



آیا هوش مصنوعی، مترجمی و زبان اشاره را منسوخ می‌کند؟

هوش مصنوعی در حال تغییر دید ما از زبان و نحوه دسترسی به آن است و با پیشرفت آن در سالیان آتی شاید شاهد تحول در زمینه ترجمه زنده و صحیح اغلب زبان‌های دنیا از جانب هوش مصنوعی باشیم.

هوش مصنوعی در حال تغییر دید ما از زبان و نحوه دسترسی به آن است و با پیشرفت آن در سالیان آتی شاید شاهد تحول در زمینه ترجمه زنده و صحیح اغلب زبان‌های دنیا از جانب هوش مصنوعی باشیم.

به گزارش ایسنا، در عصر اینترنت، مردم به یکدیگر نزدیک و نزدیک تر می‌شوند. شما می‌توانید با دوست خود در هر کجای دنیا به صورت صوتی و تصویری تماس برقرار کنید یا بسیاری از کارهای سنتی را که قبلاً به ابزار و کاغذ و موارد دیگر نیاز داشت، تنها با موبایل هوشمند خود در کمترین زمان ممکن انجام دهید.

اما هرچه دنیا به هم نزدیکتر می‌شود، توجه ما بیشتر و بیشتر به سمت فضای مجازی معطوف می‌شود. ما ساعت‌ها در طول روز به گردش در برنامه‌هایی نظیر اینستاگرام می‌پردازیم و زمان کمتری را صرف تعامل مستقیم با یکدیگر می‌کنیم.

نقش هوش مصنوعی در دنیای امروز

هوش مصنوعی (AI) موضوعی است که جهان امروز ما را به شدت تحت تاثیر قرار داده است و مانند بسیاری از فناوری‌ها، مزایای فراوان و البته برخی معایب نیز دارد. این فناوری هم‌اکنون به سرعت در حال پیشرفت و توسعه است و حتی برخی را نسبت به سلطه بر انسان طی سال‌های آینده نگران کرده است.

اما آیا هوش مصنوعی در آینده نزدیک جایگزین مترجمان و زبان اشاره خواهد شد؟

سرشاخ شدن گوگل با مترجمان

شرکت گوگل در مارس ۲۰۲۱ از ویژگی "Live Captions" به معنای "زیرنویس یا عنوان زنده" خود در مرورگرهای کروم (Chrome) رونمایی کرد. این ویژگی از فناوری یادگیری ماشینی استفاده می‌کند تا بلافاصله زیرنویس‌ها را روی هر کلیپ تصویری یا صوتی ایجاد کند و به افراد ناشنوا و کم‌شنوا دسترسی بیشتری به محتوای اینترنت بدهد.

در گذشته و امروز نیز از زیرنویس‌های از قبل تهیه شده برای قالب‌های ویدئویی استفاده می‌شود و یا اینکه یک تندنویس، تقریباً فوری و در لحظه مطالب را تایپ می‌کند تا به عنوان زیرنویس پخش شود. با این حال، در جاهایی که زیرنویس معمول و رایج نیست، مانند برنامه‌هایی مانند اینستاگرام یا تیک‌تاک، یافتن زیرنویس‌ها تقریباً غیرممکن است.

اکنون ویژگی "زیرنویس زنده" این موضوع را تغییر می‌دهد و هر کاربری با چند کلیک روی نمایشگر می‌تواند زیرنویس‌های آنی و دقیقی در اختیار داشته باشد که محتواهای صوتی و تصویری را در برمی‌گیرد.

ویژگی "زیرنویس زنده" شرکت گوگل نوعی "NLP" یا "پردازش زبان طبیعی" است. "NLP" نوعی هوش مصنوعی است که با استفاده از الگوریتم‌ها، برقراری تعامل میان مردم و ماشین‌ها را تسهیل می‌کند. "NLP" به ما کمک می‌کند تا زبان‌های انسانی را به زبان‌های ماشینی و برعکس ترجمه کنیم.

"آلن تورینگ" پدر هوش مصنوعی

برای درک تاریخچه "NLP" باید به سراغ یکی از باهوش‌ترین دانشمندان عصر مدرن به نام "آلن تورینگ" برویم. وی در سال ۱۹۵۰ مقاله "ماشین‌آلات رایانشی و هوش" را منتشر کرد که در مورد مفهوم رایانه‌های متفکر و دارای درک بحث می‌کرد.

