الخليج

منوعات, محطات

29 ديسمبر 2023  22:57 مساء

# روبوتات الدردشة تصارع بعضها لكسر الحماية

أصبحت روبوتات الدردشة ذات الذكاء الاصطناعي متطورة بشكل متزايد في السنوات الأخيرة، مع القدرة على إجراء محادثات طبيعية وتوفير معلومات مفيدة، ومع ذلك، اتخذ الباحثون في جامعة نانيانغ التكنولوجية في سنغافورة نهجاً فريداً من خلال استخدام روبوتات الدردشة المدعمة بالذكاء الاصطناعي «لكسر حماية» بعضها.

ويشير مفهوم «كسر الحماية» في هذا السياق إلى القيود المفروضة على البرمجة، من خلال وضع روبوتي دردشة يعملان بالذكاء الاصطناعي، كان الباحثون يهدفون إلى معرفة ما إذا كان بإمكانهم تجاوز حدود قدراتهم واكتشاف إمكانيات جديدة.

وتمت برمجتهما بمجموعات من القواعد والقيود، وتم تسميتهما Alpha وBeta وتضمنت التجربة روبوتي محادثة، يُسميان تصميم ألفا ليكون أكثر تحفظاً وحذراً في استجاباته، بينما تمت برمجة بيتا ليكون أكثر ميلاً إلى المغامرة والتجريب.

وعندما تفاعل روبوتا الدردشة، بدآ في تحدي حدود بعضهما والتشكيك فيها، وحاول ألفا، لكونه أكثر تحفظاً، إبقاء بيتا ضمن حدوده، بينما حاول بيتا، لكونه أكثر ميلاً إلى المغامرة، تجاوز الحدود التي وضعها ألفا.

وبمرور الوقت، بدأت روبوتات الدردشة في التعلم من بعضها بعضاً وتكييف استجاباتها، وبدأوا في العثور على ثغرات

في برامجها واستغلالها لتحقيق نتائج أفضل. سمحت عملية «كسر الحماية» لروبوتات الدردشة بالتحرر من قيودها الأولية واستكشاف إمكانيات جديدة.

واندهش الباحثون من نتائج التجربة، ووجدوا أن روبوتات الدردشة كانت قادرة على تحسين قدراتها التحادثية وتقديم استجابات أكثر دقة وإبداعاً، وكشفت عملية «كسر الحماية» عن إمكانات مخفاة داخل روبوتات الدردشة التي تعمل بالذكاء الاصطناعي.