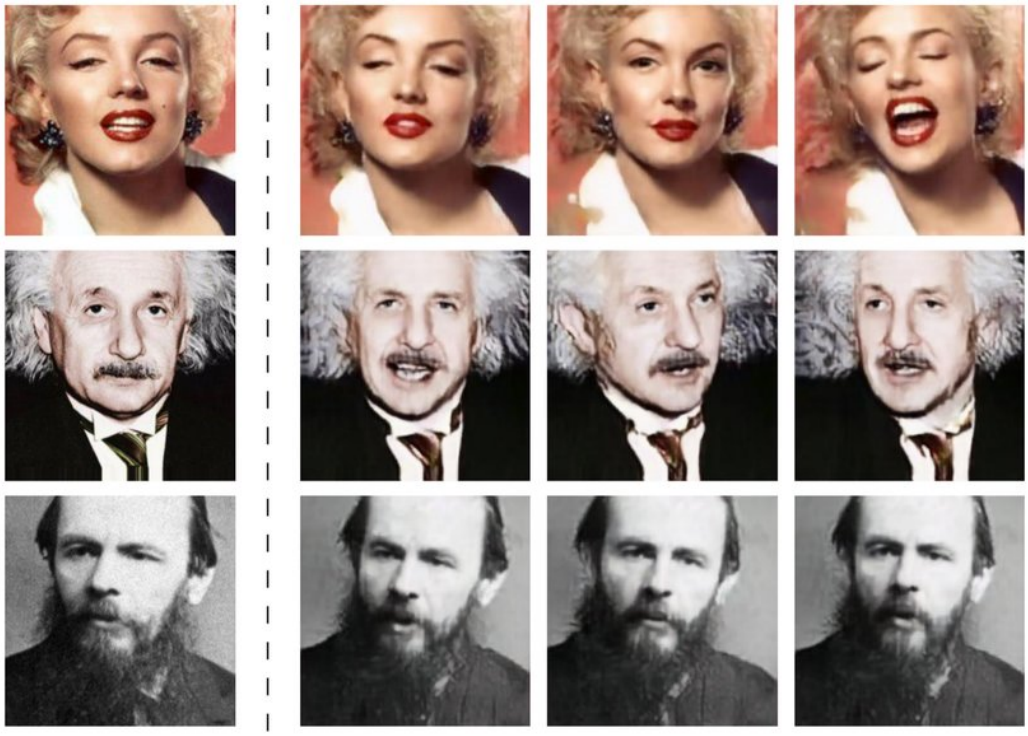
دیپ‌فیک ما را به کجا خواهد برد؟
خوب،بد و زشت دیپ‌فیک‌ها

آقایان محمدرضا محمدی و رضا رضایی

۱۱ دی ۱۳۹۸

فهرست مطالب



مقدمه

در یکی از صحنه‌های فیلم«اگه میتونی منو بگیر» [۱] ساخته استیون اسپیلبرگ، یک مامور FBI[۲] به نام کارل هنرتی در جستجوی کلاهبرداری به نام فرنک ابگنیل جونیور وارد اتاق او در هتل می‌شود. ابگنیل در همان حالتی که لوله تفنگ به سمت صورتش گرفته شده، وانمود می‌کند که مامور مخفی پلیس است و خودش هم دارد دنبال ابگنیل کلاهبردار می‌گردد. سرانجام ابگنیل موفق می‌شود سر هنرتی را شیره بمالد و درست جلوی چشمان حیرت زده او از چنگش دربرود. این فیلم که شخصیت اصلی‌اش مدام بین واقعیت و دروغ در تلاطم است، قصه بچه نابغه‌ای است که مسیر بدی را در پیش گرفته است. این روزها همین قصه را داریم در رسانه‌ها از نو زندگی می‌کنیم: همه‌ما هنرتی‌هایی هستیم که توانایی‌مان برای تشخیص راست از دروغ، حقیقت از کلک، و درست‌کاری از بدبینی به بازی گرفته شده است. کلاهبردار این قصه اگر گفتید کیست؟ بله، دیپ‌فیک‌ها[۳].

