مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

گزارش سریع
پنجاه و دوم

چالش های اخلاقی فناوری جعل عمیق و راهکارهای مواجهه با آن

Ethical Challenges of Deepfakes Technology
and the Approaches of Confronting It

بسم الله الرحمن الرحیم

گزارش سریع

گزارش شماره ۵۲
دی ۱۴۰۱

مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

چالش‌های اخلاقی فناوری جعل عمیق و راهکارهای مواجهه با آن

گزارش سریع که با عنوان Rapid Report شناخته می‌شود، نوعی گزارش کوتاه است که صرفاً برای اطلاع کلی از موضوع یا پدیده‌ای خاص در بازه زمانی محدود تهیه می‌شود. هدف عمده چنین گزارش‌هایی ایجاد تصویری اجمالی برای آشنایی ابتدایی سیاست‌گذاران و برنامه‌ریزان در موضوعات مورد علاقه آنان است.

تهیه شده در پژوهشگاه فضای مجازی
(گروه مطالعات اخلاقی فضای مجازی)

تهیه‌کننده: محمدمهدی قاسم‌پور
(کارشناسی ارشد علوم ارتباطات اجتماعی
دانشگاه علوم تحقیقات)
فاطمه صالح‌نیا
(کارشناسی ارشد
هوش مصنوعی دانشگاه مالک اشتر)
ناظر علمی: محمدمهدی نصر هرندی
(مدیر گروه مطالعات اخلاقی فضای مجازی)

نشانی: تهران، میدان آرژانتین، خیابان بیهقی، نبش
خیابان ۱۶غربی، پلاک ۲۰
تلفن: ۰۲۱-۸۶۱۵۱۰۶۱
کد پستی: ۱۵۱۵۶۷۴۳۱۱

فهرست

سخن نخست

سخن نخست

فضای مجازی با شتاب شگرف و رو به تزایدی که در حال بسط و گسترش است تمام ساحات اجتماعی، اقتصادی، سیاسی و فرهنگی زندگی بشر را درنوردیده و هر روز بخش بزرگی از زندگی واقعی را در خود فرو برده و حیات متفاوت و جدیدی به آن می‌دهد. لذا به نظر می‌رسد دو نگاه کلان به فضای مجازی وجود دارد: نگاه اول که بالاخص در ابتدای رشد و تکوین فضای مجازی مسلط شده بود، آن را همچون ابزاری کنار سایر ابزارهای بشری تصویر می‌کرد که تنها طریقیت داشت. اما نگاه دوم، در نتیجه رشد تحولات خیره‌کننده فضای مجازی و سایه گستری آن در حوزه‌ها و شئون بشر در یک دهه اخیر آن را چون سکویی می‌داند که بسیار فراتر از شأن ابزاری حیات انسان‌ها را سامان جدیدی داده و ادعای تمدن نوینی را دارد. رویکردی که از قضا از چشمان بصیر رهبر انقلاب نیز دور نمانده و انتظاری تمدنی از فضای مجازی در ایران را مطالبه داشته‌اند.

در همین راستا گزارش‌های عصرفضای مجازی تلاش می‌کند تا فهم سازمان‌ها و دستگاه‌های مرتبط با حوزه فضای مجازی را ارتقاء بخشیده و آن‌ها را برای مواجهه فعال و خردمندانه با تحولات این عرصه مهیا سازد.

سیدابوالحسن فیروزآبادی
دبیر شورای عالی و رئیس مرکز ملی فضای مجازی

چکیده

چکیده

فناوری دیپ‌فیک که با عنوان «جعل عمیق» نیز شناخته می‌شود، پدیده‌ای نوین در فضای مجازی است که با تنوع و تکثر خود دریچه‌های جدیدی را پیش روی کاربران این فضا باز کرده است. جذابیت‌های موجود در این فناوری و سهولت دسترسی به آن در بستر فضای مجازی و تلفن‌های هوشمند، موجب استفاده روزافزون از آن شده است. شاید در نگاه اول، جعل عمیق تنها یک فناوری برای استفاده در ایام فراغت کاربران به نظر برسد و به عنوان یک خطر اخلاقی جدی تلقی نشود، اما با استفاده‌های غیراخلاقی که از آن صورت گرفته، جنبه سرگرمی این فناوری به حاشیه رانده شده است. لذا با توجه به آسیب‌های اخلاقی این فناوری در جامعه، ضرورت نحوه مواجهه با آن باید بررسی و مشخص شود. در این گزارش ضمن تشریح فناوری جعل عمیق و آسیب‌های اخلاقی آن، بر ضرورت شناسایی این فناوری و رویارویی با ضدارزش‌های موجود در آن تأکید شده و نقش سازمان‌های رسانه‌ای، حاکمیتی و کاربران در این مواجهه مورد واکاوی قرار گرفته است.

واژگان کلیدی

جعل عمیق، شناسایی جعل عمیق، اخلاق فضای مجازی، چالش‌های اخلاقی، آسیب‌های اخلاقی، امنیت فضای مجازی، سواد رسانه‌ای

مقدمه

مقدمه

استفاده از ظرفیت‌های فضای مجازی هر روز ابعاد گسترده‌تری پیدا می‌کند و شکل‌های نوینی از فناوری در این فضا به وجود می‌آید. یکی از فناوری‌هایی که چندی است با فراگیری روزافزونی همراه شده، فناوری دیپ‌فیک[1] یا جعل عمیق است. در مورد این فناوری نیز به مانند هر فناوری دیگری در بستر آنلاین، می‌توان با ساده‌انگاری تنها جنبه سرگرم‌کننده و ابعاد مثبت آن را مدنظر داشت، اما با توجه به تبعات منفی آن و اثرات مخربی که بر تمام آحاد جامعه در ابعاد گوناگون دارد، لازم است تا توجه مضاعفی صرف شناسایی و مواجهه فعال و خردمندانه برای مدیریت این فناوری نوین با تمرکز بر تبعات اخلاقی آن شود.

1. Deep Fake

بخش اول

بخش اول

جعل عمیق نوپدیده‌ای رو در روی اخلاق

جعل‌های عمیق تصاویر ثابت یا متحرک و همچنین اصواتی از افراد هستند که برای به تصویر کشیدن و جایگزین کردن افراد دیگر ایجاد می‌شود و از طریق این فناوری است که می‌توان تصاویر و ویدئوها یا صوت‌های جعلی را واقعی جلوه داد (سمپل[1]، ۲۰۲۰). در واقع جعل عمیق تداوم روند شبیه سازی[2] رسانه‌ای است که منجر به دور شدن کاربران از واقعیت‌ها می‌شود (نیک ملکی، ۱۳۹۹). از منظر مفهومی، جعل عمیق، ترکیبی از یادگیری عمیق[3] و جعل[4] است. اما برای نخستین بار در اواخر سال ۲۰۱۷، این اصطلاح از نام مستعار یک کاربر رسانه اجتماعی ردیت[5] با نام deepfakes گرفته شده است که به کمک یک فناوری ماشینی متن‌باز[6]، چهره افراد شناخته شده را جایگزین تصاویر شخصیت‌های موجود در فیلم‌های پورن کرد و در آن رسانه به اشتراک گذاشت. پس از آن بود که یکی دیگر از کاربران ردیت با نام کاربری deepfakeapp برنامه‌ای با نام Fake App را تولید کرد تا افراد با دانش کمتر در این زمینه نیز بتوانند جعل‌های عمیق موردنظر خود را ایجاد کنند (کولانر و کویین[7]، ۲۰۲۰).

