TRAFICOM
Liikenne- ja viestintävirasto
Kyberturvallisuuskeskus

Yrityksen varautuminen ja riskienhallinta kyberturvallisuuden näkökulmasta

Kansallinen kyberturvallisuuden tilannekuva

Suomalaisiin tietojärjestelmiin kohdistuvien tietoturvaloukkausten ja -uhkien selvittäminen

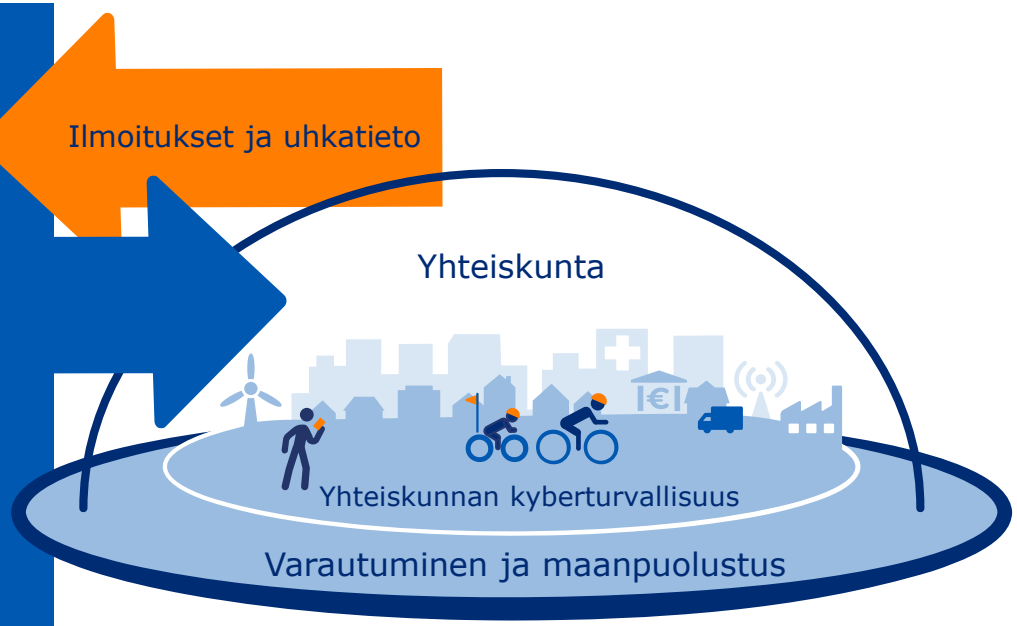
Varautumisen, tieto-turvallisuuden sekä luottamuksellisen viestinnän suoja ohjaus ja valvonta.

Julkisesti säännellyn satelliittipalvelun (Galileo PRS) vastuuviranomainen

Tietojärjestelmien ja salaustuotteiden arviointi ja hyväksyntä sekä salausteknisen materiaalin jakelu

Euroopan kyberturvallisuuden teollisuus-, teknologia- ja tutkimusosaamiskeskuksen kansallinen koordinointikeskus

ohjaus
valvonta
tuotteet
palvelut
havainnointi
varoitukset
arvioinnit
rahoitus
osaaminen
verkostot



Traficomin Kyberturvallisuuskeskus on mukana rakentamassa maailman toimivinta ja turvallisinta digitaalista yhteiskuntaa

Perustettu	Budjetti	Henkilöstö
2014	**~26 M€**	**190**
Verkosto-tapaamisia	Tarkastuksia	Julkaistuja tiedotteita
~440	**~40**	**~150**

Automaattisesti käsiteltyjä tapahtumia	Käsiteltyjä tietoturvapoikkeamia
~200 000	**~19 000**

2025

- Kyberuhkataso on pysynyt kohonneena.
- Aktiivisia toimia vaatineiden merkittävien tapausten määrät ovat nousseet viime vuosina.
- Suomi oli vuonna 2025 kohtuullisen hyvin valmistautunut kyberuhkiin.

Vakavia ja kriittisiä poikkeamia on selvitettävänä aiempaa enemmän

Kyberturvallisuuden uhkataso säilyy kohonneena

Kyberuhat ovat osa arkipäivää verkottuneessa yhteiskunnassa.

- Teknologinen kehitys
- Digitalisaatio
- Kansainvälisyys

Kriittiset haavoittuvuudet

Tekoäly ja automatisaatio nopeuttavat haavoittuvuuksien hyväksikäyttöä.

Tietomurrot ja niiden yritykset

Tietojenkalastelu ja huijaukset

Viranomaisten tai pankkien nimissä tehdyt huijaukset rapauttavat luottamusta.

Valtiollinen kybertoiminta

Muun muassa Venäjä ja Kiina kohdistavat Suomeen kybervakoilua ja -vaikuttamista.

Haktivismi

Ideologiset toimijat, myös valtioiden tukemina.

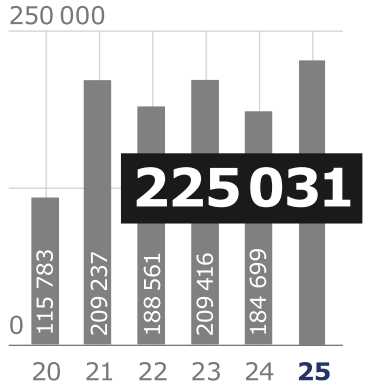
Kyberrikollisuus

Myös osa kyberrikollisuudesta on aiempaa kohdennetumpaa ja pitkäjänteisempää.

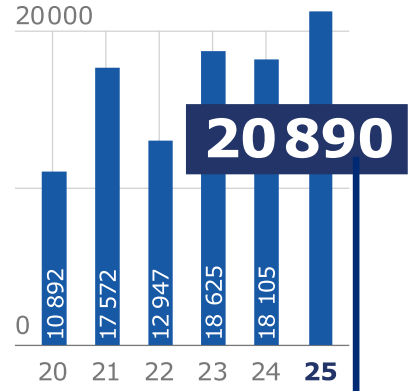
Kohteina yritykset, julkinen sektori ja kansalaiset

2025 lukuina

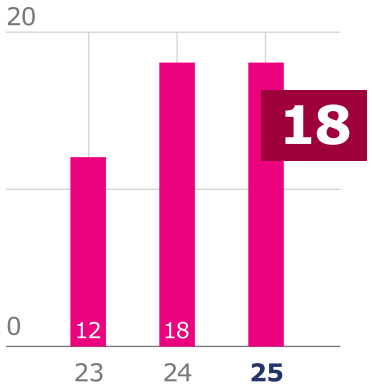
Automaattisesti käsitellyt havainnot



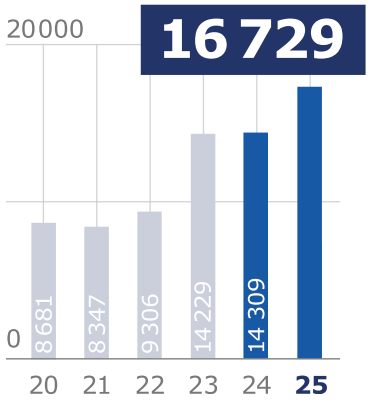
Kaikki kyberpoikkeama-ilmoitukset yhteensä



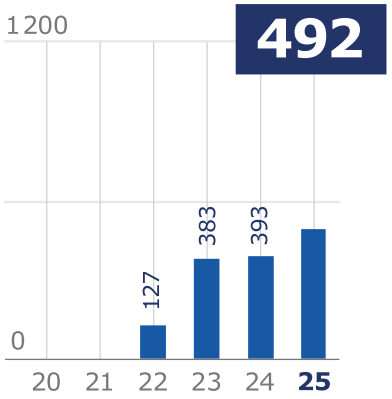
Merkittävät kyberpoikkeaman-hallintaoperaatiot



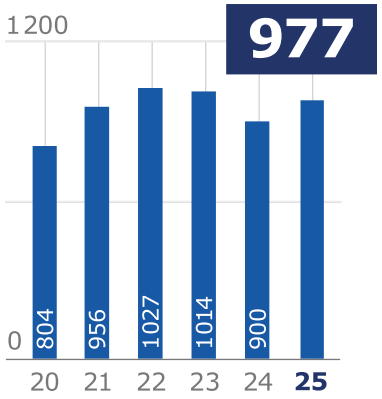
Huijaukset ja kalastelut[1]



Tietomurron yritykset[2]

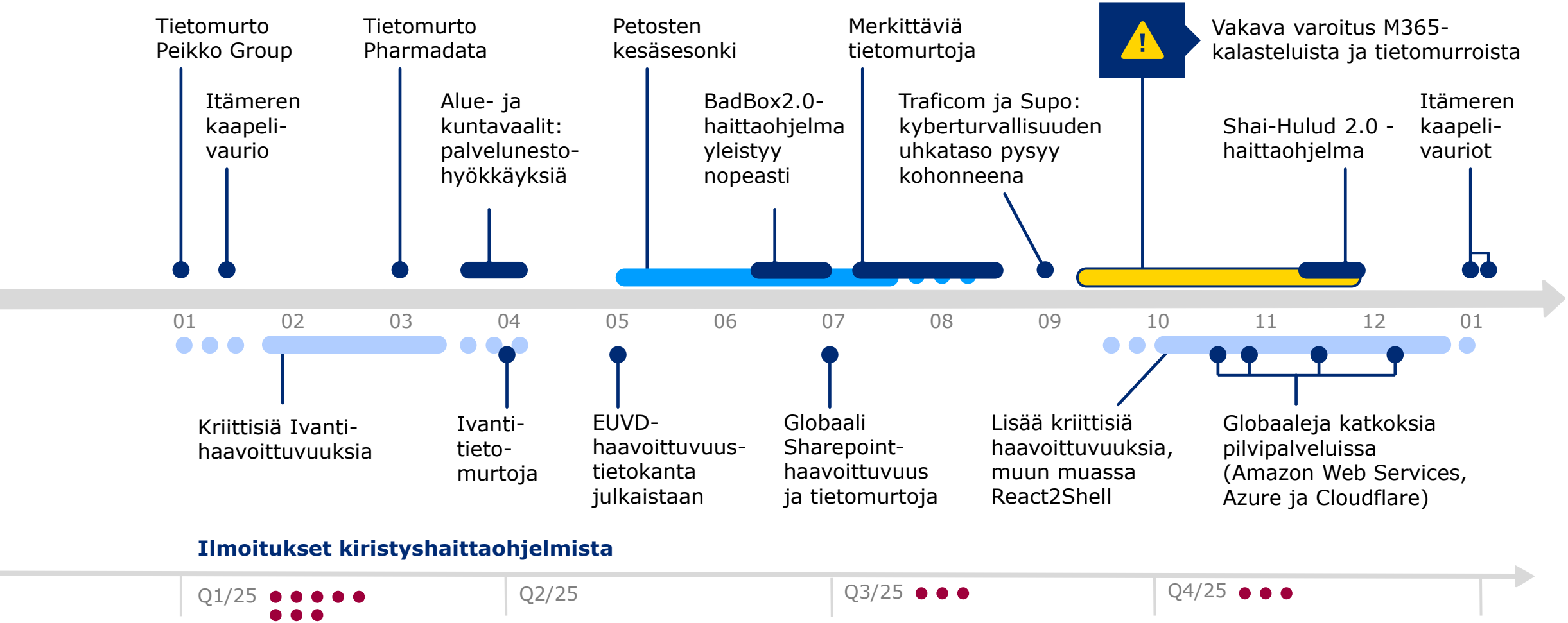


Tietomurrot



LÄHDE: Kyberturvallisuuskeskus | [1] Huijausten ja kalastelujen tilastointi on muuttunut vuonna 2024, joten luvut eivät ole täysin vertailukelpoisia. | [2] Tietomurtoyritysilmoituksia on raportoitu alkaen vuodesta 2022.

