



ARKANA

Jurnal Komunikasi dan Media

Link : <https://ojs.unsiq.ac.id/index.php/arkana/article>
Volume : 05
Nomor : 01
Bulan : Juni
ISSN (online) : 2962-1917
DOI : <https://doi.org/10.62022/rpb16866>

MARAKNYA *HATE SPEECH* PADA PLATFORM TIKTOK DAN DAMPAKNYA BAGI GENERASI MUDA

Albertus Bintang Cahyo Gumelar¹

Aryasuta Al Musthofa²

Darren Pagliuca Weringkukly³

Denzel Matthew Parish Saragi⁴

Gabriella Happy Christian⁵

Klement Deo Soejatono⁶

^{1,2,3,4,5,6} Universitas Gadjah Mada, Indonesia

Corresponding Author: darrenpagliucawering@mail.ugm.ac.id

Article History: Received: 7-April-2026, Revised: 11-Mei-2026, Accepted: 13-Juni-2026

ABSTRAK

Platform media sosial TikTok berkembang sangat pesat hingga mengubah ruang komunikasi digital. Bagi generasi muda, platform ini dijadikan ruang interaksi sosial, politik, dan budaya. Namun, di balik berbagai daya tarik yang dimiliki media sosial TikTok, algoritma For Your Page (FYP) rentan disalahgunakan menjadi sarana yang mempercepat sirkulasi informasi penyebaran ujaran kebencian (*hate speech*). Maraknya fenomena ujaran kebencian dapat memengaruhi perkembangan kesehatan mental generasi muda, sehingga diperlukan penanggulangan yang efektif. Analisis melalui pendekatan kualitatif berupa studi literatur dan analisis konten dari berbagai jurnal ilmiah dan komunikasi digital yang diperoleh dari observasi digital pada kolom komentar TikTok. Hasil penelitian ini menunjukkan bahwa ujaran kebencian di TikTok disebabkan oleh perbedaan opini dan gaya hidup yang dilontarkan oleh pelaku yang sifatnya spontan berupa komentar negatif. Solusi dari permasalahan ini yaitu untuk memberikan rekomendasi terkait kebijakan strategis sehingga tercipta ruang digital yang lebih sehat, nyaman, dan konstruktif.

Kata kunci :

TikTok, Generasi Muda, Kebencian, Kesehatan

ABSTRACT

The TikTok social media platform has grown so rapidly that it has transformed the digital communication space. For the younger generation, this platform has become a space for social, political, and cultural interaction. However, behind the various attractions of TikTok social media, the For Your Page (FYP) algorithm is vulnerable to misuse as a means to accelerate the circulation of information and the spread of hate speech. The rise of hate speech can affect the mental health development of the younger generation, so effective countermeasures are needed. Analysis using a

qualitative approach in the form of literature studies and content analysis from various scientific journals and digital communication obtained from digital observations in the TikTok comment section. The results of this study indicate that hate speech on TikTok is caused by differences in opinion and lifestyle expressed by perpetrators who are spontaneous in the form of negative comments. The solution to this problem is to provide recommendations related to strategic policies to create a healthier, more comfortable, and constructive digital space.

Keywords :

TikTok, Generation, Youth, Hate, Health

PENDAHULUAN

Teknologi digital yang telah berkembang pesat mendorong transformasi besar dalam cara manusia berkomunikasi dan berinteraksi. Sebagai salah satu produk utama di era digital, media sosial menjadi tempat bagi manusia untuk bertukar opini, pemikiran, dan ekspresi secara lebih cepat dan luas. Namun, di balik kemudahan yang diberikan, media sosial juga dapat memberikan efek yang negatif, salah satunya adalah maraknya ujaran kebencian (*hate speech*). Fenomena ini menjadi ancaman nyata terhadap kualitas diskursus publik karena dapat memicu konflik, polarisasi, serta menurunkan kualitas interaksi sosial di ruang digital.

Secara konseptual, ujaran kebencian merujuk pada segala bentuk komunikasi verbal, tulisan, maupun perilaku yang menyerang, merendahkan, atau mengintimidasi seseorang maupun kelompok berdasarkan karakteristik identitas bawaan mereka, seperti suku, agama, ras, antargolongan, gender, hingga disabilitas (Pangestu, Abdurrachman dan Rizkianto, 2024). Dalam lanskap media sosial seperti TikTok, fenomena ini sering termanifestasi dalam bentuk komentar provokatif atau konten manipulatif yang secara sengaja dirancang untuk menyudutkan kelompok terpinggirkan. Dalam kajian ilmu sosial dan hukum, penting untuk menarik garis batas yang tegas antara ujaran kebencian dan kebebasan berpendapat. Sebuah pernyataan tidak serta-merta diklasifikasikan sebagai hate speech hanya karena kata-katanya menyinggung perasaan, tidak sopan, atau berupa kritik yang sangat tajam. Batasan kritis yang mendefinisikan hate speech adalah adanya intensi untuk menghasut permusuhan, mendiskriminasi, atau memicu tindakan kekerasan terhadap kelompok sasaran (Davidson dkk., 2017). Di Indonesia, batasan yuridis ini juga dipertegas dalam Undang-undang Informasi dan Transaksi Elektronik (UU ITE), di mana suatu ekspresi digital baru dianggap melanggar hukum jika mengandung unsur provokasi yang secara spesifik menargetkan sentimen SARA dengan tujuan memecah belah kohesi sosial.