[1] catch me,if you can
[2] Federal Bureau of Investigation
[3] DeepFake

فصل ۱

دیپ‌فیک چیست؟

۱.۱ دیپ‌فیک از کجا می آید؟

کلمه دیپ فیک **Deepfake** ترکیبی از deep learning و fake به معنای یادگیری عمیق جعل کردن، تکنیکی برای تلفیق تصویر انسان بر اساس هوش مصنوعی است. این فناوری جهت ترکیب و قرار دادن تصاویر و فیلم های موجود بر روی تصاویر یا فیلم های مورد نظر با استفاده از تکنیک یادگیری ماشین تحت عنوان "شبکه مولد تخاصمی (**GAN**)"[1] بکار می رود.

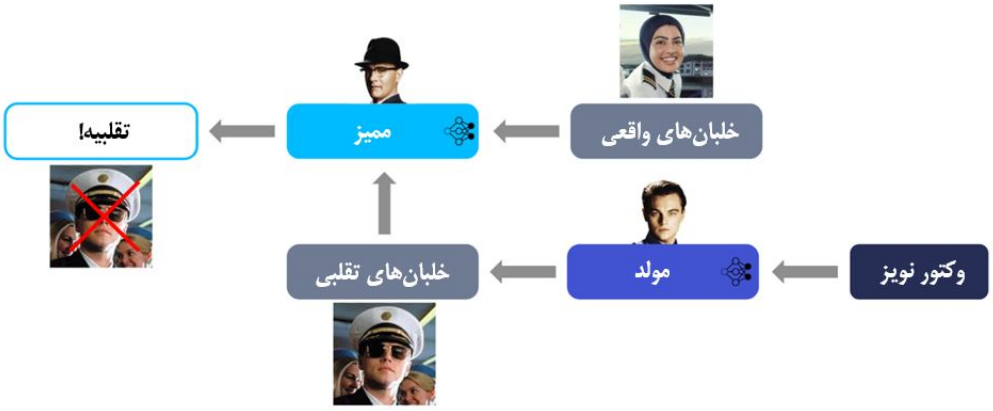
تلفیق تصاویر و فیلم های موجود با منبع مورد نظر ویدئویی به نحوی صورت میگیرد که گویی ترکیب هر دو تصویر یا هر دو فیلم یکی است و در یک صحنه رخ می دهد . این ترکیب پیچیده به طور مثال می تواند یک فرد یا افراد را به گفتن چیزها و یا انجام اقداماتی نشان دهد که هرگز در واقعیت رخ نداده اند.[۲]

GAN ۲.۱

گَن یک بخش «مولد» دارد و یک بخش «ممیز». مثلا فرض کنید گَن قصه ما، یاد گرفته باشد تصویرهای دروغین از خلبان‌ها بسازد. مولد یا همان ابگنیل قصه، دارد سعی می‌کند عکس‌های تقلبی جعل کند. از آن طرف ممیز یا هنرتی داستان، عکس‌های جعلی ساخت مولد و عکس‌های واقعی را کنار هم می‌گذارد و به خودش یاد می‌دهد که عکس واقعی را از دروغی تشخیص دهد. هرچقدر که مدل گَن بیشتر یاد بگیرد، هر دو بخش‌های مولد و ممیز هم در کارشان بهتر عمل می‌کنند و هرکدام باعث می‌شود دیگری در کار خودش ماهرتر شود. دیپ‌فیک‌ها جعل‌هایی هستند که مولد بلاخره توانسته از زیر دماغ ممیز رد کند! [۱]

شاید قبلا سایت «این فرد وجود خارجی ندارد» را دیده باشید؛ این سایت با کمک هوش مصنوعی، عکس پروفایل‌های آدم‌هایی را می‌سازد که هرگز وجود نداشته‌اند. چطوری؟ با الگوریتم شبکه مولد تخاصمی (Generative Adversarial Network) یا به طور خلاصه، گَن (GAN). این الگوریتم خودش برای خودش یک پا «اگه می‌تونی منو بگیر» است و ابگنیل و هنرتی انحصاری خودش را دارد(شکل ۱.۱) :

