JURNAL KOMUNIKASI & MEDIA DIGITAL
Volume 03, Number 01, 2025, pp. 44-52
ISSN 3025-5813 (Print) ISSN 3025-3314 (Online)
Open Access: journal.universitaspertamina.ac.id/jkmd

DEEPFAKE DAN DISTORSI RUANG PUBLIK DIGITAL: ANALISIS TEORI *PUBLIC SPHERE* JURGEN HABERMAS

Muhammad Albar Nagara[1]

[1] Program Studi Ilmu Komunikasi, UPN Veteran Yogyakarta

A B S T R A C T

This study analyzes the impact of deepfake technology on the digital public sphere through Jurgen Habermas' public sphere theory. Deepfake, an AI-driven manipulation tool, enables hyper-realistic yet deceptive videos, leading to misinformation, political propaganda, and privacy violations. Social media, particularly YouTube, has become a key platform for the spread of deepfake content, distorting rational discourse and eroding public trust. Using a qualitative, descriptive-analytical approach, this research examines how deepfake disrupts digital communication and challenge the principles of an open public sphere. The findings emphasize the urgent need for digital literacy, regulatory frameworks, and AI-powered detection systems to mitigate its negative impacts. Strengthening awareness and institutional responses is crucial in maintaining a transparent and credible digital public sphere. This study contributes to discussions on media ethics, communication strategies, and policymaking to counteract the risks posed by deepfake technology.

Keywords: *Deepfake, Public Sphere, Social Media, Digital Misinformation, Jurgen Habermas*

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*artificial intelligence/AI*), pembelajaran mendalam (*deep learning*), serta pemrosesan gambar yang semakin pesat telah menghadirkan inovasi yang dapat memberikan manfaat besar bagi berbagai bidang. Namun, di sisi lain, teknologi ini juga membawa ancaman baru, salah satunya adalah *deepfake*. Teknologi *deepfake* pertama kali muncul pada tahun 2017 dalam sebuah forum Reddit, diunggah oleh seorang pengguna anonim dengan nama "deepfakes." Pengguna ini memanfaatkan teknologi pembelajaran mendalam yang dikembangkan oleh Google untuk merekayasa video, awalnya dalam bentuk konten dewasa yang menggunakan wajah individu terkenal tanpa izin (Chatterjee, 2022). Sejak saat itu, teknologi *deepfake* berkembang pesat dan menjadi lebih canggih, sehingga semakin sulit untuk dideteksi keasliannya.

Deepfake berasal dari kata "*deep learning*" yang berarti pembelajaran mendalam dan "*fake*" yang berarti palsu. *Deepfake* adalah video hiper-realistis yang telah dimanipulasi secara digital untuk menggambarkan seseorang melakukan atau mengatakan sesuatu yang sebenarnya tidak pernah terjadi (Chadha et al., 2021). Teknologi ini memanfaatkan jaringan saraf yang dapat menganalisis kumpulan data besar untuk meniru ekspresi wajah, tindakan, suara, dan intonasi seseorang. Proses pembuatan *deepfake* melibatkan penggunaan dua individu dalam sebuah algoritma pembelajaran

Corresponding author.
Email: boredrobopunk@gmail.com

mendalam untuk menukar wajah, membuatnya tampak seolah-olah individu tersebut melakukan tindakan yang sebenarnya tidak pernah mereka lakukan (Westerlund, 2019).

Keberadaan *deepfake* semakin mengkhawatirkan karena penggunaannya tidak terbatas pada bidang hiburan semata, tetapi juga merambah ke ranah politik, sosial, dan keamanan informasi. Teknologi ini dapat digunakan untuk menyebarkan informasi palsu, menciptakan hoaks yang sulit dibedakan dari kenyataan, serta menyesatkan opini publik (Seow et al., 2022). Salah satu contoh kasus penyalahgunaan *deepfake* yang terkenal terjadi pada tahun 2018, ketika sebuah video mantan Presiden Amerika Serikat, Barack Obama, beredar luas di media sosial. Video tersebut menunjukkan Obama sedang mengucapkan kata-kata yang sebenarnya tidak pernah ia ucapkan. Video ini diunggah oleh akun *BuzzFeedVideo* di YouTube dengan tujuan menunjukkan betapa mudahnya teknologi *deepfake* digunakan untuk menyebarkan disinformasi. Dalam video tersebut, Jordan Peele, yang merupakan aktor dan komedian, menyuarakan pesan peringatan tentang potensi bahaya *deepfake* (Peele, 2018).

