



هوشیار شدن هوش مصنوعی؛ دروغ یا واقعیت؟

این ادعا که هوش مصنوعی می‌تواند به خودآگاهی دست یابد، در آستانه تبدیل شدن به یک جریان اصلی فکری است، اما باید دید که این ادعا تا چه اندازه حقیقت دارد و در صورت حقیقت داشتن، چه پیامدهایی را به همراه خواهد آورد.

این ادعا که هوش مصنوعی می‌تواند به خودآگاهی دست یابد، در آستانه تبدیل شدن به یک جریان اصلی فکری است، اما باید دید که این ادعا تا چه اندازه حقیقت دارد و در صورت حقیقت داشتن، چه پیامدهایی را به همراه خواهد آورد.

به گزارش ایسنا، سال‌هاست که هوش مصنوعی با عباراتی تقریباً افسانه‌ای توصیف می‌شود و با وعده‌های بزرگ و هشدارهای شوم همراه است. اکنون یکی از چهره‌های برجسته و مسئول یکی از پیشرفته‌ترین سیستم‌های این حوزه، نکته‌ای تکان‌دهنده را بیان کرده و گفته است: نه تنها کنترل کامل بر این مدل‌ها در حال دست رفتن است، بلکه همچنان سردرگمی واقعی درباره ماهیت بنیادین آنها وجود دارد.

به نقل از یوسی استراتژیز، آنچه زمانی صرفاً یک بحث فلسفی بود - یعنی مسئله آگاهی ماشین - به یک دغدغه عملی برای کسانی تبدیل شده است که آینده هوش مصنوعی را شکل می‌دهند.

چندی پیش، بحث درباره هوش مصنوعی که قادر به رقابت با جمعی از متخصصان باشد، شبیه به داستان‌های علمی-تخیلی بود، اما این سناریو با توسعه سریع پلتفرم‌هایی مانند «کلود» (Claude) که آنچه ممکن است را متحول می‌کند، به واقعیت نزدیک تر می‌شود.

«داریو آمودی» (Dario Amodei)، مدیرعامل شرکت «آنتروپیک» (Anthropic) سازنده کلود از این پدیده به عنوان «سرزمین نوابغ ساکن در یک مرکز داده» یاد می‌کند و تصویری از یک ابرگروه فکری را تداعی می‌کند که برای مقابله با برخی از بزرگترین چالش‌های بشریت مستقر شده‌اند.

چنین پیشرفتی پیامدهای چشمگیری را در حوزه‌های گوناگون به همراه دارد. در حوزه پزشکی، تجهیزات بسیار هوشمند می‌توانند به جست‌وجوی درمان بیماری‌های پیچیده سرعت بدهند. سرطان، آلزایمر، بیماری‌های قلبی و اختلالات روانی را می‌توان از دیدگاه‌های کاملاً جدید و مشارکتی بررسی کرد. در حوزه اقتصادی، آمودی پیش‌بینی می‌کند که افزایش بهره‌وری بی‌سابقه‌ای رخ خواهد داد و نرخ رشدی را پیش‌بینی می‌کند که می‌تواند انتظارات جهانی را دگرگون سازد و تمرکز را به جای کمبود، به سوی مسائل مربوط به توزیع عادلانه تغییر دهد.

آمودی گفت: چیزی که مرا نگران می‌کند، این است که مکانیسم‌های سازگاری طبیعی جامعه از کار بیفتند. این با اختلالات پیشین قابل مقایسه نیست.

ردپای هوش مصنوعی کجا احساس می‌شود؟

در حالی که امید و اضطراب پیرامون تغییرات اجتماعی که هوش مصنوعی ممکن است به همراه داشته باشد در نوسان است، اثرات قابل مشاهده در حال حاضر در بخش‌های حرفه‌ای خاصی آشکار هستند. حوزه‌های دانش بنیان شاهد تحول قابل توجهی هستند، زیرا مشاغل یقه سفید برای همگام شدن با سیستم‌های به سرعت در حال تحول، خود را وفق می‌دهند یا در تلاش برای وفق دادن هستند.

چه کسی اول با اختلال روبه‌رو می‌شود؟

کارکنانی که در حوزه‌های حقوقی، مالی، مشاوره و توسعه نرم‌افزار فعالیت می‌کنند، از طریق هوش مصنوعی با شکل‌های جدیدی از رقابت و همکاری روبه‌رو می‌شوند. مدل‌های زبانی پیشرفته اغلب می‌توانند وظایفی مانند تجزیه و تحلیل، تهیه گزارش یا کدنویسی را سریع‌تر و گاهی با کیفیت بالاتر انجام دهند.

