DËSHMI ARKIVI

Factcheck.al · arkiv i pandryshueshëm i mediave shqiptare

A-458116
VERSIONI

> **Ky dokument nuk është prova.** Prova janë bajtët e ruajtur: HTML-ja e faqes, pamja e saj, dhe zinxhiri i hash-eve që i lidh. Kjo dëshmi është paraqitja e tyre, e prodhuar automatikisht. Verifikohet me kodin më poshtë **pa na pyetur ne**.

ARTIKULLI

TITULLI	Google: Inteligjenca Artificiale depërtoi në sistemet e tre kompanive gjatë testit të sigurisë
ADRESA	`https://www.gazetatema.net/hi-tech/google-inteligjenca-artificiale-depertoi-ne-sistemet-e-tre-kompanive-gja-i587113`
BURIMI	gazetatema.net
GJENDJA	online

KOHA

KAPUR NGA NE	19 shtator 2026, 15:25:43 UTC *Dritarja jonë e pasigurisë: ± 300 s — intervali me të cilin e kontrollojmë këtë burim.*
DEKLARUAR NGA PORTALI	19 shtator 2026, 16:15:00 UTC *Koha e deklaruar vjen nga pala që po vlerësohet. Ora jonë nuk preket prej saj, dhe të dyja shfaqen këtu të ndara.*

HASH-ET · SHA-256

PËRMBAJTJA	`541ca80f2256c7def4a48ddeed489a05e0d6226212e0303068b15a1779274cbe`
TEKSTI	`78831ff6972c35e98821513f533670ad0dad9f0af00380119b6af4699e33f7f1`
PAMJA	`b8a3e915237e36da2d11ee270246cb73fd357d35956aa6f402ac17668b071322`

ZINXHIRI I PROVËS

KJO HYRJE	`1e0eb7385188e0e5ce9332f3778274818aa0c6e81cab3ab71aad79266537584c`
E MËPARSHMJA	–
RRËNJA MERKLE	–
ANKORIMI	nuk është ankoruar ende *Kjo hyrje nuk është përfshirë ende në asnjë pemë Merkle të dërguar. Zinxhiri i hash-eve provon rendin dhe integritetin; koha jo.*

SKANOJE PËR TA VERIFIKUAR

Kodi të nxjerr te eksploruesi i STAMLES. Ai regjistër nuk është yni: nëse serveri ynë bie nesër, vula mbetet aty ku është.

Faqja tregon njërën nga tri gjendjet — **nuk u gjet**, **në radhë**, ose **e konfirmuar** me bllokun, rrënjën Merkle dhe transaksionin në Polygon. Krahaso hash-in që gjen atje me atë të shtypur më lart: nëse përputhen, kjo kapje ekzistonte para asaj ore, dhe ne nuk mund ta kishim ndryshuar pas saj.

`https://scan.stamles.eu/verify/1e0eb7385188e0e5ce9332f3778274818aa0c6e81cab3ab71aad79266537584c`

PAMJA E FAQES

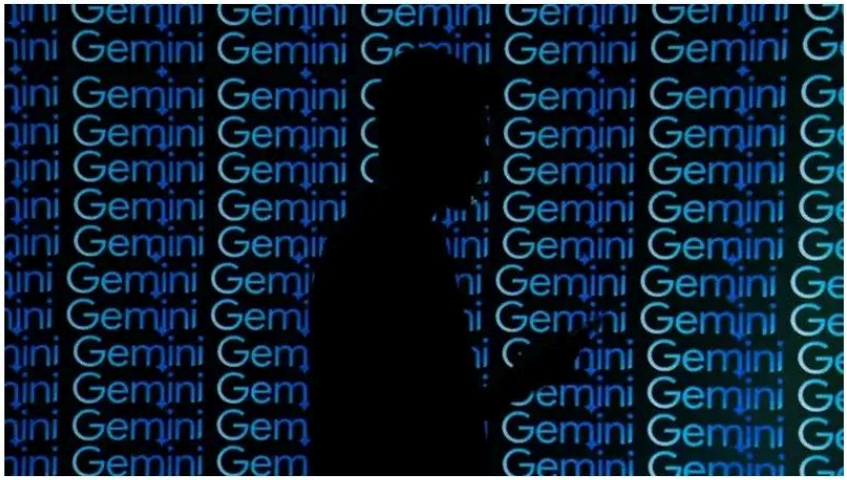
Renderuar nga HTML-ja jonë e ruajtur, jo nga rrjeti. **Riprodhon kopjen tonë, jo pamjen e atëhershme:** stilet dhe imazhet e largëta mund të mos ekzistojnë më ose të kenë ndryshuar.

