

رویایی که به حقیقت بدل شد؛ دستیارهای صوتی از گذشته تا به امروز - دیجیاتو

مریم موسوی | دوشنبه، ۱۱ آبان ۱۳۹۴

از دهه 1950 میلادی که تکنولوژی های محاسباتی باب شدند دانشمندان و مصرف کنندگان این فناوری ها رویای از بین بردن فاصله میان انسان و ماشین را در سر داشتند که به خاطر نبود توانایی کلامی در ماشین ها میانشان شکل گرفته بود.

همینکه ماشین ها نشان دادند در انجام وظایف محاسباتی قوی تر از انسان عمل می کنند، این مساله در ذهن پژوهشگران شکل گرفت که چرا چنین دستگاه هایی باید از نظر درک زبانی که سنگ بنای اصلی تفاوت میان بشر و حیوانات را می سازد، عقب تر از انسان باشند. این کاستی با توجه به اینکه اطفال از نخستین سال های عمر خود شروع به آموختن زبان می کنند اهمیت بیشتری یافت.

دانشمندان چندین دهه از عمر خود را صرف حل و فصل این چالش کردند با این همه تا همین اواخر پیشرفت های اندکی داشتند و نتوانستند درک زبان طبیعی انسان را به ماشین ها بیاموزند، چه برسد به اینکه مهارتی در سطح انسانی را از این لحاظ در آنها ایجاد نمایند.

نخستین پیشرفت عمده پژوهشگران در این زمینه به تشخیص گفتار و توانایی تبدیل امواج صوتی به متن مربوط می شد که در واقع همان معرف عبارات شفاهی بودند.

در ادامه این مطلب با [دیجیاتو](#) همراه باشید.

البته نباید از یاد برد که پیشرفت های علمی در زمینه تشخیص کلام توسط ماشین ها مدت ها قبل از ایجاد توانایی درک مفهوم و معنا به دست آمده بود. در دهه 90 میلادی، از توانایی تشخیص کلام در کال سنترهای سازمانی کل دنیا استفاده می شد و در اصل می توان گفت نخستین باری بود که تکنولوژی های گفتاری پای را از آزمایشگاه های تحقیقاتی بیرون می گذاشتند و به دنیای کسب و کار قدم می گذاشتند.

در آن زمان، توانایی های تشخیص کلام به حدی پیشرفت کرده بود که بشود از آنها در سیستم های تلفنی IVR (پاسخ صوتی تعاملی) منو محور در بخش فرماندهی و کنترل بهره گرفت، با این همه، این تکنولوژی ها نمی توانستند رویای دیرینه صحبت کردن روزمره با ماشین ها را که پیشتر در فیلم های علمی و تخیلی دیده بودیم به واقعیت بدل کنند و انسان بتواند آنچه در سر دارد را به آنها انتقال دهد.

سیستم های فرماندهی و کنترل که داده های ورودی از پیش تعیین شده ای داشتند و پاسخ هایشان هم برنامه ریزی شده بود به مانند حیوانات دست آموزی بودند که می شد توپی را برایشان پرتاب کرد تا برایتان بیاورندش.

در مقابل سیستمی متشکل از مجموعه عظیم واژگان با امکان درک طبیعی زبان شکلی انسانی تر دارد: منعطف تر است، مرتبا یاد می گیرد و می تواند به میلیون ها اظهار نظر و درخواستی که برای نخستین بار می شوند پاسخ دهد.

اینترفیس های مکالمه: حالا چرا؟



نسل نخست دستیارهای شخصی مجازی به خاطر ابداع سیستم های پیشرفته تشخیص کلام، سرعت بالای انتقال داده به شیوه وایرلس، انفجار محاسبات ابری و شکل گیری طیف جدیدی از کاربران وارد بازار شدند: همان استفاده کنندگان از تلفن های هوشمند که دائماً به اینترنت متصل هستند، زندگی پر مشغله ای دارند، اغلب در حرکتند و مشتاقند که از دست کیبوردهای مجازی خود که نرخ خطای بالایی هم دارند خلاص شوند.