1. Sample
2. Simulation
3. Deep learning
4. Fake
5. Reddit
6. Open Source Machine Technology
7. Colaner & Quinn

با مرور محتوای منتشر شده در فناوری جعل عمیق مشخص می شود که محتوای غیراخلاقی به‌خصوص جنسی، بیشترین میزان از محتوای جعل‌های عمیق را (حدود ۹۶ درصد) با بیش از ۱۳۴ میلیون بازدید در سایت‌های مرتبط در بر دارد که با ایجاد آسیب‌های اخلاقی، عاطفی و روانی، موجب به هم خوردن تعادل جنسی و آسیب به شهرت و کسب وکار افراد می‌شود. خطر اصلی آن هنگام است که با گسترش بیش از پیش آن، عادی انگاری همراه با آسیب‌های متعاقب روی دهد (لوماس[1]، ۲۰۲۰). به عنوان نمونه می‌توان در یک کارزار انتخاباتی با انتشار ویدئوهای منتسب به یک کاندیدای انتخاباتی، نشر اکاذیب کرد یا در بازار سرمایه، محتوای جعلی تولید کرد تا ضررهای مالی قابل‌توجهی متوجه بازار شود. همچنین سوءاستفاده شخصی، اجتماعی یا سیاسی از صدا و تصویر افراد فوت شده برای رسیدن به منافع مادی، تبلیغات و... نیز از دیگر انواع محتوایی است که انجام آن از طریق جعل عمیق هموارتر می‌شود (لوماس، ۲۰۲۰). نوع دیگری از جعل‌های عمیق نیز در مراکز تماس خودکار و با هدف تأمین منافع مالی، تجاری و... مورد کاربرد است که البته در عین حال می‌تواند سوءاستفاده‌های تلفنی با جعل هویت صوتی یا تصویری و کلاهبرداری جاسوسی را نیز به دنبال داشته باشد. این در حالی است که گذشت زمان و تسریع پیشرفت این فناوری، موجب تسهیل روند تولید محتوای جعلی شده و بستر شبکه‌های اجتماعی و الگوریتم‌های موجود در فناوری هوش مصنوعی، تولید جعل عمیق را بیش از پیش هموار کرده است (جیمان[2]، ۲۰۲۰). در واقع ایجاد و گسترش این فناوری با موضوعات غیراخلاقی عجین شده و تلاش‌های متعددی برای سهل‌الوصول شدن استفاده

1. Natasha Lomas
2. Jaiman, Ashish

از ایـن فنـاوری صـورت گرفتـه کـه پیامدهـای جبران‌ناپذیـر غیراخلاقـی را بـه همـراه داشـته اسـت. از ایـن روسـت کـه هم‌زمـان بـا ایجـاد جعـل عمیـق، ملاحظـات اخلاقـی پیـش روی آن و همچنیـن یافتـن نسـبت ایـن فنـاوری بـا اصـول اخلاقـی و ارزش‌هـای انسـانی، از موضوعـات مـورد توجـه در ایـن زمینـه بـوده اسـت تـا جایـی کـه پلتفرم‌هـا و شـرکت‌های رسـانه‌ای اجتماعـی بـزرگ نیـز در ایـن زمینـه اقداماتـی را در دسـتور کار قـرار داده‌انـد. البتـه نمی‌تـوان اثـرات مثبـت و کاربردهـای مفیـد ایـن فنـاوری را نیـز نادیـده گرفـت. سـینما، بـازار و تبلیغـات از بزرگ‌تریـن کاربـران ایـن فنـاوری هسـتند. یکـی از مفیدتریـن ویدئوهـای مثبتـی کـه بـه کمـک جعـل عمیـق تولیـد شـد، ویدئویـی از یـک فوتبالیسـت مشـهور خارجـی بـود کـه بـه ۹ زبـان درباره خطـرات ابتـلا بـه بیمـاری مالاریـا توضیح مـی‌داد. همچنیـن یکـی از مهم‌ترین کاربردهـای فنـاوری جعـل عمیـق در دوبلـه اسـت کـه می‌تـوان فیلم‌هـا را بـا صـدا و تصویـر بازیگـر دوبلـه کـرد و ایـن امـر منجـر بـه خداحافظـی بـا دنیـای زیرنویـس می‌شـود. یـا بـه عنـوان نمونـه دیگـر بـرای تسـکین بازمانـدگان افراد فـوت شـده، اسـتفاده از صـوت و تصویـر مصنوعـی بـا کمـک ایـن فنـاوری و ابزارهایـی ماننـد الکسـا[1]، کورتانـا [2] و سـیری[3] مفیـد فایـده خواهـد بـود. بـا وجـود اثـرات مثبـت ایـن فنـاوری امـا نتایـج منفـی آن بیشـتر قابـل توجـه اسـت، لـذا احتمـال سوءاسـتفاده و ابعـاد غیراخلاقـی آن را بایـد بـه طـور عـام در نظـر داشـت. چـرا کـه جعـل عمیـق بـا القـای فریـب و ایجـاد تخیـلات غیرقابل‌تفکیـک از واقعیـت، ارعـاب، تحقیـر، سـیاه‌نمایی و اختـلال در رونـد دموکراتیـک یـک جامعـه، کلاهبـرداری و ... را بـه دنبـال خواهـد داشـت. البتـه خطـر اصلـی زمانـی اسـت کـه بـه واسـطه واقعـی پنداشـتن نمونه‌هـای منتشـر شـده، تشـخیص جعلـی بـودن آن دشـوار باشـد و ایـن تـازه شـروع ماجراسـت (جیمـان[4]، ۲۰۲۰). در واقـع هرچنـد برخـی از مـوارد بـا کمـی

1. Alexa
2. Cortana
3. Siri
4. Jaiman, Ashish

دقت قابل تشخیص هستند اما تلاش پروژه‌هایی مانند Google's Duplex برای واقعی جلوه‌دادن هر چه تمام‌تر ادامه دارد و این مسئله نگرانی‌ها را در مورد سوءاستفاده‌های احتمالی افزایش می‌دهد.

بخش دوم

بخش دوم

فناوری جعل عمیق و خنثی بودن ارزش‌های اخلاقی

از گذشته این بحث مطرح بوده است که نمی‌توان فناوری را ذاتاً خوب یا بد دانست و سازنده Fakeapp (یکی از ابزارهای ایجاد جعل عمیق) نیز بر همین مسئله تأکید دارد. وی با وجود تصریح بر نادرستی محکوم‌کردن فناوری، اذعان کرده است که فناوری می‌تواند برای بسیاری از اهداف خوب استفاده شود، اگرچه مصارف بد هم دارد. در همین زمینه آدام آلتر[1] جامعه‌شناس معتقد است فناوری مادامی که از طرف شرکتی برای مصرف انبوه مورد استفاده قرار نگیرد، از نظر اخلاقی در دسته خوب یا بد نمی‌گنجد چرا که با انبوه مراجعات به فناوری است که کارکردهای خوب یا بد آن مشخص می‌شود. البته باید در نظر داشت که اگر بتوان فناوری را پیش از استفاده انبوه از آن، به طور کلی خنثی در نظر گرفت، باز هم نمی‌توان این ویژگی را به تمام ابعاد و دسته‌های فناوری تعمیم داد. زیرا از نظر اخلاقی، پذیرفتن چنین ادعایی از آن منظر که دست طراحان فناوری را برای بی‌توجهی به اصول اخلاقی و عدم پذیرش تبعات اخلاقی فناوری تولید شده بازمی‌گذارد، دشوار به نظر می‌رسد. از همین روست که توسعه دهندگان محصولات فناورانه مرتبط با هوش مصنوعی،

1. Adam Alter

کار خود را با دغدغه بیشتری نسبت به سوءاستفاده‌های احتمالی و عواقب ناخواسته فناوری‌ها پیش می‌برند (کولانر و کویین[1]، ۲۰۲۰). سؤال اینجاست که آیا می‌توان در مورد فناوری جعل عمیق از خنثی بودن ارزشی سخن گفت؟ یکی از شروط خنثی در نظر داشتن فناوری، قابل مقایسه و سنجش بودن منافع و آسیب‌های آن است. امکان استفاده درست یا نادرست از فناوری از یک طرف و ناچیز بودن ویژگی بالقوه جعل‌های عمیق و جنبه سرگرمی آن از طرف دیگر، لزوم توجه بیشتر به این فناوری را متذکر می شود. لذا تمرکز بیشتر بر این فناوری و پاسخگویی مسئولانه سازندگان و توسعه دهندگان مرتبط با آن مانند FakeApp، Faceswap، FaceApp، MyHeritage، DeepFaceLab، Reface و... نسبت به تبعات و آسیب‌های اجتماعی و ساختاری استفاده از آن در نهاد افراد و جوامع ضروری است (قریشی، ۱۴۰۰). ضمن این‌که با توجه به توسعه روزافزون این فناوری و غیرممکن بودن توقف کامل آن، گام منطقی برای مواجهه با جعل‌های عمیق، تنظیم مسئولیت و نظارت دقیق در قبال آن و مهم‌تر از همه کنارگذاشتن نظریه بی‌طرفی و خنثی بودن ارزشی و پیش‌بینی و تحلیل دقیق شرایطی است که باعث می‌شود این فناوری در خدمت منافع غیراخلاقی باشد (کولانر و کویین، ۲۰۲۰).

