr, “Auditing Radicalization Pathways on YouTube. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT),”

<sup>103</sup> Adewale Obadimu et al., *Identifying Toxicity within Youtube Video Comment, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11549 LNCS (Springer International Publishing, 2019), [https://doi.org/10.1007/978-3-030-21741-9\\_22](https://doi.org/10.1007/978-3-030-21741-9_22).

<sup>104</sup> Sagama, Y., & Alamsyah, “Multi-Label Classification of Indonesian Online Toxicity Using BERT and RoBERTa.”

|    |                                                                                                              |      |               |                                                                                                                                                                                                                                                                                                                     |
|----|--------------------------------------------------------------------------------------------------------------|------|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | <i>communication and network dynamics in online gambling discourse</i>                                       |      | Communication | komentar YouTube mengenai judi online Indonesia dengan 3.673 akun unik. Hasil menunjukkan interaksi minimal dan desentralisasi komunikasi; menunjukkan komentar lebih sebagai broadcast daripada diskursus deliberatif. Relevan dalam memahami struktur komunikasi publik di platform video publik <sup>105</sup> . |
| 14 | <i>Hate Speech on Social Media: Indonesian Netizens' Hate Comments of Presidential Talk Shows on YouTube</i> | 2024 | LLT Journal   | Analisis 315 komentar pemilu 2024 di talkshow "Mata Najwa." Empat tipe hate comment ditemukan: early warning (43%), dehumanisasi (21%), incitement (19%), dan offensive language (17%). Faktor anonim dan trait kepribadian juga diteridentifikasi sebagai pemicu hate comments publik <sup>106</sup> .             |
| 15 | <i>Indonesian Social Media Sentiment Analysis With Sarcasm Detection</i>                                     | 2015 | arXiv         | Studi sentiment analysis bahasa Indonesia menambah fitur deteksi sarkasme (negativity dan interjections). Hasil menunjukkan efektivitas fitur tambahan dalam mengklasifikasi sentimen sarkastik dalam komentar, berguna untuk                                                                                       |

<sup>105</sup> Nugroho and Irwansyah, "Konvergensi Konten Audio Di Media Online (Studi Kasus Podcast Detik.Com)."

<sup>106</sup> M. G. F. Tahir, I., & Ramadhan, "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), 72– (2024), <https://e-journal.usd.ac.id/index.php/LLT/article/view/8180>.

|    |                                                                                                                       |      |                             |                                                                                                                                                                                                                                                                                                                                         |
|----|-----------------------------------------------------------------------------------------------------------------------|------|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                                                       |      |                             | menghindari false-positive dalam deteksi toxicity <sup>107</sup> .                                                                                                                                                                                                                                                                      |
| 16 | <i>Long Short-Term Memory for Hate Speech and Abusive Language Detection</i>                                          | 2021 | WCSE Conference Proceedings | Dataset YouTube dijadikan testing terhadap model LSTM yang dilatih dari Twitter. Model mencapai akurasi yang baik dalam mendeteksi hate speech di YouTube meskipun domain berbeda, menunjukkan generalisasi model deep learning dari Twitter ke YouTube masih memadai <sup>108</sup> .                                                  |
| 17 | <i>Sentiment Analysis of YouTube Video Comments Review of Starlink Indonesia by Gadgetin using VADER and TextBlob</i> | 2025 | IJCS                        | Fokus pada komentar pengguna terhadap Starlink Indonesia. Meski bukan soal toxicity, analisis menemukan sentimen negatif sering disandingkan dengan kritik kasar dan skeptisisme terhadap teknologi asing. Hal ini menunjukkan bahwa platform baru menerima potensi toxic discourse terutama dalam isu politik-teknologi <sup>109</sup> |
| 18 | <i>Racism in Social Media: A Critical Discourse Analysis of YouTube Comments</i>                                      | 2022 | IJELTAL                     | Studi ini mengidentifikasi jenis-jenis rasisme dalam komentar YouTube melalui analisis kritis wacana: teks, interpretasi, dan konteks sosial. Temuan menunjukkan rasisme di YouTube                                                                                                                                                     |

<sup>107</sup> Edwin Lunando and Ayu Purwarianti, "Indonesian Social Media Sentiment Analysis with Sarcasm Detection," *2013 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2013*, 2013, 195–98, <https://doi.org/10.1109/ICACSIS.2013.6761575>.

<sup>108</sup> Obadimu et al., *Identifying Toxicity within Youtube Video Comment*, 2019.

<sup>109</sup> "Absharina, E. D., Haq, M. A., & Aditya, M. D. (2025). Sentiment Analysis of YouTube Video Comments Review: Starlink Indonesia. *International Journal of Computer Science (IJCS)*, 14(1), 45–57. <https://Ijcs.Net/Ijcs/Index.Php/Ijcs/Article/View/4686>," n.d.

|    |                                                                                                  |      |                             |                                                                                                                                                                                                                                                                                                                                                              |
|----|--------------------------------------------------------------------------------------------------|------|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                                  |      |                             | sering disampaikan melalui metafora dan asumsi implisit, bukan hanya ujaran eksplisit, sehingga perlu pendekatan analisis mendalam <sup>110</sup> .                                                                                                                                                                                                          |
| 19 | <i>Analyzing Toxicity in YouTube Comments with the Help of Machine Learning</i>                  | 2023 | Stockholm University Thesis | Sampel acak 79.577 komentar kanal Swedia dianalisis menggunakan Hatescan; hanya 0,643% komentar masuk kategori toxic. Komentar injurios terhadap kecerdasan/kompetensi paling dominan, politik cenderung umum. Temuan menyoroti pentingnya distribusi jenis toxic serta kesiapan kreator konten untuk menghadapi potensi digital harassment <sup>111</sup> . |
| 20 | <i>Abusive comment identification on Indonesian social media data using hybrid deep learning</i> | 2022 | IJ-AI Journal               | Model hybrid RNN+LSTM diuji pada dataset Twitter Bahasa Indonesia untuk mendeteksi ancaman, ujaran kebencian, dan hinaan. Model mencapai f-measure hingga 94% dan akurasi meningkat 23%. Relevan untuk menstruktur sistem deteksi otomatis komentar toxic bahasa Indonesia <sup>112</sup> .                                                                  |

<sup>110</sup> S. Shafa, "Racism in Social Media: A Critical Discourse Analysis of YouTube Comments.," *Indonesian Journal of English Language Teaching and Applied Linguistics (IJELTAL)*, no. 6(2), 145–162. (2022), <https://ijeltal.org/index.php/ijeltal/article/view/1307>.

<sup>111</sup> J. Dehkhoda, S., & Gunica, "Analyzing Toxicity in YouTube Comments with the Help of Machine Learning DiVA Portal." ([Bachelor's thesis, Stockholm University]., 2023), <https://www.diva-portal.org/smash/get/diva2:1784332/FULLTEXT01.pdf>.

<sup>112</sup> Sari et al., "Abusive Comment Identification on Indonesian Social Media Data Using Hybrid Deep Learning."

## F. Kajian Teori

### 1. Online Disinhibition Effect Theory

Teori Disinhibisi Online (*Online Disinhibition Effect Theory*) menjelaskan fenomena di mana individu cenderung berperilaku lebih bebas, ekstrem, atau agresif di dunia maya dibandingkan dengan interaksi tatap muka. Fenomena ini terutama disebabkan oleh beberapa faktor, seperti anonimitas, jarak fisik, dan kurangnya konsekuensi langsung di platform online <sup>113</sup>. Dalam konteks penelitian ini, teori ini sangat relevan untuk memahami mengapa komentar toxic, termasuk penghinaan, ancaman, dan serangan identitas, sering kali muncul di media sosial seperti TikTok. Pengguna mungkin merasa lebih bebas untuk mengekspresikan pendapat atau emosi mereka tanpa khawatir akan dampak langsung dari tindakan mereka, sehingga memicu perilaku yang lebih ekstrem dan kurang bertanggung jawab.

