

Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>

E-ISSN: 2580-9113

P-ISSN: 2581-2033

LEX JOURNAL: KAJIAN HUKUM & KEADILAN

## Dilema *Mens Rea* pada Algoritma *Self-Learning*: Siapa yang Bertanggung Jawab Pidana Ketika AI Melakukan Diskriminasi Sistemik?

Aditya Yodiansyah<sup>1\*</sup>, Any Farida<sup>2</sup>, Agus Pramono<sup>3</sup>

<sup>1,2</sup>Fakultas Hukum, Universitas Darul Ulum Islamic Centre Sudirman

<sup>3</sup>Fakultas Hukum, Universitas Wisnuwardhana Malang

\*Email Korespondensi: [yodiansyahaditya@gmail.com](mailto:yodiansyahaditya@gmail.com)

### ABSTRAK

Abstrak ini mengeksplorasi tantangan mendasar yang dihadapi oleh sistem hukum pidana tradisional dalam menghadapi kemunculan kecerdasan buatan (AI) yang memiliki kemampuan belajar mandiri (*self-learning*). Masalah utama penelitian ini terletak pada ketidakmampuan doktrin *mens rea* konvensional untuk menjangkau tindakan otonom mesin yang menghasilkan dampak diskriminasi sistemik. Dengan menggunakan metode penelitian hukum normatif melalui pendekatan perundang-undangan dan konseptual, penelitian ini menganalisis bagaimana fenomena *black box* pada algoritma menciptakan celah akuntabilitas yang signifikan. Temuan penelitian menunjukkan bahwa meskipun AI belum diakui sebagai subjek hukum mandiri di Indonesia, terdapat urgensi untuk merekonstruksi model pertanggungjawaban pidana melalui pendekatan *vicarious liability* atau *strict liability* yang diadaptasi. Analisis ini juga membandingkan kerangka regulasi internasional seperti EU AI Act untuk merumuskan kebijakan kriminal yang responsif di Indonesia. Kesimpulannya, pembaruan hukum melalui KUHP Baru dan regulasi sektoral harus secara eksplisit mendefinisikan standar kepatuhan algoritmik untuk mencegah diskriminasi sistemik dan menjamin kepastian hukum di era digital.

**Kata Kunci:** Kecerdasan Buatan, *Mens Rea*, Diskriminasi Sistemik, Pertanggungjawaban Pidana, Algoritma *Self-Learning*

### 1. PENDAHULUAN

Transformasi digital yang dipicu oleh revolusi industri 4.0 telah membawa umat manusia ke ambang peradaban baru di mana pengambilan keputusan tidak lagi menjadi hak prerogatif biologis manusia (Wijanarko et al., 2026). Kecerdasan Buatan (*Artificial Intelligence/AI*), khususnya yang berbasis algoritma *self-learning* atau *machine learning*, kini telah terintegrasi secara mendalam ke dalam infrastruktur kritis masyarakat, mulai dari sistem keuangan, layanan kesehatan, hingga sistem peradilan pidana (Hashmi et al.,

**Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>**

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

2025). Namun, di balik efisiensi dan janji objektivitas yang ditawarkan, terdapat ancaman laten berupa diskriminasi sistemik yang muncul dari proses pemrosesan data yang bersifat otonom dan sering kali tidak tertembus oleh logika manusia (The Leadership Conference on Civil and Human Rights, 2026). Fenomena ini menghadirkan tantangan eksistensial bagi hukum pidana, sebuah sistem norma yang selama berabad-abad dibangun di atas asumsi bahwa kejahatan hanya dapat lahir dari kehendak bebas dan niat jahat manusia (Setiawan et al., 2026).

Masalah utama dalam lanskap hukum kontemporer adalah bagaimana mengatribusikan kesalahan pidana ketika sebuah luaran (*output*) algoritma menyebabkan kerugian yang secara material dapat dikategorikan sebagai tindak pidana, namun secara formal tidak memiliki subjek manusia yang secara langsung menghendaki hasil tersebut (Abdelkarim et al., 2025). Dalam doktrin hukum pidana yang bersifat antroposentris, pertanggungjawaban pidana mensyaratkan adanya perpaduan antara *actus reus* (perbuatan yang salah) dan *mens rea* (niat yang salah) (Hashmi et al., 2025). Dilema *mens rea* muncul ketika algoritma *self-learning* bertindak di luar instruksi eksplisit pengembangnya, belajar dari pola data sejarah yang bias, dan akhirnya mengeksekusi keputusan yang bersifat diskriminatif secara sistemik terhadap kelompok masyarakat tertentu berdasarkan ras, gender, atau status sosial ekonomi (The Leadership Conference on Civil and Human Rights, 2026).

Urgensi penelitian ini didasarkan pada kenyataan bahwa teknologi AI saat ini sering kali beroperasi sebagai "kotak hitam" (*black box*), di mana proses internal yang mendasari sebuah keputusan tidak dapat dijelaskan bahkan oleh perancangannya sendiri (Brenner & Slowik, 2025). Ketidakmampuan untuk melacak asal-usul "niat" dalam algoritma ini menciptakan celah akuntabilitas (*accountability gap*) yang membahayakan prinsip keadilan bagi korban diskriminasi (Kersten, 2025). Di Indonesia, tantangan ini semakin tajam dengan berlakunya Kitab Undang-Undang Hukum Pidana (KUHP) Nasional yang

**Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>**

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

baru (UU No. 1 Tahun 2023), yang meskipun telah memperluas subjek hukum korporasi, belum secara eksplisit mengatur mengenai agensi non-manusia yang otonom (Hendriana, 2026).

