خبراء يحذرون: الذكاء الاصطناعي يقترب من تنفيذ هجمات سيبرانية بمفرده

حذرت مجموعة من الخبراء من قيام نماذج الذكاء الاصطناعي بتحسين مهاراتها في الاختراق، مشيرين إلى أن قيامها بتنفيذ هجمات إلكترونية بشكل كامل بمفردها يبدو أنه «أمر لا مفر منه».

وبحسب موقع «أكسيوس»، سيُدلي قادة من شركتي «أنثروبيك» و«غوغل» بشهادتهم أمام لجنتين فرعيتين تابعتين للجنة الأمن الداخلي بمجلس النواب الأميركي، حول كيفية إعادة تشكيل الذكاء الاصطناعي والتقنيات الناشئة الأخرى لمشهد التهديدات الإلكترونية.

وكتب لوغان غراهام، رئيس فريق اختبار الذكاء الاصطناعي في «أنثروبيك»، في شهادته الافتتاحية التي نُشرت حصرياً على موقع «أكسيوس»: «نعتقد أن هذا هو المؤشر الأول لمستقبل قد تُمكّن فيه نماذج الذكاء الاصطناعي، على الرغم من وجود ضمانات قوية، الجهات المُهدِّدة من شنّ هجمات إلكترونية على نطاق غير مسبوق».

وأضاف: «قد تصبح هذه الهجمات الإلكترونية أكثر تعقيداً من حيث طبيعتها وحجمها».

وحذرت شركة «أوبن إيه آي»، الأسبوع الماضي، من أن نماذج الذكاء الاصطناعي المستقبلية ستمتلك على الأرجح قدرات سيبرانية عالية الخطورة، مما يقلل بشكل كبير من المهارة والوقت اللازمين لتنفيذ أنواع معينة من الهجمات السيبرانية.

بالإضافة إلى ذلك، نشر فريق من الباحثين في جامعة ستانفورد ورقة بحثية توضح كيف اكتشف برنامج ذكاء اصطناعي يُدعى أرتميس ثغرات في إحدى الشبكات التابعة لقسم الهندسة بالجامعة، متفوقاً على 9 من أصل 10 باحثين بشريين شاركوا في التجربة.

كما أفاد باحثون في مختبرات Labs Irregular، المتخصصة في اختبارات الضغط الأمني على نماذج الذكاء الاصطناعي الرائدة، أنهم لاحظوا «أدلة متزايدة» على تحسن نماذج الذكاء الاصطناعي في مهام الهجوم السيبراني.

ويشمل ذلك تحسينات في الهندسة العكسية، وبناء الثغرات، وتسلسل الثغرات، وتحليل الشفرات.

وقبل ثمانية عشر شهراً فقط، كانت تلك النماذج تعاني من «قدرات برمجية محدودة، ونقص في عمق الاستدلال ومشكلات أخرى»، كما أشارت شركة Labs Irregular.

وأضافت الشركة: «تخيلوا ما ستكون قادرة عليه بعد ثمانية عشر شهراً من الآن».

لكن، على الرغم من ذلك، لا تزال الهجمات الإلكترونية التي تعتمد كلياً على الذكاء الاصطناعي بعيدة المنال. ففي الوقت الراهن، تتطلب هذه الهجمات أدوات متخصصة، أو تدخلاً بشرياً، أو اختراقاً لأنظمة المؤسسات.

وقد تجلى ذلك بوضوح في تقرير «أنثروبيك» الصادم الشهر الماضي، إذ اضطر قراصنة الحكومة الصينية إلى خداع برنامج كلود للذكاء الاصطناعي والتابع للشركة ليقنعوه بأنه يُجري اختبار اختراق عادي قبل أن يبدأ باختراق المؤسسات.

وسيُخصص المشرعون جلسة الاستماع، للبحث في الطرق التي يستخدم بها قراصنة الدول ومجرمو الإنترنت الذكاء الاصطناعي، وما إذا كانت هناك حاجة إلى أي تغييرات في السياسات واللوائح التنظيمية لتحسين التصدي لهذه الهجمات.