Selain ranah politik, *deepfake* juga sering digunakan untuk tujuan yang lebih meresahkan, seperti pencemaran nama baik, pemerasan, dan pelecehan berbasis teknologi. Banyak kasus di mana wajah individu, terutama perempuan, digunakan dalam video dewasa tanpa izin mereka. Hal ini menyebabkan trauma psikologis dan menciptakan ancaman serius terhadap privasi individu. Menurut Hasan & Salah (2019), *deepfake* tidak hanya merusak reputasi individu, tetapi juga dapat digunakan sebagai alat manipulasi untuk kepentingan politik dan ekonomi, termasuk dalam propaganda dan kampanye hitam.

Media sosial, terutama YouTube, telah menjadi salah satu platform utama dalam penyebaran *deepfake*. YouTube, sebagai salah satu situs berbagi video terbesar di dunia, memberikan kemudahan bagi pengguna untuk mengunggah, menonton, dan berbagi video dalam skala global. Menurut laporan dari Kready, Shimray, Hussain, & Agarwal (2020), pengguna YouTube rata-rata menghabiskan 40 menit per hari untuk menonton dan mengomentari video. Dengan jumlah pengguna yang sangat besar, YouTube menjadi lahan subur bagi penyebaran video *deepfake*, baik untuk tujuan hiburan maupun manipulasi informasi.

Dalam konteks komunikasi, media sosial seperti YouTube dapat dikategorikan sebagai ruang publik tempat masyarakat bertukar informasi, berdiskusi, dan membentuk opini. Konsep ruang publik (*public sphere*) yang dikemukakan oleh Jurgen Habermas menekankan pentingnya diskusi publik yang rasional dan bebas dari tekanan eksternal, seperti kepentingan politik atau ekonomi yang mendominasi wacana publik. Menurut Habermas, ruang publik adalah arena di mana individu dapat bertukar gagasan dan membangun opini kolektif yang berlandaskan pada rasionalitas serta klaim kesahihan (Habermas, 1989).

Namun, dalam praktiknya, ruang publik digital seperti YouTube semakin terdistorsi akibat maraknya disinformasi, termasuk *deepfake*. Fenomena ini menggambarkan apa yang disebut Habermas sebagai "refeodalisasi ruang publik," di mana ruang publik yang seharusnya menjadi tempat diskusi rasional malah dikuasai oleh aktor-aktor yang memiliki kepentingan tertentu, seperti pemerintah, korporasi, atau kelompok politik tertentu (Habermas, 1989). Dalam konteks penyebaran *deepfake*, fenomena ini tampak jelas ketika video-video manipulatif digunakan untuk menggiring opini publik tanpa adanya mekanisme verifikasi yang memadai.

Salah satu dampak utama dari penyebaran *deepfake* adalah menurunnya kepercayaan masyarakat terhadap informasi digital. Ketika publik mulai meragukan

kebenaran dari setiap video yang mereka lihat, muncul apa yang disebut sebagai "*the liar's dividend*," yaitu situasi di mana individu yang tertangkap melakukan tindakan tertentu dapat dengan mudah mengklaim bahwa video tersebut adalah *deepfake*, meskipun sebenarnya asli (Ahmed, 2021). Hal ini menimbulkan paradoks dalam ruang publik digital, di mana masyarakat semakin sulit membedakan antara informasi yang valid dan yang telah dimanipulasi.

