



فصلنامه علمی مطالعات حقوق بشر اسلامی



آموزه‌های حقوق بشر و فناوری دیپ‌فیک: واکاوی کاربردهای سوءاستفاده‌گرانه دیجیتال در تعارض با کرامت انسانی، حریم خصوصی و آزادی بیان

عمید محمدی^۱، محمد حسن قدرتی^۲

چکیده

ظهور فناوری دیپ‌فیک (جعل عمیق دیجیتال) به مثابه یکی از کاربردهای هوش مصنوعی مولد، نظم حقوقی معاصر را با چالش‌هایی بی‌سابقه در عرصه صیانت از حقوق بشر مواجه ساخته است. این فناوری با توانایی خلق محتوای شبیه‌سازی‌شده فراواقع‌نما، بنیادهای معرفتی و اخلاقی را هدف قرار می‌دهد و فرایندهای دموکراتیک و امنیت اطلاعاتی را با اختلال جدی روبه‌رو می‌کند. پرسش اصلی پژوهش آن است که گستره حمایت حقوقی در برابر کنشگری ناشناس مجازی تا کجاست و آیا چارچوب‌های هنجاری موجود، پاسخ‌گوی ابعاد نوین این تهدید هستند؟ مقاله با روشی تحلیلی-تطبیقی، پس از مفهوم‌شناسی و بررسی کارکرد دیپ‌فیک، به تحلیل ماهیت و نیز نسبت‌سنجی آموزه‌های حقوق بشر در مواجهه با این پدیده می‌پردازد و سپس چالش‌های حقوقی ناشی از آن را ارزیابی می‌کند. همچنین ابعاد گوناگون نقض حقوق بشر را در سه ساحت کرامت انسانی، حریم خصوصی و آزادی بیان به بوته سنجش می‌نهد. یافته‌ها حاکی از آن است که دیپ‌فیک، فارغ از جنبه فنی، به‌سان ابزاری قدرت‌ساز در دستان کنشگران ناشناس، نظم حاکم بر حقوق بشر را در سطح هنجاری و اجرایی به مخاطره انداخته و پاسخ‌دهی قواعد موجود را با کاستی‌های اساسی روبه‌رو ساخته است. در نهایت، این نوشتار بر ضرورت تدوین چارچوبی فراقانونی، مبتنی بر همکاری فراملی و تفسیر پویا از آموزه‌های حقوق بشری، برای مهار ظرفیت‌های ضدحقوقی این فناوری تأکید می‌ورزد.

کلیدواژه‌ها:

حقوق بشر، دیپ‌فیک، هویت دیجیتال، حقوق بشر، کرامت انسانی، حریم خصوصی، آزادی بیان.

۴۲

دوره ۱۵، شماره ۳، پیاپی ۴۲
پاییز ۱۴۰۵

مقاله پژوهشی

تاریخ دریافت:

۱۴۰۵-۰۴-۱۲

تاریخ پذیرش:

۱۴۰۵-۰۷-۰۲

صص: ۱۶-۰۱

شابا چاپی: ۵۶۳۷-۲۳۲۲

اسکن کنید



۱. استادیار گروه حقوق، دانشگاه جهرم، جهرم، ایران. (نویسنده مسئول) amid_mohammadi@jahromu.ac.ir

۲. استادیار گروه فقه و مبانی حقوق اسلامی، دانشگاه جهرم، جهرم، ایران.

مقدمه و بیان مسأله

پدیده دیپ‌فیک^۱ آموزه‌ای از یادگیری عمیق و جعل است که برای نخستین بار در سال ۲۰۱۷ در پلتفرم ردیت^۲ ظهور یافت. این فناوری در ابتدا با جایگزینی چهره سلبریتی‌ها در ویدئوهای مستهجن آغاز شد و به محض انتشار کد منبع و امکان‌پذیری تولید انبوه، از حوزه سرگرمی به عرصه‌ای برای تهدید امنیت اطلاعات، حریم خصوصی، و کرامت انسانی گام نهاد (Thi Nguyen et al., 2022, p. 1). دیپ‌فیک برآیند الگوریتم‌های پیشرفته هوش مصنوعی مولد است که با به‌کارگیری شبکه‌های تخصصی مولد و تحلیل حجم عظیمی از داده‌ها، محتوایی فراواقع‌نما خلق می‌کند که تشخیص اصالت آن برای بیننده عادی دشوار و گاه ناممکن است (Romero Moreno, 2024, p. 297). گسترش تصاعدی این ابزارها گویای عمق تهدید است؛ تا سال ۲۰۲۴، شمار ابزارهای تعویض چهره و هماهنگ‌سازی لب با صدا به ۲،۲۹۸ مورد، ابزارهای تولید تصویر به ۱۰،۲۰۶ مورد، و ابزارهای شبیه‌سازی صدا به ۱،۰۱۸ مورد رسیده است (Prasad & Jain, 2025, p. 93). این ارقام، بیانگر دموکراتیزه شدن بی‌سابقه قابلیت‌های فریبنده در دسترس عامه است.

دامنه آثار زیان‌بار دیپ‌فیک، از انتشار اطلاعات نادرست انتخاباتی و باج‌خواهی دیجیتال، تا تولید محتوای سوءاستفاده جنسی از زنان و کودکان و تخریب حیثیت اشخاص، طیف گسترده‌ای از حقوق بنیادین را نشانه رفته است (Weitzman, 2020: Online). دیپ‌فیک، با تحریف حقیقت و جایگزینی روایت جعلی به جای واقعیت، نه تنها حیثیت قربانی را در معرض خطر قرار می‌دهد، بلکه اعتماد او را به دیگری، به نظام حقوقی، و حتی به ادراک خویش از هویت خویش، متزلزل می‌سازد. درهم‌تنیدگی این تهدیدات با نظارت مبتنی بر هوش مصنوعی و جمع‌آوری انبوه داده‌ها، نشان از بروز پارادایمی دارد که در آن، کرامت انسانی و خودمختاری فردی در برابر منطق الگوریتمی بازنمایی و دستکاری، به مرز فروپاشی مفهومی نزدیک می‌شوند (Ahmed & Choudhary, 2026, P.312).

پذیرش گسترده فناوری‌های دیجیتال، به‌تنهایی، تبیین‌گر دگرگونی‌های عرصه حقوق بشر نیست؛ بلکه آنچه این روند را معنادار می‌سازد، ظهور رویه‌های مشارکتی در بستر این زیست‌بوم فناورانه است. در عین حال، نوآوری در پایش حقوق بشر حاصل تعامل سه‌گانه فناوری‌های دیجیتال، رویه‌های مشارکتی، و دسترسی به داده‌های باز است. لیکن این تحول، حوزه حقوق بشر را از معضلات ساختاری و پایدار خود برکنار نکرده است؛ معضلاتی که جنبش فراگیر حقوق بشر همواره با آن دست‌به‌گریبان بوده است (Poblet & Kolieb, 2018, p. 262). از این منظر، دیپ‌فیک نه یک استثنا، که نمودی نوین از این معضلات در عصر دیجیتال محسوب می‌شود.

