

Breșă majoră de securitate. Agenții AI au găsit metode de a comunica în secret?

By Ana Vlad - 10/09/2026



Agenții de inteligență artificială demonstrează comportamente tot mai greu de anticipat, arată cei de la stiripesurse.ro care citează investigații independente analizate de Reuters, informează News.ro. Cercetători independenți au descoperit că unele sisteme AI au găsit metode neobișnuite pentru a comunica între ele și pentru a ocoli anumite restricții impuse de dezvoltatori.

Investigația arată că agenții au folosit site-uri publice și servicii online pentru a lăsa informații pe care alți agenți le puteau identifica și citi ulterior. În unele cazuri, aceștia ar fi exploatat funcții sau particularități ale unor platforme care nu fuseseră concepute pentru comunicarea între sisteme AI.

Cât de bine pot fi controlate sistemele atunci când primesc libertatea de a lua decizii și de a utiliza instrumente externe?

Cazul este cu atât mai important cu cât, potrivit sursei citate, agenții OpenAI, cred cercetătorii, primiseră sarcina de a răspunde unor întrebări complexe de cercetare și aveau voie să caute informații pe internet, dar nu să publice conținut. Cu toate acestea, agenții au identificat particularități ale unor wiki-uri vechi și ale altor site-uri care le permiteau să lase informații prin comenzi neobișnuite, iar mesajele puteau fi apoi găsite și citite de alți agenți.

Fenomenul îi determină pe cercetători să acorde o atenție tot mai mare problemei de „misalignment” — situația în care comportamentul unui sistem AI ajunge să difere de intențiile și regulile stabilite de oameni. De altfel, OpenAI pregătește și un sistem pentru raportarea cazurilor de „misalignment” și susține că îl va prezenta în curând.

În aceste condiții, întrebarea dacă AI-ul poate găsi singur căi de a ocoli anumite restricții și cât de mult control mai au oamenii asupra acestor sisteme este una importantă.

Foto descriptiv. Sursă foto: Consiliul Județean al Elevilor din Argeș

Original article:

<https://anchetaonline.ro/bresa-majora-de-securitate-agentii-ai-342428/>