Melihat kompleksitas permasalahan ini, penelitian ini bertujuan untuk menganalisis bagaimana konsep ruang publik Jurgen Habermas dapat digunakan dalam upaya pencegahan penyebaran penyalahgunaan *deepfake* di media sosial YouTube. Dengan memahami bagaimana diskusi publik di platform digital dapat difasilitasi secara lebih sehat, diharapkan ada solusi konkret dalam menangkal penyalahgunaan *deepfake* serta meningkatkan literasi digital di kalangan masyarakat.

Penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan ilmu komunikasi dan media digital, khususnya dalam memahami bagaimana ruang publik digital dapat difungsikan secara ideal sesuai dengan prinsip-prinsip yang diajukan oleh Habermas. Selain itu, penelitian ini juga bertujuan untuk meningkatkan kesadaran masyarakat terhadap bahaya *deepfake* serta pentingnya verifikasi informasi dalam era digital.

2. METODE PENELITIAN

Penelitian ini menggunakan metode penelitian kualitatif dengan pendekatan deskriptif-analitis. Metode deskriptif-analitis adalah pendekatan penelitian yang berfokus pada pencarian fakta dengan interpretasi yang sistematis dan tepat (Achjar et al., 2023). Metode ini dipilih karena memungkinkan eksplorasi mendalam terhadap fenomena *deepfake* di YouTube dan relevansinya dengan teori ruang publik Jurgen Habermas. Sumber data penelitian ini terdiri atas bahan pustaka, termasuk buku, jurnal ilmiah, dan artikel akademik yang membahas tentang *deepfake*, filsafat komunikasi, serta ruang publik dalam konteks digital.

Teknik pengumpulan data dilakukan melalui studi literatur, yang mencakup inventarisasi dan kategorisasi data dari berbagai sumber akademik. Data yang diperoleh kemudian diklasifikasikan ke dalam kategori primer dan sekunder untuk dianalisis secara sistematis. Analisis data dilakukan dengan metode hermeneutika filosofis, yang mencakup deskripsi, interpretasi, induksi dan deduksi, holistika, serta refleksi peneliti.

Hasil dari penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang bagaimana konsep ruang publik Habermas dapat digunakan sebagai kerangka kerja dalam menangkal penyalahgunaan deepfake, serta bagaimana masyarakat dapat lebih waspada terhadap fenomena ini melalui peningkatan literasi digital.

3. HASIL DAN PEMBAHASAN

A. Penyalahgunaan Deepfake di Media Sosial Youtube

Deepfake merupakan teknologi kecerdasan buatan (*Artificial Intelligence/AI*) yang menggabungkan, mengganti, atau menimpa wajah serta suara seseorang dalam sebuah video sehingga tampak seolah-olah nyata. Teknologi ini pertama kali muncul pada tahun 2017 di forum Reddit dan berkembang pesat dalam berbagai aplikasi, termasuk hiburan, industri kreatif, serta kejahatan siber (Westerlund, 2019). *Deepfake* bekerja dengan menggunakan dua jaringan saraf buatan dalam sistem yang disebut *Generative Adversarial Networks* (GANs), yang

memungkinkan komputer untuk belajar meniru ekspresi, suara, serta gerakan seseorang dengan tingkat presisi yang tinggi (McCosker, 2022).

Meskipun teknologi ini memiliki manfaat di beberapa bidang, seperti industri film dan forensik digital, *deepfake* juga sering disalahgunakan untuk tujuan yang tidak etis, termasuk penyebaran informasi palsu, pencemaran nama baik, serta manipulasi opini publik. Dengan semakin canggihnya teknologi ini, deteksi *deepfake* menjadi semakin sulit, yang pada akhirnya meningkatkan risiko penyalahgunaan di berbagai platform media sosial, termasuk YouTube.

YouTube adalah salah satu platform berbagi video terbesar di dunia yang memungkinkan pengguna mengunggah, menonton, serta berbagi video dengan mudah. Dengan lebih dari 2 miliar pengguna aktif bulanan, YouTube menjadi ekosistem digital yang sangat berpengaruh dalam pembentukan opini publik (Kready et al., 2020). Namun, kebebasan dalam berbagi video juga membuka celah bagi penyalahgunaan teknologi *deepfake*, di mana video yang telah dimanipulasi dapat dengan cepat menyebar dan menyesatkan audiens dalam skala global.

