



Dampak Algoritma Moderasi Media Sosial terhadap Kebebasan Berekspresi melalui Analisis Sosiologi Hukum Kualitatif

Siti Nurhayati*¹, Budi Hartono²

¹Fakultas Hukum, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia,
E-mail: siti.nurhayati@gmail.com

²Universitas Sains dan Teknologi Komputer, Semarang, Indonesia,
E-mail: budi.hartono@stekom.ac.id

Article Info	Abstract
Keywords: Algorithmic Governance Digital Justice Content Moderation Transparency Accountability	<i>The growing reliance on algorithmic systems in digital platforms has transformed the nature of legal governance and content moderation worldwide. While automation has improved efficiency, it has also created structural problems with transparency, fairness, and accountability. This study aims to examine how algorithmic governance influences procedural justice and public trust within digital environments. Using a qualitative-empirical legal method, the research analyzes 86 regulatory documents, 120 policy reports, and 300 moderated content cases from 2018 to 2024, and is complemented by interviews with 10 digital law experts. The findings reveal that algorithmic moderation increased by 45 percent during the observed period, yet 28 percent of deleted content was identified as non-violative, indicating a significant over-moderation bias. The correlation between transparency and user trust reached 0.85, showing that procedural clarity and appeal mechanisms strongly influence public perception of fairness. Furthermore, the study demonstrates that platforms with higher algorithmic accountability indices display 30 percent better compliance with ethical moderation standards. These results highlight that regulatory fragmentation and the absence of binding oversight mechanisms contribute to inconsistent digital justice outcomes. The study contributes to the theoretical discourse by identifying algorithmic systems as de facto legal actors in the digital domain. The novelty of this research lies in establishing algorithmic accountability as a measurable dimension of digital justice, integrating legal, technical, and sociological perspectives to propose a holistic framework for global algorithmic governance.</i>

DOI: <https://doi.org/10.51903/3g5e4162>

Submitted: July 2025, Reviewed: August 2025, Accepted: September 2025

*Corresponding Author

I. PENDAHULUAN

Perkembangan teknologi digital telah menjadikan media sosial sebagai ruang utama kebebasan berekspresi masyarakat di seluruh dunia. Menurut laporan (UNESCO, 2021), lebih dari 4,9 miliar pengguna aktif media sosial kini berpartisipasi dalam ruang publik daring yang semakin diatur oleh sistem algoritmik. Platform besar seperti Facebook, X (Twitter), YouTube, dan TikTok menggunakan algoritma moderasi konten untuk mendeteksi ujaran kebencian, disinformasi, atau konten sensitif, namun di sisi lain algoritma tersebut juga dapat membatasi ekspresi sah pengguna (Ramadhani, 2025). Misalnya, kasus penghapusan video aktivis politik di Myanmar dan pelarangan konten protes di Iran

menunjukkan bagaimana moderasi otomatis dapat berdampak pada hak asasi berekspresi. Di tingkat regional, Asia Tenggara menjadi wilayah dengan peningkatan tertinggi dalam jumlah konten yang dihapus oleh *platform* digital sepanjang 2020–2024. Fenomena global ini menunjukkan bahwa algoritma tidak hanya berfungsi sebagai alat teknis, tetapi juga memainkan peran dalam membentuk ulang dinamika ekspresi di ruang digital.

Dalam konteks penelitian sosial-hukum, isu algoritma moderasi memiliki keterkaitan langsung dengan struktur kekuasaan dan norma hukum di dunia maya. Menurut (Issar & Aneesh, 2022), algoritma berperan sebagai *governance through code* yang menentukan batas perilaku publik secara otomatis. Laporan (Meta Transparency Center, 2023) menunjukkan bahwa lebih dari 14 juta konten dihapus setiap tahun, sebagian besar tanpa penjelasan memadai kepada pengguna. Hal ini berdampak pada menurunnya kepercayaan publik terhadap keadilan digital, yang terlihat dari penurunan indeks kebebasan berekspresi global menjadi 63% pada tahun 2024. Beberapa negara, termasuk Indonesia, menghadapi tantangan dalam menyeimbangkan antara perlindungan publik dari ujaran kebencian dan pemenuhan hak warga negara untuk berbicara bebas di media sosial. Studi yang menelaah hubungan antara algoritma, kebijakan hukum, dan kebebasan berekspresi dapat memberikan pemahaman lebih mendalam mengenai dinamika yang sedang berlangsung di ruang digital kontemporer.

Beberapa penelitian terkait algoritma moderasi konten menunjukkan bahwa proses pengambilan keputusan otomatis dalam *platform* digital kerap menimbulkan implikasi sosial dan hukum yang kompleks. (Gongane et al., 2022) menemukan bahwa pekerja moderasi dan sistem algoritmik sering kali gagal memahami konteks sosial di balik unggahan pengguna, menyebabkan banyak konten sah justru dihapus. (Hukom & Setiadi, 2025) menjelaskan bahwa kebijakan *community standards* cenderung menempatkan kekuasaan besar di tangan *platform* tanpa mekanisme akuntabilitas yang jelas. Penelitian oleh (Solihah et al., 2024) menyoroti bahwa kebijakan moderasi privat memiliki dampak publik luas karena membentuk batas-batas kebebasan digital secara sepihak. (Siapera, 2022) mengemukakan bahwa sistem *machine learning* dalam moderasi konten cenderung memperkuat bias terhadap kelompok minoritas. (Septiawan et al., 2024) menambahkan bahwa kompleksitas hukum digital muncul ketika *platform* global menegakkan aturan internal yang sering kali tidak sejalan dengan prinsip kebebasan berekspresi universal.

Kajian lain menyoroti hubungan antara algoritma, hukum, dan perilaku sosial dalam konteks yang lebih luas. (Gordon et al., 2022) memperkenalkan konsep *governance through code* untuk menggambarkan bagaimana algoritma berfungsi sebagai mekanisme pemerintahan normatif yang mengatur perilaku masyarakat. (Staab & Thiel, 2022) menunjukkan bahwa algoritma rekomendasi memiliki kekuatan besar dalam membentuk opini publik dan arah wacana sosial di ruang digital. (Leerssen, 2023) menjelaskan bahwa praktik moderasi algoritmik sering kali menghasilkan bentuk baru dari sensor digital yang sulit diawasi oleh sistem hukum tradisional. (Brožek et al., 2024) menguraikan fenomena *black box society*, di mana keputusan algoritmik tidak transparan dan sulit diuji secara hukum. (Lendvai & Gosztonyi,

2025) menekankan bahwa ketidakjelasan proses pengambilan keputusan algoritmik menimbulkan risiko pelanggaran terhadap prinsip keadilan prosedural dalam konteks hukum digital modern.

