

Tanggung Jawab Platform Digital dalam Mengatasi Hoax yang Dihasilkan Oleh AI: Analisis Kasus Penyalahgunaan Bot dan Algoritma Rekomendasi untuk Manipulasi Opini Publik

Hibrida Delima^{1*}, Hafied Alfiansyah², Muhammad Isa Anshori³, Ridho Maulana Bungas⁴, Annisa Elfina Augustia⁵

¹⁻⁵ Fakultas Teknik dan Ilmu Komputer, Teknik Informatika, Universitas Indraprasta PGRI, Jakarta, Indonesia

Email : ^{1*}hibridadelima0801@gmail.com, ²hafiedalfiansyah894@gmail.com, ³misaanshori21@gmail.com,

⁴bungasridho2@gmail.com, ⁵annisa12elfina@gmail.com

(* : coresponding author)

Abstrak—Era kecerdasan buatan (AI) telah melahirkan paradoks dalam ekosistem informasi digital. Di satu sisi, AI memberdayakan pertukaran pengetahuan, namun di sisi lain, teknologi ini memfasilitasi produksi dan penyebaran disinformasi secara masif. Penelitian ini menganalisis tanggung jawab platform digital dalam menghadapi tantangan hoax yang dihasilkan oleh AI, dengan fokus khusus pada analisis kasus penyalahgunaan bot dan algoritma rekomendasi untuk manipulasi opini publik. Menggunakan pendekatan kualitatif dengan metode studi literatur dan analisis isi terhadap data sekunder, penelitian ini mengungkapkan mekanisme simbiosis antara bot AI dan algoritma rekomendasi yang menciptakan siklus amplifikasi hoax yang swaperkuat. Bot berfungsi sebagai amplifier awal yang memanipulasi metrik engagement, sementara algoritma rekomendasi, yang didorong logika komersial untuk memaksimalkan retensi pengguna, mengangkat konten hoax tersebut ke khalayak yang lebih luas. Temuan penelitian menunjukkan bahwa klaim netralitas platform menjadi tidak relevan mengingat kekuatan kuratorial aktif yang mereka jalankan melalui algoritma. Oleh karena itu, penelitian ini berargumen bahwa tanggung jawab platform harus bergeser dari sekadar moderasi konten reaktif menuju tata kelola algoritma yang proaktif. Kesimpulan penelitian merekomendasikan perlunya transparansi algoritma, reformasi desain sistem rekomendasi, deteksi proaktif terhadap jaringan bot, dan kerangka regulasi yang jelas untuk memastikan akuntabilitas platform digital dalam menjaga integritas ruang publik di era AI.

Kata kunci: Tanggung Jawab Platform, Hoax AI, Bot, Algoritma Rekomendasi, Manipulasi Opini Publik, Tata Kelola Digital.

Abstract—The artificial intelligence (AI) era has created a paradox in the digital information ecosystem. On one hand, AI empowers knowledge exchange, yet on the other hand, this technology facilitates the mass production and dissemination of disinformation. This research analyzes the responsibility of digital platforms in facing the challenge of AI-generated hoaxes, with specific focus on case studies of bot misuse and recommendation algorithms for public opinion manipulation. Using a qualitative approach with literature study methods and content analysis of secondary data, this research reveals the symbiotic mechanism between AI bots and recommendation algorithms that creates a self-reinforcing hoax amplification cycle. Bots function as initial amplifiers that manipulate engagement metrics, while recommendation algorithms, driven by commercial logic to maximize user retention, promote these hoax contents to broader audiences. The research findings indicate that platforms' neutrality claims become irrelevant given the active curatorial power they exercise through algorithms. Therefore, this research argues that platform responsibility must shift from merely reactive content moderation toward proactive algorithmic governance. The research conclusion recommends the need for algorithm transparency, reform of recommendation system design, proactive detection of bot networks, and clear regulatory frameworks to ensure digital platform accountability in maintaining public space integrity in the AI era.

Keywords: Platform Responsibility, AI Hoaxes, Bots, Recommendation Algorithm, Public Opinion Manipulation, Digital Governance.

1. PENDAHULUAN

Era digital, yang ditandai dengan konvergensi teknologi informasi dan kecerdasan buatan (AI), telah merevolusi cara manusia mengakses informasi dan berinteraksi. Platform digital seperti media sosial, mesin pencari, dan situs berita telah menjadi infrastruktur kritikal bagi ruang publik modern,

menjadi arena utama bagi pertukaran ide, deliberasi demokratis, dan pembentukan opini. Namun, di balik manfaat demokratisasi informasinya, tersembunyi sebuah paradoks. Teknologi yang sama yang memberdayakan terutama AI ternyata menjadi pedang bermata dua yang mengancam fondasi ruang publik itu sendiri (Arly et al., 2023).