Deepfake yang tersebar di YouTube sering kali dimanfaatkan dalam berbagai konteks yang berpotensi merugikan. Salah satu penggunaannya adalah dalam manipulasi politik, di mana video *deepfake* digunakan untuk menampilkan politisi atau tokoh publik seolah-olah mengatakan atau melakukan sesuatu yang sebenarnya tidak pernah terjadi. Contoh kasus terkenal adalah video *deepfake* mantan Presiden AS, Barack Obama, yang dibuat oleh BuzzFeed pada tahun 2018 untuk menunjukkan potensi bahaya dari teknologi ini (Peele, 2018). Selain itu, *deepfake* juga digunakan sebagai alat untuk menyebarkan hoaks dan disinformasi, dengan membuat video palsu yang sulit dibedakan dari kenyataan, sehingga menurunkan tingkat kepercayaan publik terhadap informasi digital.

Tidak hanya dalam ranah politik dan informasi, teknologi *deepfake* juga dimanfaatkan untuk eksploitasi dan pencemaran nama baik, di mana individu menjadi korban dalam konten dewasa yang dibuat tanpa persetujuan mereka. Penyalahgunaan ini tidak hanya melanggar privasi, tetapi juga dapat menimbulkan dampak psikologis yang serius bagi korban. Lebih jauh, *deepfake* juga menjadi alat bagi penipuan dan pemerasan, di mana pelaku kejahatan siber menciptakan video palsu yang sangat meyakinkan untuk menipu atau mengancam korban guna mendapatkan keuntungan finansial. Dengan semakin canggihnya teknologi ini, risiko penyalahgunaan *deepfake* semakin meningkat, menuntut adanya regulasi serta peningkatan literasi digital untuk mengantisipasi dampak negatifnya.

Penyebaran *deepfake* di YouTube menimbulkan berbagai dampak negatif yang signifikan, baik dalam aspek sosial, politik, maupun keamanan digital. Salah satu dampak utamanya adalah erosi kepercayaan publik, di mana masyarakat semakin sulit membedakan antara informasi yang valid dan yang telah dimanipulasi. Fenomena ini terkait dengan konsep "*the liar's dividend*", yaitu situasi di mana seseorang dapat dengan mudah mengklaim bahwa rekaman asli adalah *deepfake* untuk menghindari pertanggungjawaban (Ahmed, 2021). Selain itu, *deepfake* juga berkontribusi terhadap polarisasi politik, terutama ketika digunakan untuk tujuan propaganda, misinformasi, dan manipulasi opini publik dalam masa pemilu, yang dapat memperburuk perpecahan sosial.

Dampak lain yang tidak kalah serius adalah ancaman terhadap privasi dan keamanan digital, di mana individu yang menjadi korban *deepfake* dapat mengalami pencemaran nama baik serta pemerasan berbasis konten palsu yang dibuat tanpa persetujuan mereka. Teknologi ini juga menjadi alat bagi penyalahgunaan ekonomi dan finansial, di mana *deepfake* dimanfaatkan dalam skema penipuan keuangan, seperti pemalsuan identitas untuk melakukan transaksi ilegal atau penipuan daring. Dengan semakin canggihnya teknologi *deepfake*, ancaman ini menjadi semakin sulit dideteksi, sehingga menuntut adanya langkah preventif, baik

dalam bentuk regulasi ketat maupun peningkatan literasi digital untuk melindungi masyarakat dari dampak negatifnya.

B. Wacana *Deepfake* dalam Teori *Public Sphere* Jurgen Habermas

Di era digital saat ini, informasi palsu berkembang dengan sangat cepat seiring dengan kemajuan teknologi. Jika sebelumnya berita atau informasi palsu hanya berbentuk teks atau gambar, kini informasi tersebut telah berevolusi menjadi format video yang dikenal sebagai *deepfake*. *Deepfake* merupakan produk dari teknologi *Artificial Intelligence (AI)* yang mampu menggabungkan, mengganti, dan menempelkan potongan gambar atau video untuk membentuk video baru yang tampak autentik. Teknologi ini memanfaatkan algoritma *Generative Adversarial Networks (GANs),* yang terdiri dari dua jaringan saraf buatan: generator dan diskriminator. Dalam prosesnya, video asli digunakan untuk melatih kedua jaringan ini agar semakin mahir dalam menciptakan video yang tampak nyata.

