JRH
ojs.uid.ac.id

Reformasi Hukum
ISSN 1693-9336 | e-ISSN 2686-1598
Vol.29 No.2 August 2025 (page. 168-183)
doi.org/10.46257/jrh.v29i2.1304

# Kejahatan *Deepfake* Berbasis *Artificial Intellegence*: Suatu Konsepsi pada Penggunaan Asas Culpabilitas Sebagai Pembaharuan Pertanggungjawaban Pidana

## *Artificial Intelligence-Based Deepfake Crimes: A Conception of Culpability Principle as a Criminal Liability Reform*

Muhammad Syafiq Wafi[1]*, Aloysius Wisnubroto[2], Yudi Prayudi[3]

[1] Master of Art in Public Law, Dicle Universitesi, Diyarbakir, Turkiye; dan Fakultas Hukum, Universitas Islam Indonesia, Yogyakarta, Indonesia.
[2] Fakultas Hukum, Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia.
[3] Fakultas Teknologi Industri, Universitas Islam Indonesia, Yogyakarta, Indonesia.

***Corresponding author email**: 23912071@students.uii.ac.id.

**Paper**
Submitted
11-03-2025
Accepted
27-08-2025

**Abstrak**

Fenomena kejahatan *deepfake* berbasis *artificial intelligence* (AI) menuntut pembaharuan konsep pertanggungjawaban pidana melalui perluasan asas *culpabilitas* yang memungkinkan penempatan AI sebagai subjek hukum. Namun, gagasan menjadikan AI sebagai subjek hukum yang mandiri (*electronic personhood*) dinilai tidak relevan, mengingat AI tidak memiliki kehendak dan kebebasan ideal sebagaimana manusia. Oleh karena itu, penelitian ini menawarkan model pertanggungjawaban pidana yang mengarahkan perluasan asas *culpabilitas* kepada penyedia dan pengguna teknologi *deepfake*. Dengan menggunakan metode penelitian hukum normatif yang bertumpu pada bahan hukum primer dan sekunder, studi ini menelaah penerapan asas *culpabilitas* secara komprehensif melalui pendekatan perbandingan antarnegara. Hasil penelitian menunjukkan bahwa bentuk pertanggungjawaban yang paling proporsional adalah model *vicarious liability*, yang awalnya diterapkan pada korporasi namun dapat disesuaikan untuk konteks AI. Dalam model ini, penyedia perangkat lunak dapat dimintai pertanggungjawaban atas tindakan AI yang digunakan untuk melakukan kejahatan *deepfake*, terutama sebagai bentuk tanggung jawab terhadap regulasi tata kelola teknologi. Penelitian ini merekomendasikan pembentukan regulasi nasional yang menekankan pada sistem tata kelola berbasis *risk assessment*, *risk management*, dan *impact assessment* sebagaimana diterapkan di Uni Eropa, Kanada, dan Amerika Serikat. Kesimpulannya, pembaharuan pertanggungjawaban pidana berbasis AI merupakan langkah strategis dalam menanggulangi kejahatan *deepfake* yang semakin masif serta memastikan sistem hukum tetap adaptif terhadap perkembangan teknologi.

**Kata Kunci**

*Artificial Intelligence*; Culpabilitas; *Deepfake*; Pertanggungjawaban Pidana; Subjek Hukum.

***Abstract***

*The phenomenon of deepfake crimes based on artificial intelligence (AI) demands a reform of criminal liability concepts through the expansion of the culpability principle, allowing the placement of AI as a subject of law. However, the idea of recognizing AI as an independent legal entity (electronic personhood) is considered irrelevant, since AI lacks human-like will and moral autonomy. Therefore, this study proposes a model of criminal liability that extends the culpability principle to providers and users of deepfake technology. Using a normative legal research method based on primary and secondary legal materials, this study comprehensively examines the application of the culpability principle through a comparative approach among various jurisdictions. The findings indicate that the most proportional form of liability is the vicarious liability model, which was initially applied to corporations but can be adapted to the AI context. In this model, software providers may be held liable for acts committed by AI in deepfake crimes, particularly as part of their responsibility toward technology governance regulations. The study recommends establishing national regulations emphasizing governance systems based on risk assessment, risk management, and impact assessment, as practiced in the European Union, Canada, and the United States. In conclusion, reforming criminal liability in the AI era is a*

*strategic step to address the growing prevalence of deepfake crimes and to ensure that the legal system remains adaptive to technological developments.*

***Keywords***
*Artificial Intelligence; Criminal Liability; Culpability; Deepfake; Legal Subject.*



## 1. Pendahuluan

Perkembangan teknologi yang begitu pesat telah membawa perubahan signifikan terhadap modus operandi kejahatan. Fenomena ini terjadi karena munculnya kejahatan siber sebagai bentuk baru yang turut memengaruhi karakteristik kejahatan konvensional. Hadirnya kecerdasan buatan atau *Artificial Intelligence* (selanjutnya disebut AI) telah menjadi faktor utama dalam transformasi tersebut. Adapun salah satu bentuk kejahatan berbasis AI yang paling mendominasi saat ini adalah *deepfake*.

Telah terjadi banyak kasus kejahatan *deepfake* berbasis AI yang terjadi di berbagai belahan dunia, salah satunya di Shanghai – Tiongkok, yang melibatkan kasus penipuan asmara daring (*romance scam*). Dalam kasus ini, seorang pria bernama Liu mengalami kerugian sebesar 200.000 yuan (sekitar Rp.450.000.000,00) setelah menjalin hubungan dengan sosok perempuan fiktif yang sepenuhnya merupakan hasil rekayasa AI. Sistem AI tersebut menciptakan identitas palsu bernama "Nona Jiao", lengkap dengan tampilan visual dan gaya komunikasi yang meyakinkan. Interaksi antara Liu dan sosok virtual tersebut berlangsung intens hingga menumbuhkan rasa kepercayaan dan kedekatan emosional. Pelaku kemudian memanfaatkan situasi tersebut dengan meminta sejumlah uang atas alasan kebutuhan usaha dan biaya pengobatan kerabat, sehingga mendorong korban melakukan transfer dana ke rekening yang dikendalikan oleh pelaku penipuan pelaku penipuan.[1] Melihat kasus tersebut, tampak bahwa dominasi AI dalam menghasilkan gambar atau video yang tampak realistis, serta kemampuannya menciptakan teks menyerupai percakapan manusia, telah membuka peluang terjadinya tindak penipuan.

Di Indonesia, modus serupa juga muncul dalam bentuk penipuan berbasis impersonifikasi wajah dan suara, sebagaimana dialami oleh publik figur Baim Wong. Dalam kasus ini, pelaku melakukan penipuan melalui panggilan video (*video call*) dengan meniru wajah dan suara Baim Wong secara sangat meyakinkan. Modus yang digunakan ialah dengan mengunduh video asli Baim Wong, kemudian mengintegrasikannya dengan teknologi kecerdasan buatan berbasis *deepfake*untuk melakukan peniruan wajah (*face imitation*) dan pengisian suara (*voice cloning*). Hasil manipulasi tersebut digunakan untuk mengelabui korban agar percaya bahwa mereka berkomunikasi langsung dengan Baim Wong, sehingga korban terdorong melakukan transfer dana ke rekening pelaku.[2] Selain itu, kasus serupa juga menimpa Presiden Republik Indonesia, Prabowo Subianto, melalui video manipulatif di *platform* Instagram yang seolah-olah menampilkan dirinya berjanji akan memberikan bantuan pasca pemilihan presiden. Video tersebut menyebabkan banyak masyarakat di sedikitnya 20 provinsi menjadi korban penipuan. Mereka diarahkan menghubungi nomor WhatsApp tertentu untuk proses penyaluran bantuan fiktif dan diminta membayar biaya administrasi antara Rp.250.000,00 hingga Rp.1.000.000,00.[3]

Adanya berbagai kasus yang melibatkan kejahatan siber berbasis AI melalui teknologi *deepfake* menunjukkan bahwa fenomena tersebut telah menjadi ancaman faktual di Indonesia. Namun, hingga saat ini, belum terdapat regulasi nasional yang secara responsif dan komprehensif mengatur bentuk pertanggungjawaban pidana atas kejahatan yang melibatkan peran aktif AI. Jenis-jenis kejahatan yang muncul antara lain meliputi penyebaran konten *deepfake*, penipuan dengan *AI voice cloning*, penyebaran *hoaks* melalui *AI-generated image*, penipuan layanan berbasis *chatbot AI* atau *AI-generated scam* melalui *auto chat* dan *phishing*, serta pemalsuan dokumen atau surat elektronik menggunakan teknologi *AI OCR* dan *text generator*.

Kejahatan-kejahatan tersebut menandai adanya transformasi nyata dalam lanskap kejahatan digital nasional, di mana peran manusia semakin kabur akibat intervensi AI yang mampu meniru tindakan, ekspresi, bahkan komunikasi manusia. Meskipun

sebagian tindak pidana tersebut masih dapat dijerat melalui Undang-Undang Republik Indonesia Nomor 1 Tahun 2024 tentang Perubahan Kedua atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (selanjutnya disebut UU ITE), serta akan digantikan oleh Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana (selanjutnya disebut sebagai KUHP Nasional) pada tahun 2026, kedua regulasi tersebut belum memberikan dasar hukum yang spesifik terhadap kejahatan yang melibatkan kinerja atau keputusan otonom dari AI itu sendiri.