Anonimitas yang disediakan oleh platform media sosial memungkinkan pengguna untuk berkomentar tanpa mengungkapkan identitas asli mereka. Hal ini sering kali menghilangkan rasa tanggung jawab pribadi atas komentar yang diungkapkan. Sebagai contoh, dalam video pemberantasan narkoba yang diunggah di akun TikTok @manangsoebeti\_official, pengguna yang mungkin tidak akan mengungkapkan pendapat kasar atau agresif dalam situasi tatap muka, merasa lebih nyaman untuk melakukannya secara online karena mereka dapat bersembunyi di balik nama samaran atau akun palsu <sup>114</sup>. Anonimitas ini menciptakan lingkungan di mana komentar toxic dapat

---

<sup>113</sup> Suler, "The Online Disinhibition Effect."

<sup>114</sup> Lapidot-Lefler and Barak, "The Benign Online Disinhibition Effect: Could Situational Factors Induce Self-Disclosure and Prosocial Behaviors?"

berkembang, merusak suasana diskusi dan mengganggu keberhasilan kampanye.

Selain anonimitas, jarak fisik yang diciptakan oleh komunikasi online juga memainkan peran penting dalam efek disinhibisi. Pengguna yang tidak berada dalam jarak dekat dengan orang lain merasa lebih terisolasi dari dampak langsung dari tindakan mereka. Dalam konteks kampanye pemberantasan narkoba, komentar-komentar yang kasar atau mengancam mungkin ditulis tanpa mempertimbangkan dampaknya terhadap individu yang terlibat langsung dalam kampanye tersebut atau masyarakat yang terpapar informasi tersebut<sup>115</sup> Jarak fisik ini membuat pengguna merasa lebih aman untuk mengekspresikan perasaan negatif mereka tanpa takut akan reaksi fisik atau sosial yang langsung.

Kurangnya konsekuensi langsung di platform online juga mendorong pengguna untuk lebih bebas dalam mengekspresikan diri, sering kali dengan cara yang tidak pantas atau bahkan berbahaya. Di platform seperti TikTok, pengguna yang memposting komentar toxic mungkin merasa tidak ada dampak langsung dari perilaku mereka, terutama jika platform tersebut tidak memiliki sistem moderasi yang ketat atau jika komentar mereka tidak segera ditindaklanjuti. Hal ini memungkinkan terjadinya eskalasi komentar toxic, yang tidak hanya merusak diskusi tetapi juga dapat mengancam keberhasilan kampanye sosial seperti pemberantasan narkoba<sup>116</sup>.

---

<sup>115</sup> S. S. Sundar, *Social Psychology of Interactivity in Human–Website Interaction*. In A. N. Joinson et Al. (Eds.), *The Oxford Handbook of Internet Psychology*, 2007.

<sup>116</sup> Lapidot-Lefler and Barak, “The Benign Online Disinhibition Effect: Could Situational Factors Induce Self-Disclosure and Prosocial Behaviors?”

Dalam artikel “*Online Disinhibition Effect: Kenapa Ya, Beberapa Orang Menjadi Berbeda Saat Di Dunia Maya?*” yang dipublikasikan oleh [binus.ac.id](http://binus.ac.id) mengatakan:

**Tabel 2 Online Disinhibition Effect**

| No | Unsur                                       | Penjelasan                                                                                                                                                                                                                                                                                                                         |
|----|---------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | <i>Dissociative anonymity</i>               | Seseorang merasa aman akan terungkapnya identitas asli diri mereka di dunia maya, sehingga mereka merasa memiliki kesempatan untuk menunjukkan perilaku yang berbeda dan memisahkan diri mereka di kehidupan nyata dengan diri mereka di dunia maya.                                                                               |
| 2  | <i>Invisibility</i>                         | Saat berada di dunia maya, tidak terjadi interaksi secara langsung antarindividu sehingga seseorang akan merasa bahwa ia invisible atau tidak terlihat oleh orang lain yang menjadi lawan bicaranya. Invisibilitas inilah yang membuat seseorang berani untuk melakukan hal-hal yang biasanya tidak mereka lakukan di dunia nyata. |
| 3  | <i>Asynchronity</i>                         | Percakapan di dunia maya tidak selalu sinkron dalam waktu. Bisa jadi ada jeda beberapa menit atau jam untuk seseorang membalas percakapan. Hal inilah yang membuat seseorang dapat mengubah arah berpikirnya.                                                                                                                      |
| 4  | <i>Solipsistic introjection</i>             | Interaksi yang tidak terjadi secara nyata di dunia maya membuat seseorang mengimajinasikan bagaimana wujud lawan bicaranya menyampaikan pesan tersebut secara langsung.                                                                                                                                                            |
| 5  | <i>Dissociative imagination</i>             | Seseorang akan berimajinasi dirinya yang di dunia maya berbeda dengan dirinya yang berada di dunia nyata, sehingga norma yang diterapkan di dunia maya juga berbeda dengan yang diterapkan di dunia nyata.                                                                                                                         |
| 6  | <i>Minimization of status and authority</i> | Saat berada di dunia maya, perbedaan akan ciri fisik, lingkungan, maupun status sosial seseorang menjadi minim atau bahkan tidak terlihat karena anonimitas yang ada. Oleh karena itu, semua orang memiliki kesempatan yang sama untuk dapat mengekspresikan diri.                                                                 |

Sumber: [Binus.ac.id](http://Binus.ac.id), 2023

Dalam kesimpulannya, Teori Disinhibisi Online menawarkan kerangka yang kuat untuk memahami dinamika di balik munculnya komentar toxic di media sosial. Faktor-faktor seperti anonimitas, jarak fisik, dan kurangnya konsekuensi langsung memberikan penjelasan mengapa pengguna merasa lebih nyaman untuk mengekspresikan perilaku negatif mereka secara online. Memahami teori ini dapat membantu dalam merancang strategi moderasi yang lebih efektif, serta meningkatkan kesadaran pengguna tentang dampak dari komentar toxic mereka di ruang digital, khususnya dalam kampanye-kampanye yang sensitif seperti pemberantasan narkoba.

## 2. Uses And Gratifications Theory (UGT)

*Uses and Gratifications Theory* (UGT) merupakan salah satu pendekatan penting dalam studi komunikasi massa yang menempatkan audiens sebagai agen aktif dalam proses komunikasi. Teori ini awalnya dikembangkan oleh Katz, Blumler, dan Gurevitch, yang menyatakan bahwa individu menggunakan media secara sengaja untuk memenuhi kebutuhan dan keinginannya. Dalam konteks ini, media tidak lagi dilihat sebagai entitas yang memiliki kekuatan mutlak atas audiens, melainkan sebagai alat yang dimanfaatkan sesuai dengan motivasi pengguna<sup>117</sup>.

UGT mengkategorikan berbagai motif penggunaan media menjadi beberapa tipe kebutuhan, seperti kebutuhan kognitif (mencari informasi), afektif (hiburan), personal *integrative* (memperkuat identitas diri), social *integrative* (menjalin relasi sosial), dan tension release (melarikan diri dari

---

<sup>117</sup> E. Blumler, J. G., & Katz, *The Uses of Mass Communications: Current Perspectives on Gratifications Research*. (Sage Publications., 1974).

tekanan)<sup>118</sup> Dalam era media digital, kelima kategori ini semakin terlihat jelas karena pengguna dapat berinteraksi langsung dengan konten dan komunitasnya. Hal ini memberikan peluang bagi individu untuk tidak hanya mengonsumsi, tetapi juga memproduksi dan mendistribusikan opini pribadi, termasuk dalam bentuk komentar.

