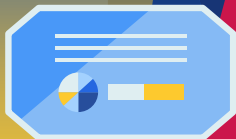
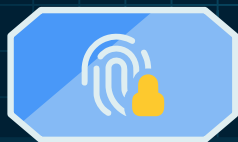


UNMASKING DECEPTIONS

Mendorong Hukum untuk Berperan
dalam Memitigasi dan Menanggulangi
Bahaya *Deepfake*





Penulis

M. Irfan Dwi Putra

Editor

Rizka Khairunissa Herdiani

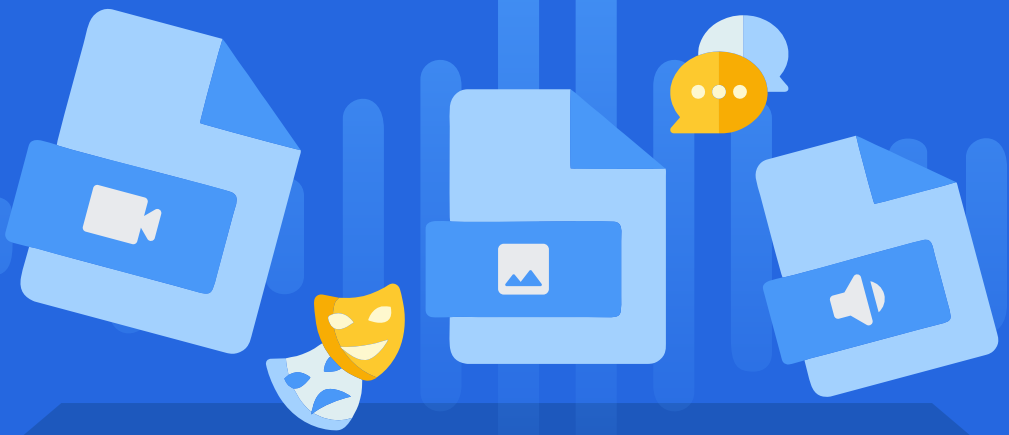
Desainer & Tata Letak

Ahmad Nur Hasan



Pada akhir Januari 2024, jagat dunia maya dihebohkan dengan kemunculan konten pornografi yang memuat gambar penyanyi pop papan atas Taylor Swift di aplikasi X—sebelumnya bernama Twitter. Konten tersebut menuai kecaman dari para Swifties—sebutan bagi para penggemar Taylor Swift—di seluruh dunia. Bahkan, Gedung Putih pun turut menanggapi kemunculan konten tersebut dengan menyebutnya sebagai fenomena yang mengkhawatirkan.¹ Sayangnya, setelah kemunculannya yang menghebohkan, X sebagai platform yang memuat konten tersebut tidak bersikap tanggap dalam melakukan penghapusan konten. Konten tersebut berseliweran di jagat X selama kurang lebih satu minggu dan dilihat sebanyak lebih dari 47 juta kali. Barulah setelah para penggemar Taylor Swift beramai-ramai melaporkan konten tersebut, konten tersebut akhirnya dihapus dari X.²

Beberapa bulan sebelumnya, tepatnya pada Oktober 2023, dunia maya Indonesia juga diramaikan dengan kemunculan video Presiden Joko “Jokowi” Widodo yang berpidato menggunakan bahasa Mandarin dengan fasih. Sekilas, secara visual tidak ada yang aneh dengan video tersebut. Namun, Kementerian Komunikasi dan Informatika (Kemenkominfo) telah menyatakan bahwa video tersebut merupakan hasil editan kecerdasan artifisial (KA).³ Sebelum video tersebut dinyatakan sebagai disinformasi, tidak sedikit orang yang percaya bahwa video tersebut asli. Dalam beberapa kasus, hal ini juga memunculkan sentimen-sentimen politik tertentu terhadap Presiden Jokowi.⁴



Kedua fenomena tersebut berakar dari satu permasalahan yang sama: *deepfake*. *Deepfake* adalah konten media sintetis berupa gambar, video, atau audio yang direkayasa secara digital di mana seseorang ditampilkan mengatakan atau melakukan suatu hal yang sebenarnya tidak pernah ada atau terjadi di dunia nyata.⁵ Terminologi *deepfake* berasal dari gabungan dua kata, yaitu *deep learning* dan *fake*. Terminologi ini pertama kali muncul pada tahun 2017 ketika seorang pengguna Reddit dengan nama pengguna Deepfakes mengunggah sejumlah video bermuatan pornografi yang menampilkan wajah artis terkenal seperti Scarlett Johansson dan Gal Gadot sebagai bintang porno. Sejak saat itu, terminologi ini digunakan untuk merujuk pada konten-konten manipulasi yang dihasilkan oleh KA. Seiring dengan kemajuan teknologi KA, *deepfake* menjadi semakin realistis dan sulit dibedakan dengan konten asli.⁶

Tidak dapat dimungkiri bahwa *deepfake* memunculkan ancaman baru bagi umat manusia. Pada Februari 2024 lalu, seorang staf keuangan di Hong Kong mengalami penipuan yang disebabkan oleh penggunaan *deepfake*. Ia mengirimkan uang sejumlah 25 juta dolar US kepada seorang penipu yang menggunakan teknologi *deepfake* untuk berpura-pura menjadi kepala keuangan perusahaan dalam suatu panggilan video.⁷ Pada kasus ini, *deepfake* digunakan sebagai sarana *social engineering* untuk melakukan *scam* atau penipuan finansial. Sementara itu, pada kasus-kasus penyebaran konten *non-consensual pornography* yang menimpa Taylor Swift dan Scarlett Johansson, *deepfake* digunakan sebagai sarana untuk menghancurkan reputasi, citra, nama baik, dan karir seseorang.⁸ Selain itu, terdapat ancaman-ancaman lain yang ditimbulkan oleh *deepfake*, di antaranya kampanye disinformasi, pencurian data pribadi, munculnya sentimen-sentimen negatif, hingga ancaman terhadap demokrasi.