1. Colaner & Quinn

بخش سوم

بخش سوم

چالش‌های اخلاقی فناوری جعل عمیق

همانگونه که بیان گردید با وجود ذات سرگرم‌کننده بودن فناوری جعل عمیق اما سوءاستفاده‌های متعددی در ابعاد مختلف به کمک آن انجام می‌شود که چالش‌های اخلاقی بزرگی را به دنبال خواهد داشت. به‌خصوص این مخاطرات وقتی بیشتر مورد توجه قرار می‌گیرد که با زندگی افراد در جامعه و زیست آن‌ها مرتبط باشد و آن را تحت تأثیر خود قرار دهد. در واقع خطر از آن جهت است که غیر واقعیتی به عنوان حقیقت پذیرفته شده و عواقب مختلفی در ابعاد گوناگون، به دنبال داشته باشد (نیک ملکی، ۱۳۹۳). از طرفی سهل‌الوصول بودن دسترسی به نرم‌افزارها و شبکه‌های اجتماعی و... نیز شدت این خطرات را مضاعف می‌کند. بنا بر تحقیقات مرکز «Sensity.ai» که در قالب مستندی به سفارش کانال ۴ با موضوع «**Deepfakes**» ارائه شده، در سال ۲۰۱۹، بیش از ۶۰ هزار فیلم جعل عمیق در اینترنت شناسایی شده است (سپهوند، ۱۳۹۹) و این مسئله زنگ خطری جدی را برای شناخت هر چه سریع‌تر و مواجهه هوشمند با این فناوری به صدا در آورده است. با ظهور این فناوری، بروز چالش‌های اخلاقی و قانونی سرعت بیشتری گرفته است چرا که در این فناوری به چیزی که اساساً وجود

واقعی ندارد، اعتماد شده و حس مشترک واقعیت در هر موضوعی زیر سؤال می‌رود. در واقع همین اعتماد ساختگی است که از بین برنده ساختارهای موجود در هر جامعه‌ای خواهد بود؛ ازاین رو، رویارویی با این فناوری اهمیت دوچندانی پیدا می‌کند (کولانر و کویین[1]، ۲۰۲۰). البته باید در نظر داشت تفسیرهای متفاوت و متناقض افراد مختلف جامعه از موضوعات و واقعیات جامعه، مطلب جدیدی نیست و هرگز نمی‌توان پایانی برای آن متصور بود. بر همین اساس حساب‌های کاربری جعلی نیز از قبل بوده و تا همیشه نیز خواهد بود، اما خبرهای جعلی مسئله‌ای فراتر از تفسیرهای متناقض است که به دنبال خود روایت‌های جعلی و متناقض را به همراه خواهد داشت که البته اثرات ناامیدکننده نه چندان جدید آن در جامعه برای همیشه باقی خواهد ماند. البته جعل‌های عمیق حوزه این تأثیر و تأثرات را گسترده‌تر کرده‌اند. حساسیت بر جعل عمیق شاید تنها به دلیل انتشار روایت دروغ نباشد، چراکه می‌تواند روایتگر یک خیانت بصری نیز باشد و حواس ذاتی انسان را تحت تأثیر قرار دهد. زیرا مغز انسان با وجود تمام ظرافت های تکامل یافته‌ای که دارد، اما شاید نتواند راه‌حل اقناع‌کننده‌ای برای شناسایی و انکار محتوای تولید شده توسط هوش مصنوعی داشته باشد و فریب آن را بخورد. بر همین اساس است که نیاز به ایجاد هنجارهای اخلاقی- فرهنگی، استانداردها و مقررات جدید بیش از پیش احساس می‌شود که البته مسیر بسیار پیچیده‌ای است (کولانر و کویین، ۲۰۲۰). با وجود اقدامات پژوهشگران برای یافتن ابزاری جهت شناسایی جعل‌های عمیق، اما باید در نظر داشت که شاید از طریق شناسایی این موارد بتوان تنها آسیب‌های اجتماعی ناشی از این مسئله را کمتر کرد و

1. Colaner & Quinn

نمی‌توان به متوقف کردن دائمی آن امید چندانی داشت. آسیب‌های فردی جعل‌های عمیق می‌تواند تنها محدود به نقش ایجادکنندگان چنین محتوایی باشد و در قیاس با آسیب‌های اجتماعی ناشی از تجربه استفاده از جعل‌های عمیق ناچیز به نظر برسد. یکی از نمونه‌های آشکار استفاده از جعل‌های عمیق، نقش بالقوه آنان در تولید اخبار جعلی است که آسیب ناشی از تولید چنین اخباری ایجاد تشکیک در خبرهای رسمی و تأیید شده است که می‌تواند در مناسبات و رفتارهای سیاسی- اجتماعی هر جامعه‌ای اثرگذار باشد. در این صورت است که فهمیده می‌شود تنها شناسایی جعل‌های عمیق کمکی به حل قضیه نمی‌کند و تأثیرات روانی که ایجاد می‌کند، موضوع مهمی به شمار می‌رود. به عنوان نمونه می‌توان به انتشار تصویر ساختگی دست دادن و لبخند زدن باراک اوباما و حسن روحانی در توییت پاول گوسار، نماینده جمهوری‌خواه آمریکایی اشاره کرد که در راستای انتقاد از سیاست‌های باراک اوباما در مورد ایران منتشر شده بود و در توضیح این تصویر نوشته بود: جهان بدون بودن این افراد در قدرت، جای بهتری برای زندگی خواهد بود. اثرات نشر این تصویر را شاید بتوان به عنوان توجیهی برای ترور ژنرال قاسم سلیمانی در نظر گرفت؛ هرچند وقتی از گوسار در ارتباط با نشر این تصویر توضیح خواسته شد وی با اعتراض به ارائه توضیح در این زمینه مدعی شد که هرگز در مورد واقعی بودن تصویر منتشر شده ادعایی نکرده است. با پژوهش‌های انجام شده در این زمینه مشخص است که نشر چنین تصاویری تأثیری به مراتب بیشتر از سایر روش‌های مبارزاتی در نبردهای سیاسی و کارزارهایی مانند انتخابات دارد. هرچند این اقدام ممکن است در تغییر نظر مشارکت‌کنندگان

چندان مؤثر نبوده باشد اما در تقویت جهت‌گیری‌های سیاسی رأی‌دهندگان و ایجاد انگیزه در آنان برای رأی دادن نقش بسزایی داشته است. شاهد مثال نیز بررسی نتایج انتخابات ۲۰۱۶ آمریکاست که بر اساس آن مشخص شد کاربران متمایل به اشتراک‌گذاری و نشر محتوایی هستند که همسو با دیدگاه‌های سیاسی آنان باشد حتی اگر آن محتوا نادرست و ساختگی باشد (کولانر و کوییـن[1]، ۲۰۲۰). به عنوان نمونه‌ای دیگر، می‌توان به استفاده از این فناوری در همه‌گیری کرونا اشاره کرد. انتشار اخبار جعلی در رابطه با راه‌های درمان کرونا و یا عوامل مؤثر در تشدید کرونا، برای مثال، ویدیوهای جعلی که ادعا می‌کنند الکل، گرمای بیش از حد یا سرمای بیش از حد می‌تواند باعث نابودی ویروس کرونا شود، موجب ایجاد هرج و مرج، افزایش فشار روحی و روانی در افراد و حساسیت‌های بیش‌ازاندازه می‌شود. در حالی‌که هیچ مدرک علمی برای حمایت از این ادعاها وجود نداشته است. در ۱۴ آوریل ۲۰۲۰، هشتگ TellTheTruthBelgium، توجه رسانه‌ها را به خود جلب کرد. ویدیویی که در آن سوفی ویلمس، نخست وزیر بلژیک، در سخنرانی حدوداً ۵ دقیقه‌ای، همه‌گیری کرونا را در نتیجه تخریب محیط‌زیست اعلام می‌کند. در این ویدیو، جنبش زیست محیطی Extinction Rebellion از فناوری جعل عمیق برای تغییر سخنرانی قبلی سوفی ویلمس و ارتباط کرونا با محیط‌زیست استفاده کرده بود (لانگوث[2] و همکاران، ۲۰۲۰). روش تخریب‌گر جعل عمیق در شبکه‌های اجتماعی با رویکرد فریب دادن ایجاد شده است که این حربه در زمینه‌های متعددی مانند کلاهبرداری، دست‌کاری و جهت دادن به نظر مردم و... انجام می‌شود و تبعات بزرگی در زندگی روزمره مردم، سوگیری افکار