Perkembangan media sosial dan platform digital seperti YouTube memperluas aplikasi UGT dalam memahami perilaku pengguna internet. Audiens tidak hanya mencari hiburan atau informasi, tetapi juga membangun narasi, menunjukkan eksistensi, dan memengaruhi opini publik melalui komentar atau konten buatan sendiri.<sup>119</sup> Dalam konteks ini, komentar yang ditinggalkan pada video yang memuat isu kontroversial atau tokoh publik sering kali menjadi bentuk aktualisasi diri maupun ekspresi afiliasi kelompok tertentu.

Selain itu, UGT membantu memahami mengapa komentar toksik muncul dalam platform digital. Komentar negatif atau provokatif sering kali muncul sebagai bentuk pelepasan emosi atau manifestasi identitas sosial yang ingin ditonjolkan kepada publik digital.<sup>120</sup> Oleh karena itu, UGT dapat menjadi alat analisis yang efektif untuk mengidentifikasi motif psikologis dan sosial dari perilaku daring pengguna, termasuk komentar yang penuh muatan ideologis, religius, atau politis.

---

<sup>118</sup> “Katz, E., Blumler, J. G., & Gurevitch, M. (1974). Utilization of Mass Communication by the Individual. In J. G. Blumler & E. Katz (Eds.), *The Uses of Mass Communications: Current Perspectives on Gratifications Research* (Pp. 19–32). Sage.,” n.d.

<sup>119</sup> Sundar, S. S., & Limperos, “Uses and Grats 2.0: New Gratifications for New Media.”

<sup>120</sup> J. Shao, G., & Wang, “Understanding User-Generated Content in Social Media: A Literature Review,” *Internet Research*, no. 27(4), 854–874. (2017), <https://doi.org/10.1108/IntR-07-2016-0201>.

Dalam kajian komunikasi penyiaran Islam, UGT dapat digunakan untuk memahami bagaimana audiens mengonsumsi dan merespons konten dakwah atau diskusi keagamaan di media sosial. Ketika figur seperti Ustadz Abdul Somad hadir dalam ruang digital bersama tokoh seperti Rocky Gerung, komentar-komentar warganet menjadi representasi dari kebutuhan ekspresif, afektif, dan identitas keagamaan yang mereka ingin tampilkan. UGT memberikan lensa untuk melihat bagaimana ruang digital bukan sekadar tempat konsumsi pesan, tetapi juga arena kontestasi makna yang sangat dipengaruhi oleh motivasi pengguna.

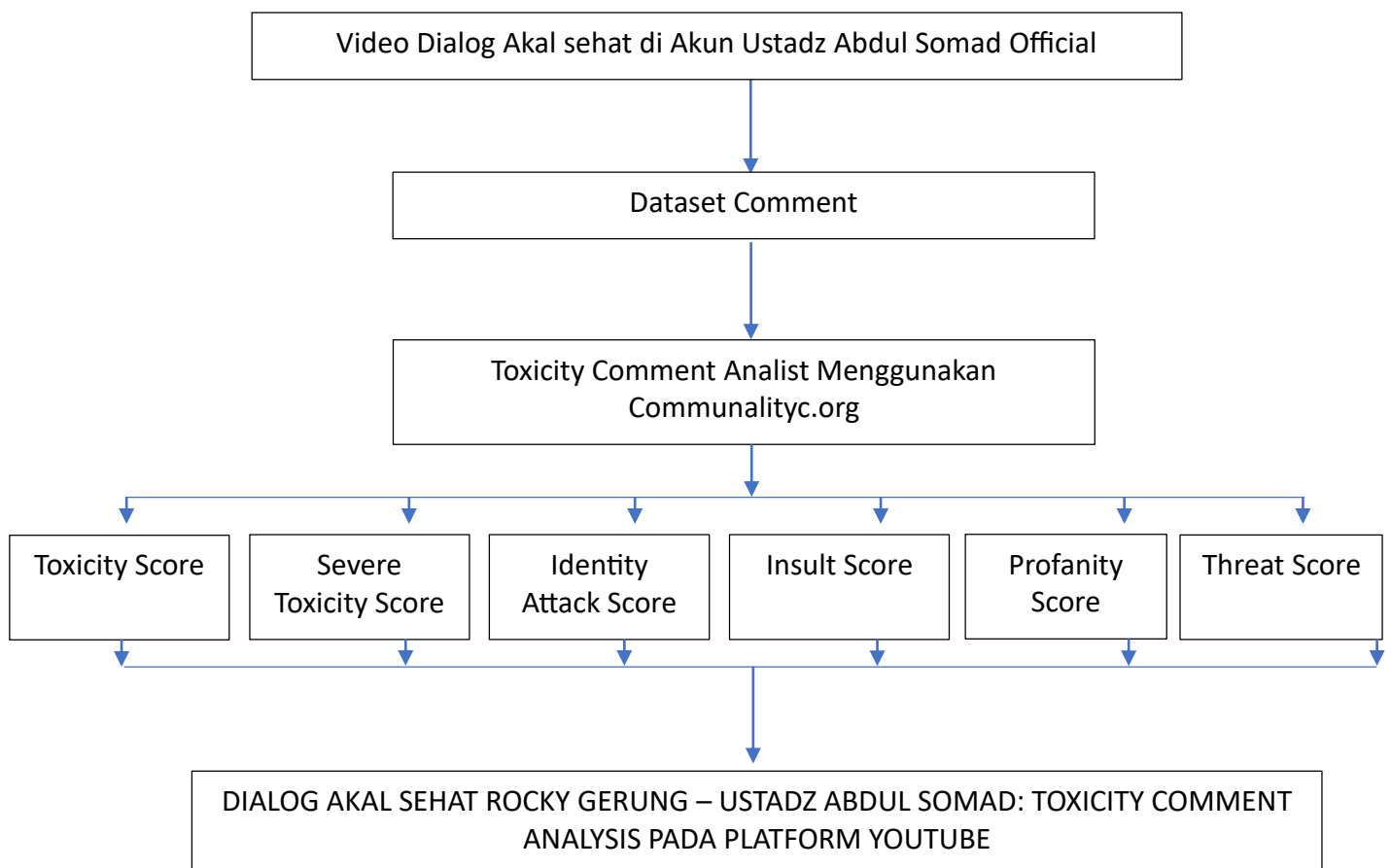
### **3. Kerangka Berpikir**

Kerangka pemikiran dalam penelitian ini berfungsi sebagai landasan konseptual yang menghubungkan antara fenomena yang diamati, teori-teori yang digunakan, dan tujuan penelitian. Dalam konteks ini, komentar toksik yang muncul dalam video dialog antara Rocky Gerung dan Ustadz Abdul Somad di platform YouTube tidak hanya dipahami sebagai ekspresi bebas pengguna, tetapi juga sebagai bentuk perilaku komunikasi yang dipengaruhi oleh berbagai faktor psikologis, sosial, dan teknologis.

Penelitian ini memadukan Teori Disinhibisi Online (Online Disinhibition Effect) dari John Suler untuk menjelaskan mengapa pengguna cenderung menunjukkan perilaku agresif di ruang digital. Efek disinhibisi, seperti anonimitas dan jarak psikologis, memberi kontribusi besar terhadap kemunculan komentar bernuansa negatif, seperti ujaran kebencian, hinaan, dan ancaman.

Melalui teori tersebut, penelitian ini membangun kerangka berpikir bahwa komentar toksik bukan hanya hasil dari opini spontan, tetapi merupakan produk dari interaksi kompleks antara struktur platform, kecenderungan psikologis pengguna, dan dinamika sosial-politik dalam ruang publik digital. Dengan menganalisis dimensi toksisitas seperti insult, hate speech, identity attack, dan threat, serta faktor penyebab dan dampaknya, penelitian ini diharapkan mampu memberikan gambaran menyeluruh tentang bagaimana komentar-komentar toksik terbentuk, menyebar, dan memengaruhi kualitas diskusi publik di media sosial.