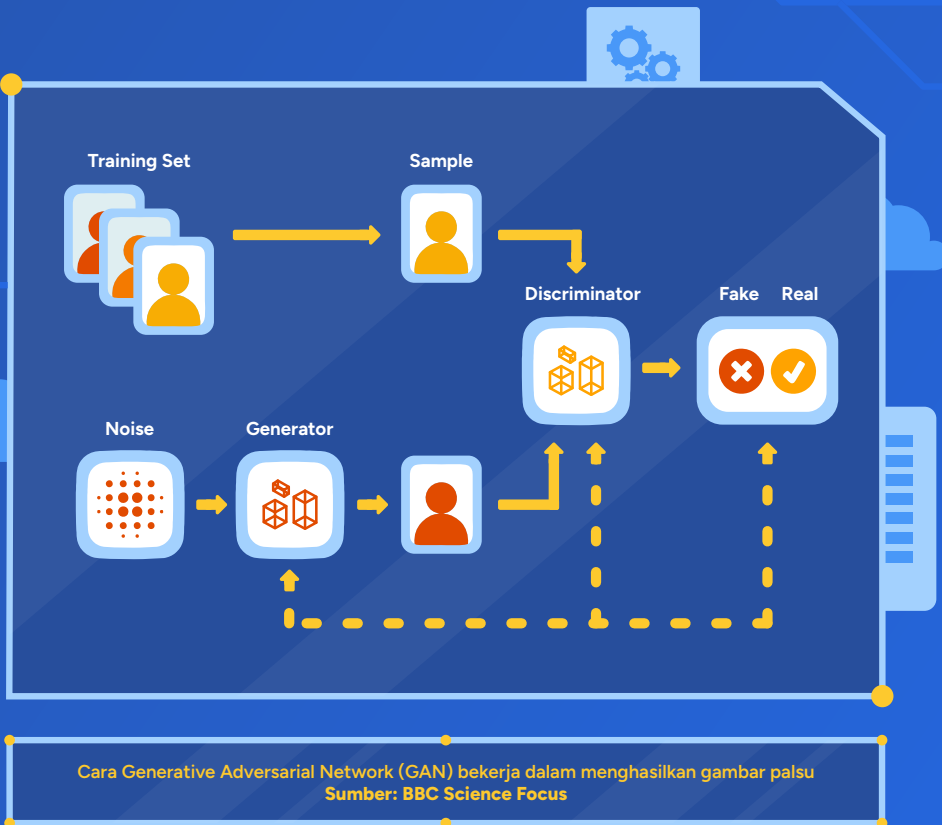
Masalah-masalah yang ditimbulkan oleh *deepfake* menjadi persoalan hukum yang harus diselesaikan. Hukum dalam hal ini memegang peranan yang sangat penting dalam meregulasi penggunaan *deepfake* serta menyelesaikan masalah yang ditimbulkan oleh *deepfake*. Tulisan ini akan membahas bagaimana *deepfake* mendistorsi realitas sosial politik masyarakat dan bagaimana seharusnya hukum berperan dalam memitigasi serta menanggulangi bahaya yang ditimbulkan oleh *deepfake*. Tulisan ini akan terdiri dari sejumlah pembahasan. Setelah pendahuluan, bagian kedua akan membahas tentang bagaimana *deepfake* mendistorsi realitas sosial politik masyarakat melalui penyebaran disinformasi dan *non-consensual pornography*. Bagian ketiga tulisan ini akan mendiskusikan tentang tiga pendekatan dalam memitigasi dan menanggulangi bahaya *deepfake*, yaitu teknis, edukasi, dan hukum dan mengapa pendekatan hukum menjadi suatu hal yang mendesak untuk diterapkan saat ini. Bagian terakhir akan memuat kesimpulan dari tulisan ini.



Deepfake, disinformasi, dan non-consensual pornography: Bagaimana KA mendistorsi realitas sosial politik masyarakat

Sebelum membahas tentang keterkaitan antara *deepfake*, disinformasi, dan *non-consensual pornography*, tulisan ini akan menjelaskan tentang bagaimana cara *deepfake* bekerja. *Deepfake* dihasilkan dari algoritma *machine learning* (ML) yang disebut sebagai *General Adversarial Networks* (GAN). GAN melibatkan proses *adversarial* dalam menghasilkan data palsu dengan mengkombinasikan dua jaringan saraf tiruan (*neural network*) yang disebut sebagai *generator* dan *discriminator*. *Generator* adalah jaringan saraf tiruan yang berfungsi untuk menghasilkan data palsu dari kumpulan data yang diinput ke dalamnya. Sementara itu, *discriminator* adalah jaringan saraf tiruan yang berfungsi untuk membedakan antara data palsu yang dihasilkan *generator* dengan data asli.⁹

Generator akan menghasilkan data palsu yang dibuat semirip mungkin dengan data asli berdasarkan data input yang dimasukkan ke dalam programnya. Kemudian, data palsu tersebut akan dideteksi oleh *discriminator* untuk membedakan antara data asli dengan data palsu. Setiap *discriminator* menemukan perbedaan dalam data palsu yang dihasilkan oleh *generator*, *discriminator* akan memberikan *feedback* kepada *generator* untuk menghasilkan data yang lebih autentik dan meyakinkan. Di sinilah *generator* dan *discriminator* akan bersaing satu sama lain sehingga proses ini disebut *adversarial*. Proses ini berlangsung secara berulang-ulang hingga *generator* menghasilkan data palsu yang terlihat sangat autentik dan meyakinkan sehingga tidak bisa dibedakan lagi oleh *discriminator* dengan data asli. Data palsu yang tidak bisa dibedakan kembali oleh *discriminator* itulah yang disebut sebagai *deepfake*.¹⁰



Terdapat beberapa jenis *deepfake* sesuai dengan konten yang ingin dihasilkan oleh pembuatnya, yaitu *deepfake* foto, audio, dan video. Pertama, *deepfake* foto, yaitu *deepfake* yang dapat menukar wajah atau bagian tubuh lain seseorang dengan wajah atau bagian tubuh orang lain. Kedua, *deepfake* audio, yaitu *deepfake* yang dapat mengubah atau meniru suara orang lain serta menghasilkan suara orang lain dengan menggunakan metode *text-to-speech*. Ketiga, *deepfake* video, yaitu *deepfake* yang dapat mengganti wajah seseorang dengan wajah orang lain dalam sebuah video dan mengubah gerakan mulut serta kata-kata yang diucapkan oleh seseorang dalam sebuah video dengan kata-kata lain.¹¹ *Deepfake pornography* Taylor Swift merupakan jenis *deepfake* yang pertama, sedangkan *deepfake* Presiden Jokowi yang berpidato menggunakan bahasa Mandarin adalah jenis *deepfake* yang ketiga.