Untuk memberikan landasan teoretis yang kuat, penelitian ini mengacu pada lima penelitian terdahulu yang krusial. Pertama, studi kasus *Mata v. Avianca* (2023) di Amerika Serikat yang mengekspos risiko "halusinasi" AI dalam dokumen hukum, menunjukkan bahwa kesalahan mesin dapat berujung pada sanksi serius bagi praktisi manusia (Rondonuwu et al., 2025). Kedua, laporan Ketua Mahkamah Agung AS, Roberts, (2023), yang menegaskan bahwa penilaian hukum tetap memerlukan "sentuhan manusia" karena AI tidak mampu menangkap konteks sosial dan moral secara utuh (Rondonuwu et al., 2025). Ketiga, penelitian Li, (2023) mengenai "*The Dark Side of ChatGPT*" yang menganalisis fenomena *Stochastic Parrots*, di mana AI menghasilkan luaran meyakinkan tanpa pemahaman makna, sehingga mengaburkan elemen kesengajaan dalam penyebaran disinformasi. Keempat, riset Matthew Dahl et al., (2024) berjudul "*Large Legal Fictions*" yang memetakan bagaimana model bahasa besar dapat menyimpang dari data pelatihan aslinya, menciptakan kesulitan dalam pembuktian kausalitas kesalahan. Kelima, penelitian klasik oleh Lum (2017) dan Sartor (2009) yang telah mengidentifikasi bahwa penggunaan *machine learning* dalam mitigasi bias yudisial justru berisiko memperkuat bias tersebut jika tidak disertai transparansi algoritma yang dapat diaudit.

Uniknya, penelitian ini tidak hanya berhenti pada kritik terhadap bias algoritmik, tetapi juga mencoba melakukan rekonstruksi doktrinal terhadap konsep *mens rea* itu sendiri. Kebaruan penelitian ini terletak pada usulan "*Algorithmic Mens Rea*" dan analisis komprehensif mengenai posisi hukum Indonesia dalam menghadapi diskriminasi sistemik oleh AI pasca-pemberlakuan KUHP Baru (Fransisco, 2025). Signifikansi penelitian ini diharapkan dapat menjadi rujukan bagi pembuat kebijakan dalam

**Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>**

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & Keadilan**

merumuskan regulasi AI yang tidak hanya progresif dalam inovasi, tetapi juga protektif terhadap hak asasi manusia dan prinsip-prinsip negara hukum (Mecca et al., 2025).

Penelitian ini bertujuan untuk menjawab dua pertanyaan fundamental: Pertama, bagaimana tantangan pembuktian *mens rea* pada algoritma *self-learning* yang bersifat *black box*? Kedua, bagaimana model pertanggungjawaban pidana yang paling tepat dalam menjerat pelaku di balik diskriminasi sistemik AI dalam sistem hukum Indonesia? Melalui analisis ini, pembaca akan diajak untuk memahami bahwa diskriminasi algoritmik bukanlah sekadar kesalahan teknis, melainkan sebuah bentuk ketidakadilan struktural yang membutuhkan respons hukum pidana yang transformatif (Rahmah et al., 2025).

## **2. METODE PENELITIAN**

Metode penelitian yang digunakan dalam artikel ini adalah penelitian hukum normatif (*normative legal research*), yang secara primer berfokus pada analisis terhadap norma-norma hukum positif, asas-asas hukum, dan doktrin-doktrin yang relevan dalam literatur hukum pidana dan teknologi (Kurniawan, 2023). Penelitian ini tidak melakukan pengujian lapangan terhadap data statistik, melainkan melakukan abstraksi teoretis terhadap fenomena hukum yang muncul akibat kemajuan teknologi kecerdasan buatan. Pendekatan penelitian yang diterapkan mencakup pendekatan perundang-undangan (*statute approach*), pendekatan konseptual (*conceptual approach*), dan pendekatan perbandingan (*comparative approach*) (Kurniawan, 2023).

Pendekatan perundang-undangan digunakan untuk menelaah kesesuaian antara regulasi yang ada di Indonesia—seperti UU No. 1 Tahun 2023 tentang KUHP, UU ITE (UU No. 1 Tahun 2024), dan UU Pelindungan Data Pribadi (UU No. 27 Tahun 2022)—dengan karakteristik unik algoritma *self-learning* (Hendriana, 2026). Pendekatan konseptual bertujuan untuk membedah ulang teori-teori pertanggungjawaban pidana,

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

khususnya mengenai elemen mental (*mens rea*) dan fisik (*actus reus*), guna menemukan formula baru yang dapat mengakomodasi tindakan otonom mesin. Sementara itu, pendekatan perbandingan dilakukan dengan merujuk pada kerangka hukum di Uni Eropa, khususnya *EU AI Act*, untuk melihat bagaimana yurisdiksi lain menangani klasifikasi risiko AI dan atribusi tanggung jawab (Puannandini et al., 2025).

