TANGGUNG JAWAB PLATFORM MEDIA SOSIAL INSTAGRAM YANG MENGALAMI KEGAGALAN PADA PROSES PENYARINGAN KONTEN MENGGUNAKAN KECERDASAN BUATAN *(ARTIFICIAL INTELLIGENCE)*

(SKRIPSI)

Oleh:

MUHAMMAD YAFI JAWAD RIADI
NPM 2212011047

FAKULTAS HUKUM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

ABSTRAK

TANGGUNG JAWAB PLATFORM MEDIA SOSIAL INSTAGRAM YANG MENGALAMI KEGAGALAN PADA PROSES PENYARINGAN KONTEN MENGGUNAKAN KECERDASAN BUATAN (*ARTIFICIAL INTELLIGENCE)*

Oleh:

MUHAMMAD YAFI JAWAD RIADI

Perkembangan kecerdasan buatan *Artificial Intelligence* (AI) telah memengaruhi praktik penyaringan konten di media sosial. AI digunakan untuk mendeteksi dan memblokir konten sensitif, namun sering kali mengalami kegagalan yang menimbulkan persoalan hukum dan sosial. Berdasarkan hal tersebut, penelitian ini membahas mengenai tanggung jawab hukum platform media sosial Instagram atas kegagalan penyaringan konten oleh kecerdasan buatan serta sanksi hukum yang dapat dikenakan apabila sistem tersebut tidak berfungsi sebagaimana mestinya dalam menyaring konten sensitif. Kegagalan tersebut kerap disebabkan oleh kesalahan klasifikasi, bias algoritmik, serta ketidakmampuan AI memahami konteks sosial dan bahasa, yang berpotensi merugikan pengguna serta menimbulkan pertanggungjawaban hukum bagi penyelenggara platform.

Penelitian ini merupakan penelitian hukum normatif dengan tipe penelitian deskriptif. Pendekatan yang digunakan meliputi pendekatan perundang-undangan dan pendekatan konseptual. Data penelitian diperoleh melalui studi kepustakaan yang kemudian dianalisis secara kualitatif untuk menarik kesimpulan yang relevan dengan permasalahan penelitian.

Hasil penelitian menunjukkan bahwa tanggung jawab hukum platform media sosial Instagram atas kegagalan penyaringan konten oleh kecerdasan buatan bersifat administratif, regulatif, dan perdata sebagaimana diatur dalam PP Nomor 71 Tahun 2019 dan Permenkominfo Nomor 5 Tahun 2020. Kegagalan moderasi konten dapat berupa *false negative*, yaitu tidak terdeteksinya konten sensitif atau ilegal, serta *false positive*, yaitu penghapusan terhadap konten yang sah. Atas kelalaian tersebut, platform dapat dikenai sanksi administratif berupa teguran tertulis, denda, pembatasan akses, hingga pencabutan izin operasional. Selain itu, tanggung jawab regulatif menuntut kepatuhan terhadap ketentuan UU ITE dan prinsip penyelenggaraan sistem elektronik yang andal dan bertanggung jawab.

Kata kunci: Penyaringan Konten, Kecerdasan Buatan, Tanggung Jawab Hukum.

ABSTRACT

THE RESPONSIBILITY OF THE INSTAGRAM SOCIAL MEDIA PLATFORM REGARDING FAILURES IN CONTENT FILTERING PROCESSES USING ARTIFICIAL INTELLIGENCEE.

By

MUHAMMAD YAFI JAWAD RIADI

The development of Artificial Intelligence (AI) has significantly influenced the practice of content moderation on social media platforms. AI is utilized to detect and block sensitive content; however, it often fails, resulting in various legal and social issues. This study discusses the legal responsibility of the social media platform Instagram for the failure of AI-based content moderation, as well as the legal sanctions that may be imposed when such systems do not function properly in filtering sensitive content. These failures are frequently caused by misclassification, algorithmic bias, and the inability of AI to understand social and linguistic contexts, which may harm users and give rise to legal liability for the platform provider.

This research is a normative legal study with a descriptive research type. The approaches used include the statutory approach and the conceptual approach. The data were obtained through library research and analyzed qualitatively.

The results of this study indicate that the legal liability of the social media platform Instagram for failures in content moderation using artificial intelligence is administrative, regulatory, and civil in nature, as stipulated in Government Regulation Number 71 of 2019 and Regulation of the Minister of Communication and Informatics Number 5 of 2020. Failures in content moderation may take the form of false negatives, namely the failure to detect sensitive or illegal content, as well as false positives, namely the removal of lawful content. As a result of such negligence, the platform may be subject to administrative sanctions in the form of written warnings, administrative fines, access restrictions, or revocation of operational licenses. Furthermore, regulatory liability requires compliance with the provisions of the Electronic Information and Transactions Law as well as the principles of reliable and responsible electronic system operation.

Keywords: Content Moderation, Artificial Intelligence, Legal Responsibility.

TANGGUNG JAWAB PLATFORM MEDIA SOSIAL INSTAGRAM YANG MENGALAMI KEGAGALAN PADA PROSES PENYARINGAN KONTEN MENGGUNAKAN KECERDASAN BUATAN *(ARTIFICIAL INTELLIGENCE)*

Oleh:

MUHAMMAD YAFI JAWAD RIADI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA HUKUM

Pada

Bagian Hukum Keperdataan
Fakultas Hukum Universitas Lampung

FAKULTAS HUKUM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

Judul Skripsi	:	**TANGGUNG JAWAB PLATFORM MEDIA SOSIAL INSTAGRAM YANG MENGALAMI KEGAGALAN PADA PROSES PENYARINGAN KONTEN MENGGUNAKAN KECERDASAN BUATAN (*ARTIFICIAL INTELLIGENCE*)**
Nama Mahasiswa	:	**Muhammad Yafi Jawad Riadi**
Nomor Pokok Mahasiswa	:	2212011047
Bagian	:	Hukum Keperdataan
Fakultas	:	Hukum

MENYETUJUI

1. Komisi pembimbing

Elly Nurlaili, S.H., M.H.
NIP 197001292006042001

Moh.Wendy Trijaya, S.H.,M. Hum.
NIP 197108252005011002

2. Ketua Bagian Hukum Perdata

Dr. Ahmad Zazili, S.H., M.H
NIP 197404132005011001

MENGESAHKAN

1. Tim Penguji

   Ketua : **Elly Nurlaili, S.H., M.H.** .....

   Sekretaris/Anggota : **Moh.Wendy Trijaya, S.H., M. Hum.** .....

   Penguji Utama : **Dewi Septiana, S,H., M.H.** .....

2. Dekan Fakultas Hukum

Dr. M. Fakih, S.H., M.S.

NIP 196412181988031002

Tanggal Lulus Ujian Skripsi : 27 Januari 2026

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama	:	Muhammad Yafi Jawad Riadi
Nomor Pokok Mahasiswa	:	2212011047
Bagian	:	Hukum Keperdataan
Fakultas	:	Hukum

Dengan ini menyatakan bahwa skripsi saya yang berjudul "**Tanggung Jawab Platform Media Sosial Instagram yang Mengalami Kegagalan Pada Proses Penyaringan Konten Menggunakan Kecerdasan Buatan *(Artificial Intelligence)***" adanya benar-benar hasil karya saya sendiri dan bukan hasil plagiat sebagaimana telah diatur dalam Pasal 43 Peraturan Akademik Universitas Lampung Nomor 2 Tahun 2025.

Bandar Lampung, 10 Feb 2026

Muhammad Yafi Jawad Riadi

NPM 2212011047

RIWAYAT HIDUP

Penulis bernama lengkap Muhammad Yafi Jawad Riadi, dilahirkan di Liwa pada 13 April 2004. Penulis merupakan anak terakhir dari pasangan Bapak Agus Triyadi dan Ibu Imelda. Penulis menyelesaikan pendidikan di Taman Kanak-Kanak Pertiwi Kab. Lampung Barat pada tahun 2010, Sekolah Dasar Negeri 1 Liwa pada tahun 2016, melanjutkan pendidikan menengah pertama di MTSN 1 Lampung Barat, dan lulus pada tahun 2019, kemudian menamatkan pendidikan menengah atas di SMAN I liwa pada tahun 2022.

Pada tahun 2022, penulis diterima sebagai mahasiswa Fakultas Hukum Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Selama menjalani perkuliahan, Penulis aktif mengikuti berbagai kegiatan organisasi dan ajang pengembangan diri di tingkat daerah maupun provinsi. Penulis tercatat sebagai anggota Ikatan Muli Mekhanai Kabupaten Lampung Barat (IMMKALAB) dan pada tahun 2023 berhasil meraih predikat Mekhanai Kabupaten Lampung Barat 2023 serta menjalankan tugasnya selama satu tahun. Pada tahun berikutnya, yakni 2024, Penulis berkesempatan mengikuti ajang pemilihan tingkat provinsi, yaitu The Face of Lampung, dan berhasil memperoleh gelar Mister National Lampung 2024, yang sekaligus menjadi representasi bagi Provinsi Lampung di tingkat nasional. Selain itu, pada tahun 2025 Penulis juga memperluas keterlibatan dalam bidang seni dan industri kreatif dengan bergabung dalam organisasi Asosiasi Perancang Pengusaha Mode Indonesia (APPMI) Lampung, yang menjadi wadah untuk mendukung bakat, minat, serta jejaring di bidang mode dan kewirausahaan.

MOTO

"Tidak ada balasan untuk kebaikan selain kebaikan (pula)"

(QS. Ar Rahman ayat 60)

"Bukanlah pegunungan di hadapan Anda yang membuat Anda menyerah untuk memanjat. Tetapi kerikil di sepatu Anda itu lah yang menyebabkan demikian."

(Muhammad Ali)

You Can't Lose If You Don't Give Up

(Muhammad Yafi Jawad Riadi)

PERSEMBAHAN

Segala puji dan syukur Penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, hidayah, kekuatan, dan kesehatan yang telah diberikan, sehingga penyusunan skripsi ini dapat diselesaikan dengan baik.
Dengan segenap ketulusan hati, karya sederhana ini Penulis persembahkan skripsi ini kepada:

Kedua Orang Tuaku tersayang,
Bapak Agus Triyadi, S.IP., M.M dan Imelda, S.ST

Kepada Abi dan Umi,
Terima kasih atas kasih sayang, doa, dan pengorbanan yang tiada henti. Terima kasih telah mendidik, membimbing, dan mengajarkan arti kesabaran, ketulusan, serta perjuangan dalam setiap langkah hidupku. Dukungan dan doa kalian adalah sumber kekuatan terbesar yang selalu mengiringi setiap perjalanan yang kujalani. Karya ini adalah wujud dari rasa hormat, cinta, dan terima kasihku kepada kalian berdua, yang senantiasa menjadi cahaya, inspirasi, dan alasan terindah dalam setiap langkah perjuanganku.

SANWACANA

Puji syukur penulis ucapkan kepada Tuhan Yang Maha Esa karena atas karunia-Nya penulis dapat menyelesaikan penulisan skripsi yang berjudul **"Tanggung Jawab Platform Media Sosial Instagram yang Mengalami Kegagalan Pada Proses Penyaringan Konten Menggunakan Kecerdasan Buatan *(Artificial Intelligence)"*** sebagai salah satu syarat untuk memperoleh gelar Sarjana Hukum pada Fakultas Hukum Universitas Lampung. Penyelesaian skripsi ini tidak terlepas dari bantuan, bimbingan, saran, serta dukungan dari berbagai pihak, dengan ini penulis ingin mengucapkan terima kasih dengan setulus-tulusnya kepada:

1. Bapak Dr. M. Fakih, S.H., M.S., selaku Dekan Fakultas Hukum Universitas Lampung;
2. Bapak Dr. Ahmad Zazili, S.H., M.H., selaku Ketua Bagian Hukum Keperdataan Fakultas Hukum Universitas Lampung;
3. Bapak Mohammad Wendy Trijaya, S.H., M.Hum., selaku Sekretaris Bagian Hukum Perdata Fakultas Hukum Universitas Lampung sekaligus Dosen Pembimbing II, terima kasih atas kesabaran serta kesediaannya dalam memberikan waktu, bimbingan, masukan, dan arahan selama proses penyelesaian skripsi ini;
4. Ibu Elly Nurlaily, S.H., M.H., selaku Dosen Pembimbing 1, terima kasih atas kesabaran dan kesediannya meluangkan waktu di sela-sela kesibukannya untuk memberikan arahan, bimbingan, serta saran dan berbagai kritik dalam proses penyelesaian skripsi ini;
5. Ibu Dewi Septiana, S.H., M.H., M.Hum., selaku Dosen Pembahas I, terima kasih atas waktu, masukan, dan kritik yang membangun selama penulisan skripsi ini;

6. Ibu Dora Mustika, S.H., M.H., selaku Dosen Pembahas II, terima kasih atas waktu, masukan, dan kritik yang membangun selama penulisan skripsi ini;
7. Seluruh dosen dan tenaga kependidikan Fakultas Hukum Universitas Lampung, khususnya Bagian Hukum Perdata, terima kasih atas ilmu yang bermanfaat bagi Penulis dan bantuan administratif yang diberikan kepada Penulis selama menempuh pendidikan di Fakultas Hukum Universitas Lampung:
8. Kedua kakak saya, Alif Panzha Riadi dan Aldellia Riadi, terima kasih atas kasih sayang, doa, dan semangat yang tak pernah putus, yang selalu menjadi dorongan besar bagi saya untuk terus berjuang hingga titik akhir;
9. Sahabat-sahabat dalam grup Secangkir Kopi SMA 1 Liwa: Kurniyawan, Rian, Dayat, Aidil, Ardi, Dimas, Ghulam, Hidayat, Yoga, Irgi, Rehan, Heru, Adeno, Agung, Arga, Arya, Fahri, Fajar, Kamal, Jeki, Kevin, terima kasih atas tawa, dukungan, serta semangat yang senantiasa mengingatkan bahwa kebersamaan adalah energi berharga dalam setiap perjuangan;
10. Kawan seperjuangan kuliah: Adi, Alif, Irfan, Fajri dan Fakhri terima kasih atas diskusi, motivasi, dan semangat pantang menyerah yang membuat langkah-langkah kecil terasa lebih ringan hingga akhirnya sampai di titik ini;
11. Grup kuliah DKM Futsalliyah: Adi, Alif, Fajri, Faqih, Ipan, Raka, Rendy, Ridho, Yaser, Kevin, Fadhil, Christoper, Dipo, Josi, terima kasih atas kebersamaan, canda tawa, serta semangat yang selalu menghidupkan perjalanan kuliah ini, baik di dalam maupun di luar kelas;
12. UKM Futsal Hukum Unila, terima kasih atas kebersamaan, sportivitas, dan semangat juang yang turut menempa mental serta memberi warna tersendiri dalam perjalanan saya di bangku kuliah;
13. Ikatan Muli Mekhanai Lampung Barat, terima kasih atas pengalaman, dukungan, dan semangat yang telah membentuk kepercayaan diri serta rasa tanggung jawab saya dalam berkarya;
14. *The Face of Lampung*, terima kasih atas kesempatan, inspirasi, serta semangat untuk terus membuktikan diri dan membawa nama daerah dengan penuh kebanggaan;

15. Kelompok KKN Sumber Agung, yaitu Retno, Nissa, Salman, Desi, Anggun, dan Hanif, Terimakasih atas kebersamaan, semangat, serta dukungan yang telah diberikan. Kehadiran kalian tidak hanya memberikan warna dalam perjalanan KKN, tetapi juga menjadi sumber motivasi dan semangat bagi penulis untuk terus berproses hingga terselesaikannya skripsi ini;
16. Dan kepada seluruh pihak yang tidak dapat saya sebutkan satu per satu, terima kasih atas doa, dorongan, dan semangat yang senantiasa menguatkan saya dalam menyelesaikan karya sederhana ini.