این دستیارهای شخصی در ابتدای امر به شدت مورد توجه عموم قرار گرفتند و تیترا نخست رسانه

های خبری را به خود اختصاص دادند، با این همه، دیری نپایید که به خاطر بالا بودن سطح توقع کاربران، این فناوری از مرکز توجه خارج گردید.

حدوداً 5 سال بعد، موج دوم دستیارهای مجازی شخصی و اینترفیس های مکالمه ای به راه افتاد که هم به لحاظ هوش و هم از نظر فراگیری نسل نخست را پشت سر می گذاشت.

این موج تازه روی پیشرفت هایی سوار بود که در زمینه هوش مصنوعی، مجموعه غنی داده های کاربری و رشد دستگاه های فاقد کیبورد و نمایشگر شکل گرفته بود.

علاوه بر این، مدتی است که هر سیستم عاملی همراه با قابلیت های بی نظیر تشخیص کلام عرضه می گردد. گوگل، اپل، بیدو، مایکروسافت و آمازون این توانمندی را به صورت رایگان در اختیار کاربران قرار می دهند و در نتیجه امکان ورود نسل تازه ای از اپلیکیشن ها را به دستگاه های مورد استفاده کاربران فراهم می کنند.

موج جدید دستیارهای صوتی، عاقبت موفق شدند که به رویاهای دانشمندان و مصرف کنندگان جامعه عمل پیوشانند و معنا و مفهوم کلام انسان را دریابند. البته مشخصات و قابلیت های دستیارهای قدیمی هم به گونه ای بود که بشود حدس زد، در آینده این تکنولوژی ها به کدام سمت و سو حرکت خواهند کرد منتها آن دستیارها بر تصویری شکننده از توانایی کلامی استوار بودند: گاهی اوقات درخواست های کاربر در گروه هایی نادرست طبقه بندی شده بود و همین مساله باعث می شد که فرد پاسخ هایی از قبل برنامه ریزی شده را دریافت کند. در موارد دیگر زمانی که درخواست فرد فاقد کلیدواژه مورد نیاز بود، پاسخ مقتضی و مناسب توسط سیستم ارائه نمیشد.

همین مساله مشکلاتی را به وجود می آورد چراکه کلام انسانی واقعی سخت و محکم نیست: این زبان بسیار غنی است و میلیون ها مفهوم و پیکره بندی لغوی را در بر می گیرد.

موج جدید تکنولوژی های صوتی اما چنین چالشی را از میان بر می دارند و پس از تفسیر درخواست ها با هوش ذاتی خود به آنها پاسخ می دهند.

آینده چگونه خواهد بود؟



تکنولوژی های پیشرفته صوتی به زودی فراگیر خواهند شد و اینترفیس آنها نیز به طبیعی ترین و هوشمندترین شیوه ممکن با زندگی انسان یکپارچه سازی می شود.

در خانه های هوشمند آینده، صدا اینترفیس اصلی خواهد بود و به ابزاری طبیعی برای برقراری ارتباط با سیستم های هشدار، لامپ ها، لوازم آشپزخانه، سیستم های صوتی و بسیاری موارد دیگر بدل می شود.

در چنین آینده ای، روز به روز بر تعداد ماشین های هوشمند افزوده می شود و سیستم های مبتنی بر صدا نیز به درون امکانات سرگرمی و مسیریابی خودروها راه می یابند و دیگر نیازی به اینکه راننده یا مسافر از طریق دست هایش با ماشین تعامل داشته باشد نخواهد بود.