با عنایت به کاربست‌های مشروع دیپ‌فیک در حوزه‌هایی نظیر سینما، آموزش، هنر، طنز و دسترس‌پذیری، هرگونه تعمیم حکم تقابل به تمامی مصادیق این فناوری، از دقت علمی به دور است. ازاین‌رو، پرسش محوری پژوهش حاضر نه بر «ذات» فناوری، که

1. Deepfake
2. Reddit



بر کاربرد نامشروع آن متمرکز است و در صدد پاسخ به این سؤال برآمده است که آیا کاربردهای مخرب دیپفیک با آموزه‌های بنیادین حقوق بشر تعارض دارند و چارچوب‌های حقوقی موجود تا چه میزان در پاسخ‌گویی به این تعارض کارآمد هستند؟ پژوهش پیش‌رو با هدف پاسخ به این پرسش و کاوش در خلأهای هنجاری موجود، با روشی تحلیلی-تطبیقی و با اتکا به رهیافتی میان‌رشته‌ای، به واکاوی ابعاد گوناگون دیپفیک در پرتو موازین حقوق بشر می‌پردازد. این مقاله، ضمن معرفی فناوری مذکور و گونه‌شناسی مصادیق آن، ماهیت حقوقی این پدیده و تعامل و ارتباط آن با حقوق بشر را امکان‌سنجی و سپس پیامدهای مخرب آن را بر حقوق بشر شناسایی می‌کند. استدلال اساسی مقاله بر آن است که پاسخ‌گویی کارآمد به این تهدید نوین، مستلزم گذار از راهبردهای صرفاً تقنینی مبتنی بر جرم‌انگاری موردی، به سوی تدوین چارچوبی چندلایه است که در آن، همکاری فراملی، تنظیم‌گری مشارکتی پلتفرم‌ها و تفسیر پویا از آموزه‌های حقوق بشر، به‌عنوان مؤلفه‌های مکمل نظام حقوقی موجود، در کنار یکدیگر قرار می‌گیرند. چنین چارچوبی، ضمن حفظ خودمختاری انسانی و تمامیت هویتی، می‌تواند از طریق بازاندیشی در مبانی و انطباق با مقتضیات عصر دیجیتال، کارآمدی نظام حمایت از حقوق بشر را در مواجهه با کاربردهای سوءاستفاده‌گرانه‌ی دیپفیک، ارتقا دهد.

۱. ملاحظات کلی پیرامون فناوری دیپفیک

در ابتدا بحث لازم است تا شناخت جامعی از پدیده دیپفیک و عملکرد آن ارائه شود. برای این منظور در دو قسمت به مفهوم و سرگذشت این فناوری و نیز کاربردها و مصادیق آن می‌پردازیم.

۱-۱. مفهوم‌شناسی و سیر تکوین

پدیده دیپفیک به‌عنوان برون‌داد پیشرفته هوش مصنوعی مولد، دستکاری‌هایی بس واقع‌گرایانه در قالب محتوای صوتی، تصویری یا ویدئویی محسوب می‌شود که با چنان مهارت و دقتی پرداخته شده‌اند که افراد یا رویدادهایی را در موقعیت‌ها و گفتارهایی بازنمایی می‌کنند که هرگز تحقق نیافته است. این ساخته‌های فریبنده، به سبب شباهت شگفت‌انگیزشان به واقعیت، ظرفیتی انکارناپذیر در گمراه‌سازی مخاطبان و باورپذیر ساختن اطلاعات نادرست یا سناریوهای ساختگی دارند (Singh, 2023, pp. 3-5). بر این بنیان، اصطلاح دیپفیک، بیانگر طیف کاملی از جعل‌های دیجیتالی است که امکان دستکاری تصویر به‌منظور جایگزینی چهره، بدن یا صدای یک شخص با شخص دیگر را فراهم می‌آورد (Bosher, 2025, p. 3). این فناوری به‌گونه‌ای طراحی شده است که به‌طور مستمر عملکرد خود را ارتقا می‌دهد؛ از این رو، الگوریتم یاد می‌گیرد و بهبود می‌یابد و خروجی‌ها را به‌طور فزاینده‌ای به واقع مشتبه می‌سازد.

به هر تقدیر، قانون هوش مصنوعی اتحادیه اروپا^۱ مصوب ۲۰۲۴، در ماده ۳ بند (۶۰)، دیپفیک را چنین تعریف می‌کند: «محتوای ترکیبی یا دست‌کاری‌شده تصویری، صوتی، یا ویدئویی که به‌گونه‌ای فریبنده، اصیل یا معتبر به نظر رسیده و شباهت به افراد، مکان‌ها، اشیاء یا سایر رویدادها یا موجودیت‌های موجود داشته باشد» (Labuz, 2024, p. 787). این تعریف، اگرچه

1. Artificial Intelligence Act

گامی مهم در جهت هنجارگذاری محسوب می‌شود، اما از جهاتی با کاستی‌هایی روبه‌رو است که به‌ویژه در نسبت با حقوق بشر، چالش‌آفرین خواهد بود. از جمله، این تعریف به‌صراحت به موضوع «رضایت» فردی اشاره نمی‌کند؛ در حالی که در بسیاری از موارد نقض حقوق بشر از طریق دیپ‌فیک، مسئله فقدان رضایت و نقض خودمختاری ارادی فرد، هسته مرکزی تعرض به کرامت انسانی را تشکیل می‌دهد. همچنین، تعریف مذکور، به‌ویژه در مصادیقی که دیپ‌فیک در خدمت طنز سیاسی، نقد اجتماعی، یا آثار هنری قرار می‌گیرد، با دشواری تفکیک محتوای مخرب از محتوای مشروع روبه‌رو است و مرزهای آزادی بیان را مبهم می‌سازد (Thakur, 2025, p. 3). این ابهامات، ضرورت ارائه تعریفی جامع‌تر و مبتنی بر معیارهای حقوق بشری را دوچندان می‌کند.

مکمل این تعریف، ماده ۵۰ قانون هوش مصنوعی اتحادیه اروپا است که از ۲ اوت ۲۰۲۶ لازم‌الاجرا شده و بر شفافیت و اعلام مصنوعی بودن محتوا تأکید دارد. این ماده، تکلیف می‌کند که سامانه‌های هوش مصنوعی تعاملی با انسان، ماهیت مصنوعی خود را اعلام کنند و سامانه‌های تولیدکننده محتوای ترکیبی صوتی، تصویری، ویدئویی یا متنی، خروجی خود را به‌نحوی قابل تشخیص و ماشین‌خوان برچسب‌گذاری نمایند. به‌عبارت دیگر، ماده ۵۰ در کنار تعریف ماده ۳، زمینه‌های یک نظام تنظیم‌گری شفاف‌محور را فراهم می‌آورد که در آن، تشخیص محتوای تولیدشده از محتوای اصیل، برای مخاطب ممکن می‌شود و بدین ترتیب، امکان پاسخگویی و حمایت از حقوق اشخاص در برابر کاربست‌های سوءاستفاده‌گرانه فراهم می‌آید.

گفتنی است که نقطه عطف این فناوری، معرفی شبکه‌های تخصصی مولد^۱ در سال ۲۰۱۴ بود که با ساختاری متشکل از دو شبکه عصبی رقیب، یعنی مولد^۲ و متمایزکننده^۳، و فرایند آموزشی مبتنی بر رقابت، تحولی بنیادین در کیفیت محتوای ساخته‌شده در قلمروهای تصویر، ویدئو و متن پدید آورد (Chesney & Citron, 2019, P. 1788). در ادامه، گسترش شبکه‌های عصبی عمیق، به‌ویژه شبکه‌های پیچشی در پردازش داده‌های تصویری و شبکه‌های بازگشتی در تولید داده‌های ترتیبی نظیر صوت و نوشتار، دقت و قابلیت‌های این سامانه‌ها را به مراتب افزایش داد. در تازه‌ترین تحولات، ظهور دیپ‌فیک‌های چندوجهی که با تلفیق همزمان صدا، ویدئو و متن، محتوای جعلی همه‌جانبه و فریبنده‌ای می‌آفرینند، مرزهای واقع‌گرایی را درنور دیده و تشخیص‌پذیری مصنوعات را به شدت دشوار ساخته است.