Semoga segala bentuk kebaikan, dukungan, dan bantuan yang telah diberikan kepada Penulis selama proses penyusunan skripsi ini mendapat balasan yang setimpal dari Tuhan Yang Maha Esa. Penulis menyadari sepenuhnya bahwa karya ini masih jauh dari kata sempurna. Meski demikian, Penulis berharap skripsi ini dapat memberi manfaat bagi berbagai pihak, khususnya dalam memberikan kontribusi bagi pengembangan ilmu hukum secara lebih luas.

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang

Kecerdasan Buatan (*Artificial Intelligence*/AI) merupakan suatu bentuk teknologi yang memiliki peran strategis dalam perkembangan masyarakat modern. Keberadaan teknologi ini menandai terjadinya transformasi signifikan dalam berbagai aspek kehidupan manusia dan memberikan pengaruh luas terhadap realitas sosial kontemporer, khususnya di tengah pesatnya kemajuan teknologi digital. Perkembangan AI yang semakin kompleks dan adaptif telah mengubah pola kerja, cara berkomunikasi, serta sistem kehidupan manusia secara menyeluruh. Dalam konteks era teknologi saat ini, kecerdasan buatan menjadi instrumen yang relevan dan krusial dalam menyediakan solusi yang efisien dan inovatif guna menjawab berbagai tantangan yang dihadapi manusia.[1]

Teknologi Kecerdasan Buatan atau AI kini telah menjadi bagian integral dari kehidupan manusia modern. Penggunaannya tidak hanya terbatas pada bidang industri atau kebutuhan profesional, tetapi juga merambah ke aktivitas sehari-hari, baik untuk hiburan maupun kemudahan hidup. Kecerdasan buatan secara perlahan mengubah tatanan dan pola interaksi manusia, terutama melalui integrasinya dalam berbagai teknologi yang digunakan masyarakat luas. Salah satu contoh penerapan AI yang paling umum adalah pada fitur pencarian *Google Search*, yang membantu pengguna memperoleh informasi dan pengetahuan dengan cepat dan efisien. Selain itu, penerapan AI juga dapat ditemukan dalam perangkat *smartphone* serta berbagai

[1] Maryani Farwati et al., "Analisa Pengaruh Teknologi Artificial Intelligence (AI) dalam Kehidupan Sehari-Hari [Analyze the Influence of Artificial Intelligence (AI) Technology in Daily Life]," *Jurnal Sistem Informatika dan Manajemen* 11, no. 1 (2023): 41–42.

platform media sosial seperti Instagram. Melalui teknologi ini, pengguna memiliki ruang untuk berekspresi, berinteraksi, dan berbagi konten secara bebas.[2]

Perkembangan teknologi kecerdasan buatan (AI) telah membawa perubahan besar dalam berbagai aspek kehidupan, termasuk dalam penyaringan konten di internet. Dengan meningkatnya volume informasi yang dipublikasikan secara daring, *platform digital* semakin bergantung pada AI untuk menyaring dan menghapus konten yang dianggap tidak sesuai, seperti ujaran kebencian, disinformasi, dan materi ilegal. AI memungkinkan proses penyaringan konten yang lebih cepat dan efisien dibandingkan dengan tenaga manusia, namun menimbulkan tantangan hukum terkait batasan kebebasan berekspresi dan tanggung jawab *platform digital*.[3]

Meskipun AI dapat membantu mengurangi penyebaran konten berbahaya, keandalan teknologi ini masih dipertanyakan. AI sering kali gagal memahami konteks dari suatu konten, yang dapat menyebabkan kesalahan dalam penghapusan informasi yang sah atau sebaliknya membiarkan konten yang berbahaya tetap beredar.[4] Kesalahan dalam penyaringan konten juga berpotensi merugikan pengguna, terutama ketika hak mereka untuk mengakses informasi atau mengekspresikan pendapatnya dibatasi tanpa proses yang transparan dan adil. Menurut *Internetworldstats* 2025 tingkat penggunaan internet di Indonesia telah mencapai 74,6% dari total popolasi.[5] Hal ini berarti bahwa di Indonesia sendiri pengguna internet sebanyak 143 juta jiwa pengguna pada tahun 2025.

Instagram sebagai salah satu platform media sosial dengan basis pengguna yang sangat besar tidak terlepas dari sorotan publik terkait kegagalan moderasi konten

---

[2] Kharisma Agustya Zahra Salsabilla et al., *"Pengaruh Penggunaan Kecerdasan Buatan terhadap Mahasiswa di Perguruan Tinggi," Prosiding Seminar Nasional Teknologi dan Sistem Informasi* 3, no. 1 (2023): 168–75, https://doi.org/10.33005/sitasi.v3i1.371.

[3] Mahyuddin Daud and Ida Madieha Abd Ghani Azmi, *"Intermediary's Liability: Towards a Sustainable Artificial Intelligence-Based Content Moderation in Malaysia," IIUM Law Journal* 31, no. 2 (2023): 155–78, https://doi.org/10.31436/iiumlj.v31i2.823.

[4] Thiago Dias Oliva, *"Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression," Human Rights Law Review* 20, no. 4 (December 9, 2020): 607–40, https://doi.org/10.1093/hrlr/ngaa032.

[5] Digital 2025: Indonesia, DataReportal, 2025, diakses 29 Maret 2025, https://datareportal.com/reports/digital-2025-indonesia.

yang sepenuhnya diotomatisasi melalui kecerdasan buatan. Penyaringan berbasis AI pada dasarnya dirancang untuk mempercepat proses identifikasi dan penyaringan terhadap konten yang dianggap melanggar kebijakan komunitas. Namun, efektivitas sistem ini masih menimbulkan perdebatan. Penelitian yang dilakukan oleh Wahyu Hadikristanto mengenai penerapan *content moderation* pada Instagram dengan menggunakan metode *machine learning* berbasis cloud menunjukkan bahwa meskipun sistem tersebut memiliki tingkat akurasi yang cukup tinggi, yakni sekitar 85%, masih ditemukan kasus kesalahan klasifikasi *(mislabeling* atau *false positives)* dalam mendeteksi konten bermuatan *cyberbullying*.[6] Kesalahan tersebut tidak hanya mengakibatkan konten yang sebenarnya tidak melanggar menjadi diblokir, tetapi juga menimbulkan risiko pelanggaran terhadap hak kebebasan berekspresi pengguna. Lebih jauh, ketergantungan mutlak pada sistem AI tanpa adanya mekanisme verifikasi dari pengawas konten berpotensi menyebabkan kerugian, baik secara psikologis maupun reputasi, bagi pengguna yang terdampak. Kondisi ini menunjukkan bahwa moderasi berbasis AI masih memerlukan pengawasan dan evaluasi berkelanjutan, terutama untuk memastikan perlindungan hukum terhadap hak-hak pengguna di ruang digital.

Selain itu, penggunaan AI dalam penyaringan konten menghadirkan tantangan hukum terkait akuntabilitas dan regulasi. Dalam beberapa yurisdiksi, seperti Uni Eropa dan Amerika Serikat, perdebatan masih berlangsung mengenai sejauh mana *platform digital* bertanggung jawab atas keputusan yang dibuat oleh sistem AI *platform*.[7] Kurangnya regulasi yang jelas dapat menimbulkan ketidakpastian hukum bagi *platform* dan pengguna, serta memperburuk risiko sensor yang tidak proporsional terhadap kebebasan berekspresi di internet.

---

[6] Wahyu Hadikristanto, "Implementasi Content Moderation dalam Social Media Instagram untuk Deteksi Cyberbullying dengan Machine Learning Berbasis Cloud," *Indonesian Journal of Business Intelligence (IJUBI)* 5, no. 2 (2022): 122–126.

[7] Ralitza Dimitrova, *"Artificial Intelligence in Content Moderation–Legal Challenges and EU Legal Framework,"* in *2022 10th International Scientific Conference on Computer Science (COMSCI)*, May 2022, 1–6 (IEEE).

AI berpotensi menghasilkan informasi yang keliru secara tidak disengaja, suatu fenomena yang dikenal dengan istilah "halusinasi." Hal ini dapat terjadi ketika AI diberikan suatu tugas, Al seharusnya menghasilkan respon berdasarkan data dunia nyata, namun dalam beberapa kasus, Al akan memalsukan sumbernya, yang berarti ia "berhalusinasi". "Halusinasi" dapat dianggap sebagai informasi yang salah. Ini adalah sesuatu yang berada di bawah kendali dan bergantung pada pengembangan dan pelatihan model AI. Pengembang harus bertanggung jawab atas halusinasi yang menyebabkan kerugian. Namun ketika pengguna *platform* yang tidak bermoral dengan sengaja menghasilkan informasi palsu dengan menggunakan AI, itu biasanya disebut sebagai disinformasi.[8]

Wakil Menteri Komunikasi dan Informatika Republik Indonesia, Nezar Patria, menyatakan kekhawatirannya bahwa konten yang dihasilkan oleh Kecerdasan Buatan atau AI berpotensi menimbulkan kesalahan analisis yang dapat menyebabkan terjadinya misinformasi atau kekeliruan penyampaian informasi, meskipun tanpa adanya niat jahat. Hal ini disebabkan karena informasi dan data yang diproses oleh teknologi AI pada dasarnya merupakan hasil ciptaan manusia, sehingga tidak terlepas dari kemungkinan adanya kesalahan sistematis. Akibatnya, hasil olahan AI juga berpotensi mengandung bias bahkan bersifat diskriminatif. Oleh karena itu, seiring dengan perkembangan zaman, teknologi AI perlu terus ditingkatkan agar mampu beradaptasi secara berkelanjutan.[9]

Meskipun demikian, pengaturan mengenai kecerdasan buatan (*Artificial Intelligence*/AI) hingga saat ini belum dirumuskan secara eksplisit dalam kerangka hukum positif di berbagai negara. Di Indonesia, pengaturan terkait teknologi AI juga belum diakomodasi secara khusus dalam suatu undang-undang tersendiri. Kendati demikian, keberadaan dan pemanfaatan teknologi AI secara implisit dapat ditelusuri dalam Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan

---

[8] Merav Ozair, *"Misinformation in the Age of Artificial Intelligence and What It Means for the Markets,"* Nasdaq, 2023, https://www.nasdaq.com/articles/misinformation-in-the-age-of-artificial-intelligence-and-what-it-means-for-the-markets (diakses 23 Maret 2025, 21:35 WIB).
[9] Lenny Septiani, *"Teknologi AI Dikhawatirkan Menimbulkan Informasi Salah,"* Katadata, 2023, https://katadata.co.id/digital/teknologi/64e462a7c45a4/teknologi-ai-dikhawatirkan-menimbulkan-informasi-salah?page=2 (diakses 23 Maret 2025, 21:37 WIB).

Transaksi Elektronik sebagaimana telah diubah dengan Undang-Undang Nomor 19 Tahun 2016 (UU ITE). Karakteristik AI yang beroperasi melalui mekanisme otomatis dalam pemrosesan informasi memungkinkan teknologi ini untuk dikualifikasikan sebagai "Agen Elektronik", sebagaimana dimaksud dalam Pasal 1 angka 8 UU ITE, yang mendefinisikan agen elektronik sebagai perangkat dalam suatu sistem elektronik yang dirancang untuk melakukan tindakan tertentu terhadap informasi elektronik secara otomatis, *yang dijalankan oleh suatu program atas nama pengguna."* Pengertian ini dapat mencakup teknologi AI yang digunakan dalam sistem moderasi konten.[10] Selain UU ITE, sejumlah peraturan lainnya juga dapat menjadi dasar hukum yang relevan.[11] Misalnya, Peraturan Menteri Komunikasi dan Informatika (Permenkominfo) Nomor 5 Tahun 2020 tentang Penyelenggara Sistem Elektronik Lingkup Privat, yang mewajibkan platform digital untuk menangani konten yang dilarang dalam waktu 24 jam atau 4 jam dalam keadaan mendesak. Ketentuan ini secara tidak langsung menuntut kemampuan penyaringan konten yang efisien dan akurat, termasuk melalui teknologi otomatis seperti AI. Apabila kewajiban ini tidak dipenuhi, maka platform dapat dikenai sanksi administratif berupa teguran, denda, hingga pemutusan akses. Adapun Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik (PP PSTE). Peraturan ini memberikan kerangka hukum bagi Penyelenggara Sistem Elektronik (PSE), termasuk kewajiban dalam menjamin keamanan sistem dan perlindungan atas informasi elektronik. Dalam Pasal 14 dan Pasal 15 Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik (PP PSTE), disebutkan bahwa PSE wajib menyediakan mekanisme perlindungan terhadap pengguna media sosial dan menjamin agar sistem yang digunakan beroperasi secara andal, aman, dan bertanggung jawab, Ketentuan tersebut sangat relevan untuk konteks moderasi konten oleh AI, karena menunjukkan bahwa meskipun tindakan moderasi dilakukan oleh sistem otomatis, tanggung jawab hukum tetap melekat pada

[10] Raisa Safina, Khalda Alifia Azzahra, and Ananda Fersa Dharmawan, *"Kajian Yuridis Penggunaan Kecerdasan Artifisial pada Pembuatan dan Penyebaran Konten Pornografi di Media Sosial dalam Hukum Positif Indonesia," Mandub: Jurnal Politik, Sosial, Hukum dan Humaniora* 2, no. 1 (2023): 302–13, https://doi.org/10.59059/mandub.v2i1.918.
[11] Grenaldo Ginting dan M. P. N. Simamora, *Hukum Teknologi Informasi dan Komunikasi* (Medan: Fakultas Hukum Universitas Muhammadiyah Sumatera Utara, 2020), 33.

penyelenggara platform sebagai pengendali sistem elektronik.[12] Artinya, jika sistem penyaringan konten berbasis AI gagal menyaring konten secara tepat dan menimbulkan kerugian, maka tanggung jawab hukum tetap dapat dibebankan kepada PSE.

