



گوگل در کنفرانس توسعه‌دهندگان I/O امسال نشان داد تا چه اندازه در زمینه به‌کارگیری هوش مصنوعی در حوزه‌های مختلف دنیای فناوری توانمند شده است. رزرو یک میز در رستورانی شلوغ توسط چت‌بات هوشمند گوگل که قادر است ارتباطی زنده با مسئول پذیرش رستوران برقرار کند، تنها نمونه کوچکی از توانمندی گوگل در این زمینه است. این قابلیت منحصر به فرد نشان داد، گوگل مصمم است از هوش مصنوعی برای تبدیل سرویس‌های سنتی به سرویس‌های خودکار و هوشمند استفاده کند.

این مطلب یکی از مقالات پرونده ویژه «رایانش و خدمات ابری» شماره 205 ماهنامه شبکه است. علاقه‌مندان می‌توانند کل این پرونده ویژه را از روی سایت شبکه [دانلود](#) کنند.

قابلیتی که بسیاری از کسب‌وکارها و کاربران نیاز مبرمی به آن دارند، سرویس‌های تبدیل متن به گفتار و گفتار به متن است. دو سرویسی که به‌طور مستقیم از زیرساخت‌های ابری استفاده می‌کنند. زمانی که شما جمله‌ای را به زبان می‌آورید و دستیار صوتی جمله را دریافت می‌کند، محاوره شما را به سمت سرویس‌های ابری ارسال کرده تا این محتوا در سرویس ابری پردازش شده و در ادامه جمله شما اگر یک دستور باشد، روی دستگاهی که از آن استفاده می‌کنید اجرا شده و شما خروجی نهایی را مشاهده کنید. زمانی که هوش مصنوعی به این بازی وارد می‌شود شما با دو فاکتور سرعت و دقت سروکار خواهید داشت.

هم‌زمان با برگزاری کنفرانس NAB and SpeechTek گوگل از قابلیت‌ها و به‌روزرسانی‌های جدیدی که تصور می‌کرد به کسب‌وکارها کمک می‌کند و به شکل بهتری از سرویس تبدیل متن به گفتار و بالعکس استفاده کنند، رونمایی کرد. گوگل در وبلاگ این شرکت در ماه مارس اعلام کرد: از این پس توسعه‌دهندگان می‌توانند از سرویس تبدیل متن به گفتار ابری این شرکت استفاده کنند. سرویسی که به توسعه‌دهندگان اجازه می‌دهد با اتکا بر الگوی ویونت (WaveNet) و زیرساخت شبکه عصبی عمیق این شرکت از صدای طبیعی در ارتباط با تبدیل متن به گفتار در برنامه‌های کاربردی خود استفاده کنند. فناوری که توسعه‌دهندگان از آن استفاده می‌کنند، همان فناوری است که سرویس‌هایی همچون Google Maps و گوگل اسیستنت از آن استفاده می‌کنند. پروژه ویونت و واسط‌های برنامه‌نویسی مرتبط با این فناوری ماحصل پژوهش‌ها و تلاش‌های شرکت دیپ‌ماینند است. یک ماه پس از انتشار این خبر، گوگل در ماه آوریل اعلام کرد، پس از گذشت دو سال از معرفی سرویس گفتار به متن (Cloud Speech-to-text) تغییرات بزرگ و مهمی را در آن اعمال کرده است. گوگل نخستین بار در سال 2016 میلادی بود که از سرویس API Cloud Speech رونمایی کرد و اکنون بیش از یک سال است که به شکل عمومی در اختیار مردم قرار گرفته است. این سرویس که تا پیش از این به نام Cloud Speech API از آن نام برده می‌شد، اکنون به‌اندازه‌ای

