دوشنبه 22 اردیبهشت 1404 - 04:28

هدفون‌های هوش مصنوعی که صحبت چندین گوینده را هم‌زمان ترجمه می‌کنند

هدفون‌های مجهز به هوش مصنوعی می‌توانند ترجمه گروهی را به طور هم‌زمان با شبیه‌سازی صدا ارائه دهند.

هدفون های مجهز به هوش مصنوعی می توانند ترجمه گروهی را به طور هم زمان با شبیه سازی صدا ارائه دهند.

به گزارش ایسنا، «توچائو چن»(Tuochao Chen) دانشجوی «دانشگاه واشنگتن» اخیرا از موزه ای در مکزیک بازدید کرد. چن اسپانیایی صحبت نمی کند. بنابراین، یک اپلیکیشن ترجمه را روی تلفن همراه خود اجرا کرد و میکروفون را به سمت راهنمای تور گرفت اما حتی در سکوت نسبی موزه، سر و صدای اطراف زیاد بود و متن حاصل فایده زیادی نداشت.

به نقل از تک اکسپلور، اخیرا فناوری های گوناگونی ظهور کرده اند که ترجمه روان را نوید می دهند اما هیچ کدام از این فناوری ها مشکل چن را در فضاهای عمومی حل نکردند. برای مثال، عینک های جدید شرکت «متا»(Meta) فقط با یک بلندگوی مجزا کار می کنند. آنها پس از به پایان رسیدن صحبت گوینده، ترجمه صوتی خودکار را پخش می کنند.

اکنون چن و گروهی از پژوهشگران دانشگاه واشنگتن یک سیستم هدفون طراحی کرده اند که هم زمان صحبت چندین گوینده را ترجمه می کند و در عین حال، جهت و کیفیت صدای افراد را حفظ می کند. این گروه پژوهشی، سیستم را با هدفون های نویزگیر موجود در بازار که به میکروفون مجهز هستند، ساخته اند. الگوریتم های این گروه پژوهشی، گویندگان متفاوت را در یک فضا جدا می کنند، آنها را هنگام حرکت دنبال می کنند، گفتار آنها را ترجمه می کنند و با تأخیر دو تا چهار ثانیه ای پخش می کنند.

«شیام گولاکوتا»(Shyam Gollakota) استاد دانشکده علوم رایانه و مهندسی دانشگاه واشنگتن و پژوهشگر ارشد این پروژه گفت: کد دستگاه برای دیگران در دسترس است تا براساس آن کار کنند. سایر فناوری های ترجمه بر این فرض ساخته شده اند که فقط یک نفر صحبت می کند اما در دنیای واقعی نمی توانید فقط یک صدای رباتیک داشته باشید که برای چندین نفر در یک اتاق صحبت کند. ما برای اولین بار صدای هر شخص و جهتی را که صدا از آن می آید، حفظ کرده ایم.

این سیستم سه نوآوری را در بر دارد. نخست این که وقتی روشن می شود، بلافاصله تشخیص می دهد چه تعداد اسپیکر در فضای داخلی یا خارجی وجود دارد. چن گفت: الگوریتم های ما کمی شبیه به رادار کار می کنند. بنابراین، آنها فضا را به صورت ۳۶۰ درجه مورد بررسی قرار می دهند و دائما به روزرسانی می کنند تا مشخص شود چند نفر در حال صحبت کردن هستند.

سپس سیستم، گفتار را ترجمه می کند و کیفیت بیان و بلندی صدای هر گوینده را هنگام اجرا روی یک دستگاه مجهز به تراشه «Apple M۲» مانند لپ تاپ ها و هدست «اپل ویژن پرو»(Apple Vision Pro) حفظ می کند. این گروه پژوهشی به دلیل نگرانی های مربوط به حریم خصوصی پیرامون شبیه سازی صدا، از به کار بردن محاسبات ابری اجتناب کردند. در نهایت، هنگامی که گوینده ها سر خود را حرکت می دهند، سیستم همچنان به ردیابی جهت و کیفیت صدای آنها همراه با تغییرات صورت گرفته ادامه می دهد.

این سیستم در ۱۰ محیط داخلی و خارجی آزمایش شد و در یک آزمایش با ۲۹ شرکت کننده، کاربران این سیستم را به مدل هایی که اسپیکرها را در فضا ردیابی نمی کردند، ترجیح دادند.

در یک آزمایش جداگانه روی کاربران، بیشتر شرکت کنندگان تأخیر سه تا چهار ثانیه ای را ترجیح دادند زیرا سیستم هنگام ترجمه با تأخیر یک تا دو ثانیه ای، خطاهای بیشتری مرتکب می شد. این گروه پژوهشی در تلاش هستند تا سرعت ترجمه را در نسخه های آینده کاهش دهند. این سیستم در حال حاضر فقط روی گفتار روزمره کار می کند، نه زبان تخصصی مانند اصطلاحات فنی. پژوهشگران در این پروژه با زبان های اسپانیایی، آلمانی و فرانسوی کار کردند اما بررسی های پیشین روی مدل های ترجمه نشان داده اند که می توان آنها را برای ترجمه حدود ۱۰۰ زبان آموزش داد.

چن گفت: این گامی به سوی از بین بردن موانع زبانی بین فرهنگ هاست. بنابراین، اگر من در خیابان مکزیک قدم بزنم، حتی اگر اسپانیایی صحبت نکنم هم می توانم صدای همه مردم را ترجمه کنم و بدانم چه کسی چه گفته است.