[1] Generative adversarial network



شکل ۱.۱: نحوه کارکرد شبکه مولد تخاصمی (گَن) به زبان «اگه می‌تونی منو بگیر»

۳.۱ دیپ فیک Deepfake چگونه ایجاد می شود؟

از آنجاییکه Deepfake با روش یادگیری ماشین کار میکند ، جهت جعل چهره ، Deepfake به داشتن چند صد تصویر از حالات مختلف چهره هدف نیاز دارد. به منظور جابجایی چهره ؛ مجموعه داده های چهره هدف و ویدئو مقصد مورد نیاز است و این یکی از دلایلی است که اشخاص معروف و سیاستمداران بیشتر در خطر هستند و با سادگی با جستجویی سریع در اینترنت می توان تمام داده های مورد نیاز را بدست آورد.

بهترین روش حفاظت در برابر Deepfake ، بستگی به خود شما دارد که آیا می خواهید عکس خود را از دید عموم حفظ کنید و یا آن را در اینترنت به صورت امن نگه دارید .

شاید تجمیع صدها تصویر از زوایای مختلف سبب ایجاد تصویری بسیار با کیفیت شود اما نبود این تعداد تصویر و زوایا نیز ، باز هم می تواند مورد استفاده Deepfake قرار گیرد با روش هایی همانند کشیدن تصاویر و فریم های چندگانه می توان شکاف ها را پر کرد. برنامه Everytime آیفون تصویری که تهیه می کند ، حداقل ۳۰ فریم در ثانیه ضبط می شود.

به هر حال برخی Deepfake ها آنقدر هوشمند شده اند که کیفیت و کمیت مورد نظر تصاویر را می توانند بدست آوردند. به خصوص اگر زوایای صورت و چهره ها با ویدیو مقصد مورد نظر هماهنگ شود و اگر هم نباشد ممکن است ویدئو و تصاویر مقصد را با مبدا هماهنگ کنند.

فصل ۲

بد: جعل، تحقیر، و دروغ

پیشرفت فناوری دیپ‌فیک حالا به جاهای نگران‌کننده‌ای رسیده است. نه فقط عکس‌های تقلبی، که حالا می‌شود با این فناوری ویدیوهای فیک هم ساخت. خوشبختانه ویدیوهای دیپ‌فیک هنوز در ابتدای مسیرند و هرکس بخواهد آنها را بسازد، باید فریم به فریم ادیتشان کند. علاوه بر این، بیشتر می‌شود تغییرات ظاهری ایجاد کرد نه حرکات واقعی. اما با همه اینها، همین دو سال پیش کاربر ناشناسی در ردیت با نام مستعار deepfakes، چهره گَل گَدو، بازیگر نقش واندر وومن را روی یک ویدیوی مستهجن گذاشت. کاربری دیگر در ردیت، اپی به نام فیک‌اپ ساخت که با آن می‌شد به سادگی ویدیوهای فیک ساخت و به اشتراک گذاشت. امروز دیگر جایگزین کردن چهره یک نفر در ویدیو با یکی دیگر نه تنها کار سختی نیست، که نتیجه آن هم تا حد خیلی خوبی غیر قابل تشخیص است.