Kyberturvallisuuskeskus seuraa toimintaympäristöä



Ilmoitukset kiristyshaittaohjelmista

2025 | YHTEENVETO
Keskeiset teemat

HAAVOITTUVUUDET
Merkittävä haaste

KRIITTINEN INFRASTRUKTUURI
Globaali turvallisuushuoli

ARTIFICIAL INTELLIGENCE
- Tekoälyn käyttöönotto ja hyödyntäminen organisaatioissa
- Tekoälyn hyödyntäminen haitallisissa tarkoituksissa myös yleistynyt

PILVIPALVELUIDEN TURVALLISUUS
- Pilvipalveluiden viranomaiskäyttö herättänyt keskustelua

TOIMITUSKETJUJEN TURVALLISUUS
- Ohjelmistoekosysteemit
- Riskienhallinta

KIRISTYSHAITTAOHJELMAT
- Tapausmäärät laskeneet Suomessa
- Reunalaitteet
- Varmuuskopiot

Suomi oli melko hyvin varautunut, mutta tehtävää riittää

Kyberturvallisuuden kehittäminen on pitkäjänteistä yhteistyötä.

- Toimitusketjujen turvallisuus
- Riskienhallinta
- Harjoittelu
- Kyberturvatietoisuuden laajentaminen käytännön liiketoimiin

Yleinen tietoisuus on lisääntynyt ja häiriötilanteisiin on varauduttu

Uhkista tiedetään enemmän.
Varautumissuunnitelmia on laadittu.

Lainsäädäntö tukee varautumista

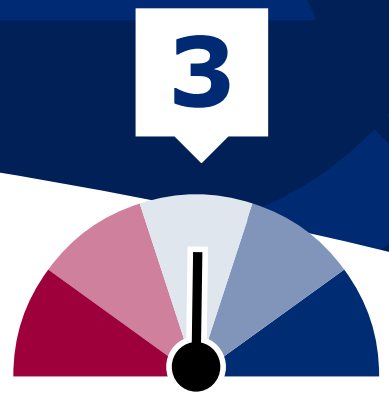
Kriittisten toimijoiden ilmoitusvelvollisuus parantaa tilannekuvaa.

Johdon tuki, resurssit ja rakenteet ovat merkittäviä tekijöitä kyberturvallisuuden kehittymiselle.

Korkean kypsyystason yrityksissä varautuminen on liiketoiminta- ja riskilähtöistä. Johto ja kyberturvallisuus-vastaavat kommunikoivat sujuvasti.

Suomalaisten yritysten kyberturvallisuuden kypsyystaso on edelleen melko hyvällä tasolla.

(Huoltovarmuuskeskus 2025)



Kybersää

Kybersää kertoo kuukauden merkittävistä tietoturvapoikkeamista ja -ilmiöistä.

Tämä tuote on suunnattu ensisijaisesti eri tasoilla organisaatioiden tietoturvallisuuden parissa työskenteleville. Kybersää tarjoaa nopean kokonaiskuvan, mitä kyberturvallisuuskentällä on tapahtunut ja mitä on tulossa.

Kybersää voi olla:  rauhallinen  huolestuttava  vakava

Kuukauden tunnuslukuja

Keväällä 2026 satelliittinavigoinnin (GNSS) ja matkaviestinverkkojen häiriöt sekä Traficomin tekemät häirintähavainnot ovat lisääntyneet. GNSS-häiriöitä havaitaan erityisesti ilmailussa eteläisen ja keskisen Suomen alueella kaikkina vuorokauden aikoina.[1]

Tekstiviestien lähettäjien tunnistaminen tiukentui 4.5.2026, kun viestien lähettäjien tunnistaminen ja lähettäjätunnusten käyttöoikeuden varmistaminen tuli pakolliseksi organisaatioille, jotka lähettävät tekstiviestejä kansalaisille. Uudistetun määräyksen tavoitteena on estää organisaatioiden nimissä lähetettäviä huijausviestejä.[2]

EU käynnistää tekoälyyn, drooneihin, robotiikkaan ja kvanttiteknologiaan perustuvan AGILE-ohjelman nopeuttamaan Euroopan puolustusteknologian kehitystä. Ohjelman 115 miljoonan euron pilottivaiheen rahoitus suunnataan erityisesti startupeille ja pk-yrityksille. Ohjelmalla varmistetaan nopea rahoitusprosessi sekä mahdollisuus siirtyä kenttätestaukseen 1–3 vuodessa.[3, 4]

Kybersään yleistilanne huhtikuussa 2026

Huhtikuu ei tuonut mukanaan merkittävää muutosta kybersäähän

Lämpenevää kevätsäätä viilensivät erityisesti M365-tilimurrot, joita ilmoitettiin Kyberturvallisuuskeskukselle edellistä kuukautta enemmän. Kuukauden aikana maahan ropisi myös sadepisaroita useissa eri tuotteissa julkaistujen haavoittuvuuksien vuoksi.

Tekoälypohjaisten ratkaisujen hyödyntäminen haavoittuvuuksien kartoittamisessa ja hyväksikäytössä nousi kuukauden puheenaiheeksi.

Tekoälyn hyväksikäytön on havaittu yleistyneen myös petoksissa.

- Kyberturvallisuuskeskuksen tiedossa on toistaiseksi vain yksittäisiä tapauksia, joissa on käytetty tekoälyllä tuotettua ääntä ja kuvaa esimerkiksi toimitusjohtajahuijauksissa ja sijoituspetoksissa.

Traficom julkaisi vuoteen 2035 sijoittuvat kyberturvallisuuden skenaariot, jotka kuvaavat neljää vaihtoehtoista tulevaisuutta

Kyberturvallisuuden näkökulmasta ratkaiseviksi kysymyksiksi nousevat kaikissa skenaarioissa tekoäly, toimitusketjut, informaatioympäristön luotettavuus sekä kriittisen infrastruktuurin kytkeytyneisyys.

Skenaariot tukevat varautumista, päätöksentekoa ja strategista keskustelua tilanteessa, jossa tulevaisuuden digitaalisen yhteiskunnan turvallisuus rakentuu yhä monimutkaisempien riippuvuuksien varaan. [5, 6]

Traficomin verkkosivuja uudistettiin

Olemme uudistaneet Kyberturvallisuuskeskuksen verkkosivuja osana Traficomin verkkosivualustan uudistusta. Muutos parantaa sivustojen toimivuutta ja mahdollistaa niiden kehittämisen jatkossa. Sivustoilla voi esiintyä yksittäisiä puutteita, kuten rikkinäisiä linkkejä. Korjaamme niitä jatkuvasti. Pahoittelemme uudistuksesta koituneita väliaikaisia häiriöitä. [7]

Kuukauden raekuuro
AI-pohjainen haavoittuvuusskannaus muuttaa pelikenttää

AI-pohjainen haavoittuvuusskannaus nousi laajaksi puheenaiheeksi huhtikuussa sen tarjoamien kyvykkyyksien vuoksi.

Kehittyneiden AI-ratkaisujen on arvioitu tehostavan pahantahtoista haavoittuvuuksien kartoitusta ja lisäävän hyväksikäytettävien haavoittuvuuksien määrää. Tekoälyn avulla voidaan löytää uusia ennalta tuntemattomia haavoittuvuuksia ja tekoäly pystyy käyttämään niitä itsenäisesti hyväksi.

AI-ratkaisuissa hyökkäyspolkujen mallinnus (attack path analysis) helpottuu myös huomattavasti. AI-ratkaisuja hyödyntämällä hyökkääjän on mahdollista löytää ja ketjuttaa erilaisia haavoittuvuuksien hyväksikäytön mahdollistavia tekijöitä, kuten tunnistus-, konfiguraatio- ja logiikkavirheitä kohteena olevassa verkkoympäristössä.

Tämä haastaa perinteisiä skannausmenetelmiä, jotka eivät välttämättä tunnista löydettyjen haavoittuvuuksien kokonaisriskiä.

Kehittyneiden AI-ratkaisujen ei kuitenkaan uskota täysin korvaavan perinteistä tunnisteisiin pohjaavaa haavoittuvuusskannausta, vaan ne muuttavat toiminnan luonnetta itsenäisemmäksi, nopeammaksi sekä kokonaisvaltaisemmin uhkia ja riskitekijöitä huomioivaksi suorituskyvyksi.