**Gambar 1 Bagan Kerangka Pemikiran Penelitian**



## **G. Metodologi Penelitian**

### **1. Jenis Penelitian**

Penelitian ini menggunakan pendekatan kualitatif dengan metode analisis isi (content analysis) yang bersifat deskriptif. Analisis isi dipilih untuk mengkaji secara mendalam makna dan kecenderungan yang terkandung dalam komentar-komentar warganet yang muncul dalam video dialog antara Rocky Gerung dan Ustadz Abdul Somad di platform YouTube. Pendekatan ini memungkinkan peneliti untuk mengidentifikasi pola-pola toksisitas, kategori ujaran, serta implikasinya terhadap ruang publik digital.

Jenis penelitian ini adalah studi kasus, karena objek kajiannya difokuskan pada satu fenomena spesifik, yaitu diskusi warganet dalam satu video YouTube yang memuat interaksi antara dua tokoh publik dengan latar belakang ideologi dan religiusitas yang berbeda (Creswell, 2016).

### **2. Fokus Penelitian**

Penelitian ini secara utama memanfaatkan Online Disinhibition Effect Theory dan Uses and Gratifications Theory (UGT) sebagai landasan teoretis untuk menganalisis perilaku komentar pengguna di media sosial, khususnya komentar toxic pada platform digital. Pemilihan kedua teori ini didasarkan pada relevansinya dalam menjelaskan dinamika perilaku pengguna internet yang bersifat aktif, ekspresif, serta cenderung tidak terkontrol dalam ruang komunikasi daring.

Online Disinhibition Effect Theory digunakan untuk memfokuskan analisis pada faktor-faktor psikologis dan situasional yang mendorong munculnya komentar toxic di media sosial. Fokus penelitian diarahkan pada bagaimana unsur anonimitas, invisibilitas, jarak fisik, serta minimnya konsekuensi langsung memengaruhi keberanian pengguna dalam mengekspresikan ujaran negatif, agresif, atau menyerang identitas pihak tertentu. Teori ini menjadi kerangka utama untuk memahami mengapa perilaku yang secara sosial tidak dapat diterima di dunia nyata justru muncul secara masif di ruang digital.

Sementara itu, Uses and Gratifications Theory (UGT) digunakan untuk menelaah motif dan kebutuhan yang mendorong pengguna dalam memproduksi komentar, baik yang bersifat netral maupun toksik. Fokus penelitian melalui UGT diarahkan pada identifikasi kebutuhan psikologis dan sosial pengguna, seperti kebutuhan ekspresi diri, pelepasan emosi, pencarian identitas kelompok, serta afiliasi ideologis atau keagamaan. Dengan demikian, komentar tidak hanya dipahami sebagai reaksi spontan terhadap konten, tetapi juga sebagai bentuk pemenuhan kebutuhan personal dan sosial pengguna media.

Integrasi kedua teori ini memungkinkan penelitian untuk melihat fenomena komentar toxic secara lebih komprehensif. Online Disinhibition Effect Theory menjelaskan *mengapa* pengguna berani mengekspresikan perilaku negatif di ruang digital, sedangkan UGT menjelaskan *untuk tujuan apa* dan *kebutuhan apa* yang ingin dipenuhi melalui perilaku

tersebut. Fokus penelitian ini tidak hanya menitikberatkan pada isi komentar, tetapi juga pada kondisi psikologis, motivasi, dan konteks sosial yang melatarbelakangi munculnya komentar tersebut.

Dengan menggunakan kedua teori ini, penelitian diharapkan mampu memberikan pemahaman yang lebih mendalam mengenai dinamika komunikasi digital, khususnya dalam ruang diskusi yang sensitif seperti isu sosial, keagamaan, dan kampanye publik. Selain itu, hasil analisis diharapkan dapat menjadi dasar dalam merumuskan strategi literasi digital dan moderasi konten yang lebih efektif di media sosial.

### **3. Teknik Pengumpulan Data**

Teknik pengumpulan data dalam penelitian ini menggunakan observasi dan dokumentasi. Kedua teknik tersebut dipilih karena penelitian berfokus pada pengamatan perilaku komunikasi warganet melalui komentar publik di platform YouTube, tanpa melibatkan interaksi langsung dengan subjek penelitian.

#### **a. Observasi**

Observasi dalam penelitian ini dilakukan dalam bentuk observasi nonpartisipan secara daring. Peneliti tidak terlibat langsung dalam interaksi dengan warganet, melainkan hanya mengamati dan merekam data komentar yang telah dipublikasikan secara terbuka pada kolom komentar video YouTube yang menjadi objek penelitian.

Observasi dilakukan dengan cara mengamati pola bahasa, bentuk ekspresi, serta kecenderungan isi komentar warganet terhadap video

dialog antara Rocky Gerung dan Ustadz Abdul Somad. Untuk mendukung efektivitas observasi dan memastikan data terekam secara sistematis, peneliti menggunakan alat bantu pengumpulan data berbasis web, yaitu Communalytic.org dan YouTube API, yang memungkinkan peneliti merekam komentar dalam jumlah besar tanpa mengubah konteks alami interaksi warganet.

Observasi ini bertujuan untuk memperoleh gambaran empiris mengenai fenomena komentar toksik yang muncul dalam ruang diskusi publik di YouTube.

#### **b. Dokumentasi**

Teknik dokumentasi digunakan untuk mengumpulkan dan menyimpan dokumen digital yang berkaitan dengan objek penelitian.

Dokumen yang dikumpulkan dalam penelitian ini meliputi:

Arsip komentar publik warganet pada video YouTube yang diteliti;

Metadata video, seperti judul, jumlah penonton, jumlah komentar, dan waktu unggah; Data hasil ekstraksi komentar yang disimpan dalam format CSV dan JSON

Dokumentasi dilakukan dengan cara mengunduh dan mengarsipkan data komentar menggunakan fasilitas scraping pada platform Communalytic.org. Seluruh dokumen yang dikumpulkan bersifat terbuka dan tidak memuat identitas pribadi pengguna, sehingga tetap memenuhi prinsip etika penelitian.

Data dokumentasi ini kemudian digunakan sebagai bahan utama dalam proses analisis isi dan pemetaan tingkat toksisitas komentar pada tahap analisis data.

#### 4. Teknik Validasi Data

Dalam penelitian ini, peneliti menggunakan triangulasi sumber data dan triangulasi teknik untuk memvalidasi data yang telah dikumpulkan dari platform YouTube guna memastikan keabsahan dan kredibilitas hasil penelitian. Validasi data dilakukan dengan membandingkan berbagai sumber dan teknik analisis agar data komentar yang diteliti benar-benar merepresentasikan fenomena komunikasi toksik dalam ruang diskusi publik digital.

Triangulasi sumber data dilakukan dengan membandingkan komentar-komentar yang berasal dari berbagai akun pengguna yang berbeda pada video *Dialog Akal Sehat Rocky Gerung – Ustadz Abdul Somad*. Keberagaman latar belakang pengguna dan variasi isi komentar memungkinkan peneliti memperoleh gambaran yang lebih komprehensif mengenai pola toksisitas, sehingga meminimalkan bias yang bersumber dari dominasi kelompok atau individu tertentu.

Selanjutnya, triangulasi teknik dilakukan dengan membandingkan hasil observasi nonpartisipan terhadap komentar warganet, dokumentasi arsip komentar dan metadata video, **serta** hasil klasifikasi tingkat toksisitas menggunakan alat bantu Perspective API. Perbandingan ini bertujuan untuk memastikan konsistensi antara

temuan kualitatif peneliti dan hasil pengukuran kuantitatif berbasis sistem analitik.

Selain itu, validasi data juga dilakukan melalui pengecekan ulang (verification) terhadap komentar yang telah dikategorikan. Komentar yang memiliki makna ambigu atau berpotensi multitafsir dianalisis kembali dengan mempertimbangkan konteks diskusi dan alur percakapan (thread komentar) agar tidak terjadi kesalahan interpretasi.

