

## ChatGPT رادیولوژیست شد



پژوهشگران مدل هوش مصنوعی ChatGPT را در یک آزمون رادیولوژی سنجیدند و مشاهده کردند که خوب عمل کرد، اما با برخی اظهارات غیرمنطقی و نادرست به نظر می‌رسد هنوز آماده جایگزینی با رادیولوژیست‌ها نیست.

پژوهشگران مدل هوش مصنوعی ChatGPT را در یک آزمون رادیولوژی سنجیدند و مشاهده کردند که خوب عمل کرد، اما با برخی اظهارات غیرمنطقی و نادرست به نظر می‌رسد هنوز آماده جایگزینی با رادیولوژیست‌ها نیست.

به گزارش ایسنا و به نقل از اس ای، پژوهشگران بیمارستان عمومی تورنتو در کانادا کاری را انجام دادند که اکثر مردم این روزها انجام می‌دهند: استفاده از ربات هوش مصنوعی ChatGPT برای پاسخ به سؤالات یک آزمون استاندارد و دیدن نحوه عملکرد آن.

این چت بات (ربات گفتگوگر) در یک آزمون ۱۵۰ سوالی که توسط گروه‌های رادیولوژی در کانادا و ایالات متحده طراحی شده بود، ۸۱ درصد امتیاز کسب کرد که بسیار بالاتر از حداقل درصد قبولی (۷۰ درصد) بود.

از زمان راه اندازی ChatGPT، کاربران از توانایی آن در درک اطلاعات و استفاده از آن برای پاسخ به سؤالات شگفت زده شده‌اند. این ربات برای پاسخگویی به سؤالات آزمون مجوز پزشکی ایالات متحده (USMLE) و همچنین آزمون MBA در مدرسه بازرگانی وارتون نیز مورد آزمایش قرار گرفته و عملکرد متوسطی داشته است.

با افزایش استفاده از ChatGPT در بخش‌های مختلف، پژوهشگران دانشگاه پزشکی تصویربرداری تورنتو نیز به این فکر افتادند که زمان آن رسیده است که توانایی‌های این چت بات را در رادیولوژی نیز آزمایش کنند.

### پاسخ ChatGPT به سؤالات رادیولوژی

پژوهشگران یک آزمون ۱۵۰ سوالی برای ChatGPT طراحی کردند، دقیقاً مانند آنچه که گروه‌های رادیولوژی در کانادا و ایالات متحده برای دانشجویان این رشته انجام می‌دهند. از آنجایی که این ربات هوش مصنوعی نمی‌تواند تصاویر را به عنوان ورودی پردازش کند، پژوهشگران فقط متن در سؤالات قرار دادند که شامل سؤالات پایه و مفهومی و سؤالات تخصصی بود.

سؤالات پایه، دانش و درک اولیه ChatGPT را در مورد موضوع مورد آزمایش می‌سنجید، در حالی که سؤالات تخصصی، تجزیه و تحلیل و ترکیب اطلاعاتی را که برای تفسیر تصاویر رادیولوژی لازم است، می‌آزمود.

از آنجایی که دو نسخه از GPT در حال حاضر موجود است، پژوهشگران هر دوی آنها را روی یک مجموعه سوال آزمایش کردند تا ببینند آیا یکی از دیگری بهتر است یا خیر.

ChatGPT ساخته شده با نسخه قدیمی‌تر یعنی GPT ۳.۵، توانست امتیاز کلی ۶۹ درصد را در مجموعه سؤالات کسب کند، در حالی که در سؤالات پایه نمره خوبی کسب کرد (۸۴ درصد معادل ۵۱ پاسخ صحیح به ۶۱ سوال) اما در پاسخ به سؤالات تخصصی فقط ۶۰ درصد امتیاز گرفت (معادل ۵۲ پاسخ صحیح به ۸۹ سوال).

پس از انتشار GPT-۴ در مارس ۲۰۲۳، پژوهشگران این نسخه بهبود یافته ChatGPT را دوباره آزمایش و مشاهده کردند که با پاسخ صحیح به ۱۲۱ سوال از ۱۵۰ سوال، امتیاز ۸۱ درصد را به دست آورد.

با این حال، چیزی که پژوهشگران را شگفت زده کرد، عملکرد GPT-۴ در سؤالات پایه است، جایی که به ۱۲ سوال پاسخ اشتباه داد، در حالی که GPT ۳.۵ به آنها به درستی پاسخ داده بود.

راجش بایانا، رادیولوژیست و سرپرست فناوری در بیمارستان عمومی تورنتو می‌گوید: در ابتدا از پاسخ‌های دقیق و مطمئن ChatGPT به برخی سؤالات چالش برانگیز حول رادیولوژی متعجب شدیم، اما سپس به همان اندازه از اظهارات بسیار غیرمنطقی و نادرست آن به برخی دیگر از سؤالات متعجب شدیم.

در حالی که تمایل به ارائه اطلاعات نادرست در GPT-۴ کاهش یافته است، اما هنوز به طور کامل از بین نرفته است. پژوهشگران

می گویند، این موضوع می تواند در پزشکی خطرناک باشد، به ویژه هنگامی که توسط افراد تازه کار استفاده می شود که ممکن است نتوانند پاسخ های نادرست را تشخیص دهند.

این پژوهش به تازگی در مجله Radiology منتشر شده است.