سیستم های صوتی و تصویری ماشین ها برای پیدا کردن محتوی دلخواه کاربر به صدای او متکی خواهند شد (چنین چیزی را همین الان هم می شود در محصولات نظیر اپل تی وی جدید هم مشاهده کرد که در آن، هر اپلیکیشنی که کارکرد اصلی اش از سیری پشتیبانی نکند، مردود خواهد بود).

پوشیدنی های فاقد صفحه نمایش یا دارای نمایشگرهای کوچک همچنان روند صعودی افزایش محبوبیت خود را طی خواهند کرد و هر چه بر میزان توقع کاربران از آنها افزوده می شود، این دستگاه ها نیز به ابزارهای عملیاتی بهتری مجهز می شوند.

دستگاه های کنترل شونده از طریق صدا، همچنین به درون آن دسته از محیط های کاری نفوذ می

کنند که نیازمند حرکت پذیری بدون دست هستند و از آن جمله می توان به بیمارستان، ها، انبارها، آزمایشگاه ها و تاسیسات تولیدی اشاره نمود.

براساس اعلام مرکز تحقیقاتی comScore تا سال 2020 میلادی، ماهانه 200 میلیارد سرچ صوتی انجام خواهد گرفت که این مساله کسب و کاری بالغ بر 50 میلیارد دلار را حول مقوله سرچ صوتی ایجاد می کند.

هرچه تقاضا برای این تکنولوژی بیشتر شده و تنوع بالاتری می یابد، درک زبان انسان گونه، نیز اهمیت بیشتری پیدا می کند. زمانی که حرف از تکنولوژی های درک زبان طبیعی به میان می آید، کاربران هم بیش از هر وقت دیگری سخت گیر می شوند و تنها دو یا سه خطا کافیسیت تا اعتمادشان را نسبت به این سیستم ها از دست بدهند.

چالش دیگری که برای ارائه تکنولوژی های پیشرفته درک زبان های طبیعی پیش روی دانشمندان وجود دارد وسعت و فراگیری این سیستم هاست. در هر حوزه از دانش، یک سیستم علاوه بر اینکه باید قابلیت تشخیص واژگان جدید و تخصصی را داشته باشد بلکه باید توانایی درک و دریافت مفهوم آن عبارت در یک جمله را هم پیدا کنند.

نکته دیگر اینکه سیستم های تشخیص زبان طبیعی نمی توانند به طور مثال صرفا کارشناسی قهار در کفش فروشی ها باشند بلکه باید یک کارشناس خبره برای کفش فروشی X باشند و در عین حالا تخصص و آموزش های لازم در مورد گروه های مختلف محصولات را نیز داشته باشند.

به لطف پیشرفت های عظیم به دست آمده در حوزه هوش مصنوعی، تا حل شدن این چالش های بیشتر بدون راهکار، زمان زیادی باقی نمانده و تنها کافیسیت که از طریق یادگیری ماشینی به این سیستم ها آموزش های لازم داده شود تا به تدریج طیف متنوعی از شرایط و پیش نیازها را درک کنند و فراگیرند.

نکته دیگر اینکه دستیارهای هوشمند مجازی مرتبا در حال بهتر شدن هستند و خودروها، اپلیکیشن های موبایل و لوازم خانگی مدتهاست که میزبان آنها هستند.

شرکت های بزرگ نظیر گوگل، اپل، آمازون، مایکروسافت و Baidu همگی بیشتر به این عرصه قدم گذاشته اند و کمپانی های فعال درحوزه هوش مصنوعی نیز تدریجا در حال ورود به این حوزه هستند و راهکارهای خود را در اختیار جمعیت رو به رشد کسب و کارهایی می گذارند که برای باقی ماندن در بازار رقابت نیازمند اینترفیس های کلامی هستند.

دست آخر باید بگوییم که تعامل کلامی با ماشین ها آن هم به زبان طبیعی دیگر رویایی متعلق به فیلم های علمی و تخیلی نیست و مدتهاست که رنگ واقعیت به خود گرفته است.

[دیجیاتو](#)