1. Colaner & Quinn
2. Langguth

آن‌ها، نتایج انتخابات و... به دنبال خواهد داشت (نیک ملکی، ۱۳۹۹). از دیگر مواردی که با استفاده از فناوری جعل عمیق منجر به ایجاد چالش‌های اخلاقی و قانونی می‌شود، استفاده از تصاویر یا ایجاد ویدیو از افرادی است که در قید حیات نیستند، بنابراین از لحاظ حقوقی و اخلاقی امکان دفاع از خود یا اثبات جعلی بودن ویدیو را ندارند. به عنوان نمونه‌ای از این موارد، می‌توان به ویدیویی اشاره کرد که در ۲۹ اکتبر ۲۰۲۰، در مکزیک خطاب به رییس‌جمهور این کشور منتشر شد. در این ویدیو، خاویر والدز، نویسنده و روزنامه‌نگار مکزیکی رییس‌جمهور و دولتش را به مبارزه با فساد و جرایم سازمان‌یافته تشویق می‌کرد، این در حالی‌است که خاویر والدز در ۱۵ می ۲۰۱۷ به قتل رسیده است، این چیزی است که در ابتدای ویدیو، آلتر ایگوی دیجیتال[1] این فرد یا به اصطلاح خود دیگر ساخته شده برای این فرد در دنیای دیجیتال، به آن اذعان دارد (لانگوث[2] و همکاران، ۲۰۲۰). علاوه بر جنبه‌های اخلاقی و حقوقی، جنبه‌های روانی و اثر مخرب این فناوری را می‌توان در صنعت پورنوگرافی نیز مشاهده کرد. استارت‌آپ هلندی Sensity.ai که در حال حاضر ویدیوهای ایجاد شده با فناوری جعل عمیق را در فضای مجازی ردیابی و شمارش می‌کند، طی گزارشی اعلام کرده است که بیش از ۸۵ درصد از تمام ویدیوهای جعل عمیق، افراد مشهور زن در صنعت ورزش، سرگرمی و مد را هدف قرار می‌دهد. در این به اصطلاح پورنوگرافی غیرارادی[3]، پیش از این به مقادیر زیادی از تصاویر یک شخص، برای ایجاد ویدیو جعلی، نیاز بوده است، به همین دلیل تنها افراد مشهور هدف قرار می‌گرفتند. اما اخیراً

۱. آلتر ایگوی دیجیتال، خود دیگر یا همزاد دیجیتال، Digital Alter ego تصویری از خود فرد در دنیای دیجیتال است که با ظهور دنیای متاورس، بیش از پیش پررنگ شده است.

2. Langguth

3 .Involuntary Pornography

فناوری جعل عمیق مبتنی بر شبکه‌های مولد متخاصم[1]، امکان هدف قرار دادن افرادی که تنها تصاویر کمی از آن‌ها وجود دارد را نیز ممکن کرده است (لانگوث[2] و همکاران، ۲۰۲۰). از دیگر مواردی که با گسترش جعل عمیق باعث ایجاد چالش‌هایی در حوزه‌های قضایی شده است، استفاده از جعل عمیق در جعل اسناد است. رسانه‌های مصنوعی و تصاویر صورت دستکاری شده دیجیتالی، رویکرد جدیدی را برای تقلب در اسناد ارائه می‌دهند. با استفاده از این روش‌ها و ابزارها، می‌توان چهره شخصی که گذرنامه به او تعلق دارد را با افرادی که قصد دریافت گذرنامه به صورت غیرقانونی دارند، تغییر داده یا ترکیب کرد. این روش ممکن است این شانس را برای متقلبان ایجاد کند تا بدون مشکل از بررسی‌های هویتی(نظیر سیستم‌های تشخیص چهره)، عبور کنند (یوروپل[3]، ۲۰۲۰).

مطالعه انجام شده توسط پانل آینده علم و فناوری[4] در سال ۲۰۲۱، نشان داد که «خطرات مرتبط با جعلیات عمیق می‌تواند ماهیت اخلاقی، روانی و مالی داشته باشد و تأثیرات آن‌ها می‌تواند از سطح فردی تا سطح اجتماعی متغیر باشد». این مطالعه توصیه می‌کند که سیاست‌گذاران از تأثیرات نامطلوب فناوری جلوگیری کرده و به آن رسیدگی کنند و گزینه‌های سیاستی را در چارچوب‌های قانونی خود بگنجانند (کابلکا[5]، ۲۰۲۲). ازاین رو با توجه به تأثیرات اخلاقی جعل عمیق بر ابعاد مختلف فرهنگی، اجتماعی، حقوقی، اقتصادی و سیاسی افراد جامعه، لازم است هشیاری لازم از طرف مسئولان ذی‌ربط، سیاست‌گذاران، اصحاب رسانه و حتی از طرف کاربران برای رویارویی و استفاده از روش‌های نوین تشخیص تقلب رسانه‌ای مورد توجه باشد.

۱. شبکه های مولد متخاصم یا تخاصمی، Generative Adversarial Networks ،که به اختصار GAN نامیده می‌شوند، قادر به بسط دادن و یا تولید محتوای جدید براساس محتوای اندک ورودی هستند.

2. Langguth
3. Europol
4. The Panel for the Future of Science and Technology
5. Kabelka

بخش چهارم

بخش چهارم

راهکارهای مواجهه با چالش‌های اخلاقی جعل عمیق

۴-۱- روش‌های تشخیص و شناسایی جعل‌های عمیق

تا پیش از انتشار جعل عمیق جنجالی اینستاگرام، اطلاع زیادی از این فناوری و تبعاتش در دسترس نبود اما با انتشار ویدئوی مربوط به توضیحات زاکربرگ که در آن از سرقت تمام اطلاعات و اسرار زندگی افراد و نحوه کنترل آن توسط یک فرد گفته بود، تلنگری جدی در زمینه اهمیت کشف و شناسایی جعل‌های عمیق به کاربران زده شد و کیفیت این شناسایی مورد توجه قرار گرفت (لویدجونز[1]، ۲۰۱۹). در واقع با توجه به اثرگذاری‌های گسترده جعل عمیق بر ابعاد مختلف زندگی افراد و تبعاتی که در سراسر جهان به واسطه فراگیر شدن آن به مدد فضای مجازی و شبکه‌های اجتماعی به دنبال داشته، شناسایی و مواجهه با این فناوری در سطوح مختلف از اهمیت ویژه‌ای برخوردار است. از این رو ابزارها و فناوری‌های نوینی برای شناسایی و تشخیص فناوری جعل عمیق ایجاد شده که امیدها را برای ایجاد امنیت بیشتر به دنبال داشته است. در واقع با ایجاد ابزارهای تشخیص این ویدئوها، با تجزیه و تحلیل سبک رفتاری و گفتاری افراد با کمک یادگیری ماشین[2] و امضای

1. Lloyd-Jones
2. Machine Learning

بیومتریکی نرم[1]، امنیت و مصونیت نسبی برای آسیب‌دیدگان ناشی از جعل عمیق -از مردم عادی گرفته تا سلبریتی‌ها و رهبران جهان- ایجاد می‌شود (نایت[2]، ۲۰۱۹).

اگرچه تشخیص نمونه های اولیه جعل عمیق کمی آسان‌تر بود اما با پیشرفت این فناوری، مسیر شناسایی و تشخیص سخت‌تر شده است. اما همچنان روش‌هایی برای شناسایی وجود دارد که از آن جمله می‌توان به موارد زیر اشاره کرد (لویدجونز[3]، ۲۰۱۹):

۱. شناسایی منابع معتبر نشر محتوا

یکی از اصول مهم و کلی زیست در فضای حقیقی و مجازی و مواجهه با هر محتوایی در آن اعم از این که جعل عمیق باشد یا نباشد دریافت اطلاعات کامل از منبع منتشرکننده آن محتواست. در این زمینه نباید به هر محتوای منتشر شده از طرف هر کاربر اعتماد کرد و لازم است از اعتبار رسانه و منتشرکننده آن اطمینان حاصل کرد که این مهم با ارتقاء سواد رسانه‌ای کاربران محقق می‌شود.

۲. تشخیص فرم صورت

گاهی با دقت بیشتر در تصویر یا ویدئوی منتشر شده می‌توان متوجه غیرعادی بودن حالت یا حرکات صورت و مصنوعی بودن آن شد. صورت در جعل‌های عمیق معمولاً کم احساس است و بین بخش‌های صورت و سایر اعضای بدن، همخوانی و یکنواختی وجود ندارد. همچنین در این زمینه می‌توان چشمک زدن افراد را مورد توجه قرار داد. در سال ۲۰۱۸ محققان آمریکایی کشف کردند که چهره‌های تقلبی جعل عمیق، در نمونه‌های ابتدایی، چشمک

۱. امضا یا اثر بیومتریکی نرم یا Soft Biometric Signature به ویژگی های شخصی برای هر فرد نظیر رنگ چشم، جای زخم، لهجه، فرم صدا، قد، وزن و ... و ویژگی های هویتی وی نظیر اثرانگشت، چهره، عنبیه و ... است

2. Will Knight
3. Lloyd-Jones

نمی‌زنند و جای تعجب نیست که در اکثر تصاویر افراد با چشمان باز نشان داده می‌شوند (سمپل[1]، ۲۰۲۰). در مطالعه‌ای مشابه نیز مشخص شد که در ویدیوهایی که نسبت به اصلاح پلک زدن در چهره‌های تقلبی اقدام نموده‌اند، باز هم با بررسی نرخ پلک زدن می‌توان جعلی بودن ویدیو را تشخیص داد. به این صورت که نرخ پلک زدن در ویدیوهای ایجاد شده از طریق جعل عمیق، معمولاً کمتر از حالت عادی است (نگوین[2] و همکاران، ۲۰۱۹). هرچند این روش دارای ایراداتی است، به خصوص زمانی که دسترسی واضح و با کیفیت به چشم‌های سوژه‌های جعل عمیق فراهم نباشد.