در دنیایی که روزانه ۹۳ میلیون سلفی در آن گرفته می‌شود، تصور کردن انواع سوءاستفاده‌ها و جرایمی که می‌شود با دیپ‌فیک انجام داد، خیلی کار سختی نیست: ساخت ویدیوهای مستهجن برای انتقام‌گیری، باج‌گیری، سرقت هویت، پخش اخبار دروغ. تازه، اینها تنها چند مثال از انبوه گزینه‌های ممکن است. در قالب تئوری، تک‌تک افرادی که عکس خودشان را در شبکه‌های اجتماعی به اشتراک بگذارند در آینده در مقابل چنین حملاتی آسیب‌پذیر خواهد بود. ویدیوهایی مثل سخنرانی فیک اوباما را با این فناوری می‌شود در انواع و اقسام تولید کرد، افرادی که مستعد گول خوردن هستند را با آنها به راحتی گول زد، و بعد اعضای جامعه دو قطبی امروز حتی بیشتر از قبل با هم دشمن می‌شوند. وقتی اینترنت و شبکه‌های اجتماعی بی‌سانسور و کنترل نشده ما سرشار از دیپ‌فیک شوند، کل جامعه و زندگی‌های خصوصی تک‌تک ما در خطر خواهد بود.

فصل ۳

زشت: گرداب بی‌پایان رقابت‌ها

۱.۳ آیا ساخت دیپ فیک آسان است؟

مثل هر نوع یادگیری ماشینی و غیر ماشینی دیگر، برای ساختن دیپ‌فیک‌های واقع‌گرایانه هم اول باید داده کافی جمع کرد. اما این خیلی هم کار ساده‌ای نیست. کوین روز (Kevin Roose)، یکی از نویسندگان ستون فناوری در نیویورک تایمز، تصمیم گرفت با دادن یک سری ویدیو از خودش به اپلیکیشن فیک‌اپ، خودش امتحان کند. دیپ‌فیکی که او با این روش از خودش ساخت بامزه بود، اما آنقدر واقعی نبود که آدم را گول بزند. بنابراین به نظر می‌رسد که فعلا، فقط از اشخاص مشهور می‌شود دیپ‌فیک «واقعی» ساخت؛ یعنی آنهایی که کلی ویدیوی باکیفیت ازشان در سطح وب وجود دارد. اما همین هم چیز کمی نیست؛ مثلا دیپ‌فیک سیاستمداران می‌تواند بین صدها میلیون آدم دست به دست بچرخد و نتایج زیانباری برای وقایع سیاسی مهم داشته باشد؛ مثلا برای نتایج انتخابات ریاست جمهوری آمریکا در سال ۲۰۲۰.

۲.۳ میتوان دیپ فیک ها را شکار کرد؟

احتمال بروز چنین اتفاقاتی، بار سنگینی بر دوش فناوری‌های تشخیص و شکار دیپ‌فیک گذاشته است. متخصصان مختلف دارند همه تلاششان را می‌کنند تا ابزارهای قدرتمندی برای تشخیص دیپ‌فیک بسازند. مثلا پژوهشگران دانشگاه واشنگتن، سایت «کدام چهره واقعی است» را ساخته‌اند که در آن می‌توانید بعد از خواندن مطلبی در مورد نحوه تشخیص تصاویر دیپ‌فیک، سعی کنید بین یک سری عکس‌های واقعی و جعلی، واقعی‌ها را تشخیص دهید. کایل مک‌دانلد هم در یک پست وبلاگ در مورد همین مسئله توضیح داده است. گروهی از محققان یک دیتاست عظیم ویدیویی جمع کرده‌اند که قرار است در علوم قانونی (Forensics) برای تشخیص جعل عکس، نقش محک و معیار را داشته باشد. استارتاپ دیپ‌تریس (Deeptrace)، مبارزه با خطرات سایبری دیپ‌فیک را به عنوان هدف خود برگزیده و گزارشی در مورد وضعیت دیپ‌فیک در سال ۲۰۱۸ ارائه کرده است. هنری فرید در دارت‌ماوت دارد نرم‌افزاری برای شناسایی دیپ‌فیک‌های ویدیویی سیاسی می‌سازد. سیوی لیو با همکاری آژانس پروژه‌های پژوهشی پیشرفته دفاعی (دارپا)، در حال توسعه نرم‌افزاری است که بتواند دیپ‌فیک‌های ویدیویی را شناسایی کرده و از انتشار و رواج آنها جلوگیری کند.

