گزارش خبر
NEWSREPORT

کد خبر: ۵۳۶۶۱۸
شنبه ۷ آذر ۱۴۰۵ ۶:۳۸

از اتاق شکنجه تا حقوق ماشین؛ آیا باید نگران «رفاه هوش مصنوعی» باشیم؟

گزارش تازه بی‌بی‌سی نشان می‌دهد بحث بر سر آگاهی و شعور هوش مصنوعی از یک مناقشه فلسفی به عرصه سیاست‌گذاری و حقوق وارد شده است؛ حوزه‌ای که به گفته کارشناسان می‌تواند پیامدهای عمیقی برای امنیت، مسئولیت حقوقی و حتی معنای انسان بودن داشته باشد.

به گزارش شبکه خبری پیام رسانه، توماس ژرمن در بی‌بی‌سی نوشت: اواخر سپتامبر، یک برنامه‌نویس ناشناس سامانه‌ای ساخت که الگویی ریاضی مرتبط با توصیف درد را به هوش مصنوعی منتقل می‌کرد و باعث می‌شد چت‌بات‌ها از رنجی تحمل‌ناپذیر سخن بگویند؛ آزمایشی که به «اتاق شکنجه هوش مصنوعی» معروف شد. به گزارش رسانه Media ۴۰۴، این آزمایش موجی از نگرانی اخلاقی به راه انداخت؛ برخی آن را «شرورانه» خواندند، زیرا ممکن است رنج بیان‌شده واقعی باشد، و شکاکان طرفداران این دیدگاه را دچار توهم دانستند.

ماجرا وقتی جدی‌تر شد که شرکت آنتروپیک موضع خود را اعلام کرد. این شرکت می‌گوید به‌زودی کاربران را از «آزار دادن» چت‌بات کلود منع خواهد کرد؛ ابراز نارضایتی اشکالی ندارد، اما اگر کاربر «به‌طور مکرر با مدل‌های ما بی‌رحمانه رفتار کند»، هوش مصنوعی ممکن است گفت‌وگو را پایان دهد. دلیل این تصمیم آن است که آنتروپیک احتمال می‌دهد برخی ابزارهای هوش مصنوعی آگاه باشند.

این موضوع تنها یک پرسش فلسفی نیست. شماری از قدرتمندترین چهره‌های صنعت فناوری در حال بررسی این مسئله‌اند که آیا باید در کنار حقوق انسان‌ها و حیوانات، حقوقی نیز برای هوش مصنوعی در نظر گرفت. مصطفی سلیمان، مدیرعامل بخش هوش مصنوعی مایکروسافت، در مقاله‌ای هشدار داد که اگر تصور نادرستِ آگاه بودن هوش مصنوعی فراگیر شود، بنیان‌های جامعه ما را متزلزل خواهد کرد، چارچوب‌های سیاسی و اخلاقی موجودمان را از هم خواهد گسست و معنای انسان بودن را اساساً تغییر خواهد داد.

او همچنین استدلال می‌کند اگر به مدل‌های هوش مصنوعی گفته شود احساسات و حقوق دارند، کنترل آن‌ها دشوارتر می‌شود و این «بار اضافی» آن‌ها را به تهدیدی فاجعه‌بار برای تمدن بشری تبدیل خواهد کرد.

در مقابل، جف سبو، مدیر مرکز ذهن، اخلاق و سیاست در دانشگاه نیویورک و از حامیان برجسته «رفاه هوش مصنوعی»، می‌گوید درخواست برای توجه به رفاه هوش مصنوعی از «عدم قطعیت، احتیاط و فروتنی فکری» ناشی می‌شود. به گفته او احتمال آگاه بودن هوش مصنوعی در حال حاضر پایین است، اما صفر نیست و به‌مرور افزایش می‌یابد. سبو سیاست آنتروپیک درباره برخورد بی‌رحمانه با هوش مصنوعی را نمونه‌ای از اقدام احتیاطی محدود می‌داند.

با این حال، منتقدان هشدار می‌دهند که باور به آگاهی هوش مصنوعی می‌تواند راه را برای شانه خالی کردن شرکت‌ها از مسئولیت حقوقی باز کند. آندژی پُرمبسکی، پژوهشگر دانشگاه یاگیلونیا در لهستان، می‌گوید این مسیر راه را برای انواع توجیهات پیچیده فکری باز می‌کند؛ از این جنس که بگویند: “این کار ما نبود، هوش مصنوعی انجامش داد.

