



هوش مصنوعی باید تصمیمات خود را توضیح بدهد!

برخی از شرکت‌های فعال در حوزه هوش مصنوعی برای پیشگیری از تعصب و غرض‌ورزی احتمالی این فناوری، باور دارند که هوش مصنوعی باید برای تصمیمات خود، دلایل قانع‌کننده و به دور از تبعیض داشته باشد.

برخی از شرکت‌های فعال در حوزه هوش مصنوعی برای پیشگیری از تعصب و غرض‌ورزی احتمالی این فناوری، باور دارند که هوش مصنوعی باید برای تصمیمات خود، دلایل قانع‌کننده و به دور از تبعیض داشته باشد.

به گزارش ایسنا و به نقل از فوربس، هوش مصنوعی هم مانند انسان، با غرض‌ورزی کار می‌کند. در حقیقت می‌توان گفت هر شخص یا چیزی که قابلیت تصمیم‌گیری داشته باشد، با غرض‌ورزی عمل می‌کند و هنگام انتخاب، به برخی از عوامل بیش از عوامل دیگر بها می‌دهد. با وجود این، همان‌گونه هوش مصنوعی در سال‌های اخیر به خاطر غرض‌ورزی و تبعیض علیه گروه‌های خاصی از جمله زنان و رنگین‌پوستان، بدنام شده، شرکت‌ها نیز به همان اندازه در غلبه بر تبعیض الگوریتم‌های آن، مهارت بیشتر به دست می‌آورند.

روشی که آنها غالباً برای این کار استفاده می‌کنند، با عنوان "هوش مصنوعی با قابلیت توضیح" (Explainable AI) شناخته می‌شود. در گذشته و حتی اکنون، عملکرد بیشتر اطلاعات مربوط به هوش مصنوعی، مانند یک جعبه سیاه ارائه می‌شده است. افرادی که کار رمزگذاری را انجام می‌دهند، شبکه‌های عصبی الگوریتم‌ها را طراحی کردند اما هنگامی که این اطلاعات منتشر شد، شبکه‌های عصبی بدون نظارت افرادی که آنها را برنامه‌ریزی کرده بودند، به کار خود ادامه دادند. در هر حال، شرکت‌ها در ابتدا متوجه این مشکل نشدند و هنگامی به آن پی بردند که بسیار دیر شده بود.

این مشکل هوش مصنوعی، موضوعی چالش‌برانگیز است. بیشتر استارت‌آپ‌ها و شرکت‌ها در حال حاضر، روش‌ها و پلتفرم‌هایی را براساس هوش مصنوعی قابل توضیح ارائه می‌دهند. یکی از جالب‌ترین این روش‌ها، روش ارائه شده توسط استارت‌آپ موسوم به "فیدلر لبز" (Fiddler Labs) است. این استارت‌آپ که در سانفرانسیسکو و توسط مهندسان پیشین فیسبوک و سامسونگ تاسیس شده است، نوعی موتور جستجوی مبتنی بر هوش مصنوعی را به شرکت‌ها پیشنهاد می‌کند که همه عوامل مرتبط با تصمیم‌گیری را نمایش می‌دهد. "آمیت پاکا" (Amit Paka)، از بنیان‌گذاران این استارت‌آپ گفت: نرم‌افزار ما موجب می‌شود که رفتار مدل‌های هوش مصنوعی، واضح و قابل درک باشند.

پاکا به عنوان نمونه توضیح داد که چگونه هوش مصنوعی قابل توضیح می‌تواند به مدل مبتنی بر هوش مصنوعی مورد استفاده بانک‌ها کمک کند.

وی افزود: ورودی‌های بسیاری وجود دارند که هنگام تصمیم‌گیری برای یک کاربرد خاص، مورد توجه قرار می‌گیرند. در یک فضای قدیمی بدون وجود فیدلر، گفتن چگونگی و دلیل تاثیر هر ورودی بر نتیجه، دشوار و یا تقریباً غیرممکن است اما شاید بانک‌ها با وجود هوش مصنوعی قابل توضیح بتوانند تاثیر هر ورودی را نسبت به خروجی مشخص کنند.

پاکا اضافه کرد: این قابلیت توضیح دادن، به ابداع‌کنندگان هوش مصنوعی، کاربران اقتصادی، تنظیم‌کنندگان و کاربران نهایی امکان می‌دهد تا درک بهتری در مورد دلیل پیش‌بینی‌های خاص داشته باشند. این موضوع، در مورد تعصب و اخلاقیات مربوط به هوش مصنوعی، از اهمیت بالایی برخوردار است زیرا به شرکت‌ها امکان می‌دهد تا تبعیض‌های احتمالی علیه گروه‌های خاص را شناسایی کنند. به علاوه، چنین قابلیت‌هایی می‌تواند به شرکت‌ها در اصلاح مدل‌های هوش مصنوعی پیش از به کار بردن آنها در مقیاس بزرگ کمک کند.

