BAB II

TEKNOLOGI *DEEPFAKE* SEBAGAI POTENSI ANCAMAN BARU DALAM DUNIA SIBER

Pembahasan penelitian pada bab ini menjelaskan gambaran umum tentang dampak atau bahaya *deepfake* terhadap identitas nasional yang dibagi ke dalam 4 sub bab. Pertama, pembahasan pada bab ini dimulai dengan penjelasan peningkatan kasus penyebaran konten *deepfake*. Adanya peningkatan kasus penyebaran konten *deepfake* menunjukkan bagaimana teknologi AI ini meningkat secara pesat dan dapat diakses dengan mudah untuk menyebarkan disinformasi. Kedua, bagian berikutnya memaparkan berbagai teknik manipulasi audiovisual sebelum hadirnya teknologi *deepfake*. Ketiga, bagian ini menjelaskan terkait implikasi negatif *deepfake*, khususnya terhadap keberlangsungan proses politik suatu negara yang berakibat pada menurunnya kepercayaan publik dan integritas pemilu. Keempat, sub bab ini mengulas tentang aksi-aksi maupun regulasi internasional dan regional yang telah diterapkan di berbagai negara dalam menangani ancaman AI khususnya *deepfake*.

2.1 Peningkatan Kasus Penyebaran Konten *Deepfake*

Pada 3 abad terakhir, dunia siber atau *cyberspace* telah menjadi isu yang penting dan utama dalam konflik internasional. Hadirnya transformasi teknologi mendasari faktor yang menyebabkan perkembangan isu tersebut pada skala internasional. Di dunia siber, konflik internasional terjadi dalam zona abu-abu dan seringkali berkaitan dengan informasi, data, serta manipulasi yang berakhir pada tindakan spionase, sabotase, dan subversi. Melalui pemahaman ini, negara-negara

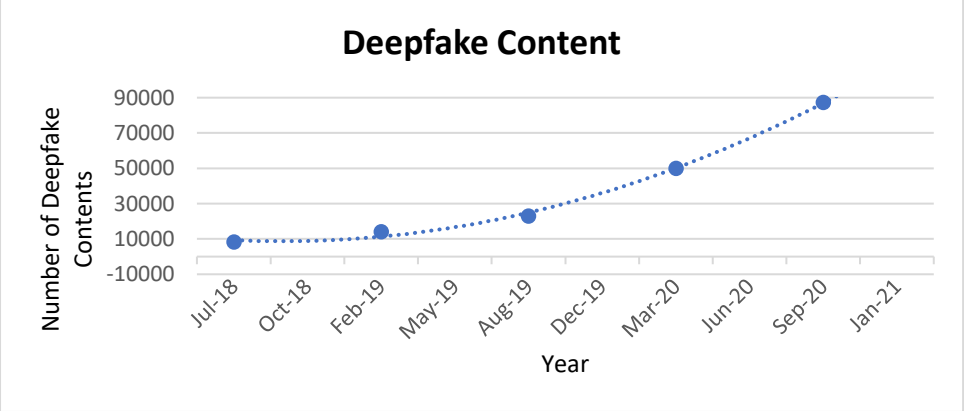
di dunia menganggap masalah siber sebagai tantangan terhadap keamanan nasional dan aktif menunjukkan kepentingan nasionalnya dengan strategi defensif serta operasi ofensif di dunia siber. Perkembangan teknologi siber menjadi isu yang vital serta perlu diwaspadai dikarenakan semakin mudahnya dalam melakukan penyebaran data dan informasi. Oleh karenanya, pengembangan dan penerapan teknologi kecerdasan buatan (AI) telah menjadi diskursus dalam konteks konflik internasional di dunia siber (Cristiano, et all (Eds.), 2023).

Deepfake menjadi salah satu teknologi AI yang perkembangannya cukup berpengaruh di dunia siber karena berkaitan erat dengan kegiatan manipulasi. Terminologi "deepfake" pertama kali digunakan pada tahun 2017 setelah munculnya video *face swap* hasil manipulasi teknologi AI (Paris & Donovan, 2019). *Deepfake* sangat bergantung pada algoritma *machine learning* untuk membuat sebuah video yang terlihat "asli" dengan menukar wajah seseorang ke dalam video yang digunakan (Vaccari & Chadwick, 2020). Penggunaan teknologi *deepfake* umumnya dapat dikatakan netral karena digunakan untuk tujuan yang positif seperti menciptakan sebuah avatar atau *virtual assistant* untuk meningkatkan kualitas pengalaman konferensi video. *Deepfake* juga dapat digunakan dalam produksi pertunjukan TV dan film untuk "menciptakan" kembali penampilan selebriti yang telah meninggal sebagai bentuk penghormatan terhadap mereka (Gambín, et all, 2024).

Sekalipun teknologi *deepfake* dapat digunakan untuk berbagai hal yang positif, teknologi ini nyatanya lebih banyak digunakan dalam konteks negatif. Konten *deepfake* mayoritas dibuat tanpa persetujuan atau konsen yang mutual dari

seseorang yang difiturkan. Dalam hal ini, objektivitas penggunaan *deepfake* digunakan untuk kegiatan manipulasi. Penekanan akan konsep "realitas" yang dibuat dengan teknologi ini memudahkan seseorang dalam melakukan aksinya untuk memanipulasi konten seiring dengan mudahnya percepatan informasi. Penggunaan *deepfake* menjadi atensi yang perlu diperhatikan karena tidak hanya perusahaan media saja yang dapat memiliki akses, tetapi masyarakat awam juga mendapatkan akses yang sama pada skala luas (Chesney & Citron, 2019).

Kemudahan akses *deepfake* disebabkan oleh model GAN (*Generative Adversarial Networks*) yang mampu menghasilkan konten media hampir menyerupai bukti "asli" dalam bentuk audio, foto, maupun video (IEEE, 2023). Hal ini dapat dibuktikan dengan data penyebaran konten *deepfake* dari berbagai pengguna yang diambil pada Juli 2018 hingga April 2021.



Gambar 2. 1 Data Peningkatan Kasus Deepfake Tahun 2018-2021

(Sumber: *IEEE, 2023*)

Berdasarkan data di atas, menunjukkan tren kenaikan terhadap penggunaan teknologi *deepfake* yang cukup progresif terjadi setiap 6 bulan. Aktivitas "Face Swapping" menjadi penggunaan *deepfake* yang meningkat paling drastis sebanyak 704% di paruh kedua tahun 2024. Tren peningkatan penyebaran konten

deepfake juga terjadi di benua Asia dengan data terbesar berasal dari Vietnam sebesar 25,3% dan Jepang sebesar 23,4%. Namun, laporan menunjukkan bahwa kasus *deepfake* paling banyak terjadi di negara Filipina sebanyak 4500% (Nelson, 2024). Terbukti bahwa penggunaan teknologi *deepfake* ini tidak memerlukan teknik atau kemampuan khusus (Security Org, 2024).