Melihat fenomena tersebut, jika terjadi kerugian yang disebabkan oleh AI pada konten atau fitur-fitur dalam media sosial salah satunya platform instagram, perlu dikaji mengenai bentuk tanggung jawab hukumnya. Di Indonesia, hingga saat ini belum terdapat regulasi maupun standar secara khusus yang mengatur terkait AI pada aplikasi media sosial. Sehingga adanya ketidakjelasan status AI di mata hukum. Pentingnya penelitian ini dikarenakan AI semakin berkembang di kalangan masyarakat, khusunya dalam konten yang ada pada media sosial. Teknologi ini akan terus berkembang di masa yang akan datang, sehingga memungkinkan terjadinya fenomena baru, AI dapat mengalami kegagalan dalam moderasi konten di media sosial, yang berpotensi menyebabkan penyebaran informasi yang keliru, lolosnya konten berbahaya dari pengawasan, atau bahkan terjadinya pembatasan berlebihan terhadap ekspresi atau diskusi yang sah secara hukum. Seiring dengan kemajuan teknologi AI, tantangan dalam menjaga keseimbangan antara kebebasan berbicara dan keamanan digital menjadi semakin rumit. Oleh karena itu, diperlukan regulasi yang lebih ketat, transparansi dalam algoritma, serta peran aktif manusia dalam proses penyaringan konten untuk meminimalkan kesalahan AI dalam menyaring dan mengevaluasi konten.

Penelitian ini dilakukan untuk meninjau tanggung jawab platform media sosial, khususnya Instagram, dalam penggunaan kecerdasan buatan AI untuk menyaring konten sensitif yang beredar di ruang digital. Penyaringan konten melalui AI dimaksudkan untuk menjaga ruang digital tetap aman, sehat, dan bebas dari konten bermuatan negatif, seperti ujaran kebencian, pornografi, maupun disinformasi. Namun, dalam praktiknya, sistem AI tidak selalu bekerja secara akurat. Terdapat sejumlah kasus di mana konten sensitif gagal disaring atau justru terjadi kesalahan

[12] S. M. T. Situmeang, *Cyber Law* (Medan: Fakultas Hukum Universitas Muhammadiyah Sumatera Utara, 2020), 11.

moderasi terhadap konten yang sebenarnya tidak melanggar. Selain itu, menarik untuk mengkaji bagaimana norma hukum yang berlaku di Indonesia mengatur perlindungan terhadap pengguna dan bentuk sanksi yang dapat dikenakan kepada penyelenggara sistem elektronik ketika gagal menjalankan kewajibannya. Penelitian ini juga akan menelaah bagaimana praktik penyaringan konten sensitif berbasis AI di Instagram dipahami secara normatif, sekaligus melihat dampak hukumnya terhadap para pihak yang dirugikan. Berdasarkan latar belakang, Penulis tertarik untuk melakukan penelitian skripsi dengan judul **"Tanggung Jawab Platform Media Sosial Instagram yang Mengalami Kegagalan Pada Proses Penyaringan Konten Menggunakan Kecerdasan Buatan *(Artificial Intelligence)*"**

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat beberapa permasalahan hukum yang menjadi fokus dalam penelitian ini, yaitu:

1. Bagaimana tanggung jawab hukum dari platform media sosial instagram atas kegaglan penyaringan konten oleh kecerdasan buatan?
2. Apa sanksi hukum bagi platform media sosial Instagram apabila kecerdasan buatan yang digunakan mengalami kegagalan dalam penyaringan konten sensitif?

1.3 Ruang Lingkup Penelitian

Berdasarkan permasalahan yang telah dipaparkan di atas, maka ruang lingkup penelitian ini yaitu:

1. Ruang Lingkup Keilmuan

   Ruang lingkup keilmuan pada penelitian ini yaitu berada dalam ranah hukum perdata, Khususnya pertanggungjawaban hukum dari platform media sosial di indonesia
2. Ruang Lingkup Kajian

   Ruang lingkup kajian pembahasan ialah pada kajian hukum (yuridis) terhadap penyaringan konten yang dilakukan oleh sistem berbasis kecerdasan buatan

(artificial intelligence) di media sosial, khususnya ketika terjadi kegagalan dalam proses penyaringan konten.

1.4 Tujuan Penelitian

Berdasarkan pada latar belakang yang telah diuraikan oleh penulis di atas, maka tujuan dari penelitian ini sebagai berikut:

1. Untuk mengetahui, memahami dan menganalisis bentuk-bentuk kegagalan kecerdasan buatan dalam penyaringan konten di media sosial, serta dampak hukum yang ditimbulkan dari kegagalan tersebut.
2. Untuk mengetahui, memahami dan menganalisis tanggung jawab hukum terkait, termasuk penyedia platform media sosial dan pengembang teknologi AI, atas kesalahan dalam penyaringan konten.

1.5 Kegunaan Penelitian

Berdasarkan kegunaan penelitian yang telah dipaparkan di atas, maka terdapat dua kegunaan penelitian yaitu:

1. Kegunaan Teoritis

   Hasil Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu hukum, khususnya dalam bidang hukum teknologi informasi dan komunikasi. Dengan membahas aspek yuridis dari penyaringan konten oleh kecerdasan buatan yang mengalami kegagalan di media sosial, penelitian ini dapat memperkaya literatur hukum yang membahas tanggung jawab hukum dan regulasi terkait penggunaan AI dalam penyaringan konten digital.
2. Kegunaan Praktis

   Hasil penelitian ini dilakukan sebagai upaya untuk mengembangkan pengetahuan hukum bagi penulis khususnya pemahaman lebih luas tentang AI yang ditinjau dari segi hukum. Selain itu, sebagai bahan tambahan informasi atau referensi bagi pihak-pihak yang memerlukan terutama mahasiswa hukum yang mana penelitiannya berkaitan dengan permasalahan dan pokok bahasan.

II TINJAUAN PUSTAKA

2.1 Tinjauan Umum Tentang *Artificial Intelligence (AI)*

2.1.1 Pengertian dan Sejarah Perkembangan *Artificial Intelligence (AI)*

Artificial Intelligence atau Kecerdasan Buatan adalah cabang ilmu komputer yang berfokus pada penciptaan sistem atau mesin yang mampu meniru kecerdasan manusia. Definisi ini pertama kali dikemukakan oleh John McCarthy pada tahun 1956, yang menyatakan bahwa AI adalah "ilmu membuat mesin melakukan hal-hal yang memerlukan kecerdasan ketika dilakukan oleh manusia" AI dirancang untuk belajar, memahami, berpikir, merasakan, dan bertindak seperti manusia. AI melibatkan pembuatan program dan algoritma yang memungkinkan mesin untuk belajar, merencanakan, dan menyelesaikan masalah, termasuk pengenalan pola, pemrosesan bahasa alami, dan pengambilan keputusan. Teknologi ini telah digunakan dalam berbagai bidang, termasuk pengenalan suara, pengenalan gambar, pemrosesan bahasa alami, pembelajaran mesin, dan mobil otonom.[13]

Menurut pendekatan ilmiah, AI adalah entitas cerdas yang diciptakan oleh manusia dan ditanamkan ke dalam mesin, sehingga mesin tersebut seolah-olah mampu berpikir sendiri untuk mengambil keputusan. Pendekatan teknik melihat AI sebagai sistem komputer yang dirancang untuk meniru kemampuan kognitif manusia.[14] Konsep AI telah ada sejak zaman kuno, namun sebagai bidang ilmu, AI mulai berkembang pada pertengahan abad ke-20. Pada tahun 1950, Alan Turing, seorang

---

[13] A. J. E. Oktavianus, L. Naibaho, and D. A. Rantung, *"Pemanfaatan Artificial Intelligence pada Pembelajaran dan Asesmen di Era Digitalisasi," Jurnal Kridatama Sains dan Teknologi* 5, no. 2 (2023): 473–86.

[14] Desi Azizah, Aji Wibawa, and Laksono Budiarto, *"Hakikat Epistemologi Artificial Intelligence," Jurnal Inovasi Teknologi dan Edukasi Teknik* 1, no. 8 (2021): 592–98, https://doi.org/10.17977/um068v1i82021p592-598.

matematikawan Inggris, mengusulkan *Tes Turing (Turing Test)* untuk mengukur kemampuan mesin dalam menunjukkan perilaku cerdas yang setara dengan manusia. Kemudian, pada tahun 1956, istilah *"Artificial Intelligence"* diperkenalkan dalam konferensi di Dartmouth College oleh John McCarthy, Marvin Minsky, Nathaniel Rochester, dan Claude Shannon, yang dianggap sebagai kelahiran resmi bidang AI.[15]

Turing Test, sebuah mesin dianggap berhasil jika mampu meyakinkan seseorang bahwa ia adalah manusia selama interaksi berlangsung. Tes ini biasanya dilakukan melalui komunikasi berbasis teks, seperti melalui keyboard. Alan Turing pernah mengusulkan konsep yang ia sebut *Child Machine* yakni sebuah mesin yang awalnya tidak terlalu pintar, tetapi kemudian dilatih dan diberi pengetahuan melalui proses pembelajaran. Ide ini berangkat dari gagasan bahwa meniru kecerdasan manusia dewasa secara langsung terlalu rumit, karena otak manusia sangat kompleks dan belum sepenuhnya dipahami. Maka, pendekatan yang lebih masuk akal adalah memulai dari bentuk kecerdasan yang lebih sederhana, seperti pada anak-anak. Meskipun hidup Turing berakhir di usia muda, kontribusinya sangat besar dan ia diakui sebagai pelopor dalam bidang Kecerdasan Buatan.[16]

Menurut Rich, *Artificial Intelligence* (AI) adalah metode yang memungkinkan komputer untuk menjalankan tugas serta menghasilkan *output* yang lebih optimal. Pandangan serupa juga diungkapkan oleh Staugaard dan Marvin Minsky, yang menyatakan bahwa AI merupakan cabang ilmu pengetahuan yang berfokus pada pengembangan mesin agar dapat melakukan seolah-olah bisa bekerja yang biasanya dilakukan oleh kemampuan manusia. AI bekerja dengan mengolah kumpulan data berukuran besar atau Big Data yang memiliki karakteristik unik dan tidak dapat ditangani oleh komputer konvensional. Dengan menggunakan kemampuan algoritma matematika, data tersebut diolah dan disimpan sebagai pengetahuan oleh

---

[15] P. A. W. Purnama, C. Fadhilah, C. A. Fadhilla, B. Wardana, J. Sumah, R. M. Thaniket, … and N. Pohan, *Artificial Intelligence* (Jakarta: Serasi Media Teknologi, 2025).
[16] Yusnaini, Y., Muhaimin, M., Firmansyah, F., Setiwan, Y. L., Bakhtiar, R., Aisyah, A., et al., *Artificial Intelligence dalam Perkembangan Teknologi Komunikasi* (Jakarta: CV. Gita Lentera, 2024), hlm. 11.

sistem AI, yang nantinya digunakan untuk mendukung proses pengambilan keputusan.

Keputusan yang diambil oleh sistem AI memiliki kemiripan dengan cara otak manusia memproses dan menghasilkan keputusan. Tidak seperti program komputer biasa, AI mampu menjalankan berbagai tugas yang pada umumnya dilakukan oleh manusia, bahkan dalam beberapa kasus, dapat melampaui kemampuan manusia dalam menyelesaikan pekerjaan yang lebih rumit. Inilah yang menjadi dasar munculnya pandangan bahwa AI memiliki kecerdasan buatan yang meniru kemampuan intelektual manusia.[17]

Artificial Intelligence dapat melakukan empat faktor berikut[18] *Acting humanly*, Merujuk pada sistem yang mampu berperilaku menyerupai kebiasaan dan respons manusia dalam berbagai situasi. *Thinking humanly*, Menggambarkan sistem yang proses berpikirnya bisa disandingkan dengan cara manusia berpikir, termasuk dalam hal persepsi dan pengambilan keputusan. *Think rationlly*, Merupakan sistem yang dapat mengolah dan menilai informasi berdasarkan logika dan objektivitas.
Act rationally, Mengacu pada sistem yang mampu membuat keputusan dan bertindak secara logis serta efisien untuk mencapai tujuan tertentu.

Perkembangan kecerdasan buatan *(artificial intelligence)* di Indonesia tidak terlepas dari kemajuan teknologi informasi dan komunikasi yang mulai berkembang sejak akhir abad ke-20. Pada tahap awal, sekitar akhir 1990-an hingga awal 2000-an, penelitian AI di Indonesia masih bersifat akademik dan terbatas pada lingkungan perguruan tinggi. Bentuk penerapan AI pada masa ini umumnya berupa sistem pakar *(expert system)*, pengolahan bahasa alami, dan pengenalan pola sederhana yang dikembangkan untuk keperluan penelitian dan pembelajaran, tanpa pemanfaatan luas dalam sektor industri maupun pelayanan publik.

---

[17] Yolanda Simbolon, "Pertanggungjawaban Perdata Terhadap Artificial Intelligence yang Menimbulkan Kerugian Menurut Hukum di Indonesia," *Veritas et Justitia* 9, no. 1 (2023): 246–73, https://doi.org/10.25123/vej.v9i1.6037.
[18] Quceny Praviyanti Anjani, Tesis: "Chatbot Menggunakan Natural Language Processing (NLP) pada Toko Bunga Online" (Jakarta Selatan, Universitas Nasional, 2023), hlm. 15.

Memasuki dekade 2010-an, perkembangan AI di Indonesia mulai mengalami percepatan seiring dengan meningkatnya penetrasi internet, kemajuan komputasi, serta ketersediaan data dalam jumlah besar *(big data).* Pada fase ini, AI mulai dimanfaatkan oleh sektor swasta, khususnya dalam bidang e-commerce, perbankan digital, transportasi daring, dan media sosial. Teknologi AI digunakan untuk sistem rekomendasi, analisis perilaku pengguna, deteksi penipuan, serta moderasi konten otomatis. Dalam konteks media sosial, AI mulai berperan penting dalam menyaring konten bermuatan pornografi, ujaran kebencian, dan disinformasi, meskipun akurasinya masih menghadapi berbagai keterbatasan teknis.

Kesadaran negara terhadap pentingnya pengembangan dan pengaturan AI secara nasional mulai menguat pada akhir dekade 2010-an. Hal ini ditandai dengan diluncurkannya Strategi Nasional Kecerdasan Artifisial (Stranas KA) oleh pemerintah Indonesia pada tahun 2020 melalui Kementerian Perencanaan Pembangunan Nasional/Bappenas. Dokumen Stranas KA menegaskan bahwa AI merupakan teknologi strategis yang harus dikembangkan secara bertanggung jawab dengan memperhatikan aspek etika, keamanan, perlindungan data, dan kepentingan publik. Stranas KA juga menetapkan lima sektor prioritas pengembangan AI, yaitu layanan kesehatan, reformasi birokrasi, pendidikan dan riset, ketahanan pangan, serta mobilitas dan kota cerdas.

