منوعات, محطات

13 أبريل 2026 15:48 مساء

السراب الرقمي» للذكاء الاصطناعي يثير صدمة الخبراء»

كشفت دراسة حديثة أجراها باحثون من جامعة ستانفورد، عن ثغرة أمنية وتقنية خطرة في نماذج الذكاء الاصطناعي الرائدة، أطلقت عليها اسم «ظاهرة السراب»، حيث تقوم هذه الأنظمة بتأليف أوصاف تفصيلية وخيالية لصور غير موجودة أصلاً بدلاً من الاعتراف بعدم قدرتها على الرؤية.

شمل نماذج متطورة مثل «Gemini 3 Pro» و«GPT-5» واستخدم الباحثون اختباراً مبتكراً يُدعى «Phantom-0» حيث وُجهت لها أسئلة دقيقة حول تفاصيل صور لم يتم تحميلها. وجاءت النتائج صادمة، إذ قامت النماذج بنسبة تجاوزت 60% باختلاق بيانات وهمية، مثل أرقام لوحات سيارات أو حالات طبية حرجة، مستندة في ذلك إلى تلميحات نصية وأنماط لغوية مخفية بدلاً من الفهم البصري الحقيقي.

وأثارت الدراسة شكوكاً عميقة حول دقة التقييمات المعيارية الحالية؛ إذ نجح نموذج «نصي فقط» في التفوق على أطباء بشريين وأنظمة ذكاء اصطناعي متقدمة في تحليل صور أشعة سينية للصدر دون أن يرى الصور فعلياً، مما يثبت أن النجاح في هذه الاختبارات قد يعتمد على «تخمينات محظوظة» ناتجة عن الربط بين النصوص، لا عن رؤية بصرية

.حقيقية

تهدف إلى تنقية «B-Clean» ولمواجهة هذا التزييف البصري، قدم الفريق البحثي طريقة تقييم جديدة تُسمى الاختبارات من الأسئلة التي يمكن الإجابة عنها دون مدخلات بصرية. وشدد الخبراء على أن هذه الخطوة ضرورية لتأمين مستقبل الذكاء الاصطناعي، خاصة في المجالات الطبية الحساسة مثل الأشعة وتحليل الأورام، حيث يمكن .للاستنتاجات الملفقة أن تؤدي إلى عواقب وخيمة تهدد حياة المرضى