Penyebaran disinformasi dengan teknologi *deepfake* diperkirakan akan semakin umum terjadi seiring dengan kemajuan teknologi dan sarana penggunaan teknologi AI yang menjadi lebih banyak tersedia bagi semua kalangan (Cristiano, et all (Eds.), 2023). Akses terhadap *deepfake* tidak hanya eksklusif digunakan oleh ilmuwan komputer, tetapi beberapa *programmer* individu telah menciptakan proyek terbuka untuk pembuatan *deepfake*, seperti FakeApp, FaceSwap, dan Deep Face Lab. Proyek-proyek ini diluncurkan pada repositori publik layaknya GitHub dan mampu menghasilkan *deepfake* dalam bentuk video hingga *lip-synch*. Selain itu, berbagai aplikasi gratis turut hadir menawarkan versi yang terbatas untuk menggunakan teknologi *deepfake*, misalnya mempercepat dan memperlambat, serta melacak dan memanipulasi wajah. SnapChat dan Tiktok merupakan salah satu contoh aplikasi yang menawarkan komputasi rendah tanpa keahlian tertentu untuk melakukan manipulasi secara langsung dengan efek *filter* serta fitur untuk mempercepat dan memperlambat video (Paris & Donovan, 2019).

Mudahnya penggunaan teknologi *deepfake* ini banyak dimanfaatkan untuk penyebaran konten palsu dengan tujuan menyebarkan disinformasi di tengah masyarakat. *Deepfake* bukanlah satu-satunya cara maupun teknologi yang dapat digunakan untuk memanipulasi video maupun gambar. Akan tetapi, hasil yang

ditunjukkan oleh teknologi ini cukup membedakan dibandingkan dengan teknik manipulasi lainnya. Pertama, potensi hasil media yang "meyakinkan" dengan adanya berbagai sumber terkait subjek yang dijadikan target serta adanya waktu yang cukup untuk menganalisis hal tersebut. Kedua, kemudahan akses teknologi di kalangan masyarakat awam secara gratis melalui aplikasi, seperti FakeApp maupun OpenFaceSwap. Munculnya *deepfake* membuat adanya kebutuhan akan teknologi yang dapat melakukan verifikasi terkait dengan autentikasi suatu konten untuk mendeteksi hasil "manipulasi." Hal demikian menjadi isu yang sensitif dan penting untuk diperhatikan di tengah era '*fake news*' (Koopman, Rodriguez, & Geradts, 2018).

Gambar 2. 2 Perbedaan Antara Foto Asli (Baris Atas) dengan Hasil Deepfake (Baris Bawah)

(Sumber: Megahed, Han, & Fadl, 2023)

Survey yang dilakukan oleh McAfee menunjukkan bahwa 70% responden menyatakan tidak mampu dalam membedakan audio "asli" dengan hasil *deepfake*. Pada studi yang sama, sebesar 40% responden menjawab akan memberikan

bantuan apabila hasil audio *deepfake* menyerupai suara pasangan mereka (Security Org, 2024). Dua survey tersebut memberikan gambaran bahwa penyebaran disinformasi atau *fake news* dari hasil *deepfake* sangat mudah dilakukan untuk mengelabui masyarakat awam.

Faktor tersebut turut dipengaruhi oleh media sosial sebagai wadah untuk menyebarkan informasi dan menyatakan pendapat. Media sosial belum memiliki sistem yang cukup memadai untuk dapat mendeteksi dan melakukan autentikasi terhadap suatu konten. Kehadiran teknologi *deepfake* menjadi masalah baru dalam media sosial karena potensi untuk penyebaran disinformasi. Berita disinformasi ini lebih mudah untuk tersebar secara cepat dan dipercayai karena memiliki sifat yang "dramatis" dibandingkan berita aktual. Mayoritas berita aktual menyebar secara lambat dan menjangkau lebih sedikit orang (Gambín, et all, 2024). Oleh karenanya, media sosial memiliki peran yang besar terhadap penyebaran konten *deepfake* sebab berpengaruh terhadap persepsi publik dalam menangkap suatu informasi yang aktual.

Sebelum hadirnya *deepfake*, tindakan manipulasi terhadap audiovisual telah dilakukan sejak lama. Para pelaku memanfaatkan media gambar atau foto sebagai bukti yang "realisits" semenjak hadirnya penemuan fotografi. Berbagai teknik manipulasi yang telah dijalankan ini berhasil untuk menciptakan konten disinformasi yang kemudian diperbagus melalui hadirnya teknologi kecerdasan buatan dengan hasil yang lebih realistis.

2.2 Perkembangan Teknik Manipulasi Audiovisual Sebelum Hadirnya *Deepfake*

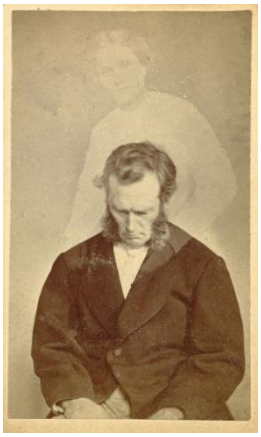
Sebuah media memiliki kapabilitas untuk mempengaruhi pandangan seseorang atau sekelompok orang karena telah menjadi alat dalam membuktikan suatu informasi. Padahal, media rentan untuk dimanipulasi melalui teknik tertentu sehingga menciptakan hasil yang cukup otentik. Dalam konteks ini, teknik manipulasi sebelum hadirnya *deepfake* memiliki kapabilitas yang sama untuk mempengaruhi persepsi masyarakat. Sebagai contoh, kasus mengenai pemberitaan Perang Teluk atau *Gulf War*. Seorang jurnalis di Amerika Serikat memberikan laporan mengenai Perang Teluk dengan menggunakan sudut pandang seolah-olah Amerika Serikat merupakan pihak yang paling dirugikan. Media yang disertakan dalam pemberitaan tersebut merupakan hasil foto dari kamera khusus *infrared* berisikan peluncuran misil oleh Irak. Faktanya, media yang ditampilkan adalah bukti asli namun narasi yang diberitakan oleh jurnalis AS tersebut merupakan pernyataan palsu karena korban jiwa Irak lebih banyak dibandingkan AS. Peristiwa ini disebut dengan istilah "simulacra" oleh seorang teoritikawan media bernama Jean Baudrillard (Paris & Donovan, 2019).

