Jurnal Riset Multidisiplin Edukasi
https://journal.hasbaedukasi.co.id/index.php/jurmie
Halaman : 1724 - 1744

KRISIS MENS REA DALAM KEJAHATAN DEEPFAKE BERBASIS ARTIFICIAL INTELLIGENCE : ANALISIS PERTANGGUNGJAWABAN PIDANA

Elisabeth Hana Gracesoldy Sihombing
Fakultas Hukum Universitas Pembangunan Nasional "Veteran" Jakarta
Email: 2410611072@mahasiswa.upnvj.ac.id

Informasi

Volume : 3
Nomor : 6
Bulan : Juni
Tahun : 2026
E-ISSN : 3062-9624

Abstract

A new type of cybercrime known as "deepfakes," which has the ability to manipulate video and audio content in a nearly lifelike manner, has emerged as a result of advances in artificial intelligence (AI) technology. On the other hand, this raises controversy with conventional criminal law doctrine, which stipulates that there must be an element of criminal intent (mens rea) on the part of the individual. The purpose of this study is to identify changes in the concept of mens rea resulting from the autonomization of AI, examine the issue of criminal liability for deepfakes under current Indonesian positive law, and propose an ideal framework for reconstructing the principle of fault for the future. The research method used in this article is normative legal analysis or a literature review, employing a statutory approach to examine relevant legislation, a conceptual approach, and a general case-based approach, with the analysis of legal materials conducted through qualitative normative reasoning using deductive reasoning. The research results show that the autonomous features and black-box nature of artificial intelligence create a disconnect between human subjective intent and the machine's automated outputs. This leads to a conceptual crisis and a multi-actor liability dilemma among developers, users, and distributors of manipulative content. Furthermore, Indonesia's current cyber regulations remain highly reactive and lack formal standards for attributing culpability in AI-generated crimes. Therefore, this article concludes by proposing a reconstruction of the principle of culpability through the application of a risk-based approach that shifts the paradigm from proving individual, anthropocentric intent toward an objective assessment of systemic risk. Thus, to ensure certain and fair digital law enforcement, progressive national criminal law reform and the standardization of digital forensics are necessary.

Keyword: *Artificial Intelligence, Deepfakes, Mens Rea, Criminal Liability*

Abstrak

Jenis kejahatan siber baru yang dikenal sebagai deepfake, yang memiliki kemampuan untuk merekayasa konten video dan audio dengan cara yang hampir nyata, telah muncul sebagai hasil dari kemajuan teknologi kecerdasan buatan atau AI. Di sisi lain, hal ini menimbulkan kontroversi dengan doktrin hukum pidana konvensional yang menetapkan bahwa ada elemen niat jahat dalam diri seseorang (mens rea). Tujuan penelitian ini adalah untuk menemukan perubahan dalam pengertian mens rea sebagai akibat dari otonomisasi AI, memeriksa masalah pertanggungjawban pidana deepfake dalam hukum positif Indonesia saat ini, dan menawarkan rumus rekonstruksi asas kesalahan yang ideal untuk masa depan. Metode penelitian yang digunakan dalam penulisan artikel ini adalah yuridis normatif atau kepustakaan dengan menggunakan pendekatan perundang-undangan menelaah peraturan perundang-undangan yang relevan/berkaitan, pendekatan konseptual, serta pendekatan kasus secara umum, yang mana analisis bahan hukumnya dilakukan secara normatif kualitatif melalui penalaran deduktif. Hasil penelitian menunjukkan bahwa fitur otonom dan sistem black box kecerdasan buatan menyebabkan bagian yang terpisah antara niat subjektif manusia dan luaran otomatis mesin. Ini

menyebabkan krisis konseptual dan dilema multi-actor liability antara pengembang (developer), pengguna (user), dan penyebar (distributor) konten manipulatif. Selain itu, regulasi siber di Indonesia saat ini masih sangat reaktif dan belum memiliki standar atribusi kesalahan formal terhadp AI-generated crime.Oleh karena itu, kesimpulan artikel ini menawarkan rekonstruksi asas kesalahan (culpabilitas) melalui penerapan pendekatan berbasis risiko (risk-based approach) yang menggeser paradigma dari pembuktian niat individual antroposentris menuju penilaian risiko sistemik yang objektif. Dengan demikian, untuk memastikan penegakan hukum digital yang berkepastian dan tetap berkeadilan, diperlukan reformasi hukum pidana nasional yang progresif dan standardisasi forensik digital.

Kata Kunci: *Kecerdasan Buatan, Deepfake, Mens Rea, Pertanggungjawaban Pidana*

A. PENDAHULUAN

Saat ini, perkembangan teknologi informasi berkembang dengan sangat cepat. Salah satu faktor pendorong utamanya adalah kecerdasan buatan atau *Artificial Intelligence* (AI). Teknologi ini beroperasi dengan basis machine learning dan deep learning, yang memiliki kemampuan untuk memproses data dalam skala besar. Sistem ini dapat mengenali pola secara mandiri dengan meniru cara otak manusia bekerja. Belakangan inii, kemampuan generatif AI telah berkembang dengan pesat. Teknologi ini sekarang sudah mampu membuat konten digital baru secara mandiri, bukan hanya untuk menganalisis data statis. Konten digital terdiri dari gambar, video, teks, dan audio yang terlihat sangat realistis. *Deepfake* adalah salah satu produk nyata dari kemajuan teknologi ini.[1] Sayangnya, kemampuan alat produksi konten digital ini sering melampaui kemampuan instrumen hukum pidana yang saat ini kita miliki.

Secara teknis, *deepfake* adalah teknik rekayasa atau manipulasi konten digital berbasis kecerdasan buatan yang menggabungkan, mengubah, dan menyunting video, gambar, atau audio asli untuk membuat konten tiruan yang sangat mirip dengan apa yang sebenarnya terjadi. Di era sekarang, *deepfake* telah berkembang dari hiburan visual semata menjadi fenomena kejahatan digital baru. Para pelaku kejahatan/kriminal memanfaatkan teknologi ini untuk melakukan berbagai tindakan ilegal.[2] Pencemaran nama baik adalah contoh nyata yang sering muncul. Untuk menjatuhkan reputasi sosial mereka, pelaku menggunakan suara atau wajah seseorang atau lebih seringnya adalah tokoh publik. Selain itu, tren penipuan finansial berskala besar pun juga muncul, dimana hal ini menjadi salah satu metode penipuan terbaru dan untuk orang yang tidak begitu mengerti teknologi, mereka akan menjadi sasaran mudah

[1] Muhammad Syafiq Wafi, Aloysius Wisnubroto, dan Yudi Prayudi. "Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana". *Jurnal Reformasi Hukum (JRH)*. Vol.29. No.2. 2025. Hal. 168-183. doi.org/10.46257/jrh.v29i2.1304

[2] Desty Aster Yansen Basah, Andika Wijaya, dan Ivans Januardy. "Kriminalisasi Pelanggaran Protokol Digital: Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial". *Journal Of Social Science Research*. Vol. 5. No.4. 2025. Hal. 386-398.https://j-innovative.org/index.php/Innovative.

dari tindakan penipuan ini. Untuk meminta transfer dana, pelaku biasanya menduplikasi wajah atau suara kerabat korban. Tak hanya itu, ada pun fenomena lainnya seperti penyebaran konten pornografi non-konsensual yang menjadi kasus yang harus diperhatikan dan dicegah/ditangani.[3] Fenomena ini berkaitan dengan tindakan pelaku yang tanpa izin menggunakan dan mengambil foto wajah korban secara digital dan menjadikannya sebagai konten asusila. Kejahatan *deepfake* memiliki fitur yang sangat mengancam. Sulit bagi masyarakat umum untuk memastikan bahwa konten ini asli. Sebaliknya, pelaku memiliki kemampuan untuk segera menyebarkan konten manipulatif melalui media sosial, yang merusak reputasi korban.