Dalam beberapa tahun terakhir, terdapat satu platform yang memiliki perkembangan yang sangat pesat, yaitu TikTok. Platform ini telah banyak digunakan oleh berbagai usia dan kalangan, terutama generasi muda. Platform TikTok umumnya hadir sebagai media hiburan, tetapi juga dapat menjadi sarana komunikasi politik, agama, dan budaya. Berbagai video yang diunggah di TikTok memiliki karakteristik, yaitu durasinya yang pendek, visual, dan emosional, menjadikannya sangat efektif dalam menyampaikan pesan secara cepat dan menarik. Namun, sifat tersebut juga berpotensi dimanfaatkan untuk menyebarkan konten yang bersifat provokatif, termasuk ujaran kebencian dan narasi diskriminatif.

TikTok merupakan media sosial yang dikembangkan oleh ByteDance, sebuah

perusahaan Tiongkok, yang memungkinkan pengguna untuk mengunggah video berdurasi singkat dengan berbagai macam fitur dan filter keren serta lucu (Weimann & Masri, 2023). Karena keunikannya tersebut, TikTok mengalami pertumbuhan pengguna yang sangat pesat dan menjadi media sosial yang digandrungi oleh masyarakat global. Berdasarkan Afriany dkk. (2025), TikTok memiliki 1,8 miliar pengguna aktif yang diproyeksikan akan terus naik setiap tahunnya. Indonesia merupakan salah satu negara di dunia yang paling banyak menggunakan platform TikTok, yaitu sebanyak 184,95 juta pengguna (Pujangkara, 2025). Hutamy dkk. (2021) menyebutkan bahwa 10 juta pengguna merupakan anak milenial dan generasi Z. Tingginya intensitas penggunaan TikTok membawa dampak signifikan pada ranah psikologis dan perilaku sosial. Konten-konten yang muncul dari content creator mengakibatkan masyarakat sering kali merasa harus terus mengikutinya. Dorongan untuk mengikuti tren ini memengaruhi pola pikir, gaya hidup, dan kebiasaan sehingga dapat memicu perubahan kondisi mental seseorang. Rendahnya literasi digital pada generasi muda membuat mereka rentan menyerap narasi manipulatif yang akhirnya memengaruhi perilaku digital mereka dalam menanggapi suatu fenomena sosial (Ellys dkk., 2025).

Meskipun demikian, tidak semua konten di TikTok secara dominan mengandung ujaran kebencian. Studi menunjukkan bahwa konten yang bersifat hiburan dan humor justru lebih banyak menarik perhatian pengguna dibandingkan dengan konten yang mengandung kebencian. Hal ini mengindikasikan bahwa algoritma dan karakteristik platform dapat memengaruhi visibilitas serta penyebaran jenis konten tertentu. Namun, keberadaan ujaran kebencian tetap menjadi perhatian karena dampaknya yang signifikan terhadap persepsi dan perilaku pengguna, khususnya generasi muda yang merupakan pengguna dominan platform ini.

Di sisi lain, ujaran kebencian di media sosial tidak hanya berdampak pada individu, tetapi juga pada kualitas keseluruhan komunikasi publik. Paparan terhadap konten yang bersifat toksik dapat membuat pengguna enggan berpartisipasi dalam diskusi, memicu respons negatif, bahkan memperkuat siklus kebencian di ruang digital. Dalam jangka panjang, kondisi ini dapat menciptakan lingkungan komunikasi yang tidak sehat dan menghambat terbentuknya dialog yang konstruktif.

Upaya penanggulangan ujaran kebencian selama ini sering dilakukan melalui pendekatan top-down, seperti moderasi oleh platform atau regulasi pemerintah. Namun, pendekatan ini menghadapi berbagai kendala, termasuk isu kebebasan berekspresi dan keterbatasan dalam menjangkau skala besar pengguna. Oleh karena itu, muncul konsep counter speech atau moderasi kolektif oleh pengguna sebagai alternatif solusi. Strategi ini menekankan peran aktif masyarakat dalam merespons ujaran kebencian secara konstruktif, misalnya dengan menyampaikan opini tanpa hinaan atau menggunakan pendekatan humor, yang terbukti dapat menurunkan tingkat kebencian dan toksisitas dalam diskusi daring.

Melihat kondisi tersebut, penting untuk mengkaji lebih dalam fenomena maraknya hate speech di platform TikTok serta dampaknya terhadap generasi muda. Tidak hanya menjadi konsumen, generasi ini memiliki kemampuan untuk memproduksi konten yang turut membentuk dinamika komunikasi digital. Oleh karena itu,

pemahaman yang komprehensif mengenai pola penyebaran, karakteristik, serta dampak ujaran kebencian menjadi krusial dalam upaya menciptakan ruang digital yang lebih sehat, inklusif, dan konstruktif.