شکل ۱.۳: تهدید دیپ‌فیک ها

۳.۳ نتیجه این شکار دیپ فیک ها چیست؟

دیپ‌فیک‌ها درست مثل ویروس‌های کامپیوتری هستند. به محض اینکه یکی راه شناسایی ویروسی را پیدا کند، فوری یک نفر دیگر می‌آید و راهی برای دور زدن راهکار نفر قبلی پیدا می‌کند. این قایم‌موشک‌بازی بین جعل‌کننده‌ها و تشخیص‌دهنده‌ها رقابتی است که تا ابد ادامه خواهد داشت. یک جورهایی، انگار خود این دو گروه با هم یک GAN عظیم و بی‌پایان می‌سازند!

۴.۳ تهدید دیپ فیک Deepfake چه قدر جدی است؟

اما تهدید deepfakes، واقعا تا چه حد پیچیده و جدی است؟ اگر در آینده اعتقاد به این پیدا کنیم که اکثر ویدئو ها جعلی هستند پس به چه چیز می توانیم اعتماد کنیم . تفاوت میان دروغ و درستی چگونه قابل تشخیص خواهد بود؟

هانی فرید می گوید: "مردم باید آگاه باشند که این چیزها وجود دارند اما باید درک کنند که ما با فناوری همراه هستیم، و اینکه چه چیزی می تواند و چه چیزی نمی تواند جعلی باشد." ، "مردم ممکن است یکباره خشمگین شوند و شروع به دیوانگی کنند. اما باید بتوانند این احساسات را کنترل کرده و سطح شناخت خود از اشخاص دور و برشان بالا ببرند به جای اینکه به ویدیوها و تصاویر جعلی اکتفا کنند ."[۲]

۵.۳ آیا Deepfakes مهمترین تهدید در آینده خواهد بود؟

بله واقعیت دارد Deepfakes خطرناک است و ممکن است به نوعی تهدیدی جدی برای اشخاص مشهور ، سیاستمداران و یا حتی مردم عادی به شمار آید. ولی این فقط Deepfakes نیست که باید از آن ترسید ،

تهدیدهای بزرگتر دیگری هستند که اکثریت مردم عادی را نیز هدف قرار داده اند . همین دو سال اخیر بود که موضوع حفظ حریم شخصی توسط فیسبوک و گوگل به شدت زیر سوال رفت و به دنبال آن باگ امنیتی فیستایم اپل و برنامه هایی که بر روی موبایل ها خواسته و ناخواسته نصب می شوند و تمام اطلاعات و فعالیت های کاربران را با اهداف تجاری یا سوء استفاده های دیگر رصد میکنند. فن آوری ها به سرعت پیشرفت می کنند و به راحتی میان مردم منتشر می گردند؛ اما هنوز قوانین جامع و کاملی برای نحوه استفاده از آن ها وجود ندارد ، هنوز حتی قانونگذاران هم اطلاعات جامعی در خصوص دستگاه هایی که میان مردم عرضه می شود ندارند. پس deepfake تنها دغدغه نخواهد بود .[۲]

فصل ۴

خوب: هوش مصنوعی شبه‌انسانی و تعاملات ارتباطی نو

جهان پر از دیپ‌فیک شاید مخوف و ناامن باشد؛ اما این تنها آینده ممکن نیست! دیپ‌فیک کاربردهای مثبتی هم دارد؛ دیپ‌فیک می‌تواند نحوه ارتباطات را دگرگون کند و شکل‌های کاملا جدیدی از آن به وجود بیاورد. به عنوان مثال، تصور کنید فناوری تولید صدا با فناوری دیپ‌فیک ترکیب شود! حتی همین حالا هم دستیار گوگل می‌تواند با استفاده از یک مدل مولد برای تولید گفتار به نام ویونت (Wavenet)، با لحن و صدای جان لجندصحبت کند. استارتاپ‌های لایربرد (Lyrebird)و ماجولیت (Modulate) می‌توانند فقط با چند ساعت آموزش دیدن، یاد بگیرند مثل شما حرف بزنند! حتی فناوری بایدو (Baidu) می‌تواند تنها در ۳.۷ ثانیه صدای افراد را شبیه‌سازی کند. در آینده‌ای نه چندان دور، سخنگوهای هوشمندی خواهیم داشت که نه تنها می‌توانند با صدای خواننده‌های مورد علاقه‌مان صحبت کنند، که بلدند وقتی خودمان سر کار نیستیم، به جای ما جواب تلفن را بدهند!