Di era KA saat ini, kehadiran *deepfake* menciptakan distorsi yang luar biasa bagi realitas sosial politik masyarakat. Penggunaan KA untuk menghasilkan konten-konten *deepfake* membuka jalan bagi era baru disinformasi. Hal ini disebabkan karena satu hal: aksesibilitas terhadap KA yang semakin mudah dan terbuka. Jika dahulu pembuatan konten manipulasi hanya mungkin dilakukan oleh orang-orang tertentu saja yang menguasai aplikasi pengeditan gambar dan video, saat ini seseorang dapat melakukan manipulasi terhadap gambar atau video hanya dengan menginputnya ke dalam suatu aplikasi berbasis KA. Akibatnya, konten-konten manipulasi akan jauh lebih mudah untuk dibuat dan kuantitasnya pun akan jauh lebih meningkat dibandingkan dengan era-era sebelumnya. Saat inilah era disinformasi baru dimulai di mana setiap orang dapat memalsukan konten dan menyebarkannya ke publik dengan narasi seolah-olah konten tersebut autentik. Terlebih lagi, demokratisasi media sosial juga membuka jalan yang lebih lebar bagi disinformasi berbasis *deepfake* tersebut untuk mencapai audiens yang lebih luas dan beragam.¹²

Kemudian, semakin realistiknya konten yang dihasilkan *deepfake* juga turut mendistorsi bagaimana masyarakat memandang suatu kenyataan. Masyarakat akan sulit untuk membedakan antara konten yang palsu dan yang asli karena batasannya yang semakin kabur. Hal ini akan mendorong lahirnya kepercayaan palsu di mana orang akan mempercayai suatu kebohongan sebagai sesuatu yang benar.¹³ Dampak ini akan lebih terasa pada masyarakat yang memiliki tingkat literasi digital yang rendah. Tanpa adanya kemampuan yang baik dalam membedakan konten asli dengan konten palsu, mereka akan cenderung mempercayai konten *deepfake* sebagai konten yang autentik.

Bahkan, suatu penelitian mengungkap bahwa pada orang-orang yang dapat mendeteksi konten *deepfake* sekalipun, *deepfake* tetap memiliki implikasi negatif. Implikasi negatif tersebut adalah munculnya keraguan dalam diri mereka tentang apakah konten tersebut asli atau palsu. Keraguan tersebut berkorelasi dengan kepercayaan terhadap berita yang beredar di media sosial. Singkatnya, semakin besar paparan seseorang terhadap konten *deepfake*, semakin meningkat pula keraguan dalam diri orang tersebut. Kemudian, keraguan ini akan berdampak pada menurunnya tingkat kepercayaan orang tersebut terhadap berita yang beredar di media sosial mengingat penyebaran *deepfake* banyak terjadi di media sosial.¹⁴

Dalam konteks politik, *deepfake* dapat membawa ancaman bagi keberlangsungan politik elektoral maupun demokrasi secara umum. Sebuah riset membuktikan bahwa konten video *deepfake* yang menampilkan seorang figur politik yang mengucapkan kalimat tertentu dapat memengaruhi persepsi dan preferensi politik masyarakat terhadap figur tersebut.¹⁵ Dalam konteks ini, *deepfake* digunakan untuk memanipulasi opini publik yang dapat memunculkan sentimen-sentimen tertentu terhadap figur yang bersangkutan. Misalnya, dalam kasus video *deepfake* Presiden Jokowi di atas, muncul sentimen bahwa Presiden Jokowi pro terhadap Tiongkok dan sentimen-sentimen lainnya.

Selain distorsi pada realitas sosial politik masyarakat akibat penyebaran disinformasi, *deepfake* juga dapat mengancam kehidupan pribadi dan sosial seseorang melalui penyebaran konten pornografi yang dihasilkan dan dimanipulasi oleh teknologi yang disebut sebagai *deepfake pornography*. Sebuah data menunjukkan bahwa dari seluruh konten video *deepfake* yang beredar, 95 persen di antaranya adalah *deepfake pornography*. *Deepfake pornography* umumnya menyasar kalangan selebritis atau figur publik yang gambar atau videonya dapat diakses dengan mudah.¹⁶ Namun, seiring dengan semakin umumnya penggunaan media sosial di tengah masyarakat, orang umum pun banyak yang menjadi korbannya. Pelaku umumnya mengambil gambar korban yang dapat diakses secara publik dan kemudian memanipulasinya menjadi konten pornografi.¹⁷

Meskipun terdapat kasus yang korbannya adalah laki-laki, mayoritas korban *deepfake pornography* adalah perempuan. Melalui *deepfake pornography*, perempuan dikonstruksikan sebagai objek seksual yang dapat dieksploitasi berdasarkan fantasi pencipta konten tersebut. Pada pertengahan tahun 2023, isu *deepfake pornography* menjadi sorotan di media sosial X setelah

seorang pengguna dengan nama @itsindahg mengunggah cuitan tentang

penyebaran formulir daring terkait *deepfake pornography*. Formulir daring tersebut berasal dari grup obrolan daring di Telegram yang memang kerap menyebarkan konten-konten *image-based sexual abuse*, termasuk salah satunya *deepfake pornography*.¹⁸ Hal ini tentu saja memunculkan kekhawatiran—khususnya bagi perempuan—karena ruang digital tidak lagi aman bagi mereka, bahkan ketika mereka tidak melakukan apapun.