2.1.2 Jenis-Jenis Artificial Intelligence (AI)

Ada berbagai macam jenis AI yang digunakan untuk memudahkan semua pekerjaan yang tidak bisa dilakukan oleh manusia normalnya, Inilah, beberapa jenis AI yang telah banyak digunakan:

1. ***Machine Learning*** **(ML)**

   Machine Learning (ML) atau pembelajaran mesin merupakan salah satu cabang dari kecerdasan buatan AI yang memungkinkan sistem komputer untuk belajar dari data dan pengalaman tanpa harus diprogram secara eksplisit untuk setiap tugas tertentu. Proses ini melibatkan pengumpulan data historis, pelatihan model

menggunakan algoritma tertentu, dan kemudian penerapan model tersebut untuk membuat prediksi atau klasifikasi terhadap data baru yang belum pernah dilihat sebelumnya. Salah satu penerapan yang sangat signifikan dari ML dalam dunia nyata adalah pada sektor keuangan, khususnya dalam hal deteksi penipuan *(fraud detection).* Penipuan dalam transaksi keuangan, seperti penyalahgunaan kartu kredit, transfer dana yang mencurigakan, dan manipulasi data keuangan, dapat menyebabkan kerugian besar bagi institusi perbankan dan nasabah. Oleh karena itu, deteksi dini terhadap pola-pola transaksi yang tidak biasa menjadi sangat penting.[19]

Teknologi ML memungkinkan sistem untuk menganalisis jutaan data transaksi keuangan elektronik yang telah terjadi, kemudian mengidentifikasi pola dan karakteristik dari transaksi-transaksi yang terbukti merupakan penipuan. Dengan memahami pola tersebut, sistem dapat mengembangkan model prediktif yang mampu mengidentifikasi transaksi mencurigakan secara *real-time*, bahkan sebelum kerugian terjadi. Keunggulan penggunaan ML dalam mendeteksi penipuan antara lain adalah kecepatan dalam memproses data dalam jumlah besar, kemampuan untuk terus belajar dari data terbaru, serta adaptasi terhadap teknik penipuan baru yang terus berkembang. Selain itu, sistem ML dapat membantu analis keuangan dalam membuat keputusan yang lebih cepat dan akurat berdasarkan hasil prediksi yang dihasilkan oleh model. Meskipun demikian, implementasi ML dalam sistem perbankan juga memiliki tantangan, seperti ketersediaan data yang berkualitas, perlindungan privasi nasabah, serta kebutuhan terhadap tenaga ahli dalam bidang data science dan keamanan siber. Oleh karena itu, pengembangan sistem deteksi penipuan berbasis ML perlu didukung dengan infrastruktur yang memadai serta kerjasama lintas disiplin antara ahli teknologi, keuangan, dan hukum.[20]

---

[19] Faried Zamachsari dan Niken Puspitasari, *"Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik," Jurnal RESTI* 5, no. 2 (2021): 204.
[20] Imanuel Adhitya Wulanata Chrismastianto, *"Efektivitas Layanan Keuangan Berbasis Machine Learning sebagai Komponen Pendukung Kebijakan Makroprudensial Pascapandemi Covid-19," Jurnal Universitas Kristen Immanuel* (2021): hlm, 258.

2. *Natural Language Processing* (NLP)

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan yang berfokus pada bagaimana mesin dapat memahami, menafsirkan, dan merespons bahasa manusia secara alami. Bahasa alami adalah jenis bahasa yang digunakan sehari-hari oleh manusia untuk berkomunikasi, termasuk saat menggambarkan suatu tempat atau kondisi. Dalam NLP, data yang diolah umumnya berupa teks yang tidak terstruktur, bukan dalam bentuk tabel atau format terorganisir lainnya.[21] Teks dan bahasa memiliki struktur serta makna yang lebih kompleks dan beragam, sehingga dinilai lebih cocok untuk pemrosesan NLP dibandingkan dengan data tabular biasa. Salah satu aplikasi umum dari NLP adalah *ChatBot* atau asisten virtual yang mampu berinteraksi dengan pengguna melalui percakapan.

3. *Computer Vision* (CV)

Computer Vision (CV) merupakan salah satu cabang dari *Artificial Intelligence* (AI) yang berfokus pada bagaimana komputer dapat memahami dan menginterpretasikan informasi dari gambar atau video digital. Misalnya, saat terdapat objek dalam sebuah gambar, sistem komputer dapat mengidentifikasi dan memberikan informasi terkait objek tersebut kepada manusia. Agar komputer mampu menjalankan fungsi ini, dibutuhkan berbagai teknik dan algoritma khusus yang dipelajari dalam bidang *Computer Vision*.[22]

Teknologi ini memungkinkan perangkat untuk mengenali dan membaca objek melalui sensor atau kamera, kemudian mengolah data tersebut menjadi perintah atau fungsi tertentu agar dapat melakukan suatu aksi sesuai informasi objek yang terdeteksi. *Computer Vision* mengintegrasikan kamera, pemrosesan komputasi (baik berbasis edge maupun cloud), perangkat lunak, dan

---

[21] Elias Hossain, R. Rana, N. Higgins, J. Soar, P. Barua, Anthony R. Pisani, and K. Turner, "Natural Language Processing in Electronic Health Records in Relation to Healthcare Decision-Making: A Systematic Review," *Computers in Biology and Medicine* 155 (2023): 106649, https://doi.org/10.1016/j.compbiomed.2023.106649.

[22] Catur Supriyanto, "Pentingnya Computer Vision pada Situs Jual Beli Online," *Kompasiana*, 2021, https://www.kompasiana.com/catursupriyanto/602206661730b90ef37dba82/computer-vision-dalam-situs-jual-beli (diakses 28 Maret 2025, 20:43 WIB).

kecerdasan buatan agar sistem mampu "melihat" dan mengenali lingkungan sekitarnya. Salah satu contoh penerapan CV dalam kehidupan sehari-hari adalah proses pemindaian kode QR, seperti yang digunakan dalam sistem pembayaran digital *QRIS*.

4. Predictive Analytics (PA)

Predictive Analytics (PA) merupakan pendekatan dalam analisis data yang memanfaatkan algoritma, metode statistik, serta teknik machine learning untuk memperkirakan kejadian yang mungkin terjadi di masa depan dengan mengacu pada pola dari data sebelumnya. Menurut SAS *Institute*, perusahaan yang mengembangkan perangkat lunak analisis data, penggunaan data historis memungkinkan organisasi untuk mengambil keputusan yang lebih tepat dan strategis.

Dalam sektor keuangan, PA sering diterapkan untuk mengidentifikasi potensi penipuan kartu kredit maupun aktivitas transaksi yang mencurigakan. Dengan menganalisis kebiasaan pengeluaran dan pola transaksi para nasabah, institusi keuangan dapat mengenali penyimpangan yang mencurigakan serta menerapkan langkah-langkah pencegahan sebelum terjadinya tindakan penipuan.[23]

5. *Blockchain* dan Teknologi Terdesentralisasi

Blockchain merupakan sebuah teknologi yang memanfaatkan sistem komputer untuk membentuk rangkaian blok data yang saling terhubung.[24] Setiap blok menyimpan informasi transaksi serta digunakan untuk melacak aset dalam suatu jaringan perusahaan. Untuk menciptakan transaksi yang transparan, teknologi *Blockchain* membutuhkan aplikasi khusus yang berfungsi mengotomatiskan proses operasional sistem dalam jaringan tersebut. Menurut Klitos Christodoulou et al., aplikasi terdesentralisasi atau *decentralized*

[23] Mavellyno Vedhitya, "Predictive Analytics, Memprediksi Masa Depan lewat Pola Data," *Marketeers*, 2023, https://www.marketeers.com/predictive-analytics-memprediksi-masa-depan-pola-data/ (diakses 28 Maret 2025, 21:15 WIB).
[24] Muh. Akbar Fhad Syahril, *Hukum Informasi dan Transaksi Elektronik* (CV. Eureka Media Aksara, 2023), hlm. 77.

applications (dApps) dijalankan melalui jaringan komputer *peer-to-peer*, bukan melalui satu server pusat, dan dirancang agar tidak dikendalikan oleh satu entitas tunggal melalui internet. Aplikasi terdesentralisasi dalam teknologi *blockchain* berperan penting dalam menjaga transparansi dalam setiap transaksi yang terjadi. Fungsinya adalah untuk memastikan bahwa sistem *peer-to-peer* dapat diterapkan secara efektif, sehingga proses transaksi berlangsung dengan baik dan sesuai dengan tujuan dari aplikasi tersebut. Selain itu, aplikasi ini juga dirancang agar dapat diakses dengan mudah oleh para pengguna dalam jaringan.[25]

6. *Robotic Process Automation* (RPA)

Robotic Process Automation (RPA) merupakan teknologi yang memungkinkan otomatisasi proses-proses bisnis yang bersifat berulang dengan bantuan *software robot* atau *bots.* RPA dirancang untuk meniru interaksi manusia dengan sistem digital, seperti memasukkan data, menavigasi sistem, dan memproses transaksi. Dengan RPA, organisasi dapat meningkatkan efisiensi operasional, mengurangi kesalahan manusia, serta menghemat waktu dan biaya. Teknologi ini biasanya diterapkan pada proses-proses administratif seperti pemrosesan faktur, entri data, dan manajemen email. Tidak seperti sistem otomatisasi tradisional, *Robotic Process Automation* (RPA) dapat diterapkan tanpa perlu melakukan perubahan besar pada sistem yang sudah ada, karena bots bekerja pada lapisan antarmuka pengguna *(user interface).*[26]

---

[25] Dimas Agung Pangestu, *Penggunaan Teknologi Blockchain dalam Transaksi Keuangan Syari'ah* (2023), hlm. 1–102.

[26] Jorge Ribeiro et al., "Robotic Process Automation and Artificial Intelligence in Industry 4.0 - A Literature Review," *Procedia Computer Science* 181 (2021): 51–58, https://doi.org/10.1016/j.procs.2021.01.104.

2.2 Tinjauan Umum Tentang Penyaringan Konten

2.2.1 Pengertian dan Dasar Hukum Penyaringan Konten

Penyaringan konten atau disebut *moderasi content* merupakan proses pengawasan dan pengendalian terhadap informasi yang disebarluaskan melalui platform digital, khususnya media sosial. Tujuan utamanya adalah untuk memastikan bahwa konten yang beredar tidak melanggar hukum, norma sosial, atau etika yang berlaku. Dalam konteks ini, penyaringan konten berperan penting dalam menjaga ruang digital yang aman dan kondusif bagi masyarakat. Penyaringan konten merujuk pada proses pengawasan dan pengendalian terhadap informasi yang disebarkan melalui platform digital, khususnya media sosial, guna memastikan kesesuaian dengan norma hukum dan etika yang berlaku. Proses ini melibatkan identifikasi dan penanganan konten yang dianggap melanggar, seperti ujaran kebencian, pornografi, dan disinformasi. Dalam era digital, penyaringan konten menjadi krusial untuk menjaga ruang publik yang sehat dan aman bagi pengguna internet.[27]

Penyaringan konten dapat dipahami sebagai upaya perusahaan dalam memfilter dan mengawasi setiap konten yang diunggah ke *platform* daring mereka, guna memastikan kesesuaiannya dengan aturan internal serta pedoman komunitas yang berlaku. Setelah melewati proses seleksi, perusahaan akan menetapkan apakah konten tersebut layak untuk dipublikasikan atau perlu ditolak. Pada saat pengguna mengirimkan konten ke situs web, konten tersebut harus melalui tahapan penyaringan untuk menjamin bahwa isi pesan tidak melanggar norma, tidak mengandung unsur penghinaan, serta selaras dengan ketentuan pedoman konten yang telah ditentukan. Dalam praktiknya, penyaringan konten banyak diterapkan di *platform* berbasis konten buatan pengguna *(user-generated content)* seperti media sosial, *e-commerce*, bursa saham daring, situs kencan, komunitas online, dan forum diskusi. Mengingat tingginya volume konten yang diproduksi secara instan, platform-platform ini memiliki potensi besar menjadi media penyebaran konten bermasalah apabila tidak diawasi dengan ketat.

---

[27] Sindy Prasetyo, "Pelanggaran Hak Asasi Manusia di Indonesia," *Indigenous Knowledge* 2, no. 1 (2023): 51–57.

Permasalahan dalam penyaringan konten tidak hanya disebabkan oleh lemahnya penegakan hukum, melainkan juga oleh keterbatasan kapasitas teknologi yang tersedia serta rendahnya tingkat koordinasi antar lembaga yang memiliki kewenangan terkait. Ketiga faktor ini menjadi tantangan utama dalam menciptakan ekosistem digital yang aman dan tertib. Menurut *Lawrence Lessig* adalah seorang akademisi dan aktivis politik Amerika, khususnya dalam berbagai penerapan teknologi, keberhasilan dalam mengatur dunia digital tidak hanya bergantung pada hukum formal, melainkan juga pada interaksi antara norma sosial, desain arsitektur teknologi, dan kekuatan pasar. Dalam konteks penyaringan konten, kolaborasi lintas sektor ini menjadi penting untuk menciptakan pendekatan yang holistik dan efektif.[28]

Peran kecerdasan buatan *(Artificial Intelligence)* dalam penyaringan konten semakin meningkat seiring dengan perkembangan teknologi. *Platform* media sosial seperti Instagram menggunakan AI untuk menyaring konten yang dianggap melanggar ketentuan. Namun, penggunaan AI juga menimbulkan tantangan, seperti kesalahan dalam identifikasi konten yang melanggar atau tidak, serta potensi bias algoritma yang dapat merugikan pengguna tertentu. Media sosial telah menjadi sarana utama dalam penyebaran informasi dan ekspresi pendapat. Namun, kebebasan ini seringkali disalahgunakan untuk menyebarkan konten negatif yang dapat merusak tatanan sosial. Oleh karena itu, moderasi konten berperan penting dalam menyeimbangkan antara kebebasan berekspresi dan perlindungan terhadap masyarakat dari dampak konten berbahaya.[29]

Kegagalan AI dalam penyaringan konten dapat berdampak pada hak kebebasan berekspresi pengguna. Beberapa studi menunjukkan bahwa regulasi yang terlalu ketat atau penggunaan AI yang tidak akurat dapat menghambat kebebasan berbicara

[28] M. Hanisch, Curtis M. Goldsby, N. Fabian, and J. Oehmichen, "Digital Governance: A Conceptual Framework and Research Agenda," *Journal of Business Research* (2023), https://doi.org/10.1016/j.jbusres.2023.113777.
[29] Jam'ul Ihsan Bambang, Nadhratun Najwa, dan Muhammad Risky Rahmadani, *"Kebebasan Berbicara di Media Sosial: Antara Regulasi dan Ekspresi," Student Research Journal* 3, no. 1 (2025): 87-96, https://doi.org/10.55606/srj-yappi.v3i1.1692.

di ruang publik virtual.[30] Oleh karena itu, diperlukan keseimbangan antara perlindungan terhadap penyebaran konten berbahaya dan penghormatan terhadap hak asasi manusia, khususnya kebebasan berekspresi. Dalam konteks hukum perdata, kegagalan moderasi konten oleh AI dapat menimbulkan tanggung jawab hukum bagi platform media sosial. Jika pengguna mengalami kerugian akibat kesalahan penyaringan, mereka dapat menuntut ganti rugi berdasarkan prinsip perbuatan melawan hukum atau wanprestasi. Hal ini menekankan pentingnya *platform* untuk memastikan bahwa sistem penyaringan mereka, termasuk yang berbasis AI, berfungsi dengan baik dan tidak merugikan pengguna.

