مدیران ارشد هوش مصنوعی DeepSeek به نیت در ده سال آینده

مدیران ارشد هوش مصنوعی DeepSeek به نیت در ده سال آینده از کاربران خواستند به آنها اعتماد کنند

□ مدیر ارشد ۱ ● هوش مصنوعی در ده سال آینده

پیشکار هفتگی

هانگ‌جو – خبرنگار ویژه پیشکار – ۱۸ مهر ۱۴۰۴: مدیران ارشد شرکت DeepSeek در نشستی خبری اعلام کردند که تا ده سال آینده هوش مصنوعی باید به ابزاری قابل اعتماد برای میلیاردها کاربر تبدیل شود. این شرکت که در سال گذشته با عرضه مدل‌های زبانی ارزان‌قیمت توجه جهانی را به خود جلب کرد، اکنون می‌گوید هدف بعدی آن ساختن سامانه‌هایی است که کاربران بتوانند به آن‌ها در تصمیم‌گیری‌های روزمره تکیه کنند.

مدیر ارشد شرکت در سخنرانی افتتاحیه گفت: «ما نمی‌خواهیم کاربران ما فقط پرسش بپرسند و پاسخ بگیرند. می‌خواهیم بدانند چه زمانی می‌توان به پاسخ ما اعتماد کرد و چه زمانی باید تردید کرد.» به گفته او، تا پایان سال ۲۰۲۷ نسخه‌ای از مدل DeepSeek طراحی شده است که عدم قطعیت خود را نشان می‌دهد و به کاربر می‌گوید چرا چیزی را درست یا نادرست می‌داند.

مدیران ارشد همچنین اعلام کردند که داده‌های کاربران ایران و چین در سرورهای جداگانه ذخیره می‌شود و هیچ کارمند خارجی به آن دسترسی ندارد. این موضوع پاسخی به نگرانی‌های مقامات اتحادیه اروپا است که از ماه ژوئیه گذشته درباره انتقال داده‌های شهروندان اروپایی به شرکت‌های چینی هشدار داده‌اند. به گفته ناظران، این سازوکار می‌تواند الگویی برای سایر شرکت‌های فناوری باشد که می‌خواهند بدون نقض مقررات وارد بازار شوند.

در بخش پرسش‌وپاسخ، یکی از مدیران ارشد پذیرفت که مدل فعلی هنوز در پاسخ به پرسش‌های پزشکی و حقوقی اشتباه می‌کند و افزود که تیم فنی روی افزودن «نمره اطمینان» به هر پاسخ کار می‌کند. به گفته او، کاربر باید ببیند که مدل در چه حوزه‌هایی مطمئن است و در کجا حدس می‌زند. این ویژگی قرار است از نیمه دوم ۲۰۲۶ به صورت آزمایشی در دسترس قرار گیرد.

ناظران بین‌المللی هوش مصنوعی گفته‌اند که این رویکرد می‌تواند الگویی برای مقررات آینده باشد، زیرا به جای ممنوع کردن فناوری، از شرکت‌ها می‌خواهد میزان خطای خود را اعلام کند. با این حال، منتقدان هشدار داده‌اند که اعلام خطا به تنهایی کافی نیست و کاربران ممکن است به پاسخ‌هایی که با لحن مطمئن ارائه می‌شوند بیش از حد اعتماد کنند.

مدیران DeepSeek اعلام کردند که نسخه آزمایشی ویژگی نمره اطمینان تا پایان ژوئن ۲۰۲۶ در دسترس برخی کاربران قرار می‌گیرد و بازخورد آنان در نسخه بعدی مدل لحاظ خواهد شد. این شرکت همچنین گفت که گزارش ماهانه‌ای درباره خطاهای مدل منتشر خواهد کرد تا پژوهشگران بتوانند روند بهبود آن را بررسی کنند.

DeepSeek به نیت در ده سال آینده هوش مصنوعی قابل اعتماد می‌سازد

مدیران ارشد شرکت هانگ‌جو اعلام کردند که تا ده سال آینده مدل‌های زبانی باید بدانند کجا اشتباه می‌کنند و به کاربر بگویند چرا چیزی را درست یا نادرست می‌دانند. این شرکت که مدل‌های ارزان‌قیمت آن توجه بازار را به خود جلب کرده، اکنون می‌گوید هدف بعدی آن ساختن سامانه‌هایی است که کاربران بتوانند بدون ترس از فریب، به آن‌ها در تصمیم‌گیری‌های روزمره تکیه کنند.

مدیر ارشد شرکت در سخنرانی افتتاحیه گفت: «ما نمی‌خواهیم کاربران ما فقط پرسش بپرسند و پاسخ بگیرند. می‌خواهیم بدانند چه زمانی می‌توان به پاسخ ما اعتماد کرد و چه زمانی باید تردید کرد.» به گفته او، تا پایان سال ۲۰۲۷ نسخه‌ای از مدل DeepSeek طراحی شده است که عدم قطعیت خود را نشان می‌دهد و به کاربر می‌گوید چرا چیزی را درست یا نادرست می‌داند.

مدیران همچنین اعلام کردند که داده‌های کاربران ایران و چین در سرورهای جداگانه ذخیره می‌شود و هیچ کارمند خارجی به آن دسترسی ندارد. این موضوع پاسخی به نگرانی‌های مقامات اتحادیه اروپا است که از ماه ژوئیه گذشته درباره انتقال داده‌های شهروندان اروپایی به شرکت‌های چینی هشدار داده‌اند. به گفته ناظران، این سازوکار می‌تواند الگویی برای سایر شرکت‌های فناوری باشد که می‌خواهند بدون نقض مقررات وارد بازار شوند.

در بخش پرسش‌وپاسخ، یکی از مدیران ارشد پذیرفت که مدل فعلی هنوز در پاسخ به پرسش‌های پزشکی و حقوقی اشتباه می‌کند و افزود که تیم فنی روی افزودن «نمره اطمینان» به هر پاسخ کار می‌کند. به گفته او، کاربر باید ببیند که مدل در چه حوزه‌هایی مطمئن است و در کجا حدس می‌زند. این ویژگی قرار است از نیمه دوم ۲۰۲۶ به صورت آزمایشی در دسترس قرار گیرد.

ناظران بین‌المللی هوش مصنوعی گفته‌اند که این رویکرد می‌تواند الگویی برای مقررات آینده باشد، زیرا به جای ممنوع کردن فناوری، از شرکت‌ها می‌خواهد میزان خطای خود را اعلام کند. با این حال، منتقدان هشدار داده‌اند که اعلام خطا به تنهایی کافی نیست و کاربران ممکن است به پاسخ‌هایی که با لحن مطمئن ارائه می‌شوند بیش از حد اعتماد کنند.

مدیران DeepSeek اعلام کردند که نسخه آزمایشی ویژگی نمره اطمینان تا پایان ژوئن ۲۰۲۶ در دسترس برخی کاربران قرار می‌گیرد و بازخورد آنان در نسخه بعدی مدل لحاظ خواهد شد. این شرکت همچنین گفت که گزارش ماهانه‌ای درباره خطاهای مدل منتشر خواهد کرد تا پژوهشگران بتوانند روند بهبود آن را بررسی کنند.

مدیران ارشد DeepSeek به نیت در ده سال آینده هوش مصنوعی قابل اعتماد می‌سازند