Fenomena ini menimbulkan konflik yang mendalam dan memicu benturan keras dengan doktrin hukum pidana klasik di Indonesia. Untuk menjatuhkan sanksi pidana, sistem hukum pidana materiil memerlukan unsur mens rea atau niat jahat. Prinsip dasar hukum pidana menegaskan bahwa tidak ada pidana tanpa kesalahan (asas culpabilitas) atau tanggung jawab. Dilansir dari Jurnal Reformasi Hukum, sistem AI tidak memiliki kesadaran moral, kehendak bebas, atau niat jahat yang independen seperti manusia, yang menyebabkan masalah fundamental.[4] Kitab Undang-Undang Hukum Pidana (KUHP) kita menuntut adanya hubungan batin yang sadar antara pelaku dengan perbuatan pidananya.Hukum menghadapi jalan buntu ketika sebuah program AI secara otomatis menghasilkan hasil yang merugikan orang lain (jika tidak ditemukan seseorang yang menggunakannya). Akan sulit untuk tidak dapat membandingkan algoritma dengan subjek hukum manusia yang memiliki kehendak batin.

Kondisi ini menyebabkan krisis konseptual yang sangat mendalam dalam hal pertanggungan pidana. Pelaku fisik, juga dikenal sebagai dader, tidak selalu melakukan perbuatan pidana secara langsung atau linier seperti dalam delik konvensional di dunia internet saat ini. Pengguna teknologi hanya memberikan perintah awal yang sifatnya umum. Selain itu, proses black box yang buram memungkinkan algoritma deep learning untuk bekerja secara mandiri. AI acap kali menghasilkan output yang merugikan yang tidak sepenuhnya

[3] Ni Putu Martina Putri, Made Sugi Hartono, dan I Dewa Gede Herman Yudiawan. "Analisis Reformulasi Pertanggungjawaban Pidana Pengguna Teknologi Deepfake Dalam Tindak Pidana Pencemaran nama Baik Berbasis Artificial Intelligence". *Jurnal Pacta Sunt Servanda.* Vol. 5. No. 4. https://ejournal2.undiksha.ac.id/index.php/JPSS

[4] Muhammad Syafiq Wafi, Aloysius Wisnubroto, dan Yudi Prayudi. "Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana". *Jurnal Reformasi Hukum (JRH)*. Vol.29. No.2. 2025. Hal. 168-183. doi.org/10.46257/jrh.v29i2.1304

dikendalikan atau diantisipasi oleh pengguna manusia[5]. Ini menimbulkan ambiguitas hukum yang signifikan. Penegak hukum akan kesulitan menentukan siapa yang sebenarnya memiliki "niat" jahat tersebut ketika sistem menghasilkan deviasi output yang ekstrem. Apakah kesalahan itu berasal dari pengembang perangkat lunak yang menulis kode, pengguna yang mengunggah bahan mentah, atau bahkan sistem otonom itu sendiri? Batas-batas pertanggungjawaban pidana konvensional menjadi kabur karena ketidakjelasan ini.

Selanjutnya, kesenjangan regulasi yang nyata dalam hukum positif Indonesia saat ini memperparah krisis konseptual tersebut. Kitab Undang-Undang Hukum Pidana Baru, yaitu Undang-Undang Nomor 1 Tahun 2023 tentang KUHP, masih membatasi subjek hukum pidana hanya pada manusia dan korporasi saja. Pasal 2 dan Pasal 45 UU 1/2023 sama sekali tidak menyinggung subjek kecerdasan buatan sebagai pelaku independen[6]. Di sisi lain, Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE) juga belum memuat definisi teknis yang spesifik mengenai *deepfake*.[7] Pasal 27A UU ITE 2024 memang melarang penyerangan kehormatan atau nama baik seseorang, dan Pasal 28 ayat (1) mengatur penyebaran berita bohong[8]. Namun, rumusan pasal-pasal tersebut masih bersifat sangat umum dan reaktif. Hukum kita belum memiliki standar atribusi kesalahan yang jelas ketika kejahatan tersebut melibatkan kontribusi otonom dari kecerdasan buatan.[9] Kekosongan hukum (*legal vacuum*) ini menyebabkan ketidakpastian hukum dan melanggar asas *lex certa* dalam hukum pidana.

Beberapa peneliti terdahulu telah mengkaji masalah teknologi ini dari berbagai sudut pandang hukum. Sebagai contoh, penelitian oleh Putri, Hartono, dan Yudiawan (2024) yang berfokus pada reformulasi hukum terkait pencemaran nama baik akibat *deepfake*. Sementara itu, Basah, Wijaya, dan Januardy (2025) melihat masalah ini dari kacamata pelanggaran protokol digital dan perlunya kriminalisasi penyebaran konten di media sosial. Di sisi lain,

---

[5] Siti Nurkholisah, Daud Rismana, Afrizal Eko Nugroho, dkk. "Tantangan Kriminalisasi Deepfake dalam Hukum Pidana Indonesia". *Jurnal USM Law Review*. Vol. 8. No. 3. 2025. Hal. 2421-2445. https://doi.org/10.26623/julr.v8i3.13060

[6] Indonesia. (2023). *Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana*. Lembaran Negara Republik Indonesia Tahun 2023 Nomor 1. Jakarta: Sekretariat Negara.

[7] Siti Nurkholisah, Daud Rismana, Afrizal Eko Nugroho, dkk. "Tantangan Kriminalisasi Deepfake dalam Hukum Pidana Indonesia". *Jurnal USM Law Review*. Vol. 8. No. 3. 2025. Hal. 2421-2445. https://doi.org/10.26623/julr.v8i3.13060

[8] Indonesia. (2024). *Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik*. Lembaran Negara Republik Indonesia Tahun 2024 Nomor 1. Jakarta: Sekretariat Negara.

[9] Desty Aster Yansen Basah, Andika Wijaya, dan Ivans Januardy. "Kriminalisasi Pelanggaran Protokol Digital: Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial". *Journal Of Social Science Research*. Vol. 5. No.4. 2025. Hal. 386-398.https://j-innovative.org/index.php/Innovative.

peneliti dari Wafi, Wisnubroto, dan Prayudi (2025) menawarkan konsep perluasan asas culpabilitas melalui model *vicarious liability* yang diadopsi dari hukum korporasi untuk menjerat penyedia perangkat lunak. Nurkholisah et al. (2025) juga telah memetakan tantangan kriminalisasi ini dengan meninjau putusan pengadilan serta membandingkannya dengan regulasi internasional seperti *EU AI Act*. Meskipun penelitian-penelitian tersebut memberikan kontribusi berharga, mayoritas studi terdahulu hanya membahas kebijakan kriminalisasi secara makro, aspek forensik digital, atau perlindungan privasi korban secara umum. Peneliti terdahulu belum membedah secara mendalam disintegrasi elemen *mens rea* ketika berhadapan dengan ketidakpastian kendali output AI generatif. Di sinilah yang menjadi letak kesenjangan penelitian (*research gap*) yang mendasar dalam pembuatan artikel ini.

Penelitian ini akan melengkapi dan menyempurnakan penelitian sebelumnya. Penulis berkonsentrasi pada rekonstruksi konseptual elemen mens rea dalam konteks hukum pidana Indonesia yang bertransisi ketika KUHP Baru diterapkan. Penemuan baru yang ditawarkan oleh penelitian ini adalah bahwa itu menawarkan kerangka teoritis untuk mengukur batasan niat jahat manusia ketika menggunakan teknologi yang memiliki derajat otonomi tinggi. Hal ini penting agar aparat penegak hukum dapat menghindari dilema antara menghukum orang yang tidak melakukan kesalahan batin atau membiarkan kejahatan tanpa hukuman. Akibatnya, tulisan ini memiliki urgensi penting untuk menyelidiki dan mengkaji secara menyeluruh terkait krisis mens rea dalam kejahatan deepfake berbasis AI untuk mewujudkan sistem hukum pidana digital yang fleksibel/adaptif, berkepastian hukum, dan berkeadilan.