Di Indonesia, moderasi konten belum diatur secara spesifik dalam satu regulasi khusus. Namun, beberapa peraturan perundang-undangan memberikan dasar hukum bagi praktik ini. Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE), yang telah diubah dengan UU Nomor 19 Tahun 2016, mengatur tentang larangan penyebaran informasi yang bermuatan negatif serta tanggung jawab penyelenggara sistem elektronik. Selain itu, Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik mengharuskan penyelenggara sistem elektronik untuk menjamin keamanan dan keandalan sistemnya, termasuk dalam hal moderasi konten.[31] Selain UU ITE, Peraturan Menteri Komunikasi dan Informatika Nomor 5 Tahun 2020 tentang Penyelenggara Sistem Elektronik Lingkup Privat juga menjadi acuan penting dalam moderasi konten, Peraturan ini mewajibkan *platform digital* untuk secara aktif memantau dan menghapus konten yang melanggar hukum, serta memberikan tanggung jawab lebih besar kepada penyelenggara *platform* dalam menjaga keamanan informasi pengguna media sosial seperti instagram.

---

[30] Danrivanto Budhijanto, *Hukum Keamanan Siber* (Bandung: Logoz Publishing, 2024), 24.
[31] Rabith Madah Khulaili Harsya, Filep Wamafma, Marius Supriyanto Sakmaf, dan Andri Triyantoro, *"Regulasi Konten Online dan Dampaknya terhadap Hak Kebebasan Berbicara di Platform Digital di Indonesia," Sanskara Hukum dan HAM* 3, no. 1 (2024): 43–52, https://doi.org/10.58812/shh.v3i01.446.

2.2.2 Jenis-jenis Penyaringan Konten Di Media Sosial

Ada berbagai macam jenis penyaringan atau moderasi di media sosial yang diterapkan untuk menjaga kualitas dan keamanan konten, inilah lima tahap utama yang dilalui data sebelum mendapatkan bentuk dan bentuk yang tepat:

1. Moderasi Pra-Publikasi *(Pre-Moderation)*

   Moderasi pra-publikasi merupakan bentuk pengawasan konten di mana setiap materi yang diunggah oleh pengguna harus melalui proses pemeriksaan terlebih dahulu sebelum dapat diakses oleh publik. Proses ini dapat dilakukan baik oleh tim moderator manusia maupun melalui sistem otomatis berbasis kecerdasan buatan. Mekanisme ini umumnya dipilih oleh platform yang memiliki segmen pengguna dengan tingkat kerentanan tinggi, misalnya *platform* yang ditujukan untuk anak-anak atau remaja, dengan tujuan memastikan bahwa konten yang beredar tidak mengandung unsur berbahaya seperti pornografi, ujaran kebencian, atau kekerasan.

   Kelebihan dari model pra-publikasi adalah memberikan jaminan perlindungan yang lebih tinggi terhadap penyebaran konten ilegal atau berbahaya. Namun, sistem ini juga memiliki kelemahan yang signifikan, antara lain keterlambatan dalam publikasi yang dapat menurunkan interaktivitas dan menghambat dinamika komunikasi antar pengguna. Dalam konteks media sosial dengan arus informasi yang cepat, model ini sering dianggap kurang efisien, meskipun tetap relevan ketika keamanan dan perlindungan pengguna menjadi prioritas utama.[32]

2. Moderasi Pasca-Publikasi *(Post-Moderation)*

   Moderasi pasca-publikasi merupakan mekanisme pengawasan di mana konten yang diunggah oleh pengguna dapat langsung ditampilkan di platform tanpa melalui proses penyaringan awal. Setelah konten tersebut dipublikasikan, barulah dilakukan proses peninjauan oleh moderator, baik secara manual

[32] Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven, CT: Yale University Press, 2018), 46–48, https://doi.org/10.12987/9780300235029.

maupun dengan bantuan sistem berbasis kecerdasan buatan. Selain itu, proses pengawasan juga dapat dilakukan melalui sistem pelaporan oleh pengguna *(user report)* yang memungkinkan komunitas turut berpartisipasi dalam menjaga keamanan platform.

Model moderasi ini banyak digunakan oleh platform media sosial berskala besar, seperti Instagram, karena dianggap lebih efisien dalam menjaga kelancaran arus komunikasi. Keunggulan utama sistem ini adalah kecepatan dalam menampilkan konten sehingga interaksi antar pengguna tetap dinamis. Namun demikian, kelemahannya terletak pada potensi risiko yang ditimbulkan, yakni konten berbahaya atau melanggar hukum dapat sempat diakses oleh publik sebelum berhasil dihapus. Hal ini menimbulkan persoalan tersendiri, terutama terkait dampak psikologis maupun sosial yang ditimbulkan akibat keterlambatan penindakan oleh moderator. Dengan demikian, meskipun *post-moderation* mendukung keterbukaan komunikasi, efektivitasnya tetap dipertanyakan dalam konteks perlindungan hak-hak pengguna dan pencegahan penyebaran konten berbahaya.[33]

3. Moderasi Reaktif *(Reactive Moderation)*

Moderasi reaktif merupakan model pengawasan konten yang sepenuhnya bergantung pada laporan dari pengguna (user-generated reports). Dalam sistem ini, moderator baru akan mengambil tindakan apabila terdapat laporan terhadap suatu konten yang dianggap melanggar ketentuan. Model ini dipandang lebih ringan dari sisi operasional, karena platform tidak perlu melakukan pemantauan secara terus-menerus terhadap seluruh konten yang diunggah.

Namun, kelemahan utama dari moderasi reaktif terletak pada ketidakpastian efektivitasnya. Apabila pengguna tidak aktif dalam melaporkan pelanggaran, maka banyak konten berbahaya dapat lolos dari pengawasan dan tetap beredar di ruang publik digital. Sebaliknya, sistem ini juga rawan disalahgunakan, misalnya ketika fitur pelaporan digunakan untuk menyerang atau menjatuhkan

[33] A. Veglis, *Moderation Techniques for Social Media Content* (Cham: Springer, 2014), 137–138.

pihak tertentu dengan melabeli konten mereka sebagai pelanggaran, padahal sebenarnya tidak demikian. Kondisi ini berpotensi menimbulkan bias dan ketidakadilan dalam praktik moderasi. Oleh karena itu, meskipun reaktif moderation efisien dari segi biaya dan sumber daya, penerapannya tetap membutuhkan mekanisme verifikasi yang ketat serta pengawasan berlapis agar tidak menimbulkan kerugian bagi pengguna yang menjadi korban penyalahgunaan laporan.[34]

4. Moderasi Komunitas *(Distributed Moderation)*

   Moderasi berbasis komunitas merupakan model pengawasan konten yang menempatkan peran aktif pengguna sebagai pihak yang menilai kesesuaian konten dengan aturan platform. Dalam sistem ini, komunitas diberi kewenangan untuk melakukan evaluasi, misalnya dengan memberikan vote positif atau negatif, menandai *(flagging)*, atau melaporkan konten yang dianggap melanggar. Model ini kerap dipandang lebih demokratis dan partisipatif, karena proses moderasi tidak hanya bergantung pada pihak penyedia platform, melainkan melibatkan pengguna secara langsung sebagai pengawas.

   Meskipun demikian, model ini memiliki keterbatasan yang signifikan. Potensi bias sangat mungkin terjadi apabila mayoritas anggota komunitas tidak objektif dalam melakukan penilaian, misalnya karena pengaruh preferensi politik, budaya populer, atau kepentingan kelompok tertentu. Hal ini dapat menyebabkan konten yang sebenarnya sah atau tidak bermasalah tetap diturunkan hanya karena tidak disukai oleh mayoritas. Selain itu, sistem ini juga rentan terhadap *brigading*, yaitu tindakan terkoordinasi oleh sekelompok pengguna untuk menjatuhkan konten atau akun tertentu. Dengan demikian, meskipun moderasi berbasis komunitas mampu meningkatkan partisipasi pengguna, penerapannya tetap memerlukan pengawasan tambahan dan

[34] Hodijah, Cucu; Merry Nirmala Yani; dan Mohamad Sajili, *Komunikasi Bisnis dalam Era Artificial Intelligence* (Padang: Takaza Innovatix Labs, 2025), hlm. 13.

mekanisme korektif dari *platform* agar tidak menimbulkan diskriminasi maupun pelanggaran terhadap kebebasan berekspresi.[35]

5. Moderasi Otomatis *(Automated Moderation)*

   Moderasi otomatis merupakan bentuk moderasi konten yang mengandalkan sistem algoritmik berbasis kecerdasan buatan (AI). Dalam model ini, konten yang diunggah oleh pengguna secara langsung dianalisis oleh sistem melalui teknologi *machine learning* maupun *natural language processing* (NLP), dengan tujuan untuk mendeteksi kata kunci, pola bahasa, gambar, maupun perilaku tertentu yang dikategorikan sebagai pelanggaran. Keunggulan utama dari sistem ini adalah efisiensi dan kecepatan, karena mampu memproses jutaan konten dalam waktu singkat, sehingga dianggap relevan untuk *platform* media sosial berskala besar seperti Instagram.

   Namun, kelemahan dari moderasi berbasis AI terletak pada keterbatasannya dalam memahami konteks yang kompleks. AI sering kali mengalami kesulitan dalam membedakan antara konten yang benar-benar melanggar dengan konten yang bersifat satir, edukatif, atau memiliki makna ganda. Kegagalan ini dapat memunculkan dua permasalahan, yaitu *false positives* (konten sah yang justru ditandai sebagai pelanggaran) dan *false negatives* (konten berbahaya yang lolos dari deteksi sistem). Kondisi tersebut tidak hanya menimbulkan kerugian bagi pengguna secara psikologis maupun reputasional, tetapi juga dapat menimbulkan pertanggungjawaban hukum bagi *platform* apabila dinilai lalai dalam melaksanakan kewajiban penyaringan konten.[36]

2.2.3 Bentuk Kegagalan Moderasi Konten oleh AI

Dalam era digital saat ini, kecerdasan buatan *(Artificial Intelligence)* telah menjadi teknologi kunci dalam proses moderasi konten di berbagai platform media sosial. Namun, penggunaan AI belum sepenuhnya mampu menggantikan peran manusia

[35] Menerapkan Strategi Moderasi yang Efektif dalam Platform Komunitas Online, *Puskomedia*, diakses 29 Maret 2025, https://puskomedia.id/blog/menerapkan-strategi-moderasi-yang-efektif-dalam-platform-komunitas-online/.
[36] "Moderasi Otomatis," *Ekrut Media*, diakses 29 Maret 2025, https://www.ekrut.com/media/content-moderation.

secara sempurna, karena terdapat sejumlah bentuk kegagalan dalam operasionalnya. Salah satu bentuk kegagalan utama adalah *false positive*, yaitu kesalahan di mana sistem AI menghapus konten yang sebenarnya sah dan tidak melanggar aturan. Kondisi ini biasanya terjadi karena AI tidak mampu memahami konteks secara mendalam, sehingga konten satire, kritik sosial, atau diskusi akademik dapat secara keliru diklasifikasikan sebagai pelanggaran.[37] Sebaliknya, terjadi pula *false negative*, yakni ketika sistem gagal mendeteksi dan menghapus konten yang mengandung unsur berbahaya seperti ujaran kebencian, pornografi anak, atau disinformasi terkait isu politik dan kesehatan. False negative sering kali dipicu oleh keterbatasan dataset pelatihan yang digunakan dalam membangun algoritma, sehingga sistem menjadi kurang efektif dalam mengenali bentuk-bentuk ekspresi baru atau bahasa-bahasa lokal yang kompleks.

Selain itu, penyaringan konten berbasis AI juga rentan terhadap bias algoritmik terjadi ketika kesalahan sistematis dalam algoritma *machine learning* menghasilkan hasil yang tidak adil atau diskriminatif. Bias ini dapat terjadi karena data yang digunakan untuk melatih AI cenderung merefleksikan ketidaksetaraan yang ada dalam masyarakat. Misalnya, AI mungkin lebih sering menandai konten dari kelompok minoritas atau komunitas tertentu sebagai pelanggaran dibandingkan dengan konten dari kelompok mayoritas.[38] Bias semacam ini bukan hanya mengancam keadilan dalam moderasi, tetapi juga memperparah ketidaksetaraan sosial yang telah ada. Kurangnya sensitivitas terhadap konteks budaya juga merupakan bentuk kegagalan lainnya. AI yang dikembangkan dalam konteks budaya tertentu mungkin gagal memahami ekspresi idiomatik, humor, atau norma sosial dari budaya lain. Hal ini menyebabkan penyaringan konten menjadi tidak efektif, bahkan kadang menimbulkan ketegangan sosial karena dianggap mengekang ekspresi budaya tertentu.

---

[37] Robert Gorwa, Reuben Binns, and Christian Katzenbach, "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance," *Big Data & Society* 7, no. 1 (2020): 1–15.

[38] Sarah Myers West, Meredith Whittaker, and Kate Crawford, *Discriminating Systems: Gender, Race and Power in AI* (New York: AI Now Institute, 2019), 12–19, https://ainowinstitute.org/discriminatingsystems.pdf.

Kegagalan penyaringan konten berbasis AI membawa dampak serius dalam berbagai dimensi, mulai dari aspek sosial, hukum, hingga ekonomi. Dari sudut pandang sosial, kesalahan dalam menghapus konten sah *(false positive)* dapat berujung pada pembatasan kebebasan berekspresi, yang merupakan hak fundamental setiap individu sebagaimana diatur dalam berbagai instrumen hak asasi manusia. Ketika pengguna merasa bahwa pendapat mereka dibungkam tanpa alasan yang jelas, hal ini berpotensi menimbulkan ketidakpercayaan terhadap *platform* digital dan mendorong migrasi pengguna ke *platform* lain. Sebaliknya, kelalaian dalam menghapus konten berbahaya *(false negative)* dapat menyebabkan penyebaran ujaran kebencian, radikalisasi online, serta meluasnya disinformasi, yang pada akhirnya mengancam keamanan nasional dan kohesi sosial masyarakat. Misalnya, kegagalan mendeteksi kampanye disinformasi politik dapat memengaruhi hasil pemilu dan menggerus legitimasi demokrasi.

