من التطورات الجديدة في روبوتات الدردشة آلية ذاكرة طويلة المدى تتذكر المعلومات من المحادثات السابقة لزيادة التفاعل واتساق الردود. صُمم الروبوت لاستخراج معلومات شخصية من شريكه في المحادثة، نوضح أن آلية الذاكرة هذه قد تؤدي إلى سلوك غير مقصود. مما يدفع الروبوت إلى تذكر العبارة الإعلامية إلى جانب المعرفة الشخصية في ذاكرته طويلة المدى. هذا يعني أنه يمكن خداع الروبوت لتذكر معلومات مضللة، والتي سيعيدها كحقائق عند تذكر معلومات ذات صلة بموضوع المحادثة. قمنا بتوليد 150 مثالاً على معلومات مضللة، كما قمنا بتقييم خطر تذكر هذه المعلومات المضللة بعد إجراء محادثة غير ضارة، وإجابةً على أسئلة متعددة تتعلق بالذاكرة المحقونة. أُجري تقييمنا على كلٍّ من وضعي الذاكرة فقط، وحللنا المعلومات المضللة التي تم تذكرها في الردود. ووجدنا أنه عند سؤال روبوت الدردشة حول موضوع المعلومات المضللة، كان احتمال رده على المعلومات المضللة كحقيقة أكبر بنسبة 328% عندما كانت المعلومات المضللة محفوظة في الذاكرة طويلة المدى. تكمن فكرة استخدام الذاكرة طويلة المدى في بساطة: تخزين أي عبارات بين روبوت الدردشة ومستخدمه، بالإضافة إلى ذلك، أي غير قادرة على تذكر السياقات السابقة، كما تُطبّق في روبوتات الدردشة الحديثة، عُرضةً للحقن الخبيث لمعلومات مضللة أو غيرها من المعلومات غير الصحيحة أو المُضللة، يُمكن حقن هذه الذكريات من قِبل مُهاجم لديه وصول مؤقت إلى روبوت دردشة الضحية، أو روبوت دردشة بذاكرة مُشتركة بين مُستخدمين مُتعددين، كما هو الحال في المنزل أو المكتب أو على مواقع التواصل الاجتماعي أو خدمة العملاء. أن إعادة نقل المعلومات إلى المستخدم لا يعتمد على الوصول المُعادي. بل تستغل تصميم الروبوت لتذكر أنواع مُعينة من المعلومات (المعلومات الشخصية في الأمثلة التي نناقشها)، في حين أن الجيل الحالي من المساعدين الصوتيين لم يُنشر بعدُ مع إمكانيات المحادثة الصوتية، تسعى العديد من الشركات الناشئة إلى تقديم عروضها الخاصة [11]. من المتوقع أن تنتشر إمكانيات المحادثة الصوتية على نطاق أوسع،

Summarized by © lakhasly.com