B. Rumusan Masalah

Teknologi AI generatif telah mengubah tatanan hukum pidana materiil melalui fenomena kejahatan deepfake. Selain menyebabkan kerugian yang signifikan bagi korban, visualisasi dan rekaman audio yang sangat realistis ini bertentangan dengan prinsip-prinsip dasar hukum pidana konvensional. Menurut prinsip utama kita, *geen straf zonder schuld*, ada hubungan batin berupa niat jahat (mens rea) yang disadari oleh pelaku manusia sebelum hukuman dijatuhkan. Namun, batasan kendali manusia atas output yang dihasilkan oleh sistem diabaikan oleh sifat otonom proses algoritma deep learning. Akibatnya, hukum positif Indonesia saat ini, baik dalam KUHP Baru maupun UU ITE, tidak memiliki konsep yang memadai dan tidak memiliki standar untuk menentukan kesalahan. Penegakan hukum pidana digital terancam kehilangan arah kepastian hukum jika tidak ada kejelasan tentang siapa yang harus memikul kesalahan atas tindakan manipulatif yang digerakkan oleh algoritma. Berdasarkan latar belakang masalah tersebut, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana konsep *mens rea* dalam doktrin hukum pidana dalam menghadapi tantangan fundamental akibat berkembangnya otonomisasi teknologi di era kecerdasan buatan (*Artificial Intelligence*)?
2. Bagaimana karakteristik kejahatan *deepfake* berbasis kecerdasan buatan dapat memicu problematika yuridis dalam penentuan pertanggungjawaban pidana berdasarkan hukum positif saat ini?
3. Bagaimana kira-kira formula rekonstruksi asas kesalahan (*culpabilitas*) yang ideal dan adaptif untuk mengatasi krisis *mens rea* dalam kejahatan *deepfake* berbasis kecerdasan buatan di masa depan?

C. Tujuan Penelitian

Secara umum, penelitian ini bertujuan untuk mengurai kebuntuan doktrinal dan juga untuk memberikan kejelasan arah mengenai bagaimana hukum pidana materiil seharusnya merespons kejahatan siber berbasis kecerdasan buatan seperti ini secara adil. Maka secara lebih spesifik, tujuan yang ingin dicapai melalui penulisan artikel ilmiah ini adalah:

1. Untuk mengidentifikasi dan menganalisis pergeseran konsep *mens rea* dalam doktrin hukum pidana, serta memetakan berbagai tantangan teoritis yang muncul akibat adanya otonomisasi sistem kecerdasan buatan (terutama untuk fenomena *deepfake*).
2. Untuk mendeskripsikan karakteristik khusus dari fenomena kejahatan siber yakni *deepfake* berbasis kecerdasan buatan sekaligus membedah problematika penentuan subjek dan pertanggungjawaban pidana dalam regulasi hukum positif di Indonesia saat ini.
3. Untuk merumuskan dan menawarkan konsep rekonstruksi dari asas kesalahan (*culpabilitas*) yang lebih kontekstual sebagai solusi alternatif dalam menyelesaikan krisis pembuktian *mens rea* pada tindak pidana *deepfake* di masa mendatang.

B. METODE PENELITIAN

Dalam rangka menjawab dan mengkaji fenomena dalam artikel ini, maka jenis penelitian yang digunakan adalah hukum yuridis normatif atau penelitian kepustakaan, yang mana fokus utamanya adalah mengkaji norma-norma hukum tertulis terkait krisis *mens rea* dalam kejahatan *deepfake* berbasis AI. Pendekatan yang digunakan penulis mencakup tiga hal, yaitu pendekatan perundang-undangan (*statute approach*) dengan menelaah UU No. 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana (KUHP Baru) dan UU No. 1 Tahun 2024 tentang Perubahan Kedua UU ITE, pendekatan konseptual (*conceptual approach*) untuk membedah

doktrin asas kesalahan, serta pendekatan kasus (*case approach*) secara umum guna melihat fenomena nyata penyalahgunaan daripada *deepfake*. Bahan hukum yang digunakan terbagi menjadi bahan hukum primer berupa undang-undang yang sudah disebutkan tadi, bahan hukum sekunder seperti jurnal-jurnal hukum pidana tentang *Artificial Intelligence* dan buku yang relevan mengenai pertangungjawaban pidana.

Kemudian, untuk teknik pengumpulan bahan/sumber dalam penelitian ini dilakukan melalui studi dokumen (*documentary study*) dengan cara menelusuri literatur tertulis baik cetak maupun digital melalui portal jurnal hukum online dan situs resmi JDIH (dalam hal mencari peraturan perundang-undangan) untuk memastikan kevalidan draf pasal. Bahan-bahan hukum yang sudah terkumpul dengan baik tersebut kemudian dianalisis menggunakan metode normatif kualitatif, yang berarti analisisnya bersifat deskriptif analitiss. Untuk dalam hal menganalisis metode yang digunakan adalah penalaran deduktif, di mana penulis menarik kesimpulan dari hal-hal yang sifatnya umum menuju ke hal yang bersifat khusus untuk menjawab bagaimana problematika pertanggungjawaban hukum kejahatan *deepfake* berbasis kecerdasan buatan di Indonesia saat ini agar didapatkan hasil yang komprehensif.

C. HASIL DAN PEMBAHASAN

A. Konsep Mens Rea dalam Hukum Pidana dan Tantangannya dalam Era AI

Dalam ranah hukum pidana doktrin mengenai mens rea atau sikap batin jahat merupakan salah satu penyangga utama yang tidak bisa dipisahkan dari keseluruhan sistem pertangungjawaban pidana. Hukum pidana secara historis mempertimbangkan tidak hanya tindakan pidana secara objektif (*actus reus*), tetapi juga memasuki wilayah subjektif pelaku, yaitu pikiran batiniahnya saat melakukan tindakan pidana tersebut. Salah satu asas hukum yang terkenal, "*Actus non facit reum nisi mens sit rea*", menyatakan bahwa suatu tindakan tidak membuat seseorang bersalah kecuali pikirannya juga bersalah. Konsep mens rea ini berkembang menjadi tiga komponen dasar. Pertama, niat atau kesengajaan (*dolus*), di mana pelaku benar-benar menginginkan perbuatan tersebut dan menginginkan akibat yang dilarang darinya. Kedua, kelalaian atau kelalaian (*culpa*), di mana pelaku sebenarnya tidak menginginkan akibat tersebut tetapi terjadi karena kurangnya kehati-hatian atau kecerobohan yang seharusnya tidak terjadi. Selain kedua elemen itu, sistem hukum juga mengakui tingkat kesembronoan, juga dikenal sebagai *recklessness*. Tingkat kesembronoan adalah ketika seseorang secara sadar mengambil risiko yang sangat besar secara tidak wajar meskipun ia

menyadari potensi bahaya yang dapat merugikan orang lain.[10] Salah satu tujuan utama kehadiran mens rea ini adalah untuk memastikan apakah pelaku melakukan kesalahan. Kesalahan inilah yang menjadi dasar pemidanaan. Jika tidak ada pembuktian tentang sikap batin ini, hukum pidana akan kehilangan moralitasnya karena menghukum orang tanpa mempertimbangkan niat atau kesadaran mereka.

Asas culpabilitas, atau dogma *geen straf zonder schuld* tiada pidana tanpa kesalahan diadopsi secara ketat di Indonesia sebagai dasar pembuktian subjektif. Seperti yang ditunjukkan oleh Pasal 36 Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana (KUHP Baru), disini asas kesalahan ini telah dijelaskan secara eksplisit dalam undang-undang terbaru, yang menetapkan bahwa setiap orang hanya dapat dipertanggungjawabkan secara pidana atas tindak pidana yang dilakukannya jika orang tersebut mempunyai kesalahan. Pasal 36 UU 1/2023 ini menegaskan bahwa hukum pidana materiil kita menolak adanya pertanggungjawaban mutlak (*strict liability*) bagi subjek hukum perorangan dalam delik-delik konvensional, sehingga pembuktian unsur kesengajaan atau kealpaan tetap menjadi syarat mutlak yang sifatnya absolut yang harus dipenuhi oleh penegak hukum sebelum menjatuhkan sanksi pidana. Masalahnya adalah doktrin kesalahan dan rumusan pasal dalam KUHP Baru didasarkan pada keyakinan/asumsi antroposentris yang kuat. Ini berarti bahwa hukum pidana kita dapat dikatakan dimaksudkan hanya untuk mengevaluasi isi alam pikiran manusia, yang dianggap sebagai subjek hukum dengan moralitas, memiliki akal budi, dan juga kehendak bebas (*free will*), untuk memutuskan mana tindakan yang baik dan mana yang buruk.