Dengan penerapan triangulasi sumber dan teknik serta proses verifikasi data, penelitian ini diharapkan mampu menghasilkan temuan yang valid, reliabel, dan dapat dipertanggungjawabkan secara ilmiah dalam menjelaskan fenomena komentar toksik pada platform YouTube, khususnya dalam konteks diskusi publik yang melibatkan tokoh dengan latar belakang pemikiran dan keagamaan yang berbeda.

## 5. Teknik Analisis Data

Setelah data komentar dari video YouTube berhasil dikumpulkan, tahap selanjutnya adalah proses analisis untuk mengidentifikasi pola-pola toksisitas yang muncul dalam interaksi warganet. Analisis dilakukan dengan pendekatan kualitatif berbasis analisis isi, yang bertujuan menggali makna dan kecenderungan diskursif dalam komentar-komentar tersebut. Untuk mendukung keakuratan klasifikasi dan meningkatkan objektivitas penilaian, analisis ini juga diperkuat dengan bantuan teknologi berbasis pembelajaran mesin, seperti *Perspective API*. Pendekatan ini

memungkinkan peneliti mengombinasikan interpretasi kontekstual secara manusiawi dengan pemetaan metrik toksisitas secara sistematis.

#### **a. Analisis Teks**

Komentar-komentar yang telah dikumpulkan akan dianalisis menggunakan pendekatan analisis isi kualitatif. Data akan dikodekan dan dikategorikan ke dalam:

**Tabel 3 Toxicity Comment Element**

| <b>No</b> | <b>Unsur Toxic</b>    | <b>Keterangan</b>                                                                                                                                                                                                   |
|-----------|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1         | Toxicity Score        | Metrik ini mengukur seberapa kasar, tidak sopan, atau tidak masuk akal nya sebuah komentar. Ini membantu dalam mengidentifikasi komentar yang dapat merusak suasana diskusi dan dinamika sosial di komunitas online |
| 2         | Severe Toxicity Score | Metrik ini mengukur seberapa besar komentar mengandung kebencian, agresivitas, dan ketidakhormatan . Ini berguna untuk mengidentifikasi komentar yang berbahaya                                                     |

|   |                       |                                                                                                                                                                                                                       |
|---|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |                       | karena mengandung elemen kebencian yang kuat.                                                                                                                                                                         |
| 3 | Identity Attack Score | Metrik ini menunjukkan apakah sebuah komentar mengandung bahasa kebencian yang menargetkan identitas seseorang, seperti ras atau agama. Ini penting untuk memahami penyebaran komentar diskriminatif di media sosial. |
| 4 | Insult Score          | Metrik ini membantu mengidentifikasi komentar yang bersifat menghina atau provokatif. Ini memberikan wawasan tentang frekuensi penghinaan dalam diskusi online dan dampaknya terhadap interaksi sosial.               |
| 5 | Profanity Score       | Metrik ini menunjukkan apakah sebuah komentar mengandung kata-kata kasar atau tidak                                                                                                                                   |

|   |              |                                                                                                                                                                                                  |
|---|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |              | senonoh. Ini penting untuk menjaga standar komunikasi yang sehat di media sosial.                                                                                                                |
| 6 | Threat Score | Metrik ini mengukur sejauh mana sebuah komentar mengandung ancaman kekerasan atau niat menyakiti. Ini penting untuk menjaga keamanan komunitas online dan melindungi pengguna dari bahaya nyata. |

#### **b. Pemanfaatan Alat Analisis**

Untuk mendukung validitas dan objektivitas dalam proses klasifikasi komentar toksik, penelitian ini memanfaatkan alat bantu berbasis machine learning, yaitu Perspective API yang dikembangkan oleh Jigsaw (Google) dan platform analitik sosial Communalytic.org. Kedua alat ini digunakan secara komplementer untuk menganalisis komentar dalam jumlah besar secara efisien, serta memberikan klasifikasi toksisitas secara kuantitatif berdasarkan parameter yang telah diuji secara ilmiah.

Selain itu, penelitian ini juga menggunakan Communalytic.org, sebuah platform riset sosial berbasis data yang dikembangkan oleh *Social Media Lab di Toronto Metropolitan*

*University*. Communalytic memungkinkan peneliti melakukan scraping komentar YouTube, menganalisis data sosial dalam format CSV, serta mengintegrasikannya secara otomatis dengan *Perspective API* untuk menghasilkan skor toksisitas yang dapat diekspor dan divisualisasikan. Platform ini juga menyediakan fitur analisis lanjutan, seperti identifikasi thread komentar, intensitas per kategori toksisitas, serta perbandingan distribusi komentar toksik antar waktu dan topik.

Dengan mengombinasikan dua alat ini, penelitian dapat memetakan komentar-komentar toksik secara lebih sistematis, baik dari sisi kuantitas (jumlah dan persentase) maupun kualitas (tingkat intensitas dan kategori toksisitas). Hasil analisis ini selanjutnya digunakan untuk menampilkan visualisasi berupa *word cloud*, grafik distribusi skor, dan heatmap korelasi yang memperjelas pola diskusi publik di platform YouTube. Pendekatan ini tidak hanya meningkatkan akurasi klasifikasi, tetapi juga memperkuat reliabilitas metodologis dalam menjelaskan fenomena komunikasi toksik di ruang digital.

## **H. Sistematika Pembahasan**

Untuk memberikan pemahaman yang utuh dan sistematis terhadap topik yang diteliti, penulisan tesis ini disusun dalam beberapa bab yang saling berkesinambungan. Setiap bab disusun secara logis untuk menjelaskan proses penelitian mulai dari latar belakang, landasan teori, metodologi, hingga hasil

analisis dan kesimpulan. Adapun sistematika pembahasan dalam penelitian ini adalah sebagai berikut:

**BAB I: PENDAHULUAN,** Bab ini berisi uraian mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat dan kegunaan penelitian, serta sistematika pembahasan. Bagian ini memberikan fondasi awal atas pentingnya penelitian dan arah yang akan dituju dalam proses analisis.

**BAB II: TINJAUAN PUSTAKA,** Bab ini memuat landasan teori yang relevan, mencakup teori komunikasi digital, media sosial dan YouTube sebagai ruang publik, konsep komentar toksik, serta pendekatan analisis teks dan pembelajaran mesin dalam studi komunikasi. Selain itu, bab ini juga menguraikan hasil-hasil penelitian terdahulu (literature review) dan menunjukkan adanya research gap yang ingin diisi oleh penelitian ini.

**BAB III: METODOLOGI PENELITIAN,** Bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk pendekatan penelitian, jenis dan sumber data, teknik pengumpulan data (web scraping), teknik analisis data menggunakan tools analisis komentar (seperti Perspective API atau CommuNalytic), serta tahapan klasifikasi komentar toksik berdasarkan kategori seperti hate speech, insult, identity attack, dan threat.

**BAB IV: HASIL PENELITIAN DAN PEMBAHASAN,** Bab ini menyajikan hasil analisis data berupa distribusi komentar toksik, jenis-jenis komentar dominan, serta pola-pola tertentu yang ditemukan dalam respons warganet terhadap dialog antara Rocky Gerung dan Ustadz Abdul Somad.

Pembahasan dilakukan secara mendalam dan dikaitkan dengan teori yang telah dibahas sebelumnya.

**BAB V: PENUTUP,** Bab ini memuat kesimpulan dari hasil penelitian dan saran yang dapat diberikan kepada pihak-pihak terkait, baik akademisi, pemerintah, pengelola platform digital, maupun masyarakat umum. Saran juga diberikan untuk penelitian lanjutan di masa mendatang.