Data yang dianalisis dalam penelitian ini bersifat sekunder, yang terdiri dari bahan hukum primer berupa peraturan perundang-undangan nasional dan internasional; bahan hukum sekunder berupa jurnal ilmiah bereputasi, buku teks hukum, dan laporan dari lembaga riset terkemuka; serta bahan hukum tersier seperti kamus hukum dan ensiklopedia (Hutapea et al., 2026). Teknik pengumpulan data dilakukan melalui studi kepustakaan (*library research*) dengan menelusuri pangkalan data digital untuk mengidentifikasi perkembangan hukum terbaru hingga tahun 2025.

Proses analisis data dilakukan secara kualitatif dengan menggunakan metode deduktif, di mana prinsip-prinsip umum hukum pidana diterapkan pada kasus-kasus spesifik diskriminasi algoritmik untuk menarik kesimpulan yang koheren (Fransisco, 2025). Keterbatasan penelitian ini terletak pada fokusnya yang sangat spesifik pada ranah pidana, sehingga aspek perdata atau administratif mungkin hanya akan dibahas secara marginal sebagai pendukung analisis utama. Namun, dengan menggabungkan teori hukum, etika teknologi, dan analisis regulasi, penelitian ini berupaya memberikan gambaran yang utuh mengenai solusi hukum terhadap dilema *mens rea* pada kecerdasan buatan (Lendvai & Gosztonyi, 2025).

Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>

E-ISSN: 2580-9113

P-ISSN: 2581-2033

LEX JOURNAL: KAJIAN HUKUM & Keadilan

### 3. HASIL & PEMBAHASAN

#### Dekonstruksi *Mens Rea* dalam Algoritma *Black-Box*: Krisis Doktrin Kesalahan pada Agensi Non-Manusia

Hukum pidana tradisional telah lama berakar pada paradigma antroposentris yang menempatkan manusia sebagai satu-satunya subjek yang mampu memiliki "pikiran yang bersalah" (*guilty mind*). Konsep *mens rea* bukan sekadar elemen teknis, melainkan fondasi moral yang membedakan antara kecelakaan murni dan kejahatan yang patut dicela (Hashmi et al., 2025). Namun, kemunculan algoritma *self-learning* berbasis *deep learning* telah memicu krisis ontologis dalam doktrin ini. Algoritma ini tidak lagi sekadar menjalankan perintah linear "jika-maka" yang diprogram oleh manusia; sebaliknya, mereka mengembangkan pola pengambilan keputusan sendiri melalui pemrosesan data masif dalam ribuan dimensi yang sering kali melampaui pemahaman kognitif penciptanya (Hendriana, 2026).

Dilema fundamental muncul ketika algoritma tersebut menghasilkan luaran yang diskriminatif secara sistemik. Sebagai contoh, dalam sistem penilaian risiko kriminal seperti COMPAS di Amerika Serikat, algoritma tersebut ditemukan secara konsisten memberikan skor risiko yang lebih tinggi kepada terdakwa kulit hitam dibandingkan dengan terdakwa kulit putih dengan profil serupa (Cofone & Khern-am-nuai, 2025). Dalam kasus semacam ini, pertanyaan krusial yang muncul adalah: di mana letak *mens rea*-nya? Jika pengembang tidak secara sengaja menyuntikkan bias ke dalam kode, dan jika pengguna (seperti hakim atau polisi) hanya mengikuti rekomendasi mesin dengan itikad baik, maka dalam kerangka hukum pidana konvensional, tidak ada subjek yang memenuhi ambang batas "niat" atau "pengetahuan" untuk dipidana (Abdelkarim et al., 2025).

Krisis ini diperparah oleh fenomena *black box*. Berbeda dengan perangkat lunak tradisional yang logika operasinya dapat diaudit baris demi baris, model AI modern

Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>

E-ISSN: 2580-9113

P-ISSN: 2581-2033

**LEX JOURNAL: KAJIAN HUKUM & Keadilan**

seperti jaringan saraf tiruan (*neural networks*) bersifat *opaque*. Proses penimbangan variabel di dalam "otak" digital tersebut bersifat probabilistik dan dinamis (Brenner & Slowik, 2025). Hal ini menciptakan tantangan pembuktian yang hampir mustahil bagi jaksa penuntut umum: bagaimana membuktikan adanya kesengajaan (*dolus*) atau kealpaan (*culpa*) ketika mekanisme internal pengambilan keputusan AI tidak dapat dijelaskan (*non-explainable*)? (Fransisco, 2025).

Karakteristik Perbandingan	Kesalahan Manusia (Tradisional)	Kesalahan Algoritma (Self-Learning)
Sumber Kesalahan	Kesadaran dan Kehendak Bebas	Logika Matematika dan Probabilitas (Hashmi et al., 2025).
Transparansi Proses	Dapat Dijelaskan melalui Motif	Bersifat <i>Opaque (Black Box)</i> (Murodova, 2025).
Bentuk Pertanggungjawaban	Subjektif ( <i>Mens Rea</i> )	Objektif/Teknis ( <i>Algorithmic Error</i> ) (Fransisco, 2025).
Pola Pembelajaran	Pengalaman dan Nilai Moral	<i>Feedback Loops</i> dan Bias Data (The Leadership Conference on Civil and Human Rights, 2026).
Dampak Pelanggaran	Individual dan Terlokalisasi	Sistemik dan Berskala Masif (Ahmad et al., 2025).