Sebagai serangan terhadap privasi dan seksualitas seseorang, penyebaran *deepfake pornography* menimbulkan dampak yang sangat merugikan bagi korbannya, baik berupa dampak psikologis maupun sosial. Secara psikologis, korban *deepfake pornography* akan merasakan tekanan sosial yang begitu besar dari orang-orang di sekitarnya ketika konten pornografi yang memuat dirinya beredar di media sosial. Tidak jarang pula orang justru menyalahkan korban (*victim blaming*) atas penyebaran konten *deepfake pornography* tersebut. Padahal, korban pun kemungkinan besar tidak tahu menahu mengenai bagaimana dirinya bisa termuat dalam konten tersebut. Tekanan sosial ini akan menambah beban psikologis korban yang dapat membuat korban menderita depresi dan dalam sejumlah kasus berujung bunuh diri.¹⁹

Secara sosial, konten *deepfake pornography* akan membuat citra, reputasi, nama baik, dan karir korban hancur. Selain itu, korban *deepfake pornography* juga akan mengalami kesulitan dalam memperoleh pekerjaan karena pemberi kerja umumnya akan melakukan pencarian informasi secara *online* tentang kandidat pekerjaanya. Ketika mereka menemukan konten pornografi yang memuat kandidat pekerjaanya, mereka kemungkinan tidak akan menerima kandidat tersebut di perusahaannya.²⁰ Begitu merugikannya dampak yang ditimbulkan dari *deepfake* membuat mitigasi dan penanggulangannya menjadi hal yang sangat mendesak untuk dilakukan. Mitigasi dan penanggulangan bahaya *deepfake* harus dilakukan dengan efektif agar dampak negatif yang ditimbulkannya tidak semakin meluas dan merugikan.



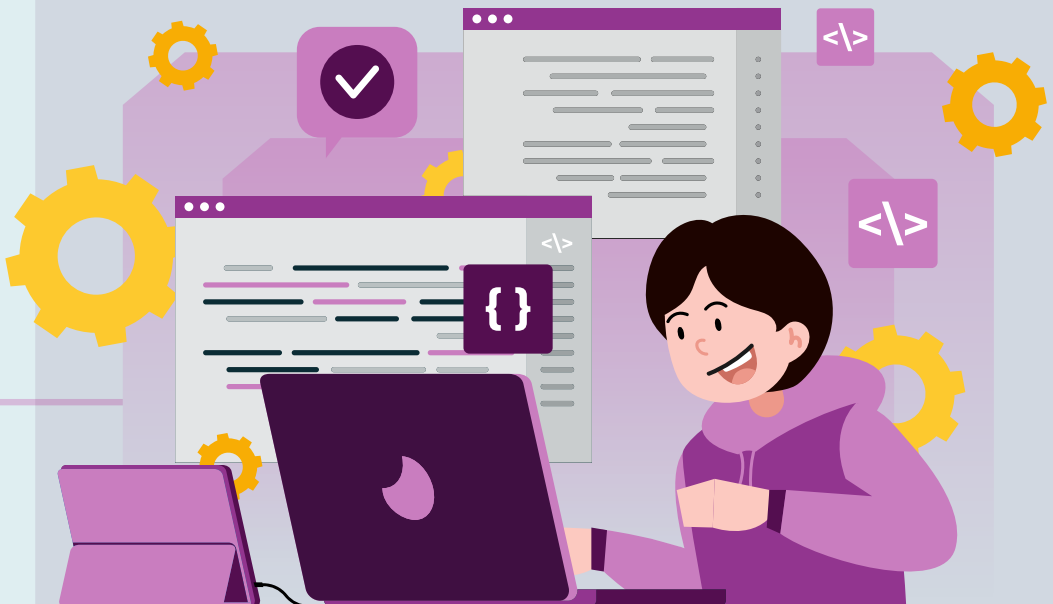
Tiga pendekatan dalam memitigasi dan menanggulangi bahaya *deepfake*

Sejauh ini, terdapat tiga pendekatan yang umum dikenal dalam memitigasi dan menanggulangi bahaya *deepfake*, yaitu teknis, edukasi, dan hukum. Pendekatan teknis adalah pendekatan yang melibatkan penggunaan teknologi untuk mendeteksi suatu konten *deepfake*. Terdapat sejumlah teknik yang dikembangkan oleh para ahli komputer dalam mendeteksi konten *deepfake*. Pertama, teknik deteksi berbasis jejak yang diusulkan oleh Li dan Lyu pada tahun 2019. Pada saat itu, algoritma *deepfake* hanya dapat menghasilkan gambar dengan resolusi yang terbatas sehingga gambar tersebut perlu ditransformasikan lebih lanjut agar dapat mirip dengan wajah asli yang akan diganti dalam video original. Transformasi ini menyisakan jejak tertentu dalam konten *deepfake* yang dihasilkan yang dapat ditangkap dengan baik oleh *convolutional neural networks* (CNNs). Jejak tersebutlah yang membuktikan bahwa konten yang dideteksi merupakan *deepfake*.²¹