پاکا ادامه داد: تعصب نژادی هوش مصنوعی در الگوریتم‌های مربوط به سلامت و تبعیض آن در تصمیم‌گیری و قضاوت، فقط چند نمونه از مشکلات مربوط به غرض‌ورزی هوش مصنوعی هستند. الگوریتم‌های پیچیده هوش مصنوعی در حال حاضر مانند جعبه سیاه عمل می‌کنند و حتی با داشتن عملکرد خوب، فعالیت داخلی آنها ناشناخته و غیرقابل توضیح دادن است.

یکی از دلایل اهمیت هوش مصنوعی قابل توضیح در غلبه بر غرض‌ورزی الگوریتم‌ها، این است که شاید جنسیت، نژاد و مقوله‌های دیگری از این دست، صریحاً رمزگذاری نشده‌اند. در چنین مواردی، هوش مصنوعی قابل توضیح لازم است تا

به شرکت‌ها در برداشتن تبعیض‌های عمیق داده‌هایی که این الگوریتم‌های پیچیده را تغذیه می‌کنند، کمک کند.

با توجه به این دلایل، ابداع مدل‌های هوش مصنوعی که قابلیت بیشتری برای توضیح داشته باشند، کلید اصلاح عواملی است که به بروز غرض‌ورزی می‌انجامد. این مدل‌ها همچنین می‌توانند رعایت مقررات توسط سیستم‌های هوش مصنوعی را نیز تضمین کنند. تنظیمات بیشتر می‌تواند به اطمینان بیشتر از عملکرد عادلانه هوش مصنوعی و قابل توضیح بودن آن کمک کند.

نکته مثبت این است که شرکت‌های دیگری به غیر از فیدلر لیز نیز راه‌حل‌ها و پلتفرم‌های مبتنی بر هوش مصنوعی را ارائه می‌دهند که قابل مقایسه و توضیح‌داده هستند. برای مثال، یک استارت‌آپ فعال در حوزه هوش مصنوعی موسوم به "کیندی" (Kyndi) که در این زمینه فعالیت می‌کند، در ماه ژوئیه سال جاری موفق شد بودجه‌ای به مبلغ ۲۰ میلیون دلار به دست بیاورد و در حال حاضر نیز از پلتفرم‌های خود برای بررسی دلیل هر تصمیم‌گیری استفاده می‌کند.

شرکت دیگری که در حوزه هوش مصنوعی قابل توضیح فعالیت دارد، شرکت آمریکایی "زد ادونسد کامپیوتینگ" (Z Advanced Computing) است. این شرکت در ماه اوت اعلام کرد که بودجه‌ای را از نیروی هوایی آمریکا دریافت کرده است تا فناوری تشخیص چهره سه بعدی را بر اساس هوش مصنوعی قابل توضیح ارائه دهد.

شرکت "ویانی سیستمز" (Vianai Systems) که در ماه سپتامبر و توسط مدیرعامل پیشین شرکت "اینفوسیس" (Infosys) بنیان گذاشته شد نیز هوش مصنوعی قابل توضیح‌داده را برای گروهی از سازمان‌ها و در طیف وسیعی از نقش‌ها به کار می‌گیرد.

شرکت‌های دیگری نیز در حوزه هوش مصنوعی قابل توضیح فعالیت دارند که نرم‌افزارها و راه‌حل‌های آنها، بهبود قابل توجهی را در نحوه عملکرد هوش مصنوعی ایجاد می‌کند.

با وجود همه فوایدی که چنین برنامه‌هایی به همراه دارند، نکات منفی استفاده از آنها غیرقابل انکار هستند. این امکان وجود دارد که هوش مصنوعی نهایتاً اثر متضادی بر شرکت‌هایی مانند فیدلر لیز و کیندی داشته باشد.

از آنجا که دولت‌ها یا شرکت‌ها می‌توانند عوامل دقیق یک الگوریتم را برای تصمیم‌گیری به کار ببرند، ممکن است برخی از سازمان‌های غیراخلاقی، از موتورهای هوش مصنوعی قابل توضیح استفاده کنند تا الگوریتم‌های آنها، غرض‌ورزی بیشتری داشته باشد.