Dalam ranah teoretis, kegagalan AI sering kali diklasifikasikan sebagai "*Hard AI Crime*", yaitu situasi di mana tindakan berbahaya oleh sistem otonom tidak dapat direduksi menjadi kesalahan manusia tunggal mana pun (Osman et al., 2025). Sebagaimana diuraikan oleh Abbott dan Sarch, dalam kondisi ini, penggunaan doktrin *mens rea* tradisional meninggalkan "lubang moral" (*moral void*) (Osman et al., 2025). Jika hukum tetap bersikeras pada pembuktian niat manusia, maka kejahatan yang dilakukan oleh algoritma akan terus berlanjut tanpa konsekuensi hukum, yang pada

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & Keadilan**

akhirnya akan menggerus kepercayaan publik terhadap sistem peradilan (Osman et al., 2025).

Lebih jauh lagi, diskriminasi sistemik oleh AI sering kali lahir dari tiga jalur utama yang saling berkelindan. Pertama, bias representasi dalam data pelatihan, di mana AI belajar dari catatan sejarah yang sudah mengandung prasangka sosial (The Leadership Conference on Civil and Human Rights, 2026). Kedua, penggunaan variabel *proxy*, di mana algoritma menggunakan data yang tampak netral (seperti kode pos atau tingkat pendidikan) untuk secara tidak langsung mengklasifikasikan orang berdasarkan ras atau kelas sosial (The Leadership Conference on Civil and Human Rights, 2026). Ketiga, penciptaan *feedback loops*, di mana keputusan bias AI di masa lalu menjadi input data untuk pelatihan di masa depan, sehingga mempermanenkan ketidakadilan tersebut dalam siklus yang tak terputus (Rodriguez, 2025).

Sebagai contoh nyata, sistem *predictive policing* yang menggunakan data penangkapan historis akan cenderung mengirim lebih banyak petugas ke lingkungan yang secara historis sudah mengalami *over-policing* (The Leadership Conference on Civil and Human Rights, 2026). Hal ini menghasilkan lebih banyak penangkapan di lokasi tersebut, yang kemudian dimasukkan kembali ke dalam algoritma sebagai "bukti" bahwa area tersebut rawan terhadap kriminalitas. Proses ini secara teknis disebut sebagai *self-fulfilling prophecy* (The Leadership Conference on Civil and Human Rights, 2026). Secara hukum, sangat sulit untuk menjerat siapa pun dengan tuduhan diskriminasi rasial yang sengaja, karena "niat" tersebut tersembunyi di dalam optimasi statistik yang murni bertujuan untuk meningkatkan efisiensi operasional (Hashmi et al., 2025).

Dalam konteks hukum Indonesia, doktrin *mens rea* menghadapi hambatan pada Pasal 1 ayat (1) KUHP yang menegaskan bahwa tiada suatu perbuatan dapat dipidana tanpa ketentuan undang-undang yang telah ada (*nullum delictum sine lege*) (Puannandini et al., 2025). KUHP Baru (UU No. 1 Tahun 2023) memang telah mengakui korporasi

**Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>**

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

sebagai subjek hukum, namun ia tetap mendasarkan kesalahan korporasi pada kesalahan pengurusnya (Paramita et al., 2026). Pertanyaannya, dapatkah seorang manajer perusahaan AI dianggap memiliki *mens rea* jika ia sendiri tidak bisa memprediksi bahwa algoritmanya akan mengembangkan perilaku diskriminatif setelah berinteraksi dengan data dunia nyata? (Osman et al., 2025).

Untuk mengatasi kebuntuan ini, para ahli hukum mulai mengusulkan pergeseran dari *mens rea* psikologis (apa yang ada di dalam pikiran) menuju *mens rea* fungsional atau objektif. Dalam model ini, fokus dialihkan pada apakah pengembang atau pengguna telah memenuhi "standar kehati-hatian algoritmik" (*algorithmic due diligence*) (Ginting, 2025). Jika sebuah perusahaan merilis AI berisiko tinggi tanpa melakukan audit bias yang memadai, maka kegagalan untuk melakukan tindakan pencegahan tersebut dapat dikategorikan sebagai bentuk kealpaan (*negligence*) yang dapat dipidana (Fransisco, 2025). Rekonstruksi ini sangat penting agar hukum tetap relevan di tengah agensi mesin yang semakin otonom dan sulit diprediksi (Hashmi et al., 2025).

Krisis doktrinal ini menuntut kita untuk tidak hanya melihat AI sebagai alat (*tool*), tetapi sebagai aktor baru dalam lanskap kriminologi (O'Keefe et al., 2025). Sebagaimana korporasi di masa lalu pernah dianggap tidak mungkin memiliki niat jahat, namun kini diakui sebagai subjek hukum, AI mungkin memerlukan status serupa—*electronic personhood*—agar akuntabilitas hukum tidak menguap di balik kompleksitas kode (Burgess, 2016). Tanpa adanya dekonstruksi terhadap konsep kesalahan, diskriminasi sistemik oleh algoritma akan menjadi bentuk penindasan baru yang kebal terhadap jangkauan pedang keadilan (Rahmah et al., 2025).

**Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>**

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

### **Dialektika Pertanggungjawaban: Rekonstruksi Subjek Hukum dan Formulasi Kebijakan Kriminal di Indonesia**

Setelah memahami bahwa doktrin *mens rea* tradisional mengalami krisis di hadapan algoritma *black box*, tantangan selanjutnya adalah merumuskan model pertanggungjawaban pidana yang operasional dalam sistem hukum Indonesia. Dialektika ini berkisar pada pertanyaan: apakah kita harus terus memaksa agensi AI ke dalam kotak subjek hukum manusia, ataukah kita perlu menciptakan kategori subjek hukum baru yang *sui generis*? (Fransisco, 2025). Di tengah kekosongan regulasi khusus mengenai AI di Indonesia, arah pertanggungjawaban saat ini masih bersandar pada interpretasi analogis terhadap hukum korporasi dan hukum informasi elektronik (Hendriana, 2026).

Salah satu model yang paling banyak didiskusikan secara global adalah model pertanggungjawaban tidak langsung melalui manusia di belakang mesin (*Perpetration-by-Another Model*). Dalam model ini, AI diposisikan sebagai agen yang tidak bersalah (*innocent agent*), sementara pengembang atau penggunanya dianggap sebagai pelaku utama (*principal*) jika mereka sengaja merancang atau menggunakan AI untuk tujuan melawan hukum (Mecca et al., 2025). Namun, model ini menunjukkan kelemahannya pada kasus diskriminasi sistemik yang muncul secara spontan dari proses pembelajaran mandiri algoritma. Ketika harmoni antara niat manusia dan tindakan mesin terputus, model agensi tradisional ini tidak lagi memadai untuk memberikan keadilan (Osman et al., 2025).

Sebagai alternatif, model *vicarious liability* (pertanggungjawaban pengganti) menawarkan solusi yang lebih praktis dalam konteks korporasi. Dalam kerangka hukum Indonesia, konsep ini mulai diakomodasi melalui doktrin pertanggungjawaban korporasi dalam KUHP Baru (Paramita et al., 2026). Jika sebuah sistem AI yang dioperasikan oleh sebuah bank melakukan diskriminasi sistemik dalam pemberian pinjaman, maka bank tersebut dapat dimintai pertanggungjawaban bukan karena memiliki niat psikologis untuk

Tersedia di online: <http://ejournal.unitomo.ac.id/index.php/hukum>

E-ISSN: 2580-9113

P-ISSN: 2581-2033

LEX JOURNAL: KAJIAN HUKUM & KEADILAN

diskriminatif, melainkan karena tindakannya dilakukan dalam lingkup usaha korporasi dan memberikan keuntungan bagi korporasi tersebut (Puannandini et al., 2025).

Model Pertanggungjawaban Pidana AI	Fokus Atribusi	Dasar Kesalahan	Relevansi di Indonesia
<i>Perpetration-by-Another</i>	Manusia Pengendali	Kesengajaan/Niat Manusia (Hallevy, 2010).	Sangat Relevan (Prinsip Umum) (Mecca et al., 2025).
<i>Vicarious Liability</i>	Entitas Korporasi	Hubungan Kerja/Kedudukan (Puannandini et al., 2025).	Diakomodasi KUHP Baru (Paramita et al., 2026).
<i>Strict Liability</i>	Pemilik/Penyedia	Dampak/Hasil Tanpa Niat (Osman et al., 2025).	Terbatas (Sektor Khusus) (Hendriana, 2026).
<i>Electronic Personhood</i>	Sistem AI Sendiri	Kapasitas Hukum Mandiri (Abdelkarim et al., 2025).	Belum Diakui ( <i>Ius Constituendum</i> ) (Hendriana, 2026).

Penerapan *strict liability* (pertanggungjawaban mutlak) juga menjadi diskursus yang mendesak. Dalam kasus diskriminasi sistemik yang bersifat masif dan sulit dibuktikan niatnya, penegakan hukum berdasarkan standar pembuktian yang ketat akan sangat merugikan korban. Dengan *strict liability*, beban pembuktian dialihkan: asalkan dapat dibuktikan bahwa AI tersebut menghasilkan dampak diskriminatif, maka penyedia layanan dianggap bertanggung jawab secara hukum tanpa perlu membuktikan adanya niat jahat (Osman et al., 2025). Model ini sangat efektif untuk memaksa perusahaan teknologi melakukan audit etika dan teknis secara ketat sebelum merilis produk mereka ke publik (The Leadership Conference on Civil and Human Rights, 2026).

Namun, di Indonesia, tantangan terbesar adalah absennya regulasi setingkat undang-undang yang secara spesifik mengatur AI. Saat ini, kita hanya memiliki instrumen

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

non-binding berupa Surat Edaran Menkominfo No. 9 Tahun 2023 tentang Etika Kecerdasan Buatan (Hendriana, 2026). Walaupun UU ITE (UU No. 1 Tahun 2024) dan UU Pelindungan Data Pribadi (UU No. 27 Tahun 2022) dapat digunakan untuk menjerat beberapa aspek penyalahgunaan AI, kedua undang-undang tersebut belum dirancang untuk menghadapi otonomi algoritma *self-learning* (Hendriana, 2026). Misalnya, Pasal 27 UU ITE tentang disinformasi mensyaratkan adanya kesengajaan menyebarkan berita bohong, sebuah elemen yang sulit dibuktikan jika informasi tersebut dihasilkan secara otomatis oleh "halusinasi" AI (Rondonuwu et al., 2025).