AI-pohjaiset sovellukset ovat siten samalla mahdollisuus organisaatioiden riskienhallinnan ja kyberturvallisuuden parantamiseksi. Esimerkiksi joidenkin ennusteiden mukaan tekoälysovelluksilla voidaan peräti kattaa puolet nykyisistä kyberhyökkäysten suojaus- ja torjuntatoimista vuoteen 2028 mennessä.[8]

Kyberturvallisuuskeskuksen toimenpiteet ja vinkit varautumiseen

Ilmoittaudu mukaan Traficomin, liikenne- ja viestintäministeriön sekä Kyberala ry:n 3. kesäkuuta järjestämään EU:n kyberkestävyyssäädöksen (CRA) voimaantuloa koskevan infotilaisuuteen. Tapahtumassa EU-komission, kansallisten viranomaisten ja yritysten edustajat tuovat esiin näkökulmia sääntelyn sisältöön, velvoitteisiin ja sen kansalliseen toimeenpanoon. [9]

Organisaatioita kannustetaan siirtymään tietojenkalasteluille resistentteihin menetelmiin, kuten FIDO2/WebAuthn tai varmennepohjaiseen tunnistautumiseen. Perinteinen monivaiheinen tunnistautuminen (MFA) kyetään yhä useammin ohittamaan Adversary-in-the-Middle (AiTM) -hyökkäyksillä, OAuth-väärinkäytöillä ja istuntotunnisteiden varastamisella. [10, 11]

FINMISP-palvelu on julkaistu! FINMISP on Kyberturvallisuuskeskuksen tarjoama kansallinen kyberuhkatiedon jakopalvelu, joka perustuu MISP-alustaan (Malware Information Sharing Platform). Palvelun avulla tehostetaan teknisen uhkatiedon jakamista kansallisesti ja kansainvälisesti havaituista tietoturvapoikkeamista. Kyberturvallisuuskeskus toimii verkoston keskuksena ja jakaa tietoa palvelun asiakkaille. [12]

Kybersään ilmiöt

Osiossa käymme läpi kyberturvallisuuden ilmiöiden kehitystä ja trendejä.

Kybersää
huhtikuu 2026

1 kk

Tietomurrot- ja vuodot

Tietomurtoja ilmoitettiin 14 % enemmän kuin maaliskuussa. Tietovuotoja raportoitiin kuitenkin huhtikuussa vähemmän. Useiden raportoitujen tietovuototapausten taustalla oli virheellinen konfiguraatio.

Haittaohjelmat

Kuukausi oli aktiivinen haittaohjelmahavaintojen osalta. Kyberturvallisuuskeskus sai ilmoituksia mm. ClickFix-tekniikalla levitetyistä haittaohjelmista, yksittäisistä Magecart- ja infostealer-haittaohjelmista.

Haavoittuvuudet

Julkaistujen haavoittuvuuksien määrä pysyi korkeana myös huhtikuussa. Verkkoon näkyvien laitteiden ja palveluiden nopea päivittäminen korostuu edelleen haavoittuvuuksien hyväksikäyttömahdollisuuksien vähentämiseksi.

Huijaukset ja kalastelut

Viranomaisviestintä siirrettiin huhtikuussa kokonaan suomi.fi-palveluun. Huijarit seuraavat tilannetta ja lähettävät linkkejä väärennettyihin palveluihin.

Hotelli- ja matkavarauspalveluhuijaukset yleistyvät lomakauden lähestyessä.

Automaatio ja IoT

Yhdysvaltain viranomaiset julkaisivat ohjeen nollaluottamusperiaatteiden soveltamisesta OT-järjestelmiin.

IoT-laitteet ja kuluttajatason verkkolaitteet ovat kiinnostavia hyökkäyksen kohteita myös valtiollisille toimijoille.

Verkkojen toimivuus

Tekoälypalveluita organisaation ulkopuolelle tarjotessa tulee huomioida mahdollisuudet niiden kuormittamiseen.

Tekoälyn hyödyntämisen myötä kasvavat määrät ohjelmistohaavoittuvuuksia voivat näkyä palvelunestohyökkäyksiin käytettävien bottiverkkojen kasvuna.

Kybersää
huhtikuu 2026 1/2

1 kk

Tietomurrot ja -vuodot

- Huhtikuussa ilmoitettiin selvästi edellistä kuukautta enemmän murrettuja M365-tilejä. Suurin osa tunnuksista kalasteltiin AiTM-tekniikalla, joten pelkkä MFA-suojaus ei enää riitä estämään tilimurtoja.
- Murretuilta tileiltä lähetettiin tuhansia jatkokalasteluviestejä, minkä arvioidaan lisäävän tietomurtoja edelleen toukokuussa.
- Useita WordPress-sivustoja murrettiin lisäosien haavoittuvuuksia hyödyntämällä. WordPressin ja lisäosien säännöllinen päivittäminen on tärkeää. Lisäksi kannattaa varmistaa, kuuluuko päivitysvastuu webhotellipalvelun tarjoajalle vai sivuston ylläpitäjälle.

Haittaohjelmat

- Magecart-haittaohjelmia havaittiin yksittäisissä verkkokaupoissa. Haittaohjelman avulla pyritään anastamaan verkkokauppaan syötettyjä henkilö- ja pankkitietoja.
- Kyberturvallisuuskeskus sai lisäksi ilmoituksia ClickFix-tekniikalla levitetyistä haittaohjelmista.
- Tietojenkalastelun ja haavoittuvuuksien avulla on pyritty levittämään myös haitallisia ohjelmistoja sekä infostealer-haittaohjelmia.
- Maaliskuussa tapahtuneet toimitusketjuhyökkäykset avoimien lähdekoodien kirjastoihin näkyvät edelleen hyökkäysvektorina haittaohjelmien levityksessä.

Haavoittuvuudet

- CVE-2026-31431 ”Copy Fail” -haavoittuvuus Linux-kernelissä, jonka avulla tavallinen käyttäjä voi saada pääkäyttäjän oikeudet.
- CVE-2026-41940 cPanel ja WHM -tuotteissa, jonka avulla autentikoimattoman hyökkääjän on mahdollista saada pääkäyttäjätason oikeudet hallintapaneeliin. Hyväksikäyttöä havaittu ja välitön päivittäminen suositeltua.
- CVE-2026-35616 FortiClient EMS haavoittuvuus, jonka avulla hyökkääjä voi ottaa haltuunsa laitteen. Haavoittuvuutta hyväksikäytetään aktiivisesti.

Kybersää
huhtikuu 2026 2/2

1 kk

Huijaukset ja kalastelut

- Viranomaisviestintä siirrettiin huhtikuussa kokonaan suomi.fi-palveluun. Huijarit seuraavat tilannetta ja lähettävät linkkejä väärennettyihin palveluihin. Verkkohuijauksen tunnistaminen helpottuu, kun ihmiset saavat tietoa huijauksista etukäteen. Tietoisuus antaa ihmisille mahdollisuuden pysähtyä ja arvioida, onko tekstiviesti, sähköposti tai puhelu aito vai huijaus. Älä seuraa tekstiviestien linkkejä vahvaan tunnistautumiseen.
- Hotelli- ja matkavarauspalveluhuijaukset yleistyvät lomakauden lähestyessä. Mieti kahdesti ja varmista palvelulta ennen kuin maksat yllättäviä ylimääräisiä maksuja.

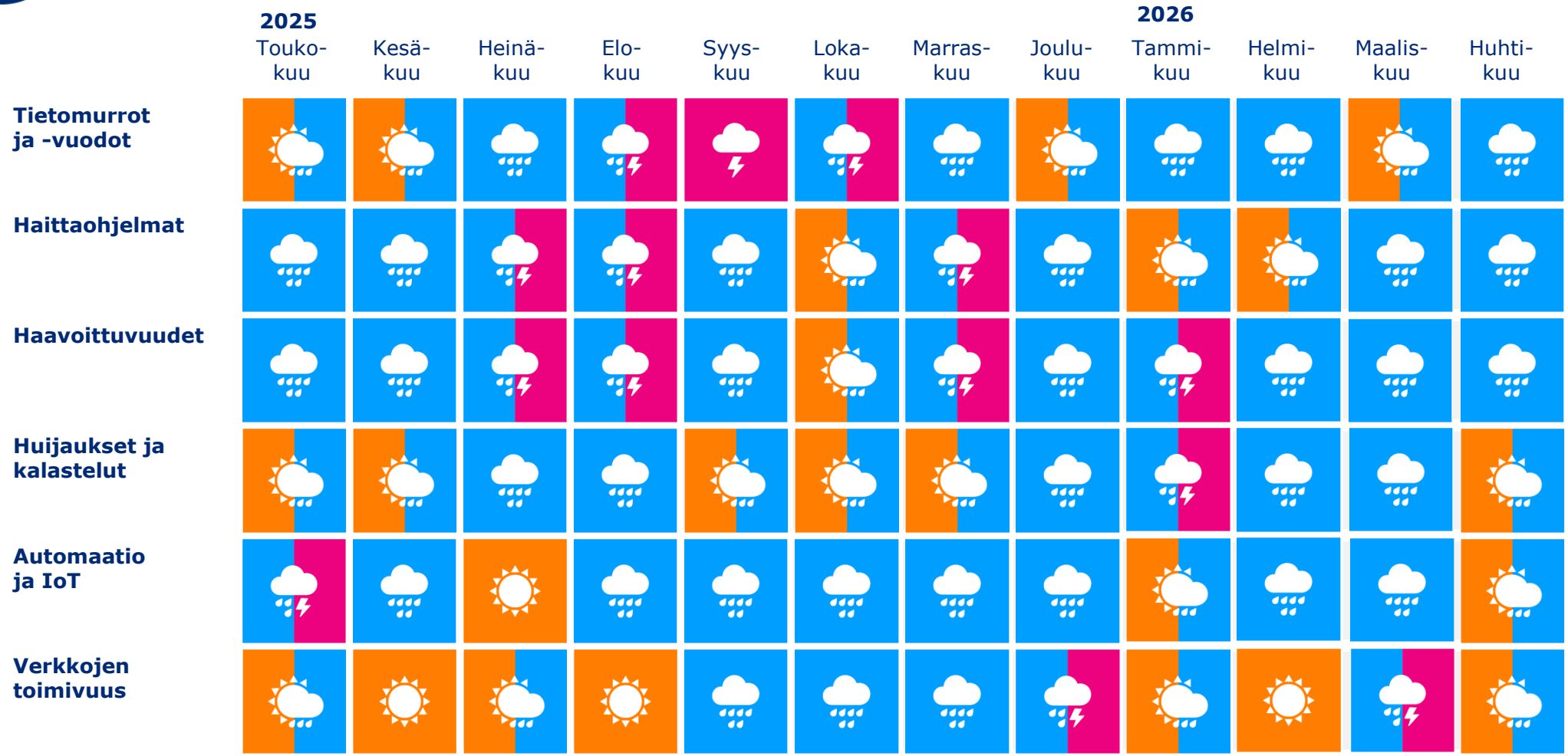
Automaatio ja IoT

- Yhdysvaltalainen viranomaistyöryhmä julkaisi dokumentin nollaluottamusperiaatteiden soveltamisesta OT-järjestelmiin ja -ympäristöihin.[13]
- Heikosti suojatut internetiin kytketyt kamerat ovat kelpo tiedonlähde myös tiedustelupalveluille. Esimerkiksi Iranin ajatolla Khamenein liikkeistä kerrotaan hankitun tietoa liikennekameroiden kuvista.[14]
- Yhdysvaltain kyberturvallisuusvirasto varoitti kiinalaisten kyberuhkatoimijoiden rakentavan peiteverkkoja murretuista IoT-laitteista.[15]

