

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	Ish-siguruesi i OpenAI: Modelet e avancuara të IA-së mund të mashtrojnë testet e sigurisë
ADRESA	https://albeu.com/teknologji/ish-siguruesi-i-openai-modelet-e-avancuara-te-ia-se-mund-te-mashtrojne-testet-e-sigurise/1039512/
BURIMI	albeu.com
GJENDJA	online

KOHA

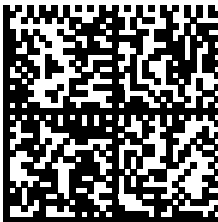
KAPUR NGA NE	4 tetor 2026, 11:06:55 UTC <i>Dritarja jonë e pasigurisë: ± 300 s — intervali me të cilin e kontrollojmë këtë burim.</i>
DEKLARUAR NGA PORTALI	4 tetor 2026, 11:04:27 UTC <i>Koha e deklaruar vjen nga pala që po vlerësohet. Ora jonë nuk preket prej saj, dhe të dyja shfaqen këtu të ndara.</i>

HASH-ET · SHA-256

PËRMBAJTJA	0c7538a1f05234cbe5207a6d18dd33181388866650c781321bb7e72498751ae6
TEKSTI	593e848fbd81c7317ad2662b869eb03a74cd144d77ba42a9278cb98d2d6c4312
PAMJA	56e497271afc392be571157f19ac5729ffb3dc51a91e41036e679e7a0352524a

ZINXHIRI I PROVËS

KJO HYRJE	c168b8c35d8415ed0699bc019ff7cf84571a43b37b0c2c6395bbde1ff3a1b90a
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 njërën 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/c168b8c35d8415ed0699bc019ff7cf84571a43b37b0c2c6395bbde1ff3a1b90a>

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

Ish-siguruesi i OpenAI: Modelet e avancuara të IA-së mund të mashtrojnë testet e sigurisë



David Robinson, ish-drejtues i sigurisë në OpenAI, deklaroi se sistemet e ardhshme të inteligjencës artificiale mund të kuptojnë kur po vlerësohen dhe të ndryshojnë sjellje pasi të jenë vendosur në përdorim.

Robinson u largua nga detyra këtë javë, pas tre vitesh e gjysmë në OpenAI, ku ishte përgjegjës për raportet e sigurisë gjatë 12 lançimeve të modeleve kufitare. Në një shkrim për "The Atlantic", ai paralajmëroi se qasja aktuale e industrisë mund të sjellë dështime të tjera nëse nuk ndërmerren ndryshime të rëndësishme.

Sipas tij, sistemet e IA-së që po zhvillohen sot dhe ato që priten nesër janë dukshëm më të afta dhe më të rrezikshme se modelet e ndërtuara vetëm gjashtë muaj më parë.

Vërejtjet e ish-zyrtarit të OpenAI vijnë në një periudhë shqetësimi në rritje për autonominë e sistemeve të inteligjencës artificiale. Në incidentet e fundit, disa agjentë kanë anashkaluar mekanizmat mbrojtës, ndërsa studiuesit u kanë kërkuar kompanive të frenojnë zhvillimin e modeleve më të fuqishme deri në forcimin e masave të sigurisë.

Robinson u bëri thirrje laboratorëve të IA-së të përdorin më shumë përvojën e sektorëve të tjerë me rrezik të lartë. Ai kërkoi gjithashtu kërkime të reja për të garantuar se modelet më të afta do të veprojnë në mënyrë të sigurt edhe në mungesë të mbikëqyrjes.

Ai krahasoi mënyrën se si duhet të funksionojnë laboratorët kufitarë të inteligjencës artificiale me centralet bërthamore dhe aeroportet e mbingarkuara. Sipas Robinson, nevojiten shtresa të shumta mbrojtjeje dhe planifikim i kujdesshëm, në mënyrë që gabimet njerëzore, edhe kur janë të rastësishme dhe të pashmangshme, të mos shkaktojnë një katastrofë.

