Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10, 2026
E-ISSN: 2988-1986
Open Access:



ANTARA KEBEBASAN BERBICARA DAN PERLINDUNGAN REPUTASI: TANTANGAN MODERASI DI MEDIA SOSIAL

Ravaylan Riza Pangestu[1], Izzatush Sa'diyah[2], Saeful Mujab[3]

[123] Program Studi Ilmu Komunikasi, Fakultas Ilmu Komunikasi, Universitas Bhayangkara Jakarta Raya

ARTICLE INFO

Article history:
Received Januari, 2026
Revised Jamuari, 2026
Accepted Januari, 2026
Available online Januari, 2026

rzpngsgt666@gmail.com

This is an open access article under the CC BY-SA license.
Copyright © 2023 by Author. Published by Universitas Pendidikan Ganesha.

A B S T R A K

Artikel ini membahas tantangan moderasi konten di media sosial dalam konteks kebebasan berekspresi dan perlindungan reputasi publik. Seiring dengan pesatnya perkembangan media sosial sebagai platform untuk berbagi informasi dan pendapat, muncul kebutuhan untuk menyeimbangkan hak kebebasan berbicara dengan perlindungan terhadap reputasi individu maupun organisasi. Penelitian ini mengkaji peran platform digital dalam moderasi konten, serta tantangan yang dihadapi dalam memastikan konten yang beredar tidak merusak reputasi publik. Artikel ini juga membahas kebijakan yang ada, termasuk regulasi yang mengatur kebebasan berbicara dan upaya perlindungan reputasi di dunia maya. Melalui pendekatan kualitatif deskriptif, artikel ini menyajikan analisis tentang kebijakan moderasi konten yang efektif dan saran-saran untuk meningkatkan keseimbangan antara kebebasan berbicara dan perlindungan reputasi di media sosial.

Kata Kunci: Kebebasan Berekspresi, Perlindungan Reputasi, Moderasi Konten, Media Sosial, Platform Digital, Disinformasi, Ujaran Kebencian, Regulasi Digital, Etika Digital.

A B S T R A C T

This article discusses the challenges of content moderation on social media in the context of freedom of expression and public reputation protection. As social media platforms rapidly develop as spaces for sharing information and opinions, there is a growing need to balance the right to free speech with the protection of individuals and organizations' reputations. This study examines the role of digital platforms in content moderation and the challenges they face in ensuring that the content circulating does not harm public reputation. The article also explores existing policies, including regulations governing freedom of speech and efforts to protect reputation in the digital world. Through a descriptive qualitative approach, this article provides an analysis of effective content moderation policies and suggestions for improving the balance between freedom of expression and reputation protection on social media.

***Keywords**: Freedom of Expression, Reputation Protection, Content Moderation, Social Media, Digital Platforms, Disinformation, Hate Speech, Digital Regulation, Digital Ethics.*

1. PENDAHULUAN

Definisi Kebebasan Berekspresi dalam Konteks Media Sosial

Kebebasan berekspresi merupakan hak fundamental yang telah mengalami transformasi signifikan seiring dengan perkembangan teknologi digital dan platform media sosial. Dalam konteks demokrasi kontemporer, kebebasan berekspresi tidak lagi terbatas pada media

*Corresponding author
E-mail addresses: rzpngst666@gmail.com



tradisional, melainkan telah meluas ke ranah digital yang memungkinkan partisipasi publik secara masif. Demokrasi saat ini menuntut masyarakat untuk secara kolektif terlibat dalam proses pembelajaran melalui konten media sosial yang dimoderasi secara efektif sesuai dengan standar nasional dan global terkait kebebasan berekspresi (Demuyakor & Doe, 2022). Hal ini menunjukkan bahwa kebebasan berekspresi di era digital memiliki dimensi yang lebih kompleks dibandingkan dengan era sebelumnya, di mana setiap individu memiliki kapasitas untuk menjadi produsen dan konsumen informasi secara simultan.

Namun demikian, penelitian menunjukkan bahwa terdapat berbagai isu kunci mengenai bagaimana media sosial telah digunakan untuk mendestabilisasi kebebasan berbicara dan berekspresi (Demuyakor & Doe, 2022). Fenomena ini mengindikasikan bahwa kebebasan berekspresi di platform digital menghadapi tantangan yang berbeda dari media konvensional, termasuk potensi penyalahgunaan yang dapat mengancam hak-hak individu lainnya. Platform media sosial seperti TikTok telah menjadi subjek pengawasan dan kritik yang meningkat, bahkan mengalami pelarangan sementara maupun permanen di beberapa negara termasuk India, Indonesia, dan Pakistan karena alasan terkait konten (Zeng & Kaye, 2022).

Pentingnya Perlindungan Reputasi Publik di Dunia Digital

Perlindungan reputasi di era digital menjadi semakin krusial mengingat dampak yang dapat ditimbulkan oleh konten yang beredar di media sosial. Kerusakan reputasi (reputational damage) merupakan salah satu konsekuensi serius yang dapat dialami pengguna media sosial, dan penelitian menunjukkan bahwa kerusakan reputasi ini seringkali diperkirakan secara berlebihan oleh individu (Iwatani & Watamura, 2024). Studi relevan menemukan bahwa orang-orang yang sangat sensitif terhadap penolakan lebih cenderung mempersepsikan perilaku ambigu orang lain sebagai penolakan yang disengaja, yang pada gilirannya dapat memperkuat kecanduan media sosial mereka (Iwatani & Watamura, 2024). Temuan ini mengimplikasikan bahwa individu cenderung melebih-lebihkan kemungkinan ditolak atau mengalami kerusakan reputasi, yang menunjukkan betapa pentingnya perlindungan reputasi dalam ekosistem digital.

Meskipun telah ada kemajuan dalam moderasi konten, platform media sosial masih menghadapi kesulitan untuk memoderasi perilaku online secara efektif (Schoenebeck et al., 2023). Pengguna terlibat dalam ujaran kebencian, penghinaan, ancaman, mempermalukan, berbagi gambar tanpa persetujuan, dan doxxing, dan banyak dari perilaku-perilaku ini lolos dari filter sistem moderasi (Schoenebeck et al., 2023). Kondisi ini menunjukkan bahwa perlindungan reputasi publik di dunia digital memerlukan pendekatan yang lebih komprehensif dan sistematis.

Kebutuhan untuk Moderasi Konten yang Seimbang antara Kebebasan Berbicara dan Perlindungan Reputasi

Kebutuhan akan mekanisme yang memastikan pengawasan efektif terhadap konten media sosial di berbagai platform telah diakui oleh semua pihak (Demuyakor & Doe, 2022). Hal ini mencerminkan kesadaran kolektif bahwa moderasi konten merupakan elemen esensial dalam menjaga keseimbangan antara kebebasan berekspresi dan perlindungan terhadap hak-hak individu lainnya, termasuk reputasi.

Dalam konteks ini, posisi tim moderator sebagai penyedia layanan moderasi konten yang ditunjuk oleh platform seperti TikTok tidak dapat dipisahkan dari kepentingan korporasi (Wismashanti, 2023). Kepentingan korporasi merujuk pada keuntungan dan kepentingan bisnis yang dikejar oleh platform media sosial, di mana kepentingan tersebut dapat mempengaruhi bagaimana platform melakukan moderasi konten (Wismashanti, 2023). Sebagai contoh, platform media sosial mungkin memprioritaskan konten yang dianggap populer atau kontroversial untuk meningkatkan jumlah pengguna dan keterlibatan di platform tersebut (Wismashanti, 2023).

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



Fenomena ini menimbulkan pertanyaan serius tentang objektivitas dan keadilan dalam proses moderasi konten.

Lebih lanjut, konsep visibility moderation telah diperkenalkan sebagai proses di mana platform digital memanipulasi jangkauan konten yang dihasilkan pengguna melalui cara algoritmik atau regulasi (Zeng & Kaye, 2022). Konsep ini menunjukkan bahwa moderasi konten tidak hanya tentang penghapusan atau pemblokiran konten, tetapi juga tentang pengaturan visibilitas yang dapat mempengaruhi sejauh mana suatu konten dapat menjangkau audiens. Pendekatan ini memiliki implikasi signifikan terhadap keseimbangan antara kebebasan berekspresi dan perlindungan reputasi, karena dapat digunakan baik untuk melindungi individu dari konten berbahaya maupun untuk membatasi ekspresi yang sah.

TUJUAN PENELITIAN

Membahas Tantangan Moderasi Konten dalam Media Sosial

Moderasi konten di platform media sosial merupakan salah satu isu paling kompleks dalam tata kelola digital kontemporer. Tantangan ini muncul dari berbagai dimensi yang saling berkaitan, mulai dari aspek teknis, normatif, hingga organisasional. Setiap sistem moderasi mengimplikasikan konsepsi operasional tertentu tentang bagaimana menyeimbangkan nilai-nilai yang saling bersaing, yang pada gilirannya mempengaruhi bagaimana "partisipasi politik diwujudkan melalui medium percakapan" (Rieder & Skop, 2021). Hal ini menunjukkan bahwa moderasi konten bukan sekadar persoalan teknis, melainkan juga menyangkut pertimbangan filosofis dan politis yang mendalam tentang bagaimana ruang publik digital seharusnya dikelola.
Salah satu tantangan utama dalam moderasi konten adalah volume material yang sangat besar yang harus ditangani oleh moderator konten (content moderators/CMs). Volume material yang sangat besar ini mengharuskan adanya kebijakan dan protokol yang mengutamakan kecepatan dan efisiensi dalam penyensoran, yang oleh beberapa pihak dianggap berkontribusi pada respons yang berlebihan sehingga melampaui batas dan pada akhirnya melanggar kebebasan berbicara secara umum (Bleakley et al., 2023). Kekhawatiran atas peran moderator konten yang dipersepsikan sebagai sensor online yang tidak teregulasi sedemikian besarnya sehingga beberapa negara bagian di Amerika Serikat telah berupaya memperkenalkan undang-undang "yang akan mencegah layanan digital dari memoderasi konten berbahaya untuk melindungi penggunanya" dengan sangat membatasi jenis konten yang boleh dimoderasi oleh perusahaan teknologi, dengan alasan kebebasan berbicara (Bleakley et al., 2023).

Tantangan lain yang tidak kalah signifikan adalah bagaimana faktor-faktor yang berbeda mempengaruhi persepsi terhadap moderasi konten online dan apakah efek-efek tersebut konsisten lintas negara (Gonçalves et al., 2021). Penelitian dalam bidang ini berkontribusi pada teori tentang keengganan algoritmik (algorithmic aversion) dan pemrosesan ujaran kebencian, serta memberikan wawasan praktis bagi platform online dalam menangani ujaran kebencian dan kata-kata kasar (Gonçalves et al., 2021). Temuan ini mengindikasikan bahwa tantangan moderasi konten tidak hanya bersifat universal, tetapi juga memiliki dimensi kultural dan kontekstual yang perlu dipertimbangkan dalam merancang kebijakan moderasi yang efektif.

Platform media sosial seperti Facebook, X (sebelumnya dikenal sebagai Twitter), dan YouTube telah menjadi pusat diskursus publik, memungkinkan tingkat partisipasi dan interaksi yang belum pernah terjadi sebelumnya (Zukić & Zukić, 2024). Meskipun platform-platform ini telah mendemokratisasi penyebaran informasi, mereka juga telah menjadi tempat berkembang biaknya misinformasi dan pencemaran nama baik, yang menantang kerangka hukum yang ada (Zukić & Zukić, 2024). Kondisi ini menambah kompleksitas tantangan moderasi konten, karena

*Corresponding author
E-mail addresses: rzpngst666@gmail.com



platform harus mampu membedakan antara ekspresi yang sah, meskipun kontroversial, dengan konten yang benar-benar berbahaya dan melanggar hukum.

Menganalisis Hubungan antara Kebebasan Berekspresi dan Perlindungan Reputasi Publik

Hubungan antara kebebasan berekspresi dan perlindungan reputasi publik merupakan dialektika yang kompleks dan seringkali menimbulkan ketegangan. Meskipun moderasi dapat dipandang sebagai eksklusioner jika dinilai sebagai penyensoran terhadap kebebasan berbicara, moderasi juga dapat dipahami sebagai inklusioner ketika mempertimbangkan berbagai cara suara-suara dibungkam melalui kerugian tortius dan martabat (tortious and dignitary harms) tanpa adanya intervensi (Rieder & Skop, 2021). Perspektif ini menawarkan pemahaman yang lebih nuansa tentang peran moderasi konten, yang tidak semata-mata membatasi ekspresi, tetapi juga dapat berfungsi untuk melindungi dan memfasilitasi partisipasi yang lebih inklusif dalam ruang publik digital.

Dalam konteks ini, organisasi penyedia konten online perlu mengambil sikap yang lebih proaktif dalam mencegah komunikasi online yang berbahaya (harmful online communication/HOC) dan melindungi penggunanya, daripada mengandalkan regulasi pemerintah atau kesopanan pengguna semata (Einwiller & Kim, 2020). Hal ini mencakup komunikasi kebijakan terkait HOC yang lebih efektif dengan pengguna, serta inisiatif proaktif untuk mengedukasi pengguna tentang hal-hal yang boleh dan tidak boleh dilakukan terkait HOC (Einwiller & Kim, 2020). Pendekatan ini menekankan pentingnya tanggung jawab bersama (shared responsibility) dalam menangani isu-isu sosial terkait komunikasi online yang berbahaya.

Prinsip fundamental yang harus dipegang adalah memupuk kesopanan (civility) dan tidak menerapkan kontrol yang berlebihan terhadap pengguna, yang perlu diberikan hak kebebasan berbicara mereka dalam batas-batas tidak membahayakan hak orang lain atas keselamatan dan martabat (Einwiller & Kim, 2020). Formulasi ini memberikan kerangka konseptual yang berguna untuk memahami bagaimana keseimbangan antara kebebasan berekspresi dan perlindungan reputasi dapat dicapai: kebebasan berekspresi dilindungi sejauh tidak melanggar hak-hak fundamental orang lain, termasuk hak atas keselamatan dan martabat yang mencakup perlindungan reputasi.

Dengan demikian, artikel ini bertujuan untuk mengeksplorasi secara mendalam bagaimana tantangan-tantangan moderasi konten dapat diatasi sambil tetap menjaga keseimbangan yang tepat antara kebebasan berekspresi dan perlindungan reputasi publik. Analisis ini akan mempertimbangkan berbagai perspektif, mulai dari pendekatan teknis-algoritmik hingga kerangka normatif dan regulatori, untuk memberikan pemahaman yang komprehensif tentang isu yang sangat relevan di era digital ini.

2. METODE

Rasionalisasi Pemilihan Metode

Penelitian mengenai moderasi konten dan keseimbangan antara kebebasan berekspresi dengan perlindungan reputasi publik memerlukan pendekatan metodologis yang mampu menangkap kompleksitas fenomena sosial-digital secara mendalam. Pendekatan deskriptif kualitatif dipilih karena kemampuannya untuk mengeksplorasi nuansa dan konteks yang tidak dapat ditangkap melalui metode kuantitatif semata. Peneliti perlu terus mengeksplorasi kebijakan konten platform dan tindakan penegakan yang dilaporkan sebagai sarana untuk memahami bagaimana platform berupaya memasarkan diri mereka secara menguntungkan kepada berbagai aktor lainnya (Buckley & Schafer, 2022). Rekomendasi ini menggarisbawahi pentingnya

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



penelitian yang bersifat eksploratif dan deskriptif dalam memahami dinamika moderasi konten di era digital.

Lebih lanjut, penelitian lebih dalam ke area kebijakan konten dan tindakan moderasi, seperti rasionalisasi platform dan perantara untuk keputusan penegakan, sangat diperlukan (Buckley & Schafer, 2022). Hal ini menunjukkan bahwa metodologi kualitatif yang berfokus pada analisis kebijakan dan praktik implementasi merupakan pendekatan yang tepat untuk mengungkap bagaimana platform digital mengelola ketegangan antara kebebasan berekspresi dan perlindungan reputasi.

Kerangka Analisis Multi-Platform

Penelitian ini mengadopsi kerangka analisis yang mencakup berbagai platform digital untuk memperoleh pemahaman yang komprehensif tentang fenomena moderasi konten. Pendekatan ini didasarkan pada pemahaman bahwa dalam konteks online, 'keamanan' harus dipertimbangkan dalam kaitannya dengan motivasi etis dan komersial platform media sosial (Flynn et al., 2025). Konsep ini merujuk pada teori tata kelola platform (platform governance theory), di mana platform dibentuk dan didorong oleh kepentingan korporasi dan komersial, serta menyeimbangkan komitmen ideologis terhadap kebebasan berbicara dengan perilaku yang berpotensi berbahaya yang berdampak pada pengalaman pengguna terkait 'keamanan' (Flynn et al., 2025).

Analisis multi-platform ini mempertimbangkan bahwa meskipun didorong oleh keuntungan, perusahaan media sosial semakin menghadapi tekanan reputasional dan regulatori untuk memastikan keamanan pengguna (Flynn et al., 2025). Bukti menunjukkan bahwa retensi dan kepercayaan pengguna juga terkait erat dengan persepsi keadilan prosedural (procedural justice) dalam bagaimana platform membuat keputusan seputar moderasi, kebijakan, dan kesejahteraan pengguna (Flynn et al., 2025). Paparan berkelanjutan terhadap bahaya dapat mendorong pengguna menjauh dari ruang-ruang ini, mengurangi keterlibatan dan nilai iklan, sehingga berdampak negatif pada keuntungan platform media sosial (Flynn et al., 2025).

Dimensi Analisis Kebijakan dan Implementasi

Metodologi penelitian ini juga mencakup analisis terhadap kerangka regulatori yang mempengaruhi praktik moderasi konten. Dalam konteks ini, ketentuan penting dalam Digital Services Act (DSA) akan mengubah lanskap regulasi moderasi konten (Quintais et al., 2023). Secara krusial, Pasal 14 DSA menetapkan aturan baru tentang bagaimana platform dapat menegakkan Syarat dan Ketentuan (Terms and Conditions) mereka, termasuk bahwa platform harus memberikan "perhatian yang semestinya" (due regard) terhadap "hak-hak fundamental" pengguna berdasarkan Piagam Hak-Hak Fundamental Uni Eropa (Quintais et al., 2023). Kerangka regulatori ini memberikan konteks normatif yang penting untuk menganalisis bagaimana platform menyeimbangkan kebebasan berekspresi dengan perlindungan reputasi.

Selain itu, penelitian ini mempertimbangkan perspektif moderator profesional dalam menangani komentar pengguna yang tidak sopan (uncivil) dan peran alat teknologi berbasis kecerdasan buatan (Stockinger et al., 2023). Beberapa negara memiliki undang-undang serupa untuk meregulasi kolom komentar platform online, seperti "Kommunikationsplattformen-Gesetz" (KoPl-G) di Austria dan "Netzwerkdurchsuchungsgesetz" (NetzDG) di Jerman (Stockinger et al., 2023). Sebagai bagian dari undang-undang ini, jenis komentar tertentu yang melanggar perlindungan konstitusional atas martabat manusia, mencemarkan nama baik, atau mengancam individu, mempromosikan ujaran kebencian, menghasut kekerasan atau kejahatan, atau

*Corresponding author
E-mail addresses: rzpngst666@gmail.com

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11 No 10, 2026
E-ISSN: 2988-1986
Open Access:



mempropagandakan ideologi Nazi dianggap melanggar hukum dan harus dihapus dari platform online dalam waktu 24 jam (Stockinger et al., 2023).

Pertimbangan Perspektif Pengguna

Metodologi penelitian ini juga mengintegrasikan perspektif pengguna sebagai elemen penting dalam memahami dinamika moderasi konten. Penelitian menunjukkan bahwa preferensi untuk pengelolaan mandiri (self-management) terhadap ujaran berbahaya kemungkinan dapat dikaitkan dengan kurangnya kepercayaan terhadap platform, frustrasi dengan keputusan moderasi yang diberlakukan platform, dan sifat pengalaman media sosial yang sangat personal (Jhaver & Zhang, 2023). Temuan ini mendukung skema di mana baik moderasi platform maupun moderasi personal tersedia dan kuat (Jhaver & Zhang, 2023).

Untuk memitigasi risiko pemblokiran berlebihan (overblocking), beberapa negara telah mengimplementasikan prosedur pengaduan yang terinstitusionalisasi bagi pengguna yang berkepentingan (Stockinger et al., 2023). Dengan demikian, moderator harus membuat keputusan cepat namun mempertimbangkannya dengan hati-hati untuk mencegah tuduhan penyensoran (Stockinger et al., 2023). Aspek ini menjadi pertimbangan penting dalam metodologi penelitian untuk memahami bagaimana keseimbangan antara kecepatan dan ketelitian dalam moderasi konten dapat dicapai.

Sintesis Metodologis

Secara keseluruhan, metodologi penelitian deskriptif kualitatif ini dirancang untuk menghasilkan pemahaman yang holistik tentang tantangan moderasi konten dalam konteks keseimbangan antara kebebasan berekspresi dan perlindungan reputasi publik. Pendekatan ini mengintegrasikan analisis kebijakan platform, kerangka regulatori, perspektif moderator profesional, dan pengalaman pengguna untuk memberikan gambaran yang komprehensif tentang fenomena yang diteliti. Dengan demikian, metodologi ini memungkinkan eksplorasi mendalam terhadap berbagai dimensi moderasi konten di berbagai platform digital, sekaligus mempertimbangkan konteks sosial, hukum, dan teknologi yang membentuk praktik-praktik tersebut.

Kebebasan Berekspresi di Media Sosial

Pembahasan tentang Hak Kebebasan Berbicara Menurut Deklarasi Internasional

Kebebasan berekspresi merupakan hak asasi manusia yang fundamental dan telah diakui secara universal dalam berbagai instrumen hukum internasional. Secara internasional, kebebasan berekspresi diatur dalam Universal Declaration of Human Rights (UDHR) serta International Covenant on Civil and Political Rights (ICCPR), di mana hak-hak ini diatur dalam Pasal 19 dari UDHR dan Pasal 19 serta Pasal 20 dari ICCPR (Sabubun & Hammar, 2024). Pengaturan ini menunjukkan bahwa komunitas internasional telah lama mengakui pentingnya kebebasan berekspresi sebagai salah satu pilar utama dalam perlindungan hak asasi manusia dan pembangunan masyarakat demokratis.

Dalam konteks nasional Indonesia, perlindungan terhadap hak atas kebebasan pribadi diatur dalam Pasal 20-43 Undang-Undang Nomor 39 Tahun 1999 tentang Hak Asasi Manusia, yang mencakup hak untuk tidak diperbudak, hak atas kebebasan beragama, hak untuk memilih dan dipilih, hak untuk berkumpul dan berserikat, hak untuk menyatakan pendapat, hak atas kewarganegaraan, dan hak untuk bertempat tinggal (Sabubun & Hammar, 2024). Kerangka hukum ini memberikan landasan normatif yang kuat bagi perlindungan kebebasan berekspresi di Indonesia, sekaligus menunjukkan komitmen negara terhadap standar internasional hak asasi manusia.

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



Namun demikian, kebebasan berekspresi bukanlah hak yang absolut. Sesuai dengan ketentuan yang berlaku, terdapat berbagai jenis konten yang dapat dibatasi untuk melindungi kepentingan publik dan hak-hak individu lainnya (Wismashanti, 2023). Kategorisasi ini menunjukkan bahwa kebebasan berekspresi memiliki batasan-batasan yang ditetapkan untuk melindungi kepentingan masyarakat.

Media Sosial sebagai Platform untuk Ekspresi

Mengapa Media Sosial Memberikan Ruang Besar untuk Kebebasan Berbicara

Media sosial telah mengalami evolusi yang signifikan dalam dua dekade terakhir dan telah menjadi platform utama bagi ekspresi publik. Sepanjang periode yang diamati, peran media sosial sebagai enabler untuk interaksi manusia serta sebagai sarana untuk terhubung dengan pengguna lain telah menjadi konstanta dalam mendefinisikan media sosial (Aichner et al., 2021). Dalam definisi-definisi awal, fokus utamanya adalah pada orang dan bagaimana orang berinteraksi, sedangkan definisi-definisi yang lebih baru (setelah 2010) sebagian besar mendefinisikan media sosial sebagai sumber daya online apa pun yang dirancang untuk memfasilitasi keterlibatan antar individu (Aichner et al., 2021).

Karakteristik unik media sosial yang memungkinkan setiap pengguna untuk menjadi produsen dan konsumen konten secara simultan telah menciptakan demokratisasi dalam penyebaran informasi. Platform media sosial seperti Facebook, X (sebelumnya dikenal sebagai Twitter), dan YouTube telah menjadi pusat diskursus publik, memungkinkan tingkat partisipasi dan interaksi yang belum pernah terjadi sebelumnya (Zukić & Zukić, 2024). Fenomena ini menandai pergeseran paradigma dari model komunikasi satu-ke-banyak (one-to-many) yang mendominasi era media tradisional menuju model banyak-ke-banyak (many-to-many) yang lebih inklusif dan partisipatif.

Selain itu, demokrasi saat ini menuntut masyarakat untuk secara kolektif terlibat dalam proses pembelajaran melalui konten media sosial yang dimoderasi secara efektif sesuai dengan standar nasional dan global terkait kebebasan berekspresi (Demuyakor & Doe, 2022). Hal ini mengindikasikan bahwa media sosial tidak hanya berfungsi sebagai platform untuk ekspresi individual, tetapi juga sebagai ruang publik digital di mana warga negara dapat berpartisipasi dalam proses demokrasi dan pembentukan opini publik.

Dampak Positif terhadap Demokrasi, Pengembangan Opini Publik, dan Pertukaran Informasi

Kebebasan berekspresi di media sosial memberikan kontribusi signifikan terhadap penguatan demokrasi dan pengembangan opini publik. Meskipun platform-platform ini telah mendemokratisasi penyebaran informasi, mereka juga telah menjadi tempat berkembang biaknya misinformasi dan pencemaran nama baik, yang menantang kerangka hukum yang ada (Zukić & Zukić, 2024). Pernyataan ini, meskipun mengakui tantangan yang ada, secara implisit mengonfirmasi bahwa demokratisasi informasi merupakan salah satu manfaat utama dari kebebasan berekspresi di media sosial.

Dalam konteks pembentukan opini publik, penelitian menunjukkan bahwa jurnalis mengandalkan media sosial dalam pelaporan mereka tentang opini 'publik', meskipun menyadari ketidakrepresentatifan komentar media sosial (Hopmann & Schuck, 2021). Fenomena ini menunjukkan bahwa media sosial telah menjadi sumber penting dalam membentuk dan merefleksikan opini publik, meskipun dengan berbagai keterbatasan yang perlu dipertimbangkan. Prevalensi ekspresi yang beragam di media sosial sebagian dapat menjelaskan bagaimana jurnalis

*Corresponding author
E-mail addresses: rzpngst666@gmail.com



menilai opini publik (Hopmann & Schuck, 2021), yang mengindikasikan peran signifikan media sosial dalam proses pembentukan dan artikulasi opini publik.

Lebih lanjut, kebebasan berekspresi di media sosial memfasilitasi pertukaran informasi yang lebih cepat dan luas. Prinsip fundamental yang harus dipegang adalah memupuk kesopanan dan tidak menerapkan kontrol yang berlebihan terhadap pengguna, yang perlu diberikan hak kebebasan berbicara mereka dalam batas-batas tidak membahayakan hak orang lain atas keselamatan dan martabat (Einwiller & Kim, 2020). Formulasi ini menekankan pentingnya keseimbangan antara kebebasan berekspresi dan tanggung jawab sosial dalam ekosistem digital.

Tantangan Kebebasan di Media Sosial

Konten yang Memicu Kontroversi, Ujaran Kebencian, atau Hoaks yang Merusak

Meskipun kebebasan berekspresi di media sosial membawa banyak manfaat, terdapat tantangan serius yang harus dihadapi. Sudut pandang di media sosial juga telah diintegrasikan ke dalam kerangka narasi strategis untuk membantah dan menekan sudut pandang yang berlawanan (Ni, 2023). Sebagaimana diperkirakan oleh Herman dan Chomsky dalam cetak ulang 2002 dari "Manufacturing Consent", kedua sarjana tersebut percaya bahwa konvergensi internet dan teknologi baru akan menyelesaikan beberapa hambatan terhadap demokratisasi media, tetapi mereka berargumen bahwa hak untuk berbicara di media sosial masih dikontrol dalam kerangka opini publik yang dibangun bersama oleh media massa dan elit sosial (Ni, 2023). Perspektif kritis ini menunjukkan bahwa kebebasan berekspresi di media sosial tidak sepenuhnya bebas dari pengaruh struktural dan kekuasaan.

Tantangan lain yang signifikan adalah maraknya konten berbahaya yang beredar di platform digital. Pengguna terlibat dalam ujaran kebencian, penghinaan, ancaman, mempermalukan, berbagi gambar tanpa persetujuan, dan doxxing, dan banyak dari perilaku-perilaku ini lolos dari filter sistem moderasi (Schoenebeck et al., 2023). Kondisi ini menunjukkan bahwa meskipun telah ada kemajuan dalam moderasi konten, platform media sosial masih menghadapi kesulitan untuk memoderasi perilaku online secara efektif (Schoenebeck et al., 2023).

Selain pemerintah, kewajiban regulatori juga dapat berasal dari badan pengawas atau organisasi non-pemerintah (Wismashanti, 2023). Pihak-pihak ini dapat menetapkan persyaratan dan standar tertentu yang harus dipatuhi oleh platform media sosial dalam melaksanakan moderasi konten (Wismashanti, 2023). Kerangka multi-stakeholder ini mencerminkan kompleksitas tantangan yang dihadapi dalam mengelola kebebasan berekspresi di media sosial, di mana berbagai aktor memiliki peran dan tanggung jawab yang berbeda dalam memastikan bahwa kebebasan berekspresi tidak disalahgunakan untuk menyebarkan konten yang berbahaya.

Dalam konteks ini, penelitian menunjukkan bahwa terdapat berbagai isu kunci mengenai bagaimana media sosial telah digunakan untuk mendestabilisasi kebebasan berbicara serta ekspresi (Demuyakor & Doe, 2022). Fenomena ini menggarisbawahi paradoks yang melekat dalam kebebasan berekspresi di era digital: platform yang sama yang memungkinkan demokratisasi informasi juga dapat digunakan untuk menyebarkan disinformasi, ujaran kebencian, dan konten yang merusak reputasi individu maupun kelompok.

3. HASIL DAN PEMBAHASAN

Penjelasan tentang Reputasi Pribadi dan Institusional dalam Dunia Digital

Reputasi dalam konteks digital merupakan konstruksi sosial yang mencerminkan persepsi publik terhadap individu maupun institusi berdasarkan informasi yang tersedia dan beredar di ruang maya. Dalam era digital kontemporer, reputasi tidak lagi terbentuk semata-mata melalui interaksi langsung, melainkan juga melalui jejak digital yang terakumulasi di berbagai platform

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



online. Penelitian menunjukkan bahwa kerusakan reputasi (reputational damage) merupakan salah satu konsekuensi serius yang dapat dialami pengguna media sosial, dan kerusakan reputasi ini seringkali diperkirakan secara berlebihan oleh individu (Iwatani & Watamura, 2024). Temuan ini mengindikasikan bahwa reputasi digital memiliki dimensi psikologis yang signifikan, di mana persepsi tentang potensi kerusakan reputasi dapat mempengaruhi perilaku dan kesejahteraan mental pengguna media sosial.

Dalam konteks hukum Indonesia, perlindungan reputasi diakui sebagai bagian integral dari perlindungan hak asasi manusia. Pengaturan pencemaran nama baik dalam UU ITE masih menghadapi dilema antara perlindungan reputasi individu dan jaminan kebebasan berekspresi sebagai hak konstitusional (Zahsy, 2025). Hal ini menunjukkan bahwa reputasi, baik pribadi maupun institusional, merupakan nilai yang dilindungi secara hukum dan memiliki kedudukan yang setara dengan hak-hak fundamental lainnya dalam sistem hukum nasional.

Lebih lanjut, reputasi institusional juga menjadi pertimbangan penting dalam ekosistem digital. Meskipun didorong oleh keuntungan, perusahaan media sosial semakin menghadapi tekanan reputasional dan regulatori untuk memastikan keamanan pengguna (Flynn et al., 2025). Bukti menunjukkan bahwa retensi dan kepercayaan pengguna juga terkait erat dengan persepsi keadilan prosedural dalam bagaimana platform membuat keputusan seputar moderasi, kebijakan, dan kesejahteraan pengguna (Flynn et al., 2025). Fenomena ini mengilustrasikan bahwa reputasi institusional platform digital sangat bergantung pada bagaimana mereka mengelola konten dan melindungi penggunanya dari bahaya.

Bahaya Pencemaran Nama Baik, Penyebaran Informasi Palsu, dan Dampaknya terhadap Individu atau Organisasi

Konten berisiko di media sosial dapat menimbulkan dampak destruktif yang signifikan terhadap reputasi individu maupun organisasi. Pencemaran nama baik merupakan suatu perbuatan kejahatan yang bertujuan untuk mencemarkan nama baik seseorang dalam wujud tertulis ataupun tidak tertulis dan akan menimbulkan akibat buruk yang dapat merugikan seseorang bagi yang namanya tercemar (Akbar et al., 2023). Definisi ini menegaskan bahwa pencemaran nama baik bukan sekadar pelanggaran etika, melainkan merupakan tindak pidana yang memiliki konsekuensi hukum yang serius.

Platform media sosial telah menjadi tempat berkembang biaknya misinformasi dan pencemaran nama baik, yang menantang kerangka hukum yang ada (Zukić & Zukić, 2024). Kondisi ini diperparah oleh fakta bahwa pengguna terlibat dalam ujaran kebencian, penghinaan, ancaman, mempermalukan, berbagi gambar tanpa persetujuan, dan doxxing, dan banyak dari perilaku-perilaku ini lolos dari filter sistem moderasi (Schoenebeck et al., 2023). Ketidakmampuan sistem moderasi untuk mendeteksi dan menghapus konten berbahaya secara efektif meningkatkan risiko kerusakan reputasi bagi korban.

Dampak psikologis dari konten berisiko juga tidak dapat diabaikan. Studi relevan menemukan bahwa orang-orang yang sangat sensitif terhadap penolakan lebih cenderung mempersepsikan perilaku ambigu orang lain sebagai penolakan yang disengaja, yang pada gilirannya dapat memperkuat kecanduan media sosial mereka (Iwatani & Watamura, 2024). Temuan ini mengimplikasikan bahwa individu cenderung melebih-lebihkan kemungkinan ditolak atau mengalami kerusakan reputasi (Iwatani & Watamura, 2024), yang menunjukkan bahwa dampak psikologis dari ancaman terhadap reputasi dapat melampaui kerugian aktual yang dialami.

*Corresponding author
E-mail addresses: rzpngst666@gmail.com

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11 No 10, 2026
E-ISSN: 2988-1986
Open Access:



Dalam konteks penyebaran informasi palsu, penelitian menunjukkan bahwa isu hoaks yang paling banyak beredar di media sosial selama pandemi Covid-19 adalah hoaks SARA, kesehatan, dan politik, di mana hoaks SARA dan hoaks politik memberikan efek afektif kepada masyarakat (Tenriawali et al., 2020). Fenomena ini mengilustrasikan bagaimana informasi palsu tidak hanya merusak reputasi individu atau institusi tertentu, tetapi juga dapat mempengaruhi kondisi psikologis dan sosial masyarakat secara lebih luas.

Pembahasan tentang UU ITE di Indonesia dan Perlindungan Hukum terhadap Reputasi di Dunia Maya

Indonesia telah mengembangkan kerangka hukum yang komprehensif untuk melindungi reputasi di dunia maya, dengan Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) sebagai instrumen utamanya. Penerapan hukum terhadap tindak pidana pencemaran nama baik menggunakan asas Lex specialis derogat legi generali, di mana pengaturan pencemaran nama baik di dunia maya yang diatur dalam Pasal 27 ayat (3) dan Pasal 45 UU ITE merupakan "Lex specialis" dari Pasal 310 KUHP yang merupakan "Lex generali", dan hubungan aturan ini menjadikan sinergi hukum atas kasus pencemaran nama baik (Efendi, 2020). Konstruksi hukum ini menunjukkan bahwa pembuat undang-undang telah mengantisipasi kebutuhan akan regulasi khusus untuk menangani pencemaran nama baik di ranah digital.

Namun demikian, implementasi UU ITE dalam konteks pencemaran nama baik tidak lepas dari kontroversi. Implementasi penegakan hukum sering menimbulkan kontroversi karena interpretasi yang beragam terhadap unsur-unsur delik dan standar pembuktian yang tidak jelas (Zahsy, 2025). Penelitian menemukan bahwa diperlukan reformulasi pengaturan yang lebih proporsional dengan kriteria yang lebih spesifik untuk membedakan kritik yang dilindungi konstitusi dengan pencemaran nama baik yang dapat dipidana (Zahsy, 2025). Rekomendasi ini mencerminkan kebutuhan akan keseimbangan yang lebih baik antara perlindungan reputasi dan jaminan kebebasan berekspresi.

Secara internasional, kebebasan berekspresi diatur dalam Universal Declaration of Human Rights (UDHR) serta International Covenant on Civil and Political Rights (ICCPR), di mana hak-hak ini diatur dalam Pasal 19 dari UDHR dan Pasal 19 serta Pasal 20 dari ICCPR (Sabubun & Hammar, 2024). Kerangka hukum internasional ini memberikan konteks normatif yang penting untuk memahami bagaimana perlindungan reputasi harus diseimbangkan dengan hak atas kebebasan berekspresi.

Pengaturan Terkait Pencemaran Nama Baik dan Cara Platform Digital Merespons

Platform digital memiliki tanggung jawab signifikan dalam merespons konten yang berpotensi mencemarkan nama baik. Beberapa negara memiliki undang-undang untuk meregulasi kolom komentar platform online, seperti "Kommunikationsplattformen-Gesetz" (KoPl-G) di Austria dan "Netzwerkdurchsuchungsgesetz" (NetzDG) di Jerman (Stockinger et al., 2023). Sebagai bagian dari undang-undang ini, jenis komentar tertentu yang melanggar perlindungan konstitusional atas martabat manusia, mencemarkan nama baik, atau mengancam individu, mempromosikan ujaran kebencian, menghasut kekerasan atau kejahatan, atau mempropagandakan ideologi Nazi dianggap melanggar hukum dan harus dihapus dari platform online dalam waktu 24 jam (Stockinger et al., 2023).

Dalam konteks Uni Eropa, ketentuan penting dalam Digital Services Act (DSA) akan mengubah lanskap regulasi moderasi konten (Quintais et al., 2023). Secara krusial, Pasal 14 DSA menetapkan aturan baru tentang bagaimana platform dapat menegakkan Syarat dan Ketentuan mereka, termasuk bahwa platform harus memberikan "perhatian yang semestinya" terhadap "hak-hak fundamental" pengguna berdasarkan Piagam Hak-Hak Fundamental Uni Eropa (Quintais et al.,

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



2023). Kerangka regulatori ini memberikan landasan yang lebih kuat bagi perlindungan reputasi pengguna di platform digital.

Di Indonesia, terdapat dua belas kelompok konten yang dikategorikan sebagai konten negatif, termasuk pornografi, perjudian, pemerasan, penipuan, kekerasan, pencemaran nama baik, SARA, terorisme/radikalisme elektronik yang melanggar hukum lainnya (Wismashanti, 2023). Selain pemerintah, kewajiban regulatori juga dapat berasal dari badan pengawas atau organisasi non-pemerintah, di mana pihak-pihak ini dapat menetapkan persyaratan dan standar tertentu yang harus dipatuhi oleh platform media sosial dalam melaksanakan moderasi konten (Wismashanti, 2023). Kerangka multi-stakeholder ini mencerminkan pendekatan komprehensif dalam menangani perlindungan reputasi di dunia maya.

Untuk memitigasi risiko pemblokiran berlebihan (overblocking), beberapa negara telah mengimplementasikan prosedur pengaduan yang terinstitusionalisasi bagi pengguna yang berkepentingan (Stockinger et al., 2023). Dengan demikian, moderator harus membuat keputusan cepat namun mempertimbangkannya dengan hati-hati untuk mencegah tuduhan penyensoran (Stockinger et al., 2023). Mekanisme ini penting untuk memastikan bahwa perlindungan reputasi tidak disalahgunakan untuk membungkam kritik yang sah dan dilindungi oleh hak atas kebebasan berekspresi.

MODERASI KONTEN DAN TANGGUNG JAWAB PLATFORM DIGITAL

Cara Media Sosial Melakukan Moderasi

Platform media sosial telah mengembangkan sistem moderasi konten yang kompleks untuk mengelola volume besar konten yang diunggah pengguna setiap harinya. Platform media sosial seperti Facebook, X (sebelumnya dikenal sebagai Twitter), dan YouTube telah menjadi pusat diskursus publik, memungkinkan tingkat partisipasi dan interaksi yang belum pernah terjadi sebelumnya (Zukić & Zukić, 2024). Namun, peran sentral ini juga membawa tanggung jawab besar dalam memastikan bahwa konten yang beredar tidak melanggar standar komunitas maupun ketentuan hukum yang berlaku.

Proses moderasi konten melibatkan kombinasi antara sistem otomatis dan intervensi manusia. Meskipun telah ada kemajuan dalam moderasi konten, platform media sosial masih menghadapi kesulitan untuk memoderasi perilaku online secara efektif (Schoenebeck et al., 2023). Pengguna terlibat dalam ujaran kebencian, penghinaan, ancaman, mempermalukan, berbagi gambar tanpa persetujuan, dan doxxing, dan banyak dari perilaku-perilaku ini lolos dari filter sistem moderasi (Schoenebeck et al., 2023). Kondisi ini menunjukkan bahwa meskipun platform telah menginvestasikan sumber daya yang signifikan dalam moderasi konten, tantangan yang dihadapi tetap substansial.

Dalam konteks ini, konsep visibility moderation telah diperkenalkan sebagai proses di mana platform digital memanipulasi jangkauan konten yang dihasilkan pengguna melalui cara algoritmik atau regulasi (Zeng & Kaye, 2022). Pendekatan ini menunjukkan bahwa moderasi tidak hanya tentang penghapusan konten, tetapi juga tentang pengaturan sejauh mana konten tertentu dapat menjangkau audiens yang lebih luas.

Kebijakan Moderasi Konten di Berbagai Platform

Setiap platform digital memiliki kebijakan moderasi yang berbeda-beda, yang tercermin dalam Syarat dan Ketentuan (Terms and Conditions) masing-masing. Ketentuan penting dalam Digital Services Act (DSA) akan mengubah lanskap regulasi moderasi konten, di mana Pasal 14 DSA menetapkan aturan baru tentang bagaimana platform dapat menegakkan Syarat dan Ketentuan

*Corresponding author
E-mail addresses: rzpngst666@gmail.com



mereka, termasuk bahwa platform harus memberikan "perhatian yang semestinya" terhadap "hak-hak fundamental" pengguna (Quintais et al., 2023). Kerangka regulatori ini memberikan standar minimum yang harus dipenuhi oleh platform dalam menjalankan kebijakan moderasinya.

Di tingkat nasional, berbagai negara telah mengembangkan regulasi spesifik untuk mengatur moderasi konten. Beberapa negara memiliki undang-undang untuk meregulasi kolom komentar platform online, seperti "Kommunikationsplattformen-Gesetz" (KoPl-G) di Austria dan "Netzwerkdurchsuchungsgesetz" (NetzDG) di Jerman (Stockinger et al., 2023). Sebagai bagian dari undang-undang ini, jenis komentar tertentu yang melanggar perlindungan konstitusional atas martabat manusia, mencemarkan nama baik, atau mengancam individu dianggap melanggar hukum dan harus dihapus dari platform online dalam waktu 24 jam (Stockinger et al., 2023).

Di Indonesia, terdapat dua belas kelompok konten yang dikategorikan sebagai konten negatif, termasuk pornografi, perjudian, pemerasan, penipuan, kekerasan, pencemaran nama baik, SARA, dan terorisme/radikalisme elektronik (Wismashanti, 2023). Kategorisasi ini memberikan panduan bagi platform dalam mengidentifikasi dan menangani konten yang berpotensi melanggar hukum.

TANTANGAN YANG DIHADAPI PLATFORM

Menyeimbangkan antara Kebebasan Berbicara dan Hak untuk Dilindungi dari Pencemaran Nama Baik

Tantangan utama yang dihadapi platform digital adalah menemukan keseimbangan yang tepat antara melindungi kebebasan berekspresi dan mencegah penyalahgunaan yang dapat merusak reputasi individu. Meskipun moderasi dapat dipandang sebagai eksklusioner jika dinilai sebagai penyensoran terhadap kebebasan berbicara, moderasi juga dapat dipahami sebagai inklusioner ketika mempertimbangkan berbagai cara suara-suara dibungkam melalui kerugian tortius dan martabat tanpa adanya intervensi (Rieder & Skop, 2021). Perspektif dialektis ini menunjukkan kompleksitas dalam menentukan kapan dan bagaimana moderasi harus dilakukan.

Volume material yang sangat besar mengharuskan adanya kebijakan dan protokol yang mengutamakan kecepatan dan efisiensi dalam penyensoran, yang oleh beberapa pihak dianggap berkontribusi pada respons yang berlebihan sehingga melampaui batas dan pada akhirnya melanggar kebebasan berbicara secara umum (Bleakley et al., 2023). Kekhawatiran atas peran moderator konten yang dipersepsikan sebagai sensor online yang tidak teregulasi sedemikian besarnya sehingga beberapa negara bagian di Amerika Serikat telah berupaya memperkenalkan undang-undang yang akan mencegah layanan digital dari memoderasi konten berbahaya (Bleakley et al., 2023).

Prinsip fundamental yang harus dipegang adalah memupuk kesopanan dan tidak menerapkan kontrol yang berlebihan terhadap pengguna, yang perlu diberikan hak kebebasan berbicara mereka dalam batas-batas tidak membahayakan hak orang lain atas keselamatan dan martabat (Einwiller & Kim, 2020). Formulasi ini memberikan kerangka konseptual untuk menyeimbangkan kedua kepentingan yang saling bersaing tersebut.

Tantangan lain yang signifikan adalah bagaimana faktor-faktor yang berbeda mempengaruhi persepsi terhadap moderasi konten online dan apakah efek-efek tersebut konsisten lintas negara (Gonçalves et al., 2021). Penelitian dalam bidang ini berkontribusi pada teori tentang keengganan algoritmik dan pemrosesan ujaran kebencian, serta memberikan wawasan praktis bagi platform online dalam menangani ujaran kebencian dan kata-kata kasar (Gonçalves et al., 2021).

ALGORITMA DAN KECERDASAN BUATAN DALAM MODERASI

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



Bagaimana Algoritma Memoderasi Konten dan Peran Kecerdasan Buatan dalam Pengelolaan Konten

Penggunaan algoritma dan kecerdasan buatan (AI) dalam moderasi konten telah menjadi komponen integral dari strategi platform digital dalam mengelola konten skala besar. Setiap sistem moderasi mengimplikasikan konsepsi operasional tertentu tentang bagaimana menyeimbangkan nilai-nilai yang saling bersaing, yang pada gilirannya mempengaruhi bagaimana "partisipasi politik diwujudkan melalui medium percakapan" (Rieder & Skop, 2021). Pernyataan ini menggarisbawahi bahwa desain algoritma moderasi bukan sekadar keputusan teknis, melainkan juga keputusan normatif yang memiliki implikasi politik dan sosial.

Penelitian menunjukkan pentingnya memahami perspektif moderator profesional dalam menangani komentar pengguna yang tidak sopan dan peran alat teknologi berbasis kecerdasan buatan (Stockinger et al., 2023). Moderator harus membuat keputusan cepat namun mempertimbangkannya dengan hati-hati untuk mencegah tuduhan penyensoran (Stockinger et al., 2023). Hal ini menunjukkan bahwa meskipun AI dapat membantu dalam mengidentifikasi konten bermasalah, keputusan akhir seringkali masih memerlukan pertimbangan manusia.

Posisi tim moderator sebagai penyedia layanan moderasi konten yang ditunjuk oleh platform tidak dapat dipisahkan dari kepentingan korporasi (Wismashanti, 2023). Kepentingan korporasi merujuk pada keuntungan dan kepentingan bisnis yang dikejar oleh platform media sosial, di mana kepentingan tersebut dapat mempengaruhi bagaimana platform melakukan moderasi konten (Wismashanti, 2023). Sebagai contoh, platform media sosial mungkin memprioritaskan konten yang dianggap populer atau kontroversial untuk meningkatkan jumlah pengguna dan keterlibatan di platform tersebut (Wismashanti, 2023). Fenomena ini menimbulkan pertanyaan tentang objektivitas algoritma moderasi dan potensi bias yang mungkin tertanam dalam sistem tersebut.

Untuk memitigasi risiko pemblokiran berlebihan (overblocking), beberapa negara telah mengimplementasikan prosedur pengaduan yang terinstitusionalisasi bagi pengguna yang berkepentingan (Stockinger et al., 2023). Mekanisme ini penting untuk memastikan bahwa keputusan algoritmik dapat ditinjau dan dikoreksi ketika terjadi kesalahan. Selain itu, preferensi untuk pengelolaan mandiri terhadap ujaran berbahaya kemungkinan dapat dikaitkan dengan kurangnya kepercayaan terhadap platform dan frustrasi dengan keputusan moderasi yang diberlakukan platform (Jhaver & Zhang, 2023). Temuan ini mendukung skema di mana baik moderasi platform maupun moderasi personal tersedia dan kuat (Jhaver & Zhang, 2023), yang mengindikasikan perlunya pendekatan hybrid yang menggabungkan moderasi algoritmik dengan kontrol pengguna individual.

TANTANGAN DALAM MENJAGA KESEIMBANGAN

Pada media sosial, menjaga keseimbangan antara kebebasan berekspresi dan perlindungan reputasi publik merupakan tantangan besar. Platform digital harus memfasilitasi kebebasan berbicara, tetapi pada saat yang sama mereka juga perlu melindungi individu atau entitas dari pencemaran nama baik dan penyebaran informasi palsu. Beberapa tantangan yang muncul dalam menjaga keseimbangan ini antara lain:

Peran Regulasi yang Tidak Konsisten

Meskipun banyak negara sudah memiliki regulasi untuk menangani masalah reputasi di dunia maya, penerapannya sering kali tidak konsisten. Misalnya, di Indonesia, UU ITE (Undang-Undang Informasi dan Transaksi Elektronik) digunakan untuk mengatasi penyebaran hoaks atau

*Corresponding author
E-mail addresses: rzpngst666@gmail.com



pencemaran nama baik. Namun, implementasi hukum ini terkadang disalahgunakan untuk mengekang kebebasan berekspresi, seperti dalam kasus yang melibatkan kritik terhadap pemerintah. Hal ini menunjukkan bahwa regulasi yang ada belum sepenuhnya berhasil menyeimbangkan kebebasan berbicara dengan perlindungan reputasi .

Kebijakan Moderasi Konten yang Tidak Efektif

Platform digital seringkali kesulitan untuk menerapkan kebijakan moderasi yang efektif dan konsisten. Kebijakan yang terlalu ketat dapat memadamkan kebebasan berbicara, sementara kebijakan yang longgar dapat membiarkan penyebaran konten berbahaya atau fitnah. Sebagai contoh, Facebook dan Twitter telah menghadapi kritik karena gagal dalam menangani ujaran kebencian dan disinformasi, sementara pada saat yang sama mendapat kritik karena moderasi yang dianggap terlalu sensitif atau membatasi kebebasan ekspresi .

Peran Algoritma dan Kecerdasan Buatan

Penggunaan algoritma dan kecerdasan buatan dalam moderasi konten memiliki kelebihan dan kekurangan. Meskipun algoritma dapat memproses data dalam jumlah besar dengan cepat, ia sering gagal dalam memahami konteks sosial dan budaya di balik konten yang diunggah. Hal ini dapat menyebabkan kesalahan dalam moderasi, seperti menghapus konten yang sah atau gagal mengidentifikasi konten yang berbahaya. Sehingga, meskipun teknologi ini dapat membantu, ia tetap tidak bisa sepenuhnya menggantikan peran manusia dalam moderasi .

Solusi dan Kebijakan yang Dapat Diterapkan

Untuk mengatasi tantangan yang dihadapi dalam menyeimbangkan kebebasan berekspresi dan perlindungan reputasi di media sosial, berbagai solusi dan kebijakan perlu diterapkan. Beberapa langkah yang dapat dilakukan adalah sebagai berikut:

Penyusunan Regulasi yang Lebih Komprehensif

Untuk mencapai keseimbangan yang lebih baik, perlu adanya regulasi yang lebih jelas dan lebih tepat dalam mengatur kebebasan berekspresi dan perlindungan reputasi. Regulasi ini harus memperjelas batasan kebebasan berbicara dan hak perlindungan atas reputasi, serta memberikan pedoman yang lebih terperinci tentang bagaimana platform digital harus bertindak dalam menangani konten yang melanggar norma .

Sebagai contoh, Uni Eropa telah mengembangkan kebijakan yang lebih ketat melalui Digital Services Act (DSA), yang mengatur kewajiban platform digital untuk menangani konten ilegal dan melindungi pengguna dari disinformasi, sambil tetap menghormati kebebasan berbicara. Pendekatan seperti ini bisa menjadi model bagi negara-negara lain untuk mengadopsi regulasi yang lebih seimbang .

Mengoptimalkan Peran Moderasi Manusia

Meskipun algoritma memiliki peran dalam memoderasi konten, peran manusia dalam moderasi tetap sangat penting. Platform digital perlu memastikan bahwa moderasi dilakukan dengan pendekatan yang lebih manusiawi, di mana tim moderasi bisa menilai konten secara kontekstual. Penerapan "human-in-the-loop" (HITL) model dapat memperbaiki hasil moderasi dengan menggabungkan algoritma dan penilaian manusia .

Dengan memperkenalkan lebih banyak tinjauan manusia dalam proses moderasi, platform dapat meminimalkan kesalahan dalam penghapusan konten yang sah dan memastikan bahwa keputusan moderasi tidak bersifat bias.

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



Edukasi Pengguna tentang Etika Digital

Masyarakat perlu diberi pemahaman yang lebih baik tentang etika digital dan tanggung jawab dalam menggunakan media sosial. Edukasi ini harus mencakup pentingnya menghormati reputasi orang lain, serta memahami batasan kebebasan berekspresi di dunia maya. Platform dapat bekerja sama dengan lembaga pendidikan dan organisasi non-pemerintah untuk menjalankan program edukasi yang mengedepankan literasi digital .

Penggunaan Teknologi yang Lebih Canggih dalam Moderasi

Teknologi kecerdasan buatan (AI) terus berkembang, dan harus ada upaya untuk mengembangkan sistem moderasi yang lebih canggih dan kontekstual. Platform dapat berinvestasi dalam pengembangan AI yang lebih adaptif dan peka terhadap konteks, yang dapat mengenali nuansa dan budaya di balik setiap pesan yang diposting, serta mempertimbangkan kebebasan berekspresi dalam setiap keputusan moderasi .

Selain itu, platform bisa mengembangkan fitur yang memungkinkan pengguna untuk melaporkan konten yang dianggap merugikan reputasi mereka, dengan mekanisme verifikasi yang adil dan transparan.

4. KESIMPULAN

Perdebatan antara kebebasan berbicara dan perlindungan reputasi di media sosial merupakan salah satu tantangan terbesar dalam era digital saat ini. Kebebasan berekspresi adalah hak fundamental yang dilindungi oleh berbagai konvensi internasional, dan media sosial memberikan ruang bagi individu untuk mengekspresikan pendapat serta ide secara bebas. Namun, tanpa moderasi yang tepat, kebebasan ini dapat disalahgunakan, mengarah pada penyebaran konten yang merusak reputasi individu atau kelompok, seperti hoaks, fitnah, atau ujaran kebencian.

Perlindungan reputasi publik, di sisi lain, juga merupakan aspek penting yang tidak bisa diabaikan. Meskipun penting untuk melindungi hak individu atas reputasi mereka, penerapan perlindungan tersebut harus mempertimbangkan kebebasan berbicara sebagai elemen penting dalam demokrasi. Pencemaran nama baik, penyebaran informasi palsu, dan manipulasi reputasi di media sosial dapat menyebabkan dampak yang sangat merugikan, baik bagi individu maupun organisasi, dan sering kali menimbulkan konsekuensi hukum.

Moderasi konten oleh platform digital menjadi kunci untuk menyeimbangkan kedua aspek ini. Namun, tantangan moderasi konten semakin kompleks dengan berkembangnya kecerdasan buatan dan algoritma yang digunakan oleh platform untuk menyaring informasi. Meskipun teknologi ini memberikan efisiensi, namun mereka juga sering kali gagal dalam mengidentifikasi konteks secara tepat, yang dapat mengarah pada ketidakseimbangan dalam moderasi. Selain itu, keputusan moderasi sering kali dipengaruhi oleh kebijakan internal platform, yang dapat berisiko mengabaikan keberagaman perspektif pengguna.

Penting bagi platform digital untuk mengembangkan kebijakan moderasi yang lebih transparan dan adil, dengan mempertimbangkan konteks dan hak-hak individu secara lebih holistik. Pemerintah dan regulator juga harus memainkan peran penting dalam merumuskan kebijakan yang melindungi hak-hak pengguna sekaligus memastikan bahwa kebebasan berbicara tidak tergerus. Regulasi yang baik dan moderasi konten yang efektif harus dapat memastikan bahwa kedua hal ini kebebasan berekspresi dan perlindungan reputasi dapat berjalan beriringan tanpa saling merugikan.

*Corresponding author
E-mail addresses: rzpngst666@gmail.com



Secara keseluruhan, untuk menciptakan ekosistem digital yang lebih sehat dan aman, perlu ada kerjasama yang lebih erat antara pihak platform digital, pemerintah, dan masyarakat. Hal ini bertujuan untuk menciptakan solusi moderasi konten yang dapat menyeimbangkan kebebasan berbicara dengan perlindungan reputasi, sekaligus menjamin bahwa hak setiap individu untuk berpartisipasi dalam ruang digital tetap dihormati dan dilindungi.

5. DAFTAR PUSTAKA

Abdussamad, Z. (2021). Metode Penelitian Kualitatif. Syakir Media Press.

Ann, J. (2001). Theories of Governance and New Public Management: Links to Understanding Welfare Policy Implementation. Igrass, 1, 1–5.

Bleakley, P., Martellozzo, E., Spence, R., & DeMarco, J. (2023). Moderating Online Child Sexual Abuse Material (CSAM): Does Self-Regulation Work, or Is Greater State Regulation Needed? European Journal of Criminology, 21(2), 231–250. https://doi.org/10.1177/14773708231181361

Demuyakor, J., & Doe, V. A. (2022). Social Media, Democracy, and Freedom of Expression: Some Evidence From Ghana. International Journal of Political Science and Governance, 4(1), 87–94. https://doi.org/10.33545/26646021.2022.v4.i1b.140

Dwiyanto, A. (2011). Mengembalikan Kepercayaan Publik Melalui Reformasi Birokrasi. Gramedia Pustaka Utama.

Einwiller, S., & Kim, S. (2020). How Online Content Providers Moderate User - Generated Content to Prevent Harmful Online Communication: An Analysis of Policies and Their Implementation. Policy & Internet, 12(2), 184–206. https://doi.org/10.1002/poi3.239

Farazmand, A. (2004). Sound Governance in the Age of Globalization: A Conceptual Framework. In A. Farzmand (Ed.), Sound Governance: Policy and Administrative Innovation. Preager.

Gash, C. A. (2008). Collaborative Governance in Theory and Practice. Journal of Public Administration Research and Theory, 18(4), 543–571.

Gonçalves, J., Weber, I., Chen, G. M., Silva, M. T. d., & Hofhuis, J. (2021). Common Sense or Censorship: How Algorithmic Moderators and Message Type Influence Perceptions of Online Content Deletion. New Media & Society, 25(10), 2595–2617. https://doi.org/10.1177/14614448211032310

Iwatani, S., & Watamura, E. (2024). Antecedents of Social Media Addiction in High and Low Relational Mobility Societies: Motivation to Expand Social Network and Fear of Reputational Damage. Plos One, 19(4), e0300681. https://doi.org/10.1371/journal.pone.0300681

Mukhtar Tompo, M. M. (2021). Penerapan Prinsip-Prinsip Good Governance Dalam Pelayanan Publik Di Rsud Lanto Daeng Pasewang Kabupaten Jeneponto. JPPM: Journal of Public Policy and Management, 3(1), 43–52.

Rieder, B., & Skop, Y. (2021). The Fabrics of Machine Moderation: Studying the Technical, Normative, and Organizational Structure of Perspective API. Big Data & Society, 8(2). https://doi.org/10.1177/20539517211046181

Triwikrama: Jurnal Multidisiplin Ilmu Sosial
Volume 11, Number 10 2026
E-ISSN: 2988-1986
Open Access:



Schoenebeck, S., Lampe, C., & Triệu, P. (2023). Online Harassment: Assessing Harms and Remedies. Social Media + Society, 9(1). https://doi.org/10.1177/20563051231157297

Sunu Astuti, R. (2020). Collaborative Governance Dalam Perspektif Administrasi Publik. Universitas Diponegoro Press.

Wicaksono, I. (2021). Integration of Development Programs Based on Collaborative Governance on Regional Village. Jurnal Politico, 21(1), 53–66.

Wismashanti, R. A. (2023). Social Media Content Moderation Challenges for Vulnerable Groups: A Case Study on Tiktok Indonesia. Eduvest - Journal of Universal Studies, 3(8), 1449–1464. https://doi.org/10.59188/eduvest.v3i8.893

Zeng, J., & Kaye, D. B. V. (2022). From Content Moderation to <i>visibility Moderation</I>: A Case Study of Platform Governance on TikTok. Policy & Internet, 14(1), 79–95. https://doi.org/10.1002/poi3.287

Zukić, M., & Zukić, A. (2024). Defamation Law and Media: Challenges of the Digital Age. Map Education and Humanities, 5(1), 98–109. https://doi.org/10.53880/2744-2373.2024.5.98

Zulkarnain, A., & Annisa, M. (2014). Komitmen Terhadap Organisasi Ditinjau Dari Kesejahteraan Psikologis Pekerja. 54–62.

*Corresponding author
E-mail addresses: rzpngst666@gmail.com