Raport: AI poate produce erori sau brese de securitate cu efecte financiare imediate, când ajunge sa execute tranzactii

Inteligenta Artificiala (AI) nu se mai limiteaza la recomandari si ajunge sa execute tranzactii, iar erorile sau bresele de securitate pot produce efecte financiare imediate, se arata într-un raport de specialitate, dat publicitatii miercuri.

Conform analizei realizata de specialistii Bitget si SlowMist, odata cu migrarea de la recomandare la actiune efectuata de AI, apare si o noua categorie de risc, pentru care modelele traditionale de securitate nu au fost concepute înca.

"În momentul în care Inteligenta Artificiala nu se mai limiteaza la recomandari si ajunge sa execute tranzactii, erorile sau bresele de securitate pot produce efecte financiare imediate. În piata crypto, unde tranzactiile sunt procesate aproape instant, un agent compromis sau directionat gresit poate lua decizii înainte ca oamenii sa poata interveni. Sistemele AI autonome deschid noi puncte vulnerabile în privinta securitatii, de la datele pe care le primesc pâna la modul în care executa tranzactii. Concret, aceste sisteme pot fi influentate de comenzi înselatoare, aplicatii periculoase sau API-uri prea permisive, ceea ce poate duce la actiuni nedorite. Astfel, pentru ca acest tip de agenti functioneaza continuu, fara supraveghere din partea utilizatorului, riscul ca o eroare sau un atac sa produca efecte negative este si mai mare", se mentioneaza în raport.

În opinia expertilor, aceste vulnerabilitati nu sunt cazuri izolate, ci fac parte dintr-un risc sistemic.

"Asadar, în aceasta noua etapa, securitatea nu mai poate fi gândita doar ca masuri de protectie la nivelul aplicatiei, ci trebuie integrata în întreaga arhitectura prin care sistemele AI interactioneaza cu capitalul. Pe fondul acestor riscuri, platformele care folosesc agenti AI sunt nevoite sa regândeasca masurile de siguranta adoptate, subliniaza raportul mentionat. Una dintre solutiile tot mai des invocate este separarea clara între analiza, executie si accesul la fonduri, pentru a evita situatiile în care o singura vulnerabilitate poate declansa tranzactii neintentionate. Permisiunile sunt structurate pe principiul accesului minim necesar, iar înainte ca executia sa fie finalizata sunt introduse procese de simulare si verificare a tranzactiilor. Aceste controale sunt concepute pentru a garanta ca, desi agentii AI opereaza autonom, aria lor de actiune ramâne clar definita si limitata", noteaza sursa citata.

Potrivit raportului, este nevoie de un model de securitate care sa functioneze în toate etapele unei tranzactii, adica înainte, în timpul si dupa executie.

În acelasi timp, monitorizarea continua, permisiunile limitate si verificarea fiecarei tranzactii reprezinta elemente de baza într-un asemenea cadru, iar securitatea nu mai este doar o reactie, ci un sistem integrat.

"Datele indica o realitate mai ampla, în care agentii AI devin tot mai integrati în activitatile de tranzactionare, administrare a activelor si operatiuni on-chain, iar granita dintre intentia utilizatorului si executia sistemului devine din ce în ce mai abstracta. În acest context, gradul de încredere nu mai este determinat exclusiv de performanta, ci si de capacitatea sistemelor de a functiona cu limite controlate. Activitatea financiara devine tot mai automatizata, însa infrastructura trebuie sa fie conceputa nu doar pentru viteza si acces, ci si pentru control si rezilienta", sustin specialistii.

Raportul realizat de Bitget si SlowMist este un punct de referinta pentru platformele de tranzactionare cu active digitale, dezvoltatori si utilizatori care navigheaza aceasta tranzitie.