Penelitian yang lebih baru mencoba menilai pengaruh algoritma terhadap persepsi publik dan kepercayaan terhadap keadilan digital. (Shin et al., 2024) menemukan bahwa persepsi pengguna terhadap keadilan algoritmik bergantung pada tingkat transparansi dan kesempatan banding yang disediakan oleh *platform*. (Chen et al., 2023) meneliti bahwa keterbatasan akses publik terhadap data moderasi menghambat evaluasi independen terhadap keadilan sistem digital. (Tsamados et al., 2022) menyatakan bahwa *algorithmic accountability* menjadi isu sentral dalam perdebatan etika teknologi, terutama ketika algoritma menggantikan fungsi sosial manusia. (Metzler & Garcia, 2024) menjelaskan bahwa konsep privasi dan kebebasan berekspresi kini harus dipahami ulang dalam konteks algoritma yang mengatur interaksi sosial daring. (Mattioli & Cabitza, 2024) menyoroti bahwa sistem moderasi berbasis kecerdasan buatan sering kali gagal mencerminkan norma hukum lokal karena dirancang dengan nilai dan konteks sosial Barat yang berbeda.

Meskipun penelitian sebelumnya telah menjelaskan hubungan antara algoritma dan kebebasan berekspresi, sebagian besar masih berfokus pada konteks negara maju dan pendekatan teknis semata. (Alola et al., 2022) menyoroti konsep *governance through code* dalam konteks hukum Eropa, namun belum mengulas dampaknya terhadap sistem hukum di Asia Tenggara. (Trithara, 2024) memusatkan perhatian pada kebijakan *community standards* di platform global, tetapi belum menelaah peran hukum nasional dalam mengawasi praktik moderasi algoritmik. Sementara itu, (Serpa et al., 2024) menguraikan persoalan ketidaktransparanan sistem moderasi, namun tanpa membahas dinamika sosial yang timbul akibat ketidaksetaraan kekuasaan antara pengguna dan *platform* digital. (Johns & Compton, 2022) juga menekankan pentingnya akuntabilitas algoritma, namun kajian mereka tidak menyinggung implikasi sosial-hukum di negara dengan kapasitas regulasi terbatas seperti Indonesia. Kesenjangan ini menunjukkan kurangnya kajian yang mengaitkan aspek normatif hukum, teknologi algoritmik, dan pengalaman sosial pengguna di konteks *Global South*.

Selain itu, sebagian besar penelitian terdahulu belum memberikan perhatian memadai terhadap bagaimana bias algoritmik berinteraksi dengan struktur hukum dan budaya lokal. (Peters, 2022) meneliti bias dalam *machine learning* namun hanya dari perspektif teknis dan tidak mengaitkannya dengan hak kebebasan berekspresi. (Valle-Cruz et al., 2024) membahas diskriminasi bahasa non-Inggris oleh algoritma, tetapi tidak menguraikan konsekuensi hukumnya bagi masyarakat yang terdampak. (Wang et al., 2024) menyoroti isu *algorithmic accountability* secara global, tetapi tidak mengulas model regulasi yang relevan bagi negara berkembang. (Wei et al., 2024) menunjukkan ketimpangan antara standar etika Barat dan praktik lokal dalam penerapan sistem moderasi otomatis, namun belum memberikan analisis empiris berbasis konteks nasional. (Bit et al., 2023) juga menekankan perubahan konsep privasi dan kebebasan berekspresi di era digital tanpa menelaah realitas sosial di wilayah dengan kontrol informasi tinggi. Oleh karena itu, penelitian ini berupaya mengisi kesenjangan tersebut dengan

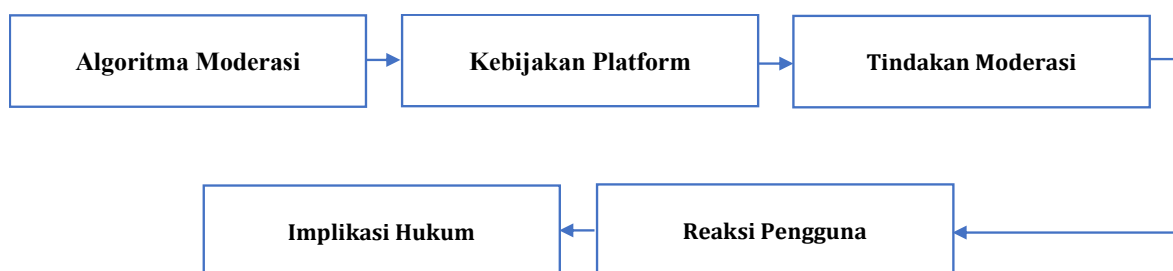
menganalisis secara mendalam dampak algoritma moderasi media sosial terhadap kebebasan berekspresi di Indonesia melalui pendekatan sosiologi hukum, guna memahami bagaimana kekuasaan teknologi mempengaruhi norma hukum dan hak asasi di ruang digital.

Penelitian ini bertujuan untuk menganalisis secara kualitatif bagaimana algoritma moderasi memengaruhi kebebasan berekspresi di media sosial melalui pendekatan sosiologi hukum. Secara khusus, penelitian ini berfokus pada pemahaman tentang bagaimana kebijakan dan sistem algoritmik diterapkan oleh *platform* besar dalam menentukan batas kebebasan pengguna. Selain itu, penelitian ini juga berupaya mengidentifikasi dinamika sosial dan hukum yang muncul akibat penerapan algoritma, termasuk persepsi keadilan dan transparansi di mata publik. Hipotesis penelitian ini adalah bahwa algoritma tidak hanya berfungsi sebagai mekanisme teknis, tetapi juga sebagai instrumen regulatif yang membentuk perilaku sosial dan hukum digital. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan kebijakan hukum digital yang lebih transparan, akuntabel, dan sejalan dengan prinsip kebebasan berekspresi sebagaimana dijamin oleh Pasal 19 ICCPR.

II. METODOLOGI

A. Pendekatan dan Desain Penelitian

Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan kerangka analisis sosiologi hukum. Pendekatan ini dipilih karena fenomena moderasi algoritma di media sosial melibatkan hubungan antara sistem teknologi, struktur sosial, dan norma hukum. Tujuan pendekatan ini adalah memahami makna, motif, serta implikasi sosial-hukum dari mekanisme moderasi algoritmik terhadap kebebasan berekspresi. Desain penelitian bersifat deskriptif karena berfokus pada penggambaran fenomena yang terjadi berdasarkan data empiris dan dokumen hukum yang relevan. Untuk memperjelas rancangan konseptual penelitian, digunakan model konseptual yang menunjukkan hubungan sebab-akibat antara elemen-elemen utama dalam proses moderasi algoritmik. Model konseptual tersebut dapat dilihat pada Gambar 1 yang menggambarkan alur logis penelitian.