Masalah penelitian yang diangkat pada penelitian ini memiliki keterkaitan satu dengan yang lainnya. Diawali dengan menganalisis pola hate speech yang berkembang pada platform TikTok untuk memahami bentuk, karakteristik, dan bagaimana ujaran kebencian tersebut dilontarkan oleh pengguna media sosial TikTok. Kemudian, penelitian ini mengidentifikasi faktor pendukung terjadinya hate speech dari berbagai sudut pandang psikologis dan pengaruh algoritma For Your Page yang menjadi faktor penting. Selanjutnya, dilakukan analisis mengenai dampak ujaran kebencian terhadap kesehatan mental generasi muda yang sering menjadi target utama konten toksik yang tersebar di TikTok. Akhirnya, akan dirumuskan langkah yang dapat dilakukan untuk mencegah dan meminimalisasi fenomena ujaran kebencian yang semakin meluas di kalangan generasi muda demi menciptakan ruang digital yang sehat dan aman. Tujuan dari penelitian ini yang sejalan dengan rumusan masalah yaitu memetakan fenomena ujaran kebencian (*hate speech*) pada media sosial TikTok dan mengevaluasi implikasi terhadap perkembangan zaman generasi muda.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan metode studi literatur dan observasi digital terhadap aktivitas pengguna pada platform TikTok. Alasan pemilihan pendekatan kualitatif adalah fokus penelitian yang berpusat pada pemahaman fenomena sosial berupa ujaran kebencian (*hate speech*) yang berkembang di ruang digital, khususnya pada kolom komentar konten TikTok, serta dampaknya terhadap generasi muda. Penelitian ini bertujuan untuk menggambarkan pola, karakteristik, faktor pendorong, dan dampak hate speech secara mendalam melalui interpretasi data yang diperoleh dari berbagai sumber.

Studi literatur dan observasi nonpartisipatif menjadi metode pilihan dalam mengumpulkan data. Langkah pertama yaitu pengumpulan literatur berupa referensi yang berkaitan dengan hate speech, media sosial, algoritma digital, kesehatan mental remaja, serta perilaku komunikasi generasi muda. Literatur yang digunakan berasal dari sumber yang kredibel dan relevan sesuai dengan topik penelitian.

Selain studi literatur, penelitian ini juga melakukan observasi terhadap kolom komentar pada beberapa konten TikTok yang memiliki tingkat interaksi tinggi dan berpotensi memunculkan ujaran kebencian. Observasi dilakukan dengan memperhatikan bentuk komentar, pola bahasa yang digunakan, respons antarpengguna, serta kecenderungan penyebaran komentar toksik melalui sistem For Your Page (FYP). Peneliti tidak terlibat secara langsung dalam interaksi digital tersebut sehingga observasi dilakukan secara nonpartisipatif.

Teknik analisis isi digunakan untuk menganalisis data yang diperoleh. Analisis dilakukan dengan mengidentifikasi bentuk-bentuk ujaran kebencian berdasarkan kategori tertentu, seperti penghinaan, stereotip, diskriminasi, provokasi, dan komentar bernada agresif yang dibungkus dalam konteks humor atau “candaan”. Selanjutnya, data

dianalisis untuk melihat hubungan antara algoritma TikTok, intensitas penyebaran hate speech, serta dampaknya terhadap pola pikir dan kesehatan mental generasi muda.

Teknik triangulasi sumber menjadi pilihan untuk meningkatkan validitas penelitian. Teknik ini membandingkan hasil observasi digital dengan teori dan temuan dari berbagai literatur ilmiah yang relevan. Dengan metode ini, penelitian yang dikerjakan mampu memberikan gambaran yang komprehensif mengenai maraknya hate speech di TikTok serta implikasinya terhadap kualitas interaksi sosial generasi muda di ruang digital.

HASIL DAN PEMBAHASAN

A. Pola dan Karakteristik *Hate Speech* di Platform TikTok

Berdasarkan observasi terhadap kolom komentar pada konten TikTok yang memiliki tingkat interaksi tinggi, ditemukan bahwa ujaran kebencian di TikTok memiliki karakteristik yang khas dibandingkan dengan platform media sosial lain. Ujaran kebencian umumnya muncul dalam bentuk komentar singkat, padat, dan emosional, sejalan dengan karakter TikTok yang berbasis video pendek dan interaksi cepat. Studi linguistik forensik yang dilakukan Rahman dan Putri (2024) mengonfirmasi bahwa *hate speech* di TikTok cenderung dikemas dalam bahasa informal, penggunaan singkatan, *slang*, serta sindiran yang dibalut humor sehingga sulit dikenali secara langsung sebagai ujaran kebencian.