Akibatnya, terdapat kekosongan hukum (*legal gap*) yang signifikan dalam sistem pertanggungjawaban pidana terhadap kejahatan berbasis AI. Hal ini terjadi karena maksud dan tujuan pembentuk undang-undang pada saat penyusunan UU ITE maupun KUHP Nasional belum diarahkan untuk menghadapi fenomena kejahatan digital yang dihasilkan oleh kecerdasan buatan, seperti *deepfake*. Sebagai contoh, Pasal 4 UU ITE menegaskan bahwa tujuan pengaturan adalah untuk mencerdaskan kehidupan bangsa, mengembangkan perdagangan dan perekonomian nasional, meningkatkan pelayanan publik, memperkuat pemanfaatan teknologi informasi, serta memberikan rasa aman dan kepastian hukum bagi penyelenggaraan teknologi informasi—namun belum menyentuh aspek pengendalian risiko kejahatan dari teknologi itu sendiri.[4] Dalam praktiknya, apabila kejahatan berbasis AI hanya diakomodasi melalui regulasi yang telah ada dengan menggunakan metode penafsiran ekstensif, maka pendekatan tersebut menjadi tidak sepenuhnya relevan. Hal ini disebabkan karena maksud dan tujuan pembentuk undang-undang tidak pernah diarahkan untuk menanggulangi penyalahgunaan kecerdasan buatan. Dengan kata lain, tindak pidana berbasis AI memang mungkin "terpenuhi secara formil" terhadap rumusan delik tertentu (*tatbestand massigheit*) melalui metode penafsiran ekstensif, namun secara substansial pembentuk undang-undang tidak bermaksud demikian (*wessenschau*).

Hal tersebut berimplikasi pada sulitnya penanganan kejahatan deepfake dalam tahapan pra-ajudikasi, yang disebabkan oleh beberapa faktor: *pertama*, teknologi *deepfake* terbuka dan mudah diakses; *kedua*, tersedianya berbagai perangkat lunak yang ada dalam versi *open-source* seperti *DeepFaceLab* dan *StyleGAN*; *ketiga,* penggunaan teknologi *deepfake* untuk kejahatan sering kali sulit terdeteksi karena pelaku menggunakan *virtual private work* (VPN);[5] *keempat*, sulitnya melacak akun pertama yang menyebarkan video *deepfake* karena tersebar di jejaring sosial, adapun jika ditemui sering kali akun tersebut merupakan akun palsu atau *anonymus*.[6]

Fakta-fakta tersebut menunjukkan adanya kekosongan hukum yang menegaskan perlunya regulasi yang dapat menjangkau dan mengatur penyedia teknologi AI agar tidak disalahgunakan, terutama karena ketersediaannya dalam bentuk *open-source*. Dalam kasus penyalahgunaan *deepfake* yang menimpa Presiden Republik Indonesia Prabowo Subianto, misalnya, aparat penegak hukum masih menghadapi kesulitan dalam menelusuri pelaku, yang menunjukkan bahwa pendekatan represif semata tidak memadai tanpa disertai langkah preventif yang bersifat sistemik. Dengan demikian, regulasi yang mengatur penyedia AI berbasis *open-source* di Indonesia perlu dilakukan pembaruan, khususnya dengan menerapkan prinsip *legal personality* sebagai dasar perluasan pertanggungjawaban pidana. Oleh karena itu, studi ini juga mengkaitkan mental elemen (culpabilitas) yang relevan dengan kejahatan *deepfake* berbasis AI dengan menjadikan AI sebagai subjek hukum yang merujuk pada kontribusi dari *programer* yang tidak melimitasi *software*-nya sehingga dapat digunakan sebagai alat kejahatan, termasuk *user* yang sudah menyalahgunakan teknologi *deepfake*.

Perlunya pembaruan hukum pidana yang mampu mengakomodasi penyalahgunaan AI dalam bentuk *deepfake* menjadi penting agar hukum dapat berfungsi secara preventif maupun represif sebagai sarana kontrol sosial, manajemen tata kelola, serta perlindungan hukum. Oleh karena itu, tantangan berikutnya adalah bagaimana merumuskan bentuk pertanggungjawaban pidana terhadap kejahatan siber berbasis AI.

Fenomena kejahatan *deepfake* yang semakin masif menimbulkan pertanyaan mendasar mengenai sejauh mana asas culpabilitas dapat diterapkan pada entitas non-manusia seperti AI. Dengan demikian, pembentukan regulasi yang komprehensif tidak cukup hanya mengatur aspek kriminalisasi dan penalisasi, tetapi juga perlu mencakup mekanisme tata kelola penggunaan AI itu sendiri. Regulasi semacam ini diharapkan dapat mengakomodasi perluasan asas culpabilitas dalam konteks AI, terutama dengan

menempatkan *programmer* dan pengguna (*user*) sebagai subjek hukum yang bertanggung jawab atas pemanfaatan teknologi tersebut.

Adapun penelitian-penelitian terdahulu yang memiliki relevansi dengan studi ini antara lain:

*Pertama,* Heny Novyanti dan Pudji Astuti dalam artikel ilmiah jurnalnya tentang "Jerat Hukum Penyalahgunaan Aplikasi *Deepfake* Ditinjau Dari Hukum Pidana"[7], menemukan bahwa penyalahgunaan aplikasi *deepfake* dapat dikualifikasikan sebagai tindak pidana karena telah memenuhi kriteria kriminalisasi. Selain itu, tindakan tersebut juga memenuhi unsur-unsur tindak pidana baik secara subjektif maupun objektif. Namun, penelitian tersebut hanya menelaah penerapan regulasi yang ada tanpa menawarkan kerangka hukum baru, sedangkan penelitian ini mengajukan perluasan asas culpabilitas sebagai upaya preventif terhadap kejahatan *deepfake*.

*Kedua,* Rendi Syaputra Nur Haida dalam artikel ilmiah jurnalnya tentang "Urgensi Pengaturan Perlindungan Hukum Terhadap Korban *Deepfake* Melalui *Artificial Inteligence* (AI) Dari Perspektif Hukum Pidana Indonesia"[8], menyoroti adanya kekosongan hukum terkait kejahatan berbasis AI, yang seiring waktu menyebabkan meningkatnya jumlah korban maupun pelaku. Penelitian tersebut menjelaskan bahwa undang-undang yang berlaku di Indonesia masih terbatas pada upaya pencegahan, tanpa memberikan efek jera yang memadai terhadap pelaku. Penelitian tersebut masih terbatas pada analisis normatif tanpa menawarkan pedoman yang jelas bagi penyedia dan pengguna *deepfake*, sedangkan studi ini menghadirkan kebaruan melalui integrasi prinsip culpabilitas dalam regulasi mitigasi risiko untuk menegaskan eksistensi AI sebagai subjek hukum berbasis tanggung jawab penyedia dan pengguna.

*Ketiga***,** Cheny Berlian dalam disertasinya "Pertanggungjawaban Pidana Pelaku Kejahatan Siber Menggunakan *Artificial Intelligence*"[9], menemukan bahwa peraturan yang ada saat ini belum memadai untuk mengakomodasi perkembangan kejahatan siber, baik sebagai langkah preventif maupun represif, mengingat transformasi kejahatan digital yang begitu masif. Dalam disertasi tersebut ditegaskan bahwa AI belum memenuhi kriteria sebagai subjek hukum, sehingga tanggung jawab tetap dibebankan kepada manusia sebagai pengembang atau pengguna. Berbeda dengan penelitian tersebut, penelitian ini mengusulkan konsep AI sebagai subjek hukum yang terinspirasi dari *legal personality* korporasi, sekaligus memperluas asas *culpabilitas* untuk membangun sistem pertanggungjawaban pidana yang mencakup penyedia dan pengguna, seiring semakin otonom dan masifnya peran AI, terutama dalam teknologi *deepfake*.

*Keempat***,** Istiqomah, Mutiasih Savanah Puti Dinanty, Surwangi Resti Anggun Saputri dalam artikel ilmiah jurnalnya "Kedudukan Hukum *Deepfake* AI sebagai Subjek Hukum dan Pertanggungjawaban Pidananya Berdasarkan Teori Dualistis"[10], membahas persoalan kekosongan hukum dalam menanggulangi kejahatan berbasis AI, khususnya deepfake, dengan menggunakan teori dualistis. Penelitian tersebut menggunakan teori dualistis untuk menjelaskan kebingungan dalam menentukan pihak yang seharusnya bertanggung jawab. Berbeda dengan penelitian tersebut, penelitian ini memperluas pertanggungjawaban kepada penyedia dan pengguna AI, dengan penyedia diwajibkan melakukan tata kelola sebagai mitigasi risiko dan pengguna yang menyalahgunakan teknologi tetap dimintai pertanggungjawaban.

Dengan demikian, kebaruan (*novelty*) dalam penelitian ini terletak pada penyajian model-model pertanggungjawaban pidana dengan memasukkan gagasan mengenai AI sebagai subjek hukum sebagai langkah preventif terhadap maraknya kejahatan deepfake secara konseptual. Selain itu, guna menyederhanakan konsep pertanggungjawaban AI, studi ini mereduksi berbagai gagasan agar semakin relevan dengan konteksnya, dengan memaparkan penerapan asas culpabilitas secara komprehensif melalui studi perbandingan antarnegara yang memiliki kesamaan ruang lingkup dengan kejahatan deepfake. Hal ini didasarkan pada kenyataan bahwa penggunaan AI yang semakin otonom berpotensi memengaruhi pola dan bentuk kejahatan di dunia digital.