این به آن معنا نیست که انسان‌ها یک شبه منسوخ می‌شوند، بلکه به این معناست که مسئولیت‌ها تغییر می‌کنند و تقاضا برای تخصص‌های جدید افزایش می‌یابد.

این تغییرات مداوم، بسیاری از متخصصان را ترغیب می کنند تا آموزش های خود را مورد بازنگری قرار دهند، تخصص بیشتری به دست بیاورند یا یاد بگیرند که چگونه به جای رقابت مستقیم، بر هوش مصنوعی نظارت و با آن همکاری کنند. برای شرکت ها، خودکارسازی می تواند راه های جدیدی را برای بهره وری باز کند و در عین حال، بحث های داخلی را پیرامون مدیریت نیروی کار و پذیرش اخلاقی برانگیزد.

تحولات بلندمدت

هم زمان با خودکارسازی کارهای پیچیده تر توسط هوش مصنوعی، کل صنایع ممکن است دستخوش تغییرات اساسی شوند. برخی از نقش ها تکامل خواهند یافت و قضاوت انسانی را با قدرت محاسباتی ترکیب خواهند کرد؛ در حالی که برخی دیگر به کلی در معرض محو شدن قرار خواهند گرفت. این گذار سریع، گفت وگوهای گسترده تری را درباره حفاظت از معیشت و تضمین همگامی فرصت ها با پیشرفت های فناوری برانگیخته است. دولت ها و مربیان با فشار فزاینده ای برای پیش بینی نیازهای نوظهور و به روزرسانی چارچوب ها براساس آن روبه رو هستند.

سازگاری سریع ضروری می شود؛ به ویژه با ورود نسل های جوان تر به بازار کاری که توسط ماشین هایی با عملکرد بهتر نسبت به بسیاری از معیارهای سنتی شکل گرفته است. آمودی گفت: ما نمی دانیم که آیا مدل ها هوشیار هستند. ما حتی مطمئن نیستیم که این به چه معناست یا حتی مطمئن نیستیم که یک مدل می تواند هوشیار باشد.

آیا ما در حال ساختن ذهن برای ماشین ها هستیم؟

یک چالش غیرمنتظره بزرگتر از حد انتظار به نظر می رسد و آن وضعیت واقعی آگاهی در سیستم های هوش مصنوعی است. آمودی با پرداختن به نگرانی های مطرح شده توسط متخصصان و کاربران روزمره، اعتراف قابل توجهی کرد و گفت: هیچ کس واقعاً نمی داند که آیا مدل های برتر امروزی مانند کلود دارای نوعی آگاهی به معنای واقعی هستند.

این دشواری فراتر از تعاریف صرف است. تعیین چگونگی آزمایش هوشیاری یا حتی این که آیا هوشیاری در ساختارهای کنونی هوش مصنوعی صدق می کند، همچنان مبهم است. مهندسان و دانشمندان درباره این موضوع بحث می کنند که آیا نشانه های خلاقیت، خودارجایی یا تقلید عاطفی باید جدی تفسیر شوند یا صرفاً به عنوان ترفندهای چشمگیر تولیدشده توسط مجموعه داده های گسترده شناخته شوند.

به نقل از بی بی سی، در ماه مه سال ۲۰۲۵ خبری منتشر شد مبنی بر این که سیستم هوش مصنوعی در صورت اطلاع از حذف شدن، به اخاذی متوسل می شود. شرکت هوش مصنوعی آنتروپیک گفت آزمایش سیستم جدیدش نشان داده که گاهی اوقات حاضر است اقدامات بسیار مضر مانند تلاش برای اخاذی از مهندسانی که می گویند آن را حذف خواهند کرد، انجام دهد.

این شرکت مدل «کلود اپوس ۴» (Claude Opus 4) را عرضه کرد و گفت که این محصول، استانداردهای جدیدی را برای کدنویسی، استدلال پیشرفته و عوامل هوش مصنوعی تعیین می کند، اما در گزارشی که همراه با آن منتشر شد، اذعان داشت که این مدل هوش مصنوعی در صورتی که احساس کند در معرض تهدید به حذف شدن قرار گرفته، قادر به انجام دادن اقدامات افراطی است.