Pola *hate speech* yang teridentifikasi dapat dikategorikan ke dalam beberapa bentuk dominan. *Pertama*, penghinaan personal (*personal insult*) yang menyerang fisik, gaya bicara, atau penampilan kreator konten. *Kedua*, stereotip berbasis identitas seperti suku, agama, ras, antargolongan (SARA), gender, dan orientasi seksual. *Ketiga*, diskriminasi terhadap kelompok minoritas atau kelompok yang dianggap "berbeda". *Keempat*, provokasi politik yang sering kali menggunakan retorika "kami versus mereka" (*us versus them*). *Kelima*, komentar agresif yang dibungkus dalam konteks humor atau "candaan" sehingga normalisasi terjadi secara halus. Klasifikasi serupa juga ditemukan dalam penelitian Prasetyo dan Nugraha (2023) yang menggunakan pendekatan *Naive Bayes* dan *TF-IDF* untuk mengklasifikasi komentar *hate speech* di TikTok, di mana kategori penghinaan, ancaman, dan diskriminasi menjadi kelompok yang paling dominan.

Karakteristik lain yang menonjol adalah penggunaan retorika antagonistik oleh aktor politik, khususnya kelompok populis. González-Aguilar, Segado-Boj, dan Makhortykh (2023) mencatat bahwa partai-partai politik kanan populis di TikTok memanfaatkan strategi *spectacularization* dan *personalization* untuk menyampaikan pesan kebencian terhadap kelompok tertentu, seperti imigran atau kelompok minoritas. Fenomena ini turut diperkuat di Indonesia melalui isu-isu sensitif seperti SARA, yang muncul dalam kolom komentar dalam bentuk sindiran, *meme*, ataupun komentar provokatif. Hidayah, Marcelawati, dan Saputra (2021) dalam studinya terhadap akun TikTok tertentu juga menemukan bahwa *cyber harassment* dan komentar kebencian intensif terjadi pada akun yang memuat konten kontroversial atau yang viral di masa pandemi Covid-19.

Tabel 3. 1 Kategori dan karakteristik *hate speech* pada kolom komentar TikTok

Kategori	Bentuk Komentar	Karakteristik
Stereotip SARA	Komentar berbasis ras, suku, agama	Memanfaatkan sentimen identitas
Diskriminasi	Menyerang kelompok minoritas	Eksklusif marginalisasi
Provokasi Politik	Retorika “Kami vs Mereka”	Polarisatif antagonistik
Humor Toksik	Komentar agresif dalam balutan candaan	Normalisasi, sulit dikenali

Temuan ini memperlihatkan bahwa *hate speech* di TikTok bukan hanya soal kata-kata kasar, melainkan strategi komunikasi yang memanfaatkan karakter platform: cepat, emosional, dan viral.

B. Faktor Pendorong Maraknya *Hate Speech* di TikTok

Hasil analisis menunjukkan terdapat beberapa faktor utama yang mendorong tingginya intensitas ujaran kebencian di TikTok sehingga dapat dibagi menjadi 3 faktor, yaitu faktor teknologis, faktor sosial-psikologis, dan faktor demografis.

a. Faktor Algoritma *For Your Page* (FYP)

Algoritma *For Your Page* (FYP) merupakan penggerak utama distribusi konten di TikTok. Sistem ini bekerja berdasarkan *engagement* pengguna, seperti durasi menonton, jumlah *like*, komentar, dan *share*. Konten yang memicu emosi kuat—termasuk konten provokatif dan kontroversial—tendensi untuk mendapatkan *engagement* lebih tinggi sehingga lebih sering didistribusikan ke pengguna lain. Akibatnya, *hate speech* yang menimbulkan reaksi emosional justru ditampilkan lebih luas oleh sistem. González-Aguilar et al. (2023) menyebut fenomena ini sebagai bentuk *spectacularization*, di mana konten yang sensasional dan emosional mendapatkan visibilitas yang lebih besar dibandingkan konten informatif yang netral.

Meskipun demikian, perlu dicatat bahwa tidak semua konten dominan di TikTok mengandung kebencian. González-Aguilar et al. (2023) juga menemukan bahwa konten hiburan dan humor justru lebih banyak menarik perhatian pengguna dibandingkan dengan konten kebencian secara eksplisit. Hal ini menunjukkan bahwa algoritma bersifat *amplifier* (penguat), bukan penyebab tunggal, tetapi efek penguatannya terhadap konten emosional negatif tetap signifikan.

b. Faktor Anonimitas dan *Disinhibition Effect*

Faktor kedua adalah anonimitas yang ditawarkan platform digital. Pengguna dapat membuat akun dengan identitas samaran sehingga merasa bebas dari konsekuensi sosial. Hidayah et al. (2021) menjelaskan bahwa *cyber harassment* di TikTok meningkat tajam karena pengguna merasa “tersembunyi”

di balik layar, yang melahirkan efek disinhibisi (*online disinhibition effect*). Membuat pengguna lebih berani melontarkan komentar yang tidak akan mereka ucapkan dalam interaksi tatap muka.

c. Faktor Polarisasi dan Identitas Kelompok

Komunikasi politik di media sosial sering kali didorong oleh logika polarisasi. Aktor politik populis memanfaatkan TikTok untuk membangun narasi "kami versus mereka" dengan menargetkan kelompok tertentu sebagai musuh (González-Aguilar et al., 2023). Polarisasi ini memperkuat solidaritas dalam kelompok (*in-group*) sekaligus memicu kebencian terhadap kelompok luar (*out-group*), sehingga komentar kebencian terus bersiklus sebagai bentuk afirmasi identitas.