## **BAB IV**

### **PENUTUP**

#### **A. Kesimpulan**

Penelitian ini menunjukkan bahwa ruang komentar dalam video “Dialog Akal Sehat” merupakan arena diskursus publik yang dinamis dan kompleks, yang mencerminkan respons emosional, ideologis, dan sosial dari para pengguna YouTube. Melalui analisis 2.921 komentar yang berhasil diproses, ditemukan bahwa mayoritas komentar memiliki nada positif, ditandai dengan dominasi emoji hati, tertawa, dan doa, serta kata-kata seperti “sehat”, “akal”, “cerdas”, “Allah”, dan “UAS”. Wordcloud dan emoji cloud memperlihatkan kuatnya dukungan terhadap narasi akal sehat, tokoh-tokoh dalam video, dan harapan terhadap masa depan bangsa.

Distribusi skor toksisitas mengungkap bahwa komentar secara umum tidak menunjukkan tingkat kebencian atau kekerasan verbal yang ekstrem. Indikator seperti Toxicity dan Insult memiliki skor rata-rata yang lebih tinggi dibandingkan kategori lain, namun masih dalam rentang yang relatif moderat. Threat dan Severe Toxicity merupakan kategori dengan skor terendah, memperlihatkan bahwa komentar eksplisit yang mengandung ancaman atau ujaran kebencian ekstrem sangat jarang ditemukan.

Grafik tren bulanan menunjukkan adanya fluktuasi skor toksisitas yang berkorelasi dengan dinamika sosial-politik nasional, terutama pada periode menjelang Pemilu 2024. Selain itu, pola waktu interaksi memperlihatkan

bahwa komentar terbanyak muncul dalam 3 minggu pertama setelah video diunggah, menandakan respons publik yang kuat namun tidak berkelanjutan.

Akhirnya, ditemukan bahwa aktivitas diskusi sangat terpusat pada beberapa akun pengguna saja. Grafik “Top Ten Posters” menunjukkan bahwa sebagian besar komentar ditulis oleh sekelompok kecil akun, salah satunya menyumbang lebih dari 35% dari total komentar. Hal ini mengindikasikan adanya potensi dominasi wacana oleh aktor-aktor tertentu, yang bisa organik atau terorganisasi.

## **B. Saran**

1. **Untuk Peneliti Komunikasi Digital**, penting untuk terus memantau dinamika diskusi di platform media sosial, tidak hanya dari sisi isi dan nada pesan, tetapi juga dari pola keterlibatan aktor-aktor dominan. Penelitian lanjutan dapat mengkaji konten komentar akun-akun paling aktif untuk melihat apakah mereka menyebarkan opini secara organik atau menjadi bagian dari ekosistem buzzer.
2. **Untuk Pengelola Platform seperti YouTube**, disarankan untuk memperkuat sistem deteksi dan moderasi komentar yang mengandung potensi toksisitas tinggi, sembari tetap menjaga kebebasan berekspresi. Peningkatan transparansi terhadap algoritma rekomendasi dan pengelolaan komentar sangat dibutuhkan.
3. **Untuk Tokoh Publik dan Intelektual**, seperti UAS dan Rocky Gerung, video seperti “Dialog Akal Sehat” sebaiknya dijadikan ruang edukatif

berkelanjutan yang mendorong diskusi sehat. Perlu juga keterlibatan aktif dalam mempromosikan etika berkomentar kepada audiens mereka.

4. **Untuk Audiens Umum**, partisipasi dalam diskusi digital perlu disertai tanggung jawab etis dan kesadaran literasi digital. Publik perlu menghindari ujaran kebencian, hoaks, dan komentar yang merendahkan martabat orang lain, sekalipun dalam perbedaan pandangan.
5. **Untuk Peneliti Kebijakan**, hasil studi ini dapat digunakan untuk merancang regulasi dan strategi nasional dalam menghadapi toksisitas digital tanpa mengorbankan hak atas kebebasan berekspresi. Regulasi yang kontekstual dan edukatif lebih diutamakan ketimbang pendekatan represif.

## DAFTAR PUSTAKA

- Ito, M., Baumer, S., Bittanti, M., boyd, d., Cody, R., Herr-Stephenson, B., ... & Tripp, L. (2010). Hanging Out. Absharina, E. D., Haq, M. A., & Aditya, M. D. (2025). Sentiment Analysis of YouTube Video Comments Review: Starlink Indonesia. *International Journal of Computer Science (IJCS)*, 14(1), 45–57. <https://Ijcs.Net/Ijcs/Index.Php/Ijcs/Article/View/4686>.” n.d.
- “Ahmadizzan. (2022). Abusive Comment Identification on Indonesian Social Media Data Using Hybrid Deep Learning. *Academia.Edu*. [https://Www.Academia.Edu/112446212/Abusive\\_comment\\_identification\\_on\\_Indonesian\\_social\\_media\\_data\\_using\\_hybrid\\_deep\\_learning](https://Www.Academia.Edu/112446212/Abusive_comment_identification_on_Indonesian_social_media_data_using_hybrid_deep_learning),” n.d.
- “Alim, S., & Dharma, A. F. (2021). YouTube Sebagai Ruang Publik Alternatif Bagi Anak Muda. *Expose: Jurnal Ilmu Komunikasi*, 4(1), 1–10. <https://Doi.Org/10.33021/Expose.V4i1.1223>,” n.d.
- “Alshamrani, S., Abusnaina, A., Abuhamad, M., Nyang, D., & Mohaisen, D. (2021). Hate, Obscenity, and Insults: Measuring the Exposure of Children to Inappropriate Comments in YouTube. *ArXiv Preprint*. <https://Arxiv.Org/Abs/2103.09050>,” n.d.
- “Alshamrani, S., Abusnaina, A., Mohaisen, D. (2020). Hiding in Plain Sight: A Measurement and Analysis of Kids’ Exposure to Malicious URLs on YouTube. *ArXiv Preprint*. <https://Arxiv.Org/Abs/2009.07923>,” n.d.
- “Barlett, C. P. (2015). Anonymously Hurting Others Online: The Effect of Anonymity on Cyberbullying Frequency. *Psychology of Popular Media Culture*, 4(2), 70–79. <https://Doi.Org/10.1037/Ppm0000013>,” n.d.
- Baudrillard, J. *Simulacra and Simulation*. University of Michigan Press., 1995.
- Bertaglia, T., Goanta, C., & Iamnitchi, A. “The Monetisation of Toxicity: Analysing YouTube Content Creators and Controversy-Driven Engagement,” no. *arXiv preprint arXi:2408.00534*. (2024). <https://arxiv.org/abs/2408.00534>.
- Blumler, J. G., & Katz, E. *The Uses of Mass Communications: Current Perspectives on Gratifications Research*. Sage Publications., 1974.
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. “Trolls Just Want to Have Fun. *Personality and Individual Differences*” 67, 97–102 (2014).
- Burgess, J., & Green, J. *YouTube: Online Video and Participatory Culture*. Polity Press., 2009.