Kedua, teknik deteksi berbasis inkonsistensi yang dikembangkan untuk mendeteksi ada atau tidaknya inkonsistensi dalam konten *deepfake*. Sejumlah penelitian menemukan teknik dalam mendeteksi tanda-tanda *deepfake* video, misalnya adanya pola kedipan yang tidak wajar, distorsi pada tampilan wajah, inkonsistensi keseluruhan gambar dalam video, dan sebagainya.²² Selain kedua teknik di atas, ada teknik-teknik lain yang dikembangkan untuk mendeteksi *deepfake* seperti teknik deteksi berbasis forensik, deteksi berbasis *deep neural network* (DNN), *GAN Redesign-based Detection*, dan lain-lain.²³ Meskipun memiliki peran penting dalam memitigasi dan menanggulangi bahaya *deepfake*, pendekatan teknis tidak akan efektif untuk digunakan dalam jangka panjang dan umumnya hanya akan bertahan sementara saja. Sebab, semakin canggih teknik deteksi yang digunakan akan membuat semakin canggih pula aplikasi *deepfake* juga dalam menciptakan konten-konten palsu. Hal ini hanya akan menciptakan permainan kucing-kucingan antara *deepfake* dan deteksinya yang tidak akan menemukan titik penyelesaian.²⁴

Pendekatan kedua adalah pendekatan edukasi, yaitu pendekatan yang melibatkan peningkatan literasi media dan digital dalam masyarakat agar dapat membedakan konten *deepfake* dengan konten asli. Pendekatan ini dapat diberikan melalui edukasi kepada masyarakat tentang apa itu konten *deepfake*, bagaimana cara konten *deepfake* diciptakan, apa bahaya yang ditimbulkan dari konten *deepfake*, bagaimana cara-cara membedakan konten *deepfake* dengan konten asli, dan sebagainya. Dengan dimilikinya pengetahuan dan kemampuan untuk mendeteksi konten *deepfake*, kemungkinan masyarakat untuk menyebarkan konten tersebut akan menurun sehingga dampak negatif disinformasi yang ditimbulkan juga akan berkurang.²⁵

Meskipun demikian, sama halnya dengan pendekatan teknis, pendekatan edukasi melalui peningkatan literasi media dan digital juga memiliki kelemahan. Agar benar-benar sukses dilaksanakan dan memberikan dampak, pendekatan ini akan memakan proses waktu yang panjang dan biaya yang besar. Proses yang panjang tersebut tidak sejalan dengan pesatnya kemajuan KA yang berpotensi menghasilkan konten-konten *deepfake* yang mustahil untuk dibedakan lagi menggunakan kemampuan inderawi. Hal ini akan membuat pengetahuan dan keterampilan dalam mendeteksi konten *deepfake* menjadi kurang relevan dari waktu ke waktu.



Pendekatan ketiga adalah pendekatan hukum, yaitu suatu pendekatan yang melibatkan hukum sebagai alat rekayasa sosial untuk meregulasi penggunaan *deepfake* di tengah masyarakat. Ada dua pendekatan hukum yang umum digunakan dalam meregulasi teknologi secara umum, yaitu pendekatan berbasis prinsip melalui *soft law* dan pendekatan berbasis aturan melalui *hard law*. Dalam konteks penggunaan KA, pendekatan pertama umumnya diaplikasikan dengan menyusun panduan etika dalam penggunaan KA. Pendekatan ini lebih menekankan pada ekspektasi yang ingin dicapai dengan menetapkan standar-standar tertentu yang harus dipenuhi oleh subjek yang diatur. Akan tetapi, pelaksanaan dan interpretasi akan standar-standar tersebut diserahkan kepada masing-masing subjek yang diatur dalam regulasi tersebut.²⁶

Di satu sisi, pendekatan berbasis prinsip memiliki kelebihan karena fleksibilitas standar yang ditetapkan membuatnya tetap relevan dengan perkembangan teknologi yang cepat. Meskipun demikian, pendekatan tersebut memiliki kekurangan yang membuatnya kurang tepat untuk diterapkan dalam merespons isu bahaya *deepfake*. Pendekatan semacam ini tidak memiliki mekanisme kontrol yang kuat untuk memastikan bahwa subjek yang diatur akan memenuhi standar-standar yang ditetapkan. Jika dikontekstualisasikan ke dalam isu *deepfake*, penetapan prinsip—misalnya—untuk tidak membahayakan orang lain (*do no harm*) tidak cukup untuk memastikan bahwa orang tidak akan membahayakan orang lain menggunakan *deepfake*.

Pendekatan hukum yang kedua adalah pendekatan berbasis aturan melalui *hard law*. Pendekatan ini lebih menekankan pada penggunaan aturan-aturan yang bersifat preskriptif dan berorientasi pada kejelasan serta kepastian hukum. Pendekatan ini umumnya dilakukan melalui pembentukan regulasi yang memuat ketentuan-ketentuan spesifik yang wajib atau dilarang dilakukan oleh subjek yang diatur. Untuk memastikan kepatuhan subjek yang diatur, pendekatan ini juga umumnya menyertakan sanksi—baik administratif maupun pidana—atas pelanggaran-pelanggaran terhadap ketentuan yang termuat di dalamnya.²⁷

Pendekatan ini memiliki kelebihan karena terdapat mekanisme kontrol yang kuat untuk memastikan bahwa subjek yang diatur akan mematuhi ketentuan-ketentuan yang diatur, terlebih apabila pendekatan ini menyertakan sanksi di dalamnya. Namun, pendekatan ini juga memiliki kekurangan jika digunakan untuk meregulasi aspek-aspek teknologi karena sifatnya yang cenderung kaku dan sulit untuk mengikuti perkembangan teknologi yang pesat.²⁸ Meskipun demikian, suatu regulasi berbasis aturan dapat diciptakan sedemikian rupa untuk mengatur *deepfake* dengan tetap memperhatikan relevansi dan keberlakuannya untuk masa yang akan datang.