۳. حالت و فرم بدن

با دقت در نحوه حرکت بدن می‌توان عدم تطابق حرکات بین اعضای بدن با صورت را شناسایی کرد. در این حالت، پیچ‌وتاب‌ها و ناسازگاری‌هایی در بافت‌های مختلف صورت یا بدن دیده می‌شود، که جعلی بودن ویدیو را قابل تشخیص می‌کنند (نگوین و همکاران، ۲۰۱۹). برخی از مواردی که در بافت و حالت بدن می‌تواند منجر به کشف تقلب شود، تار شدن لبه‌های صورت و بدن نسبت به پس‌زمینه، ناهماهنگی در مو، الگوی رگ، جای زخم و ...، انعکاس نور در چشم و ناهماهنگی در پس‌زمینه، عمق تصویر و ... است (یوروپل[3]، ۲۰۲۲).

۴. تولید صدای ساختگی

با وجود تمرکز بیشتر جعل‌های عمیق بر جلوه‌های تصویری، اما این فناوری با استفاده از ابزارهای تغییر صدا و جای گذاری صدا با صدای دیگر توسط سرویس‌های اختصاصی تولید صدا نیز انجام می‌شود.

1. Sample
2. Nguyen
3. Europol

همچنین با وجود ترکیب جلوه‌های صوتی متعدد از مجموعه قابل توجهی از صداها، گاهی حالت روباتی بودن صدا و عدم تطابق صدا و حرکات لب، به عنوان عاملی برای شناسایی جعل‌های عمیق به کار می‌رود. مدل‌های کلاسیک یادگیری ماشین، به طور گسترده در تشخیص صداهای جعلی مورد استفاده قرار گرفته‌اند. مدل های فعلی تا حدود ۹۷ درصد قادر به تشخیص صدای جعلی از اصلی هستند (المطیری و هبه[1]، ۲۰۲۲).

۵. نورهای مصنوعی

در جعل‌های عمیق عمدتاً از جلوه‌های بصری و روشنایی‌های ناهمگون، مانند نور متناقض استفاده می‌شود که بازتاب آن روی عنبیه ظاهر می‌گردد و با دقت در این بخش نیز می‌توان جعلی بودن تصویر را تشخیص داد (سمپل[2]، ۲۰۲۰). به عنوان نمونه می‌توان به تلاش پژوهشگران دانشگاه‌های برکلی[3] و کالیفرنیای جنوبی[4] اشاره کرد که برای تولید ابزاری جهت استخراج حرکات سر و صورت رهبران جهان، آزمایشی با ویدئوهایی از ترامپ و اوباما گرفته تا هیلاری کلینتون، برنی سندرز و الیزابت وارن با محتواهای جعلی ساخته‌اند و سپس با استفاده از سیستم فراگیری ماشینی و تشخیص حرکات کلی سر و بدن (و نه حالات ظریف و حرکات جزئی افراد) و همچنین با تمرکز بر نحوه صحبت و لحن آن‌ها، به موفقیت ۹۲ تا ۹۶ درصدی در تشخیص جعلی بودن ویدئوها دست یافته‌اند (نایت[5]، ۲۰۱۹). شرکت گوگل[6] و دارپا[7] (شاخه تحقیقاتی پنتاگون) تولیدکننده این فناوری تشخیص بوده و اقدامات کامل‌تری در دست انجام دارند.

1. Almutairi & Hebah
2. Sample
3. Berkeley
4. Southern California
5. Knight
6. Google
7. Defence Advanced Research Project Agency (DARPA)

مشکل اصلی پیش روی این ابزارهای تشخیص، سهولت تولید ویدیوهای جعلی و سرعت فزاینده فراگیری آن‌ها است. در کنار این مورد، عوامل جانبی دیگری نیز وجود دارد که می‌تواند منجر به اشتباه در تشخیص جعلی بودن ویدیو یا تصویر شود. به عنوان مثال می‌توان به کاهش کیفیت ویدیو یا تصویرهای ارسالی اشاره کرد. به این صورت که افراد برای ارسال تصاویر و ویدیوهای خود در فضای مجازی معمولاً از ابزارهایی برای فشرده سازی حجم استفاده می‌کنند که این ابزارها با تغییر در پیکسل‌ها می‌توانند فرآیند تشخیص را با اختلال روبرو کنند (هونگ‌منگ[1] و همکاران، ۲۰۲۰).

۴-۲- مواجهه سازمان‌های فناوری با جعل عمیق

با فراگیر شدن روزافزون فناوری جعل عمیق و تسهیل دسترسی به آن نگرانی از اثرات منفی آن بر جامعه و زیر سؤال رفتن دموکراسی و ... رو به افزایش بوده و نقش نهادهای نظارتی در شناسایی و پیگیری ابعاد قانونی آن بیش از پیش احساس می‌شود. به علاوه با توجه به اهمیت اخلاق فناوری و گستردگی موضوع اخلاق فضای مجازی در حوزه‌های مختلف به ویژه در این فناوری اهمیت روشن شدن روش‌های شناسایی و ارجاعات مستدل به ابعاد اخلاقی آن مورد توجه است. نتایج یک پژوهش که در جریان نظرسنجی از ۱۲۰۰ مدیر ارشد در ۱۲ کشور انجام شده بیانگر آن است که ۸۱ درصد از پاسخگویان، از عدم آمادگی سازمانشان برای توجه به ابعاد اخلاقی و رویارویی با فناوری‌های جدید و تبعات آن گفته‌اند و ۸۲ درصد نیز بر مبنایی بودن و توجه جدی به اخلاق در استفاده از فناوری‌هایی مانند هوش مصنوعی و ... تأکید داشته‌اند (لویدجونز[2]، ۲۰۱۹).

1. Hongmeng
2. Lloyd-Jones

در ایـن صـورت اسـت کـه اهمیـت مواجهـه هوشـمندانه و رویارویـی دقیـق بـا ابعـاد مختلـف فنـاوری و تمرکـز بـر شناسـایی آن بـا در نظـر داشـتن نقـش سـازمان‌های مختلـف بـه ویـژه متولیـان مرتبـط بـا شـبکه‌های اجتماعـی و غول‌هـای فنـاوری روشـن می‌شـود. در ایـن راسـتا دولت‌هـا، دانشـگاه‌ها و مؤسسـات فنـاوری متعـددی در سراسـر جهـان، پژوهش‌هـا و اقدامـات گسـترده خـود را بـرای شناسـایی محتـوای جعلـی انجـام داده‌انـد کـه یکـی از اولیـن نمونه‌هـا در ایـن زمینـه چالـش Deepfake Detection بـوده اسـت کـه بـا پشـتیبانی مایکروسـافت[1]، فیس‌بـوک[2] و آمـازون[3] آغـاز شـد. ایـن پـروژه شـامل تیم‌هـای تحقیقاتـی از سراسـر جهـان بـود کـه بـرای تشـخیص جعل‌هـای عمیـق رقابـت خـود را آغـاز کردنـد (سـمپل[4]، ۲۰۲۰). بـا شـروع ایـن رقابـت، یکـی از اقدامـات رو بـه رشـد در شناسـایی و مواجهـه بـا جعـل عمیـق را شـرکت مایکروسـافت بـا همـکاری چندیـن انجمـن، انجـام داده اسـت. مایکروسـافت، کار رویارویـی بـا اخبـار نادرسـت را بـه کمـک ابـزاری کـه Video Authenticator نامیـده می‌شـود، در دسـتور کار قـرار داده اسـت و از آن بـا تعبیـر «یـک درصـد شـانس یـا اطمینـان[5]» بـرای شناسـایی محتـوای جعلـی یـاد می‌کنـد. در ایـن زمینـه گفتـه شـده اسـت کـه ایـن ابـزار بـا تشـخیص مـرز مخلـوط شـده عناصـر عمیـق کاذب و محـو یـا خاکسـتری کـه ممکـن اسـت توسـط چشـم انسـان قابـل تشـخیص نباشـد، کار می‌کنـد. البتـه ایـن روش می‌توانـد بـرای مصـارف کوتاه‌مـدت ماننـد اقـدام ضربتـی کـه مدتـی پیـش از انتخابـات ۲۰۲۰ ایـالات متحـده آمریـکا انجـام شـد، کاربـرد داشـته باشـد و بـرای بلندمـدت بایـد بـه دنبـال روش‌هـای قوی‌تـری بـرای حفـظ و تأییـد صحـت مقاله‌هـای خبـری و سـایر رسـانه‌ها باشـند.