Namun, ketika perkembangan kecerdasan buatan (AI) meningkat pesat dari 2021 hingga 2026, asumsi hukum yang berpusat pada manusia bisa menjadi kurang kuat jika berhadapan dengan kasus-kasus seperti ini. Kemampuan untuk mendeteksi pola, belajar secara mandiri dari sekumpulan data besar (*big data*), bahkan membuat keputusan atau membuat konten baru secara otonom tanpa intervensi manusia adalah karakteristik utama dari AI moderen yang berbasis pada algoritma machine learning dan deep learning.[11] Di sini, masalah yang perlu dilihat adalah teknologi kecerdasan buatan sama sekali tidak memiliki kehendak bebas, moral,

---

[10] Ni Putu Martina Putri, Made Sugi Hartono, dan I Dewa Gede Herman Yudiawan. "Analisis Reformulasi Pertanggungjawaban Pidana Pengguna Teknologi Deepfake Dalam Tindak Pidana Pencemaran nama Baik Berbasis Artificial Intelligence". *Jurnal Pacta Sunt Servanda.* Vol. 5. No. 4. https://ejournal2.undiksha.ac.id/index.php/JPSS

[11] Muhammad Syafiq Wafi, Aloysius Wisnubroto, dan Yudi Prayudi. "Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana". *Jurnal Reformasi Hukum (JRH)*. Vol.29. No.2. 2025. Hal. 168-183. doi.org/10.46257/jrh.v29i2.1304

atau perasaan seperti manusia. Sebaliknya, itu hanyalah sekumpulan baris kode matematis yang bekerja berdasarkan kemungkinan data yang dimasukkan. Lebih jauh lagi, sistem kecerdasan buatan masa kini kerap dikategorikan sebagai *black box system*. Istilah "*black box*" ini mengacu pada fenomena di mana proses penalaran internal dan keputusan yang diambil oleh jaringan saraf tiruan (*neural network*) di dalam komputer menjadi amat kompleks dan tidak jelas, bahkan dari pembuat program atau insinyur yang membuat AI itu sendiri tidak dapat melacak ataupun memahami secara menyeluruh proses pembentukan output.[12] Akibatnya, terdapat perbedaan yang besar antara tindakan yang dilakukan oleh mesin dan tindakan yang dilakukan langsung oleh manusia nya.

Kombinasi antara sifat otonomisasi AI dan ketidakjelasan dari black box system inilah yang memicu terjadinya ketidaksesuaian konseptual yang mendalam dengan doktrin mens rea. Bagaimana mungkin kita dapat mengaitkan niat jahat, kelalaian, atau kesembronoan dengan alat digital yang tidak memiliki jiwa atau kesadaran psikologis? Secara teoritis, mens rea adalah konsep murni kemanusiaan yang membutuhkan keterikatan batin moral. Oleh karena itu, menyematkan mens rea secara langsung pada baris algoritma komputer merupakan sebuah kemustahilan secara konseptual.[13] Sesuai dengan Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua atas UU ITE, komputer atau sistem elektronik hanyalah alat mati (instrumentum) yang digerakkan secara mekanis dan linier oleh penggunanya amun ketika AI generatif dilepas di ruang digital dan menghasilkan luaran (*output*) destruktif seperti *deepfake* yang tidak pernah diprediksi sebelumnya oleh si pembuat program, maka terjadilah krisis hebat dalam doktrin kesalahan.[14] Karenanya hukum pidana dapat kehilangan dasar pembuktian subjektifnya karena terdapat bagian yang terputus antara niat awal si pengguna dengan tindakan nyata yang diwujudkan secara mandiri oleh algoritma mesin tersebut.

Kondisi hukum yang kurang realistis ini menghadapkan para praktisi dan akademisi hukum pada sebuah tantangan yang belum pernah terjadi sebelumnya. Beberapa penelitian terdahulu dalam kurun waktu belakangan ini telah mencoba memetakan persoalan ini dari berbagai sudut. Misalnya, studi dari Jurnal *Pacta Sunt Servanda*, mengidentifikasi bahwa

---

[12] Siti Nurkholisah, Daud Rismana, Afrizal Eko Nugroho, dkk. "Tantangan Kriminalisasi Deepfake dalam Hukum Pidana Indonesia". *Jurnal USM Law Review*. Vol. 8. No. 3. 2025. Hal. 2421-2445. https://doi.org/10.26623/julr.v8i3.13060

[13] Muhammad Syafiq Wafi, Aloysius Wisnubroto, dan Yudi Prayudi. "Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana". *Jurnal Reformasi Hukum (JRH)*. Vol.29. No.2. 2025. Hal. 168-183. doi.org/10.46257/jrh.v29i2.1304

[14] Desty Aster Yansen Basah, Andika Wijaya, dan Ivans Januardy. "Kriminalisasi Pelanggaran Protokol Digital: Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial". *Journal Of Social Science Research*. Vol. 5. No.4. 2025. Hal. 386-398.https://j-innovative.org/index.php/Innovative.

kesulitan terbesar dalam kejahatan berbasis AI terletak pada pembuktiannya, yakni apakah pengguna komputer memiliki kepastian akan pengetahuan (*scienter*) terhadap deviasi *output* yang dihasilkan oleh mesin.[15] Di sisi lain, dilansir dari Jurnal Reformasi Hukum, penulis melihat adanya kekuatan bukti bahwa asas culpabilitas klasik di Indonesia harus segera direformasi dengan meminjam konsep hukum korporasi agar hukum tidak mengalami kelumpuhan total saat menghadapi subjek non-manusia.[16] Namun, keterbatasan utama dari mayoritas literatur hukum pidana nasional saat ini sekaligus menjadi keterbatasan dalam tinjauan penulis adalah adanya risiko bias komparatif, di mana solusi yang ditawarkan sering kali terlalu memaksakan konsep *electronic personhood* barat yang jelas-jelas bertabrakan dengan struktur hukum positif Pasal 36 KUHP Baru kita yang belum mengakui subjek elektronik seperti AI sebagai subjek hukum pidana.[17] Implikasi dari krisis konseptual ini cukup sulit untuk di pecahkan bagi banyak orang misal nya aparat penegak hukum dan pencari keadilan, lalu hakim akan dipaksa melakukan pilihan sulit antara menerapkan analogi hukum yang dilarang dalam pidana, menghukum pengguna manusia secara tidak adil dengan asas pertanggungjawaban mutlak, atau terpaksa membebaskan pelaku kejahatan siber karena gagal membuktikan elemen batin kesengajaan manusianya di persidangan.

B. *Deepfake* sebagai Kejahatan Siber Berbasis AI dan Problematika Pertanggungjawaban Pidana

Memasuki pembahasan tentang anatomi kejahatannya, teknologi deepfake secara teknis beroperasi melalui mekanisme siber yang sangat canggih yang dikenal sebagai *Generative Adversarial Network* (GAN). GAN terdiri dari dua jaringan saraf tiruan yang saling bertarung, generator yang membuat konten tiruan dan discriminator yang menilai keaslian konten. Kecanggihan pemrosesan gambar ini kemudian muncul dalam dunia siber menjadi berbagai jenis tindak pidana baru yang destruktif.[18] Penipuan digital, juga dikenal sebagai *digital fraud*,

---

[15] Ni Putu Martina Putri, Made Sugi Hartono, dan I Dewa Gede Herman Yudiawan. "Analisis Reformulasi Pertanggungjawaban Pidana Pengguna Teknologi Deepfake Dalam Tindak Pidana Pencemaran nama Baik Berbasis Artificial Intelligence". *Jurnal Pacta Sunt Servanda.* Vol. 5. No. 4. https://ejournal2.undiksha.ac.id/index.php/JPSS

[16] Muhammad Syafiq Wafi, Aloysius Wisnubroto, dan Yudi Prayudi. "Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana". *Jurnal Reformasi Hukum (JRH)*. Vol.29. No.2. 2025. Hal. 168-183. doi.org/10.46257/jrh.v29i2.1304

[17] Siti Nurkholisah, Daud Rismana, Afrizal Eko Nugroho, dkk. "Tantangan Kriminalisasi Deepfake dalam Hukum Pidana Indonesia". *Jurnal USM Law Review*. Vol. 8. No. 3. 2025. Hal. 2421-2445. https://doi.org/10.26623/julr.v8i3.13060