Dari perspektif hukum, kegagalan penyaringan konten dapat menimbulkan tanggung jawab hukum terhadap *platform* penyedia layanan. Di beberapa yurisdiksi, perusahaan digital dapat dikenai sanksi atau tuntutan hukum atas konten ilegal yang beredar di platform mereka jika terbukti lalai dalam menjalankan penyaringan. Kondisi ini menempatkan perusahaan dalam posisi yang sulit, antara harus melindungi kebebasan berbicara pengguna dan memenuhi kewajiban hukum untuk mencegah penyebaran konten ilegal.[39] Dalam jangka panjang, kegagalan penyaringan juga berdampak pada reputasi perusahaan. *Platform* yang dianggap gagal melindungi pengguna dari konten berbahaya berisiko kehilangan kepercayaan publik, yang pada gilirannya memengaruhi loyalitas pengguna dan nilai pasar perusahaan. Kasus-kasus besar penyebaran informasi palsu di instagram seperti Provokasi menuju aksi kekerasan, informasi yang disebar di Instagram dapat memprovokasi terjadinya tindakan kekerasan di dunia nyata, seperti kasus pengeroyokan yang berawal dari unggahan di media sosial, menunjukkan bagaimana ketidakmampuan penyaringan dapat menyebabkan konsekuensi finansial dan reputasional yang serius.

---

[39] S. Mauludi, *Seri Cerdas Hukum: Awas Hoax! Cerdas Menghadapi Pencemaran Nama Baik, Ujaran Kebencian & Hoax* (Jakarta: Elex Media Komputindo, 2019), hlm. 57.

Oleh karena itu, pengembangan sistem AI untuk moderasi konten harus dilakukan secara hati-hati dengan mempertimbangkan faktor-faktor teknis, etis, sosial, dan hukum secara bersamaan. Diperlukan upaya kolaboratif antara perusahaan teknologi, pembuat kebijakan, akademisi, dan masyarakat sipil untuk menciptakan mekanisme moderasi yang adil, transparan, dan adaptif terhadap perubahan dinamika sosial.

2.2.4 Tanggung Jawab Hukum Platform Media Sosial

1. Teori Pertanggungjawaban *(Liability)*

Dalam hukum perdata, konsep tanggung jawab hukum *(liability)* memiliki posisi yang fundamental sebagai dasar untuk menentukan sejauh mana suatu pihak wajib menanggung akibat dari tindakan maupun kelalaian yang menimbulkan kerugian bagi orang lain.[40] Penerapan prinsip ini menjadi sangat relevan dalam era *digital*, khususnya bagi penyelenggara *platform* media sosial seperti Instagram, yang dalam operasionalnya menggunakan teknologi kecerdasan buatan (AI) untuk melakukan moderasi konten. Fungsi utama dari penggunaan AI adalah untuk menyaring, mendeteksi, dan mencegah penyebaran konten yang mengandung muatan terlarang, seperti ujaran kebencian, pornografi, hoaks, maupun bentuk pelanggaran hukum lainnya.[41]

Namun, efektivitas AI dalam praktiknya tidak selalu berjalan sesuai dengan tujuan awal. Kegagalan sistem penyaringan konten dapat terjadi dalam berbagai bentuk, seperti kesalahan klasifikasi *(misclassification)*, terjadinya *false positive* ketika konten yang sah justru dihapus, maupun *false negative* ketika konten yang berbahaya lolos dari penyaringan. Kondisi ini berimplikasi langsung pada aspek hukum karena menimbulkan kerugian baik secara materiil maupun immateriil bagi pengguna. Misalnya, kerugian ekonomi dapat timbul jika konten bisnis yang legal dihapus secara keliru, sedangkan kerugian

[40] T. N. Narwadan, S. Suyani, dan E. F. Thalib, *Buku Ajar Hukum Perdata* (Jakarta: PT. Green Pustaka Indonesia, 2025), hlm. 3.
[41] E. A. Budiman dan M. SH, *Literasi Hukum Digital di Tingkat Masyarakat* (Jakarta: Transformasi Hukum, 2025), hlm. 118.

immateriil dapat muncul dalam bentuk trauma psikologis atau pencemaran nama baik akibat beredarnya konten negatif yang seharusnya dapat dicegah..

Secara umum, dalam hukum perdata dikenal dua bentuk utama teori pertanggungjawaban, yakni pertanggungjawaban berdasarkan kesalahan *(fault liability)* dan pertanggungjawaban tanpa kesalahan *(strict liability).* Pertanggungjawaban berdasarkan kesalahan menekankan adanya unsur kelalaian, kesengajaan, atau kehendak buruk dari pihak pelaku yang menimbulkan kerugian bagi pihak lain.[42] Dalam konteks *platform* media sosial, termasuk Instagram, konsep ini berarti bahwa penyelenggara dapat dimintai pertanggungjawaban hukum apabila terbukti lalai dalam memastikan sistem kecerdasan buatan (AI) yang digunakan untuk penyaringan konten bekerja secara akurat, adil, dan tidak bias.

Kegagalan dalam menjaga akurasi sistem moderasi konten dapat dinilai sebagai bentuk kelalaian apabila *platform* tidak melakukan pengawasan, evaluasi, atau perbaikan secara berkala terhadap algoritma yang digunakan. Misalnya, jika AI secara konsisten gagal mendeteksi konten berbahaya seperti ujaran kebencian, pornografi, atau hoaks, sementara *platform* tidak mengambil langkah yang memadai untuk memperbaikinya, maka dapat dianggap terjadi pelanggaran terhadap kewajiban hukum yang melekat pada penyelenggara sistem elektronik. Dengan demikian, unsur kesalahan dalam *fault liability* dapat terpenuhi melalui bukti adanya kelalaian tersebut.

Sebaliknya, pada teori pertanggungjawaban tanpa kesalahan *(strict liability),* tanggung jawab hukum tetap melekat pada pihak pelaku meskipun tidak terbukti adanya unsur kesalahan atau kelalaian. Prinsip ini biasanya diterapkan pada aktivitas berisiko tinggi, di mana potensi kerugian yang ditimbulkan sangat besar meskipun telah dilakukan upaya pencegahan. Dalam konteks AI di Instagram, prinsip strict liability menjadi relevan mengingat platform

[42] L. O. Hasan dan M. SH, *Perbuatan Melawan Hukum, Wanprestasi, Ganti Rugi Materiil dan Immateriil dalam Kasus-Kasus Perdata* (Yogyakarta: Jejak Pustaka, 2025), hlm. 6.

memperoleh keuntungan ekonomi dari penggunaan teknologi tersebut, sekaligus menanggung risiko sosial dan hukum yang ditimbulkan oleh kegagalan sistem moderasi konten.

Dengan adanya dua teori ini, dapat disimpulkan bahwa tanggung jawab hukum *platform* seperti Instagram bersifat berlapis. Pada satu sisi, kegagalan AI dapat menimbulkan tanggung jawab karena kelalaian *(fault liability),* sementara di sisi lain, risiko inheren dari penggunaan AI juga dapat menimbulkan tanggung jawab tanpa kesalahan *(strict liability).*

Beberapa ahli seperti Wahyudi (pengamat/regulator/akademisi) juga mendorong penerapan (*strict liability)* untuk *platform digital*, terutama ketika dampak kerugian yang ditimbulkan sangat besar dan meluas. Pendekatan ini digunakan untuk mempertegas bahwa tanggung jawab platform bersifat mutlak atas kerugian yang ditimbulkan oleh sistem otomatisnya, terlepas dari ada atau tidaknya kesalahan langsung yang dilakukan oleh operator. Pendekatan ini relevan dalam konteks penyaringan konten berbasis AI, di mana kegagalan sering kali tidak dapat dikaitkan dengan kelalaian manusia secara langsung, melainkan akibat dari ketidaksempurnaan teknologi itu sendiri.

Penerapan sistem penyaringan berbasis AI tidak serta-merta membebaskan *platform* dari kewajiban hukum. Justru, penggunaan AI menciptakan tantangan baru terkait kontrol, pengawasan, dan evaluasi algoritma yang digunakan. Jika platform tidak memastikan bahwa teknologi yang mereka gunakan berfungsi dengan baik, maka mereka dapat dianggap telah lalai secara hukum. Seiring meningkatnya ketergantungan pada sistem otomatis, teori pertanggungjawaban perlu berkembang untuk memasukkan aspek teknologi sebagai bagian dari kewajiban kehati-hatian.[43]

2. Analisis Perbuatan Melawan Hukum (PMH) bila penyaringan konten Gagal

[43] Robert Gorwa, Reuben Binns, dan Christian Katzenbach, "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance," *Big Data & Society* 7, no. 1 (2020): 8, https://doi.org/10.1177/2053951719897945.

Dalam Kitab Undang-Undang Hukum Perdata Indonesia, Pasal 1365 menyebutkan bahwa *"Setiap perbuatan yang melanggar hukum dan membawa kerugian kepada orang lain, mewajibkan orang yang karena kesalahannya* menerbitkan *kerugian itu, mengganti kerugian tersebut."* Ketentuan ini menjadi dasar utama untuk menilai pertanggungjawaban perdata melalui skema perbuatan melawan hukum (PMH). Kegagalan dalam penyaringan konten oleh sistem kecerdasan buatan juga dapat dikaji melalui pendekatan perbuatan melawan hukum (PMH) sebagaimana dimaksud dalam Pasal 1365 KUHPerdata.[44] Pasal tersebut menyatakan bahwa setiap perbuatan yang melanggar hukum dan menimbulkan kerugian bagi orang lain mewajibkan pelakunya untuk memberikan ganti rugi. Untuk dapat dikatakan sebagai PMH, terdapat empat unsur penting: adanya perbuatan melanggar hukum, kesalahan atau kelalaian, kerugian, dan hubungan kausalitas antara perbuatan dan kerugian.[45]

Dalam kasus penyaringan konten yang gagal, unsur perbuatan melawan hukum dapat muncul ketika platform membiarkan konten berbahaya seperti ujaran kebencian, pornografi anak, atau fitnah tersebar luas tanpa tindakan yang memadai. Unsur kesalahan atau kelalaian dapat dibuktikan melalui ketidaksiapan sistem penyaringan, kurangnya pengawasan internal terhadap kinerja AI, atau tidak adanya mekanisme banding yang layak bagi pengguna yang kontennya salah dihapus.

Kerugian yang ditimbulkan akibat kegagalan moderasi konten oleh kecerdasan buatan (AI) pada platform media sosial dapat bersifat materiil maupun immateriil. Kerugian materiil misalnya dialami oleh pengguna yang kontennya secara keliru dihapus oleh sistem AI *(false positive)*, sehingga kehilangan potensi pendapatan dari aktivitas komersial berbasis media sosial, atau mengalami penurunan jumlah pengikut yang berdampak pada nilai ekonomis akun mereka.

---

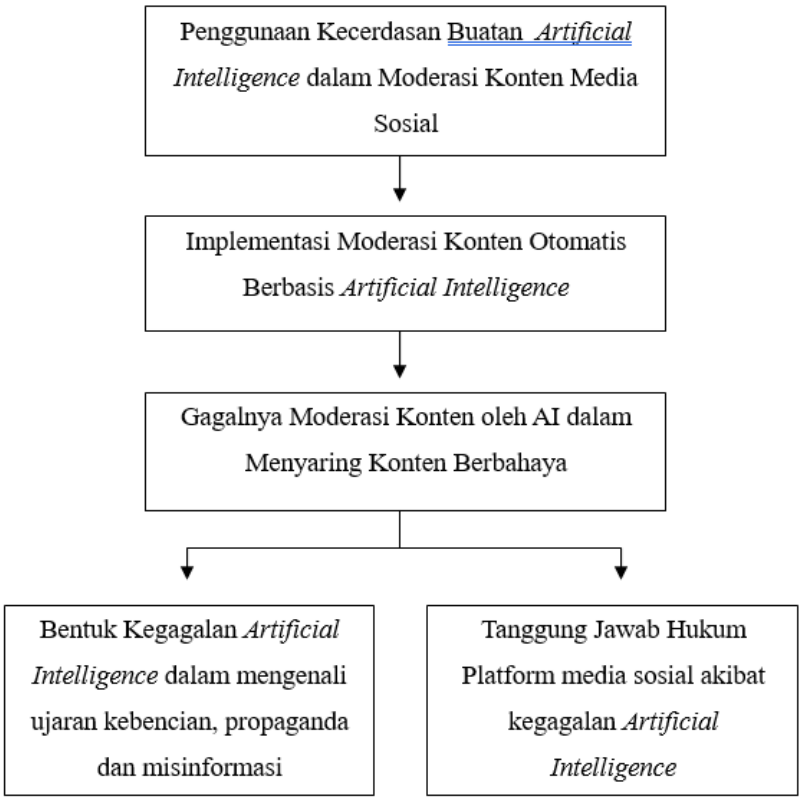
[44] Gunawan Widjaja dan Kartini Muljadi, *Perikatan yang Lahir dari Undang-Undang* (Jakarta: Buku Dosen, 2021), hlm. 81.
[45] A. Budiman, *Perbuatan Melawan Hukum* (Jakarta: Hukum Perdata, 2024), hlm. 73–74.

Sebaliknya, kegagalan penyaringan konten berbahaya *(false negative)*, seperti ujaran kebencian, kekerasan, atau disinformasi, dapat menimbulkan dampak yang lebih luas terhadap masyarakat. Misalnya, penyebaran ujaran kebencian yang tidak tersaring berpotensi memicu polarisasi politik, menimbulkan ketegangan sosial, atau bahkan berujung pada tindak kekerasan di dunia nyata. Dampak ini menunjukkan bahwa kegagalan moderasi AI tidak hanya berimplikasi pada individu, tetapi juga dapat mengganggu ketertiban umum dan stabilitas sosial.

Unsur kausalitas atau hubungan sebab-akibat dalam hukum perdata juga dapat dipenuhi apabila terdapat bukti langsung antara kegagalan sistem AI dengan kerugian yang dialami pengguna atau masyarakat. Contoh konkret adalah ketika AI gagal mendeteksi dan memblokir konten bermuatan kekerasan, kemudian pengguna lain menjadi korban akibat konsumsi atau paparan konten tersebut.

Dengan demikian, platform media sosial seperti Instagram dapat dimintai tanggung jawab perdata berdasarkan prinsip Perbuatan Melawan Hukum (PMH) sebagaimana diatur dalam Pasal 1365 KUHPerdata. Hal ini memberikan dasar bagi pengguna untuk menuntut ganti rugi baik atas kerugian materiil maupun immateriil. Penerapan prinsip ini sekaligus menegaskan bahwa tanggung jawab hukum platform tidak dapat diabaikan hanya dengan alasan kesalahan bersumber pada algoritma, karena pada hakikatnya penyedia *platform* tetap memiliki kewajiban hukum untuk menjamin keamanan layanan yang mereka sediakan.