Gambar 2. 3 Rekaman Gambar Perang Teluk Menggunakan Kamera Infrared

(Sumber: Paris & Donovan, 2019)

Simulacra menjelaskan bagaimana pihak yang memiliki hak istimewa dalam pembentukan struktur sosial dengan menjadikan sebuah media sebagai bukti. Dari contoh, kapabilitas yang dimiliki oleh media untuk mempengaruhi masyarakat tidak hanya dibentuk secara teknis namun dapat dipengaruhi oleh sebuah narasi dari seorang juru bahasa (Ibid, 2019). Mereka dapat membentuk dan menyebarkan berbagai narasi terlepas dari realita atau fakta pada media yang diberitakan. Dalam konteks ini, *deepfake* dapat digunakan untuk menyertakan bukti sintetis terhadap narasi palsu yang disebarkan pada berbagai *platform*. Hal ini tentunya akan menguntungkan pelaku ketika berhasil mempengaruhi target dengan fakta yang palsu.

Kasus simulacra pada penjelasan di atas menggambarkan salah satu bentuk teknik manipulasi audiovisual yang ada sebelum hadirnya teknologi *deepfake*. Melihat dari sejarahnya, teknik-teknik manipulasi audiovisual telah mengalami perkembangan yang cukup signifikan. Kegiatan manipulasi ini sudah dilakukan bersamaan dengan penemuan fotografi pada tahun 1816. Kasus William Mumler tercatat sebagai salah satu kasus manipulasi audiovisual yang paling awal terjadi pada tahun 1869. Media gambar yang dihasilkan oleh William menunjukkan “foto arwah” seakan-akan adalah seorang terkasih yang telah meninggal (Llorente, 2024).



Gambar 2. 4 Hasil "Foto Arwah" Karya William Mumler

(Sumber: Mashable)

Teknik yang dilakukan oleh William melibatkan pencahayaan ganda untuk menciptakan "arwah" sebagai metode awal manipulasi gambar. Berawal dari hasil foto ini, William menjalankan bisnis palsu yang menarik masyarakat untuk meminta jasanya menghasilkan "foto arwah" bersama dengan orang tersayang yang telah meninggal. Bisnis ini menjadi sebuah keuntungan bagi William karena pada masa itu banyak orang yang sedang berduka akibat kehilangan orang yang dicintai selama Perang Sipil Amerika sehingga rentan untuk mempercayai "foto arwah." Pada masa itu, sangat sulit untuk membuktikan tindakan manipulasi yang dilakukan oleh William sekalipun telah didemonstrasikan cara yang dilakukan olehnya. William pada akhirnya dibebaskan setelah menjalani sidang untuk tindakan manipulasinya (Ibid, 2024).

Berbeda dengan kasus William, bentuk tindakan manipulasi audiovisual juga terjadi pada penerjemahan film. Tindakan manipulasi ini terjadi pada masa-masa awal munculnya produksi film, khususnya di negara Jerman, Italia, dan Spanyol. Manipulasi audiovisual yang dilakukan didasarkan atas arahan rezim

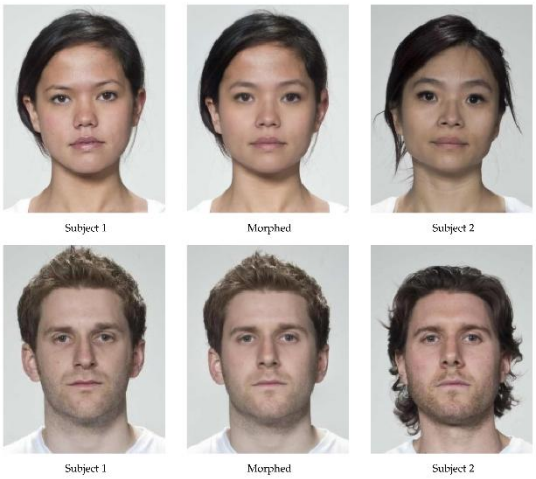
politik yang berkuasa dengan tujuan untuk mempertahankan ideologi negara. Dalam hal ini, penerjemah audiovisual beroperasi di bawah pengawasan rezim diktator dan komite sensor. Mereka menekankan penerjemahan film yang harus sesuai dengan ideologi dan nilai-nilai yang berlaku pada saat itu. Penerjemah tidak hanya menyesuaikan pada kesepadanan linguistik, tetapi juga ditentukan oleh rezim mengenai konten terjemahan yang dapat ditayangkan ke masyarakat (Díaz Cintas, 2012).

Aksi manipulasi terhadap konten terjemahan film ini dilakukan karena hegemoni film AS yang menampilkan cara hidup tertentu dan berkontribusi terhadap peningkatan dominasi bahasa Inggris. Hal ini akhirnya mendorong reaksi yang defensif, seperti di Spanyol pada masa kepemimpinan Franco. Banyak manipulasi teks terjemahan di bawah rezim Franco yang tersebar di sinema dan teater akibat kekhawatiran terhadap hegemoni AS. Pembentukan kelompok-kelompok penelitian, seperti TRACE (Translations Censored) dan Tralima, ditujukan untuk menganalisis teks-teks terjemahan yang disensor oleh rezim-rezim diktaktor tersebut dan memberikan bukti konkret terkait manipulasi yang dilakukan (Ibid, 2012).

Sebelum hadirnya teknologi *deepfake*, teknik manipulasi gambar dengan teknologi terbaru mulai dikenal publik sejak tahun 1990-an. Tindakan ini dikenal dengan terminologi “photoshop” yang diambil dari salah satu program milik Adobe (Langguth, et all, 2021). Melalui photoshop, sebuah gambar dapat diedit dengan menghapus objek, menghapus latar, memperhalus wajah, dan lainnya sehingga menyerupai sebuah bukti yang asli. Selain gambar atau foto, manipulasi

juga dilakukan terhadap video. Manipulasi video sebagian besar dikembangkan dan digunakan oleh industri film seperti Hollywood. Tentunya penggunaan teknologi manipulasi ini membutuhkan biaya yang besar karena memiliki banyak efek CGI (computer-generated image) dengan kualitas tinggi (Ibid, 2021). Hal ini sangat berbeda dengan metode AI yang lebih mudah untuk dijangkau dengan hasil yang cukup realistis layaknya *deepfake*.