[18] Ni Putu Martina Putri, Made Sugi Hartono, dan I Dewa Gede Herman Yudiawan. "Analisis Reformulasi Pertanggungjawaban Pidana Pengguna Teknologi Deepfake Dalam Tindak Pidana Pencemaran nama Baik Berbasis Artificial Intelligence". *Jurnal Pacta Sunt Servanda.* Vol. 5. No. 4. https://ejournal2.undiksha.ac.id/index.php/JPSS

adalah jenis tindak pidana yang paling umum terjadi di masyarakat saat ini. Hal tersebut dengan cara-cara yang terdiri dari meniru suara atau video orang penting untuk memaksa korban mengirimkan uang. Selain penipuan, ada jenis kriminal lainnya, seperti pencemaran nama baik secara digital yang merusak reputasi sosial korban, dan juga eksploitasi seksual melalui pembuatan pornografi non-konsensual, dengan merekayasa wajah korban ke dalam konten asusila tanpa persetujuan mereka.[19] Berbagai bentuk manifestasi delik digital ini menunjukkan bahwa deepfake sekarang bukan lagi sekadar komoditas hiburan melainkan sebuah ancaman nyata terhadap hak-hak hukum perorangan yang dilindungi oleh hukum pidana.

Kejahatan *deepfake* sangat ganas, terlepas dari fitur siber yang dibawa oleh teknologi AI, yang membedakannya dari kejahatan konvensional di dunia nyata. Karakteristik pertama yang paling menyulitkan hukum adalah adanya tingkat anonimitas pelaku yang sangat tinggi, di mana pelaku dapat dengan mudah menyembunyikan identitas aslinya di balik jaringan dark web atau akun-akun palsu yang dibuat secara otomatis oleh bot.[20] Selain anonim, kejahatan siber berbasis kecerdasan buatan generatif ini sangat skalable, yang berarti konten manipulatif yang dibuat dalam hitungan menit dapat disalin dan didistribusikan ke jutaan pengguna internet dalam waktu yang sangat singkat. Karakteristik ini bertentangan dengan dampak luas yang ditimbulkannya, yang dapat menyebabkan kerusakan psikologis dan penghancuran martabat sosial korban secara abadi. Ini karena jejak digital yang tertinggal di internet sangat sulit untuk dihapus sepenuhnya. Lebih lanjut, antara tahun 2023 dan 2026, tren data dari berbagai lembaga keamanan siber di seluruh dunia menunjukkan peningkatan kasus penipuan berbasis manipulasi audio dan video yang merugikan sektor perbankan hingga jutaan dolar.[21]

Adanya karakteristik teknologi otonom tersebut pada akhirnya menimbulkan seuah problematika pertanggungjawaban pidana yang sangat membingungkan ketika dihadapkan pada subjek hukum pidana. Pertanyaan mendasar yang selalu muncul dan belum terjawab secara tuntas dalam ruang persidangan adalah, siapakah pelaku kejahatan yang sesungguhnya yang harus dimintai pertanggungjawaban hukum? Mengingat rantai penciptaan konten

---

[19] Desty Aster Yansen Basah, Andika Wijaya, dan Ivans Januardy. "Kriminalisasi Pelanggaran Protokol Digital: Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial". *Journal Of Social Science Research*. Vol. 5. No.4. 2025. Hal. 386-398.https://j-innovative.org/index.php/Innovative.

[20] Siti Nurkholisah, Daud Rismana, Afrizal Eko Nugroho, dkk. "Tantangan Kriminalisasi Deepfake dalam Hukum Pidana Indonesia". *Jurnal USM Law Review*. Vol. 8. No. 3. 2025. Hal. 2421-2445. https://doi.org/10.26623/julr.v8i3.13060

[21] Muhammad Syafiq Wafi, Aloysius Wisnubroto, dan Yudi Prayudi. "Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana". *Jurnal Reformasi Hukum (JRH)*. Vol.29. No.2. 2025. Hal. 168-183. doi.org/10.46257/jrh.v29i2.1304

manipulatif ini melibatkan banyak pihak dan hukum pidana terjebak dalam pusaran *multi-actor liability* yang melibatkan pembuat AI (*developer*) yang merancang baris kode algoritma dasar, pengguna (*user*) yang memberikan instruksi spesifik untuk memanipulasi wajah korban, serta penyebar (*distributor*) yang membagikan konten merugikan tersebut di media sosial. Berdasarkan kerangka berpikir antroposentris dalam Pasal 36 UU No. 1 Tahun 2023 tentang KUHP, pertanggungjawaban pidana hanya bisa dilekatkan pada subjek hukum manusia yang memiliki kesalahan subjektif.[22] Namun, ketika pengembang perangkat lunak menciptakan sistem AI yang bersifat terbuka (*open-source*) dan kemudian disalahgunakan oleh pihak ketiga untuk membuat *deepfake*, maka pembagian porsi kesalahan di antara para aktor siber tersebut menjadi sangat kabur dan ambigu, sehingga menyulitkan jaksa penuntut umum untuk merumuskan dakwaan pertanggungjawaban pidana yang proporsional.

Saat penegak hukum mencoba membawa kasus ini ke ranah peradilan pidana, tingkat kesulitan pembuktian yang cukup tinggi membuat masalah ini menjadi rumit untuk diselesaikan. Seperti yang telah disebutkan sebelumnya, dalam delik siber berbasis kecerdasan buatan, akan sangat sulit bagi penegak hukum untuk membuktikan niat jahat (mens rea) dari pelaku secara langsung karena proses otomatisasi sistem komputasi terputus. Selain itu, melacak pelaku asli alias pelaku sebenarnya seringkali tidak berhasil karena penggunaan teknologi enkripsi berlapis, VPN, atau bahkan karena *software* AI itu sendiri beroperasi di luar yurisdiksi hukum Indonesia. Kompleksitas alat bukti digital, juga menjadi tantangan tersendiri bagi forensik karena sifat data digital yang mudah dimanipulasi. Dalam undang-undang hukum acara pidana, pembuktian video asli yang dikompresi berbeda dengan video yang dihasilkan dari rekayasa algoritma GAN. Pada akhirnya, tantangan teknis dan konseptual ini menempatkan kelompok kunci terutama korban kejahatan siber dalam posisi yang tidak menguntungkan. Laporan hukum mereka seringkali tidak dapat diproses karena kurang memenuhi standar pembuktian hukum pidana.

Salah satu faktor utama yang menyebabkan kelumpuhan penegakan hukum ini adalah kelemahan regulasi hukum positif di Indonesia, yang masih sangat tidak stabil dan tidak memiliki rencana masa depan. Instrumen hukum siber kita saat ini, yaitu Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua atas UU ITE, masih menggunakan paradigma hukum pidana siber konvensional yang belum secara spesifik mengatur mengenai *AI-generated crime* atau kejahatan yang dihasilkan secara otonom oleh mesin. Sebagai contoh, dalam Pasal

[22] Indonesia. (2023). *Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana*. Lembaran Negara Republik Indonesia Tahun 2023 Nomor 1. Jakarta: Sekretariat Negara.

27A UU ITE 2024 melarang tindakan yang melanggar kehormatan atau nama baik, dan Pasal 27 ayat (1) melarang penyebaran konten yang melanggar kesusilaan. Namun, rumusan pasal-pasal tersebut dibuat dengan asumsi bahwa alat yang digunakan hanyalah media transmisi biasa dan bukan entitas pintar yang dapat berubah sendiri.[23] Asas legalitas *lex certa dan lex stricta* juga dapat dilanggar jika ada kekosongan hukum formal mengenai standar atribusi kesalahan dalam konteks AI ini. Akibatnya, putusan hakim dapat menjadi bias jika penegak hukum terpaksa melakukan interpretasi yang dipaksakan untuk menghukum pelaku. Oleh karena itu, keterbatasan regulasi saat ini menunjukkan betapa pentingnya pembentuk undang-undang untuk merekonstruksi doktrinal terhadap asas kesalahan agar hukum pidana kita tidak tertinggal jauh dari kemajuan teknologi kecerdasan buatan.