حتی تولید ویدیوهای فیک هم الزاما چیز بدی نیست. شرکت سینتزیا (Synthesia) ، ویدیوی دیپ‌فیکی از دیوید بکهام را برای یک کمپین حمایتی مالاریا تولید کرده است. استارتاپ دیتاگرید (DataGrid) توانسته تصویر کل بدن آدم‌هایی که وجود خارجی ندارند را کامل بسازد، و پژوهشگران دانشگاه برکلی کالیفرنیا با کمک دیپ‌فیک توانسته‌اند حرکات مختلف رقص را روی بدن افراد مختلف اجرا کنند.

تصور کنید شخصیت‌های تاریخی‌ای که رویتان تاثیر گذاشته‌اند با شما حرف بزنند. تصور کنید عزیزانی که از دنیا رفته‌اند دوباره به پیشتان برگردند. و اگر هنوز دارید در آرزوی رفتن به دنیای هری‌پاتر می‌سوزید، تصور کنید همه آن پرتره‌های متحرک هاگوارتز واقعا وجود داشته باشند!

فصل ۵

جمع‌بندی

فرنک ابگنیل فیلم «اگه میتونی منو بگیر»، یک شخصیت واقعی است که وقتی سرانجام از زندان آزاد شد، چهار دهه از عمرش را به کار کردن در FBI [1] روی جعل، اختلاس، جرم‌های مالی، و جرم‌های امنیت سایبری گذراند. ابگنیل در سخنرانی‌ای در گوگل گفته که «فناوری، زاینده جرم است. همیشه همین بوده و همیشه هم چنین خواهد بود». با اینحال، او می‌گوید که این هکرها نیستند که امنیت اطلاعات شرکت‌ها را پایین می‌آورند؛ بلکه مقصر آن کارمندان سهل‌انگاری در آن شرکت هستند که کاری که قرار نبوده بکنند را کرده‌اند، یا اینکه نتوانسته‌اند کاری که قرار بوده بکنند را انجام دهند.

همین اتفاق برای دیپ‌فیک هم می‌افتد: مهم نیست دیپ‌فیک‌ها چقدر واقع‌گرایانه‌تر شوند یا دقت فناوری‌های ضد دیپ‌فیک تا چه حد افزایش پیدا کند؛ آسیب اصلی ناشی از این دیپ‌فیک‌ها کار انسان‌هایی هست که آنها را می‌سازند، ساخته‌های دروغین را باور می‌کنند، و چیزی که بدون تحقیق، درست فرض کرده‌اند را نشر می‌دهند. به جای اینکه انگشت اتهام را به سمت خود فناوری دیپ‌فیک بگیریم، باید ببینیم چطور می‌شود کاری کرد که افراد در مورد چیزهایی که در اینترنت می‌بینند با دید انتقادی‌تری قضاوت کنند و هنگام به اشتراک‌گذاری در شبکه‌های اجتماعی هوشمندانه‌تر عمل نمایند. اثرات منفی دیپ‌فیک را نمی‌شود انکار کرد؛ اما باید چشم‌هایمان را به بخش‌های مثبت‌تر هوش مصنوعی بدوزیم و پتانسیلی که دیپ‌فیک برای ایجاد روش‌های ارتباطی جدید و بهتر کردن زندگی‌هایمان دارد را به مسیر درست هدایت کنیم.

[1] Federal Bureau of Investigation

مراجع

[1] https://medium.com/twentybn/deepfake-the-good-the-bad-and-the-ugly-8b261ecf0f52

[2] https://www.asriran.com/fa/news/673335/deepfake . .