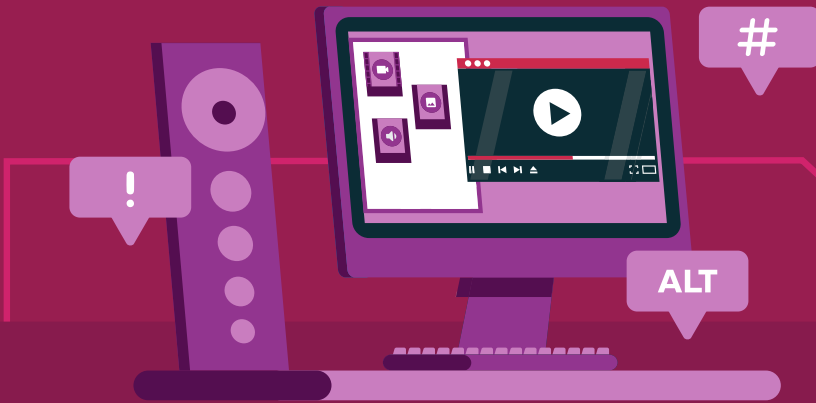
Dalam konteks di Indonesia, sejatinya terdapat sejumlah aturan yang beririsan dengan penggunaan *deepfake* sebagaimana dijelaskan pada bagian sebelumnya, di antaranya adalah Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi (UU PDP), Undang-Undang Nomor 12 Tahun 2022 tentang Tindak Pidana Kekerasan Seksual (UU TPKS), Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE) beserta kedua perubahannya, Undang-Undang Nomor 44 Tahun 2008 tentang Pornografi, dan Undang-Undang Nomor 1 Tahun 2023 tentang Kitab Undang-Undang Hukum Pidana yang baru akan berlaku secara efektif pada tahun 2026 mendatang.

Meskipun demikian, penerapan aturan-aturan di atas dalam merespons isu *deepfake* sebagian besar hanya bersifat kuratif. Dengan kata lain, regulasi-regulasi tersebut baru akan berlaku ketika muncul kasus-kasus *deepfake* yang merugikan. Misalnya, dalam konteks UU PDP, terdapat larangan memalsukan data pribadi yang dapat berakibat pada kerugian orang lain dengan ancaman hukuman maksimal 6 tahun penjara dan/atau denda maksimal 6 miliar rupiah.²⁹ Atau misalnya dalam konteks *deepfake pornography* yang masuk ke ranah kekerasan seksual. UU TPKS hanya menentukan bahwa kekerasan seksual yang menggunakan sarana elektronik akan dikenakan pidana yang ditambah sepertiga dari pidana awal.³⁰ Kedua peraturan tersebut hanya merespons isu *deepfake* ketika teknologi tersebut telah menimbulkan kerugian dan korban. Belum ada regulasi di Indonesia yang mengatur tentang bagaimana mekanisme yang sepatutnya digunakan untuk mencegah agar dampak yang ditimbulkan *deepfake* tidak semakin masif dan dapat ditanggulangi sejak awal.



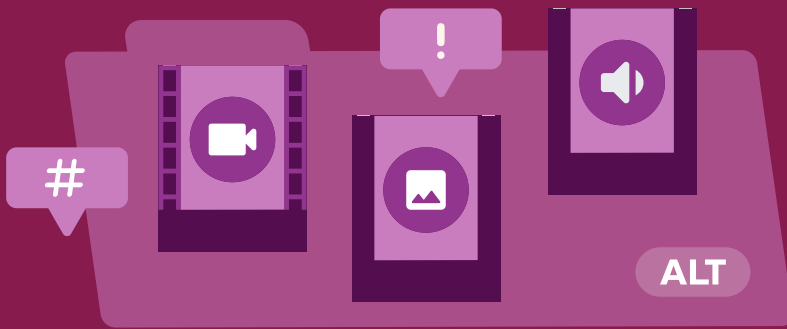
Di negara-negara lain, mekanisme hukum yang banyak diadopsi untuk meregulasi *deepfake* adalah kebijakan pengungkapan (*disclosure policy*). Kebijakan pengungkapan merupakan salah satu bagian dari penerapan prinsip transparansi yang umum dalam etika KA. Dalam konteks *deepfake*, sejumlah negara mewajibkan penerapan kebijakan pengungkapan dengan menyatakan bahwa konten *deepfake* sebagai konten yang dihasilkan oleh KA. Dengan kata lain, konten *deepfake* wajib dilabeli dengan keterangan bahwa konten tersebut adalah konten yang telah dimanipulasi secara artifisial. Kebijakan ini salah satunya diadopsi oleh Uni Eropa dalam EU AI Act yang baru saja disepakati oleh Parlemen Uni Eropa pada 13 Maret 2024 lalu.³¹

Dalam Pasal 50 ayat (4) EU AI Act, terdapat ketentuan yang mewajibkan penyedia sistem KA yang menghasilkan gambar, audio, atau video yang dimanipulasi oleh KA (*deepfake*) untuk mengungkapkan bahwa konten tersebut telah dibuat atau dimanipulasi secara artifisial.³² Kemudian, pada bagian *Recital 134* dijelaskan bahwa kewajiban pengungkapan tersebut dilakukan dengan memberikan label pada konten *deepfake* tersebut dan mengungkapkan konten original yang telah dimanipulasi.³³ Tiongkok, sebagai salah satu negara terdepan dalam pengembangan KA juga telah memiliki kebijakan serupa yang bahkan lebih dulu disahkan dibandingkan EU AI Act. Di dalam kebijakan tersebut, Tiongkok melarang publikasi konten *deepfake* tanpa adanya pengungkapan yang wajar bahwa konten tersebut adalah hasil manipulasi KA.³⁴ Di Amerika Serikat, rancangan undang-undang yang mengatur tentang *deepfake* juga diusulkan ke Kongres pada tahun 2023 yang di dalamnya memuat ketentuan mengenai kewajiban pengungkapan sebagai bagian dari penerapan prinsip transparansi.³⁵



Kewajiban pengungkapan ini merupakan mekanisme kontrol yang baik untuk memitigasi dampak *deepfake* sejak dini. Dengan adanya mekanisme ini, aplikasi penghasil konten *deepfake* akan diwajibkan untuk memberikan label pada konten yang dihasilkan. Dengan demikian, orang yang melihat konten tersebut akan mengetahui bahwa konten tersebut merupakan hasil manipulasi KA dan bukan konten orisinal. Hal ini tentunya akan meminimalisasi dampak-dampak yang jauh lebih merugikan dari *deepfake* seperti meluasnya disinformasi. Mekanisme ini seharusnya dapat diterapkan juga di Indonesia dengan mewajibkan adanya pelabelan pada konten-konten *deepfake* sehingga masyarakat tidak akan tertipu dengan konten palsu tersebut dan tidak akan terdorong untuk membagikan konten *deepfake* tersebut kepada orang lain sehingga dampak yang lebih luas bisa diminimalisasi.