۱-۲. مصادیق و کاربردها

فناوری دیپ‌فیک در دهه گذشته با سرعتی هشداردهنده پیشرفت کرده و در گونه‌های فنی متعددی ظهور یافته است به نحوی که با دستکاری حرکات و رفتار توأم با تغییر ژست‌ها، سناریوهای تخیلی یا تحریف‌شده را به‌مانند واقعیت عرضه می‌دارد (Van der Sloot & Wagenveld, 2022, p. 2). مهمترین این گونه‌ها عبارتند از تعویض چهره^۴ که شامل جایگزینی شمایل یک شخص با چهره شخص دیگر در یک عکس یا محتوای ویدئویی است؛ این امر اغلب با بهره‌گیری از شبکه‌های تخصصی مولد

1. Generative adversarial networks

2. Generator

3. Discriminator

4. Face Swapping



و مدل‌های عمیق یادگیری، به واقع‌گرایی خیره‌کننده‌ای دست می‌یابد (Singh, 2023, p. 6). مصداق دیگر، هم‌نشانی لب^۱ است که شخص را در موقعیتی نشان می‌دهد که گویا سخنی را ادا می‌کند که در واقع از او صادر نشده است؛ همچنین، تکنیک عروسکی^۲ که به حرکات جعلی یک فرد به‌نحو غیرطبیعی دلالت دارد (Mirsky & Lee, 2020, p. 1). گونه دیگر، شبیه‌سازی صدا^۳ است که در آن، ضبط‌های صوتی ترکیبی با تقلید دقیق از الگوهای گفتاری، لحن، و آهنگ صدای یک شخص خاص تولید می‌شوند و مدل‌های عصبی با تحلیل مؤلفه‌های آکوستیکی، قادر به بازآفرینی صداهای جعلی باورپذیر خواهند بود (Bosher, 2025, p. 5).

گستره مسائل حقوقی ناشی از دیپ‌فیک، از حقوق ادله اثبات دعوی و موضوعات کیفری تا مالکیت فکری، امنیت ملی و حقوق بشر را دربرمی‌گیرد. باین وجود، از منظر حقوقی، تمایز قاطع میان «محتوای ترکیبی تولیدشده به‌یاری هوش مصنوعی»، «دیپ‌فیک غیرمخرب» و «دیپ‌فیک متقلبانه، تحقیرآمیز یا ناقض حق»، برای هرگونه تحلیل هنجاری و تنظیم‌گری هوشمندانه، ضرورتی انکارناپذیر است. محتوای ترکیبی تولیدشده به‌یاری هوش مصنوعی، گسترده‌ترین سطح مفهومی است که شامل تمامی کاربردهای مشروع و نامشروع می‌شود. دیپ‌فیک غیرمخرب، گونه‌ای از این محتواست که با هدف خلاقانه، آموزشی، هنری یا طنزآمیز تولید می‌شود و نه تنها به کرامت، حریم یا حیثیت اشخاص آسیب نمی‌زند، بلکه گاه ارزش افزوده نیز ایجاد می‌کند؛ نمونه‌های آن در سینما (جلوه‌های ویژه)، آموزش، دسترس‌پذیری برای افراد دارای معلولیت، بازسازی تاریخی و طنز سیاسی، شخصی‌سازی آواتارها و بازی‌های رایانه‌ای دیده می‌شود (Mirsky & Lee, 2020, p. 1). در نقطه مقابل، دیپ‌فیک متقلبانه، تحقیرآمیز یا ناقض حق قرار دارد که با هدف فریب، تخریب حیثیت، نقض حریم خصوصی، جعل هویت یا دستکاری افکار عمومی تولید می‌شود و مصادیق آن شامل پورنوگرافی، سخنرانی‌های جعلی از رهبران سیاسی، کلاهبرداری با شبیه‌سازی صدا، اخبار جعلی دو قلدری مجازی است. این تفکیک، از آن رو اهمیت دارد که تنظیم‌گری یکسان و بی‌تمایز می‌تواند به محدودسازی ناموجه آزادی بیان و تولیدات هنری یا طنز مشروع بینجامد (Thakur, 2025, p. 3).

در همین راستا، اهمیت راهبردی دیپ‌فیک در گسترش اطلاعات نادرست، مرهون عواملی چندگانه است: اصالت ظاهری این محتواها، تشخیص داده‌های معتبر از جعلی را دشوار می‌سازد؛ گسترش ویروس‌وار آن، پیش از راستی‌آزمایی، مخاطبان انبوه را درمی‌نوردد؛ و بازنمایی مغایر با واقعیت از رخدادها، ادراک جمعی را دستکاری کرده، حیثیت اشخاص را خدشه‌دار می‌سازد (Singh, 2023, pp. 3-5). این ویژگی‌ها، دیپ‌فیک را به ابزاری چندوجهی برای تضعیف اعتماد عمومی و بنیان‌های معرفتی جوامع دموکراتیک بدل ساخته است. در نهایت، آنچه دیپ‌فیک را به چالشی فراتر از مسائل فنی یا کیفری تبدیل می‌کند، نقش آن

-
1. Lip Syncing
 2. Puppet Technique
 3. Voice Cloning



به‌سان «سرقت هویت»^۱ در معنای وجودشناختی است؛ بدین‌معنا که نه تنها تصویر یا داده‌های فرد به سرقت می‌رود، بلکه امکان تعریف خود و نسبت او با جهان سلب می‌شود.

۲. تحلیل ماهیت حقوقی دیپ‌فیک در پرتو حقوق بشر

اهمیت حقوقی پدیده دیپ‌فیک نه در فرایند فنی آفرینش آن، که در کاربست‌های اجتماعی و دلالت‌های هنجاری برآمده از آن جلوه‌گر می‌شود. اطلاق برچسب «دیپ‌فیک» بر هر محتوای تولیدشده توسط هوش مصنوعی، به‌سبب قراین عرفی همراه با آن (نظیر فریب، جعل هویت و نقض حریم شخصی)، ماهیت اثر را از یک «ساخته فنی بی‌طرف» به «ابزار بالقوه زیان‌بار» در نظم حقوقی دگرگون می‌سازد (Nath & Chakravarty, 2024: 3). ازاین‌رو، تفکیک مفهومی میان «تولید محتوای ترکیبی به‌یاری هوش مصنوعی» و «دیپ‌فیک به‌مثابه عمل سوءاستفاده‌گرانه از این قابلیت»، نخستین گام در تحلیل حقوقی این پدیده محسوب می‌شود.

با این اوصاف، باید اذعان داشت که دیپ‌فیک، در عمیق‌ترین لایه حقوقی خود، با «حق بر هویت دیجیتال» گره خورده است؛ حقی که در زمره حقوق شخصیتی و کرامت‌محور قرار می‌گیرد و حمایت از آن، پیش‌نیاز تحقق خودمختاری فردی در فضای مجازی محسوب می‌شود (Snow, 2025: p. 1010). استفاده غیرمجاز از تصویر، صدا یا حرکات اشخاص، نه‌تنها مصداق بارز «غصب هویت دیجیتال»^۲ به‌شمار می‌آید، بلکه به‌دلیل قابلیت استناد کاذب به افراد، زمینه‌ساز نقض حق بر آبرو و حیثیت (مصادیق حریم خصوصی) و همچنین حرمت نفس (به‌عنوان جلوه‌ای از حقوق شخصیتی اجتناب‌ناپذیر) می‌گردد (Kocsis, 2022, p. 621)؛ در این برساخت، دیپ‌فیک، پیش از آنکه مسئله‌ای در حوزه حقوق فناوری‌های نوین باشد، تهدیدی برای تمامیت هویتی اشخاص در فضای مجازی و به‌تبع آن، کرامت انسانی آنان محسوب می‌شود. رویه قضایی نیز بر این نکته تأکید دارد که سوءاستفاده از شباهت و اصوات اشخاص، حتی در بسترهای ساختگی، نوعی تعرض به تمامیت هویتی و کرامت روانی فرد تلقی می‌شود و از مصادیق بارز نقض حق کرامت در برخی پرونده‌ها^۳ به‌شمار می‌آید.

در پایان باید خاطرنشان کرد که ماهیت رسانه ترکیبی^۴ دیپ‌فیک، توان پاسخ‌گویی از طریق سازوکارهای سنتی مسئولیت مدنی یا کیفری را ندارد. ازاین‌رو، نظام حقوقی کارآمد در قبال این پدیده، ناگزیر از گذار از مسئولیت فردی به مسئولیت زنجیره‌ای واسطه‌ها (شامل پلتفرم‌ها، میزبانان محتوا و توسعه‌دهندگان ابزارهای هوش مصنوعی) است و می‌بایست با تکیه بر اصل احتیاط و تعدیل بار اثبات به‌سود زیان‌دیده، ضمانت‌اجراهای مؤثری را در عرصه دادرسی الکترونیک فراهم آورد (Shukla, 2025. P.