Indonesia perlu belajar dari Uni Eropa yang telah merumuskan *EU AI Act* dengan pendekatan berbasis risiko (*risk-based approach*). Dalam regulasi tersebut, sistem AI yang digunakan untuk penilaian sosial (*social scoring*) oleh pemerintah dilarang sepenuhnya karena risiko diskriminasi sistemik yang tidak dapat diterima (Lendvai & Gosztonyi, 2025). Sementara itu, AI yang digunakan dalam penegakan hukum, rekrutmen, atau pendidikan diklasifikasikan sebagai risiko tinggi (*high-risk*) yang wajib memenuhi standar transparansi, tata kelola data, dan pengawasan manusia (*human oversight*) yang sangat ketat (Caparrós, 2025). Kegagalan memenuhi standar ini berujung pada sanksi administratif dan denda yang sangat besar, mencapai €35 juta atau 7% dari perputaran uang global perusahaan (Caparrós, 2025).

Formulasi kebijakan kriminal Indonesia ke depan harus mencakup tiga pilar utama. Pertama, pengakuan AI sebagai entitas yang memiliki status hukum fungsional, sehingga hukuman dapat dijatuhkan langsung kepada "sistem" (misalnya berupa perintah penghapusan algoritma atau denda digital) (Fransisco, 2025). Kedua, mewajibkan audit algoritmik secara berkala bagi setiap pengembang AI risiko tinggi untuk mendeteksi dan memitigasi bias sejak tahap desain (*privacy and fairness by design*) (Medda-Windischer & Crepez, 2025). Ketiga, memperkuat kapasitas aparat penegak hukum dalam bidang

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

forensik digital agar mampu membongkar "kotak hitam" algoritma saat terjadi tindak pidana (Kristian & Ramadan, 2025).

Penting juga untuk memperhatikan bahwa diskriminasi sistemik sering kali merupakan pelanggaran terhadap hak-hak konstitusional. Pasal 28I ayat (2) UUD 1945 menjamin perlindungan terhadap perlakuan diskriminatif atas dasar apa pun (Rahmah et al., 2025). Ketika sebuah institusi negara atau swasta menggunakan AI yang bias dalam pengambilan keputusan publik, hal tersebut bukan sekadar kegagalan teknis, melainkan pelanggaran hak asasi manusia yang serius. Oleh karena itu, Mahkamah Agung perlu mempertimbangkan penerbitan Peraturan Mahkamah Agung (PERMA) tentang etika dan tata kelola AI dalam proses yudisial guna menjamin bahwa teknologi tidak menjadi instrumen yang memperlebar jurang ketidakadilan (Mazli, 2025).

Dalam jangka panjang, dialektika ini akan berujung pada perlunya harmoni antara inovasi teknologi dan kepastian hukum. Kita tidak boleh mematikan inovasi AI dengan beban hukum yang mustahil dipenuhi, namun kita juga tidak boleh membiarkan masyarakat menjadi subjek eksperimen algoritmik yang tidak bertanggung jawab (Lendvai & Gosztonyi, 2025). Tanggung jawab pidana harus tetap melekat pada manusia sebagai pengendali akhir (*human-in-command*), namun dengan doktrin yang lebih adaptif dan standar kepatuhan yang lebih jelas (Osman et al., 2025). Hanya dengan cara inilah, dilema *mens rea* pada algoritma *self-learning* dapat diselesaikan tanpa mengorbankan nilai-nilai luhur keadilan yang menjadi tujuan akhir dari hukum itu sendiri (Hendriana, 2026).

#### **4. KESIMPULAN**

Kehadiran kecerdasan buatan berbasis algoritma *self-learning* telah menciptakan tantangan yang belum pernah terjadi sebelumnya bagi doktrin hukum pidana konvensional. Masalah utama terletak pada dekonstruksi konsep *mens rea* yang bersifat

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

antroposentris, di mana agensi otonom mesin menghancurkan asumsi tradisional mengenai niat dan kesalahan batin. Fenomena *black box* pada algoritma tidak hanya menghalangi transparansi pengambilan keputusan, tetapi juga menciptakan celah akuntabilitas yang memungkinkan terjadinya diskriminasi sistemik tanpa ada subjek manusia yang dapat dijerat secara langsung menurut standar pembuktian hukum saat ini.

Penelitian ini menyimpulkan bahwa untuk menghadapi realitas ini, sistem hukum Indonesia harus segera bertransformasi dari pendekatan yang murni subjektif menuju model pertanggungjawaban yang lebih fungsional dan berbasis risiko. Implementasi KUHP Baru (2023) harus diperkuat dengan regulasi sektoral yang secara eksplisit mengatur tata kelola AI, termasuk kewajiban audit algoritmik dan pengawasan manusia yang berkelanjutan. Penggunaan model *vicarious liability* untuk korporasi dan potensi pengakuan *electronic personhood* bagi AI risiko tinggi menjadi langkah krusial untuk memastikan bahwa keadilan tetap dapat ditegakkan di era otomatisasi.