1. Microsoft
2. Facebook
3. Amazon
4. Sample
5. Confidence Interval

مایکروسافت گفته است که ابزار Video Authenticator آن با استفاده از یک مجموعه داده عمومی FaceForensic++ ایجاد شده و در مجموعه داده‌های چالشی DeepFake Detection آزمایش شده است که هر دو مدل برای آموزش و آزمایش فناوری‌های تشخیص عمیق جعلی پیشرو هستند. این غول فناوری همچنین خاطرنشان کرده است که هم‌زمان با این ابزار به کمک تبلیغ یک خدمت عمومی[1] (PSA) در آمریکا مردم را ترغیب می‌کند تا قبل از به اشتراک گذاشتن یا تبلیغ اطلاعات منتشر شده در رسانه‌های اجتماعی، از سازمان‌های خبری معتبر صحت محتوای تولید شده را به خصوص در ایام پیش از انتخابات استعلام کنند (لوماس[2]، ۲۰۲۰). همچنین، صندوق سرمایه‌گذاری خطرپذیر مایکروسافت[3]، با نام M12، در سال ۲۰۲۱ با هدف مبارزه با موج نوظهور عکس‌ها و ویدیوهای دیجیتالی تغییر یافته، که حاصل فناوری جعل عمیق هستند، روی استارت‌آپ Truepic که در سن‌دیگو استقرار یافته، سرمایه‌گذاری به مبلغ ۲۶ میلیون دلار، انجام داده است. شرکت‌های دیگری نظیر سونی[4] و ادوبی[5] نیز در این استارت‌آپ سرمایه‌گذاری کرده‌اند. Truepic با ارائه فناوری دوربین ایمن برای گوشی‌های همراه، به جای اثبات جعلی بودن یک عکس، نسبت به ثبت واقعیت اقدام کرده و تصاویر واقعی عکس‌برداری شده از طریق دوربین را برچسب‌گذاری می‌کند (مک گرگور[6]، ۲۰۲۱). فیس‌بوک[7] نیز در آستانه انتخابات ۲۰۲۰ آمریکا روندی را برای شناسایی فیلم‌های جعل عمیق که احتمال گمراه کردن کاربران در فرایندهای انتخابات را به دنبال دارد، پیش برد. بر اساس اطلاعیه

1. Public Service Anouncement
2. Lomas
۳. M12 صندوق سرمایه گذاری خطرپذیر مایکروسافت است که روی شرکت های فناوری در مراحل اولیه قرار دارند،سرمایه گذاری می کند.در این هدف،تمرکز روی شرکت‌هایی است که در حوزه‌های هوش مصنوعی کاربردی برنامه‌های کاربردی تجاری،زیرساخت‌ها،امنیت وفناوری‌های پیشتاز فعال هستند
4.Sony
5. Adobe
6. McGregor
7. Facebook

این شرکت فناوری، سیاست در پیش گرفته شده در حذف این محتواها تنها شامل اطلاعات نادرست تولید شده به کمک هوش مصنوعی می‌شود و برخی از محتواهای جعلی که به شدت جعل‌های عمیق نیستند، مجوز انتشار داشته‌اند (سمپل[1]، ۲۰۲۰). توییتر[2] نیز در اقدامی مشترک با فیس‌بوک استفاده از جعل عمیق مخرب را ممنوع و ناقض قوانین پلتفرم خود اعلام کرد و شرکت گوگل[3] هم اقداماتی را در ابزار تبدیل متن به گفتار جهت تأیید و اعتبارسنجی محتواهای صوتی و منفک کردن آن‌ها از محتوای جعلی در دستور کار خود قرار داده است. ادوبی نیز سیستمی ایجاد کرده که به کاربر این امکان را می‌دهد تا به نوعی امضاء و شناسه کاربر به محتوایی که تولید می‌کند، مرتبط شود تا جزئیاتی از شناسنامه آن برای تأیید محتوا ایجاد شود. این شرکت همچنین به طور هم‌زمان در حال توسعه ابزاری برای تعیین دست‌کاری و جعلی بودن یا نبودن محتوا بوده است. همچنین در یک مرکز پژوهشی در دانشگاه بوفالو[4]، ابزاری طراحی شده که به کمک آن تشخیص نقص‌های تصویری محتوا با هدف شناسایی عکس‌های پرتره جعل عمیق با دقت ۹۴ درصد انجام شده است. دانشمندان این مرکز با استفاده از این ابزار و با تجزیه و تحلیل بازتاب نور در چشم، به طور خودکار عکس‌های عمیق جعلی را شناسایی کرده‌اند. بر این اساس گفته شده است که دو چشم الگوهای انعکاسی بسیار مشابهی داشته‌اند که معمولاً در نگاه اول به چهره، قابل مشاهده نیست اما به کمک این ابزار که البته همراه با درصدی از خطا و محدودیت بوده، شناسایی جعل عمیق امکان‌پذیر می‌شود. پژوهشگران مرکز بوفالو همچنین در

1. Sample
2. Twitter
3. Google
4. Buffalo

اقدامی دیگر با کمک به شرکت فیس‌بوک ابزاری تحت عنوان «Deepfake-o-meter» ایجاد کردند که در بستری آنلاین به کاربران برای تشخیص محتوای جعل عمیق و جعلی بودن ویدئوها یا تصاویر کمک می‌کند (مرکز پژوهشی بوفالو[1]، ۲۰۲۱). در پروژه‌ای مشابه، از طریق مجموعه داده ارائه شده در چالش شناسایی جعل عمیق (DFDC)[2] و همکاری با شرکت‌های فناوری، راه‌حل‌هایی برای تشخیص جعل‌های عمیق ایجاد شده است. این مجموعه داده، در بر دارنده ۱۲۴هزار فیلم به اشتراک گذاشته شده و دارای هشت الگوریتم برای اصلاح صورت است. به عنوان نمونه‌ای دیگر، Deeptrace یک شرکت تازه‌وارد مستقر در آمستردام در حال توسعه ابزارهای خودکار تشخیص جعل عمیق برای انجام اسکن پس‌زمینه از رسانه‌های سمعی و بصری است که برای شناسایی، مشابه یک آنتی‌ویروس عمل می‌کند. به علاوه آژانس پروژه‌های تحقیقاتی دفاعی ایالات متحده (دارپا)، بودجه پژوهشی اختصاصی را برای توسعه غربالگری خودکار فناوری جعل عمیق از طریق توسعه برنامه‌ای به نام MediFor یا Media Forensics تأمین کرده است (جانسن[3]، ۲۰۲۰). مک‌آفی نیز به منظور غلبه بر مشکلات ناشی از گسترش جعل عمیق، در اکتبر ۲۰۲۰ نسبت به راه‌اندازی آزمایشگاه جعل عمیق خود اقدام نمود. در این آزمایشگاه از تخصص علم داده برای تشخیص ویدیوهای جعلی استفاده می‌شود (مک‌آفی[4]، ۲۰۲۰).

۴-۳- اقدامات حاکمیتی و سیاست‌گذاری فناوری جعل عمیق

باید در نظر داشت هرچند تلاش‌های گسترده‌ای در امر شناسایی و تشخیص جعل‌های عمیق از طرف شرکت‌های فناوری و مؤسسات

1. Buffalo
2. Deep Fake Detection Challenge
3. Johansen
4. McAfee

پژوهشی مرتبط انجام شده و بخشی از اهداف محقق شده است، اما به هیچ عنوان نتایج این اقدامات اطمینان‌بخش نبوده و لازم است اقدامات حاکمیتی و سیاست‌گذاری دقیقی نیز در این زمینه در دستور کار باشد که البته این مهم نیز تا حدودی انجام شده است. در این زمینه دولت‌ها و حاکمیت‌ها برای صیانت از حقوق کاربران باید در مورد افترا، تقلب و سوءاستفاده‌های در حال انجام در قوانین تجدید نظر داشته باشند. همچنین لازم است با نزدیک‌تر کردن رابطه خود با صنایع و انعطاف‌پذیری بیشتر، برای هموار کردن مسیر تدوین راه‌حل‌های فنی بهتر که به شناسایی جعل‌های عمیق کمک می‌کند اقدام کنند. چرا که شناسایی بهتر جعل‌های عمیق به عوامل رسانه‌های اجتماعی کمک می‌کند تا سیاست‌های یکنواخت‌تری در مورد آن ایجاد نمایند که این مورد در محافظت از حق مردم و صیانت از آنان مؤثر خواهد بود. به علاوه دولت‌ها باید رهنمودهای شفاف خود در مورد چگونگی و زمان استفاده مجموعه دولت از فناوری جعل عمیق را تعیین کنند. همان‌طور که شهروندان حق دارند در برابر اقدامات ضد اطلاعات خارجی محافظت شوند، آن‌ها همچنین حق دارند بدانند که دولت خود چه زمانی و چگونه یادگیری ماشین را به کار می‌گیرد. چرا که سیاست‌گذاری مناسب نیز می‌تواند علاوه بر جامعه معاصر، نسل‌های آینده جوامع را از آثار سوء فناوری‌های نوین نظیر جعل عمیق محافظت کند. بر این اساس با توجه به گسترش این فناوری، سیاست گذاران در سراسر جهان با درک خطرات پیش روی جعل عمیق برای جامعه اقداماتی را انجام داده‌اند. تاکنون اتحادیه اروپا آینده‌نگرترین مجموعه‌ای بوده که اقداماتی را در برابر انواع اشکال نادرست به کار رفته از طریق جعل عمیق،