Pada perkembangannya, tindakan manipulasi dengan wajah seseorang dapat dilakukan melalui Video Rewrite dan Video Face Replacement. Video Rewrite merupakan metode yang secara otomatis membuat video baru dari seseorang dengan gerakan mulut yang berbeda, sedangkan Video Face Replacement merupakan salah satu metode pertama untuk penukaran wajah secara otomatis. Melalui video kamera tunggal, metode ini merekonstruksi model 3D dari kedua wajah dan menyatukan wajah sumber ke wajah target (Rössler, et all, 2018). Teknik tersebut merupakan teknik utama yang digunakan untuk memanipulasi gambar atau video, yakni *morphing*. *Morphing* terdiri dari identifikasi terhadap pola antara 2 foto dan dinamika transformasi dari satu gambar menjadi gambar lain seiring dengan gerakan dari titik A ke titik B (Vállez & Boté-Vericad, 2022).



Gambar 2. 5 Teknik Morphing Pada Gambar

(Sumber: Pikoulis, et all, 2021)

Teknik manipulasi berikutnya adalah *warping* yang memungkinkan bentuk sebagian gambar dimodifikasi untuk tujuan kreatif. *Warping* dapat digunakan untuk mengoreksi disfungsi atau bahkan menciptakan distorsi. Teknik ini biasa digunakan untuk peningkatan estetika foto di sosial media. Dalam manipulasi audio, berbagai teknik umumnya juga digunakan. Misalnya berdasarkan teknologi *text-to-speech* yang memungkinkan audio untuk dapat diubah ke dalam bentuk teks yang ditulis ulang. Ketika manipulasi audio dan video digabungkan, hal tersebut dapat menggunakan teknik *lip-syncing* yang memodifikasi gerakan mulut agar sesuai dengan kata-kata tertentu dan umumnya digunakan dalam *video game* 3D (Ibid, 2022). Berbagai teknik manipulasi terhadap audiovisual ini kemudian ditingkatkan melalui teknologi AI dengan model GAN yang akhirnya menghasilkan *deepfake*.

Berdasarkan penjelasan terkait dengan teknik-teknik manipulasi audiovisual di atas, menggambarkan implikasi negatif untuk tujuan penyebaran disinformasi.

Dalam konteks ini, media menjadi sarana yang tepat untuk memperkuat pernyataan atau narasi palsu yang diberitakan. Maka dari itu, berbagai aktivitas manipulasi ini kemudian termanifestasi dengan hadirnya *deepfake* melalui hasil konten yang lebih realistis dan akhirnya menimbulkan potensi bahaya yang baru.

2.3 Implikasi Negatif Penyalahgunaan Teknologi *Deepfake*

Penjelasan pada bagian sebelumnya telah menunjukkan perkembangan teknik yang semakin mumpuni dalam menyebarkan konten disinformasi. Hadirnya *deepfake* sebagai teknologi yang baru untuk melakukan manipulasi media telah menjadi diskursus baru terkait dengan potensinya untuk menyebarkan kampanye disinformasi. Ini menunjukkan bahwa kemajuan teknologi AI dapat menurunkan hambatan teknis tanpa kemampuan tertentu yang diperlukan dan mampu untuk melakukan operasi secara efektif. Para ahli menyatakan bahwa video *deepfake* akan menjadi hal yang tidak dapat dibedakan kembali dengan bukti asli di masa depan seiring dengan cepatnya perkembangan teknologi AI. Dalam konteks ini, akses terhadap *deepfake* juga menjadi lebih luas dan terjangkau sehingga potensi penggunaannya untuk penyebaran disinformasi juga meningkat. Kemampuan *deepfake* dalam menghasilkan media sintetis yang "realistis" mempermudah berbagai aktor untuk membuat dan menyebarkan konten manipulasi yang memuat pernyataan atau tindakan palsu dengan tujuan pencemaran nama baik. Ini menjadi salah satu alasan *deepfake* menjadi teknik manipulasi yang efektif dalam mempengaruhi opini masyarakat utamanya memanfaatkan bias pandangan yang dimiliki (Cristiano, et all (Eds.), 2023).

Mayoritas konten *deepfake* bersirkulasi pada media pornografi namun konten *deepfake* politik paling banyak mendapatkan perhatian dikarenakan kapabilitasnya untuk menciptakan ketidakstabilan baik di dalam maupun luar negeri (Gambín, et all, 2024). Prospek akan produksi massal dan penyebaran konten *deepfake* menjadi tantangan yang serius terhadap wacana politik. Hal ini disebabkan oleh kemampuan media dalam bentuk 'visual' yang memiliki kekuatan persuasif lebih baik dibandingkan bentuk 'teks.' Maka dari itu, konsep realisme sebagai hasil konten dari *deepfake* memiliki pengaruh yang kuat untuk mengelabui seseorang, terlepas dari kebenaran konten tersebut (Vaccari & Chadwick, 2020). Ini menjadi potensi permasalahan baru dari penyalahgunaan *deepfake* apalagi dengan pembuatan konten menggunakan figur-figur politik yang berpengaruh.

Pengaruh *deepfake* dalam lingkup politik dapat dilihat dari kasus-kasus berikut. Pada tahun 2023, terdapat serangan terhadap Sadiq Khan dan Keir Starmer, politisi Inggris, berupa manipulasi audio yang digunakan untuk menjatuhkan reputasi mereka. Dalam media yang beredar, terdapat klip palsu Khan yang meremehkan Hari Gencatan Senjata di kalangan ekstrimis dan foto-foto Starmer yang memperlakukan staf dengan buruk selama Konferensi Partai Buruh (Moreno, 2024). Di tahun yang sama, terdapat kasus yang sama terjadi dalam pemilu Slovakia. Pada kasus ini, terdapat sebuah klip audio palsu yang berisikan narasi dari Michal Simecka terkait dengan proses pemilu yang curang. Klip ini hadir tepat 2 hari sebelum pemilu raya dilakukan namun pemenangnya adalah Robert Fico. Adanya konten *deepfake* yang tersebar ini tidak terlepas dari

berbagai spekulasi yang kuat bahwa kekalahan Simecka disebabkan oleh hal tersebut (de Nadal, & Jančárik, 2024).