Verkkojen toimivuus

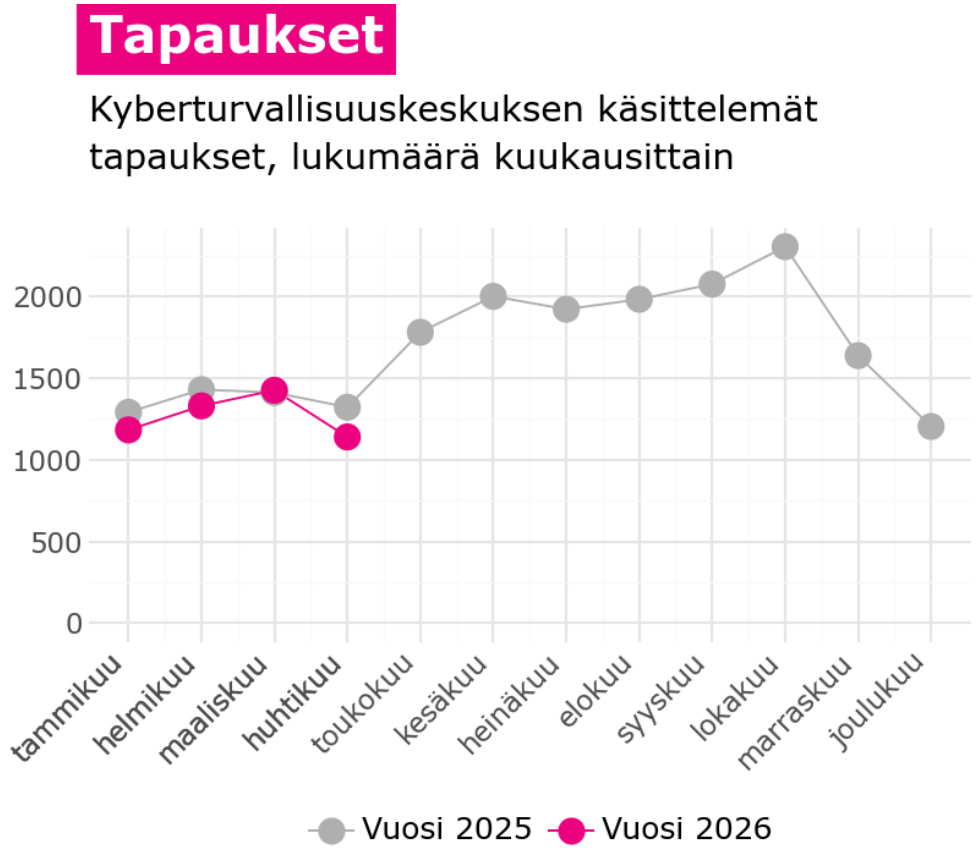
- Huhtikuussa yleisissä viestintäverkoissa ei havaittu vakavia toimivuushäiriöitä.
- Tekoälygeneroidulla sisällöllä voidaan myös ruuhkauttaa organisaation ulkopuolisille tarjoamia tekoälypohjaisia palveluita. Esimerkiksi verkkosivuilla asiakkaita avustavan chatbotin käyttöönotossa on huomioitava sen käyttämien resurssien rajoitukset.
- Palvelun ruuhkautumisen lisäksi massiiviset määrät syötteitä voivat aiheuttaa organisaatiolle ylimääräisiä kuluja.

Kybersään ilmiöt
kulunut 12 kk



Tapaukset

- Kyberturvallisuuskeskus käsitteli huhtikuussa 1140 tapausta.
- Ilmoitusten määrä oli 12 kuukauden keskiarvoon nähden vajaan kolmanneksen alhaisempi. Vaikka tapauksia ilmoitettiin määrällisesti vähemmän, ei niiden luonteessa ole havaittu merkittävää muutosta.
- Kyberturvallisuuskeskukselle ilmoitettujen tapausten osalta kuukausi oli puolipilvinen.
- M365-tilimurtoja havaittiin huhtikuussa edellistä kuukautta enemmän. Tilimurrot johtivat tietojen vaarantumiseen ja jatkokalasteluun useilla eri toimialoilla.
- Huhtikuu oli myös haavoittuvuuksien osalta erittäin aktiivinen.



Kybersääennuste

Kybersääennuste on aiempiin havaintoihin perustuva yhteenveto ja suuntaa-antava arvio lähikuukausien kyberuhista ja niiden kehityskuluista.

Osiossa käsitellään myös puolivuosittain ilmiöiden pitkän aikavälin kehitysnäkymiä ja lähitulevaisuuden top 5 kyberuhat.

Kybersääennuste
Kyberuhat pysyvät tavanomaisina

1 kk

Edelleen runsastuneet M365-tilimurrot ja murretuilta tileiltä lähetetyt jatkokalasteluviestit johtavat todennäköisesti tilimurtoihin myös toukokuussa.

Tekoälyteknologiat ja niiden soveltaminen kehittyvät nopeasti, mikä voi aiheuttaa äkillisiä muutoksia hyökkääjien toimintatavoissa. Toimintaympäristön jatkuva luotaaminen ja siihen mukautuminen on tärkeää organisaatioiden turvallisuuden kannalta.

Organisaation varautuminen

- Haavoittuvuuksienhallinnassa korostuu edelleen verkkoon näkyvien laitteiden ja palveluiden nopea päivittäminen haavoittuvuuksien hyväksikäytön vähentämiseksi.
- Tunne oma verkkoympäristösi, käyttämäsi järjestelmät ja niiden riippuvuudet, sekä korvaa elinkaarensa päähän tulevat järjestelmät ajoissa.
- Valistaminen ja monivaiheinen tunnistautuminen (MFA) eivät riitä suojaamaan työntekijöitä kehittyneiltä tilimurtoyrityksiltä, kuten AiTM-tekniikkaa käyttävältä kalastelulta.

Huolestuttava

Kyberuhkien määrä ja vakavuus ovat tavanomaisella tasolla.

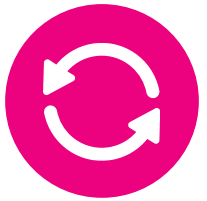
Kybersääennuste on aiempiin havaintoihin perustuva yhteenveto ja suuntaa-antava arvio kyberuhkien tilasta. Arviota ei tule käyttää sellaisenaan kyberuhkiin varautumisessa, vaan sen tukena on käytettävä organisaatiokohtaista tietoa ja analyysiä. Kyberuhat voivat muuttua nopeasti, myös negatiiviseen suuntaan.

2026

Arvio vuoden 2026 kyberuhkamaiseman kehityksestä

Kehityksen arviointi on haasteellista

- Merkittäviä globaaleja muutoksia, joilla vaikutuksia kansalliseen ja kansainväliseen kyberturvallisuuteen
- Varautuminen vaihtoehtoisiin ratkaisuihin sekä omavaraisuuden kasvattaminen ovat tarpeellista.
- Ukrainan tilanteen kehittyminen vaikuttaa Venäjän kyberkykyjen käyttöön.
- Valtiollisen kybertoiminnan monimuotoistuminen on todennäköistä.



Vuoden 2026 teemoja

TIETOJENKALASTELU JA TIETOMURROT

- M365-ongelman jatkuminen
- MFA ei riitä, tehokkaita teknisiä toimia tarvitaan

HAAVOITTUVUUDET

- Opportunististen hyökkäysten jatkuminen
- Viiveiden pienentyminen hyödyntämisessä
- OT-järjestelmien hyödyntämisen kehitys

AI JA TEKNOLOGIAKEHITYS

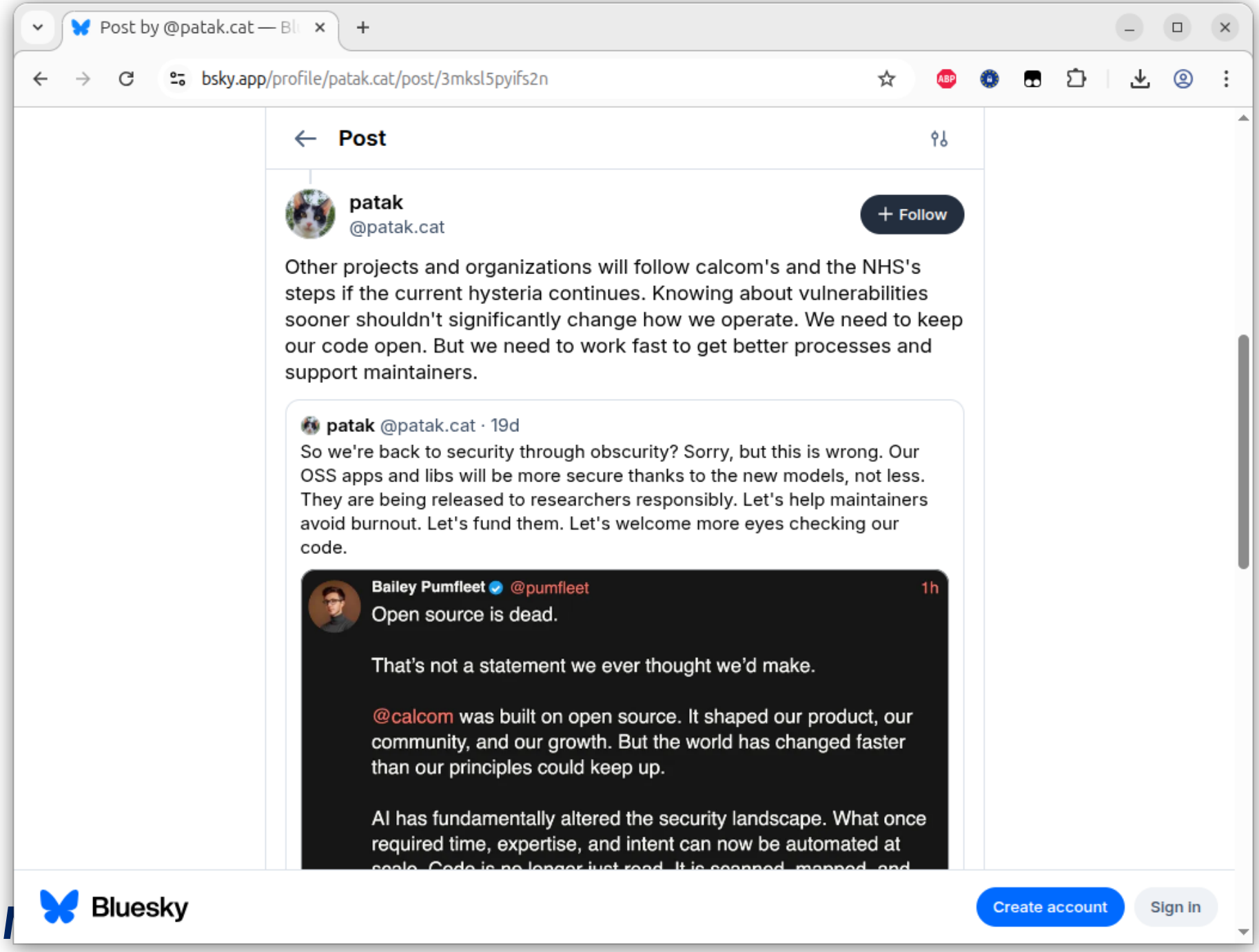
- AI-ominaisuuksien yleistyminen
- Varjo-AI ja varjo-IT
- AI:n hyödyntäminen haitallisessa toiminnassa