انجام داده است. به عنوان نمونه در آگوست ۲۰۱۹، آژانس اجرای قانون اتحادیه اروپا، یوروپل[1]، اولین جرائم اینترنتی مرتکب شده با استفاده از هوش مصنوعی را گزارش داد. طبق گزارش منتشر شده، مجرمان با استفاده از برنامه‌ای با جعل صدای یک رئیس اجرایی از یک شرکت انرژی مستقر در انگلیس، حدود ۲۰۰هزار یورو کلاهبرداری کردند (ولچ[2]، ۲۰۱۹). استراتژی اتحادیه اروپا، ایجاد یک شبکه مستقل ضروری حقیقت‌یاب در اروپا برای کمک به تجزیه و تحلیل منابع و فرایندهای تولید محتوا بوده است. در آوریل ۲۰۲۲، یوروپل نسبت به ارائه گزارشی در رابطه با موارد سوءاستفاده از جعل عمیق، اقدام کرده است. در این گزارش اعلام شده که جعل‌های عمیق می‌توانند با دستکاری مصنوعی یا تولید فایل‌های جعلی، بر روند قانون تأثیر منفی بگذارند، به عنوان مثال، در یکی از پرونده‌های اخیر حضانت کودک، مادر یک کودک در تلاش برای متقاعد کردن دادگاه مبنی بر رفتار خشونت‌آمیز شوهرش، نسبت به دستکاری فایل صوتی ضبط شده از صدای وی اقدام کرده است. در این گزارش یوروپل، از سازمان‌های مجری قانون درخواست کرده تا مهارت‌ها و فناوری‌های خود را در رابطه با تشخیص موارد جعل عمیق تقویت کنند. همچنین اعلام می‌کند که سیاست‌گذاران باید قوانین بیشتری را در تنظیم دستورالعمل‌ها و اجبار برای استفاده از آن‌ها، تدوین نمایند (یوروپل، ۲۰۲۲). همچنین در اوایل سال ۲۰۱۹، بروکسل استراتژی مقابله با اطلاعات نادرست را منتشر کرد که شامل رهنمودهایی مربوط به دفاع در برابر جعل عمیق است. این رهنمودها بر نیاز به مشارکت عمومی که به مردم جهت درک و شناسایی منبع محتوا، چگونگی

1. Europol
2. Welch

تولید آن و آگاهی از اعتبار آن کمک می‌کند، تأکید دارند. در پایان سال ۲۰۲۰، اتریش کارگروهی شامل صدراعظم فدرال، وزارت دادگستری، وزارت دفاع و وزارت امور خارجه را برای رسیدگی به چالش‌های جعل عمیق، ایجاد کرد. این کارگروه با تشریح موضوع فناوری جعل عمیق و تلاش برای تدوین برنامه عملیاتی در چهار حوزه ساختارها و فرایندها، حکمرانی، تحقیق و توسعه و همکاری بین‌المللی، کار خود را آغاز کرد. در نهایت، در ۲۵ مه ۲۰۲۲، دولت اتریش برنامه اقدام این کشور برای مبارزه با جعلیات عمیق را منتشر کرد (کابلکا[1]، ۲۰۲۲). در ایالات متحده آمریکا نیز نمایندگان مجلس از هر دو حزب جمهوری‌خواه و دموکرات و همچنین تمامی جریان‌های موجود در کنگره، نگرانی‌های خود را در مورد جعل عمیق ابراز کرده‌اند. پیش از این، نمایندگان آدام شیف و استفانی مورفی و همچنین نماینده سابق کارلوس کربلو با نگارش نامه‌ای از رئیس اطلاعات ملی خواسته‌اند تا این مهم را درک کند که چگونه دولت‌های خارج از این کشور، آژانس‌های اطلاعاتی و افراد عادی می‌توانند از جعل عمیق برای آسیب رساندن به منافع آمریکا بهره ببرند و بر این اساس بر لزوم چگونگی جلوگیری از آن‌ها تأکید کرده‌اند. اولین اقدامات در آمریکا و در سطح فدرال در ایالت‌های کالیفرنیا و تگزاس در سال ۲۰۱۹ صورت گرفته است. طبق قوانین جدید کالیفرنیا (AB ۷۳۰)، توزیع محتوای دستکاری شده نامزدهای سیاسی در دوره‌ای ۶۰ روزه پیش از انتخابات، که با قصد صدمه‌زدن به شهرت نامزد یا فریب رأی‌دهنده برای رأی دادن به یا علیه نامزد انجام شود، غیرقانونی است. قانون تگزاس نیز شبیه به همین قانون است با این تفاوت که طول دوره در این ایالت ۳۰ روزه است. در سایر

1. Kabelka

ایالت‌ها نیز قوانینی در رابطه با برخورد با انتشار پورنوگرافی تصویب شده است. به‌عنوان‌مثال، در ویرجینیا در سال ۲۰۱۹، قانونی تصویب شده که انتشار چنین محتوایی را در صورتی‌که به قصد آزار و اذیت، اخاذی یا تعرض به فرد صورت بگیرد، جرم تلقی می‌کند. از دیگر قوانین کالیفرنیا در این رابطه حق اقدام برای ثبت شکایت خصوصی برای افرادی است که در پورنوگرافی جعل عمیق، دیده می‌شوند تا روال شکایت برای این افراد سریع و آسان شود. در اواخر سال ۲۰۲۰ نیز در نیویورک قانون شکایت خصوصی علیه پورنوگرافی جعل عمیق، تصویب گردید. اما در این قانون مورد دیگری نیز لحاظ شده که بنا به آن، حق انتشار جعل عمیق یا استفاده از آن در مورد افرادی که از دنیا رفته‌اند تا ۴۰ سال پس از مرگ آن‌ها وجود ندارد. این مورد به منظور جلوگیری از استفاده و استثمار تجاری غیرمجاز از این افراد در نظر گرفته شده است (پنل آینده علم و فناوری، ۲۰۲۱).

جعل عمیق در حال حاضر در چین نیز به یک چالش اخلاقی تبدیل شده است. برنامه‌های مختلفی در چین برای تولید جعل عمیق استفاده می‌شوند. بنابراین مانند سایر کشورها، چین نیز در تلاش است تا اثرات منفی بالقوه ناشی از جعل عمیق را مهار کند. به همین منظور دولت چین در ژانویه ۲۰۲۰ قانونی را تصویب کرده است که براساس آن، همه ویدیوها، تصاویر و محتوای صوتی که با استفاده از الگوریتم‌های یادگیری عمیق یا فناوری‌های واقعیت مجازی تولید می‌شوند باید توسط ارائه‌دهندگان برنامه برچسب‌گذاری شوند. همچنین اپراتورها در پلتفرم‌ها موظف هستند تا محتوای بدون برچسب را شناسایی، برچسب‌گذاری یا حذف نمایند. براساس این قانون جدید، تولید و انتشار اخبار جعلی ممنوع بوده و باید بلافاصله پس از شناسایی

حذف شود. مسئولیت رسیدگی به این مسائل به عهده اداره فضای مجازی چین[1] نهاده شده است (پنل آینده علم و فناوری، ۲۰۲۱). به طور کلی اقدام مهم، حاکمیت کشورها در این زمینه این است که با تدوین قوانین و سیاست‌گذاری‌های مدون در امر اشتراک اطلاعات، موارد استفاده نامناسب از جعل‌های عمیق را تعریف کنند. از آنجا که جعل عمیق در زمینه‌های مختلف و برای اهداف مختلف -خوب و بد- مورد استفاده قرار می‌گیرد، برای جامعه بسیار مهم است تا بر اساس این تعاریف، محتوای قابل قبول را به کار بگیرد. انجام این مهم با ابلاغ سیاست‌ها به شرکت‌های رسانه‌های اجتماعی میسر است تا با کمک ابزارهای فنی خود و ظرفیت نظارتی پلیس فضای مجازی با مواجهه هوشمندانه، کنترل قابل قبولی بر محتوا داشته باشند.