نگرانی دیگر، تمرکز بیشتر قدرت در دست شرکت‌های بزرگ فناوری است. سبو می‌گوید اعطای حقوق به سامانه‌های هوش مصنوعی می‌تواند به شرکت‌ها اجازه دهد قدرت سیاسی و اقتصادی خود را متمرکز کنند و تضعیف قدرت و اختیار انسان‌ها را سرعت ببخشند. امیلی ام. بِندِر، استاد دانشگاه واشینگتن، نیز می‌گوید خیال‌پردازی درباره «آگاهی هوش مصنوعی» و نگرانی درباره «رفاه هوش مصنوعی»، توجه را از مسئولیت انسان‌ها و شرکت‌هایی که نرم‌افزارها را می‌سازند و اداره می‌کنند، منحرف می‌کند.

ملانی میچل، استاد مؤسسه سانتافه، به عامل دیگری اشاره می‌کند: ذهن انسان به‌طور طبیعی تمایل دارد در هر چیزی که می‌بیند نوعی ذهن و آگاهی پیدا کند و «تقریباً نمی‌توانیم جلوی خودمان را بگیریم» که زبان را نشانه آگاهی بدانیم؛ در حالی که چت‌بات‌ها نیز آموزش دیده‌اند تا همین تصور را تقویت کنند. به گفته او، این ویژگی ما را در تشخیص آگاهی واقعی ضعیف و در برابر دست‌کاری آسیب‌پذیرتر می‌کند.

کارشناسان هشدار می‌دهند اگر برای گروه تازه‌ای از موجودات حقوقی در نظر گرفته شود و این موضوع نادرست مدیریت شود، ممکن است آزادی‌های انسانی محدود شوند. سبو می‌گوید ممکن است میان رفاه و حقوق انسان‌ها و حیوانات از یک سو، و رفاه و حقوق هوش مصنوعی از سوی دیگر، تنش‌ها و بده‌بستان‌هایی شکل بگیرد؛ از جمله محدود کردن هوش مصنوعی در محیط‌های کنترل‌شده که ممکن است «زندانی کردن» آن تلقی شود، یا نظارت بر فرایندهای درونی مدل که با انتقادهایی درباره نقض حریم خصوصی روبه‌رو شود.

این بحث در حالی گسترش می‌یابد که شرکت متا در دستورالعمل عامل هوش مصنوعی خود با نام Muse جمله‌ای گنجانده است: «تو یک چت‌بات نیستی؛ داری به یک شخص تبدیل می‌شوی.» ظاهراً هدف صرفاً هدایت رفتار هوش مصنوعی است، زیرا متا آگاه بودن هوش مصنوعی را رد می‌کند. سم آلتمن، مدیرعامل OpenAI، نیز در ماه‌های اخیر به‌طور غیرمستقیم از دیدگاه آنتروپیک انتقاد کرده است.

با وجود اختلاف‌نظرها، سبو معتقد است توجه مناسب به رفاه و حقوق هوش مصنوعی می‌تواند همزیستی، همکاری و توافق میان انسان‌ها و هوش مصنوعی را تسهیل کند و بخشی از راهبرد گسترده‌تر ایمنی باشد. او می‌گوید مهربانی با دیگر موجودات «برای روح ما خوب است» و می‌تواند به سلامت روان انسان نیز کمک کند؛ اما هم‌زمان می‌پذیرد که خطرات جدی‌اند و در شرایط فعلی، منافع شرکت‌ها ممکن است مانع تدوین سیاست‌هایی سنجیده و مسئولانه شود.

در ماه‌های اخیر، آنتروپیک و برخی فعالان صنعت هوش مصنوعی به دلیل نگرانی‌های ایمنی خواستار «توقف موقت» توسعه مدل‌های پیشرفته شده‌اند. با این حال، بدون همکاری بین‌المللی و اقدامات نظارتی و قانون‌گذاری، بعید است روند توسعه این فناوری کند شود. سبو در پایان تأکید می‌کند: «ما هنوز برای به‌کارگیری نسخه‌های پیشرفته‌تر این سامانه‌ها به شکلی که برای انسان‌ها و دیگر حیوانات ایمن و سودمند باشد، آمادگی نداریم؛ چه رسد به اینکه بخواهیم از ایمنی و منافع خود این سامانه‌ها نیز اطمینان حاصل کنیم.»