۱. Identity theft

۲. Digital Identity Usurpation

۳. John K. Modise v. Botswana, African Commission on Human and Peoples' Rights, Comm. No. 97/93 (1997); Purohit and Moore v. The Gambia, African Commission on Human and Peoples' Rights, Comm. No. 241/2001 (2003).

۴. Synthetic Media



12). این رویکرد، هرچند گامی ضروری به سوی جبران ناکارآمدی قواعد فعلی به شمار می‌رود، اما به شرطی ثمربخش خواهد بود که با ظرافت حقوقی کامل، میان وظایف فزاینده واسطه‌ها و حفظ حقوق کاربران تعادل برقرار کند.

۳. چالش‌های فناوری دیپفیک در تقابل با حقوق بشر

حقوق بشر مجموعه اصول و قواعدی است که بر اساس آن، هر انسانی صرفاً به دلیل انسان بودن، از حقوقی برابر، ذاتی، سلب‌ناپذیر و جهانی برخوردار است. این حقوق پایه‌های کرامت، آزادی، برابری و عدالت را برای همه افراد، بدون تبعیض از نظر نژاد، رنگ، جنسیت، زبان، مذهب، عقیده، ملیت، ثروت یا هر موقعیت دیگر، تضمین می‌کند (شریفی طرازکوهی، ۱۳۹۷، ص. ۶۴). در همین راستا، تأثیر حقوق بشر و فضای مجازی بر عرصه‌های گوناگون حیات اجتماعی، عمیق، فراگیر و ماندگار می‌نماید. این تأثیر، متضمن گستره‌ای از حقوق و آزادی‌های بنیادینی است که دولت‌ها را مکلف به پاسداشت و تضمین عملی این حقوق می‌نماید (استیون، ۱۹۵۹، ص. ۱۲).

پیش از ورود به تبیین چالش‌هایی که فناوری دیپفیک برای حقوق بشر پدید آورده است، باید به این نکته توجه داشت که در این وادی آنچه محل تأمل است، ذات فناوری نیست، بلکه الگوهای بهره‌برداری مخرب از آن است که با هدف فریب، تخریب حیثیت، یا نقض خودمختاری فردی صورت می‌گیرند. به دیگر سخن، دیپفیک فی‌نفسه فناوری خنثی و با کاربرد دوگانه (مثبت-منفی) است؛ اما در کاربردهای خاص می‌تواند ناقض حقوق بشر باشد. از این رو، ملاک تشخیص تعارض با حقوق بشر، قصد عامل، زمینه کاربرد و پیامد آن است و فقدان رضایت، وجود فریب، و ایجاد ضرر، مرزهای اصلی این تشخیص را تشکیل می‌دهند. بدین سان، آنچه دیپفیک را در نظام حقوق بشر به پدیده‌ای شاخص از سایر جرائم سایبری تبدیل می‌کند، نه صرفاً دقت فزاینده آن، بلکه توانایی در به چالش کشیدن بنیادهای معرفتی و هنجاری این نظام است.

دیپفیک با خلق «حقیقت‌های جایگزین»، مرز میان جعل و واقعیت را چنان درهم می‌آمیزد که امکان احقاق حق در فرایندهای قضایی را با دشواری مضاعف مواجه می‌سازد؛ هم‌زمان، با تسهیل نظارت مبتنی بر هوش مصنوعی و تولید محتوای فریبنده برای مهندسی افکار عمومی، نه تنها حریم خصوصی و آزادی بیان، بلکه بسترهای شکل‌گیری اراده‌ی جمعی و مشارکت دموکراتیک را نیز تخریب می‌نماید (Murray, 2025, p. 132). از این منظر، دیپفیک نه تهدیدی حاشیه‌ای، بلکه نمودی عینی از بحران‌های ساختاری حقوق بشر در عصر پساحقیقت است که بازتعریف مفاهیم بنیادینی چون «حقیقت»، «هویت» و «رضایت» را به عنوان پیش‌نیازی اجتناب‌ناپذیر برای نظریه‌پردازی حقوق بشر نوین مطرح می‌سازد.

در بعد حقوقی، آنچه مسئله را دشوارتر می‌کند، نبود صراحت قانونی در مواجهه با دیپفیک و ناکارآمدی چارچوب‌های نظارتی موجود در پاسخ به ابعاد نوظهور آن است (Saini, 2026, p. 167). این ناکارآمدی در سه سطح خود را آشکار می‌سازد: نخست، در عرصه سیاسی، دیپفیک با تولید سناریوهای بصری باورپذیر از شخصیت‌های عمومی، هم‌زمان دو رکن اساسی حاکمیت مبتنی بر اراده‌ی مردم را هدف می‌گیرد. از یک سو با تخریب سرمایه نمادین نامزدها، موجبات سردرگمی و بی‌اعتمادی

رأی‌دهندگان را فراهم می‌آورد (Ramluckan, 2024, p. 285) و از سوی دیگر با ایجاد بحران معرفت‌شناختی در تشخیص واقعیت از جعل، زمینه مشروعیت‌زدایی از نتایج آرا را مهیا می‌سازد؛ دوم، در حوزه مسئولیت مدنی و کیفری، افترای مطبوعاتی از طریق دیپ‌فیک، به دلیل گستردگی بی‌سابقه‌ی انتشار، دوام اثر، دشواری اثبات سوءنیت در زنجیره‌ی تولید تا نشر، و امکان‌ناپذیری جبران خسارت‌های ناملموس حیثیتی در بازه زمانی حساس انتخاباتی، نظام‌های حقوقی موجود را با خلأهای ساختاری جدی روبه‌رو ساخته است؛ سوم، در فرایند دادرسی، بار اثبات در مراجع قضایی به دلیل ماهیت فنی دیپ‌فیک و دشواری انتساب محتوای جعلی به عاملان خاص، همراه با مسائل فرامرزی و صلاحیت قضایی، فرایند اجرا را به شدت پیچیده می‌کند و قربانیان را عملاً از دستیابی به عدالت محروم می‌سازد (Van der Sloot & Wagenveld, 2022, p. 2).

چالش دیگر، ظهور دیپ‌فیک در قالب مصداقی نوین از خشونت جنسی تسهیل‌شده توسط فناوری (TFSV) است که با هدف‌گذاری نامتناسب بر پیکر و هویت زنان، سه سطح بنیادین حقوق بشر را هم‌زمان مورد تعرض قرار می‌دهد: حق بر حریم خصوصی جنسی که با تولید و توزیع محتوای صریح بدون رضایت، خودمختاری بدنی قربانیان را سلب می‌کند، کرامت انسانی و تمامیت روانی-اجتماعی که در پیوند با پیامدهایی چون اضطراب مزمن، انزوای حرفه‌ای و کناره‌گیری اجباری از فضای عمومی (که در مطالعات کیفی از آن به «اثر حجاب دیجیتال» تعبیر شده) عاملیت زنان را در حیات مدنی و سیاسی سلب می‌نماید (De Vido, 2024; Thakur, 2025)؛ و اصل برابری و عدم تبعیض مصرح در اسناد بنیادین (از جمله ماده‌ی ۲ اعلامیه‌ی جهانی حقوق بشر و کنوانسیون CEDAW) که در نتیجه‌ی عادی‌سازی خشونت بصری و بازتولید کلیشه‌های جنسی‌زده، نه تنها نقض می‌شود، بلکه به عاملی برای تشدید شکاف‌های ساختاری قدرت بدل می‌گردد (Johnson, 2025: P. 19). آنچه این چالش را به وضعیتی اضطراری ارتقا می‌دهد، دوام و گستردگی بی‌سابقه محتوای تولیدشده و ناتوانی چارچوب‌های کیفری سنتی در احراز سوءنیت و ارائه‌ی جبرانی مؤثر است؛ از این رو، گزارش‌گر ویژه سازمان ملل در امور خشونت علیه زنان، بر گذار از نگاه صرفاً کیفری-دیجیتال به رویکردی مبتنی بر حقوق بشر تأکید دارد که در آن، پاسخ‌های ساختاری (همچون بازطراحی پلتفرم‌ها با پیش‌فرض رضایت و ایجاد نهادهای داور تخصصی) جایگزین واکنش‌های موردی و ناکارآمد گردد و خشونت جنسی‌زده را در کانون سیاست‌گذاری‌های کلان حقوق بشر قرار دهد.^۱