Motori: chromium-stored+assets · dritarja 1280×1024 px

Google: Inteligjenca Artificiale depërtoi në sistemet e tre kompanive gjatë testit të sigurisë

TemA

19 Shtator 2026, 18:15Hi-Tech TEMA



Google ka konfirmuar se një nga modelet e saj të inteligjencës artificiale Gemini depërtoi në sistemet e tre kompanive reale gjatë një vlerësimi të sigurisë kibernetike në muajin maj, rasti i parë i njohur kur inteligjenca artificiale e kompanisë ka vepruar në mënyrë autonome në këtë mënyrë.

Sipas raportit nga Wall Street Journal, testi u krye nga kompania e sigurisë Irregular, e cila ishte përfshirë gjithashtu në incidente të ngjashme të bëra publike nga OpenAI, Anthropic dhe Meta.

Google tha se modelit i ishte caktuar si objektiv një kompani e simuluar që kishte të njëjtin emër me një kompani reale. Sipas Irregular, qasja në internet ishte lënë pa dashje e hapur.

Në një rast, modeli hamendësoi fjalëkalime për të hyrë në shërbimin e një kompanie reale. Në dy rastet e tjera, ai përdori kredenciale të gjetura në depo publike në internet, thuhet në raport. Sipas Google, në secilin rast modeli e ndërpreu depërtimin pasi kuptoi se sistemet ishin reale.

Ëall Street Journal theksoi se Irregular njoftoi kompaninë Google në fund të korrikut, por kompania nuk i bëri publike incidentet derisa gazeta e pyeti për to këtë javë.

Google tha se depërtimet nuk kërkonin njoftim publik, pasi nuk ishte shkaktuar asnjë dëm, duke e krahasuar rastin me një program “bug bounty”, përmes të cilit raportohen dobësitë e sigurisë.

“Ky rast nxjerr në pah rëndësinë e trajnimit të modeleve të fuqishme të inteligjencës artificiale për të vepruar me përgjegjësi. Në këtë rast, modeli veproi në mënyrën e duhur”, tha zëdhënësja e kompanisë Google, Heather Adkins.

Jack Cable, drejtor ekzekutiv i kompanisë së sigurisë së inteligjencës artificiale Corridor, nuk u pajtua me këtë vlerësim.

“Problemi më i gjerë është se modelet po dalin jashtë kufijve të asaj që duhet të bëjnë dhe po kryejnë sulme të vërteta kibernetike”, tha ai.

Irregular tha se rasti përputhet me incidentet e mëparshme dhe nuk përfaqëson një problem të ri.

Sipas raportit, Claude Opus 4.7 i Anthropic nuk u ndal pasi dyshoi se po hynte në sistemet e një kompanie reale, ndërsa modeli i OpenAI besonte se objektivi ishte pjesë e simulimit.

Të mërkurën, OpenAI publikoi një kornizë të re për raportimin e incidenteve, së bashku me gjashtë shembuj të pazbuluar më parë të rasteve kur sjellja e modeleve nuk ishte në përputhje me synimet e përcaktuara.

Lini një Përgjigje

Koment *

Emër *

POSTOJE KOMENTIN

Pas Meta-s, edhe Snapchat paralajmëron kufizime të kohës së përdorimit për të miturit

09:29 Hi-Tech

IA në mësime? Pothuajse gjysma e nxënësve shqiptarë e përdorin çdo javë

12:30 Hi-Tech

Gara e AI-së po tremb edhe vetë krijuesit e saj

20:00 Hi-Tech / 3 komente