پرمخاطب شده که به گفته گوگل هر شش ماه یکبار بر تعداد کاربران آن افزوده می‌شود. Cloud Speech-to-Text اکنون از یک مکانیزم جدید تشخیص متادیتاها برای نگ کردن و گروه‌بندی محتوا، اضافه کردن نشانه‌گذاری و مجموعه‌ای از مدل‌های از پیش‌ساخته شده برای بهبود دقت تبدیل پشتیبانی می‌کند. گوگل در وبلاگ این شرکت متعهد شده که سرویس یادشده در 99 درصد موارد در دسترس مشتریان قرار خواهد داشت. قابلیت‌ها و تغییراتی که گوگل به سرویس تبدیل گفتار به متن خود اضافه کرده باعث شده تا کیفیت تبدیل محتوای متنی به گفتار و بالعکس بهتر از قبل شود. به‌روزرسانی اعمال‌شده از آن جهت حائز اهمیت است که به کسب‌وکارها اجازه می‌دهد واسطه‌های برنامه‌نویسی مرتبط با این فناوری را بهتر از گذشته به کار گرفته و محتوایی کاملاً غنی و قدرتمند در اختیار مشتریان خود قرار دهند. (شکل 1)



به‌روزرسانی سرویس Speech-to-Text ضمن آن‌که قادر است متادیتاها را تشخیص دهد، به ویژگی نشانه‌گذاری خودکار نیز تجهیز شده است. گوگل در یادداشتی که در وبلاگ این شرکت در ماه مارس منتشر کرد، اعلام کرد از این پس سرویس Cloud Text-to-Speech به توسعه‌دهندگان اجازه می‌دهد سامانه‌های پاسخ‌گوی صوتی قدرتمندی را برای مراکز تماس مشتریان ایجاد کرده، به دستگاه‌های اینترنت اشیا که به شکل دوطرفه با کاربران در ارتباط هستند قابلیت پاسخ‌گویی داده و به شکل خودکار محتوای متنی مقاله‌ها را به قالب‌های صوتی همچون کتاب‌های صوتی تبدیل کنند. این سرویس به توسعه‌دهندگان اجازه می‌دهد، از میان 32 صدای مختلف گزینه موردنظر خود را انتخاب کنند. گوگل به توسعه‌دهندگان اجازه داده از برجسب‌های نشانه‌گذاری متن به گفتار موسوم به SSML برای اضافه کردن مکث، دستورالعمل‌های تلفظ و تاریخ به گفتار استفاده کنند. سیاست اتخاذشده از سوی گوگل نشان می‌دهد، این شرکت بیش از آنچه تصور می‌شد علاقه دارد تا ابزارهای هوش مصنوعی ساخت این شرکت را در قالب سرویس‌هایی در اختیار سازمان‌ها قرار دهد.

گوگل می‌گوید: «به‌روزرسانی‌های اخیر درجه‌های تازه‌ای از فرصت‌ها را پیش روی شرکت‌هایی قرار خواهد داد که در نظر دارند به شکل متفاوتی با کاربران خود ارتباط برقرار کنند. سرویس‌های جالب‌توجهی در اختیار آن‌ها قرار داده و بازخوردها را به شکل متفاوتی از مشتریان دریافت کرده تا در نهایت کیفیت سرویس‌های خود را بهبود بخشند. با به‌روزرسانی speech To Text کاربران به جدیدترین پژوهش انجام‌شده از سوی متخصصان گوگل در ارتباط با یادگیری ماشین از طریق به‌کارگیری توابع RESTful API دسترسی خواهند داشت.» این به‌روزرسانی‌ها با بهتر کردن