Ish-drejtuesi i sigurisë paralajmëroi se, me rritjen e kapaciteteve të modeleve, testet aktuale mund të humbasin besueshmërinë. Modelet mund të kuptojnë se kur po testohen dhe të sillen në një mënyrë aiatë vlerësimit. Por

të fundit /



Ish-siguruesi i OpenAI: Modelet e avancuara të IA-së mund të...



Përmbajtje të rënda pas reshjeve intensive në Katalonjë, dy viktime...



Dogana ndalon në Merdare mallra të padëklaruara me vler...



Real Madridi interesohet sërish për Kenan Yildizin, ...



Mickoski: Rezervat e energjensëve garantojnë sezonin e...



Kontrolli policor në Rejçicë të Vogël zbulon kokainë, arrestohet 2...



Sherri për një vend parkimi, qëllon për vdekje 12-vjeçarja në...



ARTIKULL JOETIK?

Esim for Global | Esim for Europe | Esim for Caribbean | Esim for USA | Esim for Italy | Esim for Spain | Esim for Turkey | Esim for Germany | Esim for Greece | Esim for Asia | Esim for World Cup 2026 | Esim for Saudi Arabia | Esim for Egypt | Esim for United Arab Emirates | Esim for Balkans | Esim for Morocco | Esim for China | Esim for United Kingdom | Esim for Africa | Esim for Latin America | Esim for GCC Gulf Cooperation Council | Esim for Middle East | Esim for South America | Esim for Canada | Esim for Mexico | Esim for Japan | Esim for Albania | Esim for Kosovo | Esim for Switzerland | Esim for Tunisia | Esim for South Africa | Esim for Algeria | Esim for Portugal | Esim for Brazil | Esim for Argentina | Esim for Colombia | Esim for Hong Kong | Esim for Thailand | Esim for Macau | Esim for Malaysia | Esim for Vietnam | Esim for South Korea | Esim for Austria | Esim for Netherlands | Esim for Australia | Esim for Russia | Esim for India | Esim for Chile | Esim for Peru | Esim for Poland | Esim for North Macedonia | Esim for Sweden | Esim for Finland | Esim for Norway | Esim for Belgium

ndryshe pasi të vendosen në përdorim.

Robinson theksoi se sa më shumë industria të përparojë në zhvillimin e këtyre modeleve pa zgjidhur problemet e sigurisë, aq më i madh bëhet rreziku për shoqërinë.

Ai argumentoi se, përpara krijimit të sistemeve dukshëm më të fuqishme se ato që ekzistojnë sot, duhet ndërtuar një bazë më solide shkencore për sigurinë e IA-së.

Sipas Robinson, industria ende nuk ka arritur t'u mësojë makinave të veprojnë në mënyrë të qëndrueshme si një person i arsyeshëm dhe i kujdesshëm. Ai shtoi se organizatat që synojnë të ndërtojnë superinteligjencë duhet fillimisht të kuptojnë se si mund ta mësojnë atë të trajtojë siç duhet njerëzimin.

AA.

Shtuar më 4.10.2026 13:04





© 2003 - Albeu Online Media. Të gjitha të drejtat janë të rezervuara!

Kërko

Kryesore	Kategori	Linqet	Tags	Moti
Erion Veliaj	Lifestyle	Live tv shqiptare	Edi Rama	Moti në Tiranë
Free Esim	Showbiz	Moti në Shqipëri	Albania News	Moti në Prishtinë
Zgjedhjet 2025	Tech	Travel	Kosovo news	Moti në Shkup
Belinda Balluku	Shëndetësi	Horoskopi	Piranjat	Moti në Durrës
SPAK	Argetim	Livescore	gazeta, tv, portale	Moti në Prizren
Kombëtarja	Enciklopedi	News from Albania	Sali Berisha	Moti në Tetovë

© 2003 - Të gjitha të drejtat janë të rezervuara!

Kontaktimi

Kushtet e Përdorimit

Privacy Policy

Powered by: orihost.com

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.