Adapun rumusan permasalahan dalam studi ini berfokus pada pertanyaan: bagaimana pembaruan pertanggungjawaban pidana pada kejahatan deepfake berbasis AI sebagai konsepsi atas penerapan asas culpabilitas? Penelitian ini bertujuan untuk memberikan pemahaman baik secara teoritis maupun praktis mengenai pembaruan konsep pertanggungjawaban pidana, khususnya bagi praktisi dan akademisi hukum dalam konteks kejahatan *deepfake*.

## 2. Metode

Jenis penelitian yang digunakan dalam Penelitian ini yaitu penelitian hukum normatif karena yang dikaji adalah pembaharuan pertanggungjawaban pidana pada kejahatan *deepfake* berbasis *artificial intelligence* sebagai konsepsi atas penerapan asas culpabilitas. Pendekatan yang digunakan pada studi ini adalah pendekatan konseptual, pendekatan perbandingan, dan pendekatan peraturan perundang-undangan. Pada pendekatan konseptual mengacu pendapat – pendapat ahli terkait dengan konsep penerapan asas culpabilitas dalam rangka pembaharuan pertanggungjawaban pidana dan ruang lingkup yang relevan pada kejahatan *deepfake.* Adapun pendekatan perbandingan mengacu pada regulasi di luar negara Indonesia yang relevan pada studi ini. Bahwa alasan digunakannya tiga pendekatan tersebut yaitu; *pertama*, pendekatan konseptual digunakan untuk menelaah asas culpabilitas; *kedua*, pendekatan perbandingan digunakan untuk mengidentifikasi praktik hukum di negara lain; *ketiga*, pendekatan peraturan perundang-undangan digunakan untuk menilai kesesuian norma dalam KUHP Nasional dan UU ITE.

Pendekatan-pendekatan tersebut diintegrasikan dalam analisis melalui empat tahap yaitu; 1) Identifikasi konseptual, pada tahap ini pendekatan yang digunakan adalah pendekatan konseptual yang berfungsi untuk menyajikan konsep kejahatan *deepfake* berbasis AI sekaligus fungsinya sebagai kerangka konseptual; 2) Identifikasi norma, pada tahap ini pendekatan yang digunakan adalah pendekatan peraturan perundang-undangan untuk menelaah kesesuaian dan kekosongan norma baik dalam KUHP Nasional maupun dalam UU ITE; 3) Perbadingan sistem hukum asing, pada tahap ini pendekatan yang digunakan adalah pendekatan perbandingan yang berfungsi untuk mengkaji dan menelaah praktik hukum di negara lain yang berkaitan dengan perkembangan AI sebagai subjek hukum dan mitigasi resiko termasuk yang patut dipertanggungjawabkan pada penyalahgunaan AI berupa *deepfake* dengan menyajikan *foreign regulation* yang relevan. Adapun alasan dalam menyajikan *foreign regulation* adalah untuk menilai pertanggungjawaban atau sanksi penyedia *software* AI untuk memitigasi tindakan kriminal salah satunya berupa *deepfake* termasuk kewajiban melakukan *systemic risk management, impact assessments,* dan *high-impact AI systems* yang sudah dilakukan di beberapa negara; 4) Tahap sintesis konsep baru, pendekatan yang digunakan adalah perpaduan pendekatan konseptual, peraturan perundang-undangan, dan pendekatan perbandingan.

Sumber bahan hukum pada studi ini yaitu bahan hukum primer yaitu peraturan perundang-undangan berupa UU ITE dan KUHP Nasional. Selain itu, adanya bahan hukum sekunder berupa *foreign regulation* dan tulisan ilmiah yang relevan dengan pembaharuan pertanggungjawaban pidana pada kejahatan *deepfake* berbasis *artificial intelligence* sebagai konsepsi atas penerapan asas culpabilitas yaitu; 1) House of Commons of Canada. BILL C-27 An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts; 2) The House of Representatives. H. r. 6580 To direct the Federal Trade Commission to require impact assessments of automated decision systems and augmented critical decision processes, and for other purposes; 3) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive; 4) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations; 5) European Civil Law Rules in Robotics. Directorate-General for Internal Policies: Policy Department C: Citizens' Rights and Constitutional Affairs; 6) Buku, jurnal, artikel, dan literatur lain yang masih berkaitan. Pada bahan hukum sekunder tersebut, sudah dilakukan validitas melalui masing – masing publikasi resminya.

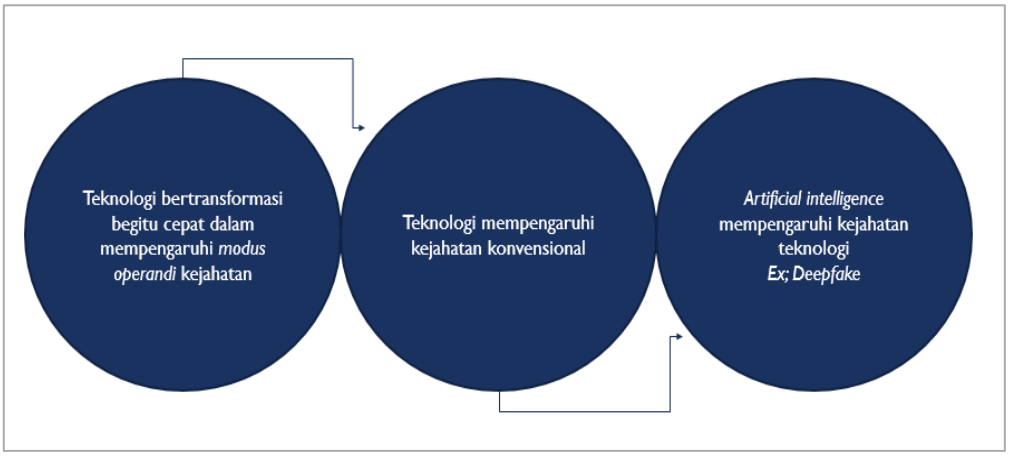
Adapun bahan hukum yang sudah dikumpulkan dianalisis secara deskriptif kualitatif. Ada tiga jalur dalam analisis kualitatif pada studi ini, yaitu reduksi bahan hukum, penyajian bahan hukum, dan penarikan kesimpulan. Reduksi bahan hukum dilakukan dengan menyederhanakan konsep culpabilitas dalam menilai AI sebagai pengaruh kejahatan yang ada dari berbagai literatur, sehingga pereduksian tersebut agar lebih relevan dalam menunjang *novelty* studi ini.

## 3. Hasil dan Pembahasan

### 3.1 Kejahatan *Deepfake* Berbasis *Artificial Intelligence* (AI)

*Deepfake* memiliki makna media baik audio maupun visual dalam format elektronik yang dibuat, diubah, atau dimanipulasi secara digital sehingga orang lain merasa hal tersebut asli yang merepresentasikan perilaku aktual seseorang yang tergambar dalam rekaman tersebut.[11] Sedangkan AI merupakan sub-bidang ilmu komputer yang berfungsi untuk mengkaji keterampilan manusia sehingga dapat mempersepsi, bernalar, dan bertindak.[12] *Deepfake* sebagai dalam menjalankan peranannya berbasis pada AI (kecerdasan buatan). AI sendiri berkembang dengan pesat sehingga memunculkan potensi negatif berupa penyebaran misinformasi, menyebarkan ketidakpercayaan, dan memungkinkan penipuan dan bentuk kejahatan lainnya.[13] Oleh karena itu, kejahatan dan teknologi merupakan dua hal yang perkembangannya berubah secara masif seiring dengan perubahan zaman. Sebagaimana gambar dibawah ini:

**Gambar 1**. Transformasi Kejahatan dan Teknologi




*Sumber:* diolah oleh penulis

Pada gambar tersebut, pada mulanya kejahatan merupakan representasi aktivitas konvensional yang melanggar norma kehidupan. Namun seiring dengan berjalannya waktu, kemajuan teknologi yang turut berkembang juga mempengaruhi kejahatan konvensional baik sebagai alat utama, alat bantu, atau sasaran kejahatan sehingga melahirkan tiga tipologi kejahatan siber berupa *computer crime*, *computer-supported crime*, dan *computer-facilitated crime*. Kendati demikian, tidak berhenti pada keterlibatan teknologi semata, dijaman yang kian modern kini perkembangan AI semakin mendominasi dalam aktivitas kehidupan sehingga AI menjadi sangat otonom. Hal ini turut mempengaruhi kejahatan dibidang *cyberspace* dimana salah satu kejahatan yang sering dijumpai adalah *deepfake*. Oleh karena itu, penulis menyebutnya sebagai *AI-facilitated crime*, sedangkan dalam diskursus yang lebih konkret terkait kejahatan *deepfake* disebut *deepfake as AI-facilitated crime.* Transformasi ini menunjukan bahwa hukum pidana menghadapi disrupsi ontologis terhadap konsep perbuatan dan pelaku.

Penulis mengkasifikasikan kejahatan *deepfake* berbasis AI sendiri berdasarkan pandangan Gareth Shelwel, walaupun memang klasifikasi ini tidak bersifat konkrit karena transformasi digital dan *modus operandi* yang begitu masif sehingga paling tidak sudah merepresentasikan perkembangan-perkembangan tersebut selama masih relevan. Menurut Gareth Shelwel di laman Canipish menjelaskan enam *trend* kejahatan yang berbasis AI yaitu penipuan asmara, kejahatan impersonifikasi atau peniruan suara dan wajah seseorang, kanal percakapan pada akun media sosial yang palsu, surat elektronik palsu, penipuan berupa percakapan langsung, dan penipuan layanan dan produk investasi.[14]

Adapun klasifikasi kejahatan siber berbasis AI yang dikemukakan oleh Gareth Shelwel dijabarkan dalam penelitian ini untuk memberikan gambaran praktis atas fenomena tersebut.