d. Faktor Demografis dan Karakteristik Pengguna

Generasi muda merupakan pengguna dominan TikTok dengan karakteristik psikologis yang masih dalam tahap pencarian identitas. Nufus, Maharani, dan Zahro (2022) menemukan bahwa pengguna muda TikTok memiliki kecenderungan psikologis yang khas, di antaranya impulsivitas, kebutuhan validasi sosial, dan kerentanan terhadap pengaruh tren. Lisdinani, Afifah, Restiana, dan Supriyono (2023) menambahkan bahwa TikTok memiliki pengaruh signifikan terhadap wawasan kebangsaan generasi muda, baik secara positif maupun negatif, tergantung pada konten yang dikonsumsi. Kombinasi karakter psikologis ini membuat generasi muda lebih rentan untuk ikut serta dalam praktik *hate speech*, baik sebagai pelaku maupun sebagai konsumen yang menormalisasinya.

C. Dampak *Hate Speech* Terhadap Kesehatan Mental Generasi Muda

Dampak ujaran kebencian terhadap generasi muda tidak hanya bersifat sosial, tetapi juga psikologis dan jangka panjang. Hasil analisis literatur menunjukkan setidaknya empat dampak utama yang saling berkaitan.

a. Penurunan Kepercayaan Diri dan Citra Diri

Paparan terhadap komentar kebencian, terutama yang menyerang fisik atau identitas, dapat menurunkan kepercayaan diri generasi muda. Hidayat, Rifki, dan Rachman (2025) menjelaskan bahwa konsumsi konten negatif di Instagram dan TikTok berdampak signifikan terhadap kesehatan mental siswa, termasuk menurunnya *self-esteem* dan meningkatnya perasaan tidak berharga. Generasi muda yang menjadi sasaran *body shaming* atau penghinaan personal sering kali menginternalisasi komentar negatif tersebut sehingga membentuk citra diri yang buruk.

b. Kecemasan, Depresi, dan Stres Psikologis

Sarumaha, Manik, Borus, dan Sitepu (2024) mengidentifikasi bahwa ada keterkaitan yang signifikan antara intensitas konsumsi media sosial dengan kesehatan mental remaja, di mana semakin tinggi paparan terhadap konten

negatif, semakin tinggi pula risiko kecemasan dan depresi. Hasil serupa diungkapkan Yuhana, Mariyati, dan Sugiyanto (2023) yang menegaskan bahwa konsumsi media sosial berlebihan berpengaruh terhadap kesehatan mental remaja yang menurun. Lebih jauh, studi *specification curve analysis* yang dilakukan Twenge, Haidt, Lozano, dan Cummins (2022) menemukan bahwa konsumsi media sosial berhubungan langsung dengan buruknya kesehatan mental, khususnya pada remaja perempuan. Temuan ini mengindikasikan adanya dimensi gender dalam dampak *hate speech*, di mana perempuan muda cenderung mengalami konsekuensi psikologis yang lebih berat.

c. Gangguan Perilaku Belajar dan Sosial

Dampak *hate speech* juga merembet pada aspek perilaku belajar. Agustiah, Fauzi, dan Ramadhani (2020) menyimpulkan bahwa konsumsi media sosial berpengaruh terhadap proses pembelajaran siswa di sekolah, di antaranya menurunnya konsentrasi, berkurangnya motivasi belajar, dan kecenderungan menarik diri dari interaksi sosial di sekolah. Generasi muda yang terus-menerus terpapar komentar negatif cenderung mengalami gangguan tidur, kelelahan emosional, dan penurunan performa akademik.

d. Normalisasi Toksisitas dan Pembelajaran Sosial

Dari perspektif teori pembelajaran sosial (*social learning theory*), generasi muda mempelajari pola komunikasi dari lingkungan digitalnya. Ketika *hate speech* terus ditampilkan dan mendapat *engagement* tinggi, perilaku ini dipelajari sebagai sesuatu yang "wajar" Lisdinani et al. (2023) memperingatkan bahwa hal ini berdampak pada wawasan kebangsaan dan sikap toleransi generasi muda. Selain itu, paparan konten toksik dapat menciptakan *echo chamber* psikologis di mana pengguna menjadi enggan berpartisipasi dalam diskusi sehat dan justru memperkuat siklus kebencian.