C. Rekonstruksi Asas Kesalahan dalam Kejahatan Deepfake Berbasis AI

Kebutuhan untuk melakukan rekonstruksi terhadap asas kesalahan di dalam hukum pidana kita saat ini sudah menjadi hal yang harus diperhatikan menurut penulis. Doktrin terkait *mens rea* terbukti tidak lagi cukup untuk menjangkau kompleksitas kejahatan siber yang digerakkan oleh algoritma cerdas otonom seperti *deepfake*. Kita membutuhkan pendekatan baru yang revolusioner agar hukum tidak menjadi terhambat saat menghadapi dampak destruktif teknologi. Ada beberapa alternatif pendekatan pertanggungjawban pidana yang bisa menjadi di diskusikan sebagai jalan keluar dari krisis doktrinal ini. Pertama adalah skema *strict liability* atau pertanggungjawaban mutlak, di mana pelaku dapat langsung dihukum tanpa perlu membuktikan adanya elemen niat jahat batiniah sama sekali, asalkan perbuatan dan dampaknya sudah nyata terjadi (akan tetapi perlu pengaturan yang tepat agar tidak salah menjatuhkan hukuman dan perlu penyidikan yang ketat agar tersangka benar-benar adalah pelaku kejahatan).[24] Alternatif kedua adalah *vicarious liability* yang menggeser tanggung jawab pidana ke pihak lain yang memiliki kendali sistemik, misalnya perusahaan teknologi yang menjadi penyedia platform atau korporasi pengembang perangkat lunak AI tersebut.[25] Selain itu, kita juga bisa menerapkan pendekatan *negligence-based liability* yang menitikberatkan kesalahan pada kelalaian pengguna atau pengembang dalam mengawasi, memitigasi risiko, dan

[23] Desty Aster Yansen Basah, Andika Wijaya, dan Ivans Januardy. "Kriminalisasi Pelanggaran Protokol Digital: Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial". *Journal Of Social Science Research*. Vol. 5. No.4. 2025. Hal. 386-398.https://j-innovative.org/index.php/Innovative.

[24] Basya Radyananda, Faisal Abdaud, dan Gamlan Dagani. "Analisis Limitasi Pertanggungjawaban Pidana Dalam Kejahatan Manipulasi Digital di Indonesia". *Nomos: Jurnal Penelitian Ilmu Hukum*. Vol. 6, No. 1. 2026. 4. https://journal.actual-insight.com/index.php/nomos/article/download/4273/3406.

[25] Muhammad Syafiq Wafi, Aloysius Wisnubroto, dan Yudi Prayudi. "Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana". *Jurnal Reformasi Hukum (JRH)*. Vol.29. No.2. 2025. Hal. 168-183. doi.org/10.46257/jrh.v29i2.1304

menggunakan sistem AI secara serampangan. Penerapan ragam model pertanggungjawaban alternatif ini dapat menjadi langkah untuk menutup celah lolosnya pelaku kriminal siber dari jeratan hukum pidana.

Langkah rekonstruksi doktrin kesalahan ini menuntut adanya perluasan konseptual mengenai arti dari sebuah kesalahan hukum itu sendiri. Maka, kita harus mulai berani menggeser paradigma pertanggungjawaban hukum yang awalnya bersifat individual murni antroposentris menuju ke arah pertanggungjawaban yang bersifat sistemik (atau menjadi kekhususan tersendiri bagi kejahatan jenis seperti ini), serta mengubah tolok ukur dari pembuktian niat subjektif menjadi penilaian risiko objektif. Pendekatan berbasis risiko (*risk-based approach*) ini berfokus penuh pada kalkulasi potensi bahaya yang melekat pada suatu sistem kecerdasan buatan sebelum dan saat sistem tersebut dilepaskan ke ruang publik. Model pendekatan ini sebenarnya sudah mulai digunakan secara luas di dalam regulasi modern internasional, salah satu contoh konkritnya adalah *European Union Artificial Intelligence Act* atau UU AI Uni Eropa (Regulation (EU) 2024/1689). Melalui regulasi tersebut, sistem kecerdasan buatan diklasifikasikan secara rigid berdasarkan tingkat kerentanan risikonya terhadp hak-hak fundamental warga negara.[26] Dengan memfokuskan hukum pada baagian pengelolaan dan mitigasi risiko teknis, hukum pidana dapat menilai sejauh mana aktor manusia di balik sistem tersebut abai dalam mengelola tingkat bahaya teknologi yang mereka operasikan.

Di tengah upaya rekonstruksi doktrin ini, wacana mengenai kemungkinan menempatkan Artificial Intelligence sebagai subjek hukum mandiri atau yang biasa kita kenal dengan istilah *electronic personhood* masih menjadi perdebatan terutama di kalangan akademisi. Orang-orang yang mendukung mengatakan bahwa dengan memberikan status subjek hukum kepada AI akan membuat penuntutan kejahatan deepfake yang dilakukan secara mandiri lebih mudah. Sebaliknya, *electronic personhood* dianggap sebagai konsep yang sangat kontroversial secara teoretis dalam hukum pidana oleh mereka yang menentangnya. Namun, bagaimanapun juga kecerdasan buatan tidak memiliki kesadaran moral, rasa bersalah, dan kemampuan untuk memahami nilai sanksi pidana penjara, sehingga rasanya tidak rasional untuk menyamakan kecerdasan buatan dengan subjek hukum perorangan. Perdebatan yang belum tuntas ini menunjukkan bahwa memaksakan AI sebagai subjek hukum independen di pengadilan pidana

[26] Ronit Justo-Hanani. "*Risk-based approach to EU AI act: benefits and challenges of co-regulation*". *Policy Design and Practice.* Vol. 9. No. 2. 2026. 173, https://doi.org/10.1080/25741292.2025.2610869.

saat ini berpotensi menimbulkan kekacauan dogmatis yang baru, sehingga perluasan asas kesalahan manusia yang mengendalikan teknologi dinilai jauh lebih realistis untuk diterapkan.

Semua perubahan teoritis dan rekonstruksi asas kesalahan ini penting bagi masa depan penegakan hukum Indonesia. Agar tidak teus-menrus bersifat reaktif terhadap kejahatan siber saat ini, reformasi hukum pidana nasional harus dilakukan secara bertahap.[27] Dibutuhkan perubahan substansial dalam peraturan positif, terutama untuk menyempurnakan batasan pertanggungjawaban pidana yang ditetapkan dalam UU No.1 Tahun 2023 tentang KUHP (Baru) dan UU No.1 Tahun 2024 tentang Perubahan Kedua UU ITE. Pasal 36 KUHP Baru yang mewajibkan adanya unsur kesalahan pada subjek hukum manusia harus mulai direformulasikan interpretasinya agar dapat mengakomodasi standar pertanggungjawaban berbasis risiko dan kelalaian digital (*digital negligence*) dalam penggunaan sistem otonom.[28] Jika tidak ada peraturan khusus yang mengatur kejahatan yang dihasilkan oleh kecerdasan buatan ini, penegak hukum kita akan terus mengalami kelumpuhan dogmatis dalam menangani manipulasi *deepfake*. Oleh karena itu, rekonstruksi asas culpabilitas ini menjadi jawaban atas solusi hukum yang diperlukan untuk menyelesaikan krisis mens rea dan memberikan perlindungan hukum yang berkeadilan bagi seluruh masyarakat Indonesia di era digital.

Nah, selain berfokus pada rekonstruksi asas kesalahan secara teoretis, penyelesaian krisis *mens rea* ini pada akhirnya akan sangat bergantung pada kemampuan nyata aparat penegak hukum di lapangan untuk mendeteksi serta menemukan siapa pelaku sesungguhnya di balik layar (*the real actor behind the screen*) yang membuat/menginput data awal dalam pembuatan *deepfake* diluar algoritma dari AI itu sendiri. Masalah fundamental dalam kejahatan *deepfake* berbasis AI seringkali bukan cuma soal mengaburnya konsep niat, tapi juga karena kegagalan teknis saat mengatribusikan tindakan otonom mesin kepada individu asli tertentu akibat tingginya tingkat anonimitas siber. Akan menjadi lebih sulit bagi jaksa penuntut umum bisa membuktikan adanya kesalahan batin (*schuld*) berdasarkan Pasal 36 UU No. 1 Tahun 2023 tentang KUHP Baru, jika identitas asli dari operator yang menginput data mentah atau mengarahkan algoritma GAN tersebut gagal dilacak oleh penyidik forensik digital. Karenanya, pembaruan hukum positif kita baik dalam sinkronisasi KUHP Baru maupun Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua UU ITE, sama-sama harus mengintegrasikan

[27]Desty Aster Yansen Basah, Andika Wijaya, dan Ivans Januardy. "Kriminalisasi Pelanggaran Protokol Digital: Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial". *Journal Of Social Science Research*. Vol. 5. No.4. 2025. Hal. 386-398.https://j-innovative.org/index.php/Innovative.