Kemunculan *Generative AI* (AI Generatif) yang mampu memproduksi teks, gambar, audio, dan video yang hiper-realistic dalam skala massal dan waktu singkat, telah membawa ancaman disinformasi ke level yang baru. Jika dahulu hoax disebarluaskan secara manual atau semi-otomatis, kini AI memungkinkan produksi dan personalisasi hoax dengan kecepatan, skalabilitas, dan daya persuasi yang belum pernah terjadi sebelumnya. Ancaman ini diperparah oleh dua elemen struktural dalam platform digital itu sendiri: pertama, penggunaan bot AI yang canggih untuk menyebarluaskan konten secara otomatis dan menciptakan ilusi konsensus (*astroturfing*); dan kedua, desain algoritma rekomendasi yang secara inherent cenderung memprioritaskan konten yang sensasional, emosional, dan divisif—ciri khas hoax—untuk memaksimalkan *engagement* dan retensi pengguna.

Konvergensi antara kemampuan AI penghasil hoax dan logika operasional platform digital menciptakan sebuah "badai sempurna" untuk manipulasi opini publik. Algoritma rekomendasi, yang dirancang untuk memberi pengguna apa yang mereka "inginkan", secara tidak sengaja menjadi amplifier bagi konten-konten bermasalah yang dihasilkan AI. Bot-bot yang menyamar sebagai pengguna manusia kemudian memanipulasi algoritma ini dengan memberikan *engagement* buatan (like, share, komentar), sehingga mendorong konten hoax tersebut masuk ke *feed* pengguna nyata dalam skala besar. Pada akhirnya, siklus ini tidak hanya mencemari ekosistem informasi tetapi juga menggerus kepercayaan sosial, mempolarisasi masyarakat, dan mengancam integritas proses demokratis seperti pemilu (Chairunnisa et al., 2021).

Dalam konteks inilah, muncul pertanyaan mendasar: di manakah letak tanggung jawab platform digital dalam mengatasi gelombang hoax yang dihasilkan oleh AI ini? Apakah platform dapat terus bersembunyi di balik statusnya sebagai "penyedia layanan netral" (*neutral intermediaries*), ataukah mereka harus memikul tanggung jawab yang lebih besar sebagai *gatekeeper* dan pengelola ruang publik digital yang memiliki kekuatan dan pengaruh signifikan?

Paper ini berargumen bahwa dalam menghadapi ancaman hoax berbasis AI, platform digital tidak dapat lagi bersikap pasif dan hanya mengandalkan solusi teknis yang bersifat reaktif. Mereka harus memikul tanggung jawab prosedural dan substantif yang lebih besar, yang mencakup transparansi algoritma, audit eksternal yang independen, investasi dalam deteksi *AI-generated content* yang lebih canggih, dan peninjauan ulang model bisnis yang mendorong penyebaran konten berbahaya.

Untuk membongkar kompleksitas masalah ini, analisis akan difokuskan pada dua kasus penyalahgunaan utama: (1) penggunaan jaringan bot untuk menyebarluaskan dan mengamplifikasi konten hoax yang dihasilkan AI, dan (2) peran algoritma rekomendasi dalam menciptakan ruang gema (*echo chambers*) dan memfasilitasi viralnya manipulasi opini publik. Dengan menganalisis mekanisme ini, paper ini bertujuan untuk memberikan peta jalan yang jelas mengenai bentuk dan urgensi tanggung jawab platform digital dalam menjaga kesehatan informasi di era AI.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan desain studi kasus untuk menyelidiki secara mendalam fenomena tanggung jawab platform digital dalam menangani hoax yang dihasilkan oleh AI. Pendekatan kualitatif dipandang tepat karena memungkinkan peneliti untuk memahami kompleksitas permasalahan yang tidak hanya bersifat teknis, tetapi juga menyangkut aspek regulasi, etika, dan dinamika sosial-budaya di balik operasi platform digital. Studi ini berfokus pada dua unit analisis utama, yaitu penyalahgunaan bot AI dan peran algoritma rekomendasi, yang menjadi locus dimana masalah hoax dan manipulasi opini publik termanifestasi.