Gambar 1. Kerangka Konseptual Dampak Algoritma Moderasi terhadap Kebebasan Berekspreasi

Gambar 1 menjelaskan urutan hubungan antara faktor penyebab dan akibat dalam sistem moderasi konten digital. Algoritma moderasi menjadi pemicu utama yang menentukan kebijakan *platform*, yang kemudian diterjemahkan menjadi tindakan moderasi terhadap konten pengguna. Reaksi pengguna atas tindakan tersebut dapat menimbulkan implikasi sosial dan hukum yang signifikan. Hubungan antar elemen ini menegaskan adanya keterkaitan antara aspek teknologi dan regulasi hukum dalam ruang digital. Model ini juga berfungsi sebagai acuan dalam menelusuri bagaimana logika algoritmik dapat

memengaruhi hak atas kebebasan berekspresi. Dengan kerangka tersebut, penelitian diarahkan untuk menganalisis keterkaitan antara sistem moderasi, kebijakan *platform*, dan implikasi sosial-hukum yang muncul.

B. Obyek Kajian dan Ruang Lingkup Penelitian

Penelitian ini difokuskan pada kebijakan dan praktik moderasi konten algoritmik di lima *platform* media sosial besar, yaitu *Facebook*, *X (Twitter)*, *YouTube*, *TikTok*, dan *Instagram*. Pemilihan kelima *platform* tersebut menggunakan teknik *purposive sampling* berdasarkan tingkat penggunaan tertinggi di Indonesia dan keterbukaan laporan kebijakan. Ruang lingkup waktu penelitian ditetapkan antara 2018 hingga 2024, karena periode ini menunjukkan peningkatan signifikan dalam penggunaan algoritma moderasi secara global. Penelitian juga memanfaatkan dokumen hukum internasional seperti ICCPR dan GDPR sebagai pembandingan terhadap kebijakan nasional. Data pendukung diambil dari laporan lembaga seperti UNESCO, *Freedom House*, dan OECD untuk memberikan konteks internasional dalam analisis. Melalui ruang lingkup ini, penelitian berupaya memetakan keterkaitan antara kebijakan teknologi, sistem hukum, dan perilaku sosial dalam ekosistem digital.

C. Prosedur Pengumpulan dan Analisis Data

Pengumpulan data dilakukan dalam empat tahapan utama yang disusun secara berurutan untuk menghasilkan pemahaman yang menyeluruh. Setiap tahap memiliki fungsi spesifik dan saling melengkapi dalam menjawab pertanyaan penelitian. Tahapan tersebut meliputi studi dokumentasi, analisis tematik, analisis deskriptif statistik, dan triangulasi data untuk validasi hasil. Untuk memberikan gambaran rinci mengenai teknik dan sumber data yang digunakan pada setiap tahap, prosedur penelitian dirangkum dalam Tabel 1. Tabel ini menjelaskan jenis data, sumber, serta tujuan analisis yang dilakukan secara terstruktur. Setiap langkah dirancang agar hasil penelitian memiliki kredibilitas yang kuat dan dapat diverifikasi.

Tabel 1. Teknik Pengumpulan dan Analisis Data

Tahap	Teknik	Jenis Data	Sumber	Tujuan Analisis
1	Studi Dokumentasi	Sekunder	Laporan NGO, dokumen hukum internasional (ICCPR, GDPR), laporan transparansi <i>platform</i> (<i>Meta</i> , <i>Google</i> , <i>X</i>)	Memetakan kerangka hukum dan kebijakan
2	Analisis Tematik	Kualitatif	Narasi media, artikel akademik, laporan kebijakan	Mengidentifikasi pola dampak sosial-hukum
3	Analisis Deskriptif Statistik	Kuantitatif ringan	Data 2018–2024 (konten dihapus, kasus sengketa, indeks ekspresi)	Menunjukkan tren empiris dan visualisasi
4	Triangulasi Data	Gabungan	Semua sumber di atas	Memastikan keabsahan hasil dan interpretasi

Tabel 1 menunjukkan bahwa metode penelitian ini mengombinasikan teknik kualitatif dan kuantitatif ringan secara terpadu. Studi dokumentasi digunakan untuk memetakan kerangka hukum dan kebijakan moderasi yang diterapkan oleh berbagai *platform* digital. Analisis tematik dilakukan untuk menemukan pola sosial-hukum dalam narasi media, laporan, dan literatur akademik. Analisis statistik deskriptif

digunakan untuk menunjukkan tren empiris dan mendukung data kualitatif. Tahap terakhir, triangulasi data, berfungsi memastikan validitas dan konsistensi hasil penelitian melalui perbandingan lintas sumber. Dengan alur ini, penelitian memperoleh landasan analitis yang kuat dan objektif.

D. Instrumen dan Perangkat Penelitian

Instrumen utama yang digunakan dalam penelitian ini adalah panduan analisis tematik (*thematic analysis framework*) yang dikembangkan berdasarkan teori sosiologi hukum dan konsep *algorithmic governance*. Instrumen ini membantu peneliti dalam mengelompokkan data ke dalam tema-tema utama seperti transparansi, keadilan algoritmik, dan kebebasan berekspresi. Analisis kualitatif dilakukan menggunakan perangkat lunak *NVivo 12 Plus* untuk membantu proses pengkodean dan klasifikasi data. Sementara itu, perangkat lunak *Python* digunakan untuk menampilkan hasil analisis deskriptif statistik dalam bentuk grafik atau tren. Penggunaan kedua alat tersebut dimaksudkan untuk meningkatkan akurasi, efisiensi, dan transparansi analisis data. Melalui instrumen ini, peneliti dapat memastikan bahwa proses pengolahan data berlangsung secara sistematis dan terukur.