Terkait dengan *deepfake pornography*, hukum juga perlu berperan dalam menanggulangi penyebaran konten pornografi manipulatif yang beredar di platform media sosial. Berkaca dari lambatnya X dalam merespons kasus Taylor Swift, hukum seharusnya dapat membebani platform dengan pertanggungjawaban untuk memastikan bahwa ruangnya aman bagi semua orang. Imunitas platform dari pertanggungjawaban hanya akan membuat korban-korban *deepfake pornography* semakin bertambah. Salah satu cara pertanggungjawaban platform adalah dengan mewajibkan mereka untuk menghapus konten *deepfake pornography* dari jaringan mereka langsung setelah mereka mendeteksi konten tersebut sebagai konten pornografi yang dibuat oleh KA. Selain itu, hukum juga seharusnya memberikan mekanisme yang jelas dan mudah bagi korban untuk menggugat platform atas kegagalannya menghapus konten *deepfake pornography* yang merugikan dirinya.³⁶



Kesimpulan

Kasus-kasus yang melibatkan *deepfake* merefleksikan gambaran bahwa teknologi tidak semata-mata membawa manfaat bagi umat manusia, tetapi juga kerugian. *Deepfake* menjadi distorsi yang luar biasa bagi realitas sosial politik umat manusia abad ini dengan digunakannya *deepfake* sebagai sarana untuk menyebar disinformasi. Bahkan, *deepfake* juga dapat mengancam keselamatan dan nyawa manusia ketika digunakan untuk menyebarkan *non-consensual pornography*. Kompleksnya persoalan *deepfake* tentunya membutuhkan pemecahan masalah yang ideal agar dampak yang ditimbulkannya dapat dimitigasi dan ditanggulangi agar tidak menjadi semakin merugikan.

Dari ketiga pendekatan yang ada untuk memitigasi dan menanggulangi dampak *deepfake*—teknis, edukasi, dan hukum, menggabungkan ketiga pendekatan tersebut tentunya menjadi pemecahan masalah yang ideal. Namun, mengingat mendesaknya permasalahan yang harus diselesaikan, pendekatan hukum merupakan pemecahan masalah yang paling tepat untuk diterapkan mengingat kemampuannya sebagai alat rekayasa sosial dan mekanisme kontrolnya yang kuat. Pendekatan hukum ini dapat dilakukan melalui mekanisme kebijakan pengungkapan yang wajib dilakukan oleh penyedia sistem KA dengan melabeli konten *deepfake* dan membebankan pertanggungjawaban platform dalam konteks *deepfake pornography*. Dengan adanya pelabelan dan pertanggungjawaban platform tersebut, dampak-dampak yang lebih merugikan dari *deepfake* setidaknya dapat dimitigasi dan ditanggulangi.

Daftar Pustaka

- ¹ Witt, A. dan Henry, N. (2024). 'Taylor Swift deepfakes: new technologies have long been weaponised against women. The solution involves us all' Tersedia pada: <https://theconversation.com/taylor-swift-deepfakes-new-technologies-have-long-been-weaponised-against-women-the-solution-involves-us-all-222268>. Diakses pada 12 Maret 2024.
- ² Saner, E. (2024) 'Inside the Taylor Swift Deepfake Scandal: "It's Men Telling a Powerful Woman to Get Back in Her Box"' Tersedia pada: www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box. Diakses pada 12 Maret 2024.
- ³ Afifa, L. (2023) 'Kominfo Ministry Asserts Jokowi's Speech in Chinese Produced by Deepfake' Tersedia pada: en.tempo.co/read/1789306/kominfo-ministry-asserts-jokowis-speech-in-chinese-produced-by-deepfake. Diakses pada 12 Maret 2024.
- ⁴ Dewi, I.R. (2023) 'Video Pidato Mandarin Jokowi Telan Korban, Kominfo Bongkar' Tersedia pada: www.cnbcindonesia.com/tech/20231208094512-37-495581/video-pidato-mandarin-jokowi-telan-korban-kominfo-bongkar. Diakses pada 12 Maret 2024.
- ⁵ Mustak, M. et al. (2023) 'Deepfakes: Deceptions, Mitigations, and Opportunities', *Journal of Business Research*, 154, pp. 1-15. doi:<https://doi.org/10.1016/j.jbusres.2022.113368>.
- ⁶ Dasilva, J.P., Ayerdi, K.M., dan Galdospin, T.M. (2021) 'Deepfakes on Twitter: Which Actors Control Their Spread?', *Media and Communication*, 9(1), pp. 301-312. <https://doi.org/10.17645/mac.v9i1.3433>.
- ⁷ Chen. H. dan Magramo, K. (2024) 'Finance worker pays out \$25 million after video call with deepfake 'chief financial officer' Tersedia pada: <https://www.google.com/amp/s/amp.cnn.com/cnn/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>. Diakses pada 13 Maret 2024.