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, yang dikumpulkan melalui teknik studi literatur dan dokumentasi. Sumber data primer penelitian ini terdiri dari laporan investigatif

dari lembaga riset terpercaya seperti Reuters Institute dan Pew Research Center, laporan transparansi dari platform media sosial itu sendiri (misalnya Meta dan X/Twitter), artikel jurnal akademis yang ditelaah sejawat, serta pemberitaan media massa yang kredibel terkait kasus-kasus spesifik penyalahgunaan AI. Dokumen-dokumen hukum dan kebijakan, seperti Rancangan Undang-Undang Uni Eropa tentang AI (*AI Act*) dan undang-undang tata kelola konten digital di berbagai negara, juga dianalisis untuk memahami kerangka tanggung jawab hukum yang sedang berkembang.

Proses analisis data dilakukan dengan menggunakan metode analisis isi kualitatif (*qualitative content analysis*). Data-data sekunder yang telah dikumpulkan kemudian diorganisir, dikategorisasikan, dan ditafsirkan secara sistematis untuk mengidentifikasi pola, tema, dan hubungan antar konsep. Analisis difokuskan pada tiga dimensi utama. Pertama, menganalisis modus operandi dan dampak dari penggunaan bot AI serta algoritma rekomendasi dalam siklus hidup hoax, mulai dari produksi, amplifikasi, hingga konsumsi. Kedua, mengevaluasi respons dan kebijakan yang selama ini diterapkan oleh platform digital, serta mengidentifikasi celah dan kelemahannya. Ketiga, menganalisis wacana dan argumentasi yang berkembang mengenai bentuk dan batas tanggung jawab platform dalam kerangka tata kelola internet yang lebih luas. Melalui triangulasi sumber data, validitas dan reliabilitas temuan penelitian ini diupayakan dapat dipertahankan.

3. HASIL DAN PEMBAHASAN

3.1. Siklus Amplifikasi Hoax Berbasis AI: Konvergensi Bot dan Algoritma

Penelitian ini mengungkap bahwa penyebaran hoax berbasis AI tidak beroperasi dalam ruang hampa, melainkan melalui sebuah siklus amplifikasi yang dimungkinkan oleh konvergensi antara akun bot yang canggih dan logika operasi algoritma rekomendasi. Siklus ini dimulai dari produksi massal konten oleh *Generative AI*. Teknologi ini memungkinkan aktor badai untuk memproduksi ratusan hingga ribuan varian narasi hoax yang berbeda secara tekstual, namun memiliki inti pesan yang sama, dalam waktu yang sangat singkat. Hal ini mengatasi keterbatasan manual dalam kampanye disinformasi tradisional.

Konten yang telah dihasilkan kemudian memasuki tahap amplifikasi buatan. Di sini, jaringan bot AI berperan sebagai "pasukan shock digital" yang memberikan *engagement* awal. Bot-bot ini tidak hanya melakukan retweet atau share secara masif, tetapi juga telah dirancang untuk menghasilkan komentar-komentar yang tampak natural guna menciptakan ilusi diskusi publik yang hidup dan mendukung. Praktik *astroturfing* ini berhasil memanipulasi sinyal-sinyal (*engagement metrics*) yang menjadi masukan bagi algoritma rekomendasi platform. Algoritma, yang pada dasarnya dirancang untuk mempromosikan konten yang populer dan menarik perhatian, kemudian menafsirkan *engagement* buatan ini sebagai indikator kualitas dan relevansi. Akibatnya, konten hoax tersebut diangkat (*boosted*) dan mulai disebarluaskan ke khalayak pengguna nyata yang lebih luas melalui *feed berita*, *recommended videos*, atau *trending topics* (Andika & Soemarno, 2023).

Pada tahap akhir, internalisasi oleh pengguna manusia, algoritma rekomendasi memperkuat efek ruang gema (*echo chambers*). Pengguna yang telah berinteraksi dengan konten sejenis akan terus-menerus disuguhi varian hoax yang sama, yang pada akhirnya menormalisasi wacana palsu tersebut dan mempersulit koreksi fakta. Dengan demikian, terciptalah sebuah siklus yang swaperkuat: bot memicu algoritma, algoritma memperluas jangkauan kepada manusia, dan *engagement* manusia selanjutnya mengonfirmasi kepada algoritma bahwa konten tersebut "berkualitas", sehingga memperkuat penyebarannya lebih lanjut (Juditha, 2018).