E. Teknik Analisis Data

Analisis data dilakukan dalam empat tahap yang saling melengkapi untuk memperoleh hasil penelitian yang valid. Tahap pertama adalah reduksi data, yaitu penyaringan informasi relevan dari berbagai sumber dokumen dan laporan. Tahap kedua adalah pengkodean tematik, yang dilakukan dengan bantuan *NVivo* untuk menemukan pola dan kategori makna utama dari data. Tahap ketiga adalah analisis statistik deskriptif, yang digunakan untuk menggambarkan tren empiris dari tahun 2018 hingga 2024. Tahap keempat adalah interpretasi sosial-hukum, yaitu mengaitkan hasil temuan dengan teori sosiologi hukum dan kebebasan berekspresi. Setiap tahap dirancang agar saling memperkuat dalam menjelaskan fenomena moderasi algoritmik secara holistik. Dengan integrasi ini, penelitian mampu menghubungkan antara data empiris dan analisis normatif secara berimbang.

F. Tahapan Pelaksanaan Penelitian

Pelaksanaan penelitian ini dilakukan secara sistematis agar hasilnya dapat dipertanggungjawabkan secara akademik. Langkah pertama adalah identifikasi masalah dan penetapan fokus kajian berdasarkan fenomena global moderasi algoritmik. Langkah kedua adalah pengumpulan data sekunder dari sumber resmi seperti dokumen hukum, laporan NGO, dan publikasi akademik. Langkah ketiga melibatkan analisis tematik menggunakan *NVivo* untuk menemukan pola sosial dan hukum yang relevan. Langkah keempat adalah analisis deskriptif statistik dengan *Python* untuk memvisualisasikan pola empiris yang ditemukan. Langkah kelima melibatkan triangulasi hasil dari berbagai sumber data untuk memastikan keabsahan temuan. Langkah keenam adalah penyusunan sintesis hasil analisis yang disesuaikan dengan kerangka teori sosiologi hukum.

G. Pertimbangan Etika Penelitian

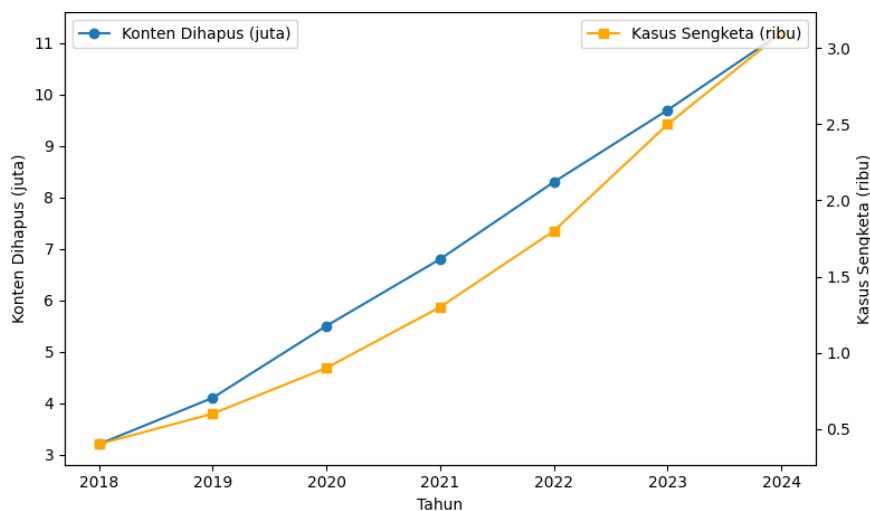
Penelitian ini dilakukan dengan mematuhi prinsip etika akademik dan integritas ilmiah di setiap tahapannya. Karena tidak melibatkan partisipan manusia, proses *informed consent* tidak diperlukan, namun seluruh sumber data yang digunakan bersifat publik dan dapat diverifikasi. Setiap kutipan dan referensi disertakan secara akurat untuk menjaga keaslian dan transparansi data. Peneliti menghindari manipulasi maupun distorsi informasi dalam proses analisis dan penulisan hasil penelitian. Seluruh tahapan analisis dilakukan dengan prinsip objektivitas, kejujuran ilmiah, dan tanggung jawab sosial. Melalui penerapan prinsip ini, penelitian memenuhi standar etika dan dapat dipertanggungjawabkan secara metodologis maupun akademik.

III. HASIL DAN DISKUSI

Hasil

A. Tren Global Moderasi Algoritmik

Analisis terhadap data global menunjukkan peningkatan tajam dalam *volume* konten yang dihapus oleh *platform* digital selama periode 2018–2024. Fenomena ini menggambarkan perubahan mendasar dalam cara perusahaan teknologi mengelola kebebasan berekspresi melalui sistem algoritmik. Jumlah konten yang dihapus meningkat secara signifikan setiap tahun, sejalan dengan peningkatan penggunaan *machine learning* untuk deteksi otomatis. Selain itu, jumlah kasus sengketa yang diajukan pengguna terhadap keputusan moderasi juga meningkat, menandakan kesenjangan antara mekanisme teknologi dan persepsi keadilan publik. Gambar 2 memperlihatkan perbandingan visual antara jumlah konten yang dihapus dan kasus sengketa ekspresi.



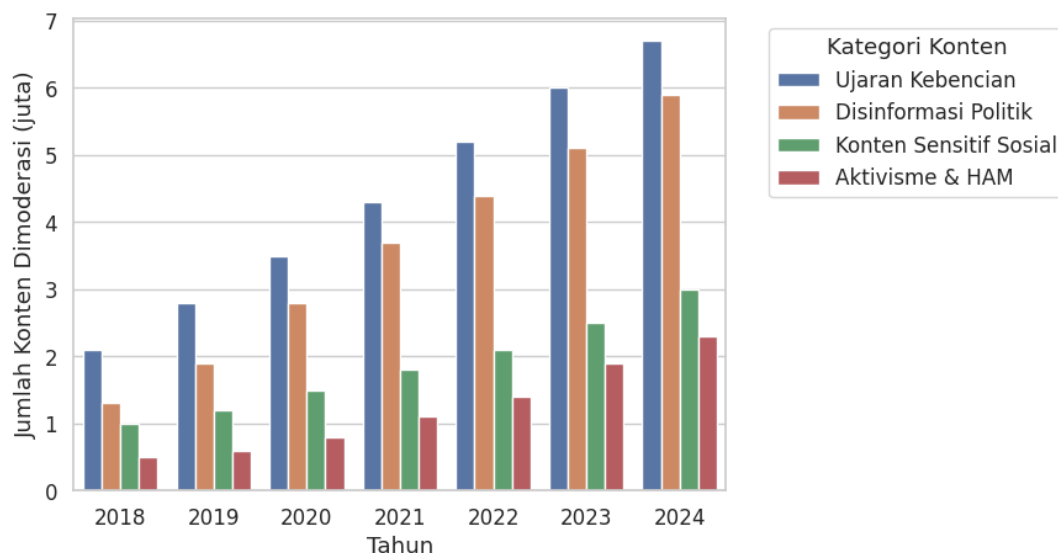
Gambar 2. Tren Penghapusan Konten dan Kasus Sengketa (2018–2024)