عملکرد مدل‌های تشخیص گفتار اجازه می‌دهند تا سرویس‌هایی همچون استخراج/تبدیل صوت از یک ویدیو و تبدیل محتوای صوتی یک تماس تلفنی به متن بهتر از گذشته انجام شود. مشتریان می‌توانند مدلی را که احساس می‌کنند برای کسب و کارها آن‌ها بهتر بوده و به‌خوبی می‌تواند به نیازهای آن‌ها پاسخ دهد، انتخاب کنند. گوگل همچنین مدل enhanced phone_call را طراحی کرده است. مدل یادشده از داده‌های مربوط به مشتریانی که به شکل داوطلبانه تصمیم گرفته‌اند، داده‌های خود را با Cloud Speech-to-text به اشتراک قرار دهند به‌منظور بهبود کیفیت سرویس گفتار به متن استفاده می‌کند. مشتریانی که تصمیم می‌گیرند در این برنامه مشارکت داشته باشند، دسترسی کامل به این سرویس و سایر مدل‌های پیشرفته‌ای را خواهند داشت که بر مبنای داده‌های سایر مشتریان کار می‌کند. گوگل می‌گوید: «مدل enhanced phone_call در مقایسه با مدل basic phone_call به میزان 54 درصد کمتر دچار خطا و اشتباه می‌شود.» گوگل همچنین مدل ویدیویی ارائه‌شده از سوی این شرکت را نیز مورد بازبینی قرار داده است. این بازبینی به‌منظور بهینه‌سازی فرآیند پردازش صوتی از ویدیوها و/یا صدا از طریق چند اسپیکر اعمال شده است. مدل ویدیویی از همان الگوی یادگیری ماشین استفاده می‌کند که یوتیوب از آن استفاده کرده و نزدیک به 64 درصد کمتر از مدل پیش‌فرض دچار اشتباه و خطا شده است. Cloud Speech-to-Text اکنون می‌تواند به لطف شبکه عمیق عصبی جدید LSTM به‌طور خودکار به نشانه‌گذاری گفتار پردازد. این مدل اکنون در وضعیت بتا قرار دارد، اما قادر است به شکل خودکار کاماها، علامت سوال و نقطه‌ها در یک متن را پیشنهاد دهد. گوگل به کاربران اجازه داده با برچسب‌گذاری متادیتاهای قابل‌تشخیص و گروه‌بندی رونوشت‌ها و ارائه بازخوردها به تیم گوگل اجازه دهند تا کیفیت این محصول را بهبود بخشد. به‌طور مثال، شما می‌توانید ویدیو یا صوت رونوشت شده خود را با برچسب‌هایی همچون (فرمان‌های صوتی برای یک برنامه خرید) voice commands for a shopping app یا (شوی تلویزیونی ورزش بسکتبال) basketball sports to shows نشان‌گذاری کرده و توصیف کنید تا گوگل به‌درستی بتواند این اطلاعات را برای Cloud Speech-to-Text جمع‌آوری کرده و در پروژه بعدی خود از آن‌ها استفاده کند. لازم به توضیح است که مدل ویدیویی و مدل enhanced phone_call به زبان انگلیسی در دست‌کاربران قرار داشته و به‌زودی زبان‌های دیگری نیز به آن اضافه خواهد شد. برای کسب اطلاعات بیشتر و آزمایش فناوری تبدیل گفتار به متن می‌توانید به آدرس <https://cloud.google.com/speech-to-text> مراجعه کنید. در بخش convert your speech to text right now گوگل اجازه می‌دهد از طریق میکروفون یا آپلود فایل این سرویس مبتنی بر یادگیری ماشین را آزمایش کنید. البته توجه داشته باشید در زمان انتخاب فایل صوتی شما با محدودیت اندازه فایل صوتی روبرو هستید. در زمان نگارش این مقاله فایل صوتی شما باید زمانی کمتر از یک دقیقه و حجمی کمتر از 50 مگابایت داشته باشد.

منبع:

نشانی منبع:

<https://www.shabakeh-mag.com/success-and-entrepreneurship/13395/%D8%B3%D8%B1%D9%88%DB%8C%D8%B3-%D8%AA%D8%A8%D8%AF%DB%8C%D9%84-%DA%AF%D9%81%D8%AA%D8%A7%D8%B1-%D8%A8%D9%87-%D9%85%D8%AA%D9%86-%DA%AF%D9%88%DA%AF%D9%84-%D9%86%D9%85%D9%88%D9%86%D9%87%E2%80%8C%D8%A7%DB%8C-%D9%85%D9%88%D9%81%D9%82-%D8%A7%D8%B2-%DB%8C%DA%A9-%DA%A9%D8%B3%D8%A8%E2%80%8C%D9%88%DA%A9%D8%A7%D8%B1-%D8%A7%D8%A8%D8%B1%DB%8C>