3.2. Tanggung Jawab Platform di Tengah Paradigma Netralitas yang Usang

Temuan utama dalam penelitian ini adalah kegagalan paradigma netralitas platform dalam menjawab tantangan ini. Platform sering kali berargumen bahwa mereka hanyalah "*pasif pipe*" atau saluran netral yang tidak bertanggung jawab atas konten yang dibawa oleh penggunanya. Namun, analisis terhadap mekanisme algoritma rekomendasi membantah klaim ini. Algoritma bukanlah entitas

yang netral; ia adalah sebuah sistem kuratorial yang aktif melakukan seleksi, pemeringkatan, dan amplifikasi konten berdasarkan serangkaian nilai komersial—yakni memaksimalkan waktu pengguna di platform (Haripriya & Manikandan, 2020).

Oleh karena itu, tanggung jawab platform justru lahir dari kekuatan prosedural yang mereka pegang. Ketika sebuah platform secara aktif memutuskan untuk menampilkan konten A di atas konten B kepada jutaan pengguna, ia telah melakukan tindakan kuratorial yang memiliki dampak nyata terhadap opini publik. Dalam konteks ini, tanggung jawab mereka bergeser dari sekadar *moderasi konten* (*content moderation*) reaktif menuju tata kelola algoritma (*algorithmic governance*) yang proaktif. Penelitian ini mengidentifikasi tiga dimensi tanggung jawab yang sering kali diabaikan:

- **Tanggung Jawab Transparansi:** Platform sangat tertutup mengenai cara kerja algoritma rekomendasinya. Tanpa transparansi dan kemampuan audit yang memadai, pihak eksternal—termasuk peneliti dan regulator—tidak dapat secara independen menilai apakah suatu platform secara tidak sengaja atau diam-diam memfasilitasi penyebaran hoax.
- **Tanggung Jawab Desain:** Model bisnis yang bertumpu pada *engagement* maksimal telah menciptakan *misalignment* atau ketidakselarasannya antara kepentingan platform (laba) dan kepentingan masyarakat (ekosistem informasi yang sehat). Tanggung jawab desain menuntut platform untuk tidak hanya mendeteksi hoax, tetapi juga meninjau ulang desain algoritmik yang secara inherent memberi imbalan pada sensasionalisme.
- **Tanggung Jawab Deteksi Proaktif:** Ketergantungan pada sistem *flagging* oleh pengguna dan mitigasi *post-hoc* (setelah konten viral) terbukti tidak memadai. Platform memiliki akses ke data perilaku yang dapat digunakan untuk mendeteksi pola kecurangan, seperti ledakan *engagement* yang tidak wajar dari akun-akun bot, secara lebih dini dan akurat.

3.3. Kelemahan Kebijakan yang Berfokus pada Konten, Bukan pada Sistem

Analisis terhadap kebijakan platform saat ini menunjukkan bahwa pendekatan mereka masih terlalu terfokus pada gejala (konten hoax individual) dan bukan pada akar penyebab (sistem amplifikasi). Sebagian besar sumber daya dihabiskan untuk mengembangkan model AI guna mendeteksi *deepfakes* atau teks buatan AI, serta menghapus konten dan akun yang melanggar. Meskipun penting, upaya ini bagaikan memadamkan api yang menyala di tengah badai bensin (Jayus, 2023).

Strategi yang lebih efektif dan berkelanjutan adalah dengan memutus rantai amplifikasinya. Ini dapat dilakukan dengan, pertama, mendeteksi dan menonaktifkan *jaringan bot* secara sistematis, bukan sekadar akun individual. Kedua, dan yang lebih penting, adalah mereformasi algoritma rekomendasi agar tidak terlalu rentan terhadap manipulasi *engagement* buatan. Beberapa langkah praktis yang dapat dipertimbangkan adalah memasukkan integritas informasi sebagai salah satu metrik dalam pelatihan algoritma, di samping *engagement*, serta memberikan opsi kepada pengguna untuk mengurutkan *feed* mereka berdasarkan kronologi atau sumber tepercaya, sehingga mengurangi ketergantungan mutlak pada kurasi algoritmik (Hussein & Hejase, 2022).

Dengan demikian, pembahasan ini menyimpulkan bahwa tanggung jawab platform digital dalam mengatasi hoax AI tidak dapat lagi dipisahkan dari tanggung jawab mereka atas desain dan tata kelola sistem algoritmik yang mereka ciptakan. Tanpa perubahan fundamental pada paradigma dan model bisnis ini, upaya platform dalam memerangi hoax akan tetap seperti menambal kebocoran di sebuah bendungan yang retak di mana-mana.