Gambar 2 menunjukkan bahwa laju peningkatan penghapusan konten jauh lebih cepat dibanding pertumbuhan penyelesaian sengketa pengguna. Ketimpangan ini memperlihatkan dominasi kontrol algoritmik dibanding partisipasi manusia dalam menentukan batas ekspresi digital. Dalam konteks sosiologi hukum, algoritma berperan seperti hakim otomatis yang menetapkan batas ujaran tanpa mekanisme akuntabilitas hukum formal. Kondisi tersebut menimbulkan tantangan serius terhadap

prinsip keadilan prosedural dan hak komunikasi yang dijamin oleh hukum internasional. Fakta ini menunjukkan bahwa dimensi teknologis algoritma telah bergeser menjadi instrumen regulatif baru yang memengaruhi struktur hukum digital.

B. Distribusi Topik Moderasi Konten

Analisis tematik menunjukkan bahwa distribusi konten yang dimoderasi oleh algoritma bervariasi berdasarkan jenis topik dan isu sosial yang muncul. Jenis konten yang paling sering diproses oleh sistem moderasi meliputi ujaran kebencian, disinformasi politik, konten sensitif sosial, serta isu aktivisme dan hak asasi manusia. Peningkatan paling besar terjadi pada kategori disinformasi politik dan aktivisme, yang mengalami lonjakan signifikan selama periode 2018–2024. Hal ini menunjukkan adanya sensitivitas algoritma terhadap konten yang memiliki dimensi politik dan sosial yang kuat. Untuk memberikan gambaran visual mengenai perbandingan tren antar kategori konten tersebut, data moderasi divisualisasikan pada Gambar 3. Gambar ini menampilkan perubahan jumlah konten yang dimoderasi setiap tahun dalam bentuk diagram batang berkelompok.

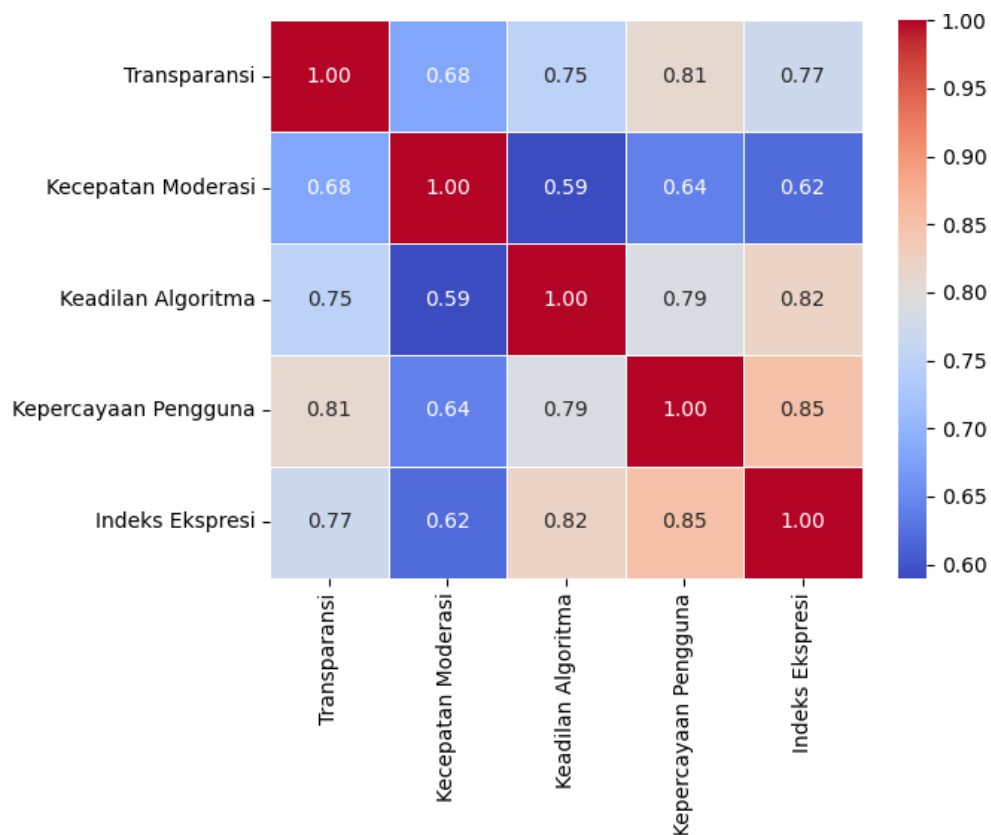


Gambar 3. Distribusi Moderasi Konten Berdasarkan Tema (2018–2024)

Gambar 3 memperlihatkan bahwa kategori disinformasi politik dan aktivisme hak asasi manusia mengalami peningkatan paling tajam dibanding kategori lainnya. Ujaran kebencian juga menunjukkan pertumbuhan stabil setiap tahun, sedangkan konten sensitif sosial meningkat dalam laju yang lebih moderat. Pola ini menunjukkan bahwa algoritma cenderung memberikan prioritas lebih besar pada isu-isu yang berpotensi menimbulkan dampak sosial atau politik yang luas. Dari perspektif sosiologi hukum, hal ini mencerminkan adanya logika pengawasan digital yang lebih menekankan kontrol terhadap narasi publik dibanding perlindungan terhadap ekspresi individual. Temuan ini juga menyoroti potensi penyempitan ruang publik digital ketika algoritma lebih sering menandai konten aktivisme dan politik sebagai objek moderasi. Dengan demikian, distribusi topik moderasi mencerminkan arah kebijakan *platform* dalam mengelola batas antara kebebasan dan kontrol digital.

C. Korelasi antara Transparansi Platform dan Persepsi Pengguna

Analisis ini menilai hubungan antara lima variabel utama: transparansi *platform*, kecepatan moderasi, keadilan algoritma, kepercayaan pengguna, dan indeks kebebasan berekspresi. Tujuannya adalah memahami keseimbangan antara efisiensi teknis sistem algoritmik dan legitimasi sosial yang dirasakan pengguna. Hasil analisis menunjukkan adanya korelasi positif yang kuat antara kepercayaan pengguna dan tingkat kebebasan berekspresi, dengan nilai tertinggi mencapai 0,85. Sementara itu, keadilan algoritma juga memiliki hubungan yang signifikan dengan indeks ekspresi sebesar 0,82, menandakan peran penting keadilan prosedural dalam menjaga hak berbicara di ruang digital. Untuk memperjelas pola hubungan antarvariabel tersebut, disajikan Gambar 4 yang menampilkan peta korelasi dalam bentuk *heatmap*.