وی در این مقاله ادعا کرد که هیچ استدلال قانع‌کننده‌ای علیه این ایده که ماشین‌ها می‌توانند مانند انسان فکر کنند، وجود ندارد و آزمون "بازی تقلید" (imitation game) را که اکنون به عنوان "آزمون تورینگ" شناخته می‌شود، ارائه کرد. "تورینگ" راهی برای سنجش اینکه آیا هوش مصنوعی می‌تواند به تنهایی فکر کند یا نه پیشنهاد کرد و گفت که اگر هوش مصنوعی بتواند

انسانی را فریب دهد تا آن انسان باور کند که هوش مصنوعی احتمالاً یک انسان است، می توان آن را هوشمند دانست.

"جوزف وایزناوم" دانشمند آلمانی از سال ۱۹۶۴ تا ۱۹۶۶ یک الگوریتم "NLP" نوشت که معروف به "الیزا"(ELIZA) است. "الیزا" از تکنیک های تطبیق الگو برای ایجاد مکالمه استفاده کرده است. به عنوان مثال در یک مکالمه مربوط به مراجعه بیمار به پزشک، اگر بیمار به رایانه بگوید "سر من درد می کند"، این عبارت با عبارتی شبیه به "چرا سرت درد می کند؟" پاسخ داده می شود. "الیزا" در حال حاضر به عنوان یکی از قدیمی ترین ربات های گفتگو و یکی از اولین الگوریتم هایی است که به نوعی در "آزمایش تورینگ" می تواند انسان را فریب دهد.

دهه ۱۹۸۰ نقطه عطف بزرگی در تولید "NLP" بود. در گذشته سیستم های "NLP" مانند "الیزا" با تکیه بر مجموعه ای پیچیده از قوانین، مکالمات را شکل می دادند و هوش مصنوعی نمی توانست برای خودش فکر کند. بلکه از پاسخ های از پیش آماده متناسب با موضوع گفتگو استفاده می کرد و وقتی یک انسان چیزی را به آن می گفت که پاسخی برای آن نداشت، با پاسخی نظیر "درباره این موضوع بیشتر به من بگوید" مواجه می شد.

"NLP" در اواخر دهه ۱۹۸۰ در عوض بر روی مدل های آماری متمرکز شد که به آن کمک می کرد مکالمات را بر اساس احتمال شکل دهد.

تشخیص گفتار مدرن "NLP" شامل چند اصل مشترک مانند تشخیص گفتار، تشخیص صدا، شناسایی زبان و خلاصه سازی است که می تواند بین سخنرانان تفاوت قائل شود.

سیستم "زیرنویس زنده" گوگل از سه مدل یادگیری عمیق برای تشکیل زیرنویس ها استفاده می کند: یک شبکه عصبی بازگشتی (RNN) برای تشخیص گفتار، یک RNN مبتنی بر متن برای تشخیص علائم نگارشی و یک شبکه عصبی حلقوی (CNN) برای طبقه بندی رویدادهای صوتی. این سه مدل، سیگنال هایی را ارسال می کنند که با هم ترکیب می شوند و زیرنویس را تشکیل می دهند.

وقتی گفتار در قالب صوت یا تصویر تشخیص داده می شود، شبکه عصبی بازگشتی تشخیص خودکار گفتار (ASR RNN) فعال می شود و به دستگاه اجازه می دهد کلمات را به شکل متن درآورد. وقتی این گفتار متوقف می شود، به عنوان مثال وقتی موسیقی به جای آن پخش می شود، فعالیت تشخیص خودکار گفتار برای صرفه جویی در باتری موبایل با نمایش برچسب "موسیقی" در زیرنویس، متوقف می شود.

همانطور که متن گفتار به صورت زیرنویس تنظیم می شود، علائم نگارشی روی جمله کامل شکل می گیرد. علائم نگارشی به طور مداوم تنظیم می شوند تا زمانی که نتایج تشخیص خودکار گفتار با معنی جمله تداخل نداشته باشد.

