



ارزیابی ChatGPT به عنوان یک سیستم پرسش و پاسخ: تحلیلی جامع و مقایسه با مدل‌های موجود



حسین باهک آ، فرزانه طاهری آ، زهرا زجاجی آ*، عارفه کاظمی ب

آ دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران.
ب مرکز تحقیقات ADAPT، دوبلین، ایرلند.

چکیده

اطلاعات مقاله

تاریخچه مقاله:

دریافت: 12 February 2024

اصلاح: 19 September 2024

پذیرش: 28 September 2024

انتشار آنلاین: 16 June 2025

کلمات کلیدی:

ChatGPT، سامانه‌های پاسخ‌گویی به

پرسش، مدل‌های زبانی بزرگ، ارزیابی

عملکرد، توهم (Hallucination)

در عصر حاضر، مدل‌های زبانی متعددی برای پاسخگویی به پرسش‌های کاربران توسعه یافته‌اند. مدل زبانی «Chat-3.5-Turbo» به عنوان فناوری پایه «ChatGPT» توجه قابل توجهی را به خود جلب کرده است. این مدل با بهره‌گیری از پارامترهای گسترده، به طیف وسیعی از پرسش‌ها به خوبی پاسخ می‌دهد. با این حال، به دلیل وابستگی به دانش داخلی، دقت پاسخ‌ها ممکن است کافی نباشد. این مقاله، «ChatGPT» را به عنوان یک سیستم پرسش و پاسخ (QAS) بررسی کرده و عملکرد آن را با سایر سیستم‌های موجود مقایسه می‌کند. تمرکز اصلی بر ارزیابی کارایی ChatGPT در استخراج پاسخ از پاراگراف‌های داده شده به مدل است که قابلیت اصلی یک QAS سیستم پرسش و پاسخ به شمار می‌رود. همچنین، عملکرد آن در سناریوهای فاقد متن زمینه بررسی و تحلیل شده است. آزمایش‌های متعددی برای ارزیابی ChatGPT با تمرکز بر جنبه‌های مختلفی چون توهم (Hallucination) و پیچیدگی پرسش‌ها انجام شده است. ارزیابی با استفاده از مجموعه داده‌های شده پرسش و پاسخ شامل، NewsQA SQuAD و PersianQuAD به زبان‌های انگلیسی و فارسی صورت گرفته است. معیارهایی مانند F-score، انطباق دقیق و صحت در ارزیابی به کار رفته است. نتایج به دست آمده نشان می‌دهد اگرچه ChatGPT به عنوان یک مدل مولد توانمند است، در مقایسه با مدل‌های خاص وظیفه پرسش و پاسخ ضعیف‌تر عمل می‌کند. همچنین بررسی‌ها نشان داد ارائه متن زمینه به این مدل، عملکرد آن را بهبود می‌بخشد و مهندسی ورودی (prompt) به ویژه برای پرسش‌های فاقد پاسخ صریح در پاراگراف‌ها، دقت را افزایش می‌دهد. از سوی دیگر مدل GPT در پاسخ به پرسش‌های ساده و واقع نما (مثلاً «چه» و «کی») نسبت به پرسش‌های تحلیلی (مانند «چگونه» و «چرا») عملکرد بهتری از خود نشان داد. در نهایت ارزیابی‌ها وجود توهم در پاسخ به پرسش‌هایی که در متن زمینه، پاسخی برای آنها وجود ندارد را تایید نمود.

© 2024. تمامی حقوق محفوظ است.

* نویسنده مسئول.

آدرس رایانامه: z.zojaji@eng.ui.ac.ir (زهرا زجاجی)

تمامی حقوق محفوظ است. ISSN: 2322-4460 © 2024