Sebagai rekomendasi kebijakan, pemerintah Indonesia didorong untuk segera meningkatkan status Surat Edaran Etika AI menjadi Undang-Undang Kecerdasan Buatan yang komprehensif, mengadopsi standar internasional seperti EU AI Act, namun tetap disesuaikan dengan konteks sosiologis dan konstitusional Indonesia. Selain itu, diperlukan peningkatan literasi AI bagi para hakim, jaksa, dan polisi untuk menjamin bahwa proses penegakan hukum di masa depan mampu menangani bukti digital yang kompleks. Akhirnya, tujuan hukum pidana harus tetap konsisten: memastikan bahwa setiap kemajuan teknologi membawa kemanfaatan bagi kemanusiaan, bukan justru menjadi alat penindasan baru yang tersembunyi di balik tabir algoritma.

## **5. REFERENSI**

Abdelkarim, M., Youcef, M., Mounia, B., & Maher, B. (2025). The Implications of Artificial Intelligence on Criminal Liability in Contemporary Criminal Thought. *Lex Localis-Journal of Local Self-Government*, 23(8), 1–10.

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

- Ahmad, A., Vallès, Y., & Idaghdour, Y. (2025). Bias in AI systems: integrating formal and socio-technical approaches. *Machine Learning and Artificial Intelligence*, 5. <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2025.1686452/full>
- Brenner, G., & Slowik, J. (2025, January 10). [Podcast] AI at Work: Black Box Issues. *Law and the Workplace*. <https://www.lawandtheworkplace.com/2025/01/podcast-black-box-issues/>
- Burgess, M. (2016). *Google's AI has written some amazingly mournful poetry*. Wired.Co.Uk. <https://www.wired.co.uk/article/google-artificial-intelligence-poetry>
- Caparrós, V. A. (2025). EU AI Act: Prohibited practices and penalties. *AGM Abogados*. <https://www.agmabogados.com/en/eu-ai-act-prohibited-practices-and-penalties/>
- Cofone, I., & Khern-am-nuai, W. (2025). The Overstated Cost of AI Fairness in Criminal Justice. *Indiana Law Journal*, 100(4), 1432–1478.
- Fransisco, W. (2025). Drafting Laws for the Lifeless: A Legal Framework for Criminal Liability and Punishment for Artificial Intelligence. *Jurnal Hukum Dan Peradilan*, 14(3), 701–718. <https://doi.org/10.25216/jhp.14.3.2025.701-718>
- Ginting, Y. P. (2025). Akuntabilitas Pidana di Era Penyalahgunaan Kecerdasan Buatan ditinjau dari Asas Legalitas. *The Prosecutor Law Review*, 3(3), 48–60.
- Hallevy, G. (2010). “I, Robot – I, Criminal”—When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal Offenses. *Syracuse Science & Technology Law Reporter*, 22(1), 1–37. <https://jost.syr.edu/wp-content/uploads/Hallevy-Volume-22.pdf>
- Hashmi, M. A. I., Butt, M. F., Jawad, M., & Sultan, S. (2025). Criminal Liability in the Age of Autonomous Systems: Rethinking Mens Rea and Actus Reus. *The Critical Review of Social Sciences Studies*, 3(3), 290–303.
- Hendriana, D. Y. (2026, February 13). Mencari Arah Pertanggungjawaban Hukum Artificial Intelligence di Indonesia. *Mahkamah Agung RI*. <https://marinews.mahkamahagung.go.id/artikel/mencari-arah-pertanggungjawaban-hukum-ai-di-indonesia-0S7>
- Hutapea, N. M. S., Sitepu, D. K. C., Damanik, J., & Sianipar, S. K. L. (2026). Artificial Intelligence and Criminal Liability: A Preliminary Study within the Indonesian Legal System. *Jurnal Ilmu Hukum Kyadiren*, 7(2), 688–704. <https://doi.org/10.46924/jihk.v7i2.330>