2.3 Kerangka Pikir



Gambar 1. Kerangka Pikir

Keterngan:

Seiring dengan meningkatnya penggunaan media sosial dalam kehidupan masyarakat, jumlah konten yang bersifat sensitif, provokatif, dan berpotensi merugikan pihak tertentu juga semakin bertambah. Untuk menghadapi permasalahan ini, platform media sosial seperti Instagram mengadopsi teknologi kecerdasan buatan *(Artificial Intelligence)* sebagai sarana utama dalam melakukan moderasi konten. AI diharapkan mampu menyaring dan mengidentifikasi konten yang melanggar ketentuan komunitas secara cepat dan efisien, sehingga lingkungan digital dapat tetap aman serta kondusif bagi para penggunanya.

Namun, meskipun AI menawarkan efisiensi dan kecepatan, penerapannya dalam moderasi konten tidak dapat dipisahkan dari berbagai keterbatasan. Kecerdasan buatan bekerja berdasarkan data pelatihan dan algoritma tertentu, yang sering kali

belum sepenuhnya mampu memahami nuansa bahasa, konteks sosial, serta perbedaan budaya. Akibatnya, sistem moderasi AI kerap menghadapi tantangan berupa kesalahan dalam klasifikasi konten *(misclassification),* baik berupa kelolosan konten yang seharusnya diblokir, maupun penghapusan konten yang sebenarnya sah. Situasi ini menimbulkan implikasi serius, bukan hanya bagi pengalaman pengguna, tetapi juga bagi legitimasi platform dalam menjalankan kewajibannya sebagai penyelenggara sistem elektronik.

Dalam perspektif hukum perdata, permasalahan tersebut menjadi relevan karena menyangkut tanggung jawab hukum platform media sosial atas kegagalan sistem yang mereka gunakan. Tanggung jawab ini tidak hanya dipandang sebagai kewajiban moral, tetapi juga memiliki dimensi yuridis yang menuntut adanya kepastian hukum bagi para pengguna. Kegagalan AI dalam menyaring konten sensitif menimbulkan ruang bagi pengujian terhadap prinsip pertanggungjawaban hukum, baik yang berbasis kesalahan *(fault liability)* maupun yang bersifat tanpa kesalahan *(strict liability).*

Dengan demikian, penelitian ini memposisikan diri untuk menganalisis secara yuridis kegagalan kecerdasan buatan dalam moderasi konten pada Instagram. Fokus utamanya adalah menguraikan sejauh mana tanggung jawab hukum dapat dibebankan kepada platform ketika AI yang digunakan tidak berfungsi secara optimal. Melalui pendekatan normatif yang mengacu pada doktrin, teori hukum, dan kerangka peraturan yang berlaku, penelitian ini berupaya memperluas pemahaman mengenai hubungan antara teknologi kecerdasan buatan, moderasi konten, dan prinsip pertanggungjawaban hukum di era digital.

III. METODE PENELITIAN

Penelitian hukum pada dasarnya merupakan suatu proses berpikir dan bertindak yang dilakukan secara logis, terencana, serta sistematis untuk mengkaji suatu peristiwa hukum, fenomena yuridis, maupun fakta-fakta empiris yang berkembang di masyarakat.[46] Tujuan dari kegiatan penelitian ini adalah untuk memberikan analisis yang mendalam terhadap suatu masalah hukum sehingga dapat ditemukan solusi maupun pemahaman baru yang relevan dengan isu yang diteliti. Oleh karena itu, pemilihan metode penelitian yang tepat memiliki peran penting agar pembahasan dapat terarah sesuai dengan rumusan masalah dan tujuan penelitian. Metode penelitian hukum sendiri merupakan bagian dari ilmu hukum yang berfokus pada tata cara melakukan riset secara benar, terukur, dan sistematis. Proses ini meliputi penentuan pendekatan terhadap masalah hukum, pemilihan teknik pengumpulan data yang sesuai, serta langkah-langkah analisis dan pengolahan data yang digunakan dalam penelitian.[47]

3.1 Jenis Penelitian

Jenis Penelitian dalam penelitian ini adalah penelitian hukum normatif. Penelitian hukum normatif merupakan metode penelitian yang memposisikan hukum sebagai norma atau kaidah yang berlaku dalam kehidupan bermasyarakat dan berfungsi sebagai pedoman perilaku bagi setiap subjek hukum. Fokus utama penelitian hukum normatif terletak pada pengkajian konsep hukum, asas-asas hukum, serta norma hukum yang mengatur suatu permasalahan tertentu. Berdasarkan pandangan doktrinal yang berkembang, penelitian hukum normatif dapat dipahami sebagai

---

[46] Mahmud Marzuki, *Penelitian Hukum* (Jakarta: Kencana, 2023), hlm. 136.
[47] Abdulkadir Muhammad, *Hukum dan Penelitian Hukum* (Bandung: PT. Citra Aditya Bakti, 2004), hlm. 52.

metodologi penelitian hukum yang bertumpu pada analisis terhadap peraturan perundang-undangan yang berlaku dan memiliki keterkaitan dengan isu hukum yang menjadi objek kajian.[48]

Pemilihan jenis penelitian hukum normatif dalam penelitian ini didasarkan pada tujuan untuk menelaah dan menganalisis penerapan ketentuan hukum normatif, baik dalam bentuk kodifikasi maupun peraturan perundang-undangan, secara faktual (*in action*) terhadap peristiwa hukum tertentu, khususnya yang berkaitan dengan proses penyaringan konten melalui pemanfaatan kecerdasan buatan pada platform media sosial.

3.2 Tipe Penelitian

Tipe penelitian yang digunakan dalam penelitian ini adalah tipe penelitian deskriptif, yaitu penelitian bersifat pemaparan dan bertujuan untuk memperoleh gambaran (deskripsi) lengkap tentang keadaan hukum yang berlaku di tempat tertentu dan pada saat tertentu yang terjadi dalam masyarakat.[49] Untuk itu, pada tujuan penelitian ini memperoleh pemaparam lengkap, rinci dan sistematis tentang Tanggung Jawab Platform Media Sosial Dalam Penyaringan Konten Sensitif Pada Penggunaan Kecerdasan Buatan *Artificial Intelligence* Yang Mengalami Kegagalan.

3.3 Pendekatan Masalah

Pendekatan masalah merupakan proses pemecahan atau penyelesaian masalah melalui tahap-tahap yang telah ditentukan sehingga mencapai tujuan penelitian. Dalam penelitian ini, penulis menggunakan pendekatan perundang-undangan. Pendekatan Perundang-undangan *(statute approach)* adalah pendekatan penelitian yang dilakukan dengan melakukan telaah terhadap semua undang-undang dan regulasi yang bersangkut paut dengan isu hukum yang sedang ditangani oleh

[48] Hari Sutra Disemadi, "Lensa Penelitian Hukum: Esai Deskriptif tentang Metodologi Penelitian Hukum," *Journal of Judicial Review (JJR)* 24, no. 2 (2022).
[49] Abdulkadir Muhammad, *Op.Cit.*, hlm. 53.

peneliti.[50] Pendekatan perundang-undangan *(statute approach)* akan dilihat hukum sebagai suatu sistem yang tertutup yang mempunyai sifat sebagai berikut:

1) *Comprehensive* artinya norma-norma hukum yang ada didalamnya terkait antara yang satu dengan yang lainnya secara logis;
2) *All-iclusive* bahwa kumpulan norma hukum tersebut cukup mampu menampung permasalahan hukum yang ada sehingga tidak akan ada kekurangan hukum;
3) *Sistematic*, bahwa di samping bertautan antara satu dengan yang lain, norma-norma hukum tersebut juga tersusun secara sistematis.

Melalui pendekatan ini, penulis mengumpulkan informasi serta menganalisis berkaitan dengan permasalahan yang akan dibahas yaitu *Artificial Intelligence* yang Mengalam Kegagalan di media sosial.

3.4 Data dan Sumber Data

Berdasarkan jenis penelitian dan pendekatan masalah yang digunakan, maka data yang digunakan untuk memecahkan permasalahan dalam penelitian ini adalah data sekunder. Data sekunder adalah data yang sejatinya sudah tersedia dan terkompilasi sehingga peneliti dipermudahkan dalam memperoleh data karena ia tinggal mencari dan mengumpulkan data ini dari sumber yang menyediakannya, serta tidak perlu lagi mencari data tersebut dari sumber aslinya.[51] Data ini diperoleh dari studi kepustakaan, dengan cara mengumpulkan dari berbagai sumber bacaan yang berhubungan dengan masalah yang diteliti.[52] Data sekunder terdiri dari:

---

[50] Muhammad Teguh Arifiawan, *Implementasi Lisensi Creative Commons pada Ciptaan Lagu yang Diunggah dalam Platform Musik Digital* (Skripsi, Universitas Lampung, Bandar Lampung, 2022), hlm. 37.
[51] David Tan, "Metode Penelitian Hukum: Mengupas dan Mengulas Metodologi dalam Menyelenggarakan Penelitian Hukum," *Nusantara: Jurnal Ilmu Pengetahuan Sosial* 8, no. 8 (2021): 2472.
[52] I Ketut Dharma Putra Yoga, *Implementasi Konsep Creating Shared Value (CSV) sebagai Program Corporate Social Responsibility (CSR) dalam Upaya Peningkatan Kesejahteraan Stakeholder (Studi pada PT Nestle Indonesia Panjang Factory)* (Skripsi, Universitas Lampung, Bandar Lampung, 2018), hlm. 44.

a. Bahan hukum primer yaitu, bahan hukum yang bersifat otoritatif atau mengikat seperti peraturan perundang-undangan dan regulasi yang dibentuk secara formal oleh lembaga yang berwenang di mana berhubungan dengan penelitian ini, antara lain:
    1. Kitab Undang-Undang Hukum Perdata.
    2. Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik.
    3. Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi Dan Transaksi Elektronik.
    4. Surat Edaran Menteri Komunikasi dan Informatika Republik Indonesia Nomor 9 Tahun 2023 Tentang Etika Kecerdasan Artifisial.
    5. Panduan Kode Etik Kecerdasan Buatan (Artificial Intelligence) yang Bertanggung Jawab dan Terpercaya di Industri Teknologi Finansial.
    6. Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik
    7. Peraturan Menteri Komunikasi dan Informatika Nomor 5 Tahun 2020 tentang Penyelenggara Sistem Elektronik Lingkup Privat
b. Bahan Hukum sekunder yaitu bahan hukum yang eksistensinya berfungsi untuk menyediakan elaborasi lebih lanjut terhadap bahan hukum primer. Yang mana berupa literatur-literatur mengenai penelitian ini, meliputi buku-buku ilmu hukum, hasil karya dari kalangan hukum dan lainnya yang berkaitan dengan penelitian ini.[53]
c. Bahan Hukum Tersier yaitu, bahan hukum yang bersifat pelengkap yang menyediakan petunjuk ataupun elaborasi lebih lanjut terhadap bahan hukum primer dan bahan hukum sekunder, yakni berupa kamus, ensiklopedia, dan artikel pada majalah, surat kabar atau internet.

3.5 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah Studi

[53] Sri Mamuji, *Teknik Menyusun Karya Tulis Ilmiah* (Jakarta: UI Press, 2006), hlm 12.

Kepustakaan, yang mana merupakan teknik untuk memperoleh data sekunder melalui serangkaian kegiatan yaitu membaca, menelaah, dan mengutip peraturan perundang-undangan, buku-buku, dan literatur yang berkaitan dengan permasalahan yang diteliti.

3.6 Metode Pengolahan Data

Pengolahan data dilakukan dengan beberapa tahapan yakni:

1. Pemeriksaan Data

   Pada tahap ini, diperoleh melalui studi kepustakaan untuk memastikan data yang terkumpul sudah lengkap, relevan, jelas, dan tidak berlebihan, dengan permasalahan yang dibahas dan kejelasan jawaban terhadap permasalahan yang dibahas.

2. Klasifikasi Data

   Penempatan, pengelompokkan atau penggolongan data menurut jenis dan sumbernya, dengan tujuan untuk menyajikan data secara sempurna, yang memudahkan rekontruksi serta analisis data

3. Sistematika Data

   Data yang telah diperiksa dan diklasifikasikan kemudian disusun secara sistematis, untuk mempermudah dalam membahas dan menganalisis permasalahannya.

3.7 Analisis Data

Analisis data dalam penelitian ini menggunakan analisis kualitatif. Analisis ini dilakukan dengan cara menafsirkan atau menginterpretasikan data dalam bentuk kalimat yang tersusun sistematis sehingga memudahkan dalam pemahaman hasil dan mendapat gambaran yang jelas sesuai dengan permasalahan untuk kemudian ditarik berbagai kesimpulan.

V. PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang dilakukan oleh penulis terhadap Platform media sosial instagram mengenai pertanggungjawaban platform instagram yang mengalami kegagalan pada proses pemyaringan konten menggunakan kecerdasan buatan (*artificial intelligence),* maka dapat disimpulkan beberapa hal sebagai berikut:

1. Tanggung jawab hukum platform media sosial Instagram atas kegagalan penyaringan konten oleh kecerdasan buatan (AI) terletak pada kewajiban sebagai Penyelenggara Sistem Elektronik (PSE) sesuai ketentuan peraturan perundang-undangan di Indonesia, khususnya UU ITE dan PP No. 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik. Instagram, meskipun menggunakan teknologi AI dalam moderasi konten, tetap memikul tanggung jawab hukum penuh untuk memastikan bahwa konten sensitif atau ilegal tidak tersebar di ruang digital. Kegagalan AI dalam melakukan penyaringan tidak dapat dijadikan alasan pembebasan tanggung jawab, karena tanggung jawab hukum tetap melekat pada penyelenggara *platform*.
2. Sanksi hukum bagi *platform* media sosial Instagram apabila AI yang digunakan gagal dalam penyaringan konten sensitif meliputi sanksi administratif dan perdata, serta berpotensi mengarah pada sanksi pidana. Berdasarkan Permenkominfo No. 5 Tahun 2020, Instagram dapat dikenai teguran tertulis, denda administratif, pembatasan fitur, hingga pemutusan akses apabila tidak segera menindaklanjuti laporan konten terlarang. Di samping itu, pengguna yang dirugikan akibat kegagalan moderasi berhak menuntut ganti rugi melalui mekanisme perdata berdasarkan Pasal 1365 KUH Perdata. Bahkan dalam kasus tertentu, jika konten yang gagal difilter mengandung muatan pidana, maka

tanggung jawab pidana dapat ditujukan kepada *platform*. Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE). yang telah diperbarui dengan Undang-Undang Nomor 19 Tahun 2016. Sanksi meliputi pidana penjara hingga enam tahun dan denda maksimal Rp1 miliar bagi pelanggaran Pasal 27 sampai pasal 29 UU ITE. Hal ini menegaskan bahwa tanggung jawab hukum Instagram bersifat komprehensif, mencakup aspek administratif, perdata, dan pidana, sehingga mampu memberikan perlindungan hukum bagi pengguna.