[28] Siti Nurkholisah, Daud Rismana, Afrizal Eko Nugroho, dkk. "Tantangan Kriminalisasi Deepfake dalam Hukum Pidana Indonesia". *Jurnal USM Law Review*. Vol. 8. No. 3. 2025. Hal. 2421-2445. https://doi.org/10.26623/julr.v8i3.13060

penguatan hukum materiil dengan standardisasi sistem forensik digital nasional yang mampu menembus manipulasi data otonom. Menemukan aktor manusia utama ini menjadi kunci mutlak untuk mengembalikan elemen pembuktian subjektif dari kesengajaan manusia yang menyalahgunakan teknologi, sehingga penegakan hukum pidana siber tidak lagi berjalan pincang di era digital. Dengan perpaduan antara rekonstruksi doktrin pertanggungjawaban hukum dan ketajaman pelacakan pelaku siber inilah, maka seluruh problematika krisis hukum yang telah diuraikan di atas dapat bermuara pada simpulan serta rekomendasi konseptual.

D. KESIMPULAN

Berdasarkan seluruh uraian penilitian yang telah dibahas sebelumnya, dapat disimpulkan bahwa doktrin *mens rea* konvensional benar menghadapi tantangan konseptual yang serius akibat berkembangnya otonomisasi kecerdasan buatan. Asumsi antroposentris hukum pidana materiil kita, seperti yang tertuang secara eksplisit dalam Pasal 36 UU No. 1 Tahun 2023 tentang KUHP Baru, mengalami kelumpuhan ketika harus berhadapan dengan karakteristik otonom dan buramnya *black box system* yang dimiliki dalam sistem AI. Sistem kecerdasan buatan tidak memiliki kesadaran moral, akal budi, atau hubungan batiniah layaknya manusia, melainkan bekerja mandiri berdasarkan probabilitas kode matematis dan data input awal. Hal ini menyebabkan terjadinya bagian yang terputus antara niat awal pengguna manusia dengan luaran destruktif yang dihasilkan secara mandiri oleh algoritma mesin, sehingga asas *geen straf zonder schuld* kehilangan landasan pembuktian subjektifnya yang absolut di ruang pengadilan.

Selanjutnya, penelitian ini membuktikan bahwa karakteristik khusus kejahatan *deepfake* yang diproduksi melalui mekanisme *Generative Adversarial Network* (GAN) memicu problematika yuridis yang *tricky* dalam menentukan pertanggungjawaban pidana. Bentuk tindak pidana berupa penipuan digital (*digital fraud*), pencemaran nama baik, hingga pornografi non-konsensual memiliki tingkat anonimitas pelaku yang sangat tinggi serta skalabilitas yang masif, sehingga dampak kerusakan reputasi sosial korban bersifat abadi karena jejak digital sulit untuk dihapus permanen secara keseluruhan (apalagi jika sudah tersebar luas). Di sisi lain, instrumen hukum positif Indonesia saat ini, khususnya UU No. 1 Tahun 2024 tentang Perubahan Kedua UU ITE, masih bersifat reaktif dan belum memiliki definisi teknis yang spesifik mengenai *AI-generated crime*. Kekosongan hukum formal ini menyebabkan batas kesalahan di antara para aktor siber menjadi kabur, sehingga hukum pidana terjebak dalam pusaran *multi-actor liability* yang membingungkan antara pengembang (*developer*), pengguna (*user*), dan penyebar (*distributor*) konten manipulatif tersebut.

Untuk mengatasi krisis pembuktian *mens rea* tersebut, artikel ini menawarkan formula rekonstruksi asas kesalahan (*culpabilitas*) yang ideal melalui perluasan konsep kesalahan dari individual murni menuju ke arah sistemik, serta menggeser tolok ukur dari pembuktian niat subjektif ke penilaian risiko objektif (*risk-based approach*). Penerapan ragam alternatif model pertanggungjawaban pidana seperti *strict liability*, *vicarious liability*, maupun *negligence-based liability* dinilai jauh lebih realistis untuk digunakan dalam menutup celah kekosongan hukum, ketimbang memaksakan konsep *electronic personhood* bagi AI yang masih sangat kontroversial secara teoretis. Implikasi nyata dari penelitian ini menegaskan bahwa Indonesia membutuhkan reformasi hukum pidana digital secara bertahap agar mampu mengakomodasi standar kelalaian digital (*digital negligence*) dalam penggunaan sistem otonom secara adil dan berkepastian hukum.

Pada akhirnya, seluruh bangunan rekonstruksi teori pertanggungjawban hukum berbasis risiko ini tidak akan berjalan maksimal jika tidak dibarengi dengan kemampuan nyata aparat penegak hukum untuk melacak dan menemukan pelaku sesungguhnya di balik layar (*the real actor behind the screen*). Menemukan aktor manusia utama yang menginput data awal tetap menjadi kunci mutlak untuk mengembalikan elemen pembuktian subjektif dari kesengajaan manusia yang menyalahgunakan otonomi AI, yang mana hal ini tentu memerlukan standardisasi sistem forensik digital nasional yang kuat. Berdasarkan pada simpulan-simpulan objektif yang telah dipaparkan di atas, penulisan artikel ilmiah ini nantinya akan merumuskan beberapa sumbangsih pemikiran serta rekomendasi konseptual yang lebih praktis, yang mana poin-poin strategis tersebut akan dijabarkan secara rinci.

SARAN

Menyikapi krisis konseptual yang terjadi pada elemen pembuktian batiniah hukum pidana, saran pertama adalah dengan melakukan reformasi hukum pidana materiil secara progresif dan futuristik. Pembuat undang-undang tidak boleh membiarkan kekosongan hukum ini berlarut-larut. Pemerintah perlu menyusun regulasi turunan yang memperluas interpretasi asas culpabilitas dengan memasukkan doktrin *negligence-based liability* atau kesalahan berbasis kelalaian digital (*digital negligence*) bagi pengguna teknologi otonom. Di samping itu, Pemerintah perlu segera melakukan revisi terhadap Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua UU ITE dengan mencantumkan definisi teknis yang spesifik mengenai kejahatan rekayasa visual dan audio berbasis algoritma *Generative Adversarial Network* (GAN). Langkah kodifikasi yang spesifik ini sangat krusial agar penegakan hukum di masa depan memiliki standar atribusi kesalahan yang jelas dan tidak lagi dipaksa menabrak

asas *lex certa* akibat tumpang tindihnya penerapan pasal kesusilaan konvensional dalam menjerat manipulasi siber.[29]

Saran kedua adalah untuk meningkatkan kapasitas teknis dan standardisasi penanganan alat bukti siber. Penegak hukum sebaiknya tidak terus terjebak pada metode pembuktian niat batin klasik yang bersifat verbal, melainkan harus mulai mengadopsi mekanisme pelacakan jejak digital yang sekiranya mampu menembus kerumitan sistem *black box* kecerdasan buatan demi menemukan pelaku asli di balik layar komputer (untuk kasus seperti *deepfake* ini). Aparat penegak hukum dapat merumuskan standardisasi nasional yang rigid terkait penanganan alat bukti digital (*digital evidence*) yang bersifat mudah berubah (*volatile*) dengan mengadopsi standar SNI ISO/IEC 27037:2014 melalui penguatan laboratorium forensik digital.[30] Standardisasi penanganan ini sangat penting agar APH seperti jaksa penuntut umum dapat secara presisi membuktikan hubungan kausalitas hukum serta sikap batin kesengajaan dari pengguna manusia yang secara sadar memberikan instruksi input data mentah kepada mesin untuk memproduksi konten destruktif.[31] Peningkatan keahlian forensik ini akan sangat membantu kelompok kunci, terutama para korban kejahatan siber, agar hak-hak hukum mereka dapat dilindungi secara adil di persidangan.