Sebenarnya, teknik manipulasi gambar dan video telah digunakan sejak lama, khususnya dalam industri perfilman. Teknologi *Computer Generated Imagery* (CGI) telah digunakan untuk menggantikan aktor yang berhalangan hadir, seperti dalam film *Fast and Furious 7*, di mana aktor Paul Walker yang meninggal dunia "dihidupkan kembali" dengan bantuan CGI dan pemeran pengganti. Proses ini membutuhkan biaya besar dan teknologi canggih, tetapi seiring berjalannya waktu, teknik serupa kini dapat dilakukan dengan jauh lebih mudah dan murah. Dengan bantuan ponsel pintar, komputer, dan akses internet, siapa pun dapat membuat video manipulasi, yang menyebabkan teknologi *deepfake* berkembang pesat.

Ancaman dari teknologi *deepfake* tidak bisa dianggap remeh. Beberapa risiko yang ditimbulkan antara lain pencurian identitas, penipuan, manipulasi realitas, distorsi kebenaran, dan kebingungan di tengah masyarakat. Penyebaran *deepfake* semakin sulit dikendalikan karena platform media sosial menjadi saluran utama distribusinya, memungkinkan konten manipulatif ini menyebar dengan cepat dan luas. Pada titik ini, penyalahgunaan teknologi *deepfake* menjadi isu yang harus ditangani bersama oleh berbagai pihak.

Dalam teori *public sphere* yang dikemukakan oleh Jurgen Habermas, ruang publik adalah arena diskusi di mana masyarakat dapat membahas isu-isu yang menyangkut kepentingan bersama (Ismoyo, 2025). Penyebaran *deepfake* merupakan salah satu isu yang perlu didiskusikan secara luas dalam ruang publik, karena dampaknya terhadap kepercayaan publik, politik, dan keamanan digital sangat signifikan. Selain partisipasi masyarakat dalam diskusi dan literasi digital, peran birokrasi dan pemerintah juga menjadi krusial dalam menangani penyalahgunaan *deepfake*. Menurut Habermas, intervensi pemerintah dalam kehidupan masyarakat dapat dijelaskan melalui konsep Strukturwandel der Öffentlichkeit atau perubahan struktur ruang publik, di mana kebijakan dan regulasi yang diterapkan dapat membantu mencegah dampak negatif teknologi *deepfake*. Oleh karena itu, sinergi antara masyarakat, media, dan pemerintah sangat diperlukan untuk menciptakan ekosistem digital yang lebih aman dan terpercaya.

Komunikasi yang efektif dalam suatu diskusi dapat tercapai dengan baik apabila didasarkan pada klaim kesahihan (*validity claims*), yaitu klaim kebenaran (*truth*), klaim ketepatan (*rightness*), klaim kejujuran (*sincerity*), dan klaim komprehensibilitas (*comprehensibility*). Media sosial, sebagai ruang komunikasi digital, seharusnya berfungsi sebagai sarana untuk berbagi informasi yang akurat dan membangun pemahaman kolektif (Scroeder, 2018). Namun, di era digital saat ini, media sosial justru menghadapi

tantangan besar, seperti pengaburan fakta, penyebaran disinformasi, serta distorsi realitas, yang salah satunya dipicu oleh keberadaan teknologi *deepfake*.