5.2. Saran

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, penulis memberikan beberapa rekomendasi sebagai berikut:

1. Bagi *Platform* Media Sosial Instagram, perlu dilakukan penguatan sistem kecerdasan buatan (AI) yang digunakan dalam moderasi konten sensitif dengan meningkatkan akurasi algoritma dan transparansi mekanisme penyaringan. Instagram juga diharapkan menyeimbangkan peran AI dengan pengawasan manusia *(human oversight)* agar kesalahan dalam penyaringan konten dapat diminimalisasi, sehingga hak-hak pengguna tetap terlindungi.
2. Bagi Pemerintah dan Regulator, khususnya Kementerian Komunikasi dan Informatika (Kominfo), disarankan untuk memperkuat regulasi terkait penggunaan AI dalam moderasi konten melalui kebijakan yang lebih tegas dan adaptif terhadap perkembangan teknologi. Selain itu, pemerintah perlu memperjelas standar tanggung jawab hukum bagi platform digital, termasuk mekanisme pertanggungjawaban apabila terjadi kegagalan penyaringan konten.
3. Bagi Pengguna Media Sosial, penting untuk lebih kritis dan bijak dalam memanfaatkan layanan *digital*, serta berperan aktif dalam melaporkan konten sensitif atau ilegal yang lolos dari sistem moderasi. Partisipasi aktif pengguna akan membantu terciptanya ruang digital yang aman, sehat, dan bermanfaat.

DAFTAR PUSTKA

BUKU

Andari, T. *Hukum Informasi dan Transaksi Elektronik di Indonesia*. Jakarta: Sinar Grafika, 2022.

Budhijanto, Danrivanto. *Hukum Keamanan Siber*. Bandung: Logoz Publishing, 2024.

Budiman, A. *Perbuatan Melawan Hukum*. Jakarta: Hukum Perdata, 2024.

Budiman, E. A., dan M. SH. *Literasi Hukum Digital di Tingkat Masyarakat*. Jakarta: Transformasi Hukum, 2025.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press, 2018.

Ginting, Grenaldo, dan M. P. N. Simamora. *Hukum Teknologi Informasi dan Komunikasi*. Medan: Fakultas Hukum Universitas Muhammadiyah Sumatera Utara, 2020.

Hasan, L. O., dan M. SH. *Perbuatan Melawan Hukum, Wanprestasi, Ganti Rugi Materiil dan Immateriil dalam Kasus-Kasus Perdata*. Yogyakarta: Jejak Pustaka, 2025.

Indra, S. *Etika dan Hukum dalam Era Digital*. Jakarta: Citra Aditya Bakti, 2023.

Mamuji, Sri. *Teknik Menyusun Karya Tulis Ilmiah*. Jakarta: UI Press, 2006.

Marzuki, Mahmud. *Penelitian Hukum*. Jakarta: Kencana, 2023.

Mauludi, S. *Seri Cerdas Hukum: Awas Hoax! Cerdas Menghadapi Pencemaran Nama Baik, Ujaran Kebencian & Hoax*. Jakarta: Elex Media Komputindo, 2019.

Muhammad, Abdulkadir. *Hukum dan Penelitian Hukum*. Bandung: PT. Citra Aditya Bakti, 2004.

Narwadan, T. N., S. Suyani, dan E. F. Thalib. *Buku Ajar Hukum Perdata*. Jakarta: PT. Green Pustaka Indonesia, 2025.

Purnama, P. A. W., C. Fadhilah, C. A. Fadhilla, B. Wardana, J. Sumah, R. M. Thaniket, dan N. Pohan. *Artificial Intelligence*. Jakarta: Serasi Media Teknologi, 2025.

Situmeang, S. M. T. *Cyber Law*. Medan: Fakultas Hukum Universitas Muhammadiyah Sumatera Utara, 2020.

Syahril, Muh. Akbar Fhad. *Hukum Informasi dan Transaksi Elektronik*. CV. Eureka Media Aksara, 2023.

Veglis, A. *Moderation Techniques for Social Media Content*. Cham: Springer, 2014.

Widjaja, Gunawan, dan Kartini Muljadi. *Perikatan yang Lahir dari Undang-Undang*. Jakarta: Buku Dosen, 2021.

Yusnaini, Y., Muhaimin, M., Firmansyah, F., Setiwan, Y. L., Bakhtiar, R., Aisyah, A., et al. *Artificial Intelligence dalam Perkembangan Teknologi Komunikasi*. Jakarta: CV. Gita Lentera, 2024.

JURNAL

Azizah, Desi, Aji Wibawa, dan Laksono Budiarto. "Hakikat Epistemologi Artificial Intelligence." *Jurnal Inovasi Teknologi dan Edukasi Teknik* 1, no. 8 (2021): 592–98. https://doi.org/10.17977/um068v1i82021p592-598.

Daud, Mahyuddin, dan Ida Madieha Abd Ghani Azmi. "Intermediary's Liability: Towards a Sustainable Artificial Intelligence-Based Content Moderation in Malaysia." *IIUM Law Journal* 31, no. 2 (2023): 155–78. https://doi.org/10.31436/iiumlj.v31i2.823.

Dimitrova, Ralitza. "Artificial Intelligence in Content Moderation–Legal Challenges and EU Legal Framework." In *2022 10th International Scientific Conference on Computer Science (COMSCI)*, 1–6. IEEE, 2022.

Fajrina, R. M. "Pencegahan Tindak Pidana Pornografi Online Melalui Penerapan Etika Digital di Media Sosial." *Jurnal Dinamika Sosial dan Sains* 2, no. 5 (2025): 738–44.

Farwati, Maryani, et al. "Analisa Pengaruh Teknologi Artificial Intelligence (AI) dalam Kehidupan Sehari-Hari [Analyze the Influence of Artificial Intelligence (AI) Technology in Daily Life]." *Jurnal Sistem Informatika dan Manajemen* 11, no. 1 (2023): 41–42.

Febryani, E. "The Impact of Content Moderation Policy on the Spread of Fake News on Social Media in Indonesia." *The Easta Journal Law and Human Rights* 3, no. 3 (2025): 176–83.

Gorwa, Robert, Reuben Binns, dan Christian Katzenbach. “Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance.” *Big Data & Society* 7, no. 1 (2020): 1–15. https://doi.org/10.1177/2053951719897945.

Hadikristanto, Wahyu. “Implementasi Content Moderation dalam Social Media Instagram untuk Deteksi Cyberbullying dengan Machine Learning Berbasis Cloud.” *Indonesian Journal of Business Intelligence (IJUBI)* 5, no. 2 (2022): 122–126.

Hanisch, M., Curtis M. Goldsby, N. Fabian, dan J. Oehmichen. “Digital Governance: A Conceptual Framework and Research Agenda.” *Journal of Business Research* (2023). https://doi.org/10.1016/j.jbusres.2023.113777.

Harsya, R. M. K. “Tinjauan Yuridis terhadap Tanggung Jawab Platform Digital atas Konten Ilegal Menurut Hukum Indonesia.” *Sanskara Hukum dan HAM* 4, no. 1 (2025): 276–286.

Hossain, Elias, R. Rana, N. Higgins, J. Soar, P. Barua, Anthony R. Pisani, dan K. Turner. “Natural Language Processing in Electronic Health Records in Relation to Healthcare Decision-Making: A Systematic Review.” *Computers in Biology and Medicine* 155 (2023): 106649. https://doi.org/10.1016/j.compbiomed.2023.106649.

Judijanto, L., A. S. Utama, dan H. Setiyawan. “Implementation of Ethical Artificial Intelligence Law to Prevent the Use of AI in Spreading False Information (Deepfake) in Indonesia.” *The Easta Journal Law and Human Rights* 3, no. 2 (2025): 101–109.

Oliva, Thiago Dias. “Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression.” *Human Rights Law Review* 20, no. 4 (December 9, 2020): 607–40. https://doi.org/10.1093/hrlr/ngaa032.

Prasetyo, Sindy. “Pelanggaran Hak Asasi Manusia di Indonesia.” *Indigenous Knowledge* 2, no. 1 (2023): 51–57.

Puannandini, D. A., R. Fabian, R. A. P. Firdaus, M. Z. Mustopa, dan I. Herdiyana. “Liabilitas Produk AI dalam Sistem Hukum Indonesia: Implikasi bagi Pengembang, Pengguna, dan Penyedia Layanan.” *Iuris Studia: Jurnal Kajian Hukum* 6, no. 1 (2025): 24–33.

Ribeiro, Jorge, et al. “Robotic Process Automation and Artificial Intelligence in Industry 4.0 - A Literature Review.” *Procedia Computer Science* 181 (2021): 51–58. https://doi.org/10.1016/j.procs.2021.01.104.

Rodliyah, R., A. Suryani, dan L. Husni. “Konsep Pertanggungjawaban Pidana Korporasi (Corporate Crime) dalam Sistem Hukum Pidana Indonesia.” *Jurnal Kompilasi Hukum* 5, no. 1 (2020): 191–206.

Safina, Raisa, Khalda Alifia Azzahra, dan Ananda Fersa Dharmawan. "Kajian Yuridis Penggunaan Kecerdasan Artifisial pada Pembuatan dan Penyebaran Konten Pornografi di Media Sosial dalam Hukum Positif Indonesia." *Mandub: Jurnal Politik, Sosial, Hukum dan Humaniora* 2, no. 1 (2023): 302–13. https://doi.org/10.59059/mandub.v2i1.918.

Salsabilla, Kharisma Agustya Zahra, et al. "Pengaruh Penggunaan Kecerdasan Buatan terhadap Mahasiswa di Perguruan Tinggi." *Prosiding Seminar Nasional Teknologi dan Sistem Informasi* 3, no. 1 (2023): 168–75. https://doi.org/10.33005/sitasi.v3i1.371.

Savitri, A. M., dan T. N. Fatihah. "Tinjauan Yuridis Mengenai Perlindungan Data Pribadi dan Pencegahan Kekerasan Seksual terhadap Anak di Bawah Umur dalam Menggunakan Gawai dan Media Sosial di Indonesia." *IBLAM Law Review* 5, no. 2 (2025): 58–68.

Simbolon, Yolanda. "Pertanggungjawaban Perdata Terhadap Artificial Intelligence yang Menimbulkan Kerugian Menurut Hukum di Indonesia." *Veritas et Justitia* 9, no. 1 (2023): 246–73. https://doi.org/10.25123/vej.v9i1.6037.

Siregar, A. G. "Implementasi Asas Ultimum Remedium Terhadap Penerapan Sanksi Pidana dalam Undang-Undang Administratif." *Innovative: Journal of Social Science Research* 3, no. 4 (2023): 10271–10285.

Suryoutomo, M., Solekhan, M., dan S. Murni. "Tanggung Jawab Perdata dalam Kasus Wanprestasi dan Perbuatan Melawan Hukum." *Jurnal Kolaboratif Sains* 8, no. 4 (2025): 2018–2023.

SKRIPSI

Anjani, Quceny Praviyanti. *Chatbot Menggunakan Natural Language Processing (NLP) pada Toko Bunga Online. Tesis, Universitas Nasional, Jakarta Selatan, 2023.*

Arifiawan, Muhammad Teguh. Implementasi Lisensi Creative Commons pada Ciptaan Lagu yang Diunggah dalam Platform Musik Digital. Skripsi, Universitas Lampung, Bandar Lampung, 2022.

Chrismastianto, Imanuel Adhitya Wulanata. "Efektifitas Layanan Keuangan Berbasis Machine Learning Sebagai Komponen Pendukung Kebijakan Makroprudensial Pascapandemi Covid-19." Jurnal Universitas Kristen Immanuel, 2021.

Immanuel, I. G. D. Pertanggungjawaban Platform Digital dalam Mengatasi Konten Ilegal. Disertasi Doktor, Universitas Islam Sultan Agung Semarang, 2024.

Yoga, I Ketut Dharma Putra. Implementasi Konsep Creating Shared Value (CSV) sebagai Program Corporate Social Responsibility (CSR) dalam Upaya

Peningkatan Kesejahteraan Stakeholder (Studi pada PT Nestle Indonesia Panjang Factory). Skripsi, Universitas Lampung, Bandar Lampung, 2018.

PERATURAN PERUNDANG-UNDANGAN

Kitab Undang-Undang Hukum Perdata

Peraturan Menteri Komunikasi dan Informatika Nomor 5 Tahun 2020 tentang Penyelenggara Sistem Elektronik Lingkup Privat. Berita Negara Republik Indonesia Tahun 2020 Nomor 159.

Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik. Lembaran Negara Republik Indonesia Tahun 2019 Nomor 185, Tambahan Lembaran Negara Republik Indonesia Nomor 6400.

Surat Edaran Menteri Komunikasi dan Informatika Republik Indonesia Nomor 9 Tahun 2023 tentang Etika Kecerdasan Artifisial.

Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik. Lembaran Negara Republik Indonesia Tahun 2008 Nomor 58, Tambahan Lembaran Negara Republik Indonesia Nomor 4843.

Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik. Lembaran Negara Republik Indonesia Tahun 2016 Nomor 251, Tambahan Lembaran Negara Republik Indonesia Nomor 5952.

ARTIKEL

Digital 2025: Indonesia. DataReportal. 2025. Diakses 29 Maret 2025. *https://datareportal.com/reports/digital-2025-indonesia.*

Lenny Septiani. "Teknologi AI Dikhawatirkan Menimbulkan Informasi Salah." Katadata, *2023. https://katadata.co.id/digital/teknologi/64e462a7c45a4/teknologi-ai-dikhawatirkan-menimbulkan-informasi-salah?page=2 (diakses 23 Maret 2025, 21:37 WIB).*

Merav Ozair. "Misinformation in the Age of Artificial Intelligence and What It Means for the Markets." Nasdaq, *2023. https://www.nasdaq.com/articles/misinformation-in-the-age-of-artificial-*

intelligence-and-what-it-means-for-the-markets (diakses 23 Maret 2025, 21:35 WIB).

Myers West, Sarah, Meredith Whittaker, dan Kate Crawford. Discriminating Systems: Gender, Race and Power in AI. *New York: AI Now Institute, 2019. https://ainowinstitute.org/discriminatingsystems.pdf*.

Puskomedia. "Menerapkan Strategi Moderasi yang Efektif dalam Platform Komunitas Online." Diakses 29 Maret 2025. https://puskomedia.id/blog/menerapkan-strategi-moderasi-yang-efektif-dalam-platform-komunitas-online/.

Supriyanto, Catur. "Pentingnya Computer Vision pada Situs Jual Beli Online." Kompasiana, *2021. https://www.kompasiana.com/catursupriyanto/602206661730b90ef37dba82/computer-vision-dalam-situs-jual-beli (diakses 28 Maret 2025, 20:43 WIB).*

Universitas Palangka Raya. "Sejarah Artificial (AI) dan Fungsi dalam Kehidupan Sehari-Hari: Pengantar Teknik Informatika Sejarah Artificial Intelligence (AI) dan." No. November (2023).

Vedhitya, Mavellyno. "Predictive Analytics, Memprediksi Masa Depan lewat Pola Data." Marketeers, *2023. https://www.marketeers.com/predictive-analytics-memprediksi-masa-depan-pola-data/* (diakses 28 Maret 2025, 21:15 WIB).