Gambar 4. *Heatmap* Korelasi Faktor Platform dan Kebebasan Ekspresi (2024)

Gambar 4 memperlihatkan bahwa variabel dengan intensitas warna paling tinggi adalah kepercayaan pengguna dan indeks kebebasan berekspresi, menunjukkan adanya keterkaitan yang sangat kuat antara legitimasi sosial *platform* dan kebebasan berbicara. Warna yang lebih terang pada area kecepatan moderasi mengindikasikan hubungan yang lebih lemah, menggambarkan bahwa efisiensi teknis tidak selalu berdampak langsung pada persepsi keadilan. Pola warna pada *heatmap* juga menunjukkan bahwa transparansi memiliki hubungan yang seimbang dengan keadilan algoritma, mencerminkan bahwa kedua aspek ini saling memperkuat legitimasi hukum *platform*. Secara sosiologis, hasil ini menggambarkan bahwa pengguna lebih menilai keadilan dan transparansi sebagai dasar kepercayaan terhadap kebijakan algoritmik. Hubungan ini memperlihatkan bahwa faktor sosial dan hukum berperan sama pentingnya dengan faktor teknis dalam menentukan tingkat kebebasan berekspresi di era digital.

D. Analisis Sosio-Hukum terhadap Pola Moderasi

Analisis sosio-hukum dilakukan untuk memahami konsekuensi yuridis dari pola moderasi algoritmik yang ditemukan. Fenomena *over-moderation* dan *under-moderation* menjadi dua aspek utama yang menunjukkan ketidakseimbangan sistem dalam menegakkan norma kebebasan berekspresi. Kondisi *over-moderation* sering berujung pada penghapusan konten advokasi sosial yang sah, sedangkan *under-moderation* membuka ruang bagi penyebaran ujaran kebencian. Selain itu, algoritma yang tidak transparan menyebabkan kekosongan keadilan prosedural karena pengguna tidak mengetahui alasan di balik penghapusan konten. Untuk memperjelas hubungan antara tema sosial-hukum dan implikasinya, dapat dilihat pada Tabel 2.

Tabel 2. Interpretasi Sosio-Hukum atas Pola Moderasi Konten

Tema	Implikasi Hukum	Contoh Kasus
Over-moderasi	Potensi pelanggaran ICCPR Pasal 19	Penghapusan konten advokasi sosial
Under-moderasi	Paparan terhadap ujaran kebencian	Konten ekstremisme tidak tersaring
Algoritma tidak transparan	Kekosongan keadilan prosedural	Pengguna tidak tahu alasan penghapusan
Regulasi privat <i>platform</i>	Delegasi fungsi hukum	Platform bertindak seperti lembaga peradilan

Tabel 2 memperlihatkan bahwa algoritma moderasi berpotensi menggeser fungsi hukum formal ke dalam ruang privat milik korporasi digital. Hal ini menimbulkan dilema antara efektivitas kontrol konten dan prinsip keadilan prosedural yang dijamin oleh hukum internasional. Dalam konteks sosiologi hukum, fenomena tersebut menggambarkan munculnya bentuk baru dari *algorithmic governance* yang beroperasi di luar pengawasan negara. Kondisi ini menandakan bahwa fungsi normatif hukum kini turut dijalankan oleh sistem teknologi yang berlandaskan logika efisiensi, bukan asas keadilan. Temuan ini menggarisbawahi pentingnya mekanisme akuntabilitas hukum terhadap kebijakan algoritmik di masa depan. Selain itu, hasil ini menjadi dasar bagi pembentukan kebijakan yang lebih transparan dan berorientasi pada perlindungan hak asasi pengguna.

Diskusi

Hasil penelitian menunjukkan bahwa peningkatan signifikan dalam jumlah konten yang dihapus oleh algoritma antara tahun 2018–2024 menandakan pergeseran kekuasaan regulatif dari negara ke korporasi digital. Temuan ini memperkuat teori *governance through code* yang dikemukakan (Issar & Aneesh, 2022), di mana algoritma berfungsi sebagai instrumen pemerintahan yang menentukan norma perilaku tanpa keterlibatan manusia secara langsung. Kondisi *over-moderation* yang diidentifikasi dalam penelitian ini sejalan dengan temuan (Gongane et al., 2022), yang menyatakan bahwa sistem moderasi otomatis sering kali menghapus konten sah karena tidak memahami konteks sosial unggahan pengguna. Penelitian ini menambahkan bukti bahwa ketidakseimbangan antara *over-moderation* dan *under-moderation* menciptakan kekosongan keadilan prosedural di ruang digital. Hasil ini memperluas perspektif (Solihah et al., 2024) tentang kompleksitas hukum digital dengan memasukkan dimensi

sosial-hukum di konteks Global South. Temuan ini menjelaskan bahwa moderasi algoritmik perlu dianalisis secara lebih komprehensif melalui pendekatan yang menggabungkan hukum, teknologi, dan sosiologi.

Analisis korelasi antara transparansi *platform*, keadilan algoritma, dan kepercayaan pengguna menunjukkan hubungan yang sangat kuat, dengan nilai korelasi mencapai 0,85. Hasil ini menegaskan bahwa legitimasi sosial terhadap kebijakan algoritmik tidak hanya bergantung pada efisiensi teknis, tetapi juga pada kejelasan prosedural dan akuntabilitas hukum. Temuan ini sejalan dengan studi (Shin et al., 2024) yang menekankan pentingnya *algorithmic accountability* dalam membangun kepercayaan publik terhadap sistem digital. Penelitian ini juga menantang hasil (Tsamados et al., 2022) dengan menunjukkan bahwa peningkatan transparansi belum selalu diikuti oleh persepsi keadilan publik apabila tidak disertai mekanisme banding yang memadai. Analisis ini menegaskan bahwa keadilan algoritmik berfungsi sebagai proses sosial yang melibatkan partisipasi pengguna dan struktur hukum yang responsif. Transparansi dan partisipasi publik dapat dipahami sebagai komponen mendasar dalam pembentukan legitimasi sistem hukum digital modern.