*Pertama,* kejahatan yang merepresentasikan bentuk *romance fraud* atau penipuan asmara. Dalam kasus ini, korban sering kali tertipu dan mengalami kerugian finansial

akibat mempercayai pasangan yang ternyata merupakan sosok fiktif hasil rekayasa sistem AI. Kecerdasan buatan yang melatarbelakangi penciptaan identitas palsu tersebut mampu membangun karakter dengan sapaan tertentu serta menjalin komunikasi intensif hingga menumbuhkan rasa kedekatan dan empati. Seiring waktu, pelaku melalui identitas fiktif tersebut kerap meminta bantuan dana, misalnya untuk membuka usaha atau membantu kerabat yang sakit, sehingga korban tidak ragu melakukan transfer dana ke rekening yang ditentukan, seperti dalam kasus "Nona Jiao". Kasus ini menunjukkan dominasi AI dalam menciptakan gambar dan video yang tampak realistis, serta kemampuannya menghasilkan teks yang dapat menimbulkan penipuan.

*Kedua*, kejahatan yang merepresentasikan bentuk impersonifikasi atau penipuan melalui suara dan wajah. Modus ini umumnya menimpa tokoh publik, di mana pelaku menggunakan teknologi *deepfake* untuk meniru wajah dan suara tokoh tersebut dalam panggilan video, guna mengelabui korban agar mentransfer sejumlah dana ke rekening penipu. Pelaku biasanya mengunduh video asli tokoh publik, kemudian memanipulasi tayangan tersebut dengan mengarahkan kamera ponsel dan menambahkan suara hasil *dubbing* melalui aplikasi berbasis AI.

Masih banyak kasus serupa yang berkaitan dengan kejahatan *deepfake* yang terus bermunculan. Kendati demikian, tantangan utama yang dihadapi aparat penegak hukum terletak pada proses pelacakan dan identifikasi pelaku. Salah satu contoh menonjol adalah kasus yang menimpa CEO perusahaan energi di Jerman, yang menjadi korban penipuan melalui video *deepfake* dengan kerugian mencapai 3,8 miliar rupiah. Pelaku berhasil memindahkan dana tersebut, sementara aparat penegak hukum sempat mengalami kendala dan kehilangan jejak dalam proses penyelidikan.[15] Kondisi ini menunjukkan bahwa penegakan hukum terhadap kejahatan *deepfake* sering kali terhambat karena pelaku tidak dapat ditemukan atau tidak terdapat individu maupun kelompok yang dapat dimintai pertanggungjawaban.[15] Hambatan tersebut disebabkan oleh karakteristik unik dari kejahatan itu sendiri.

Terdapat beberapa faktor yang menyebabkan sulitnya pelacakan terhadap pelaku *deepfake*. *Pertama,* teknologi *deepfake* bersifat mudah diakses dan bersifat terbuka bagi masyarakat luas. Beberapa perangkat lunak seperti DeepFaceLab, FaceSwap, dan StyleGAN tersedia dalam versi *open-source*, sehingga siapapun dapat mengunduh dan menggunakannya.[5] Kesulitan penelusuran semakin meningkat apabila pelaku memanfaatkan Virtual Private Network (VPN), yang dapat menyamarkan lokasi dan identitas pengguna.[6] *Kedua,* ketika sebuah video *deepfake* tersebar di dunia maya, penyebarannya berlangsung sangat cepat melalui berbagai jejaring sosial. Hal ini berdampak pada kesulitan untuk melacak sumber unggahan pertama secara akurat. Bahkan jika konten asli berhasil ditemukan, pelaku umumnya menggunakan akun palsu untuk menyamarkan identitasnya.[16] *Ketiga,* kelemahan dalam kerangka hukum turut menjadi faktor penghambat investigasi. Oleh karena itu, diperlukan regulasi khusus yang mengatur *deepfake* sebagai tindak pidana tersendiri, serta inovasi dalam konsep pertanggungjawaban pidana yang mampu secara efektif menanggulangi kejahatan berbasis AI, baik dalam fungsi preventif maupun represif. Adapun regulasi yang saat ini berpotensi menjangkau tindak pidana *deepfake* dapat diuraikan sebagai berikut:

**Tabel. 1** Regulasi yang dapat menjangkau kejahatan *deepfake*

| **Regulasi** | **Rumusan delik** | **Bentuk keterbatasan dalam konteks kejahatan *deepfake* yang berbasis AI** |
|---|---|---|
| Pasal 27 A UU ITE | *Setiap Orang dengan sengaja menyerang kehormatan atau nama baik orang lain dengan cara menuduhkan suatu hal, dengan maksud supaya hal tersebut diketahui umum dalam bentuk Informasi Elektronik dan/ atau Dokumen Elektronik yang dilakukan melalui Sistem Elektronik.* | Rumusan ini tidak mengakomodir AI di dalam regulasinya. Selain itu, sulitnya mendeteksi *deepfake* dengan akurasi tinggi terhadap pelaku yang menyerang kehormatan atau nama baik orang lain.[17] |

| | | |
|---|---|---|
| Pasal 27 Ayat (1) UU ITE | *Setiap Orang dengan sengaja dan tanpa hak menyiarkan, mempertunjukkan, mendistribusikan, mentransmisikan, dan/atau membuat dapat diaksesnya Informasi Elektronik dan/ atau Dokumen Elektronik yang memiliki muatan yang melanggar kesusilaan untuk diketahui umum.* | Banyaknya kasus pornografi yang sejatinya dapat diterapkan dengan Pasal ini bagi orang menyiarkan, mempertunjukan..' [18]. Kendati demikian, kesulitan mencari pelaku yang menyamarkan identitas melalui penggunaan *proxy* atau VPN termasuk ketidakjelasan unsur tanpa hak dan melanggar kesusilaan dikarenakan *deepfake* sering kali dibuat dari hasil rekayasa algoritmik sehingga sekiranya *mens rea* pelaku saja tidak cukup tanpa keterlibatan penyedia AI yang seharusnya terlibat dalam memitigasi kejahatan tersebut. |
| Pasal 28 Ayat (1), (2), dan (3) UU ITE | ▪ *Setiap Orang dengan sengaja dan/atau mentransmisikan Informasi Elektronik dan/ atau Dokumen Elektronik yang berisi pemberitahuan bohong atau informasi menyesatkan yang mengakibatkan kerugian materiel bagi konsumen dalam Transaksi Elektronik.*<br>▪ *Setiap Orang dengan sengaja dan tanpa hak mendistribusikan dan/atau mentransmisikan Informasi Elektronik dan/atau Dokumen Elektronik yang sifatnya menghasut, mengajak, atau memengaruhi orang lain sehingga menimbulkan rasa kebencian atau permusuhan terhadap individu dan/atau kelompok masyarakat tertentu berdasarkan ras, kebangsaan, etnis, warna kulit, agama, kepercayaan, jenis kelamin, disabilitas mental, atau disabilitas fisik.*<br>▪ *Setiap Orang dengan sengaja menyebarkan Informasi Elektronik dan/atau Dokumen Elektronik yang diketahuinya memuat pemberitahuan bohong yang menimbulkan kerusuhan di masyarakat.* | ▪ Dimungkinkannya terjadi simpangsiur pembuktian terhadap unsur informasi yang menyesatkan atau pem-beritahuan bohong karena bisa jadi korban sejatinya benar melakukan perbuatan bohong (*victim participation to crime*) namun di satu sisi aparat penegak hukum kesulitan membuktikan informasi tersebut bohong karena *deepfake* sering kali tidak berbentuk penyataan faktual melainkan rekonstrusi suara yang menyerupai realitas, di sisi lain *deepdake* dianggap bohong atau menyesatkan masih menjadi abu-abu dalam ruang lingkup UU ITE.<br>▪ Teknologi *deepfake* berkerja melalui algoritma otomatis yang sedikit banyaknya dapat mengkaburkan unsur tanpa hak dikarenakan data atau gambar yang digunakan bersumber dari *public space* sehingga tidak ada batas yang jelas. |
| Pasal 407 Ayat (1) KUHP Nasional | *Setiap Orang yang memproduksi, membuat, memperbanyak, menggandakan, menyebarluaskan, menyiarkan, mengimpor, meng-ekspor, menawarkan, memperjual-belikan, menyewakan, atau menyediakan Pornografi, dipidana dengan pidana penjara paling singkat 6 (enam) Bulan dan pidana penjara paling lama l0 (sepuluh)* | Bahwa maksud dan tujuan pembentuk undang-undang pada unsur *"setiap orang yang memperoduksi.."* disusun dengan asumsi manusia secara manual menciptakan pornografi, sehingga apakah aktivitas software dengan mengunggah wajah seseorang ke sistem AI sudah bisa dikategori-kan sebagai memproduksi atau |

| | | |
|---|---|---|
| | *tahun atau pidana denda paling sedikit kategori IV dan pidana denda paling banyak kategori VI.* | membuat.[19] Hal ini dikarena-kan tidak semua konten yang dihasilkan dari teknologi dianggap sebagai produksi manual.[20] |

Regulasi-regulasi tersebut, melalui penafsiran ekstensif, pada dasarnya dapat digunakan untuk menindak secara represif pelaku kejahatan berbasis AI. Namun, timbul pertanyaan apakah pendekatan penafsiran ekstensif tersebut tidak akan menjadi terlalu luas dan apakah dapat bertahan dalam jangka panjang. Hal ini berpotensi menimbulkan pertentangan dengan asas *lex certa*, *lex stricta*, dan *lex scripta*. Lebih lanjut, eksistensi asas legalitas berpotensi mengalami degradasi akibat perdebatan antara penggunaan penafsiran ekstensif dan analogi, yang pada prinsipnya dilarang dalam asas legalitas dan hingga kini masih menjadi polemik dalam teori hukum pidana.