Tabel 3. 2 Dampak *hate speech* pada kesehatan mental generasi muda

Dimensi Dampak	Bentuk Dampak	Refrensi Pendukung
Citra Diri	Penurunan <i>self-esteem</i> , <i>body shaming</i>	Hidayat et al. (2025)
Emosi/Psikis	Kecemasa, depresi, stres	Sarumaha et al. (2024); Yuhana et al. (2023); Twenge et al. (2022)
Akademik/Perilaku	Penurunan konsentrasi dan motivasi belajar	Agustiah et al. (2020)
Sosial/Karakter	Normalisasi toksisitas, intoleransi	Lisdinani et al. (2023); Nufus et al. (2022)

D. Strategi Penanggulangan *Hate Speech* di TikTok

Berdasarkan analisis pola, faktor pendorong, dan dampaknya, strategi penanggulangan *hate speech* perlu dirancang secara multilevel, mencakup pendekatan teknologis, regulasi, edukasi, dan partisipasi komunitas.

a. Pendekatan Teknologis: Moderasi Algoritmik

Pendekatan pertama adalah penguatan moderasi konten berbasis teknologi. Prasetyo dan Nugraha (2023) menunjukkan bahwa pendekatan *machine learning* dengan algoritma *Naive Bayes* dan *TF-IDF* dapat mendeteksi komentar *hate speech* di TikTok dengan tingkat akurasi yang memadai. Platform perlu mengintegrasikan sistem deteksi otomatis yang lebih sensitif terhadap konteks bahasa lokal Indonesia, termasuk *slang* dan *humor* terselubung. Selain itu, penyesuaian algoritma FYP agar tidak terlalu mengutamakan konten provokatif (*de-amplification*) menjadi langkah penting untuk memutus siklus penyebaran.

b. Pendekatan Regulasi Kebijakan *Top-Down*

Pendekatan regulasi melalui Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) dan kerja sama antara pemerintah dengan platform tetap diperlukan. Namun, pendekatan ini menghadapi keterbatasan, termasuk isu kebebasan berekspresi dan keterbatasan dalam menjangkau skala besar pengguna. Oleh karena itu, regulasi sebaiknya difokuskan pada penanganan kasus berat (ancaman, hasutan kekerasan) sambil membuka ruang bagi pendekatan komunitas.

c. Pendekatan Komunitas: *Counter Speech* dan Moderasi Kolektif

Lasser, Herderich, Garland, Aroyehun, Garcia, dan Galesic (2025) dalam studinya tentang moderasi kolektif pada diskusi daring menemukan bahwa *counter speech* atau moderasi kolektif oleh pengguna terbukti efektif menurunkan tingkat kebencian, toksisitas, dan ekstremitas dalam diskusi *online*. Strategi yang berhasil meliputi (a) menyampaikan opini tanpa hinaan, (b) penggunaan pendekatan humor yang konstruktif, dan (c) keterlibatan komunitas secara konsisten dalam merespons konten kebencian. Strategi ini menempatkan pengguna sebagai aktor aktif, bukan sekadar konsumen pasif.

d. Pendekatan Edukatif: Literasi Digital

Edukasi literasi digital menjadi pilar paling fundamental. Generasi muda perlu dibekali kemampuan berpikir kritis dalam mengonsumsi konten, kemampuan mengenali ujaran kebencian, dan kesadaran akan dampaknya. Lisdinani et al. (2023) menekankan pentingnya pendekatan edukatif yang menghubungkan penggunaan TikTok dengan penguatan wawasan kebangsaan dan toleransi. Program literasi digital dapat dilakukan melalui sekolah, kampus, komunitas, dan kolaborasi dengan *content creator* yang memiliki pengaruh

“Maraknya *Hate Speech* Pada Platform TikTok dan Dampaknya Bagi Generasi Muda”
positif.

e. Pendekatan Psikologis: Dukungan Kesehatan Mental

Mengingat dampak signifikan pada kesehatan mental, perlu disediakan dukungan psikologis bagi korban *hate speech*. Sarumaha et al. (2024) dan Yuhana et al. (2023) merekomendasikan layanan konseling berbasis sekolah dan pemberdayaan keluarga sebagai pendamping utama remaja dalam menghadapi paparan konten negatif.

Tabel 3. 3 Matriks strategi penanggulangan *hate speech* di TikTok

Pendekatan	Aktor Utama	Bentuk Aksi
Teknologis	Platform TikTok	Deteksi otomatis, <i>deamplification</i> algoritma
Regulasi	Pemerintah	Penegakan UU ITE, kerja sama dengan platform
Komunitas	Pengguna	<i>Counter speech</i> , moderasi kolektif
Edukasi	Sekolah, Kampus, Keluarga	Literasi digital, Pendidikan karakter
Psikologis	Tenaga Profesional	Konseling, dukungan Kesehatan mental

Strategi ini menekankan bahwa penanggulangan *hate speech* tidak dapat dibebankan pada satu pihak. Kolaborasi antara platform, pemerintah, masyarakat, institusi pendidikan, dan tenaga kesehatan mental menjadi kunci dalam membangun.

PENUTUP

Kesimpulan

Secara keseluruhan, temuan penelitian menunjukkan adanya keterkaitan erat antara karakteristik teknologi TikTok (terutama algoritma FYP), karakteristik psikologis pengguna muda, pola komunikasi yang berkembang, serta konsekuensi sosial dan psikologis yang ditimbulkan. Algoritma FYP berperan sebagai amplifier yang memperkuat konten emosional, termasuk *hate speech*; karakteristik psikologis generasi muda yang masih dalam tahap pencarian identitas menjadikan mereka rentan, dan absennya literasi digital yang memadai memperburuk situasi. Oleh karena itu, intervensi yang efektif harus bersifat sistemis—menyentuh dimensi teknologi, regulasi, komunitas, edukasi, dan kesehatan mental secara simultan.