- ⁸ Pascale, E. (2023) 'Deeply Dehumanizing, Degrading, and Violating: Deepfake Pornography and the Path to Legal Recourse', *Syracuse Law Review*, 73, pp. 335-366.
- ⁹ Yadav, D. dan Salmani, S. (2019) 'Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network', *Proceedings of the International Conference on Intelligent Computing and Control Systems*, pp. 852-857.
- ¹⁰ Ibid.
- ¹¹ Masood, M. et al. (2023) 'Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward', *Applied Intelligence*, 53, pp. 3974-4026. <https://doi.org/10.1007/s10489-022-03766-z>.
- ¹² Gosse, C. dan Burkell, J. (2020) 'Politics and porn: how news media characterizes problems presented by deepfakes', *Critical Studies in Media Communication*, 37(5), pp. 497-511. <https://doi.org/10.1080/15295036.2020.1832697>.
- ¹³ Ibid.
- ¹⁴ Vaccari, C. dan Chadwick, A. (2020) 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News', *Social Media + Society*, 6(1), pp. 1-13. doi:10.1177/2056305120903408.
- ¹⁵ Ray, A. (2021) 'Disinformation, Deepfakes and Democracies: The Need for Legislative Reform', *UNSW Law Journal*, 44(3), pp. 983-1013. <https://doi.org/10.53637/DELS2700>.
- ¹⁶ Story, D. dan Jenkins, R. (2023) 'Deepfake Pornography and the Ethics of Non-Veridical Representations', *Philosophy & Technology*, 56, pp. 36-56. <https://doi.org/10.1007/s13347-023-00657-0>.
- ¹⁷ Solichah, I.I., Sulistio, F., dan Istiqomah, M. (2023) 'Protection of Victims of Deep Fake Pornography in a Legal Perspective in Indonesia', *International Journal of Multicultural and Multireligious Understanding*, 10(1), pp. 383-390. <http://dx.doi.org/10.18415/ijmmu.v10i1.4409>.
- ¹⁸ Pertiwi, S.P. (2023) 'Waspada KBGO, Ada Grup Obrolan Bikin Deepfake Porn via Formulir Daring' Tersedia pada <https://www.konde.co/2023/07/darurat-kbgo-ada-temuan-group-chat-dan-permintaan-deepfake-porn-lewat-google-form/>. Diakses pada 21 Maret 2024.

¹⁹ Pascale, E. (2023) 'Deeply Dehumanizing, Degrading, and Violating: Deepfake Pornography and the Path to Legal Recourse', *Syracuse Law Review*, 73, pp. 335-366.

²⁰ Ibid.

²¹ Li, Y. dan Lyu, S. (2019) 'Exposing DeepFake Videos By Detecting Face Warping Artifacts', *CVPR Workshop*, pp. 46-52.

²² Engler, A. (2019) 'Fighting deepfakes when detection fails' Tersedia pada: <https://www.brookings.edu/articles/fighting-deepfakes-when-detection-fails/>. Diakses pada 14 Maret 2024.

²³ Zobaed, S. et al. (2021) 'DeepFakes: Detecting Forged and Synthetic Media Content Using Machine Learning', dalam Montasari, R. dan Jahankhani, H. (ed.) *Artificial Intelligence in Cyber Security: Impact and Implications. Advanced Sciences and Technologies for Security Applications*. Cham: Springer. https://doi.org/10.1007/978-3-030-88040-8_7.

²⁴ Farid, H. (2022) 'Creating, Using, Misusing, and Detecting Deep Fakes', *Journal of Online Trust and Safety*, 1(4), pp. 1-33. <https://doi.org/10.54501/jots.v1i4.56>.

²⁵ Hwang Y., Ryu J.Y., dan Jeong S.H. (2021) 'Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education', *Cyberpsychology, Behavior, and Social Networking*, 24(3), pp. 188-193. doi: 10.1089/cyber.2020.0174.

²⁶ Black, J., Hopper, M., dan Band, C. (2007) 'Making a success of Principle-based regulation', *Law and Financial Markets Review*, 1(3), pp. 191-206. <https://doi.org/10.1080/17521440.2007.11427879>.

²⁷ Decker, C. (2018) 'Goals-based and Rules-based Approaches to Regulation', *BEIS Research Paper*, 8, pp. 1-67. <https://ssrn.com/abstract=3717739>.

²⁸ Ibid.

²⁹ Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi, Pasal 66 jo. Pasal 68.

³⁰ Undang-Undang Nomor 12 Tahun 2022 tentang Tindak Pidana Kekerasan Seksual, Pasal 15 ayat (1) huruf l.

- ³¹ Gilchrist, K. dan lordache, R. (2024) 'World's first major act to regulate AI passed by European lawmakers' Tersedia pada <https://www.cnn.com/2024/03/13/european-lawmakers-endorse-worlds-first-major-act-to-regulate-ai.html>. Diakses pada 17 Maret 2024.
- ³² EU Artificial Intelligence Act, Pasal 50 ayat (4).
- ³³ EU Artificial Intelligence Act, Recital 134.
- ³⁴ Statt, N. (2019) 'China makes it a criminal offense to publish deepfakes or fake news without disclosure' Tersedia pada <https://www.theverge.com/2019/11/29/20988363/china-deepfakes-ban-internet-rules-fake-news-disclosure-virtual-reality>. Diakses pada 17 Maret 2024.
- ³⁵ Ramstack, T. (2023) 'Congress Wants Laws to Confront Deepfakes That Deceive Consumers' Tersedia pada <https://www.thewellnews.com/technology/congress-wants-laws-to-confront-deepfakes-that-deceive-consumers/>. Diakses pada 17 Maret 2024.
- ³⁶ Gieseke, A.P. (2020) "The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography', *Vanderbilt Law Review*, 73(5), pp. 1479-1516.





Center for Digital Society

Fakultas Ilmu Sosial dan Ilmu Politik
Universitas Gadjah Mada
Gedung BC Lt. 2, Ruang BC 201-203
Jalan Socio Yustisia 1
Bulaksumur, Yogyakarta, 55281, Indonesia

Telepon: (0274) 563362, Ext. 116
Email: cfds.fisipol@ugm.ac.id
Situs Web: cfds.fisipol.ugm.ac.id



facebook.com/cfdsugm



Center for Digital Society (CfDS)



[cfds_ugm](https://www.instagram.com/cfds_ugm)



[@cfds_ugm](https://twitter.com/cfds_ugm)



[@cfds_ugm](https://x.com/cfds_ugm)



[CfDS UGM](https://www.youtube.com/CfDSUGM)