Terakhir, saran ketiga kita dapat menerapkan tata kelola teknologi berbasis risiko (*risk-based approach*). Pemerintah dapat meningkatkan status regulasi yang awalnya hanya berupa imbauan moral seperti misal yang ada dalam Surat Edaran Menteri Komunikasi dan Informatika Nomor 9 Tahun 2023 tentang Etika Kecerdasan Buatan menjadi sebuah regulasi pengawasan yang mengikat secara hukum.[32] Pemerintah harus mewajibkan setiap pengembang aplikasi AI generatif lokal maupun internasional yang beroperasi di yurisdiksi Indonesia untuk menanamkan sistem pengaman otomatis berupa tanda air digital secara terenkripsi pada setiap luaran video atau audio yang dihasilkan oleh mesin.[33] Langkah

---

[29] S. A. U. Sijabat dan D. Lukitasari. "Konten Gambar dan Video Pornografi Deepfake Sebagai Suatu Bentuk Tindak Pidana Pencemaran Nama Baik". *Recidive: Jurnal Hukum Pidana Dan Penanggulangan Kejahatan*. Vol. 13. No. 2. 2024. 185, https://jurnal.uns.ac.id/recidive.

[30] Badan Standardisasi Nasional."SNI ISO/IEC 27037:2014 Teknologi informasi - Teknik keamanan - Pedoman identifikasi, pengumpulan, perolehan, dan preservasi bukti digital". Badan Standardisasi Nasional. 2014. hlm. 2. https://aksessni.bsn.go.id/.

[31] Budi Widodo."Deteksi Deepfake: Tantangan Penegakan Hukum Di Indonesia". *Jurnal Kriminologi Indonesia*. Vol. 8. No. 2023. 155. https://journals.ui.ac.id/index.php/jki.

[32] Kementerian Komunikasi dan Informatika."Surat Edaran Menteri Komunikasi dan Informatika Nomor 9 Tahun 2023 tentang Etika Kecerdasan Buatan". JDIH Kominfo. 2023. hal. 3, https://jdih.kominfo.go.id/produk_hukum/view/id/871/t/surat+edaran+menteri+komunikasi+dan+informatika+nomor+9+tahun+2023+tentang+etika+kecerdasan+buatan.

[33] Ronit Justo-Hanani. "*Risk-based approach to EU AI act: benefits and challenges of co-regulation*". *Policy Design and Practice.* Vol. 9. No. 2. 2026. 173, https://doi.org/10.1080/25741292.2025.2610869.

preventif ini sejalan dengan regulasi modern internasional *European Union Artificial Intelligence Act* yang membagi klasifikasi risiko teknologi secara ketat guna memitigasi penyalahgunaan platform sebelum konten manipulatif menyebar luas di media sosial. Melalui kolaborasi sinergis antara pembaruan undang-undang pidana siber, ketajaman forensik aparat, dan tanggung jawab etis para pengembang teknologi, tatanan hukum digital Indonesia barulah dapat berjalan secara adaptif, berkepastian hukum, serta senantiasa mampu memberikan perlindungan yang berkeadilan bagi seluruh masyarakat.

E. DAFTAR PUSTAKA

Jurnal

Wafi, M. S., Wisnubroto, A., & Prayudi, Y. (2025). Kejahatan Deepfake Berbasis Artificial Intellegence: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana. Jurnal Reformasi Hukum, 29(2), 168–183. https://doi.org/10.46257/jrh.v29i2.1304

Basah, D. A. Y., Wijaya, A., & Januardy, I. (2025). Kriminalisasi Pelanggaran Protokol Digital: Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial. Innovative: Journal Of Social Science Research, 5(4), 386–398. https://j-innovative.org/index.php/Innovative

Putri, N. P. M., Hartono, M. S., & Yudiawan, I. D. G. H. (2024). Analisis Reformulasi Pertanggungjawaban Pidana Pengguna Teknologi Deepfake dalam Tindak Pidana Pencemaran Nama Baik Berbasis Artificial Intelligence. Jurnal Pacta Sunt Servanda, 5(2), 118–128. https://ejournal2.undiksha.ac.id/index.php/JPSS

Nurkholisah, S., Rismana, D., Nugroho, A. E., Munjiyah, A., & Ayunisa, Q. (2025). Tantangan Kriminalisasi Deepfake dalam Hukum Pidana Indonesia. Jurnal USM Law Review, 8(3), 2421–2444. https://journals.usm.ac.id/index.php/usm-law-review

Widodo, B. (2023). Deteksi Deepfake: Tantangan Penegakan Hukum Di Indonesia. Jurnal Kriminologi Indonesia, 8(2), 150–165. https://journals.ui.ac.id/index.php/jki

Justo-Hanani, R. (2026). Risk-based approach to EU AI act: benefits and challenges of co-regulation. Policy Design and Practice, 9(2), 171–180. https://doi.org/10.1080/25741292.2025.2610869

Radyananda, B., Abdaud, F., & Dagani, G. (2026). Analisis Limitasi Pertanggungjawaban Pidana Dalam Kejahatan Manipulasi Digital di Indonesia. Nomos: Jurnal Penelitian Ilmu Hukum, 6(1), 1–7. https://journal.actual-

insight.com/index.php/nomos/article/download/4273/3406

Sijabat, S. A. U., & Lukitasari, D. (2024). Konten Gambar dan Video Pornografi Deepfake Sebagai Suatu Bentuk Tindak Pidana Pencemaran Nama Baik. Recidive: Jurnal Hukum Pidana Dan Penanggulangan Kejahatan, 13(2), 179-194. https://jurnal.uns.ac.id/recidive

Suryokencana, B., & Naufal, I. (2024). Deepfake dan Tantangan Disinformasi di Era AI. Literasi Nusantara.

Widjaja, R. (2025). Harmonisasi UU ITE Dan UU PDP Dalam Kasus Deepfake. Jurnal Legislasi Indonesia, 22(1), 33–48. https://doi.org/10.22334/jli.v22i1.7890

Internet

Badan Standardisasi Nasional. (2014). SNI ISO/IEC 27037:2014 Teknologi informasi - Teknik keamanan - Pedoman identifikasi, pengumpulan, perolehan, dan preservasi bukti digital. Badan Standardisasi Nasional. https://aksessni.bsn.go.id/

Kementerian Komunikasi dan Informatika. (2023). Surat Edaran Menteri Komunikasi dan Informatika Nomor 9 Tahun 2023 tentang Etika Kecerdasan Buatan. JDIH Kominfo. https://jdih.kominfo.go.id/produk_hukum/view/id/871/t/surat+edaran+menteri+komunikasi+dan+informatika+nomor+9+tahun+2023+tentang+etika+kecerdasan+buatan

Institute for Criminal Justice Reform. (2024). Catatan Atas Revisi Kedua UU ITE: Ruang Gelap Kriminalisasi di Dunia Maya. ICJR Articles. https://icjr.or.id/

UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Hukumonline. (2024). Menjerat Pelaku Kejahatan Deepfake dengan KUHP Baru: Sebuah Tantangan Pembuktian. Hukumonline Articles. https://www.hukumonline.com/

Lembaga Studi dan Advokasi Masyarakat. (2024). Tantangan Regulasi Kecerdasan Buatan dan Perlindungan Hak Korban di Era Digital. ELSAM Policy Brief. https://elsam.or.id/

Center for Digital Society. (2024). Menelisik Ancaman Deepfake: Keamanan Siber dan Implikasinya terhadap Hukum di Indonesia. CfDS Research Report. https://cfds.fisipol.ugm.ac.id/

Perundang-undangan

Indonesia. (2023). Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana. Lembaran Negara Republik Indonesia Tahun 2023 Nomor 1. Jakarta: Sekretariat Negara.

Indonesia. (2024). Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik.

Lembaran Negara Republik Indonesia Tahun 2024 Nomor 1. Jakarta: Sekretariat Negara.

Indonesia. (2022). Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi. Lembaran Negara Republik Indonesia Tahun 2022 Nomor 245. Jakarta: Sekretariat Negara.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj