

مزایا، معایب و آینده تکنولوژی دیپفیک صوتی - دیجیاتو

شایان ضیایی | پنجشنبه، ۱۴ اسفند ۱۳۹۹

از تکنولوژی دیپفیک صوتی که منجر به کلون کردن اصوات یا بازتولید تقریباً بی نقص صدای افراد می‌شود می‌توان هم برای مقاصد خوب و هم مقاصد بد استفاده کرد. برای مثال می‌توان آن را در سرویس‌های ترکیب صدا به کار گرفت و صدا را به افرادی که این توانایی را به خاطر عارضه‌هایی مانند اسکروز جانبی آمیوتروفیک، آپراکسی، ضربه‌های مغزی تروماتیک، سکته یا هرچیز مشابهی از دست داده‌اند بازگرداند.

از این تکنولوژی همین حالا در صنایع فیلم و تلویزیون، گیمینگ و مراکز تماس استفاده می‌شود و بی‌شمار پتانسیل نیز برای استفاده از آن در امور رمزنگاری و تراپی نیز وجود ندارد. در عین حال، نمی‌توان این را زیر سوال برد که تکنولوژی دیپفیک صوتی تهدیدی بزرگ برای پروسه‌های دموکراتیک است، خصوصاً پروسه‌هایی که به حریم شخصی مربوط می‌شوند. در صورت استفاده بدخواهانه، تکنولوژی دیپفیک صوتی می‌تواند به کلاهبرداری، فریب و آزار منجر شود.

دقیقاً به خاطر همین پتانسیل‌های خوب و بد، اکنون زمانش رسیده که توجهی مضاعف به تکنولوژی دیپفیک صوتی نشان دهیم و به طریقی به بررسی به آن بپردازیم که منجر به کاهش خطر فریب دادن مردم شود، خصوصاً این فریب که شخصی سرشناس، چیزی را گفته که در حقیقت بیانش نکرده است.

موارد استفاده برای کسب‌وکارها چه خواهد بود؟ ارقام نشان می‌دهند که فرصتی بزرگ برای سازمان‌های اقتصادی فراهم شده تا از تکنولوژی‌های صوتی استفاده کرده و وارد تجارت‌های تازه شوند. موسسه AppDynamics در سال ۲۰۱۸ گزارش کرد که تا سال ۲۰۲۰، بیش از نیمی از سرچ‌های انجام شده در سطح وب، با صدا صورت خواهند گرفت.

از سوی دیگر نیز، ۶۱ درصد از تصمیم‌گیران برجسته حوزه فناوری اطلاعات گام را فراتر گذاشته و انتظار دارند فرمان‌های صوتی به صورت کامل جایگزین فرمان‌های تایپ شده به صورت دستی برای یافتن اطلاعات در سراسر اینترنت شوند. و نسل جوان در این حوزه پیشرو است - ۸۴ درصد از متولدین قرن بیست و یکم همین حالا از دستیارهای صوتی برای پایش برنامه روزانه و مسئولیت‌های خود استفاده می‌کنند.

بیاید به مرور مزایا، معایب و آینده که در انتظار تکنولوژی دیپفیک صوتی است بپردازیم.

مزایا



امکان صحبت به صورت طبیعی از سوی افرادی که از مشکلات پزشکی رنج می‌برند

امکان برقراری ارتباط و اشتراک‌گذاری افکار و احساسات با کلمات، از اهمیتی بسیار زیاد برخوردار بوده و در واقع این یکی از معدود ویژگی‌هایی است که ما انسان‌ها را منحصر به فرد می‌کند. اما تکنولوژی واگردانی صوتی سنت‌شکنانه کنونی، کارهای بسیار بیشتری می‌تواند برای افرادی انجام دهد که از توانایی صحبت برخوردار نیستند.

برای مثال گسترش تکنولوژی اتوماسیون خانگی را کنترل صوتی را در نظر بگیرید. کلون کردن صدا می‌تواند باعث شود افرادی که قادر به صحبت به صورت طبیعی نیستند استقلال بیشتری داشته باشند و بهتر به استفاده از دیوایس‌هایی بپردازند که از با فرمان‌های صوتی کنترل می‌شوند.

دستیارهای صوتی

بنابر گزارش «پیش‌بینی وضعیت دستیار دیجیتال و دیوایس صوتی مبتنی بر هوش مصنوعی» شرکت Ovum، تا سال ۲۰۲۱ میلادی شمار دستیارهای صوتی از تعداد انسان‌های حاضر روی کره خاکی بیشتر خواهد بود. بیا بید برای مثال به دستیار صوتی گوگل نگاه کنیم. صدای این دستیار با استفاده از سیستم نوشتار به گفتار Tacotron 2 تولید می‌شود که خود بر دو شبکه عصبی عمیق

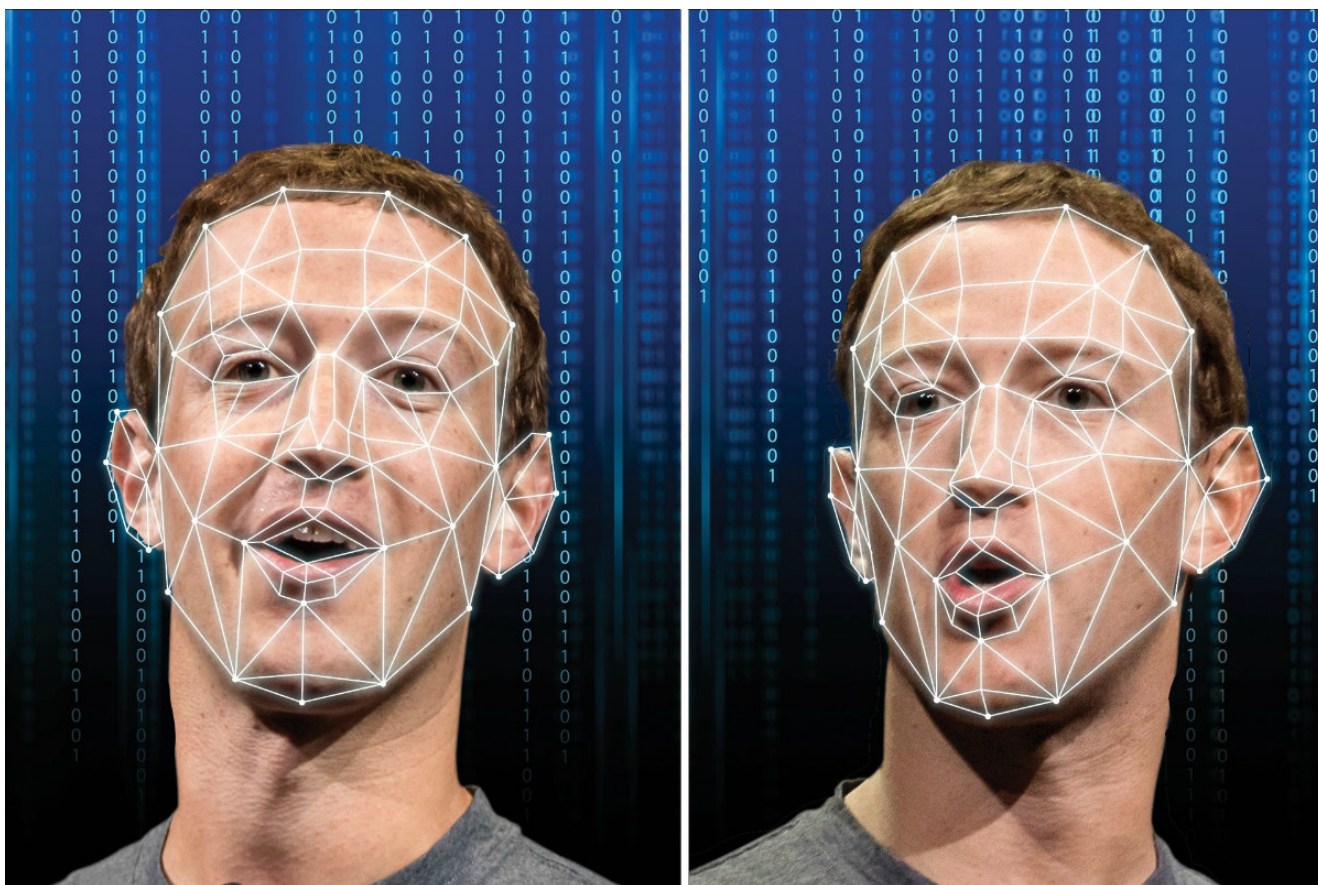
متکی است.

نخستین شبکه، متن را به تصویری بصری (یا طیف نگاره) از فرکانس‌های صدا به مرور زمان تبدیل می‌کند و سپس یک سیستم WaveNet به تحلیل طیف نگاره پرداخته و المان‌های صوتی را شکل می‌دهد. خروجی، گفتاری است که تفاوت چندانی با گفتار عادی انسانی ندارد و حتی می‌تواند کلمات بسیار دشوار را به خوبی تلفظ کند.

محتوای تعاملی برای کورس‌های یادگیری آنلاین

کلون کردن صدا با هوش مصنوعی باعث می‌شود که دیگر در هر کلاس درس نیازی به نوت‌برداری یا بازنگری در نوت‌ها برای یافتن اشتباهات بالقوه نباشد. از سوی دیگر، هزینه پولی و زمانی سخنرانی‌هایی که به صورت حرفه‌ای ضبط می‌شوند نیز کاهش می‌یابد و کورس‌های یادگیری آنلاین شکلی دسترس‌پذیرتر به خود می‌گیرند. این اتفاقی کوچک نیست، خصوصاً در دوران بحران کنونی که به خاطر پاندمی کووید-۱۹، محدودیت‌های فراوان برای مردم به وجود آمده است.

معایب



اخاذی

از ویدیوهای جعلی اما شدیداً واقع‌گرایانه با صحنه‌های جنسی یا خشونت‌آمیز که با ترکیب تکنولوژی دیب‌فیک ویدیویی و صوتی ساخته شده‌اند می‌توان برای اخاذی از افراد استفاده کرد.

ایمیل‌های اسپم

اگر تا به حال ایمیل برایتان آمده باشد که در آن نوشته شده «با بانک X ارتباط برقرار کنید تا اطلاعات بیشتر راجع به روند حواله پول در اختیارتان قرار بگیرد» به احتمال زیاد نام ایمیل‌های اسپم به گوشتان خورده و هیچ توجهی هم به آن‌ها نشان نداده‌اید. اما یک تماس تلفنی از سوی کسی که صدایی بسیار شبیه به یکی از مخاطبان مورد اعتمادتان دارد و شما را به پاسخ دادن به یک ایمیل ترغیب می‌کند ممکن است نظرتان را عوض کرده و باعث شود دست به انجام کاری بزنید که بعداً آرزو می‌کردید هیچوقت انجامش نمی‌دادید.

رقابت غیر قانونی

یک نفر ممکن است خود را مدیر عامل فلان کمپانی جای بزند و در این جایگاه مهم، شروع به نشر اطلاعات دروغین در تماس‌های تلفنی مربوط به گزارش‌های مالی کند. سهام‌داران و سرمایه‌گذاران فریب خواهند خورد و قیمت سهام را از آنچه در واقعیت هست خواهند پنداشت. از همین تکنیک بدخواهانه می‌توان برای تخریب رقبای صنعتی نیز استفاده کرد.

آینده



استفاده در مدل بنگاه-به-بنگاه

با در نظرگیری یافته‌های موجود در گزارش AppDynamics که بالاتر به آن‌ها اشاره شد، استفاده از تکنولوژی بازتولید صدا به صورت گسترده، آنقدرها دور نیست. ۶۹ درصد از تصمیم‌گیرندگان حوزه فناوری اطلاعات در شرکت‌هایی کار می‌کنند که یا همین حالا روی تکنولوژی سرمایه‌گذاری کرده‌اند یا می‌خواهند طی سه سال آتی روی آن سرمایه‌گذاری کنند.

چه به دنبال بازتولید صدا باشید، چه درمان برای مشکلات کلامی، چه گیمینگ، چه رمزنگاری و چه هرچیز دیگر، از تکنولوژی کلون صدا می‌توان در گستره وسیعی از صنایع نفع برد. برای مثال با توسعه یک سیستم بلادرنگ در دنیای گیمینگ می‌توان به گیمرها اجازه داد که از صداهای مختلف در چت‌های داخل بازی استفاده کنند. وقتی صحبت از تکنولوژی دیپ‌فیک صوتی باشد، جهان دارد با سرعت بسیار زیادی تغییر می‌کند.

کلون صدا برای مراکز تماس

کمپانی Reespecher اکنون مشغول کار روی تکنولوژی‌هایی سنت‌شکن است که به اپراتورها در آن‌سوی آب‌ها اجازه می‌دهد صدایی شبیه به افراد بومی داشته باشند. بنابراین هنگام تماس با یکی از این مراکز، احساس خواهید کرد که با یکی از شهروندان کشور خود صحبت می‌کنید. از سوی دیگر، اپراتورهای رباتیک هم به زودی صدایی انسانی‌تر خواهند داشت.

جمع‌بندی

تکنولوژی دیپ‌فیک صوتی بدون هیچ تردید ریسک‌های خود را به همراه می‌آورد، اما نخستین گام برای تمرکز مضاعف روی جنبه‌های مثبت تکنولوژی اینست که جنبه‌های منفی را شناسایی کرده

و در صدد تقلیل دادن آن‌ها برآییم. صداهای کلون شده‌ای که هیچ تفاوتی با صدای سخنگوی اصلی ندارند می‌توانند بهترین ابزار ممکن برای فیلم‌سازان، بازی‌سازان و دیگر تولیدکننده مدیا باشند.

خیلی زود صداهای بازتولید شده به صورت دیجیتالی، احساسات و حالات گوناگون را نیز در بر می‌گیرند، همانطور که شواهدش را با دستیارهای صوتی نظیر بیکسبی سامسونگ، سیری اپل یا الکسای آمازون دیده‌ایم. و با از راه رسیدن صداهایی که قادر به انتقال احساسات باشند، مشاوره و مصاحبت شکلی کاملاً جدید به خود خواهند گرفت.

[دیجیاتو](#)