Penelitian ini menganalisis fenomena *deepfake* dalam media sosial YouTube melalui empat klaim kesahihan yang dikemukakan oleh Jurgen Habermas, sebagai berikut:

a. Klaim Kebenaran (Truth)
   Klaim ini menyoroti apakah informasi yang disampaikan di YouTube mengenai *deepfake* sesuai dengan fakta yang sebenarnya. Dalam konteks ini, peneliti menganalisis apakah video yang dibagikan memberikan pemahaman yang benar mengenai pengertian deepfake, proses pembuatannya, serta dampak yang ditimbulkan, baik dari sisi positif maupun negatif.
b. Klaim Ketepatan (*Rightness*)
   Klaim ini berfokus pada keberhasilan dalam mencapai konsensus terhadap nilai-nilai sosial. Informasi mengenai *deepfake* yang dibagikan di YouTube seharusnya dapat membangun kesepahaman bersama tentang dampak sosial yang ditimbulkan oleh teknologi ini, terutama dalam konteks penyebaran disinformasi, manipulasi politik, dan ancaman terhadap kepercayaan publik.
c. Klaim Kejujuran (*Sincerity*)
   Klaim autentisitas ini mengacu pada sejauh mana koherensi antara niat, ekspresi, dan tindakan seseorang dalam menyampaikan informasi. Dalam konteks *deepfake*, klaim ini mengkaji keaslian ekspresi dan penyampaian informasi oleh aktor atau narator dalam video di YouTube, apakah mereka menyampaikan informasi dengan jujur atau justru memanfaatkan *deepfake* untuk menyesatkan audiens.
d. Klaim Komprehensibilitas (*Comprehensibility*)
   Klaim ini menekankan pada kejelasan dan keterpahaman informasi yang disampaikan. Informasi mengenai deepfake yang dibagikan di YouTube harus dapat dipahami dengan jelas oleh audiens, sehingga masyarakat dapat mengenali ancaman yang ditimbulkan dan mengembangkan kesadaran akan dampak teknologi ini terhadap kepercayaan publik dan demokrasi.

Penerapan klaim kesahihan dalam analisis fenomena *deepfake* di YouTube menunjukkan bahwa keakuratan, kejelasan, dan kejujuran informasi sangat menentukan bagaimana publik memahami serta merespons ancaman teknologi ini. Oleh karena itu, diperlukan upaya peningkatan literasi digital serta regulasi yang lebih ketat untuk memastikan bahwa media sosial tidak menjadi sarana penyebaran informasi yang menyesatkan, tetapi tetap berfungsi sebagai ruang diskusi yang sehat dan demokratis.

B. *Deepfake* dan Ruang Publik: Peran Media Sosial dalam Mewujudkan Diskusi Publik

Deepfake adalah teknologi manipulasi video berbasis *Artificial Intelligence (AI)* yang dapat mengubah perilaku seseorang dalam video secara hiper-realistis. Teknologi ini memungkinkan pemetaan wajah dan suara seseorang untuk disesuaikan dengan individu lain, sehingga orang tersebut tampak seolah-olah mengatakan atau melakukan sesuatu yang sebenarnya tidak pernah terjadi. Proses *deepfake* melibatkan algoritma pembelajaran mendalam (*deep learning*) yang melatih dua rekaman wajah untuk bertukar ekspresi dan gerakan secara meyakinkan (Westerlund, 2019).

Seiring dengan meningkatnya ancaman yang ditimbulkan oleh *deepfake*, ruang publik (*public sphere*) hadir sebagai arena sosial yang berperan dalam melawan

penyebaran disinformasi berbasis teknologi ini. Salah satu bentuk ruang publik modern yang aktif dalam diskusi ini adalah media sosial, terutama YouTube, yang memungkinkan masyarakat berbagi informasi mengenai karakteristik *deepfake*, proses pembuatannya, cara mendeteksinya, serta dampak negatif yang ditimbulkannya.

Menurut Jurgen Habermas, ruang publik merupakan suatu wilayah kehidupan sosial tempat opini-opini terbentuk melalui diskusi yang terbuka bagi semua warga negara. Ruang publik dalam pemikirannya mengacu pada tempat atau forum di mana masyarakat dapat membahas isu-isu yang menjadi kepentingan bersama. Habermas menekankan bahwa ruang publik menghubungkan aspek pribadi (*idion*) dengan aspek komunal (*koinon*) melalui dialog yang rasional (Supartiningsih, 2013).