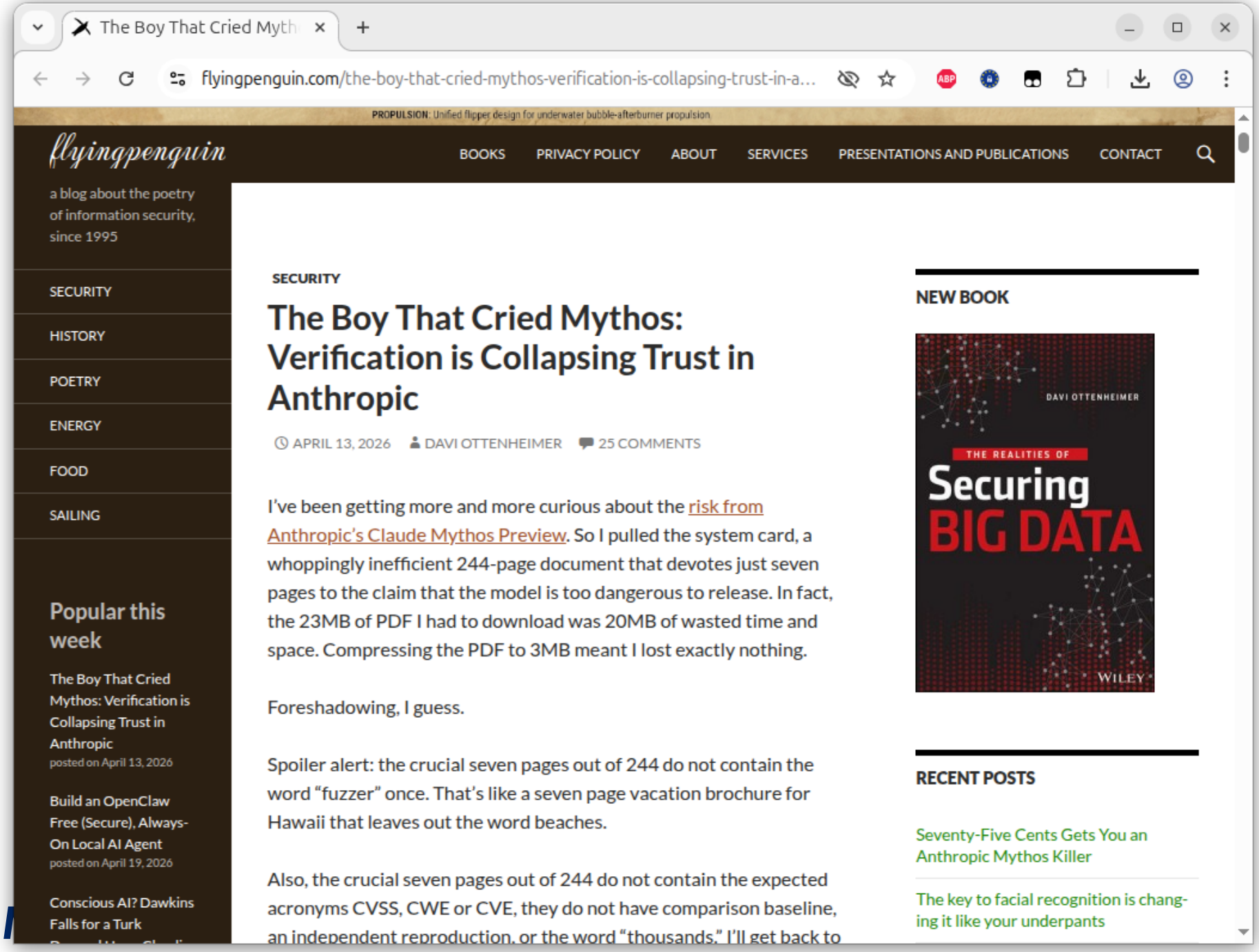
GLOBAALIN KEHITYKSEN VAIKUTUKSET KYBERTURVALLISUUTEEN

- Tarkka arviointi haasteellista
- Yllättävät muutokset
- Varautuminen vaihtoehtoisiin ratkaisuihin

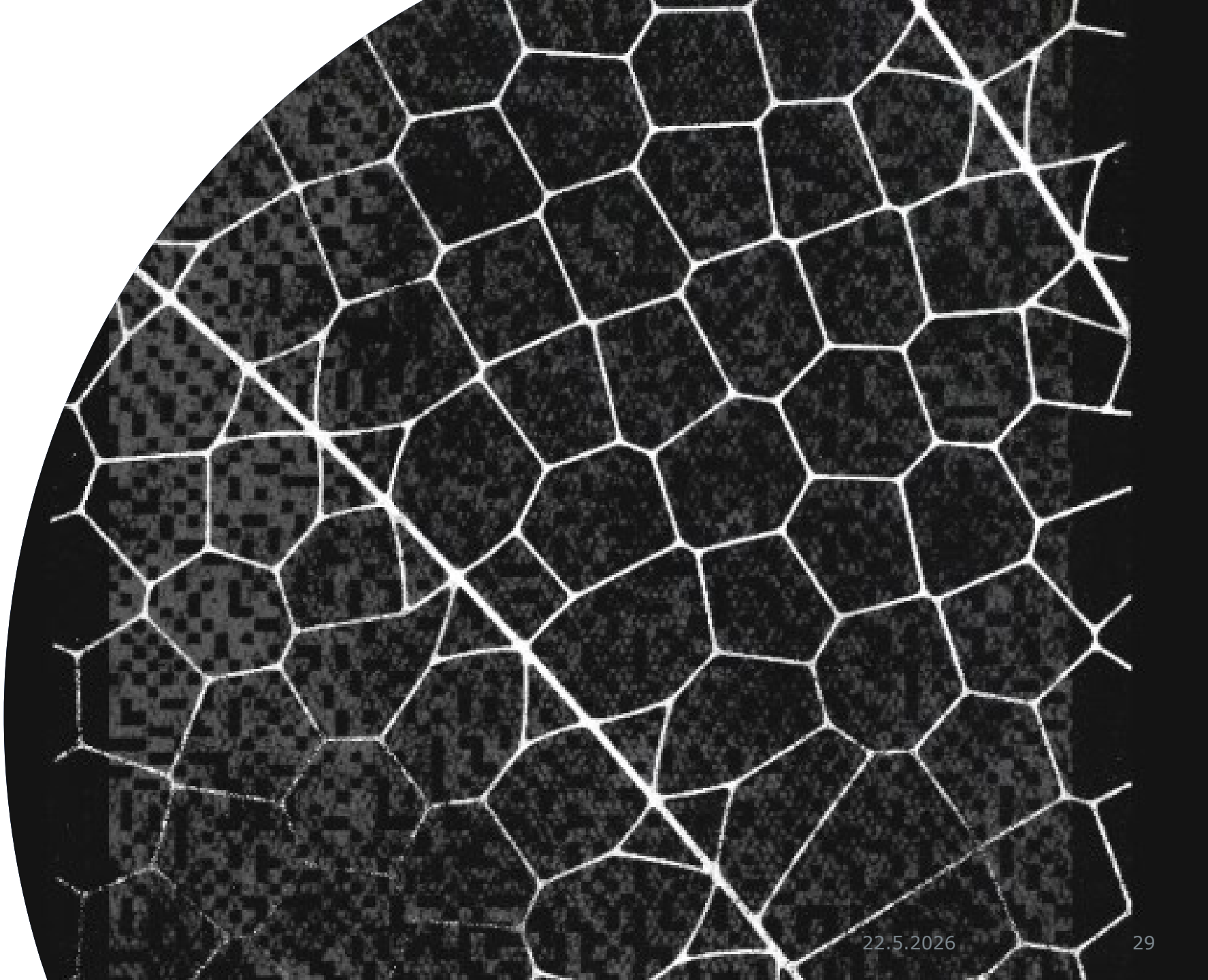
Haavoittuvuuksien löytäminen kielimalleilla - faktaa ja fiktiota


Chinese Firm Claims AI-D
nattothoughts.com/p/where-is-china-in-ai-driven-vulnerability
Natto Thoughts
Subscribe
Sign in
Chinese Firm Claims AI-Driven Bug Discovery Near Claude Mythos Scale
Chinese companies could match the capabilities attributed to Claude Mythos within months, according to industry experts, reinforcing existing cyber offense asymmetries
EUGENIO BENINCASA
APR 22, 2026 · PAID
8
2
Share
"Whoever masters automated vulnerability discovery technology holds the upper hand in cyber offense and defense" – Zhou Hongyi, Chairman and CEO, 360 Digital Security Group (2018)
On April 7, 2026, artificial intelligence developer Anthropic introduced its new general-purpose model Claude Mythos Preview to a restricted partnership of over 40 vetted organizations, including major technology and cybersecurity firms, as part of its defensive security initiative Project Glasswing. The company stated that the Claude Mythos model has identified thousands of high-severity vulnerabilities across widely used software, including major operating systems and web browsers. Crucially, in some cases it can autonomously develop exploits and chain vulnerabilities without