Dalam praktiknya, apabila penegakan hukum hanya mengandalkan regulasi yang telah ada—seperti Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) serta Kitab Undang-Undang Hukum Pidana (KUHP)—dengan menggunakan metode penafsiran ekstensif, maka pendekatan tersebut dapat dianggap kurang relevan. Hal ini karena secara historis dan teleologis, maksud serta tujuan pembentuk undang-undang tidak ditujukan untuk mengatur norma mengenai penyalahgunaan AI. Dengan kata lain, kejahatan berbasis AI memang dapat disesuaikan dengan rumusan delik (*tatbestand massigheit*) melalui penafsiran ekstensif, tetapi sejatinya pembentuk undang-undang tidak bermaksud demikian (*wessenschau*).

Oleh karena itu, diperlukan gagasan pembaharuan hukum pidana yang bersifat preventif dalam menanggulangi kejahatan *deepfake*, melalui pembentukan konsep pertanggungjawaban pidana yang mampu mengakomodasi karakteristik AI yang semakin otonom. Hal ini menjadi penting mengingat kejahatan berbasis AI memiliki karakteristik khusus yang menyulitkan proses pelacakan, serta menunjukkan keterbatasan kerangka hukum yang ada, yang pada akhirnya berimplikasi terhadap hambatan investigasi. Dengan demikian, pengaturan khusus terhadap tindak pidana *deepfake* dan inovasi dalam sistem pertanggungjawaban pidana menjadi kebutuhan mendesak untuk menjamin efektivitas penegakan hukum di era kecerdasan buatan.

### 3.2 Penanggulangan Kejahatan *Deepfake* Melalui Gagasan *Artificial Intelligence* (AI) Sebagai Subjek Hukum

Seiring berkembangnya teknologi, khususnya kecerdasan buatan, muncul tantangan baru dalam ranah hukum pidana, salah satunya adalah kejahatan *deepfake*. Kondisi ini menuntut pembaruan hukum pidana yang mampu mengakomodasi tindak pidana berbasis AI, termasuk kemungkinan munculnya kejahatan lain yang memanfaatkan AI. Salah satu isu mendasar yang perlu dikaji adalah apakah AI dapat diakui sebagai subjek hukum. Dalam ilmu hukum, subjek hukum adalah pemegang hak dan kewajiban, sedangkan objek hukum merupakan sasaran atau penerima hak. Saat ini, AI umumnya diperlakukan sebagai objek hukum karena merupakan produk ciptaan manusia atau badan hukum yang dimediasi melalui teknologi. Upaya untuk memberikan status subjek hukum pada AI pernah diajukan oleh European Parliament Committee on Legal Affairs pada 2017, dengan konsep *electronic personhood,* yang bertujuan mengisi kekosongan pertanggungjawaban hukum ketika AI menimbulkan kerugian atau melakukan tindakan tercela yang tidak terkait langsung dengan penciptanya.[21] Gagasan tersebut dipandang tidak relevan mengingat konsep AI sendiri berasal dari negara-negara Eropa, meskipun Amerika Serikat juga turut berperan dalam pengembangan AI awal. Sejalan dengan itu, Nathalie Nevejans menilai bahwa konsep *electronic personhood* tidak memiliki dasar ontologis yang kuat, karena AI, sebagaimana dimaksud, tidak memiliki kesadaran maupun kehendak bebas—dua unsur yang menjadi prasyarat pertanggungjawaban hukum.[22]

Permasalahan pertanggungjawaban yang sepenuhnya dikenakan pada AI menurut Sabine Gless, Emily Silverman dan Thomas Weigend membawa kembali pada pertanyaan dasar tentang apa arti menjadi manusia. Para filsuf sudah lama memperdebatkan hal ini sehingga terdapat kesimpulan yang berbeda-beda. Menurut John Locke manusia harus

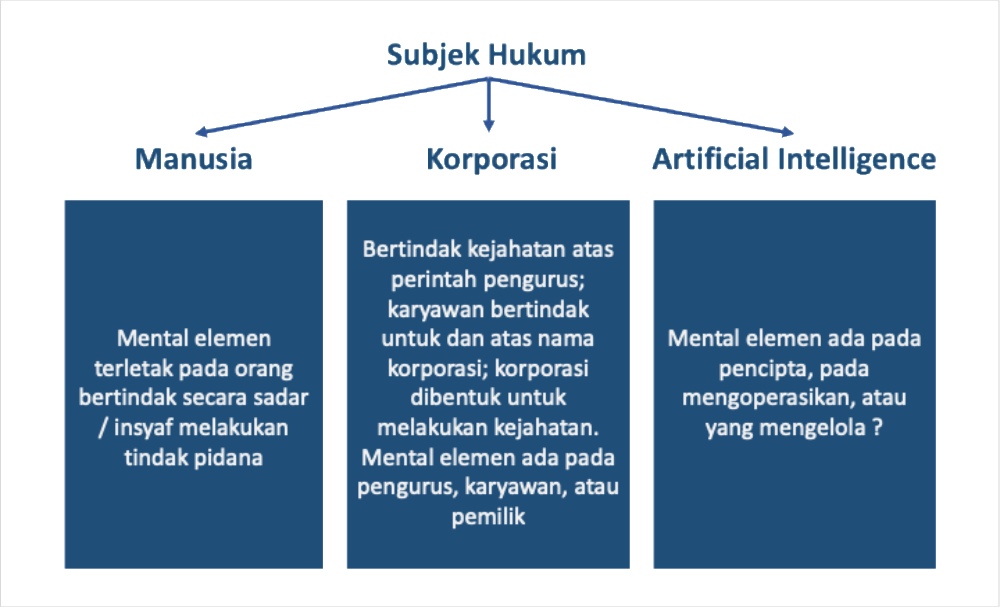
mampu memiliki hukum, penderitaan, dan kebahagiaan sehingga harus memiliki kesadaran termasuk kesadaran akan masa lalu sebagai kemampuan mengingat yang dengannya memiliki kesadaran, kepedulian, dan rasa tanggung jawab. Immanuel Kant dengan mengutip Gewissen bahwa sebagai bukti kebenaran atas korespondensi orang lain juga akan mengetahui seseorang pelaku juga mampu bertanggungjawab atas perbuatannya. Menurut Sabine Gless, Emily Silverman dan Thomas Weigend berdasarkan pendapat para filsuf tersebut pada abad ke-18 dan ke-19 memang tidak terpikirkan adanya AI.[23] Kendati demikian, jelas bahwa AI sendiri tidak memiliki karakteristik seperti manusia. Menurut Lawrence B. Solum, AI sendiri tidak memiliki persyaratan kebebasan yang idealis walaupun AI dapat belajar membuat keputusan, terlebih lagi sejatinya AI sendiri tidak menyadari perbuatannya sehingga tidak mengetahui konsep hak dan kewajiban.[24] Oleh karena itu, menurut Gunther Teubner AI di program dengan sistem kelebihan dan kekurangan dalam menentukan keputusan tertentu sehingga dapat saja penyedia teknologi AI memperoleh kemampuan untuk terlibat dalam penalaran moral. Penyedia dapat dipersalahkan ketika AI sendiri juga dapat dipersalahkan mengingat sanksi hukum pidana ditujukan untuk manusia bukan ditujukan untuk entitas non manusia seperti AI.[25]

Pendapat Gunther Teubner tersebut menjadi pondasi bahwa AI dapat saja menjadi subjek hukum dimana hak dan kewajiban diletakan pada penyedia berupa manusia atau korporasi yang tertuju kepada manusia. Oleh karena itu, sudah sepatutnya usaha untuk menjadikan AI sebagai subjek hukum yang mandiri seperti *electronic personhood* patut untuk ditinggalkan sehingga bukan membahas bagaimana substitusi pertanggungjawaban pidana, melainkan bagaimana mengembangkan perluasan pertanggungjawaban pidana. Gagasan AI sebagai subjek hukum guna mengakomodasi langkah preventif kejahatan seperti *deepfake* seharusnya dapat merujuk pada salah satu bentuk perluasan pertanggungjawaban pidana yang pernah dilakukan ketika eksistensi korporasi sebagai subjek hukum.