- Castells, M. *The Rise of the Network Society (2nd Ed.)*. Wiley-Blackwell., 2011.
- Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. *Antisocial Behavior in Online Discussion Communities. Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM)*, 2017.
- Chou, H. T. G., & Edge, N. “‘They Are Happier and Having Better Lives than I Am’: The Impact of Using Facebook on Perceptions of Others’ Lives. *Cyberpsychology, Behavior, and Social Networking*” 15(2), 117 (2012). <https://doi.org/10.1089/cyber.2011.0324>.
- CNN Indonesia. “Youtube Jadi Raja Media Sosial Di Indonesia, Diakses 94 Persen Warga.” *CNN/Teknologi Internet*, 2022. <https://www.cnnindonesia.com/teknologi/20220610164924-192-807472/youtube-jadi-raja-media-sosial-di-indonesia-diakses-94-persen-warga>.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. *Introduction to Algorithms (3rd Ed.)*. MIT Press., 2009.
- Cormode, G., & Krishnamurthy, B. “Key Differences between Web 1.0 and Web 2.0.” *First Monday* 13(6). (2008). <https://doi.org/10.5210/fm.v13i6.2125>.
- Covington, P., Adams, J., & Sargin, E. “Deep Neural Networks for YouTube Recommendations.” *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016, 191–198. <https://doi.org/10.1145/2959100.2959190>.
- Darmajaya. “Deteksi Kalimat Toxic Dalam Postingan Media Sosial Menggunakan Algoritma Decision Tree.” 2022. <https://repo.darmajaya.ac.id/15460/>.
- Dehkhoda, S., & Gunica, J. “Analyzing Toxicity in YouTube Comments with the Help of Machine Learning DiVA Portal.” [Bachelor’s thesis, Stockholm University]., 2023. <https://www.diva-portal.org/smash/get/diva2:1784332/FULLTEXT01.pdf>.
- Dewan Kesenian Jakarta. “Pidato Kebudayaan Akhir Tahun: Memelihara Republik, Mengaktifkan Akal Sehat. Taman Ismail Marzuki.” 2010.
- Dina Nisa Aulia, and Dunan Amri. “Konvergensi Media Dan New Media Dalam Transformasi” 6, no. 1 (2025): 48–63.
- Djoko Purwanto. *Komunikasi Bisnis*. Penerbit Erlangga, 2003.
- Figueiredo, F., Almeida, J. M., Gonçalves, M. A., & Benevenuto, F. “On the Dynamics of Social Media Popularity: A YouTube Case Study.” *ACM Transactions on Internet Technology*, no. 14(4), 1–23. (2014). <https://doi.org/10.1145/2633121>.

- Gillespie, T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press., 2018.
- Google Transparency Report. "YouTube Community Guidelines Enforcement.," 2023. <https://transparencyreport.google.com/youtube-policy/removals>.
- Griffith University. "The Impact of YouTube on Mental Health.," 2023. <https://news.griffith.edu.au/2023/10/10/the-impact-of-youtube-on-mental-health/>.
- Habermas, J. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. MIT Press., 1989.
- Hanscom, R., Lehman, T. S., Lv, Q., & Mishra, S. "The Toxicity Phenomenon Across Social Media." *ArXiv Preprint ArXiv:2410.21589*, 2023. <https://arxiv.org/abs/2410.21589>.
- Hogan, B. "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. Bulletin of Science, Technology & Society," no. 30(6), 377–386. (2010). <https://doi.org/10.1177/0270467610385893>.
- Jane, E. A. "Flaming? What Flaming? The Pitfalls and Potentials of Researching Online Hostility. Ethics and Information Technology, 17(1)," 2015, 65–87. <https://doi.org/10.1007/s10676-015-9362-0> <https://doi.org/10.1007/s10676-015-9362-0>.
- Jayus, J., Abdullah, A., Mustafa, A., & Sumaiyah, S. "YouTube, Public Discourse, and the 'Makan Siang Gratis' Program: An Analysis of Toxicity Comments on the Liputan6 Channel." *Proceeding Jogja Communication Conference (JCC-Indonesia)*, 8(1), 1–20., 2015. <https://jcc-indonesia.id/proceeding/index.php/jcc/article/view/383>.
- Jenkins, H. *Convergence Culture: Where Old and New Media Collide*. New York University Press., 2006.
- Jhaver, S., Bruckman, A., & Gilbert, E. "Does Transparency in Moderation Really Matter? Proceedings of the ACM on Human-Computer Interaction, 3(CSCW)," 2019, 1–27.
- Juncosa, G., Yasseri, T., Koltai, J., & Iniguez, G. "Toxic Behavior Silences Online Political Conversations." *ArXiv Preprint ArXiv:2412.05741*, 2024. <https://arxiv.org/abs/2412.05741>.
- Kaplan, A. M., & Haenlein, M. "Users of the World, Unite! The Challenges and Opportunities of Social Media." *Business Horizons*, no. 53(1), 59–68. (2010). <https://doi.org/10.1016/j.bushor.2009.09.003>.

- “Katz, E., Blumler, J. G., & Gurevitch, M. (1974). Utilization of Mass Communication by the Individual. In J. G. Blumler & E. Katz (Eds.), *The Uses of Mass Communications: Current Perspectives on Gratifications Research* (Pp. 19–32). Sage,” n.d.
- “Kemp, S. (2023). Digital 2023: Indonesia. DataReportal. <https://Datareportal.Com/Reports/Digital-2023-Indonesia>,” n.d.
- Kominfo. “Pedoman Pengendalian Konten Negatif Di Internet,” 2022. <https://kominfo.go.id/>.
- Kompas TV. “Rocky Gerung: Pemerintah Produsen Hoaks Terbaik. [Video]. YouTube. Diakses 6 Agustus 2025, Dari,” 2017. <https://www.youtube.com>.
- Kurtz, Howard. “Final Edition: Rocky Mountain News to Shut Down Today.” *The Washington Post Democracy Dies in Darkness*, 2009. <https://www.washingtonpost.com/archive/business/2009/02/27/final-edition-rocky-mountain-news-to-shut-down-today/25fa709b-f799-425f-98c9-e2656fc713d2/>.
- Lapidot-Lefler, Noam, and Azy Barak. “The Benign Online Disinhibition Effect: Could Situational Factors Induce Self-Disclosure and Prosocial Behaviors?” *Cyberpsychology* 9, no. 2 (2015). <https://doi.org/10.5817/CP2015-2-3>.
- Lunando, Edwin, and Ayu Purwarianti. “Indonesian Social Media Sentiment Analysis with Sarcasm Detection.” *2013 International Conference on Advanced Computer Science and Information Systems, ICACIS 2013*, 2013, 195–98. <https://doi.org/10.1109/ICACIS.2013.6761575>.
- Maddock, Jim, Kate Starbird, Haneen Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. “Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures.” *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 2015, 228–41. <https://doi.org/10.1145/2675133.2675280>.
- Mangold, W.G., & Faulds, D. J. “Social Media: The New Hybrid Element of the Promotion Mix.” *Business Horizons* 52(4), 357 (2009). <https://doi.org/10.1016/j.bushor.2009.03.002>.
- Mardatrisna, S. R., & Banowo, E. “Analisis Isi Komentar Model Analisis Harold Laswell Pada Akun YouTube Channel Metro TV Mengenai Konten ‘HOTROOM – Naturalisasi Untuk Prestasi?’” *Triwikrama: Jurnal Ilmu Sosial* 5 (2024). <https://ejournal.warunayama.org/index.php/triwikrama/article/download/6133/5659/18700>.

- Mardhiyyah Millatul. "Konvergensi Media (Analisis Transformasi Media Konvensional Dalam Perspektif Ekonomi Kritis)." *Jurnal An-Nida* 15, no. 2 (2023): 130.
- Marwick, Alice E., and Danah Boyd. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media and Society* 13, no. 1 (2011): 114–33. <https://doi.org/10.1177/1461444810365313>.
- Munger, K., & Phillips, J. "Right-wing YouTube: A Supply and Demand Perspective." *International Journal of Press/Politics*, no. 25(3), 480–500. (2020). <https://doi.org/10.1177/1940161219892519>.
- Napoli, P. M. *Social Media and the Public Interest: Media Regulation in the Disinformation Age*. Columbia University Press., 2019.
- Nasrullah, M. *Ekonomi Politik Media Dan Demokrasi*. Prenada Media Group., 2012.
- Nasrullah, R. *Media Sosial: Perspektif Komunikasi, Budaya, Dan Sosioteknologi*. Simbiosis Rekatama Media., 2015.
- Novitasari, Lusy. "Penyimpangan Perilaku Seks Waria Dalam Novel Taman Api Karya Yonathan Rahardjo." *Deiksis* 10, no. 02 (2018): 125. <https://doi.org/10.30998/deiksis.v10i02.2339>.
- Nugroho, Irwan, and Irwansyah Irwansyah. "Konvergensi Konten Audio Di Media Online (Studi Kasus Podcast Detik.Com)." *Jurnal Komunikasi* 15, no. 1 (2021): 55–70. <https://doi.org/10.21107/ilkom.v15i1.9753>.
- Obadimu, Adewale, Esther Mead, Muhammad Nihal Hussain, and Nitin Agarwal. "Identifying Toxicity within Youtube Video Comment." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11549 LNCS (2019): 214–23. [https://doi.org/10.1007/978-3-030-21741-9\\_22](https://doi.org/10.1007/978-3-030-21741-9_22).
- . *Identifying Toxicity within Youtube Video Comment. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11549 LNCS. Springer International Publishing, 2019. [https://doi.org/10.1007/978-3-030-21741-9\\_22](https://doi.org/10.1007/978-3-030-21741-9_22).
- Otoni, R., Cunha, E., Magno, G., Bernadina, P., Meira Jr., W., & Almeida, V. "Analyzing Right-Wing YouTube Channels: Hate, Violence and Discrimination." *ArXiv Preprint*, 2018. <https://arxiv.org/abs/1804.04096>.
- Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You*. 2011. Penguin Books., n.d.