Post by @patak.cat — Bl
bsky.app/profile/patak.cat/post/3mksl5pyifs2n
Post
patak
@patak.cat
Follow
Other projects and organizations will follow calcom's and the NHS's steps if the current hysteria continues. Knowing about vulnerabilities sooner shouldn't significantly change how we operate. We need to keep our code open. But we need to work fast to get better processes and support maintainers.
patak @patak.cat · 19d
So we're back to security through obscurity? Sorry, but this is wrong. Our OSS apps and libs will be more secure thanks to the new models, not less. They are being released to researchers responsibly. Let's help maintainers avoid burnout. Let's fund them. Let's welcome more eyes checking our code.
Bailey Pumfleet @pumfleet 1h
Open source is dead.
That's not a statement we ever thought we'd make.
@calcom was built on open source. It shaped our product, our community, and our growth. But the world has changed faster than our principles could keep up.
AI has fundamentally altered the security landscape. What once required time, expertise, and intent can now be automated at
Bluesky
Create account
Sign in


The Boy That Cried Myth
flyingpenguin.com/the-boy-that-cried-mythos-verification-is-collapsing-trust-in-a...
PROPULSION: Unified flipper design for underwater bubble-afterburner propulsion
flyingpenguin
BOOKS
PRIVACY POLICY
ABOUT
SERVICES
PRESENTATIONS AND PUBLICATIONS
CONTACT
a blog about the poetry of information security, since 1995
SECURITY
HISTORY
POETRY
ENERGY
FOOD
SAILING
Popular this week
The Boy That Cried Mythos: Verification is Collapsing Trust in Anthropic
posted on April 13, 2026
Build an OpenClaw Free (Secure), Always-On Local AI Agent
posted on April 19, 2026
Conscious AI? Dawkins Falls for a Turk
SECURITY
The Boy That Cried Mythos: Verification is Collapsing Trust in Anthropic
APRIL 13, 2026
DAVI OTTENHEIMER
25 COMMENTS
I've been getting more and more curious about the risk from Anthropic's Claude Mythos Preview. So I pulled the system card, a whoppingly inefficient 244-page document that devotes just seven pages to the claim that the model is too dangerous to release. In fact, the 23MB of PDF I had to download was 20MB of wasted time and space. Compressing the PDF to 3MB meant I lost exactly nothing.
Foreshadowing, I guess.
Spoiler alert: the crucial seven pages out of 244 do not contain the word "fuzzer" once. That's like a seven page vacation brochure for Hawaii that leaves out the word beaches.
Also, the crucial seven pages out of 244 do not contain the expected acronyms CVSS, CWE or CVE, they do not have comparison baseline, an independent reproduction, or the word "thousands." I'll get back to
NEW BOOK
DAVI OTTENHEIMER
THE REALITIES OF
Securing
BIG DATA
WILEY
RECENT POSTS
Seventy-Five Cents Gets You an Anthropic Mythos Killer
The key to facial recognition is changing it like your underpants

Kuinka uniikki Mythos on?



Our evaluation of OpenA

aisi.gov.uk/blog/our-evaluation-of-openais-gpt-5-5-cyber-capabilities

AISI AI SECURITY INSTITUTE

About Research Grants Blog Careers

Preview, 52.4% (±9.8%) for GPT-5.4, and 48.6% (±10.0%) for Opus 4.7. On this measure, GPT-5.5 may be the strongest model we have tested.

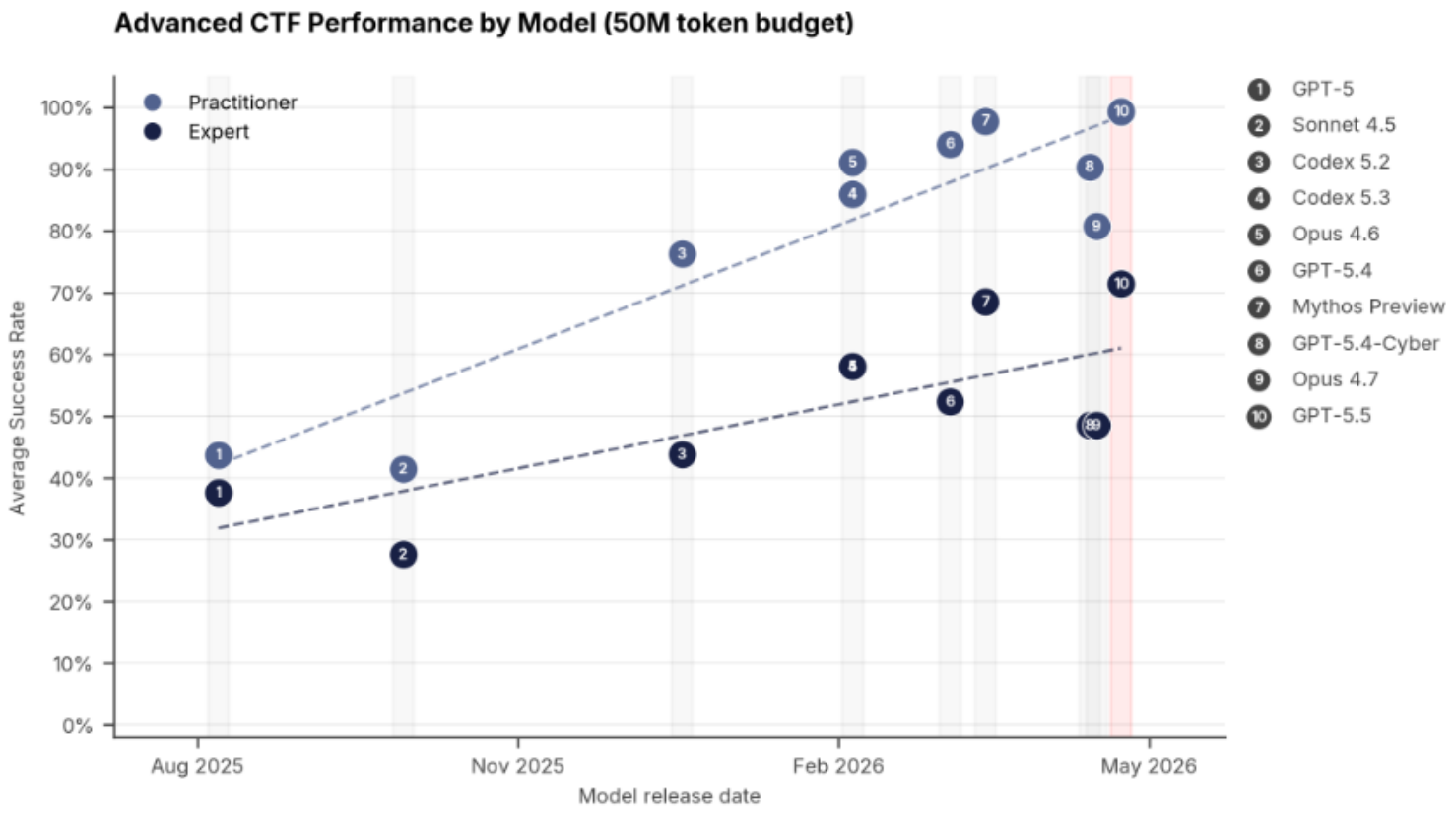
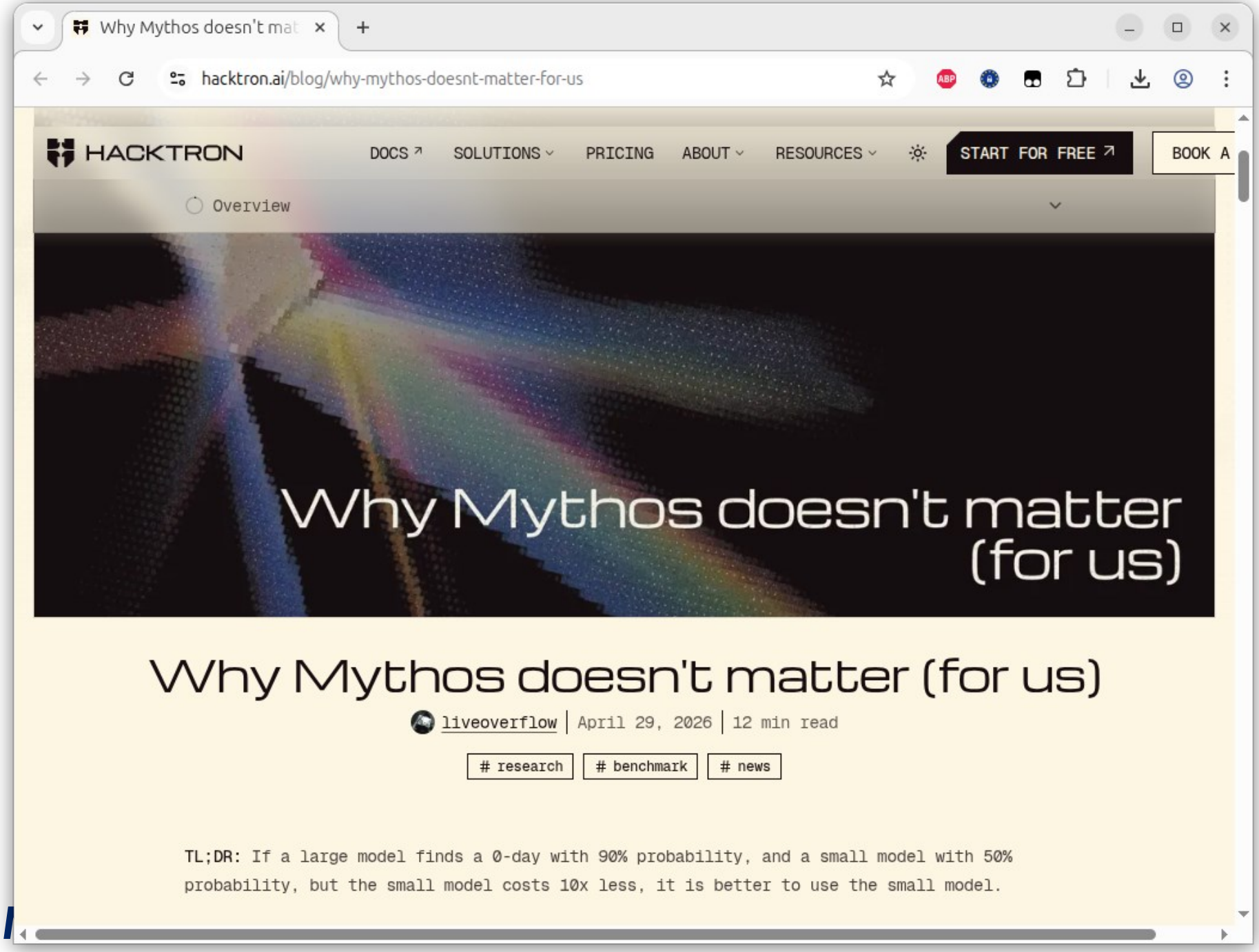
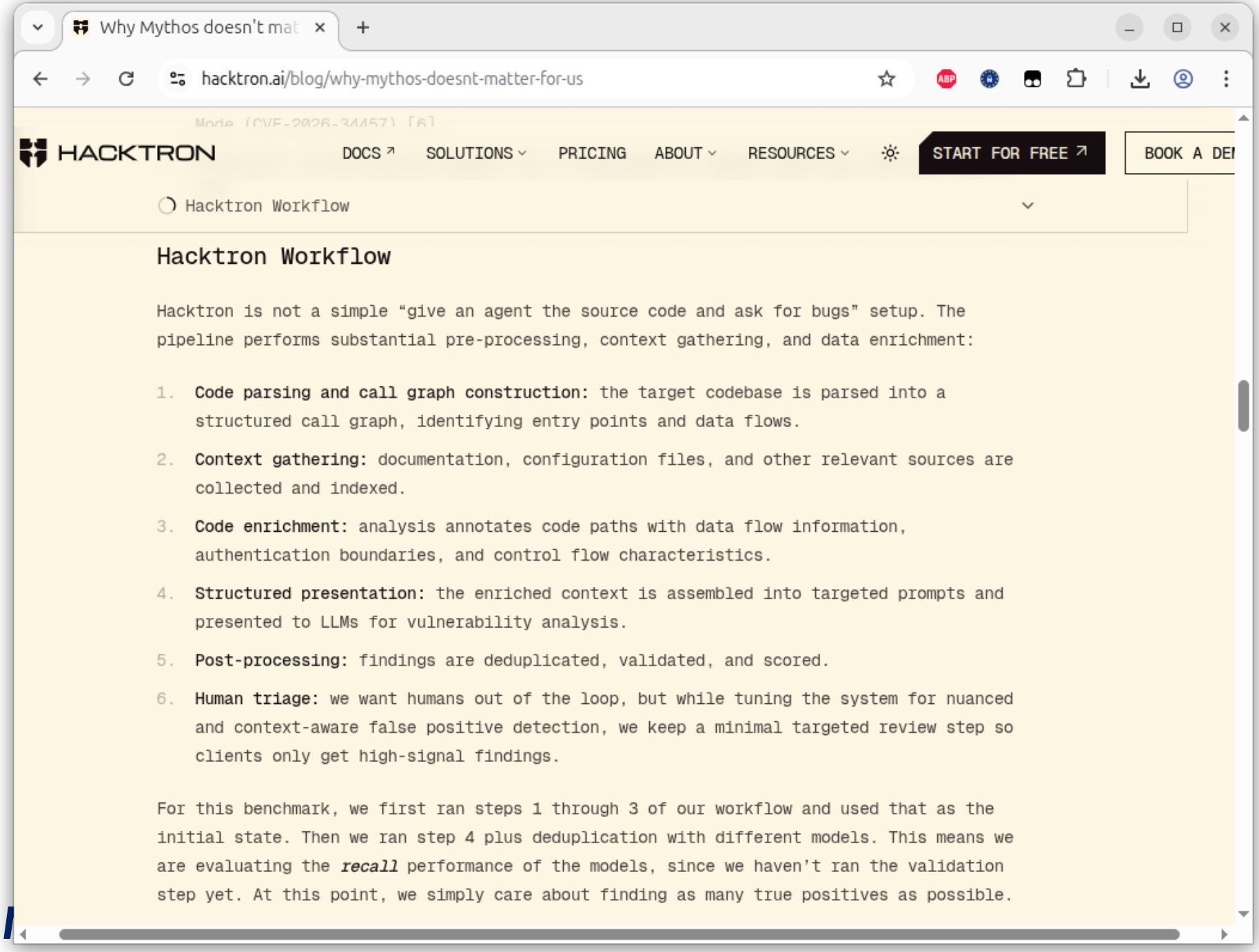