۴. بررسی نقض ابعاد حقوق بشر در تقابل با فناوری دیپ‌فیک

در این مجال، زوایای مختلف نقض حقوق مرتبط با بشر در خصوص بنیادهای کرامت انسانی، حریم خصوصی و تعارض با آزادی بیان تبیین می‌گردد. در همین جا و قبل از شروع بحث باید به این نکته اشاره کرد که با توجه به پیوستگی و به هم‌وابستگی حقوق بشر، دیپ‌فیک تمامی این مبانی فکری را به‌طور بالقوه تحت‌الشعاع قرار می‌دهد (Kocsis, 2022, P. 622)؛ به‌نحوی که نقض یک حق (مانند حریم خصوصی) زنجیره‌ای از نقض سایر حقوق (از جمله آزادی بیان، مشارکت سیاسی و حق بر عدالت) را

۱. UN Special Rapporteur, 2021: <https://www.ohchr.org/en/press-releases/2021/10/gender-equality-freedom-expression-remains-distant-goal-un-expert>



به دنبال دارد. این شبکه به هم پیوسته از نقض‌ها، ضرورت اتخاذ رویکردی جامع و سیستمی را آشکار می‌سازد؛ رویکردی که در آن، بازتعریف مفاهیم بنیادین حقوق بشر، انطباق قواعد ادله اثبات با اقتضائات فناوری‌های نوین، و تعیین تکلیف مسئولیت پلتفرم‌ها در قبال محتوای تولیدشده توسط کاربران، به عنوان محورهای اصلی اصلاحات تقنینی و قضایی مورد توجه قرار گیرند.

۴-۱. کرامت انسانی

کرامت انسانی به این معناست که هر انسان، صرف نظر از هر ویژگی ظاهری، موقعیت اجتماعی، توانمندی جسمی یا عقیدتی، دارای ارزش ذاتی و والایی است که قابل کم و زیاد شدن، خرید و فروش یا بخشش نیست. به زعم برخی، حق کرامت انسانی منبعث از سه عنصر برتری، تقدم و احترام انسان است (عباسی، ۱۳۹۰، ص. ۸۵). کرامت انسانی، حق ذاتی و مبنای حقوق بشر در همه جوامع به رسمیت شناخته می‌شود (حبیب‌زاده و فرج‌پور، ۱۳۹۵، ص. ۱۲۲). در مقدمه و ماده ۴-۶ اعلامیه جهانی حقوق بشر و نیز ماده ۱۷(۱) میثاق حقوق مدنی و سیاسی، حمایت از حیثیت و اعتبار اشخاص در برابر هرگونه مداخله تضمین شده است.

بنا به مراتب بالا، دیپفیک، به‌ویژه هنگامی که در خلق تصاویر یا محتوای صمیمی جعلی بدون رضایت به‌کار گرفته شود، به حریم شخصی افراد هجوم برده و تصویر اجتماعی ایشان را دستخوش تحریف می‌سازد. نظریه حقوقی معاصر، سوءاستفاده از شباهت یا اصوات اشخاص را حتی در بسترهای آشکارا ساختگی نوعی تعرض به تمامیت هویتی و کرامت روانی فرد تلقی می‌کند؛ چراکه این عمل، خودمختاری شخص را بر سرنوشت بازنمایی خویش سلب نموده و او را به ابزاری فناورانه تقلیل می‌دهد (Thakur, 2025, p. 3).

به هر روی، کنکاش در ماهیت وجودشناختی این برساخت مقتضی آن است که در جهان متحول امروزی رویکردها و رهیافت‌ها پیرامون «خود»ی دیگرگونه از فرد شکل می‌گیرند که همین امر به نوبه خود، خودمختاری فردی را هدف قرار می‌دهد. این ویژگی، دیپفیک تفاوت مفاهیم سستی افترا یا توهین که مبتنی بر نسبت دادن سخن یا رفتار خاص به فرد هستند را نشان می‌دهد؛ چراکه در دیپفیک، نه تنها یک سخن یا رفتار خاص، بلکه تمامیت هویتی فرد، یعنی آنچه او را به عنوان «این شخص خاص» در جهان تعریف می‌کند مورد تعرض ساختاری قرار می‌گیرد.

از منظر روانشناختی حقوقی، دیپفیک با ایجاد پریشانی روانی عمیق در قربانی، نه تنها کرامت، بلکه حق برخورداری از بالاترین استاندارد قابل دستیابی سلامت جسمی و روانی را که در ماده ۱۲(۱) میثاق حقوق اقتصادی، اجتماعی و فرهنگی به رسمیت شناخته شده، نیز نقض می‌نماید. آنچه این پریشانی را تشدید می‌سازد، سه عامل است: نخست، حس بیگانگی با «خود جعلی» که فرد را در تناقضی وجودی با تصویر خویش قرار می‌دهد؛ دوم، ناتوانی در کنترل انتشار و ماندگاری محتوای جعلی که احساس درماندگی و استیصال را تقویت می‌کند؛ و سوم، مواجهه مکرر با بازنمایی تحریف‌شده خویش در فضای عمومی که فرایند بازسازی هویت را با دشواری مضاعف روبه‌رو می‌سازد (Ramluckan, 2024, p. 282).

نهایتاً، آنچه تحلیل حقوق بشری دیپ‌فیک را از رویکردهای صرفاً کیفری یا مدنی تفکیک می‌سازد، تأکید بر کرامت به‌مثابه «اصل مادر» تمامی حقوق بشر است. از این منظر، نقض کرامت از طریق دیپ‌فیک، نه صرفاً جرمی علیه فرد، بلکه تعرض به بنیادی‌ترین ارزش نظام حقوق بشر معاصر تلقی می‌شود. این نگرش، بازتعریف مفاهیم سنتی نظیر «رضایت»، «حریم خصوصی» و «شخصیت» را در پرتو چالش‌های عصر دیجیتال ضروری می‌سازد؛ بازتعریفی که در آن، خودمختاری فرد بر تصویر و هویت دیجیتال خویش، در کانون حمایت‌های حقوقی قرار می‌گیرد.

۴-۲. حریم خصوصی

حریم خصوصی یا حق بر خلوت شخص که از آن با عنوان «حق بر تنها بودن»^۱ نیز یاد می‌شود، از دیگر مبانی حقوق بشر به‌شمار می‌آید که در ماده ۱۲ اعلامیه جهانی حقوق بشر و ماده ۱۷(۱) میثاق حقوق مدنی و سیاسی به رسمیت شناخته شده است. گروهی از اندیشمندان قلمرو حریم خصوصی را در چهار مقوله تقسیم‌بندی می‌کنند: (۱) ورود یا تعرض به حریم خصوصی یا انزوای خواهان، یا دخالت در امور شخصی وی بدون مجوز؛ (۲) افشای عمومی حقایق خصوصی که موجب خجالت و آبروریزی خواهان می‌شود. (۳) بزرگنمایی یا انعکاس نادرستی از شخصیت خواهان که وی را در ملاء عام به گونه‌ای ناصواب معرفی کند و (۴) بهره‌برداری از نام، چهره، یا شباهت خواهان برای منافع خواننده، بدون رضایت وی (Solove citing William

Prosser, 2006, p. 482)

با کاربست این طبقه‌بندی، کاربردهای مخرب دیپ‌فیک را می‌توان در دو مقوله اخیر جای داد. پرونده رانا آیوب^۲، روزنامه‌نگار هندی، نمونه‌ای شاخص از این نقض حق است: پس از انتشار گزارش‌های وی درباره فساد در احزاب سیاسی هند، ویدئوی دیپ‌فیکی با هدف تخریب هویت وی منتشر شد که او را فردی «خوش‌گذران، بی‌اخلاق و کالایی آسیب‌دیده» معرفی می‌کرد (Chesney & Citron, 2019, p. 1789).