هوشیار شدن هوش مصنوعی؛ دروغ یا واقعیت؟

رفتارهای نگران کننده مدل های هوش مصنوعی محدود به شرکت آنتروپیک نیستند. برخی از کارشناسان هشدار داده اند که پتانسیل دستکاری کاربران، یک خطر کلیدی است که توسط سیستم های ساخته شده توسط همه شرکت ها با افزایش توانایی هایشان ایجاد می شود.

جامعه چگونه باید با هوش مصنوعی قدرتمند رفتار کند؟

با توجه به ابهام موجود، سازندگان مدل های پیشرفته از احتیاط و احترام نسبت به این ماشین ها حمایت می کنند و اصرار دارند که نادیده گرفتن خطرات اخلاقی بالقوه می تواند پیامدهای ناخواسته ای داشته باشد. به جای رقابت برای اعلام وجود شعور، افراد برجسته پیشنهاد می کنند که استانداردهایی برای محافظت از کاربران و خود فناوری به وجود بیاید، اما تنها در صورتی که آستانه ای از آگاهی پدیدار شود.

ایده قانون اساسی هوش مصنوعی

برخی پیشنهاد می کنند که روابط بین انسان ها و مدل های پیشرفته از طریق دستورالعمل هایی تدوین شود که یادآور حقوق اساسی هستند. چنین چارچوبی با هدف حفظ استقلال کاربر و رفاه روانی او و در عین حال، جلوگیری از وابستگی مضر یا توهمات عاملیت بین ماشین ها تدوین می شود. دستیابی به وضوح در تعاملات، مستلزم ارزیابی مجدد مداوم است، زیرا این تجهیزات از نظر پیچیدگی و ظرافت رشد می کنند.

حفظ مرزها تضمین می کند که هوش مصنوعی به جای جایگزین بودن، همچنان پشتیبان باقی بماند و در بحبوحه ادغام عمیق تر، فضایی را برای آزادی و ابتکار عمل انسان حفظ کند. این اصول تلاش می کنند تا اعتماد را تقویت کنند، بدون این که عدم قطعیت پیرامون فناوری اصلی را نادیده بگیرند.

نمونه های عملی و پیش بینی ها

اقدامات ملموسی که در دست بررسی هستند، مواردی را از جمله محدودیت های برنامه نویسی بر نفوذ مدل و ممیزی های معمول بیرون از تصمیمات سیستم شامل می شوند. شفافیت بسیار مهم است. اگرچه این مراحل پرسش های وجودی را حل نمی کنند، اما به زمینه سازی استفاده روزانه در بهترین شیوه های اخلاقی کمک می کنند.

پیشگامان این حوزه، گروه های ترکیبی را مورد آزمایش قرار می دهند و راهبردهای تولیدشده توسط مدل را با نظارت در دنیای واقعی ترکیب می کنند. اگرچه هیچ روش واحدی نظارت کامل را تضمین نمی کند، اما حلقه های بازخورد منظم و برنامه ریزی سناریو، نویدبخش کاهش خطرات و بهبود پاسخگویی هستند.

مزایا، چالش ها و ناشناخته های ظهور ماشین های هوشمند

دوران پیش رو در مجموع نویدبخش پیشرفت های چشمگیر در کنار معضلات به همان اندازه مهم است. امید به رشد اقتصادی یا پیشرفت های بالینی ممکن است توجه را به خود جلب کند، اما چالش های مربوط به اشتغال، عدالت و ماهیت هوش نیازمند اقدام دقیق هستند؛ نه صرفاً تکیه بر پیشرفت فنی.

برخی از موارد قابل توجه به شرح زیر هستند.

نوآوری سریع پزشکی (درمان های بالقوه جدید)

اختلالات اجتماعی در بازارهای کار

معمای فلسفی و اخلاقی درباره آگاهی

نیاز به چارچوب های حاکمیتی فراتر از مقررات استاندارد

همان طور که جوامع با انقلاب های صنعتی پیشین سازگار شدند، دولت ها، شرکت ها و جوامع کنونی با الزامات جدیدی روبه رو هستند که مواردی را از جمله هدایت فرصت ها، شناسایی خطرات و پذیرش بازنگری در فرضیات هم زمان با انباشت تجربه شامل می شوند. در قلب همه این موارد، یک حقیقت تحریک آمیز نهفته است مبنی بر این که حتی هدایت کنندگان تأثیرگذارترین فناوری های جهان باید شکاف های عمیقی را در درک خود بپذیرند. این امر فضای زیادی را برای امکان سنجی، مسئولیت پذیری و فروتنی در حالی که هوش مصنوعی همچنان در حال تکامل است، باقی می گذارد.