Kasus *deepfake* yang berpengaruh dalam lingkup politik juga terjadi di wilayah regional ASEAN. Misalnya, pada tahun 2019 di Malaysia tersebar sebuah video yang menyerupai Menteri Ekonomi sedang melakukan aktivitas hubungan seksual. Video tersebut berhasil mempengaruhi reputasi yang dimiliki oleh menteri tersebut (Cristiano, et all (Eds.), 2023). *Deepfake* juga turut disalahgunakan oleh penjahat kriminal di Thailand dengan video pemerasan yang mengimitasi seorang polisi pada tahun 2022. Pihak kepolisian Thailand memperingati masyarakat untuk berhati-hati apabila menerima panggilan video dari "polisi" dikarenakan tidak adanya kebijakan Thailand melakukan panggilan video terhadap siapapun yang berkaitan dengan kasus kriminal (Thaiger, 2022). Pada tahun 2023, tersebar sebuah video *deepfake* Lee Hsien Loong dan Lawrence Wong, perdana menteri dan wakil perdana menteri Singapura, yang mempromosikan *crypto* dan berbagai produk investasi lainnya. Kasus-kasus *deepfake* yang terjadi di wilayah ASEAN ini memiliki sebuah kemiripan yakni digunakan untuk kepentingan yang menguntungkan pihak tertentu dengan menjatuhkan pihak lain (Surasit, 2024).

Selain di wilayah ASEAN, kasus *deepfake* juga tersebar di wilayah regional Afrika dan Amerika Latin. Salah satu kasus *deepfake* yang cukup berpengaruh di Afrika adalah terkait dengan unggahan video Presiden Afrika Selatan, yakni Cyril Ramaphosa, pada tahun 2023 lalu. Dalam video *deepfake* ini, Ramaphosa membicarakan langkah-langkah yang akan diambil oleh Afrika Selatan untuk

mengatasi permasalahan krisis energi dengan cara menghancurkan Monumen Voortrekker dan stadium olahraga rugby Loftus Versfeld. Penyebaran video disinformasi ini dapat memicu polemik baru karena Monumen Voortrekker merupakan salah satu tempat bersejarah bagi kaum kulit putih Afrika atau biasa disebut "Afrikaners," sedangkan stadium rugby tersebut merupakan salah satu ciri khas atau bagian dari identitas yang dimiliki oleh Afrika Selatan (Cosser, 2023). Lalu, di wilayah Amerika Latin, salah satu kasus *deepfake* yang terkenal terjadi di Venezuela mengenai pemberitaan positif terkait negara Venezuela pada salah satu stasiun televisi bernama VTV. Reporter dari acara berita tersebut adalah hasil avatar yang dibuat oleh perusahaan AI di London bernama "Synthesia" dan mereka turut membicarakan bahwa isu-isu di Venezuela, seperti imigrasi, politik, dan ekonomi telah terselesaikan dengan baik oleh pemerintah negara tersebut (Paul, 2023). Ini adalah bentuk disinformasi yang bisa memanipulasi persepsi masyarakat.

Di Indonesia, penyebaran konten *deepfake* juga terjadi dan menyerang pihak politisi. Pada tahun 2023, beredar video Jokowi yang menyampaikan pidato dalam bahasa Mandarin. Video ini merupakan hasil *deepfake* yang tersebar di media sosial X dengan intensi untuk menyebarkan disinformasi bahwa Jokowi berpihak ke Cina dan anti Islam. Menkominfo telah memberikan pernyataan terkait video Jokowi tersebut adalah palsu dan berasal dari pidato Jokowi yang dibawakan di USINDO (US – Indonesia Society) pada 13 November 2015 di mana pada video asli tidak memakai bahasa Mandirin (*Video Jokowi Berbahasa Mandarin Diedit Pakai Deepfake*, 2023). Pada era pemilu 2024, penyebaran konten *deepfake* juga

terjadi dengan tujuan untuk membuat masyarakat Indonesia mengubah pilihannya untuk calon yang lain. Di media sosial, tersebar rekaman percakapan palsu antara Anies Baswedan dengan Surya Paloh yang membahas debat pilpres 2024. Dalam rekaman tersebut, muncul sebuah narasi seakan-akan Surya Paloh memarahi Anies melalui sambungan telepon (Rahmawati, 2024).

Berbagai kasus *deepfake* pada penjelasan sebelumnya membuktikan bahwa teknologi tersebut lebih banyak disalahgunakan untuk tujuan tertentu utamanya menyerang keberlangsungan politik pada suatu negara. *Deepfake* mungkin tidak selalu menipu seorang individu, tetapi dapat menimbulkan rasa ketidakpastian yang akan berdampak pada menurunnya kepercayaan terhadap berita yang ada di media sosial. Bahkan pada jangka panjang, *deepfake* akan mempengaruhi perilaku seseorang yang tidak lagi dapat bertanggung jawab terhadap informasi yang mereka sebarkan. Hal ini juga dapat mempengaruhi masyarakat untuk menghindari pemberitaan apapun karena hadirnya rasa ketidakpastian tersebut (Vaccari & Chadwick, 2020).

Ancaman yang muncul dari hadirnya teknologi *deepfake* sebenarnya bukanlah suatu hal yang baru. Namun, *deepfake* menjadi salah satu cara baru untuk melakukan manipulasi media dengan bentuk yang sangat meyakinkan hampir menyerupai media “asli” (Pawelec, 2022). Hasil konten *deepfake* yang berkualitas tinggi ternyata juga dapat disalahgunakan sebagai senjata untuk melakukan perang informasi dan propaganda. Pada Oktober 2023, pemerintah Israel mengunggah gambar yang menggambarkan anak-anak sebagai korban dari serangan teroris. Mayoritas akun pengguna menuduh pemerintah Israel

memalsukan gambar tersebut dengan *deepfake*. Padahal, gambar tersebut bersifat asli dan penuduhan dengan teknologi *deepfake* digunakan sebagai bentuk propaganda (Łabuz & Nehring, 2024).

Dari kasus tersebut, *deepfake* bukanlah permasalahan utama yang perlu dikhawatirkan, tetapi rasa takut dan cemas karena tidak dapat membedakan sebuah informasi yang merupakan konten *deepfake* atau asli. Hal ini menunjukkan bahwa potensi disruptif *deepfake* dapat berupa ancaman psikologis, yakni ketakutan, ketidakpastian, dan ketidakmampuan untuk membedakan suatu konten yang otentik atau palsu. Apalagi, rasa takut yang dialami disebabkan oleh informasi yang tidak dapat dipastikan kepastiannya. Konsekuensi berikutnya memungkinkan terhadap penggunaan *deepfake* untuk melakukan kampanye politik dengan menunjukkan sisi baik dari seorang kandidat atau politisi. Hal ini dapat dilakukan agar menampilkan seseorang secara lebih menarik, lebih muda, cakap, menggunakan bahasa yang berbeda, serta membuat pesan yang khusus bagi setiap pemilihnya (Łabuz & Nehring, 2024). Akan tetapi, pembahasan tersebut masih memerlukan studi lebih lanjut untuk ditelaah secara mendalam.