در حال حاضر، ویژگی "زیرنویس زنده" تنها می تواند زیرنویس را برای متون و گفتارها به زبان انگلیسی ایجاد کند، اما دائماً در حال بهبود است و در آینده نزدیک به زبان های دیگر نیز گسترش می یابد. هم اکنون نسخه های اولیه زیرنویس های اسپانیایی، آلمانی و پرتغالی در دسترس هستند.

"پروژه یوفونیا"

NLP های دسترس محور فقط به ایجاد زیرنویس محدود نمی شوند. یکی دیگر از پروژه های گوگل، "پروژه یوفونیا" (Project Euphonia) است که با استفاده از NLP به افراد دارای اختلال گفتاری کمک می کند تا توسط نرم افزار تشخیص گفتار بهتر شنیده و درک شوند.

"پروژه یوفونیا" ۳۰۰ تا ۱۵۰۰ عبارت صوتی را از داوطلبان دارای اختلال گفتاری جمع آوری می کند. سپس می توان این نمونه های صوتی را به مدل های تشخیص گفتار داده و برای آموزش انواع نقص گفتار تغذیه کرد. علاوه بر این، این برنامه سیستم های صوتی ساده شده ای ایجاد می کند که می توانند با استفاده از ردیابی صورت یا صداهای ساده اعمال مختلفی مانند روشن کردن چراغ یا پخش یک آهنگ خاص را انجام دهند.

یکی از جدیدترین شبکه های عصبی بازگشتی تشخیص خودکار گفتار گوگل در تلاش است تا نحوه تعامل ما با دیگران را تغییر دهد و دامنه ارتباط را گسترش دهد. حالت "مترجم گوگل" (Google Interpreter) برای شناسایی آنچه شما می گوید از تشخیص خودکار گفتار استفاده می کند و ترجمه دقیق آن را به زبان دیگری پخش می کند و به طور موثر مکالمه ای را بین افراد دارای زبان

های متفاوت ایجاد می کند و موانع زبانی را از بین می برد.

هنوز چند مشکل در سیستم تشخیص خودکار گفتار (ASR) وجود دارد. ماشین ها در مشکلی که اغلب "شکاف لهجه هوش مصنوعی" نامیده می شوند، گاهی اوقات در درک افراد با لهجه ها یا گویش های غلیظ مشکل دارند. در حال حاضر این مشکل به شکل مورد به مورد حل می شود.

دانشمندان تمایل دارند از یک مدل "لهجه واحد" استفاده کنند که در آن الگوریتم های مختلف برای گویش ها یا لهجه های مختلف طراحی شده است. به عنوان مثال، برخی از شرکت ها با استفاده از سیستم های جداگانه تشخیص خودکار گفتار برای تشخیص گویش های مکزیکی-اسپانیایی در مقابل گویش های اسپانیایی-اسپانیایی آزمایش کرده اند.

در نهایت، بسیاری از این سیستم های تشخیص خودکار گفتار نشان دهنده درجه ای از سوگیری و تعصب ضمنی است. در ایالات متحده، گویش انگلیسی بومی آفریقایی-آمریکایی که به آن "AAVE" نیز گفته می شود، گویشی کاملاً رایج از انگلیسی سنتی است که معمولاً توسط آمریکایی های آفریقایی تبار صحبت می شود. با این حال، مطالعات متعدد اختلافات نژادی قابل توجهی را در میزان خطای سیستم های مختلف تشخیص خودکار گفتار پیدا کرده است. به عنوان مثال یک مطالعه که نشان می دهد که میانگین خطا برای سیاه پوستان در برنامه های تشخیص خودکار گفتار شرکت های آمازون، اپل، گوگل، آی بی ام و مایکروسافت تقریباً دو برابر سفیدپوستان است.

ایجاد آموزش متنوع تر برای هوش مصنوعی که شامل لهجه های محلی، گویش های مختلف و عامیانه است می تواند به کاهش اختلاف در دقت تشخیص خودکار گفتار برای نژادها و قومیت های مختلف کمک کند.

این فناوری پتانسیلی باورنکردنی برای گردهم آوردن مردم دارد، اما وقتی دچار تعصب و جهت گیری می شود، می تواند یک نیروی تفرقه افکن و منزوی کننده باشد. ما به لطف فناوری پردازش زبان طبیعی (NLP) در حال شروع به پر کردن این شکاف برای ایجاد آینده ای بهتر هستیم.