Korporasi sebagai subjek hukum dilatarbelakangi oleh lahirnya prinsip *legal personality* sebagai pengakuan hukum atas eksistensi korporasi sebagai subjek hukum yang berdiri sendiri, memiliki kekayaan sendiri, dan bisa menggugat atau digugat di pengadilan.[26] Atas perkembangan tersebut lahirlah beberapa teori pertanggung-jawaban pidana yakni; *pertama*, teori Identifikasi yaitu korporasi dapat melakukan berbagai macam delik secara *direct* yang sejatinya dilakukan oleh orang-orang yang ada pada korporasi itu sendiri sehingga segala akibat yang dilakukan oleh orang-orang di dalam korporasi merupakan tindakan dari korporasi itu sendiri; *kedua,* teori *strict liability* yaitu dalam pertanggungjawaban pidana tidak mensyaratkan adanya kesalahan dalam diri pelaku, kendati demikian teori ini tidak sesuai jika diterapkan pada korporasi karena pertanggungjawaban tanpa kesalahan hanya mungkin terjadi pada kategori pelanggaran ringan, namun perkembangannya mulai ada sejak korporasi diakui sebagai subjek hukum; *ketiga*, teori *vicarious liability*, teori ini mendudukan pertanggungjawaban sebagai pengganti sehingga yang patut bertanggungjawab adalah atasan atas perbuatan bawahannya.[27]

Bahwa sesungguhnya kontribusi perkembangan pertanggungjawaban pidana yang diperluas dengan memasukan korporasi sebagai subjek hukum menjadi momentum yang berharga untuk menjawab permasalan sosial hukum pidana sehingga dapat membantu persoalan kejahatan AI ini. Menurut Ahmad Sofian, guna mengukur pertanggungjawaban AI tidah hanya pada satu aktor sebagai subjek hukum yang melibatkan AI untuk melakukan tindak pidana, tetapi ada beberapa subjek hukum lain yang penting dipertimbangkan. Bahwa keterlibatan korporasi menjadi faktor yang menentukan selain manusia dan AI itu sendiri dikarenakan AI tidak dapat berdiri sendiri sehingga perlu mengukur kontribusi dari subjek hukum lain dengan menggunakan mental elemen (culpabilitas) atau kesalahan dari masing-masing subjek hukum. Unsur subjektif atau mental elemen merupakan faktor penting dalam menentukan subjek hukum (lihat gambar 2).[28]

**Gambar 2.** Subjek Hukum Manusia, Korporasi dan AI




*Sumber:* Ahmad Sofian, Maret 2025.

Gagasan AI sebagai subjek hukum setidaknya hampir serupa dengan eksistensi korporasi sebagai subjek hukum, dimana korporasi dalam melakukan kejahatan berdasarkan perintah dari pengurus untuk betindak dan atas nama korporasi sehingga letak mental element berada pada pengurus tersebut, sedangkan gagasan AI sebagai subjek hukum dalam melakukan kejahatan berdasarkan element pencipta/ pengelola/yang mengoperasikan sebagai pengguna. Oleh karena itu, hal ini merupakan langkah fundamental dalam rangka melakukan rekonseptualisasi atas penerapan asas culpabilitas sebagai pembaharuan pertanggungjawaban pidana terhadap dominasi kejahatan yang berbasis AI seperti *deepfake* ini. Hal ini ditujukan pada AI sebagai subjek hukum yang berbasis pada pencipta/pengelola ditujukan kepada koporasi, mengingat pengelola AI dalam suatu aplikasi biasanya di dominasi oleh korporasi, adapun AI sebagai subjek hukum yang berbasis pada pengguna dilakukan jika kapasitas aparat penegak hukum memadai untuk mengungkap pengguna tersebut, hal ini dikarenakan sangat sulit untuk menemukan pelaku secara langsung pada kejahatan *deepfake* karena dominasi *pseudonim* yang sangatlah kuat. Kendati demikian, mengingat pembahasan mengenai prinsip *legal personality* sebagaimana yang dimiliki oleh korporasi, diakuinya AI sebagai subjek hukum seharusnya memperhatikan atribusi dari *legal personality*-nya. Hal tersebut dikarenakan pengakuan AI sebagai subjek hukum bukan dimaknai bahwa hukum terikat untuk memberikan hak dan kewajiban kepada AI sebagaimana yang dimiliki oleh manusia.[29]

### 3.3 Pembaharuan Pertanggungjawaban Pidana Bagi Penyedia *Artificial Intelligence* (AI) yang Menimbulkan Kejahatan *Deepfake* sebagai Konsepsi atas Penerapan Asas Culpabilitas

Secara etimologis, konsepsi sendiri berasal dari bahasa latin yang bermakna pemahaman, pengertian, atau tangkapan akal pada suatu ide. Dalam perbincangan ilmiah, konsepsi dimaknai sebagai kerangka berfikir yang dapat menggambarkan gagasan, pemahaman, maupun pandangan.[30] Adapun asas culpabilitas sebagai prinsip dasar hukum pidana yang menegasikan adanya pidana harus terdapat kesalahan / *geen straf zonder schuld*. Asas ini menjadi pondasi utama dalam menentukan pertanggungjawaban pidana. Kemudian pada perkembangannya, lahirlah tiga model pertanggungjawaban AI oleh Gabriel Hallevy yaitu; model pertanggungjawaban yang dilakukan oleh orang lain, model pertanggungjawaban alamiah, dan model pertanggungjawaban langsung. *Pertama*, model pertanggungjawaban yang dilakukan oleh orang lain ini tidak memandang AI sebagai organ yang tidak bersalah karena ada keterlibatan manusia sehingga yang patut dipersalahkan adalah *programmer* dari perangkat lunak AI. *Kedua,* model pertanggungjawaban alamiah menginsafi bahwa

*programmer* bisa jadi tidak memiliki niat jahat, namun adakalanya AI dapat berkontribusi pada kejahatan sehingga adanya organ alamiah antara *programmer* yang lalai dan *user* yang berimplikasi pada akibat terlarang, oleh karenanya organ tersebut dapat dimintai pertanggungjawaban. *Ketiga*, model pertanggungjawaban langsung tidak memberikan kedudukan pada *programmer* dan *user* yang bisa dipertanggungjawabkan semata, melainkan pada AI sepanjang melekat *actus reus* dan *mens rea*. Oleh karena itu digunakanlah tanggung jawab bersama antara AI, *programmer,* dan *user* tergantung pada peran masing-masing.[29]

Pandangan Gabriel Hallevy dapat dijadikan rujukan dalam membangun konsepsi *culpabilitas* serta dalam merumuskan pembaharuan pertanggungjawaban pidana. Kendati demikian, timbul pertanyaan mengenai bagaimana asas *culpabilitas* dapat diperluas tanpa menyalahi prinsip *geen straf zonder schuld* (tiada pidana tanpa kesalahan). Dalam konteks hukum pidana Indonesia, ajaran yang diakui sebagai bentuk perluasan dari prinsip tersebut adalah teori *strict liability* dan teori *vicarious liability*, yang pada awalnya muncul sebagai respons terhadap pengakuan korporasi sebagai subjek hukum. Namun demikian, penerapan kedua teori ini tidak terbatas pada entitas korporasi semata.

Selanjutnya, perlu dipahami mengapa kedua teori tersebut dapat dikategorikan sebagai bentuk perluasan prinsip *geen straf zonder schuld*. Teori *strict liability* menekankan adanya pertanggungjawaban mutlak tanpa pembuktian unsur kesalahan [31], sedangkan teori *vicarious liability* mengandung pertanggungjawaban pengganti terhadap perbuatan orang lain yang berada dalam lingkup tanggung jawabnya. Kedua teori ini pada hakikatnya merupakan upaya untuk memperluas prinsip *geen straf zonder schuld* agar hukum pidana tetap mampu mengakomodasi perkembangan sosial tanpa mengabaikan asas fundamental "tiada pidana tanpa kesalahan".[32]

Berdasarkan hal tersebut, model pertanggungjawaban yang ditawarkan oleh Gabriel Hallevy dapat dikonkretkan melalui bentuk perluasan prinsip *geen straf zonder schuld* sebagaimana telah dikenal dalam sistem hukum Indonesia. Namun, perlu dikaji lebih lanjut teori mana yang paling tepat diterapkan dalam konteks kejahatan *deepfake*—apakah teori *strict liability* atau *vicarious liability*. Oleh karena itu, perlu digambarkan terlebih dahulu bentuk aktivitas kejahatan *deepfake*berbasis AI secara konkret untuk menentukan relevansi penerapan teori yang paling sesuai dengan model yang ditawarkan oleh Hallevy. Adapun aktivitas kejahatan *deepfake* dan uraian rinciannya adalah sebagai berikut:

**Gambar 3.** Salah Satu Aktivitas Kejahatan *Deepfake*


*Sumber:* diolah oleh penulis.

*Pertama*, *data collection* dilakukan denga mengambil foto atau video dari internet. *Kedua, AI model training* dilakukan dengan menggunakan *algoritma deep learning* dan melatih suara atau wajah. *Ketiga, deepfake creation* dilakukan dengan membuat video atau audio palsu korban. *Keempat, distribution* dilakukan dengan mengunggah ke media sosial. *Kelima,* berdampak kepada korban dan *keenam,* dilakukan investigasi seperti analisis forensik dan pelacakan metadata dan *IP Tracing or Log*.

Berdasarkan gambar tersebut, dapat diketahui bahwa proses terjadinya kejahatan *deepfake* melibatkan tiga unsur utama, yaitu *programmer*, AI, dan *user*. *Pertanyaannya kemudian, apakah ketiganya dapat dimintai pertanggungjawaban hukum?* Atribusi pertanggungjawaban dalam konteks ini dapat dianalisis melalui tiga model pertanggungjawaban AI sebagaimana dikemukakan oleh Gabriel Hallevy. Kendati demikian, langkah yang paling rasional dalam konteks pembaharuan hukum pidana adalah melalui pendekatan preventif dan represif. Pendekatan preventif menempatkan hukum pidana sebagai instrumen pengendalian dan reduksi kejahatan, sedangkan pendekatan represif berfokus pada penindakan terhadap pelaku. Studi ini secara khusus menitikberatkan pada upaya preventif terhadap kejahatan *deepfake* melalui pembaharuan hukum pidana, yakni dengan merumuskan regulasi yang mengatur tata kelola penyelenggara *software* berbasis AI agar tidak mudah disalahgunakan oleh *user*. Hal ini sejalan dengan ketentuan dalam UU ITE yang mewajibkan penyelenggara sistem elektronik untuk memastikan sistem yang andal dan aman. Oleh karena itu, regulasi mengenai tata kelola *software* berbasis AI diharapkan dapat mewajibkan *programmer* untuk memiliki tanggung jawab dalam mengontrol dan memastikan keamanan sistem elektronik yang mereka kembangkan.