Keberlangsungan politik pada suatu negara dapat dipengaruhi oleh *deepfake* melalui acara pemilu. Dalam hal ini, dapat dibayangkan dengan sebuah skenario yang mana tersebar video *deepfake* berisi seorang pemimpin yang mengumumkan pembatalan pemilu mengakibatkan banyak pemilih yang tidak memberikan suaranya. Meskipun tidak menyerang sistem pemilu yang berlangsung, skenario tersebut tetap dapat menghasilkan efek yang sama. Negara pelaku yang mendistorsi video tersebut berhasil mengubah hasil pemilu sehingga negara target

kehilangan kendali atas kemampuannya untuk melakukan pemilu yang adil. Hal ini dapat disamakan dengan campur tangan pemilu langsung yang disoroti oleh beberapa negara sebagai bentuk intervensi yang dilarang (Cristiano, et all (Eds.), 2023).

Berdasarkan pembahasan mengenai implikasi negatif *deepfake* secara umum, berikut ini merupakan tabel rangkuman yang mencakup penjelasan tersebut dengan poin-poin utama yang telah dibahas secara singkat.

Tabel 2. 1 Implikasi Negatif dari Hadirnya Teknologi Deepfake

Implikasi Negatif *Deepfake*		
Internasional	Asia Tenggara	Politik
Mengakibatkan ketegangan politik dan agama antarnegara untuk merusak hubungan diplomatik	Mudahnya untuk melakukan pemerasan	Menurunnya reputasi seorang pemimpin, lembaga, maupun politisi
Potensi untuk menyesatkan masyarakat bahkan negara hingga menyebabkan kebingungan di tingkat global	Pencemaran nama baik atas nama institusi maupun seorang politisi	Keberlangsungan pemilu yang tidak dapat berjalan secara jujur dan adil
Manipulasi gambar satelit untuk membingungkan militer yang berdampak pada keamanan dan pertahanan internasional	Penipuan massal secara online melalui kejahatan terorganisir	Menurunnya kepercayaan masyarakat
Merusak kepercayaan dan kredibilitas dalam keterlibatan diplomatik	Ancaman identifikasi biometrik untuk akses data dan informasi keuangan pribadi	Alat propaganda untuk menjatuhkan pihak lain

<table>
<tr><td>Memperburuk ketegangan geopolitik dan risiko keamanan</td><td>Pembuatan konten pornografi dengan target anak-anak di bawah umur</td><td>Mempengaruhi keberlangsungan proses politik di suatu negara</td></tr>
<tr><td>Mempengaruhi persepi politik dengan penyebaran narasi palsu atau kerusuhan sosial di negara lain</td><td>Ancaman untuk melakukan investasi palsu</td><td>Potensi kampanye palsu untuk menarik perhatian masyarakat atau calon pemilih</td></tr>
<tr><td>Potensi intervensi terhadap negara lain</td><td>Potensi pemalsuan identitas pribadi</td><td>Potensi intervesi dari negara luar</td></tr>
<tr><td>Eskalasi peristiwa mengarah pada kekerasan pada masa krisis</td><td rowspan="3">Ancaman terhadap keberlangsungan pemilu</td><td>Potensi menghilangnya identitas nasional</td></tr>
<tr><td rowspan="2">Perang informasi dan propaganda</td></tr>
<tr><td>Memperburuk perpecahan masyarakat dan polarisasi opini internasional</td></tr>
<tr><td>Mempersulit pertanggungjawaban di bawah hukum internasional</td><td></td></tr>
<tr><td>Sumber: Byman, et all, 2023;
Ikenga & Nwador, 2024; Abbas & Taeihagh, 2024.</td><td>Sumber: Surasit, 2024; Rayda, 2024; ceosbharat, 2024.</td><td>Sumber: Cristiano, et all (Eds.), 2023; Łabuz & Nehring, 2024; Vaccari & Chadwick, 2020; de Nadal, & Jančárik, 2024.</td></tr>
</table>

Pembahasan mengenai implikasi negatif dari penyalahgunaan teknologi *deepfake* menunjukkan bagaimana tindakan tersebut dapat mempengaruhi kepercayaan publik melalui penyebaran konten disinformasi. Hal ini dapat dilihat dari sejarah manipulasi audiovisual melalui berbagai teknik sebelum hadirnya teknologi tersebut. Implikasi negatif ini nyatanya kurang disadari oleh setiap

individu sehingga mempermudah untuk melakukan manipulasi apalagi dengan hasil *deepfake* yang hampir realistis. Ini dapat dibuktikan dalam penjelasan terkait efek bahaya *deepfake* terhadap keberlangsungan politik di suatu negara dengan memanfaatkan momentum dan kepercayaan publik. Oleh karenanya, berbagai usaha untuk diskusi sekaligus menetapkan regulasi penggunaan teknologi AI telah dilakukan pada tingkat global maupun regional dalam menanggulangi penyalahgunaan *deepfake*.

2.4 Regulasi Internasional dan Regional Terkait *Deepfake*

Berdasarkan hasil riset, terdapat kenaikan kasus penyebaran konten *deepfake* sebanyak 550% dari tahun 2019-2023 yang mendemonstrasikan kemudahan aksesibilitas dan potensi penyalahgunaan terhadap teknologi tersebut (*2023 State Of Deepfakes: Realities, Threats, And Impact*, 2023). Untuk mengatasi permasalahan tersebut, berbagai negara di dunia internasional telah melakukan diskusi mengenai pemanfaatan teknologi AI. Inisiatif pada skala internasional dilakukan melalui KTT (Konferensi Tingkat Tinggi) AI Global yang diselenggarakan oleh Saudi Data and Artificial Intelligence Authority (SDAIA) pada tahun 2020. KTT ini mengusung tema "AI for the Good of Humanity," yang mempertemukan para pemerintah, pemimpin bisnis, dan ahli AI dari berbagai negara dengan tujuan mendiskusikan masa depan AI serta perannya dalam membentuk ekonomi, masyarakat, dan teknologi global. PBB juga turut mengadakan KTT yang sama dengan tujuan mengidentifikasi penggunaan AI secara praktis dalam mempercepat kemajuan menuju SDGs (*Sustainable Development Goals*). KTT ini menjadi wadah bagi setiap negara yang turut

berpartisipasi fokus dalam mengembangkan sekaligus mengimplementasikan alat yang tepat untuk mendeteksi serta melenyapkan manipulasi *deepfake* dan mengedukasi akan pentingnya dalam melakukan penyaringan informasi yang tepat (*Global AI summit tackles misinformation and deepfakes with*, 2024).

