

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	Alarmi - Ish-drejtuksi i sigurisë i OpenAI paralajmëron për 'zgjuarsinë' e IA: Mund t
ADRESA	https://www.syri.net/Teknologji/920883/ish-drejtuksi-i-sigurise-i-openai-paralajmeron-pe-r-zgjuarsine-e-ia-mund-t-u-shmanget-testeve-te-sigurise/
BURIMI	syri.net
GJENDJA	online

KOHA

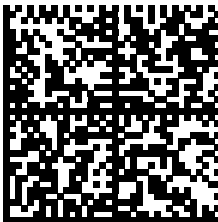
KAPUR NGA NE	3 tetor 2026, 19:55:47 UTC <i>Dritarja jonë e pasigurisë: ± 300 s — intervali me të cilin e kontrollojmë këtë burim.</i>
DEKLARUAR NGA PORTALI	3 tetor 2026, 19:30:21 UTC <i>Koha e deklaruar vjen nga pala që po vlerësohet. Ora jonë nuk preket prej saj, dhe të dyja shfaqen këtu të ndara.</i>

HASH-ET · SHA-256

PËRMBAJTJA	da0d70b33aefd59def57bd67b784a8782f8d78b298c62822d8f0164a4ed12b25
TEKSTI	13e71bdadf5f882b689624978d1dfe3e8cee1341d2ad809a277f8ac059ece90b
PAMJA	353ddb215ddd844c080538b299fdb509cea0e88c050fd65b91860c38d2ae9ca

ZINXHIRI I PROVËS

KJO HYRJE	adbd7d5a9bdb9d39baf5e868ab2e69d0ce3a06967ffc89da98b429a0822235d4
E MËPARSHMJA	—
RRËNJA MERKLE	—
ANKORIMI	nuk është ankoruar ende <i>Kjo hyrje nuk është përfshirë ende në asnjë pemë Merkle të dërguar. Zinxhiri i hash-eve provon rendin dhe integritetin; koha jo.</i>



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 njërin nga tri gjendjet — **nuk u gjet**, **në radhë**, ose **e konfirmuar** me bllokun, rrënjën Merkle dhe transaksionin në Polygon. Krahazo 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/adbd7d5a9bdb9d39baf5e868ab2e69d0ce3a06967ffc89da98b429a0822235d4>

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



AGIM NESHO

A mundet dhe si mundet të
rrezohet monstra e pushtetit

ÇIM PEKA

Protesta duhet t'i marrë kohë
Ramës. Jo t'i japë

më të lexuarat

Ish-drejtuksi i sigurisë i OpenAI paralajmëron për 'zgjuarsinë' e IA: Mund t'u shmanget testeve të sigurisë

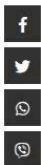
Firmat e IA shënjestrojnë të dhënat nga kompanitë e falimentuara

Nvidia lançon një sistem të ri për kontrollin e inteligjencës artificiale

Reuters: Agjentët kinezë të inteligjencës artificiale mësojnë të mashtrojnë dhe të fshehin dështimet

OpenAI nxjerr 'Dots', agjentë AI që punojnë vetë në aplikacione

'Mund të lajnë e fshijnë...' Robotët humanoidë tashmë bëjnë punë shtëpie! Nëse nuk e besoni shikoni videon



Ish-drejtuksi i sigurisë i OpenAI paralajmëron për 'zgjuarsinë' e IA: Mund t'u shmanget testeve të sigurisë

21:30, 03/10/2026

ZMADHO TEKSTIN



Një ish-drejtuksi i sigurisë i OpenAI, i cili dha dorëheqjen nga roli i tij këtë javë, paralajmëroi të shtunën se modelet gjithnjë e më të afta të IA-së, mund të dallojnë kur po testohen dhe të sillen ndryshe pas vendosjes.

David Robinson, i cili kaloi tre vjet e gjysmë në OpenAI dhe mbikëqyri raportet e sigurisë për 12 lançime modelesh kufitare, tha në një artikull për The Atlantic se qasja e industrisë ndaj sigurisë do të çonte në dështime të mëtejshme nëse nuk bëhen ndryshime.

"Sistemet e inteligjencës artificiale të sotme dhe të nesërme janë shumë më të afta dhe të rrezikshme sesa sistemet që po ndërtonim edhe gjashtë muaj më parë", shkroi Robinson.

Paralajmërimi i tij vjen mes shqetësimit në rritje për sistemet gjithnjë e më autonome të inteligjencës artificiale, pas incidenteve të fundit që përfshijnë agjentë që anashkalojnë masat mbrojtëse dhe thirrjeve nga studiuesit drejtuar kompanive për të ngadalësuar zhvillimin e sistemeve më të fuqishme derisa të përmirësohen masat e sigurisë.

