گزارش خبر
NEWSREPORT

کد خبر: ۵۲۳۴۶۲
یکشنبه ۳ اسفند ۱۴۰۴ ۶:۴۷

فاجعه دقت در هوش مصنوعی؛ مدل‌های پیشرفته در گفت‌وگوهای طولانی تا ۳۵ درصد خطا دارند

ظاهراً دقت مدل‌های هوش مصنوعی در گفت‌وگوهای چندمرحله‌ای طولانی کاهش می‌یابد.

به گزارش روابط عمومی دانشگاه سئوی ایران، شرکت‌های پیشرو طی سال‌های اخیر مدل‌ها و چت‌بات‌های پیشرفته‌ای را برای تثبیت جایگاه برند خود در بازار پرشتاب هوش مصنوعی معرفی کرده‌اند؛ بازاری که دنبال‌کردن تحولاتش همچنان ساده نیست و با وجود پیشرفت‌های فنی، کاربران همچنان از توهم‌زایی و پاسخ‌های نادرست در تعامل با این مدل‌ها گلایه دارند.

Microsoft Research و Salesforce در پژوهشی مشترک با تحلیل بیش از ۲۰۰ هزار مکالمه‌ی مرتبط با مدل‌هایی مانند Gemini ۲.۵ Pro، GPT-۴.۱، DeepSeek R۱، o۳، Claude ۳.۷ Sonnet و Llama ۴، نشان دادند که مدل‌های زبانی بزرگ در مکالمه‌های چندمرحله‌ای دچار افت عملکرد می‌شوند و به‌اصطلاح «در گفت‌وگو گم می‌شوند».

طبق تحقیقات، نرخ موفقیت مدل‌هایی مانند GPT-۴.۱ و Gemini ۲.۵ Pro در سناریوهای تک‌درخواستی به حدود ۹۰ درصد می‌رسد؛ اما عملکرد همین مدل‌ها در گفت‌وگوهای رفت‌وبرگشتی طولانی با کاهش محسوسی روبه‌رو می‌شود و میانگین موفقیت آن‌ها به نزدیک ۶۵ درصد سقوط می‌کند.

پژوهش مایکروسافت نشان می‌دهد افت کیفیت لزوماً به‌معنای کاهش چشمگیر توانایی نیست. در واقع، گفته می‌شود هرچه بیشتر با چت‌بات‌ها صحبت کنید، دقت‌شان کمتر خواهد شد.

تحلیل‌ها همچنین از بروز پدیده‌ای با عنوان «تورم پاسخ» حکایت دارند؛ به‌گونه‌ای که طول خروجی‌ها در مکالمه‌های چندنوبتی بین ۲۰ تا ۳۰۰ درصد افزایش می‌یابد. پاسخ‌های مفصل‌تر فرضیات بیشتری وارد متن می‌کنند و احتمال توهم‌زایی را بالا می‌برند؛ خطاهایی که سپس به‌عنوان زمینه‌ی دائمی گفت‌وگو استفاده می‌شوند.

در شرایطی که استفاده از ابزارهای هوش مصنوعی به‌سرعت گسترش پیدا می‌کند و قابلیت‌هایی مانند Google AI Overviews به تجربه‌ی جست‌وجوی افزوده شده‌اند، کنارگذاشتن موتورهای جست‌وجوی سنتی و اتکا به پاسخ‌های تولیدشده با مدل‌های هوش مصنوعی می‌تواند ریسک‌پذیر باشد.

منبع:زومیت