Ruang publik memiliki tiga ciri utama. Pertama, aktor-aktor yang terlibat tidak berasal dari birokrasi negara, melainkan dari kalangan masyarakat sipil, seperti akademisi, profesional, atau pengusaha (Gamham, 2020). Hal ini sejalan dengan karakteristik YouTube sebagai ruang publik digital, di mana mayoritas penyampai informasi mengenai *deepfake* bukan berasal dari lembaga pemerintahan, melainkan individu independen atau komunitas tertentu. Kedua, dalam ruang publik terjadi pemberdayaan melalui konsep *public use of reason* yang dikemukakan oleh Kant, yaitu penggunaan rasionalitas dalam diskusi tanpa rasa takut. Di YouTube, diskusi tentang *deepfake* sering kali muncul karena adanya keprihatinan terhadap dampak teknologi ini, terutama dalam konteks disinformasi, manipulasi politik, dan kejahatan digital. Ketiga, ruang publik berfungsi sebagai jembatan antara isu-isu pribadi dan kepentingan kolektif, di mana masyarakat dari berbagai latar belakang dapat berdiskusi untuk mencari solusi bersama. Dalam konteks ini, YouTube menjadi salah satu media sosial yang memungkinkan diskusi terbuka mengenai risiko, regulasi, dan mitigasi deepfake, sehingga mendorong kesadaran dan kewaspadaan publik.

Dengan demikian, media sosial seperti YouTube dapat menjadi ruang publik digital yang efektif dalam menyebarkan kesadaran dan mendorong diskusi tentang *deepfake*. Namun, untuk memastikan bahwa informasi yang dibagikan di platform ini tetap valid dan akurat, diperlukan penguatan literasi digital, regulasi yang lebih ketat, serta peran aktif masyarakat dalam mendeteksi dan menangkal *deepfake* guna melindungi kebenaran dan transparansi dalam ekosistem digital.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa *deepfake* merupakan salah satu bentuk disinformasi berbasis kecerdasan buatan yang memiliki potensi merusak kepercayaan publik terhadap informasi digital. Teknologi *deepfake* yang berbasis pada *Generative Adversarial Networks* (GANs) dapat menciptakan video yang sangat meyakinkan dan sulit dibedakan dari yang asli. Dalam konteks teori *public sphere* Jurgen Habermas, media sosial seperti YouTube seharusnya menjadi ruang diskusi publik yang rasional dan transparan. Namun, keberadaan *deepfake* dapat mengancam prinsip ini dengan menyebarkan informasi yang menyesatkan dan mengurangi kredibilitas komunikasi di ruang publik digital. Oleh karena itu, teori *public sphere* dapat digunakan sebagai dasar untuk membangun mekanisme pencegahan penyalahgunaan *deepfake* melalui penguatan literasi digital, pengawasan kebijakan, dan partisipasi aktif masyarakat dalam diskusi publik.

Sebagai langkah konkret dalam upaya pencegahan penyebaran *deepfake*, diperlukan peningkatan literasi digital agar masyarakat mampu mengenali dan

menganalisis informasi secara kritis. Pemerintah dan platform media sosial harus lebih aktif dalam menetapkan regulasi serta mengembangkan teknologi deteksi *deepfake* yang lebih canggih. Selain itu, diperlukan kerja sama antara akademisi, praktisi media, dan pengembang teknologi untuk menciptakan sistem yang lebih aman dalam menyaring konten digital. Masyarakat juga didorong untuk lebih aktif dalam berdiskusi dan berbagi informasi yang benar guna membangun ekosistem informasi yang lebih sehat. Dengan mengadopsi pendekatan ini, diharapkan penyebaran *deepfake* dapat diminimalisir, sehingga ruang publik digital tetap menjadi tempat yang kredibel dan dapat dipercaya oleh masyarakat luas.