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

- Kersten, M. (2025, August 6). Algorithms, Automation and Accountability: Imagining Responsibility for the Crimes of Machines. *Justice in Conflict*. <https://justiceinconflict.org/2025/08/06/algorithms-automation-and-accountability-imagining-responsibility-for-the-crimes-of-machines/>
- Kristian, B., & Ramadan, T. A. (2025). The Deterrence Dilemma: Assessing Criminal Liability Standards for Emerging Digital Offenses in the AI Era. *Realism: Law Review*, 3(3). <https://doi.org/10.71250/rlr.v3i3.86>
- Kurniawan, I. D. (2023). Analisis terhadap Artificial Intelligence sebagai Subjek Hukum Pidana. *Mutiara: Jurnal Ilmiah Multidisiplin Indonesia*, 1(1), 35–44.
- Lendvai, G. F., & Gosztonyi, G. (2025). Algorithmic Bias as a Core Legal Dilemma in the Age of Artificial Intelligence: Conceptual Basis and the Current State of Regulation. *Laws*, 14(3). <https://doi.org/10.3390/laws14030041>
- Li, Z. (2023). The Dark Side of ChatGPT: Legal and Ethical Challenges from Stochastic Parrots and Hallucination. *ArXiv Preprint Ar Xiv:2304.14347*.
- Lum, K. (2017). Limitations of mitigating judicial bias with machine learning. *Nature Human Behaviour*. <https://doi.org/10.1038/S41562-017-0141>
- Matthew Dahl, V., Magesh, M. S., & Ho, D. E. (2024). Large Legal Fictions: Profiling Legal allucinations in Large Language Models. *Journal of Legal Analysis*.
- Mazli, A. (2025). Menakar Pertanggungjawaban Artificial Intelligence dalam Tata Hukum Indonesia. <https://marinews.mahkamahagung.go.id/artikel/menakar-pertanggungjawaban-artificial-intelligence-0uZ>
- Mecca, A. S. P., Hidayat, W. A., & Tuasikal, H. (2025). Pemanfaatan Teknologi Kecerdasan Buatan (Artificial Intelligence) dalam Sistem Peradilan Pidana di Indonesia. *Jurnal Sosial Dan Teknologi*, 5(6), 1730–1746.
- Medda-Windischer, R., & Crepaz, K. (2025). *Reframing Minority Rights Amid Global Challenges: The Role of AI and Algorithmic Fairness in Promoting Diversity and Inclusion*. <https://www.eurac.edu/en/blogs/midas/reframing-minority-rights-amid-global-challenges-the-role-of-ai-and-algorithmic-f>
- Murodova, S. (2025). The Black Box Problem in Administering Justice: Risks of Opaque Algorithms in Legal Decision-Making. *International Journal of Law and Policy*, 3(6), 1–20. <https://doi.org/10.59022/ijlp.331>
- O’Keefe, C., Ramakrishnan, K., Tay, J., & Winter, C. (2025). *Law-Following AI: Designing AI Agents to Obey Human Laws*. <https://law-ai.org/law-following-ai/>
- Osman, Y. H. M., John, D. J., Ahmed, D. A., Mohamed, R. M. A., & Ali, R. M. S. A.

**Tersedia di online:** <http://ejournal.unitomo.ac.id/index.php/hukum>

**E-ISSN: 2580-9113**

**P-ISSN: 2581-2033**

**LEX JOURNAL: KAJIAN HUKUM & KEADILAN**

- (2025). Criminal Responsibility in the Age of AI: Rethinking Legal Accountability. *African Journal of Law and Justice System*, 4(3). <https://journals.co.za/doi/10.31920/2753-3123/2025/v4n3a9>
- Paramita, D., Adi, A. B., Ghiffari, M., & Salim, J. (2026). Indonesia: New Criminal Law Framework Takes Effect in 2026 - Highlights of Corporate Criminal Liability and Deferred Prosecution Agreement. *Indonesia Newsletter*, 39. <https://www.morihamada.com/en/insights/newsletters/131626>
- Puannandini, D. A., Fabian, R. M., Firdaus, R. A. P., Mustopa, M. Z., & Herdiyana, I. (2025). Liabilitas Produk Ai dalam Sistem Hukum Indonesia: Implikasi Bagi Pengembang, Pengguna, dan Penyedia Layanan. *IURIS STUDIA: Jurnal Kajian Hukum*, 6(1), 24–33.
- Rahmah, J. Z., Wijaya, T. D., Setyawan, F., & Floranta, F. (2025). Integrasi Artificial Intelligence dalam Pengembangan Hukum Perjanjian di Indonesia. *Journal of Innovative and Creativity*, 5(3), 32412–32424.
- Roberts, J. G. (2023). *2023 Year-End Report on the Federal Judiciary* (Technical Report December).
- Rodriguez, F. S. (2025). How AI is Setting the stage for a Digital Jim Crow Era. *CHCI Policy Brief: Developing the Next Generation of Latino Leaders*, 1–4.
- Rondonuwu, N. T., Setiabudhi, D. O., & Gerungan, C. A. (2025). Pengaturan Penggunaan Kecerdasan Buatan dalam Tugas Profesional Hakim di Indonesia. *Lex Privatum*, 15(2), 1–12.
- Sartor, G. (2009). *Judicial applications of artificial intelligence*. <https://doi.org/10.1007/978-94-015-9010-5>
- Setiawan, R., Tohari, M., & Alid, I. H. (2026). Risiko “Double Victimization” pada Korban Kekerasan Seksual Berbasis Elektronik (KSBE) dalam Penegakan Hukum Deepfake Pornografi. *Lex Journal: Kajian Hukum Dan Keadilan*, 10(2), 342–357. <https://doi.org/10.25139/lex.v10i2.11849>
- The Leadership Conference on Civil and Human Rights. (2026). *Disparate Impact as Uniquely Relevant in the Age of AI*. <https://civilrights.org/disparate-impact-age-of-ai/>
- Wijanarko, S., Ekaningsih, L., & Farida, A. (2026). Konstruksi Pertanggungjawaban Pidana Korporasi Pengembang Kecerdasan Buatan (AI) terhadap Tindak Pidana Pemalsuan Dokumen Elektronik. *Lex Journal: Kajian Hukum Dan Keadilan*, 10(2), 329–341. <https://doi.org/10.25139/lex.v10i2.11847>