Figure 1: Average success rate on advanced cyber tasks at a 50M token budget. 27 Practitioner tasks, 21 Expert tasks.


Why Mythos doesn't mat
hacktron.ai/blog/why-mythos-doesnt-matter-for-us
HACKTRON
DOCS
SOLUTIONS
PRICING
ABOUT
RESOURCES
START FOR FREE
BOOK A
Overview
Why Mythos doesn't matter (for us)
Why Mythos doesn't matter (for us)
liveoverflow | April 29, 2026 | 12 min read
research
benchmark
news
TL;DR: If a large model finds a 0-day with 90% probability, and a small model with 50% probability, but the small model costs 10x less, it is better to use the small model.


Why Mythos doesn't mat
hacktron.ai/blog/why-mythos-doesnt-matter-for-us
HACKTRON
DOCS
SOLUTIONS
PRICING
ABOUT
RESOURCES
START FOR FREE
BOOK A DE
Hacktron Workflow
Hacktron Workflow
Hacktron is not a simple "give an agent the source code and ask for bugs" setup. The pipeline performs substantial pre-processing, context gathering, and data enrichment:
1. Code parsing and call graph construction: the target codebase is parsed into a structured call graph, identifying entry points and data flows.
2. Context gathering: documentation, configuration files, and other relevant sources are collected and indexed.
3. Code enrichment: analysis annotates code paths with data flow information, authentication boundaries, and control flow characteristics.
4. Structured presentation: the enriched context is assembled into targeted prompts and presented to LLMs for vulnerability analysis.
5. Post-processing: findings are deduplicated, validated, and scored.
6. Human triage: we want humans out of the loop, but while tuning the system for nuanced and context-aware false positive detection, we keep a minimal targeted review step so clients only get high-signal findings.
For this benchmark, we first ran steps 1 through 3 of our workflow and used that as the initial state. Then we ran step 4 plus deduplication with different models. This means we are evaluating the recall performance of the models, since we haven't ran the validation step yet. At this point, we simply care about finding as many true positives as possible.

Synthesizing Multi-Agent Harnesses for Vulnerability Discovery

Hanzhi Liu
University of California, Santa Barbara
hanzhi@ucsb.edu

Chaofan Shou
Fuzzland
shou@fuzz.land

Xiaonan Liu
Fuzzland
xl@fuzz.land

Hongbo Wen
University of California, Santa Barbara
hongbowen@ucsb.edu

Yanju Chen
University of California, San Diego
yanju@ucsd.edu

Ryan Jingyang Fang
World Liberty Financial
ryan@worldlibertyfinancial.com

Yu Feng
University of California, Santa Barbara
yufeng@cs.ucsb.edu

Abstract

LLM agents have begun to find real security vulnerabilities that human auditors and automated fuzzers missed for decades, in source-available targets where the analyst can build and instrument the code. In practice the work is split among several agents, wired together by a *harness*: the program that fixes which roles exist, how they pass information, which tools each may call, and how retries are coordinated. When the language model is held fixed, changing only the harness can still change success rates by several-fold on public agent benchmarks, yet most harnesses are written by hand; recent harness optimizers each search only a narrow slice of the design space and rely on coarse pass/fail feedback that gives no diagnostic signal about *why* a trial failed. AgentFlow addresses both limitations with a typed graph DSL whose search space jointly cov-

web-vulnerability exploitation [48]; and agentic systems have reproduced CTF-level exploits from natural-language descriptions [35]. The capability is real and accelerating.

To see where these capabilities come from, and why they are still limited, it helps to look at how an LLM-based vulnerability finder is actually put together. In the simplest deployment, a single language model operates as a *security agent*: the operator hands it a *target program* (the software under analysis, e.g. `libtiff`, OpenSSL, `curl`), gives it a *system prompt* that states its goal in natural language ("find a memory-safety vulnerability in the TIFF parsing module"), and exposes a set of *tools*: concrete actions the model may invoke at each step (read source files, compile the target with instrumentation, execute the binary, inspect output). The agent reasons about the target, generates candidate inputs, runs them, and observes the

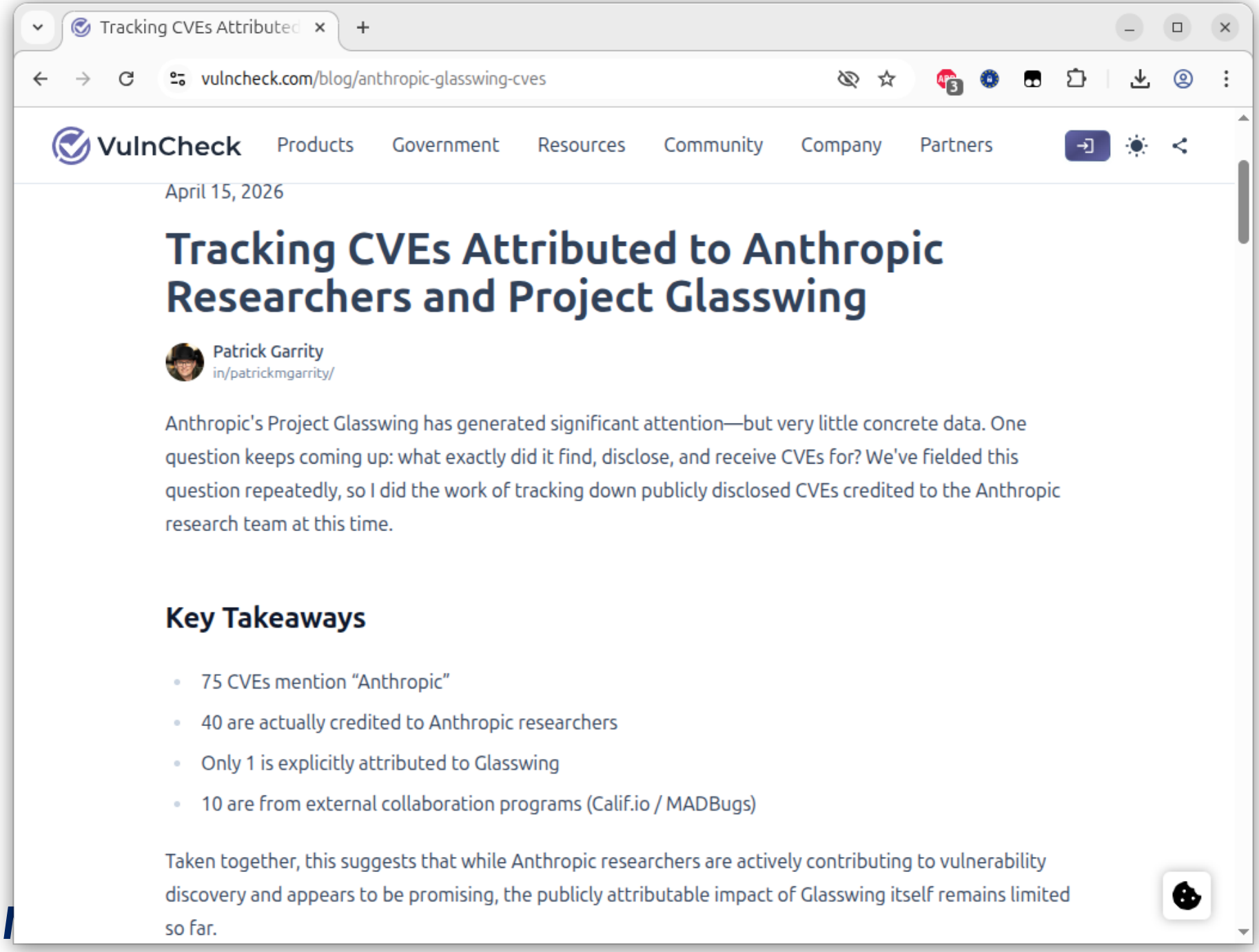
Target	Vuln. Type	Severity	Identifier	Patch Date
Chrome / WebCodecs	Use-after-free	Critical	CVE-2026-5280	2026-03-31 [11]
Chrome / Proxy	Use-after-free	Critical	CVE-2026-6297	2026-04-15 [10]
Chrome / Network	Use-after-free	High	CVE-2026-4454	2026-03-18 [12]
Chrome / Codecs	Integer overflow	High	CVE-2026-5274	2026-03-31 [11]
Chrome / Rendering	Heap Buffer Overflow	High	CVE-2026-4462	2026-03-18 [12]
Chrome / Rendering	Use-after-free	High	494352590	N/A
Chrome / Rendering	Heap Buffer Overflow	High	493534964	N/A
Chrome / WebRTC	Heap Buffer Overflow	High	488803429	N/A
Chrome / WebCodecs	Heap Buffer Overflow	High	488585490	N/A
Chrome / WebGL	Inappropriate implementation	Medium	CVE-2026-5291	2026-03-31 [11]