4. KESIMPULAN

Berdasarkan analisis yang telah diuraikan, dapat disimpulkan bahwa tantangan hoax yang dihasilkan oleh Artificial Intelligence (AI) merepresentasikan ancaman kualitatif baru terhadap ekosistem informasi. Ancaman ini tidak lagi sekadar tentang volume konten palsu, melainkan tentang kecepatan, skalabilitas, dan personalisasi disinformasi, yang diperkuat oleh infrastruktur platform digital itu sendiri. Penelitian ini menunjukkan bahwa inti permasalahannya terletak pada simbiosis mutualistik

antara akun bot AI dan algoritma rekomendasi, yang menciptakan sebuah siklus amplifikasi yang swaperkuat. Bot memanipulasi algoritma dengan memberikan sinyal *engagement* buatan, sementara algoritma, yang didorong oleh logika memaksimalkan perhatian pengguna, secara otomatis dan sistematis mendorong konten hoax tersebut ke ruang publik yang lebih luas, sehingga memfasilitasi manipulasi opini secara massal.

Dalam menghadapi realitas ini, klaim netralitas platform digital menjadi usang dan tidak lagi dapat dipertahankan. Argumen utama dari penelitian ini adalah bahwa tanggung jawab platform beraser secara fundamental dari peran pasif sebagai “penyedia layanan” menjadi peran aktif sebagai “pengelola ruang publik digital” yang harus memikul tanggung jawab prosedural. Tanggung jawab ini lahir dari kekuatan kuratorial yang mereka jalankan melalui algoritma rekomendasi. Oleh karena itu, upaya mitigasi yang hanya berfokus pada penghapusan konten (*content moderation*) bersifat reaktif dan tidak memadai.

UCAPAN TERIMA KASIH

Puji dan Syukur atas kehadiran tuhan yang Maha Esa atas segala rahmat dan karunia-nya penulis panjatkan sehingga dapat menyelesaikan penyusunan *paper* dengan judul “Tanggung Jawab Platform Digital Dalam Mengatasi Hoax Yang Dihasilkan Oleh Ai: Analisis Kasus Penyalahgunaan Bot Dan Algoritma Rekomendasi Untuk Manipulasi Opini Publik” dengan lancar dan tepat waktu.

Penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada Annisa Elfina Augustia M.Kom, selaku pembimbing utama, atas bimbingan, masukan, dan dukungannya dalam menyelesaikan *paper* ini.

Penulis juga menyadari bahwa *paper* ini jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang diharapkan demi perbaikan dimasa mendatang. Semoga *paper* ini dapat memberikan manfaat bagi pembaca dan menambah wawasan.

DAFTAR PUSTAKA

- Andika & M. Soemarno. (2023). Masalah Privasi dan Keamanan Data Pribadi pada Penerapan Kecerdasan Buatan. *INNOVATIVE: Journal Of Social Science Research*, 3, 4917–4929.
- Arly, A., Dwi, N., & Andini, R. (2023). Implementasi Penggunaan Artificial Intelligence Dalam Proses Pembelajaran Mahasiswa Ilmu Komunikasi di Kelas A. *Prosiding Seminar Nasional*, 362–374.
- Chairunnisa, C., Radityo, H., Wicaksono, H. R., & Ayyasy, S. T. (2021). Penerapan Algoritma pada Artificial Intelligence sebagai Upaya Menangani Penyebaran Hoax. *Cakrawala*, 15(2), 174–187. <https://doi.org/10.32781/cakrawala.v1i5.2.316>.
- Haripriya, S., & Manikandan, L. C. (2020). A Study on Artificial Intelligence Technologies and Their Applications. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3307, 336–344. <https://doi.org/10.32628/cseit206455>.
- Hussein, F., & Hejase, H. J. (2022). Artificial Intelligence and Fake News. *Žurnalistikos Tyrimai*, 16 April, 39–71. <https://doi.org/10.15388/zt/jr.2022.2>
- Jayus. (2023). Disrupsi Teknologi Digital dalam Penanganan Krisis. *Jurnal Pendidikan Tambusai*, 7(1), 1602–1610.
- Juditha, C. (2018). Hoax Communication Interactivity in Social Media and Anticipation (Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya). *Journal Pekommas*, 3(1), 31. <https://doi.org/10.30818/jpkm.2018.2030104>.