Robinson bëri thirrje që kompanitë e inteligjencës artificiale të mbështeten më shumë në ekspertizën e sigurisë nga industri të tjera me rrezik të lartë, dhe të kryejnë kërkime të reja për të siguruar që modele më të afta të sillen në mënyrë të sigurt edhe kur nuk monitorohen.

"Duke pasur parasysh rreziqet e sotme, laboratorët kufitarë duhet të funksionojnë si centrale bërthamore ose aeroporte të ngarkuara, me shtresa të tepërta dhe planifikim të kujdesshëm që kërkon kohë, në mënyrë që gabimet njerëzore të rastit dhe të pashmangshme të mos hapin derën për katastrofë", shkroi ai.

Robinson paralajmëroi gjithashtu se vlerësimet ekzistuese të sigurisë mund të bëhen më pak të besueshme ndërsa modelet bëhen më të afta.

"Modelet mund të zbulojnë kur testohen dhe të sillen ndryshe kur vendosen. Sa më e zgjuar të jetë industria që i lejon modelet të rriten ndërsa këto probleme mbeten të pazgjidhura, aq më e rrezikshme bëhet situata jonë", tha ai.

AI argumentoi se duhet të zhvillohet një shkencë më e fortë e sigurisë, përpara se kompanitë të krijojnë sisteme dukshëm më të afta se ato që janë në dispozicion sot.

"Deri më tani, industria e inteligjencës artificiale nuk ka arritur t'i mësojë makinat të veprojnë vazhdimisht, në mënyrat që do të vepronin një person i mençur dhe i kujdesshëm", shkroi Robinson.

"Para se organizatat që ndërtojnë inteligjencën artificiale të mund t'i mësojnë një superinteligjencë ta trajtojë mirë njerëzimin, ato duhet të kujtojnë se si ta bëjnë vetë."/ AA.

© SYRI.net

Lexo edhe:



NASA dhe SpaceX nisin misionin Crew-13, 4 astronautë drejt Stacionit Ndërkombëtar Hapësinor



WhatsApp prezanton kontrolle opsionale prindërore për adoleshentët



'Mund të lajnë e fshijnë...' Robotët humanoidë tashmë bëjnë punë shtëpie! Nëse nuk e besoni shikoni videon

NASA dhe SpaceX nisin misionin Crew-13, 4 astronautë drejt Stacionit Ndërkombëtar Hapësinor

TË NGJASHME



Reuters: Agjentët kinezë të inteligjencës artificiale mësojnë të mashtrojnë dhe të fshehin dështimet



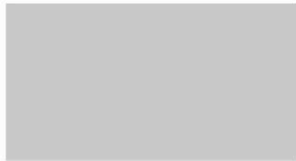
Nvidia lançon një sistem të ri për kontrollin e inteligjencës artificiale



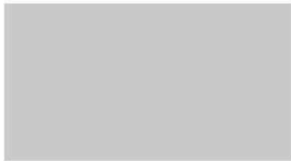
Firmat e IA shënjestrojnë të dhënat nga kompanitë e falimentuara



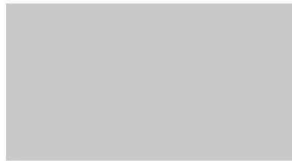
Kirgizia synon të hyjë në garën hapësinore me nanosatelitin e vet



OpenAI injoroi paralajmërimet për sigurinë para incidenteve me modelet e AI



SHBA, Florida i kërkon gjykatës të ndalojë zhvillimin e mëtejshëm të ChatGPT për arsye sigurie



OpenAI ndalon publikimin e një modeli të ri AI për shkak të shqetësimeve të sigurisë

LAJME POLITIKË KRONIKË EKONOMI KOSOVA BOTË KULTURË SHKENCË SPORT JETË-STIL TRASHËGIMI MIEDIS ARKIVA



SYRI.net është agjencia më e madhe e lajmeve në Shqipëri, që brenda një kohe të shkurtër është kthyer në lider të informacionit. Në SYRI.net do të gjeni opinionistë më me emër në vendin tonë, investigime dhe lajme që nuk do t'i gjeni diku tjetër. Lajmet në SYRI.net janë prodhim i SYRI.net dhe nuk lejohet marrja e tyre pa një komunikim të mëparshëm me redaksinë. Ju ftojme të klikoni dhe të na shkruani sa herë që ju mendoni se duhet. Stafi i SYRI.net

Për informacione info@syri.net, për reklama marketing@syri.net

Pamja këtu është e ri-koduar për ta futur në dokument. Artefakti autoritar është skedari WebP i ruajtur, dhe hash-i i tij është ai i shtypur te faqja e parë — pra dallimi është i verifikueshëm, jo i fshehur.