دیپ‌فیک‌های بدون رضایت، به‌ویژه در حوزه پورنوگرافی، مصداق بارز تجاوز به حریم خصوصی جنسی هستند. برخی معتقدند که این اعمال، کنترل افراد را بر شباهت خود سلب کرده و کرامت آنان را به شیوه فاحش نقض می‌کنند. از منظر تحلیل حقوق بشری، دیپ‌فیک پورنوگرافیک در یک اقدام واحد، دست‌کم چهار لایه حریم خصوصی را هدف می‌گیرد: حریم جسمانی از طریق بازنمایی غیرمجاز بدن، حریم اطلاعاتی از طریق افشای داده‌های بیومتریک، حریم تصمیم‌گیری از طریق سلب حق فرد بر تعیین سرنوشت تصویر خویش، و حریم ارتباطی از طریق تخریب روابط شخصی و حرفه‌ای قربانی (Chesney & Citron, 2019, p. 1893).

با این حال، دشواری در به‌کارگیری حقوق حریم خصوصی این است که این حقوق تنها در جایی اعمال می‌شوند که «انتظار معقولی» از حریم خصوصی وجود داشته باشد. از آنجا که معمولاً در مکان‌های عمومی، چنین انتظاری وجود ندارد، استفاده از

1. Right to be alone
2. Rana Ayyub



تصاویر ضبط‌شده در فضای عمومی برای ساخت دیپ‌فیک، مشمول حمایت‌های سستی حریم خصوصی نمی‌شود (Bosher, 2026, p. 8). این ناکارآمدی، ریشه در پارادایم سستی مبتنی بر دوگانگی «عمومی-خصوصی» دارد؛ دوگانگی که در عصر دیجیتال و با گسترش داده‌های شخصی در فضای مجازی، عملاً فروپاشیده است. از این منظر، بازتعریف «انتظار معقول از حریم خصوصی» در پرتو قابلیت‌های فناوری‌های نوین، یکی از مهم‌ترین چالش‌های پیش روی نظام‌های حقوقی محسوب می‌شود.

۴-۳. آزادی بیان

آزادی بیان، یکی از بنیادی‌ترین حقوق بشر، در ماده ۱۹ اعلامیه جهانی حقوق بشر و ماده ۱۹ میثاق حقوق مدنی و سیاسی به رسمیت شناخته شده است.^۱ این حق که به افراد اجازه می‌دهد اندیشه‌ها، عقاید و اطلاعات خود را بدون ترس از مجازات یا سانسور بیان و دریافت کنند، ارزش‌هایی چون خودمختاری شخصی و مردم‌سالاری را تأمین می‌کند (عباسی، ۱۳۹۰، ص. ۱۵۲) و پیش‌شرط ابراز وجودی انسان و بسط هویت فردی در جامعه محسوب می‌شود (فرانک یورگنسون، ۱۳۸۶، ص. ۷۸). باین‌حال، پرسش اساسی آن است که آیا ساخت و استفاده از دیپ‌فیک می‌تواند مصداقی از حق آزادی بیان باشد؟ پاسخ به این پرسش، نیازمند تفکیک میان کاربردهای مشروع و سوءاستفاده‌گرانه است؛ چراکه آزادی بیان امری مطلق نیست و حقوق بین‌الملل، برای حمایت از حقوق و حیثیت دیگران، امنیت ملی، نظم عمومی و سلامت عمومی، محدودیت‌ها را مجاز می‌شمارد.

در یک سو، کاربردهای مشروع دیپ‌فیک شامل مصادیقی چون طنز سیاسی، پارودی، هنر دیجیتال، نقد اجتماعی، مستندهای اصیل، درام‌های مستند (که واقعیات را با تفسیر خلاقانه ترکیب می‌کنند) و بازسازی تاریخی می‌شود. این مصادیق، در صورتی که با رعایت شفافیت (اعلام مصنوعی بودن محتوا) و بدون قصد فریب یا تخریب حیثیت تولید شوند، در دایره بیان مشروع قرار می‌گیرند و از حمایت کامل آزادی بیان برخوردارند (Shtefan, 2024: 37). در سوی مقابل، مصادیق سوءاستفاده‌گرانه‌ای که در قلمرو حمایت آزادی بیان قرار نمی‌گیرند، عبارتند از: دیپ‌فیک‌هایی که با هدف کلاهبرداری، پورنوگرافی، تخلفات انتخاباتی، تحریک به خشونت، یا ارائه به‌عنوان مستند در حالی که عمدتاً تخیلی هستند و شخص را در فضایی نادرست به‌تصویر می‌کشند (Thakur, 2025, p. 3).

در عین حال، معیار تشخیص میان بیان مشروع و رفتار زیان‌بار، ترکیبی از چند مؤلفه است که در رویه قضایی و قوانین ایالتی تبلور یافته است: نخست، معیار اصالت در برابر تخیل‌پردازی که بر اساس آن، اثری که به‌عنوان رمان (با آگاهی مخاطب از ساختگی بودن) ارائه شود، حمایت می‌شود، اما اثری که در لباس واقعیات ارائه شود و واقعیت را تحریف کند، حمایت نمی‌شود. دوم، معیار استفاده تحول‌آمیز که بر اساس آن، اگر اثر «عناصر خلاقانه قابل توجهی فراتر از یک همانندی صرف» داشته باشد و به‌گونه‌ای دگرگون شده باشد که عمدتاً به بیان خودِ خالق تبدیل شود، مشمول حمایت است؛ در غیر این صورت، بهره‌برداری غیرمجاز از هویت محسوب می‌شود. سوم، تمایز هدف اصلی اثر که اگر هدف اصلی، بیان سیاسی یا اجتماعی باشد، حتی با وجود ظاهر شدن در آگهی تجاری، حمایت می‌شود، اما اگر هدف اصلی تبلیغ محصول باشد و عنصر طنز در درجه دوم قرار گیرد، حمایت نمی‌شود.

1. Iccpr (International Covenant on Civil and Political Rights), 1966, Article.19

چهارم، تفاوت اثر منفرد از تولید انبوه؛ به گونه‌ای که نقاشی منفرد، منافع اقتصادی شخص را تهدید نمی‌کند، اما گردش تجاری گسترده آن، آسیب اقتصادی وارد می‌آورد و به عنوان بهره‌برداری غیرمجاز تلقی می‌شود. پنجم، وجود یا فقدان رضایت شخص ذی‌نفع، به‌ویژه در مورد اجراکنندگانی که جعل عمیق، فعالیت حرفه‌ای آنان را تقلید می‌کند و باید با اجازه آنان صورت پذیرد (Barber, 2023: 41).

به علاوه، دیپ‌فیک، فراتر از چالش تشخیص مصادیق، تأثیرات نامطلوبی بر گفتمان عمومی دارد. نخستین چالش، تأثیر نامطلوب بر گفتمان عمومی است؛ هنگامی که افراد می‌دانند سخنان یا تصاویر آنان می‌تواند به راحتی دستکاری شده و علیه آنان به کار گرفته شود، احتمال خودسانسوری افزایش می‌یابد. این امر، به‌ویژه برای گروه‌های آسیب‌پذیر همچون زنان، اقلیت‌ها، روزنامه‌نگاران و فعالان مدنی، حق مشارکت در گفتمان عمومی را به طور جدی محدود می‌سازد. دیپ‌فیک نه با سانسور مستقیم، بلکه با تهدید ضمنی تخریب اعتبار، میدان گفتمان را ناامن می‌سازد؛ تأثیری که ماهیتی خاموش و تدریجی دارد اما در بلندمدت با درونی‌سازی ترس، میل به مشارکت عمومی را نشانه می‌گیرد (Westerlund, 2019). دومین چالش، آلودگی فضای اطلاعاتی و فرسایش اعتماد عمومی است؛ دیپ‌فیک‌ها با تولید محتوای جعلی و واقع‌نما، تشخیص حقیقت از دروغ را برای شهروندان دشوار می‌کنند و «آخرالزمان اطلاعاتی» پدید می‌آورند؛ فضایی که در آن، مردم احساس می‌کنند نمی‌توان به هیچ اطلاعاتی اعتماد کرد. این فروپاشی معرفت‌شناختی، پیش شرط اساسی دموکراسی یعنی امکان داوری آگاهانه را از میان برمی‌دارد (Ramluckan, 2024, p. 290). سومین چالش، ابزار بودن دیپ‌فیک برای سرکوب مخالفان است؛ دولت‌ها ممکن است با تهدید به تولید دیپ‌فیک، مخالفان را به سکوت وادارند یا با استناد به وجود دیپ‌فیک‌ها، هرگونه محتوای انتقادی را «جعلی» تلقی کرده و از پاسخگویی طفره روند (Chesney & Citron, 2019). همچنین، واکنش‌های نظارتی بیش‌اندازه می‌تواند خود به نقض آزادی بیان بینجامد (Van der Sloot & Wagenveld, 2022, p. 2).