Implikasi penelitian ini mencakup aspek teoritis, praktis, dan kebijakan yang berkontribusi terhadap pemahaman tentang *deepfake* serta perannya dalam ruang publik digital, khususnya di media sosial seperti YouTube. Secara teoritis, penelitian ini memperkaya kajian mengenai teori *public sphere* Jurgen Habermas, dengan menunjukkan bagaimana media sosial dapat berfungsi sebagai ruang publik modern yang memungkinkan diskusi terbuka, tetapi sekaligus rentan terhadap disinformasi yang disebabkan oleh *deepfake*. Selain itu, penelitian ini juga menguji relevansi klaim kesahihan informasi dalam komunikasi digital, menyoroti *bagaimana truth, rightness, sincerity*, dan *comprehensibility* menjadi tantangan dalam mendeteksi dan menangkal video manipulatif berbasis AI. Dari sisi praktis, penelitian ini menekankan pentingnya literasi digital, terutama dalam meningkatkan kesadaran masyarakat terhadap bahaya *deepfake* dan membekali mereka dengan kemampuan untuk membedakan antara konten asli dan hasil manipulasi. Penelitian selanjutnya dapat berfokus pada pengembangan metode deteksi *deepfake* menggunakan AI dan *machine learning* untuk meningkatkan keamanan di media sosial. Selain itu, studi lanjutan dapat mengeksplorasi dampak psikologis dan sosial *deepfake* terhadap kepercayaan publik, khususnya dalam politik dan keamanan digital. Dari aspek hukum, penelitian dapat mengkaji regulasi dan kebijakan terkait *deepfake*, termasuk perlindungan privasi dan hak cipta. Terakhir, penelitian mengenai strategi peningkatan literasi digital diperlukan untuk membantu masyarakat mengenali dan menangkal manipulasi *deepfake* secara lebih efektif.

5. REFERENSI

Achjar, K. A. H., Rusliyadi, M., Zaenurrosyid, A., Rumata, N. A., Nirwana, I., & Abadi, A. (2023). *Metode penelitian kualitatif: Panduan praktis untuk analisis data kualitatif dan studi kasus*. PT. Sonpedia Publishing Indonesia.

Ahmed, S. (2021). *Navigating the Maze: Deepfakes, Cognitive Ability, and Social Media News Skepticism. SAGE Journals*, 1-22. doi:10.1177/14614448211019198.

Chadha, A., Kumar, V., Kashyap, S., & Gupta, M. (2021). *Deepfake: An overview.* In *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020* (pp. 557-566). Springer Singapore.

Chatterjee, A. (2022). *Deepfake technology: How and why China is planning to regulate it.* The Hindu.

Garnham, N. (2020). The media and the public sphere. In *The information society reader* (pp. 357-365). Routledge.

Habermas, J. (1989). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society.* Polity Press, Great Britain.

Hasan, R. H., & Salah, K. (2019). *Combating Deepfake Videos Using Blockchain. IEEE Access*, 1.

Ismoyo, S. L. (2025). Kajian Seni Rupa di Ruang Publik dan Pengaruhnya Terhadap Citra Kota Yogyakarta. *Askara: Jurnal Seni dan Desain*, *3*(2), 113-129.

Kready, J., Shimray, S. A., Hussain, M. N., & Agarwal, N. (2020). *YouTube Data Collection Using Parallel Processing. IEEE International Parallel and Distributed Processing Symposium Workshops*, 1119-1122. doi:10.1109/IPDPSW50202.2020.00185.

McCosker, A. (2022). *Making Sense of Deepfakes: Socializing AI and Building Data Literacy on GitHub and YouTube. SAGE Journal*.

Peele, J. (2024, Des 17). *You Won't Believe What Obama Says In This Video!* Retrieved from YouTube: https://www.youtube.com/watch?v=cQ54GDm1eL0.

Schroeder, R. (2018). *Towards a Theory of Digital Media. Information, Communication & Society, 21*(3), 323-339.

Seow, J. W., Lim, M. K., Phan, R. C., & Liu, J. K. (2022). A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing*, *513*, 351-371.

Supartiningsih (2013). *Pluralitas Agama dalam Ruang Publik: Komunikasi Jurgen Habermas Melihat Keberagaman Agama.* Yogyakarta: Fakultas Filsafat UGM.

Westerlund, M. (2019). *The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9*(11), 39-52. doi:10.22215/1282.