۴-۴- ارتقاء سواد رسانه‌ای پادزهری فراتر از هر رویارویی

وضعیت کنونی و چشم‌انداز جهانی در حوزه رسانه و رویارویی با فناوری جعل عمیق، با فقدان استانداردهای منسجم برای تنظیم محتوا و قوانین دقیق حاکمیتی همراه است؛ از این رو لازم است هر کاربر، فارغ از نقص قوانین و سایر خلأهای موجود، ارتقاء سواد رسانه‌ای خود را در اولویت قرار دهد. در این زمینه صرف نظر از این که کاربر در چه سنی باشد، لازم است تا آموزش کافی در مورد چگونگی دریافت و انتشار اطلاعات نادرست با آگاهی‌بخشی عمومی در مورد آسیب‌های اخلاقی ناشی از محتوای جعل عمیق در دستور کار باشد، تا نه تنها به آنان برای شناسایی جعل عمیق کمک کند بلکه باعث شود افراد بتوانند تمامی اطلاعات گمراه‌کننده در فضای مجازی و شبکه‌های اجتماعی را تشخیص دهند. در این زمینه لازم است اصحاب رسانه

1. The Cyberspace Administration of China (CAC)

بـا دقـت در نشـر و توجـه بـه سـرعت و صحـت بـه طـور هم‌زمـان از یـک طـرف و کاربـران بـا دقـت در دریافـت محتـوا از طـرف دیگـر، ایـن اولویـت مهـم را مـورد توجـه قـرار دهنـد. نکتـه حائـز اهمیـت دیگـر در ایـن حـوزه تدویـن سـازوکار و اسـتراتژی دقیـق رسـانه‌ای بـرای ارتقـاء آن، خـروج از فقـر ادبیـات سـواد رسـانه‌ای و تولیـد محتـوای مناسـب اسـت (نیـک ملکـی، ۱۳۹۹). بـر ایـن اسـاس لازم نیسـت مانـع اسـتفاده کاربـران از برنامـه‌ای شـد، بلکـه می‌تـوان دانـش اسـتفاده صحیـح را بـه آن‌هـا در هـر سـطح آموزشـی آموخـت تـا هـر کاربـر بتوانـد رفتـار هوشـمندانه‌ای در قبـال فناوری‌هـای بـه کار رفتـه داشـته باشـد.

جمع بندی

جمع‌بندی

فنـاوری جعـل عمیـق نیـز ماننـد هـر فنـاوری مزایـا و معایبـی دارد کـه البتـه تبعـات منفـی آن بـر ابعـاد مثبتـش غلبـه دارد. بـر ایـن اسـاس بـا توجـه بـه اهمیـت ملاحظـات اخلاقـی فناوری‌هـا بـه خصـوص فنـاوری نویـن جعـل عمیـق و اثـرات منفـی آن بـر وجـوه اخلاقـی زندگـی افـراد در ابعـاد مختلـف، لازم اسـت شناسـایی و رویارویـی دقیـق بـا آن از طـرف حاکمیت‌هـا و سـازمان‌های رسـانه‌ای بـه طـور خـاص و کاربـران بـه طـور عـام در اولویـت قـرار گیـرد. از طـرف دیگـر گسـترش روزافـزون اسـتفاده از فنـاوری جعـل عمیـق و تسـهیل و سـرعت دسترسـی بـه آن از بسـتر شـبکه‌های اجتماعـی، اهمیـت ایـن موضـوع را دو چنـدان کـرده اسـت. بـر ایـن اسـاس جعـل عمیـق، بـا فراوانـی تولیـد محتـوای جعلـی و نشـر بـدون رضایـت، آسـیب‌های اخلاقـی، روانـی و بی‌ثباتی‌هـای متعـددی را در ابعـاد سیاسـی، فرهنگـی، اجتماعـی، اقتصـادی، امنیتـی و ... بـه همـراه خواهـد داشـت؛ لـذا لازم اسـت تدویـن قوانیـن و چارچوب‌هـای اخلاقـی مشـخص و نظـارت بـر آن همـراه بـا فرهنگ‌سـازی و ابـلاغ دسـتورات لازم‌الاجـرا جهـت ارتقـاء سـواد رسـانه‌ای در جوامـع بـا هـدف رویارویـی هوشـمندانه و خردمندانـه بـا ایـن فنـاوری بیـش از پیـش مـورد توجـه قـرار گیـرد.

منابع

[1] Dormehl, Luke. 2020. Researchers have found a new way to spot the latest deepfakes. Available at: www.digitaltrends.com

[2] Colaner, Nathan and Quinn, Michael J. 2020. Deepfakes and the Value-Neutrality Thesis. Available at: www.seattleu.edu

[3] Knight, Will. 2019. A new deepfake detection tool should keep world leaders safe—for now. Available at: www.technologyreview.com

[4] Lloyd-Jones, Chris. 2019. How to spot deepfake videos – and why you should care. Available at: www.avanade.com

[5] Jaiman, Ashish. 2020. Debating the ethics of deepfakes. Available at: www.orfonline.org

[6] Lomas, Natasha. 2020. Microsoft launches a deepfake detector tool ahead of US election. Available at: www.techcrunch.com

[7] Sample, Ian. 2020. What are deepfakes – and how can you spot them? Available at: www.theguardian.com

[8] University at Buffalo. 2021. New Deepfake Spotting Tool Proves 94% Effective – Here's the Secret of Its Success. Available at: www.scitechdaily.com

[9] Johansen, Alison. 2020. How to spot deepfake videos — 15 signs to watch for. Available at: us.norton.com

[10] Welch, Clare. 2019. Artificial Intelligence: What can government do about deepfakes? Available at: www.apolitical.co

[۱۱] نیک ملکی، محمد. ۱۳۹۹. جعل عمیق (Deep Fake)، جعل واقعیت با هوش مصنوعی. موجود در www.iribnews.ir

[۱۲] سپهوند، علیرضا. ۱۳۹۹. چرا «جعل عمیق» نگران‌کننده شده است؟ موجود در www.farsnews.ir

[۱۳] قریشی، محمد. ۱۴۰۰. جعبه ابزار: بهترین اپلیکیشن‌های تغییر چهره و ساخت جعل عمیق. موجود در www.digiato.com

[14] Langguth, Johannes &Pogorelov, Konstantin &Brenner, Stefan &Filkuková, Petra &Schroeder, Daniel. 2021. Don't Trust Your Eyes: Image Manipulation in the Age of DeepFakes. Frontiers in Communication. 6. 632317. 10.3389/fcomm.2021.632317.

[15] Nguyen, Thanh & Nguyen, Cuong M. & Nguyen, Tien & Duc, Thanh & Nahavandi, Saeid. (2019). Deep Learning for Deepfakes Creation and Detection: A Survey.

[16] Almutairi, Zaynab & Hebah Elgibreen. 2022. "A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions." Algorithms Journal, vol 15.

[17] McGregor, Jeffrey. 2021. Announcing our $26M Series B Fundraise. Available at: truepic.com

[18] Europol. 2022. Facing reality? Law enforcement and the challenge of deepfakes. Available at: https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes

[19] McAfee. 2020. The Deepfakes Lab: Detecting & Defending Against Deepfakes with Advanced AI. Available at: https://www.mcafee.com/blogs/enterprise/security-operations/the-deepfakes-lab-detecting-defending-against-deepfakes-with-advanced-ai/

[20] Kabelka, Laura. 2022. Austria to combat deep fakes amid increasing use of the technology. Available At: https://www.euractiv.com

[21] Hongmeng, Zhang & Zhiqiang, Zhu & Lei, Sun & Xiuqing, Mao & Yuehan, Wang. 2020. A Detection Method for DeepFake Hard Compressed Videos based on Super-resolution Reconstruction Using CNN. Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence. P 98-103. doi: 10.1145/3409501.3409542.

[22] STUDY Panel for the Future of Science and Technology. 2021. Tackling deepfakes in European policy. Available At: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf

csri.majazi.ir

حـوزه فضـای مجـازی بـه انـدازه انقـلاب اسـلامی اهمیـت دارد. ایـن فضـا مثـل یـک رودخانـه پـر از آب و خروشـان اسـت کـه مـی آیـد و دائمـاً هـم بـر آب آن افـزوده و خروشـان تـر مـی شـود. اگـر مـا بـر ایـن رودخانـه تدبیـر کنیـم و برنامـه داشـته باشـیم، زهکشـی کنیـم و هدایـت کنیـم ایـن رودخانـه را تـا بـه سـد بریـزد، می‌شـود فرصـت. اگـر رهایـش کنیـم و برنامـه‌ای بـرای آن نداشـته باشـیم می‌شـود یـک تهدید.

csri.majazi.ir