- Purwaningsih, R. "Aktivisme Digital Dalam Konten Politik YouTube Indonesia." *Jurnal Komunikasi Indonesia*, 2022, 11(2), 150–165.
- Putri, Adel Andila, and Jurnal Data. "Tantangan Baru Industri Media Di Indonesia." Jakarta, 2023. <https://goodstats.id/article/tantangan-baru-industri-media-di-indonesia-XIPEa>.
- Reinert Yosua Rumagit. "Multilabel Classification for Toxic Comments in Indonesian." *Engineering Mathematics and Computer Science (EMACS) Journal*, no. (1):29-34 (2020). <https://doi.org/DOI:10.21512/emacsjournal.v2i1.6256>.
- Rheingold, H. *The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press., 2000.
- Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira Jr, W. "Auditing Radicalization Pathways on YouTube. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT)," 2020, 131–141. <https://doi.org/10.1145/3351095.3372879>.
- Rumagit, Reinert Yosua. "Multilabel Classification for Toxic Comments in Indonesian." *Engineering, Mathematics and Computer Science (EMACS) Journal* 2, no. 1 (2020): 29–34. <https://doi.org/10.21512/emacsjournal.v2i1.6256>.
- Russell, S. J., & Norvig, P. "Artificial Intelligence: A Modern Approach." *Pearson Education*. 3rd ed (2010).
- Sagama, Y., & Alamsyah, A. "Multi-Label Classification of Indonesian Online Toxicity Using BERT and RoBERTa." *ResearchGate*, 2023. [https://www.researchgate.net/publication/373036835\\_Multi-Label\\_Classification\\_of\\_Indonesian\\_Online\\_Toxicity\\_using\\_BERT\\_and\\_RoBERTa](https://www.researchgate.net/publication/373036835_Multi-Label_Classification_of_Indonesian_Online_Toxicity_using_BERT_and_RoBERTa).
- Santoso, Y., & Widodo, B. "Analisis Komentar Toksik Pada Video Politik Di YouTube." *Jurnal Teknologi Informasi Dan Komunikasi*, no. 10(1), 71–85. (2023).
- Sari, Tiara Intana, Zalfa Natania Ardilla, Nur Hayatin, and Ruhaila Maskat. "Abusive Comment Identification on Indonesian Social Media Data Using Hybrid Deep Learning." *IAES International Journal of Artificial Intelligence* 11, no. 3 (2022): 895–904. <https://doi.org/10.11591/ijai.v11.i3.pp895-904>.
- Sasan, D., & Gunica, J. "Analyzing Toxicity in YouTube Comments with the Help of Machine Learning [Undergraduate Thesis, Stockholm University]." *DiVA Portal*, 2023. <https://www.diva-portal.org/smash/get/diva2:1784332/FULLTEXT01.pdf>.

- “Setara Institute. (2024). Tentang Kami. Diakses 6 Agustus 2025, Dari [Https://Setarainstitute.Org](https://Setarainstitute.Org),” n.d.
- Shafa, S. “Racism in Social Media: A Critical Discourse Analysis of YouTube Comments.” *Indonesian Journal of English Language Teaching and Applied Linguistics (IJELTAL)*, no. 6(2), 145–162. (2022). <https://ijeltal.org/index.php/ijeltal/article/view/1307>.
- Shao, G., & Wang, J. “Understanding User-Generated Content in Social Media: A Literature Review.” *Internet Research*, no. 27(4), 854–874. (2017). <https://doi.org/10.1108/IntR-07-2016-0201>.
- Shirky, C. *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin Press., 2008.
- Suler, J. “The Online Disinhibition Effect.” *CyberPsychology & Behavior*, no. 7(3), 321–326. (2004). <https://doi.org/10.1089/1094931041291295>.
- Sundar, S. S., & Limperos, A. M. “Uses and Grats 2.0: New Gratifications for New Media.” *Journal of Broadcasting & Electronic Media* 57(4), 504 (2023). <https://doi.org/10.1080/08838151.2013.845827>.
- Sundar, S. S. *Social Psychology of Interactivity in Human–Website Interaction*. In A. N. Joinson et Al. (Eds.), *The Oxford Handbook of Internet Psychology*, 2007.
- Sunstein, C. R. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press., 2018.
- Tahir, I., & Ramadhan, M. G. F. “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), 72– (2024). <https://e-journal.usd.ac.id/index.php/LLT/article/view/8180>.
- Tubbs. Stewart L dan Moss Sylvia. *Human Communication : Prinsip - Prinsip Dasar*. PT. Remaja Rosdakarya, 2008.
- Tufekci, Z. “Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency.” *Colorado Technology Law Journal* 13, 203–21 (2015).
- Tufekci, Z. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press., 2017.
- Turkle, S. *Alone Together: Why We Expect More from Technology and Less from Each Other.*, 2011. Basic Books.
- Ulum, Nuzzulul. “Pola Komunikasi Kyai Dalam Membentuk Karakter Santri Di

Pesantren Ash-Shiddiqi Putri Talangsari Jember.” *Fajar: Jurnal Pendidikan Islam* 1, no. 1 (2021): 48–62.

Verywell Mind. (n.d.). “Mental Health Effects of Reading Negative Comments Online.” 2025.

Wardle, C., & Derakhshan, H. “Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking.” *Council of Europe.*, 2017.

Weißmann, S., Philipp, A., Verwiebe, R., Krauter, C. O., Fritsch, N. S., & Buder, C. “Clicks, Comments, Consequences: Are Content Creators’ Socio-Structural and Platform Characteristics Shaping Exposure to Negative Sentiment, Offensive Language, and Hate Speech on YouTube?” *ArXiv Preprint.*, 2025. <https://arxiv.org/abs/2403.00745>.

“Wikipedia. (2024, Agustus 6). Abdul Somad. Wikipedia Bahasa Indonesia. [https://Id.Wikipedia.Org/Wiki/Abdul\\_Somad](https://Id.Wikipedia.Org/Wiki/Abdul_Somad),” n.d.

“Wikipedia. (2024, Agustus 6). Rocky Gerung. Wikipedia Bahasa Indonesia. Diakses 6 Agustus 2025, Dari [https://Id.Wikipedia.Org/Wiki/Rocky\\_Gerung](https://Id.Wikipedia.Org/Wiki/Rocky_Gerung),” n.d.

YouTube. “Eksklusif Dialog Akal Sehat Rocky Gerung Ustadz Abdul Somad YouTube.” 2023. <https://www.youtube.com/watch?v=B2WtOAwl158>.

Zhou, X., Zafarani, R., Shu, K., & Liu, H. “Characterizing Online Abuse: A Survey.” no. arXiv preprint arXiv:2101.12552. (2021). <https://arxiv.org/abs/2101.12552>.

t, *Messing Around, and Geeking Out: Kids Living and Learning with New Media.*