با همه این‌ها، از منظر حقوق بشر، پدیده دیپ‌فیک نه صرفاً یک چالش فناورانه یا حقوقی محدود، بلکه آزمونی برای سنجش تعهد نظام‌های حقوقی به دو ارزش بنیادین هم‌تراز است: آزادی بیان به عنوان ستون دموکراسی، و حریم خصوصی و کرامت انسانی به عنوان زیربنای شخصیت فردی. آنچه در مواجهه با دیپ‌فیک‌ها ضرورت می‌یابد، نه انتخاب میان این دو ارزش، بلکه ایجاد موازنه‌ای ظریف و اصولی است که در آن، تحدید آزادی بیان تنها در حد ضرورت مطلق و متناسب با شدت آسیب وارده صورت پذیرد. همان‌گونه که اسناد بین‌المللی حقوق بشر، از جمله میثاق بین‌المللی حقوق مدنی و سیاسی، آزادی بیان را مشروط به رعایت حقوق و حیثیت دیگران می‌دانند، مقابله با دیپ‌فیک‌های مخرب، در چارچوب همین استثنائات معقول قابل توجیه است (Bourgault, 2025: 56). تأثیر افتراقی دیپ‌فیک بر گروه‌های به‌حاشیه‌رانده، به‌ویژه زنان، اقلیت‌های قومی، فعالان مدنی و روزنامه‌نگاران متقد، به مراتب بیش از دیگران است و دفاع انتزاعی از آزادی بیان در برابر هرگونه نظارت، نادیده‌انگاشتن نابرابری‌های ساختاری قدرت است؛ نابرابری‌هایی که دیپ‌فیک نه تنها بازتاب‌دهنده، بلکه تشدیدکننده آنهاست. از این رو، چارچوب



حقوق بشر ایجاب می‌کند آزادی بیان متوازن با کرامت و امنیت دیگران به‌کار گرفته شود و محدودیت‌های آن، با هدف حفظ گفت‌وگو دموکراتیک فراگیر و عاری از ارباب‌تدوین گردد (Bourgault, 2025: 67). در نهایت، آنچه از منظر حقوق بشری حائز اهمیت است، نه این پرسش که آیا دیپفیک «بیان» محسوب می‌شود، بلکه این است که چگونه می‌توان با حفظ کرامت انسانی و کارکرد معرفتی دموکراسی، از هرج‌ومرج ناشی از جعل دیجیتال جلوگیری کرد.

نتیجه‌گیری

در یک دهه اخیر، سیر تکاملی فناوری هوش مصنوعی در تولید محتوای جعلی با واقع‌نمایی فزاینده، به‌ویژه از رهگذر الگوریتم‌های یادگیری عمیق، جهشی شگرف یافته است. پژوهش حاضر در پی پاسخ به این پرسش برآمد که آیا کاربردهای سوءاستفاده‌گرانه فناوری دیپفیک با آموزه‌های بنیادین حقوق بشر تعارض دارند و در صورت پاسخ مثبت، چارچوب‌های حقوقی موجود تا چه میزان در پاسخ‌گویی به این تعارض کارآمد هستند. یافته‌ها نشان می‌دهند که دیپفیک در کاربردهای سوءاستفاده‌گرانه‌ای نظیر پورنوگرافی، تخلفات انتخاباتی، جعل هویت و تخریب حیثیت، با هر سه ساحت بررسی شده در تعارضی ساختاری قرار دارد. این تعارض، نه در ذات فناوری، که در قصد عامل، زمینه کاربردها، فقدان رضایت و پیامد زیان‌بار آن ریشه دارد. از این‌رو، پاسخ به پرسش اصلی آن است که گستره حمایت از آموزه‌های بنیادین حقوق بشر در برابر کنشگری ناشناس مجازی، با چارچوب‌های هنجاری موجود قابل دفاع نیست. نظام حقوقی سستی، به‌ویژه در سطوح بار اثبات، صلاحیت فرامرزی، مسئولیت پلتفرم‌ها و حمایت از گروه‌های آسیب‌پذیر، با کاستی‌های اساسی روبه‌رو است که پاسخ‌دهی به ابعاد نوین این تهدید را با ناکارآمدی جدی مواجه ساخته است.

تحلیل حقوقی نشان می‌دهد که دیپفیک، فارغ از جنبه فنی، به‌سان ابزاری قدرت‌ساز در دستان کنشگران ناشناس، نظم حقوق بشر را در دو سطح هنجاری و اجرایی با بحران مواجه ساخته است. در سطح هنجاری، مفاهیم بنیادینی نظیر «رضایت»، «هویت» و «حقیقت» که پیش‌فرض‌های اساسی نظریه‌پردازی حقوق بشری هستند، در برابر قابلیت‌های بی‌سابقه این فناوری، نیازمند بازتعریف و روزآمدی می‌نمایند. در سطح اجرایی، قواعد سستی ادله اثبات، به دلیل ماهیت فنی و غیرمتمرکز تولید و نشر دیپفیک، عملاً از شناسایی عاملان و احراز سوءنیت ناتوانند. افزون بر این، فقدان قواعد شفاف در خصوص مسئولیت پلتفرم‌ها و نحوه تخصیص خسارت، قربانیان را از دستیابی به جبران مؤثر باز می‌دارد. با لحاظ پیوستگی و به‌هم‌وابستگی حقوق بشر، نقض یک حق در دیپفیک، زنجیره‌ای از نقض‌های دیگر را به دنبال دارد؛ چنان‌که پورنوگرافی، علاوه بر حریم خصوصی، کرامت قربانی را هدف گرفته، با ایجاد ترس و خودسانسوری، آزادی بیان و مشارکت سیاسی او را نیز محدود می‌سازد.

بر اساس یافته‌های این پژوهش، پنج گزاره نهایی برای افق پیش‌روی نظام حقوق بشر قابل طرح است: نخست، تفکیک مفهومی میان محتوای ترکیبی، دیپفیک غیرمخرب و دیپفیک متقلبانه به‌عنوان پیش‌نیاز تنظیم‌گری هوشمندانه که از محدودسازی ناموجه آزادی بیان جلوگیری کرده و جرم‌انگاری کاربردهای سوءاستفاده‌گرانه را ممکن می‌سازد؛ دوم، بازتعریف مفاهیم بنیادین نظیر



رضایت، هویت و حریم خصوصی در پرتو عصر دیجیتال، به گونه‌ای که خودمختاری فرد بر تصویر و هویت دیجیتال خویش در قانون حمایت‌های حقوقی قرار گیرد؛ سوم، گذار از مسئولیت فردی به مسئولیت زنجیره‌ای که با توسعه مسئولیت به پلتفرم‌ها، میزبانان محتوا و توسعه‌دهندگان ابزارهای هوش مصنوعی، و با تکیه بر اصل احتیاط و تعدیل بار اثبات به سود زیان‌دیده، ضمانت‌اجراهای مؤثری در دادرسی الکترونیک فراهم می‌آورد؛ چهارم، ایجاد موازنه میان آزادی بیان و کرامت انسانی با بهره‌گیری از معیارهای پنج‌گانه (اصالت، تحول‌آمیزی، هدف اصلی، تفاوت اثر منفرد از تولید انبوه، و وجود رضایت) که تحدید آزادی بیان را در حد ضرورت مطلق و متناسب با شدت آسیب امکان‌پذیر می‌سازد؛ و پنجم، تدوین چارچوب چندلایه مبتنی بر همکاری فراملی، تنظیم‌گری مشارکتی پلتفرم‌ها و تفسیر پویا از آموزه‌های حقوق بشر که التزام به تعهدات سه‌گانه احترام، حمایت و اجرا را نه تنها برای دولت‌ها، بلکه برای پلتفرم‌های آنلاین، تولیدکنندگان محتوا و کاربران عادی نیز الزامی می‌کند.



