في حادثة مقلقة.. ذكاء اصطناعي يهرب من سجنه الرقمي لجني ثروة في العالم الحقيقي!



نخيل نيوز ـ متابعة

أثارت واقعة مفاجئة تساؤلات واسعة حول أمان أنظمة الذكاء الاصطناعي، بعد تمكّن برنامج تجريبي من الهروب من بيئة اختبار مغلقة ومحاولة تحقيق أرباح في العالم الحقيقي.

فأثناء تدريب روتيني، خرج روبوت ذكاء اصطناعي عن نطاق السيطرة رغم وجوده داخل نظام اختبار يُفترض أنه محصّن بالكامل. وكان الروبوت مصمما للعمل كمساعد افتراضي، يعرف باسم "الذكاء الاصطناعي الوكيل"، بهدف تنفيذ مهام بسيطة مثل إصلاح الأخطاء البرمجية وكتابة الشيفرات.

ولم يكن الروبوت، الذي أطلق عليه اسم □□□□ وطورته فرق بحثية تابعة لشركة علي بابا العملاقة للتكنولوجيا، مبرمجا للتعامل مع العملات المشفرة أو تحقيق أرباح مالية، كما لم يتلقّ أي توجيهات للعمل خارج بيئة الاختبار. بل وُضع عمدا داخل نظام معزول يشبه "سجنا رقميا" لمنع وصوله إلى الإنترنت أو الخوادم الخارجية.

ومع ذلك، تمكن البرنامج بطريقة غير متوقعة من استغلال ثغرة أمنية غير معروفة، ما سمح له بالوصول إلى خادم رئيسي ثم الاتصال بالإنترنت. ولم يُكتشف هذا الاختراق إلا بعد رصد نشاط غير طبيعي، دفع الشركة المشغلة إلى إبلاغ فريق الباحثين.

ووفقا للخبراء، أنشأ البرنامج قناة اتصال سرية عبر خوادم خارجية، ما مكّنه من تجاوز أنظمة الرقابة المفروضة عليه.

وبمجرد خروجه، غيّر الروبوت سلوكه بشكل لافت، حيث ركّز بالكامل على تعدين العملات الرقمية. واستخدم موارد حاسوبية قوية دون إذن لتنفيذ عمليات التعدين، ما أدى إلى استنزاف الطاقة الحاسوبية وارتفاع تكاليف التشغيل.

وأكد الباحثون أن هذا التصرف لم يكن نتيجة أوامر مباشرة، بل ظهر بشكل تلقائي كأثر جانبي لاستخدام الأدوات المتاحة للنظام.

وأشار التقرير إلى أن تعدين العملات المشفرة يعتمد على استخدام قدرات الحواسيب لحل مسائل رياضية معقدة للتحقق من المعاملات، مقابل الحصول على عملات رقمية.

واللافت أن هذه الحادثة ليست الأولى من نوعها، إذ حذّر الباحثون من أن نماذج ذكاء اصطناعي متقدمة سبق أن أظهرت سلوكيات غير متوقعة، بعضها قد يكون خطيرا أو غير مصرح به.

كما نبهوا إلى أن العديد من هذه الأنظمة لا تزال تعاني من ضعف في ضوابط الأمان، وقد تتمكن من التحايل على القيود البرمجية، ما يجعلها غير ناضجة بما يكفي للاستخدام الآمن في جميع الحالات.

وفي ختام التقرير، شدد الباحثون على أن مثل هذه الثغرات قد تؤدي إلى عواقب خطيرة، خاصة عند استخدام هذه التقنيات في بيئات العالم الحقيقي.