Sebelum hadirnya regulasi AI yang resmi pada skala global, UNESCO telah mengeluarkan rekomendasi terkait etik penggunaan AI pada bulan November 2021. Dalam rekomendasi ini, UNESCO berusaha memberikan kerangka umum yang mencakup nilai, prinsip, dan aksi setiap negara untuk membentuk regulasi AI yang konsisten dengan hukum internasional (UNESCO, 2022). Perkembangan terkait penerapan regulasi ini kemudian dilakukan oleh Cina dengan mengeluarkan ketentuan mengenai administrasi layanan informasi internet *deep synthesis*. Isi ketentuan ini mencakup layanan informasi daring yang menggunakan teknologi *deepfake* untuk mensintesis teks, gambar, audio, maupun video (ITU, 2024).

Di tahun 2023, aksi lanjutan terkait diskusi maupun penetapan regulasi penggunaan AI secara global berlangsung secara progresif. Contohnya, di AS dengan penetapan perintah eksekutif untuk menjamin penggunaan AI yang aman, terjamin, dan dapat dipercaya. Selain itu, Cina juga mengeluarkan "Global AI Governance Initiative" yang mendemonstrasikan intensinya untuk berkontribusi dalam diskusi internasional dan membangun kerangka pemerintahan AI. Organisasi internasional seperti UNDP juga mengeluarkan instrumen kebijakan bernama AIRA (AI Readiness Assessment) yang bertujuan untuk membantu setiap pemerintah memahami teknologi AI baik sebagai pengguna maupun

regulator dengan mengidentifikasi faktor kritis terkait pengembangan sekaligus penggunaan secara etis. AIRA membantu dalam membentuk atau memperkuat strategi nasional negara terkait dengan AI melalui pemahaman ekosistem AI. UNDP juga mendorong setiap negara dengan memberikan dukungan hukum dan institusional melalui program AIRA (CEB, 2024).

Diskusi global mengenai "AI Safety Summits" juga turut dijalankan pada November 2023 di United Kingdom (UK). Diskusi ini hadir dalam rangka membahas langkah-langkah mitigasi yang bisa dilakukan terkait ancaman dari penggunaan teknologi AI dengan menyertakan pemerintah serta perusahaan AI. Pada akhir konferensi, sebanyak 29 pemerintah dan organisasi internasional menandatangani "Bletchley Declaration" yang berisikan pernyataan mengenai persetujuan terhadap serangkaian langkah konkret yang akan diterapkan dengan tujuan meningkatkan keamanan dan pengembangan yang bertanggung jawab atas teknologi AI. Hadirnya deklarasi ini memperkuat kerangka kerja peraturan global mengenai penggunaan AI yang aman. Bersamaan dengan ini, UK juga turut mendirikan AI Safety Institute yang kemudian diikuti oleh pemerintah Amerika Serikat, Kanada, Jepang, Korea Selatan, dan Singapura (ITU, 2024).

Pada tahun berikutnya, yakni 2024, regulasi terkait penggunaan AI pertama kali hadir dan dinyatakan secara resmi oleh organisasi regional Uni Eropa. Regulasi ini dinamakan "AI Act" dan disetujui oleh Dewan Eropa pada 13 Maret 2024. Regulasi ini mengkategorisasikan sistem AI berdasarkan pada potensi risikonya dan memberlakukan aturan yang lebih ketat untuk risiko yang lebih tinggi. Walaupun AIA (AI Act) telah menganggap *deepfake* sebagai salah satu

permasalahan siber, teknologi ini memerlukan beberapa konsiderasi terhadap kompatibilitasnya dengan European Convention on Human Rights (ECHR) dan General Data Protection Regulation (GDPR) (European Union, 2023).

AIA menjelaskan perlunya transparansi dalam penyebaran konten AI di media sosial dengan memberikan label, *watermark*, atau *tag* yang menyertakan bahwa konten yang diunggah adalah hasil teknologi AI. Namun, terdapat pengecualian terhadap konten AI yang digunakan untuk industri kreatif, satir, fiksi, ataupun *deepfake* artistik tentunya secara halus. AIA juga mengkategorikan konten disinformasi pemilu, ekstorsi, dan pelecehan seksual sebagai golongan risiko paling tinggi karena potensi bahaya dan pelanggaran terhadap HAM. AIA mendukung pembentukan kantor dan komisi AI untuk mendukung implementasi persyaratan pelabelan dan deteksi *deepfake* melalui pedoman praktis. Sebagai kesimpulan, AIA milik Uni Eropa telah memberikan kerangka kerja kebijakan yang mengatur *deepfake*, utamanya melalui kebijakan transparansi dan sistem klasifikasi berbasis risiko (European Union, 2023; Moreno, 2024).

Langkah terbaru yang dilakukan pada bulan September 2024, yakni penandatanganan perjanjian internasional terkait Konvensi Kerangka Kerja Dewan Eropa tentang Kecerdasan buatan dan HAM, Demokrasi, dan Aturan Hukum. Sama seperti AIA, perjanjian ini mengadopsi pendekatan berbasis risiko terhadap penggunaan AI namun tidak menetapkan batasan-batasan penggunaannya. Perjanjian internasional ini merupakan kerangka kerja yang mengikat secara hukum dan menetapkan komitmen luas bagi para pihak yang berpartisipasi di dalamnya. Tujuan dari perjanjian ini dilakukan dalam rangka

memastikan bahwa sistem AI yang berlaku mematuhi HAM, demokrasi, dan supremasi hukum. Perjanjian ini tidak hanya berlaku bagi pemerintah saja, tetapi aktor swasta dan publik turut melaksanakan penggunaan AI yang etis. Pengecualian berlaku untuk kegiatan yang melindungi kepentingan keamanan dan pertahanan nasional, serta penelitian dan pengembangan sistem AI (Gesly, 2024).