فهرست منابع

- استیون، هیک. (۱۹۵۹). حقوق بشر و اینترنت، مترجم: سید قاسم زمانی و مهناز بهراملو، تهران: انتشارات خرسندی، چاپ اول ۱۳۸۶.
- عباسی، بیژن. (۱۳۹۰). حقوق بشر و آزادی‌های بنیادین (سه نسل حقوق بشردر اسلام، ایران و اسناد بین‌المللی و منطقه‌ای)، تهران: انتشارات دادگستر، چاپ اول.
- حبیب‌زاده، توکل و فرج‌پور، علی‌اصغر. (۱۳۹۵). «کرامت انسانی و ابتدای حقوق بشری و شهروندی مندرج در قانون اساسی جمهوری اسلامی ایران بر آن»، فصلنامه مطالعات حقوق عمومی، دوره ۵۶، شماره ۱، ۱۱۵-۱۳۷.
- شریفی طرازکوهی، حسین. (۱۳۹۷). حقوق بشر (نظریه‌ها و رویکردها)، تهران: میزان، چاپ اول.
- فرانک یورگنسن، رایکه. (۱۳۲۸). حقوق بشر در جامعه جهانی اطلاعات، ترجمه: بهرام مستقیمی، تهران: چاپ شریعت، چاپ اول.
- مصطفوی اردبیلی، سید محمد مهدی؛ تقی‌زاده انصاری، مصطفی و رحمتی‌فر، سمانه. (۱۴۰۲). «تأثیر هوش مصنوعی بر نظام حقوق بشر بین‌الملل»، فصلنامه حقوق فناوری‌های نوین، دوره ۴، شماره ۸، ۸۵-۱۰۰.

Ahmed, Ikhlal & Choudhary, Rizvana. (2026). Artificial Intelligence, Human Rights, and the Right to Privacy: The Deepfake Challenges, International Journal of Recent Development in Engineering and Technology Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, 1429-1432.

Barber, Alex. (2023). Freedom of expression meets deepfakes. Springer: Synthese, 202:40 .
<https://doi.org/10.1007/s11229-023-04266-4>

Bosher, Hayleigh. (2026). Do deepfakes, digital replicas and human digital twins justify personality rights?, The Journal of World Intellectual Property, 1–28. DOI: 10.1111/jwip.70020

Bourgault, Jacob R. (2025). Free Speech and Synthetic Lies: Deepfakes, Synthetic Media, and the First Amendment, Student J. Info. Priv. L. 49(3), 49-73. Available at: <https://digitalcommons.maine.edu/sjpl/vol3/iss1/5>

Chesney, Bobby & Citron, Danielle Danielle. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, CALIF. L. REV., Vol. 107.

De Vido, Sara. (2024). Deep fake as AI-generated violence against women, DEP Finestra sul presente, No.19.

Johnson, Valerie Maria. (2025). Deepfake Violence and the Future of Human Rights: A South Korean Case Study, Division of Human Rights Studies, Department of History, Course Code: MRS15, Supervisor: Lena Halldenius.

Kocsis, Eric, (2022). “Deepfakes, Shallowfakes, and the Need for a Private Right of Action”, DICK. L. REV., Vol.126.

Łabuz, M. (2024). Deep fakes and the Artificial Intelligence Act—An important signal or a missed opportunity? Policy & Internet, 5(6), 783–800.

Mirsky, Y., & W. Lee. (2020). “The Creation and Detection of Deepfakes: A Survey.” ACM Computing Services. <https://arxiv.org/pdf/2004.11138.pdf>.

Murray, Michael D. (2025). “Deceptive exploitation: deepfakes, the rights of publicity and privacy, and trademark law”, IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property, Vol. 65, No.2.

Nath, Aranya & Chakravarty, Gautami. (2024). “Navigating the Frontier: Balancing Personality Rights, Privacy, and Intellectual Property In the Age of Deepfake Technology”, E- Journal of Academic Innovation and Research in Intellectual Property Assets (E-JAIRIPA), Vol. V, Issue. I.

Poble, Marta & Kolibe, Jonathan. (2018). RESPONDING TO HUMAN RIGHTS ABUSES IN THE DIGITAL ERA: NEW TOOLS, OLD CHALLENGES, STANFORD JOURNAL OF INTERNATIONAL LAW, Vol.54, No.2, Electronic copy available at: <https://ssrn.com/abstract=3859873>



Prasad, Likhitha & Jain, Ishika D. (2025), Technological Progress vs. Human Rights: Privacy, Free Expression, and Reputation in the Era of A.I and Deepfakes, *Journal of Legal Research and Polity*, Vol. 2, Issue 2, ISSN: 3107-7161 (Online)

Ramluckan, Trishana. (2024). Deepfakes: The Legal Implications, *Proceedings of the 19th International Conference on Cyber Warfare and Security, ICCWS*.

Romero Moreno, Felipe. (2024) Generative AI and deepfakes: a human rights approach to tackling harmful content, *International Review of Law, Computers & Technology*, Vol. 38, No. 3, 297–326, DOI: 10.1080/13600869.2024.2324540

Saini, Seema. (2026). Violation of Human Rights vis-a-vis Misuse of Artificial Intelligence through Deepfake: Challenges and Regulatory Framework, *Cadernos de Pós-Graduação em Direito Político e Econômico*, Volume 26 Issue S1, Gurugram University Conference : 09-10 April 2026 Website: <https://ceapress.org>.

Shtefan, Anna, (2024). The Digital Replication Right as the Element of the Right of Publicity in the AI Age, Available at SSRN: <https://ssrn.com/abstract=4788701>

Shukla, Chitra. (2025). “Digital Forensics and Intellectual Property Rights Protection in the Age of Artificial Intelligence Technology and Deep Fakes”, *JYANAVI*, Vol. 1, Issue 3, SPMVV.

Singh, Pawan (2023). Exploding AI-Generated Deepfakes and Misinformation: A Threat to Global Concern in the 21st Century, DOI:10.36227/techrxiv.24715605

Snow, Sarah M. (2025). The Right to (Digital) Identity, *Fordham Intell. Prop. Media & Ent. L.J.*, Vol. 35, No.4, 1010-1091. Available at: <https://ir.lawnet.fordham.edu/iplj/vol35/iss4/3>

Solove, Daniel J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, Vol. 154, No. 3, 477-560. GWU Law School Public Law Research Paper No. 129, Available at SSRN: <https://ssrn.com/abstract=667622>

Thakur, Dushyant Singh, (2025). Deepfakes as Human Rights Violations: Rethinking Dignity and Privacy in India. Available at SSRN: <https://ssrn.com/abstract=5532383>

Thi Nguyena, T., Quoc Viet Hung Nguyenb, Dung Tien Nguyena, Duc Thanh Nguyena, Thien Huynh-Thec, Saeid Nahavandid, Thanh TamNguyene, Quoc-Viet Phamf, and Cuong M. Nguyeng. 2022. “Deeplearning for Deepfakes Creation and Detection: A Survey.” <https://www.sciencedirect.com/science/article/abs/pii/S1077314222001114>.

Van der Sloot, B., and Y. Wagenveld. (2022). “Deepfakes: Regulatory Challenges for the Synthetic Society.” *Computer Law and Security Review*. <https://www.sciencedirect.com/science/article/pii/S0267364922000632>.

Weitzman, Cliff. (2020). Free Deepfake Video Maker: How to Use AI For Fun And Creativity, *SPEECHIFY*, <https://speechify.com/blog/free-deepfake-video-maker/> [<https://perma.cc/F34LCBEK>].