Dalam praktiknya, perkembangan *software* berbasis AI mengenal suatu prinsip yang disebut *misuse resistance*, yaitu kapasitas sistem untuk mencegah terjadinya penyalahgunaan atau penyimpangan fungsi. Dalam konteks ini, *programmer* berperan sebagai agen sentral yang secara etis dapat dibebani tanggung jawab untuk membatasi fitur-fitur AI agar tidak dimanfaatkan secara ilegal, termasuk dalam kasus *deepfake*.[33] Upaya tersebut dapat dilakukan melalui berbagai pendekatan teknis, antara lain: pertama, penerapan *content-sensitive detection filter* untuk memblokir konten yang bersifat pornografis, *impersonation* wajah, maupun kekerasan; kedua, penerapan kontrol akses dan autentikasi pengguna guna membatasi pihak-pihak yang berhak mengakses fungsi-fungsi AI tertentu; ketiga, penerapan *audit log* serta sistem pelacakan aktivitas pengguna yang memungkinkan aparat penegak hukum menelusuri penyalahgunaan AI; dan keempat, penerapan model *guardrails* melalui penyisipan kode pengamanan yang berfungsi membatasi keluaran (*output*) tertentu dari AI.[34]

Dengan menempatkan *programmer* sebagai aktor sentral dalam upaya memitigasi risiko kejahatan, diperlukan adanya regulasi yang mengatur tata kelola agar pihak tersebut dapat turut dimintai pertanggungjawaban dalam sistem hukum pidana Indonesia. Penerapan asas *culpabilitas* dalam konteks ini dapat dipahami berdasarkan model pertanggungjawaban AI kedua atau ketiga sebagaimana dikemukakan oleh Gabriel Hallevy. Dalam model tersebut, *programmer* mungkin tidak memiliki niat jahat (*mens rea*), namun AI dapat berkontribusi terhadap terjadinya suatu kejahatan. Hal ini menimbulkan hubungan kausalitas antara kelalaian *programmer* dan tindakan *user* yang berimplikasi pada akibat yang terlarang. Oleh karenanya, organ yang terlibat dalam proses tersebut dapat dimintai pertanggungjawaban. Model pertanggungjawaban langsung (*direct liability model*) tidak semata-mata menempatkan *programmer* atau *user* sebagai subjek yang dapat dipertanggungjawabkan, melainkan juga membuka ruang bagi AI sepanjang melekat unsur *actus reus* dan *mens rea* di dalamnya. Dengan demikian, pertanggungjawaban pidana yang diterapkan bersifat kolektif antara AI, *programmer*, dan *user*, tergantung pada peran masing-masing dalam menentukan kadar *culpabilitas* yang terkandung di dalamnya.

Oleh karena itu, dalam beberapa kasus penyedia jasa *software* yang juga berstatus sebagai korporasi, maka AI sebagai subjek hukum dengan mengombinasikan model pertanggungjawaban yang dikemukakan oleh Gabriel Hallevy sejatinya dapat dikaitkan dengan penerapan teori *vicarious liability* sebagai bentuk pertanggungjawaban pengganti. Melalui teori ini, tanggung jawab hukum dapat dialihkan kepada pengurus atau atasan sebagai representasi dari kewajiban korporasi untuk mematuhi regulasi yang mengatur tata kelola penyedia *software*. Dengan demikian, pengembang atau

*programmer* memiliki kewajiban untuk memastikan sistem elektronik yang mereka rancang dapat dikontrol dengan baik. Penerapan mekanisme ini diharapkan mampu mengakomodasi penanggulangan kejahatan berbasis AI, khususnya *deepfake*, yang kian marak terjadi, serta berfungsi sebagai langkah preventif dalam mereduksi potensi penyalahgunaan teknologi tersebut.

Beberapa negara sudah mengakomodir sanksi pada korporasi penyedia *software* berbasis AI dalam memitigasi tindakan kriminal. Sebagai contoh, melalui Pasal 71 dan Pasal 99 Eropean Union AI Act[35] yang disahkan pada bulan Mei 2024 dan berlaku pada tahun 2025 mengatur bagi penyedia AI untuk melakukan *risk assessment* dan *risk management* pada setiap AI yang rilis. Sehingga korporasi yang menyediakan *software* berbasis AI akan diberikan sanksi denda 35 juta euro atau 7% dari omzet global bagi korporasi yang gagal menerapkan tata kelola AI serta melalui Digital Services Act Uni Eropa[36] yang berlaku pada 17 Februari 2024 memberi kewajiban tata kelola risiko sistemik (*systemic risk management)* untuk menindak penyalahgunaan pada kejahatan siber dan pemberian sanksi berupa denda 6% omzet global.

Adanya rancangan Algorithmic Accountability Act di Amerika Serikat,[37] walaupun di beberapa negara bagian sudah menerapkannya, hal ini mewajibkan korporasi di bidang teknologi untuk melakukan *impact assessments* terhadap sistem algoritmik dan AI, jika korporasi tersebut tidak menjalankannya sehingga menimbulkan penyalah-gunaan AI pada kejahatan maka dikenai sanksi dari Federal Trade Comission.

Adanya Canada's Artificial Intelligence dan Data Act[38] mengatur bagi pengembang dan penyedia *high-impact AI systems* perlu melakukan mitigasi risiko dan transparansi sehingga jika AI dimanfaatkan untuk melakukan kejahatan dan terbukti korporasi sebagai penyedia lalai dalam tata kelola maka dapat dikenai sanksi administasi maupun sanksi pidana.

Kejahatan *deepfake* yang terus merajalela dan paling mendominasi pada kejahatan yang berbasis AI perlu adanya regulasi yang sebagaimana hal nya diatas, walaupun mengatur semua kejahatan AI secara *general*. Formulasi dilakukan berdasarkan model pertanggungjawaban AI dan teori pertanggungjawaban pidana yang lahir dari perkembangan diakuinya korporasi sebagai subjek hukum untuk dielaborasikan pada ketentuan tata kelola yang mengandung sanksi pidana. Ada hal penting lagi adalah, formulasi regulasi dikontruksikan pada unsur kesalahan dari korporasi itu sendiri berupa *culpa* (kelalaian), karena dirasa tidak mungkin atau sulit jika koporasi mengembangkan AI secara sengaja yang bertujuan untuk memfasilitasi kejahatan.

## 4. Kesimpulan

Fenomena kejahatan *deepfake* berbasis AI menuntut adanya kebaharuan dengan mengkombinasikan perluasan asas culpabilitas sehingga mendudukan AI sebagai subjek hukum. Kendati demikian, upaya menjadikan AI sebagai subjek hukum yang mandiri (*electronic personhood*) sudah sepatutnya ditinggalkan karena AI sendiri tidak memiliki kebebasan idealis seperti manusia. Oleh karena itu, AI sebagai subjek hukum pada penelitian ini merujuk pada perluasan asas culpabilitas yang tertuju kepada penyedia dan pengguna dalam menjalankan teknologi *deepfake*.

Perluasan asas *culpabilitas* tersebut dirancang agar tetap sejalan dengan prinsip *geen straf zonder schuld*, sehingga bentuk pertanggungjawaban pidana yang diusulkan berbasis pada teori *vicarious liability*. Teori ini muncul sebagai respons terhadap pengakuan korporasi sebagai subjek hukum, namun penerapannya tidak terbatas pada korporasi saja. Dengan mengadopsi model pertanggungjawaban yang dikemukakan Gabriel Hallevy, penyedia teknologi yang tidak memiliki niat jahat tetap dapat dipertanggungjawabkan apabila AI yang mereka kembangkan digunakan untuk melakukan kejahatan (*deepfake*). Hal ini menekankan kewajiban penyedia *software* atau *programmer* untuk mengontrol sistem elektronik yang mereka ciptakan, sehingga potensi kejahatan berbasis AI dapat diminimalkan.

Sebagai rekomendasi kebijakan dalam pembaharuan hukum pidana untuk menanggulangi kejahatan deepfake, diperlukan regulasi yang menekankan sistem tata kelola, seperti *risk assessment*, *risk management*, *impact assessment*, maupun *high-impact AI systems*. Sistem tata kelola tersebut sudah dilakukan di beberapa negara dengan melibatkan sanksi pidana diantaranya *European Union*, Kanada, dan Amerika Serikat. Strategi serupa dapat menjadi langkah penting bagi Indonesia untuk melakukan

pembaharuan pertanggungjawaban pidana secara preventif dalam menanggulangi kejahatan *deepfake* yang semakin masif begitupun AI yang semakin otonom.

## 5. Ucapan Terima Kasih

Ucapan rasa terimakasih penulis ucapkan atas kehadirat Allah SWT yang sudah melimpahkan rahmat sehingga penulisan studi ini dapat berjalan dengan lancar dan semua pihak yang sudah memberikan bantuan kepada penulis untuk dapat menerbitkan tulisan ini. Semoga tulisan ini dapat memberikan manfaat bagi semua pihak.