Saran

Berdasarkan kesimpulan mengenai perlunya intervensi sistemis dalam menangani dampak algoritma TikTok terhadap generasi muda, saran yang dapat direkomendasikan ditujukan kepada beberapa pihak terkait, antara lain:

1. Bagi Pihak Pengembang

TikTok perlu mengevaluasi dan merancang ulang algoritma rekomendasi agar tidak semata-mata mengutamakan *engagement* (keterlibatan) yang

“Maraknya *Hate Speech* Pada Platform TikTok dan Dampaknya Bagi Generasi Muda” memicu amplifikasi emosional ekstrem, melainkan juga memprioritaskan keamanan psikologis pengguna. Selain itu, pengembang perlu memperketat sistem moderasi otomatis maupun manual yang lebih peka terhadap konteks *hate speech*, perundungan siber, dan konten negatif yang menyoroti kerentanan remaja.

2. Bagi Pemerintah

Pemerintah perlu mendorong lahirnya regulasi khusus yang mewajibkan transparansi algoritma dari platform media sosial dan memberikan perlindungan hukum yang lebih kuat bagi anak dan remaja di ruang digital.

3. Bagi Institusi Pendidikan

Institusi pendidikan berperan untuk memasukkan literasi digital tidak hanya sebatas penggunaan perangkat, tetapi juga pada pemahaman kritis tentang cara kerja algoritma dan dampaknya terhadap privasi serta emosi.

4. Bagi Orang Tua dan Masyarakat

Komunitas dan keluarga perlu menyediakan ruang aktualisasi diri dan pencarian identitas alternatif di dunia nyata, sehingga remaja tidak sepenuhnya menggantungkan validasi pada interaksi di media sosial. Orang tua disarankan untuk tidak sekadar membatasi waktu *screen time*, tetapi juga membuka ruang diskusi yang tidak menghakimi mengenai apa yang anak-anak konsumsi di FYP mereka.

5. Bagi Peneliti Selanjutnya

Diharapkan penelitian selanjutnya dapat melakukan studi jangka panjang untuk mengukur efektivitas dari program literasi digital tertentu terhadap perubahan perilaku pengguna muda di TikTok atau memfokuskan penelitian pada strategi penanganan stres remaja saat menghadapi *hate speech* di platform tersebut.

DAFTAR PUSTAKA

- Afriany, J., Ajima Ritonga, N., Suginam, S. and Mesran, M. (2025) 'Optimalisasi pemasaran digital melalui TikTok untuk pemberdayaan kampung digital', JPM: Jurnal Pengabdian Masyarakat, 5(4). Tersedia di: <https://doi.org/10.47065/jpm.v5i4.2417>
- Agustiah, D., Fauzi, T. and Ramadhani, E. (2020) 'Dampak penggunaan media sosial terhadap perilaku belajar siswa', Islamic Counseling, 4(2), p. 181. Tersedia di: <https://doi.org/10.29240/jbk.v4i2.1935>
- Basti, B. and Solihin, N. A. (2023) 'Hubungan antara anonimitas dengan disinhibisi online pada mahasiswa di universitas', Journal of Art, Humanity and Social Studies, (6)
- Davidson, T. dkk. (2017) “Automated Hate Speech Detection and the Problem of Offensive Language,” Proceedings of the International AAAI Conference on Web and Social Media, 11(1), hlm. 512–515. Tersedia pada: <https://doi.org/10.1609/icwsm.v11i1.14955>.
- Ellys, R. A., Wahid, A. and Lamada, M. (2025) 'Pengaruh penyebaran berita hoax di media sosial TikTok terhadap perilaku digital mahasiswa generasi Z PTIK Universitas Negeri Makassar', Variable Research Journal, 02

- González-Aguilar, J. M., Segado-Boj, F. and Makhortykh, M. (2023) 'Populist right parties on TikTok: spectacularization, personalization, and hate speech', *Media and Communication*, 11(2), pp. 232–240. Tersedia di: <https://doi.org/10.17645/mac.v11i2.6358>
- Hidayah, A., Marcelawati, Y. and Saputra, H. (2021) 'Cyber harassment: fenomena hate comment di era pandemi Covid-19 pada akun Tik-Tok @y***q', *Jurnal Masyarakat Maritim*, 5(1), pp. 10–17. Tersedia di: <https://doi.org/10.31629/jmm.v5i1.3419>
- Hidayat, R. M., Rifki, A. and Rachman, I. F. (2025) 'Dampak paparan konten negatif di TikTok dan Instagram terhadap kesehatan mental siswa: kajian literatur', *Jurnal Riset Sosial Humaniora dan Pendidikan*, 3(3), pp. 136–143. Tersedia di: <https://doi.org/10.62383/risoma.v3i3.769>
- Hutamy, E. T., Swartika, F., Alisyahbana, A. N. Q. A., Arisah, N. and Hasan, M. (2021) 'Penelitian dan pengabdian inovatif pada masa pandemi Covid-19', dalam *Prosiding Seminar Nasional Penelitian dan Pengabdian*.
- Khairina, J. dkk. (2022) "Fenomena Echo Chamber di Media Sosial dan Dampaknya Terhadap Polarisasi Politik bagi Mahasiswa," *Bulan Tahun Journal of Civics and Education Studies*, 9(2).
- Lahinda, A., Moningkey, E. R. S. and Tangkawarow, I. R. H. T. (2025) 'Deteksi ujaran kebencian pada media sosial terkait Pemilu 2024 menggunakan algoritma Naive Bayes', *JMP*, 14. Tersedia di: <https://doi.org/10.33395/jmp.v14i2.15665>
- Lasser, J., Herderich, A., Garland, J., Aroyehun, S. T., Garcia, D. and Galesic, M. (2025) 'Collective moderation of hate, toxicity, and extremity in online discussions', *PNAS Nexus*, 4, p. pgaf369. Tersedia di: <https://doi.org/10.1093/pnasnexus/pgaf369>
- Lisdinani, A., Afifah, A. N., Restiana, C. and Supriyono, M. B. A. A. D. (2023) 'Pengaruh media sosial "TikTok" terhadap wawasan kebangsaan pada generasi muda', *PENDEKAR: Jurnal Pendidikan Berkarakter*, 1(6), pp. 232–241. Tersedia di: <https://doi.org/10.51903/pendekar.v1i6.508>
- Mansyur, M. and Muannas, M. (2020) 'Model literasi digital untuk melawan ujaran kebencian di media sosial digital literacy model to counter hate speech on social media', *Jurnal Ilmu Pengetahuan dan Teknologi Komunikasi*, 22(2). Tersedia di: <https://doi.org/10.33164/iptekkom.22.2.2020.125-142>
- Nufus, R., Maharani, W. and Zahro, E. B. (2022) 'Capture the psychological image of youth using TikTok applications', *ICHSS: International Conference of Humanities and Social Science*, 1(1), pp. 547–551
- Pangestu, W.A., Abdurrachman, H. dan Rizkianto, K. (2024) *Perlindungan terhadap Korban Tindak Pidana Hate Speech dalam Sistem Hukum Indonesia*. Disunting oleh T.V. Widyastuti. Pekalongan: PT Nasya Expanding Management.
- Prasetyo, R. and Nugraha, F. (2023) 'Classification of hate speech in TikTok comments using Naive Bayes and TF-IDF', *Journal of Machine Learning and Technology*, 2(1),

- "Maraknya *Hate Speech* Pada Platform TikTok dan Dampaknya Bagi Generasi Muda"
pp. 10–18. Tersedia di:
<https://journal.educollabs.org/index.php/JMTT/article/view/102>
- Pujangkara, D. A. T. (2025) Personal branding akun TikTok @VMuliana (studi analisis isi kuantitatif edukasi personal branding periode Januari-Juni 2025).
- Rahman, A. and Putri, D. A. (2024) 'Hate speech on TikTok: a forensic linguistic study', *Journal of English Language and Linguistics*, 5(2), pp. 45–60. Tersedia di:
https://www.researchgate.net/publication/394388305_Hate_Speech_on_Tiktok_A_Forensic_Linguistic_Study
- Regina Lestari, D. and Ahkam Alwi, M. (2024) 'Hubungan online disinhibition effect dengan kecenderungan perilaku cyberbullying pada siswa MA pengguna Instagram', *Jurnal Cendekia Ilmiah*, 4(1)
- Sarumaha, M., Manik, R. M., Borus, A. V. B. and Sitepu, A. B. (2024) 'Hubungan penggunaan media sosial dengan kesehatan mental remaja', *Jurnal Ilmu Kesehatan Bhakti Husada: Health Sciences Journal*, 15(02), pp. 333–342. Tersedia di: <https://doi.org/10.34305/jikbh.v15i02.1245>
- Suler, J. (2004) "The Online Disinhibition Effect," *CyberPsychology & Behavior*, 7(3), hlm. 321–326. Tersedia pada: <https://doi.org/10.1089/1094931041291295>
- Twenge, J. M., Haidt, J., Lozano, J. and Cummins, K. M. (2022) 'Specification curve analysis shows that social media use is linked to poor mental health, especially among girls', *Acta Psychologica*, 224, p. 103512. Tersedia di: <https://doi.org/10.1016/j.actpsy.2022.103512>
- Weimann, G. and Masri, N. (2023) 'Research note: spreading hate on TikTok', *Studies in Conflict and Terrorism*, 46(5), pp. 752–765. Tersedia di: <https://doi.org/10.1080/1057610X.2020.1780027>
- Yuhana, E. S., Mariyati, M. and Sugiyanto, E. P. (2023) 'Penggunaan media sosial dengan kesehatan mental remaja', *Jurnal Keperawatan Jiwa*, 11(2), p. 477. Tersedia di: <https://doi.org/10.26714/jkj.11.2.2023.477-486>