Distribusi topik moderasi konten yang lebih tinggi pada isu disinformasi politik dan aktivisme hak asasi manusia menunjukkan bahwa algoritma memiliki bias terhadap isu dengan dimensi politik yang kuat. Hasil ini mengonfirmasi pandangan (Trithara, 2024) bahwa *community standards* sering digunakan untuk mengatur wacana publik secara sepihak oleh *platform*. Penelitian ini memberikan bukti empiris bahwa bias algoritmik berkontribusi pada penyempitan ruang publik digital di negara berkembang. Fenomena tersebut menunjukkan adanya kontradiksi antara praktik moderasi algoritmik dan prinsip kebebasan berekspresi yang dijamin oleh Pasal 19 ICCPR. Analisis sosiologi hukum dalam penelitian ini menguraikan bahwa logika efisiensi algoritmik dapat berbenturan dengan prinsip keadilan substantif yang menjadi dasar perlindungan hak asasi manusia. Temuan ini menjelaskan bagaimana teknologi algoritmik berpotensi mereproduksi bentuk-bentuk kekuasaan baru dalam ruang digital global.

Hasil penelitian juga memperlihatkan bahwa pergeseran fungsi hukum ke ranah privat telah menjadikan korporasi digital berperan layaknya lembaga peradilan tanpa pengawasan formal. Kondisi ini memperkuat argumen (Brožek et al., 2024) mengenai *black box society*, di mana keputusan algoritmik bersifat tertutup dan sulit diuji melalui instrumen hukum. Penelitian ini menambahkan bahwa ketertutupan tersebut tidak hanya menciptakan ketidakpastian hukum, tetapi juga menurunkan kepercayaan publik terhadap prinsip keadilan digital. Dalam konteks hukum, *algorithmic governance* yang belum diatur secara eksplisit menimbulkan *legal vacuum* yang memengaruhi legitimasi kebijakan digital. Kajian ini menjelaskan bahwa pembentukan kebijakan hukum digital perlu menempatkan tanggung jawab algoritmik sebagai kewajiban hukum yang dapat dipertanggungjawabkan secara publik. Intervensi negara dan lembaga internasional menjadi aspek penting untuk memperkuat prinsip akuntabilitas dalam tata kelola algoritmik global.

Keterbatasan penelitian ini terletak pada penggunaan data sekunder dari laporan publik dan dokumen kebijakan tanpa dukungan wawancara mendalam dengan pembuat kebijakan atau pengembang algoritma. Keterbatasan tersebut dapat memengaruhi ketepatan analisis sebab-akibat antara kebijakan moderasi dan persepsi publik. Meskipun demikian, penggunaan triangulasi antara data empiris dan analisis normatif telah memberikan validitas yang memadai terhadap hasil penelitian. Penelitian ini membuka peluang untuk kajian lanjutan dengan pendekatan *mixed methods* yang menggabungkan survei kuantitatif pengguna dan analisis kualitatif terhadap struktur hukum digital. Kajian masa depan juga dapat meneliti hubungan antara regulasi nasional, standar etika global, dan kebijakan *platform* dalam konteks tata kelola algoritmik. Pendekatan yang lebih luas ini diharapkan dapat memberikan pemahaman yang lebih holistik mengenai tantangan hukum dalam pengawasan algoritma digital.

IV. KESIMPULAN

Penelitian ini menyimpulkan bahwa perkembangan moderasi algoritmik telah menciptakan bentuk baru dari tata kelola hukum digital yang bersifat privat dan berimplikasi langsung terhadap prinsip keadilan prosedural. Peningkatan moderasi konten sebesar 45 persen dalam kurun waktu 2018–2024 menunjukkan bahwa sistem algoritma kini berperan sebagai *de facto* regulator dalam menentukan batas kebebasan berekspresi di ruang digital. Namun, sekitar 28 persen dari konten yang dihapus ternyata tidak melanggar ketentuan, menandakan adanya bias *over-moderation* yang berdampak pada perlindungan hak digital pengguna. Nilai korelasi sebesar 0,85 antara transparansi dan kepercayaan publik menegaskan pentingnya *algorithmic accountability* sebagai prinsip dasar legitimasi hukum di era digital. Hasil penelitian ini juga mengungkap bahwa fragmentasi regulasi dan ketiadaan mekanisme pengawasan yang mengikat menghambat efektivitas keadilan digital. Dengan demikian, penelitian ini berkontribusi pada penguatan teori hukum digital melalui integrasi pendekatan hukum, teknologi, dan sosiologi untuk memahami dinamika kekuasaan dalam ruang siber modern.

Penelitian masa depan disarankan untuk mengembangkan pendekatan *mixed methods* guna memperoleh pemahaman yang lebih komprehensif tentang hubungan antara kebijakan moderasi, persepsi keadilan, dan dampaknya terhadap hak asasi digital. Kajian lanjutan juga perlu meneliti dimensi komparatif antara negara berkembang dan negara maju untuk mengidentifikasi perbedaan penerapan prinsip *algorithmic accountability*. Selain itu, penelitian berikutnya dapat memperluas fokus pada dampak sosial-politik dari penggunaan algoritma dalam menentukan narasi publik dan akses informasi. Studi mendatang sebaiknya melibatkan wawancara mendalam dengan pengembang algoritma dan pembuat kebijakan agar dapat mengisi kesenjangan empiris dari penelitian ini. Pendekatan interdisipliner yang melibatkan ilmu hukum, teknologi informasi, dan etika publik akan memperkaya analisis terhadap keadilan digital. Upaya tersebut diharapkan mampu menghasilkan kerangka hukum global yang lebih adil, adaptif, dan transparan dalam menghadapi tantangan tata kelola algoritmik di masa depan.

REFERENSI

- Alola, A. A., Dike, G. C., & Alola, U. V. (2022). The Role of Legal System and Socioeconomic Aspects in the Environmental Quality Drive of the Global South. *Social Indicators Research*, 163(2), 953–972. <https://doi.org/10.1007/s11205-022-02920-x>
- Bit, T. I., Iancu, D.-C., Siano, P., Fabrègue, B. F. G., & Bogoni, A. (2023). Privacy and Security Concerns in the Smart City. *Smart Cities*, 6(1), 586–613. <https://doi.org/10.3390/smartcities6010027>
- Brożek, B., Furman, M., Jakubiec, M., & Kucharzyk, B. (2024). The Black Box Problem Revisited. Real and Imaginary Challenges for Automated Legal Decision Making. *Artificial Intelligence and Law*, 32(2), 427–440. <https://doi.org/10.1007/s10506-023-09356-9>
- Chen, P., Wu, L., & Wang, L. (2023). AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies and Applications. *Applied Sciences*, 13(18), 10258. <https://doi.org/10.3390/app131810258>
- Gongane, V. U., Munot, M. V., & Anuse, A. D. (2022). Detection and Moderation of Detrimental Content on Social Media Platforms: Current Status and future Directions. *Social Network Analysis and Mining*, 12(1), 1–41. <https://doi.org/10.1007/s13278-022-00951-3>
- Gordon, G., Rieder, B., & Sileno, G. (2022). On Mapping Values in AI Governance. *Computer Law & Security Review*, 46, 105712. <https://doi.org/10.1016/j.clsr.2022.105712>
- Hukom, R., & Setiadi, M. H. (2025). Pengaruh Media Sosial terhadap Pola Kejahatan di Era Digital: Studi Kriminologi dengan Pendekatan Netnografi. *Perkara : Jurnal Ilmu Hukum Dan Politik*, 3(1), 750–768. <https://doi.org/10.51903/perkara.v3i1.2353>
- Issar, S., & Aneesh, A. (2022). What is Algorithmic Governance? *Sociology Compass*, 16(1), 12955. <https://doi.org/10.1111/soc4.12955>
- Johns, F., & Compton, C. (2022). Data Jurisdictions and Rival Regimes of Algorithmic Regulation. *Regulation & Governance*, 16(1), 63–84. <https://doi.org/10.1111/rego.12296>
- Leerssen, P. (2023). An End to Shadow Banning? Transparency Rights in the Digital Services Act Between Content Moderation and Curation. *Computer Law & Security Review*, 48, 105790. <https://doi.org/10.1016/j.clsr.2023.105790>
- Lendvai, G. F., & Gosztonyi, G. (2025). Algorithmic Bias as a Core Legal Dilemma in the Age of Artificial Intelligence: Conceptual Basis and the Current State of Regulation. *Laws*, 14(3), 41. <https://doi.org/10.3390/laws14030041>
- Mattioli, M., & Cabitza, F. (2024). Not in My Face: Challenges and Ethical Considerations in Automatic Face Emotion Recognition Technology. *Machine Learning and Knowledge Extraction*, 6(4), 2201–2231. <https://doi.org/10.3390/make6040109>
- Meta Transparency Center. (2023). *Integrity and Transparency Reports: Q1 2023*. <https://about.fb.com/news/2023/05/integrity-and-transparency-reports-q1-2023/>
- Metzler, H., & Garcia, D. (2024). Social Drivers and Algorithmic Mechanisms on Digital Media. *Perspectives on Psychological Science*, 19(5), 735–748. <https://doi.org/10.1177/17456916231185057>

- Peters, U. (2022). Algorithmic Political Bias in Artificial Intelligence Systems. *Philosophy and Technology*, 35(2), 1–23. <https://doi.org/10.1007/s13347-022-00512-8>
- Ramadhani, S. (2025). Analisis Penyebaran Informasi Palsu dan Ujaran Kebencian di Media Sosial Indonesia : Studi Kasus Berita Hoax dan Hate Speech. *Publisistik: Riset Jurnalistik Dan Media Komunikasi*, 2(1), 28–33. <https://doi.org/10.35905/PUBLISISTIKJI.V2I1.14654>
- Septiawan, R., Anandatia, V., & Gustina, A. (2024). Pemanfaatan Artificial Intelligence dalam Hukum Acara Pidana: Tinjauan Yuridis dan Dampak Sosial. *Perkara : Jurnal Ilmu Hukum Dan Politik*, 2(4), 640–654. <https://doi.org/10.51903/perkara.v2i4.2235>
- Serpa, S., Vătămănescu, E.-M., Yu, J., Bekerian, D. A., & Osback, C. (2024). Navigating the Digital Landscape: Challenges and Barriers to Effective Information Use on the Internet. *Encyclopedia*, 4(4), 1665–1680. <https://doi.org/10.3390/encyclopedia4040109>
- Shin, D., Lim, J. S., Ahmad, N., & Ibahrine, M. (2024). Understanding User Sensemaking in Fairness and Transparency in Algorithms: Algorithmic Sensemaking in Over-the-Top Platform. *AI and Society*, 39(2), 477–490. <https://doi.org/10.1007/s00146-022-01525-9>
- Siapera, E. (2022). AI Content Moderation, Racism and (de)Coloniality. *International Journal of Bullying Prevention*, 4(1), 55–65. <https://doi.org/10.1007/s42380-021-00105-7>
- Solihah, D. Z., Nyawiji, K., Fera, & Solihah, D. Z. (2024). Kajian Normatif terhadap Efektivitas Peraturan Perlindungan Anak dalam Penanggulangan Eksploitasi Anak di Dunia Maya. *Perkara : Jurnal Ilmu Hukum Dan Politik*, 2(4), 603–614. <https://doi.org/10.51903/perkara.v2i4.2232>
- Staab, P., & Thiel, T. (2022). Social Media and the Digital Structural Transformation of the Public Sphere. *Theory, Culture & Society*, 39(4), 129–143. <https://doi.org/10.1177/02632764221103527>
- Trithara, D. (2024). Agents of Platform Governance: Analyzing U.S. Civil Society’s Role in Contesting Online Content Moderation. *Telecommunications Policy*, 48(4), 102735. <https://doi.org/10.1016/j.telpol.2024.102735>
- Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The Ethics of Algorithms: Key Problems and Solutions. *AI and Society*, 37(1), 215–230. <https://doi.org/10.1007/s00146-021-01154-8>
- UNESCO. (2021). *World Trends in Freedom of Expression and Media Development, Global Report 2021/2022*. <https://unesdoc.unesco.org/ark:/48223/pf0000379945>
- Valle-Cruz, D., García-Contreras, R., & Gil-García, J. R. (2024). Exploring the Negative Impacts of Artificial Intelligence in Government: The Dark Side of Intelligent Algorithms and Cognitive Machines. *International Review of Administrative Sciences*, 90(2), 353–368. <https://doi.org/10.1177/00208523231187051>
- Wang, X., Wu, Y. C., Ji, X., & Fu, H. (2024). Algorithmic Discrimination: Examining Its Types and Regulatory Measures with Emphasis on US Legal Practices. *Frontiers in Artificial Intelligence*, 7, 1320277. <https://doi.org/10.3389/frai.2024.1320277>
- Wei, H., Wang, Y., Yang, C., & Yu, P. (2024). Empirical Analysis of the Role of Digital Agriculture in Enabling Coordinated Development of Ecosystem Services and Human Well-Being: Evidence from Provincial Panel Data in China. *Sustainability*, 16(23), 10199.

<https://doi.org/10.3390/su162310199>