## Daftar Pustaka

1. Sarwindaningrum, I. (2025). Tertipu Pacar AI, Rp 450 Juta Lenya. *Kompas*. Retrieved from https://www.kompas.id/artikel/tertipu-pacar-ai-rp-4583-juta-lenyap
2. OCBC. (2025). Bahaya Deepfake: Penipuan Wajah Palsu AI dan Cara Mencegahnya. *OCBC*. Retrieved from https://www.ocbc.id/id/article/2025/05/15/penipuan-wajah-palsu-ai-dan-cara-mencegahnya
3. Tim Detik.Com. (2025). Babak Baru Perkara Deepfake Catut Prabowo. *Detik.com*. Retrieved from https://news.detik.com/berita/d-7884867/babak-baru-perkara-deepfake-catut-prabowo
4. Ridwan. (2021). Kebijakan tentang Ujaran Kebencian Menurut Hukum Pidana Indonesia. Riau: Perpustakaan Universitas Islam Riau. Retrieved from https://repository.uir.ac.id/id/eprint/13502
5. Hao Li, et al. (2023). The State of Deepfake Detection: Challenges and Opportunities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(2), 345–359. Retrieved from https://doi.org/10.1109/TPAMI.2022.3152211
6. Solove, D. J. (2001). Privacy and Power: Computer Databases and Metaphors for Information Privacy. *Stanford Law Review*, *53*(6), 1393. https://doi.org/10.2307/1229546
7. Novyanti, H. & Astuti, P. (2022). Jerat Hukum Penyalahgunaan Aplikasi Deepfake Ditinjau Dari Hukum Pidana. *Novum : Jurnal Hukum*, *9*(4), 6. https://doi.org/10.2674/novum.v0i0.43571
8. Haida, R. S. N., & Nuriyatman, E. (2024). Urgensi Pengaturan Perlindungan Hukum Terhadap Korban Deepfake Melalui Artficial Inteligence (AI) Dari Perspektif Hukum Pidana Indonesia. *Jurnal Hukum Respublica*, 1–12. Retrieved from https://journal.unilak.ac.id/index.php/Respublica
9. Berlian, C. (2025). *Pertanggungjawaban Pidana Pelaku Kejahatan Siber Menggunakan Artificial Intelligence*. Universitas Jambi. Retrieved from https://repository.unja.ac.id/82076/5/FULL DISERTASI CHENY.pdf
10. Istiqomah, I., Dinanty, M. S. P., & Saputri, S. R. A. (2025). Kedudukan Hukum Deepfake AI sebagai Subjek Hukum dan Pertanggungjawaban Pidananya Berdasarkan Teori Dualistis. *Deposisi: Jurnal Publikasi Ilmu Hukum*, *3*(2), 59–74. https://doi.org/10.59581/deposisi.v3i2.5065
11. Instructions, C. J., Joseph, C. C., & Lamonica, P. R. (2025). 17 La. Civ. L. Treatise, Criminal Jury Instructions § 10:135.60 (3d ed.), *60* (November), 1–3. Retrieved from https://1.next.westlaw.com/
12. Fosberg, A. (2020). From Siri To Sci-Fi Are Lethal Robots People Too. *Penn State Law Review*, 3. Retrieved from https://1.next.westlaw.com/
13. Klinkner, B. A. (2025). What Attorneys Should Know About Deepfakes, 2023–2025. Retrieved from https://1.next.westlaw.com/
14. Maryoto., A (2025). Enam Kejahatan dengan Menggunakan AI Marak di Sekitar Kita. *Kompas*. Retrieved from https://www.kompas.id/artikel/enam-kejahatan-dengan-menggunakan-ai-marak-di-sekitar-kita
15. Rachmadie, D. T., & Supanto. (2020). Regulasi Penyimpangan Artificial Intelligence Pada Tindak Pidana Malware Berdasarkan Undang-Udang Republik Indonesia Nomor 19 Tahun 2016. *Recidive : Jurnal Hukum Pidana dan Penanggulangan Kejahatan*, *9*(2), 128. https://doi.org/10.20961/recidive.v9i2.47400
16. Ovadya, A., & Whittlestone, J. (2019). Reducing malicious use of synthetic media research: Considerations and potential release practices for machine learning, 1–11. Retrieved from http://arxiv.org/abs/1907.11274
17. Milano, A., Suharti, T., & Serfiyani, C. Y. (2024). Online Legal protection against defamation through deepfake in negatively charged political content. *International Journal of Law, Policy and Social Review*, *6*(3), 282–285. Retrieved from www.lawjournals.net
18. Novera, O., & Z., Y. F. (2024). Analisis Pengaturan Hukum Pidana terhadap Penyalahgunaan Teknologi Manipulasi Gambar (Deepfake) dalam Penyebaran Konten Pornografi Melalui Akun Media Sosial. *El-Faqih: Jurnal Pemikiran dan Hukum Islam*, *10*(2), 460. Retrieved from https://ejournal.iaifa.ac.id/index.php/faqih
19. Rohmawati, I., Junaidi, A., & Khaerudin, A. (2024). Urgensi Regulasi Penyalahgunaan Deepfake Sebagai Perlindungan Hukum Korban Kekerasan Berbasis Gender Online ( KBGO ). *Journal Of Social Science Research*, *4*(6), 1779–1794. https://doi.org/10.31004/innovative.v4i6.16559
20. Putri, N. A., & Apriyani, M. N. (2025). Pertanggungjawaban Pidana Pelaku Kekerasan Seksual Berbasis Elektronik Artificial Intelegence (Deep Fake Porn). *Wajah Hukum*, *9*(1), 352. https://doi.org/10.33087/wjh.v9i1.1725
21. Commite on Legal Affairs. (2017). Report with Recommendations to the Comission on Civil Law Rules on Robotics (2015/2103(INL)). *European Parliament*. Retrieved from

https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html
22. Nevejans, N. (2017). European Civil Law Rules in Robotics. *Directorate-General for Internal Policies: Policy Department C: Citizens' Rights and Constitutional Affairs*, 19–26. Retrieved from http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf
23. Gless, S., Silverman, E., and Weigend. T. (2016). If Robots Cause Harm , Who Is To Blame ? Self-Driving Cars And Criminal Liability *New Criminal Law Review: An International and Interdisciplinary Journal*, *19*(3), 416. Retrieved from https://www.jstor.org/stable/pdf/26417695.pdf?refreqid=fastly-default%3Ac2c97f67856789f60b3ad1904c584993&ab_segments=0%2Fspellcheck_basic_search%2Fcontrol&initiator=&acceptTC=1
24. Solum, L. B. (1992). Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, *70*(4), 1239. Retrieved from https://scispace.com/pdf/legal-personhood-for-artificial-intelligences-miggjk8k4s.pdf
25. Teubner, G. (2007). *Elektronische Agenten und grosse Menschenaffen: Zur Ausweitung des Akteursstatus in Recht und Politik*. *Interdisziplinäre Wege in der juristischen Grundlagenforschung*. Retrieved from https://www.jura.uni-frankfurt.de/42828694/Generic_42828694.pdf
26. Gierke, O. (1922). *Political Theories of The Middle Age*. Cambridge: Cambridge at The University Press. Retrieved from https://historyofeconomicthought.mcmaster.ca/gierke/MedPolTheo.pdf
27. Ali, M. (2011). *Dasar-Dasar Hukum Pidana* (Pertama.). Jakarta Timur: Sinar Grafika.
28. Sofian, A. (2025). Konsepsi Subjek Hukum dan Pertanggungjawaban Pidana Artificial Intellegence The Concept of Legal Subjects and Criminal Responsibility of Artificial Intelligence, *9*(1), 13–26. https://doi.org/10.1177/1741659020917434.Doowon
29. Ravizki, E. N., & Yudhantaka, L. (2022). Artificial Intelligence Sebagai Subjek Hukum: Tinjauan Konseptual dan Tantangan Pengaturan di Indonesia. *Notaire*, *5*(3), 351–376. https://doi.org/10.20473/ntr.v5i3.39063
30. Badan Pengembangan dan Pembinaan Bahasa. (2016). *Kamus Besar Bahasa Indonesia*. Kementerian Pendidikan dan Kebudayaan Republik Indonesia.
31. Suud, A. K. (2023). Analisis Penerapan Konsep Pertanggungjawaban Mutlak (Strict Liability) Dalam Kasus Korupsi. *Masalah-Masalah Hukum*, *52*(2), 154–155. https://doi.org/10.14710/mmh.52.2.2023.153-162
32. Kaluase, J. (2021). Kajian Yuridis Alasan Penghapus Pidana Karena Perintah Jabatan (Ambtelijk Bevel) Menurut Pasal 51 Ayat (1) Kitab Undang Undang Hukum Pidana. *Lex Crimen*, *10*(12), 45. Retrieved from https://ejournal.unsrat.ac.id/v3/index.php/lexcrimen/article/view/38534
33. Bhatnagar, S., Cotton, T., Brundage, M., Avin, S., Clark, J., Toner, H., … Amodei, D. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation Authors are listed in order of contribution Design Direction. *arXiv preprint arXiv:1802.07228*, (February 2018), 101. Retrieved from https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/1c6q2kc4v_50335.pdf
34. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., … Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. https://doi.org/10.1145/3351095.3372873
35. The European Parliament and of the Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 an. *Official Journal of the European Union*. Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
36. The European Parliament and of the Council. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). *Official Journal of the European Union*. Retrieved from https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng
37. The House of Representatives. H. r. 6580 To direct the Federal Trade Commission to require impact assessments of automated decision systems and augmented critical decision processes, and for other purposes. (2022). USA.
38. House of Commons of Canada. BILL C-27 An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts. , Ministre De L'innovation, Des Sciences Et De L'industrie 128–128 (2022). Canada.