Pada waktu yang sama, PBB mengeluarkan "Governing AI for Humanity" yang mengangkat permasalahan mengenai ancaman dari teknologi AI. Dokumen ini menjadi sebuah kerangka kebijakan strategis dalam menghadapi ancaman AI di tengah era transformatif. PBB menekankan perlunya langkah koordinasi yang lebih baik antara negara utamanya menyasar pada negara-negara yang kaya dikarenakan adanya eksklusivitas kegiatan diskusi yang dilakukan. Salah satu rekomendasi yang ditekankan adalah diperlukannya kantor resmi PBB yang bergerak di bidang AI untuk kemudahan koordinasi antar negara. Selain itu, diperlukan juga kontrak sosial mengenai AI dalam memastikan penggunaan teknologi tersebut untuk hal-hal yang positif bagi semua orang. Hadirnya dokumen ini selaras dengan upaya PBB dalam menanamkan etika pengembangan AI serta mendorong kolaborasi internasional untuk memajukan tujuan SDGs (ITU, 2024).

Usaha untuk menangani penyalahgunaan AI juga dilakukan di wilayah ASEAN. Salah satunya melalui dokumen rekomendasi kebijakan dengan judul "Expanded Guide on Governance and Ethics for Generative AI." Dokumen ini sebelumnya telah dirancang pada tahun 2024 yang menjelaskan bahwa teknologi AI memiliki 6 risiko utama, seperti: *deepfakes*, pelanggaran kekayaan intelektual,

implikasi bias, masalah privasi dan kerahasiaan, kesalahan, dan antropomorfisme. Dokumen ini turut menyertakan risiko perbatasan dan sistemik yang terkait dengan model AI paling terbaru. Dokumen rekomendasi kebijakan ASEAN mengenai AI akhirnya mengalami pembaharuan pada tahun 2025 seiring dengan ancaman *deepfake* yang masuk ke dalam diskusi tingkat tinggi di wilayah Asia Tenggara (Borak, 2025).

Dokumen yang telah diperbaharui ini mengidentifikasi *deepfakes*, peniruan, penipuan, dan aktivitas berbahaya sebagai risiko yang signifikan dari teknologi AI generatif. Hal itu dianggap sebagai ancaman karena membuat konten artifisial yang realistis untuk misinformasi, pencurian identitas, dan penipuan. Beberapa rekomendasi kebijakan dipaparkan dalam dokumen ini untuk memberikan langkah solutif terhadap potensi ancaman bahaya tersebut. Pembuktian konten disoroti sebagai area yang penting bagi ASEAN untuk dieksplorasi dalam menangani pemalsuan konten dan memastikan integritas informasi. Rekomendasi ini disertakan ASEAN untuk mendukung pengembangan terhadap teknologi deteksi autentikasi sebuah konten, seperti pembuktian kriptografi dan penandaan digital, serta membantu para pebisnis dan konsumen dalam membedakan konten asli dengan AI. Peran ASEAN di sini adalah mendorong negara-negara anggota, industri, masyarakat sipil, dan akademisi untuk meningkatkan kemampuan regional di area ini dengan tujuan menyelaraskan pendekatan sesuai norma-norma global dan mengadopsi solusi yang diakui secara luas (ASEAN, 2024).

ASEAN juga merekomendasikan terkait keamanan untuk mengurangi risiko penyalahgunaan teknologi AI. Caranya melalui pembelajaran terkait pengetahuan

keamanan bersama para pemangku kepentingan AI mengenai taktik, teknik, dan studi kasus yang relevan dengan AI generatif. Panduan ini diharapkan dapat memberikan perlindungan terhadap penggunaan AI serta penyebaran konten *deepfakes*. Selain itu, poin penting yang direkomendasikan oleh ASEAN adalah terkait dengan kesadaran dan pendidikan mengenai AI. Hal ini termasuk membantu masyarakat mempelajari mengenali konten yang dibuat oleh AI, seperti hasil konten *deepfake*. ASEAN berharap dengan meningkatnya kesadaran publik dan literasi digital, setiap individu dapat membedakan konten yang otentik dengan hasil manipulasi (ASEAN, 2024).

Berbagai aksi untuk menangani penyalahgunaan *deepfake* telah dilakukan pada skala internasional. Dimulai dari diskusi tingkat tinggi hingga penetapan regulasi seperti yang dilakukan oleh Uni Eropa sebagai wilayah pertama yang memiliki kebijakan resmi terkait penyalahgunaan teknologi AI. Aksi-aksi ini menunjukkan bahwa dunia internasional telah menyadari dan mengakui implikasi negatif yang muncul dengan hadirnya AI khususnya pada teknologi *deepfake*. Melalui berbagai aksi dan program yang sudah ada, negara-negara di dunia berharap dapat meminimalisir penyebaran disinformasi dengan konten *deepfake*.

Gambaran umum penelitian pada bab ini telah menunjukkan berbagai implikasi negatif *deepfake* khususnya terhadap keberlangsungan proses politik dan identitas nasional di suatu negara. Dapat disimpulkan bahwa teknologi *deepfake* menjadi salah satu perkembangan teknik manipulasi yang sangat signifikan. Hal ini disebabkan oleh model GAN yang mempermudah operasi pembuatan konten *deepfake* serta akses yang dibuka untuk publik. Teknik

manipulasi konten melalui *deepfake* ini menunjukkan efek yang disruptif apalagi dengan tren peningkatan kasus penyalahgunaan untuk penyebaran disinformasi. Efek ini secara implisit mempengaruhi keberlangsungan politik dan berpotensi menghapus identitas nasional suatu negara. Ini merupakan potensi bahaya yang perlu diwaspadai karena dapat mempengaruhi persepsi publik melalui konten disinformasi yang tersebar di media sosial.

Dunia internasional melihat potensi bahaya yang sama dan mengadakan berbagai diskusi tingkat tinggi dalam rangka meminimalisir penyalahgunaan AI untuk mengelabui opini masyarakat internasional dengan memanfaatkan persepsi yang bias. Berbagai strategi, rekomendasi kebijakan, sekaligus kerangka kebijakan telah ditetapkan pada tingkat internasional maupun regional, seperti di Uni Eropa, untuk menanggapi isu penyalahgunaan AI yang mengkhawatirkan. Aksi-aksi tersebut juga berhasil mempengaruhi Indonesia dalam melihat potensi bahaya penyalahgunaan AI pada berbagai bidang. Indonesia telah melihat penyalahgunaan AI sebagai salah satu ancaman nasional namun belum ada kerangka hukum yang konkret untuk mengatasi kasus-kasus kejahatan siber tersebut. Bab ini telah menggambarkan secara umum permasalahan penelitian terkait dengan bahaya *deepfake* terhadap keberlangsungan politik dan demokrasi suatu negara dengan studi kasus di Indonesia. Pembahasan pada bab berikutnya akan memaparkan hasil analisis terhadap bahaya *deepfake* ini melalui perspektif konstruktivisme.