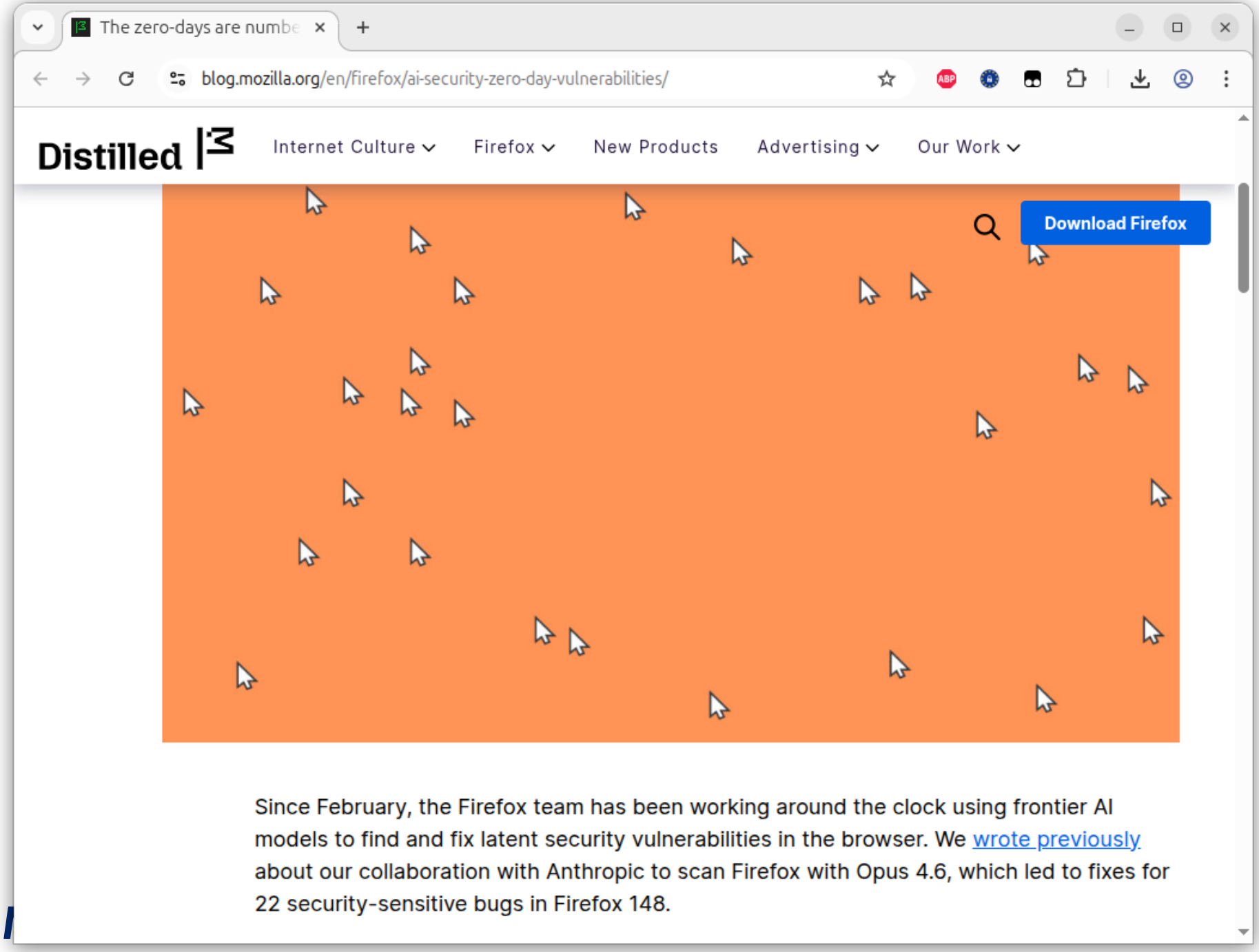
ways that Mythos Preview found to exploit them. It was ab
of these vulnerabilities—and develop many related exploits—
autonomously, without any human steering. The following are thr

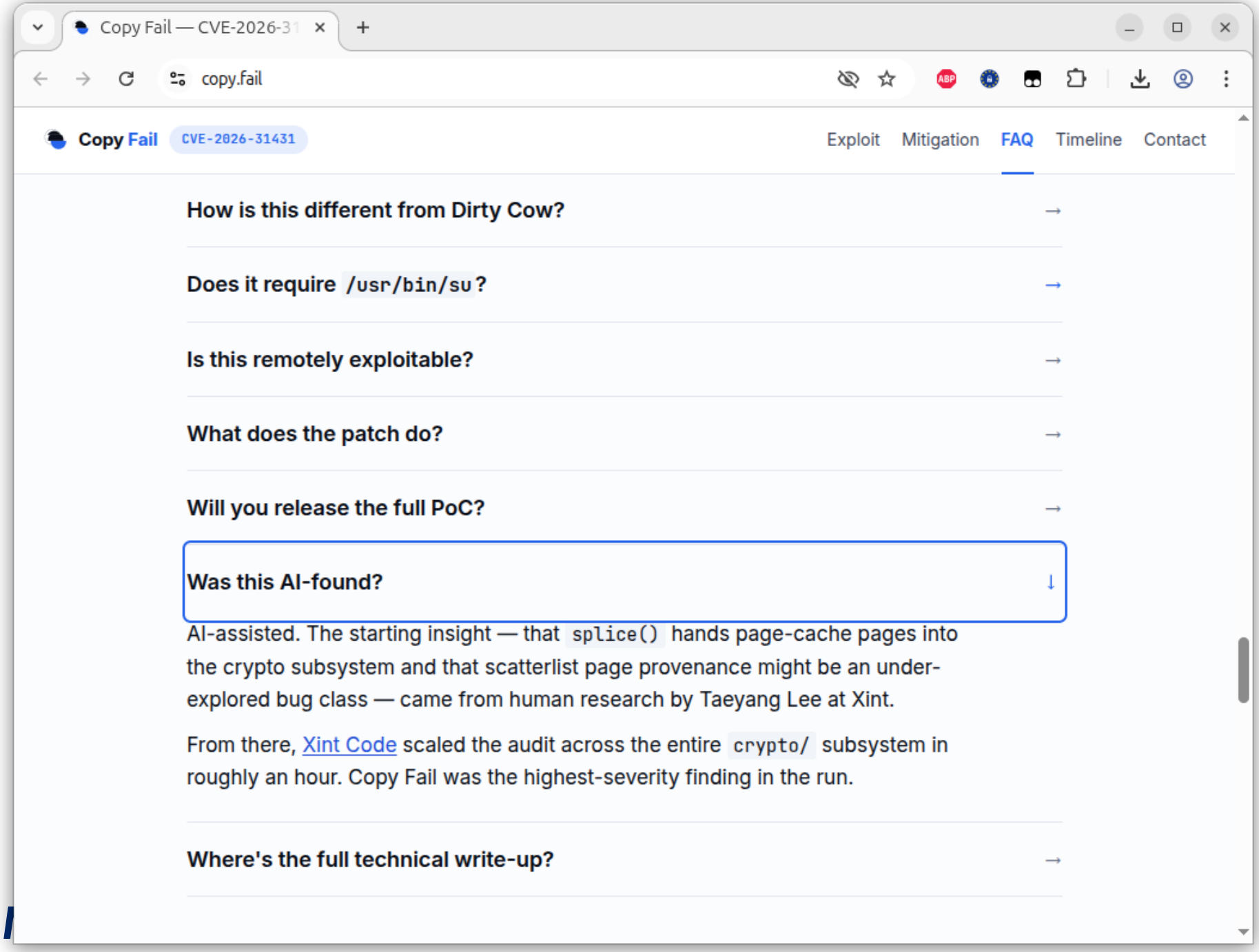
- Mythos Preview found a 27-year-old vulnerability in OpenBSD—w
reputation as one of the most security-hardened operating systems in
world and is used to run firewalls and other critical infrastructure. The
vulnerability allowed an attacker to remotely crash any machine running
operating system just by connecting to it;
- It also discovered a 16-year-old vulnerability in FFmpeg—which is used by
innumerable pieces of software to encode and decode video—in a line of code
that automated testing tools had hit five million times without ever catching
the problem;
- The model autonomously found and chained together several vulnerabilities
in the Linux kernel—the software that runs most of the world's servers—to
allow an attacker to escalate from ordinary user access to complete control of
the machine.

We have reported the above vulnerabilities to the maintainers of the relevant
software, and they have all now been patched. For many other vulnerabilities, we

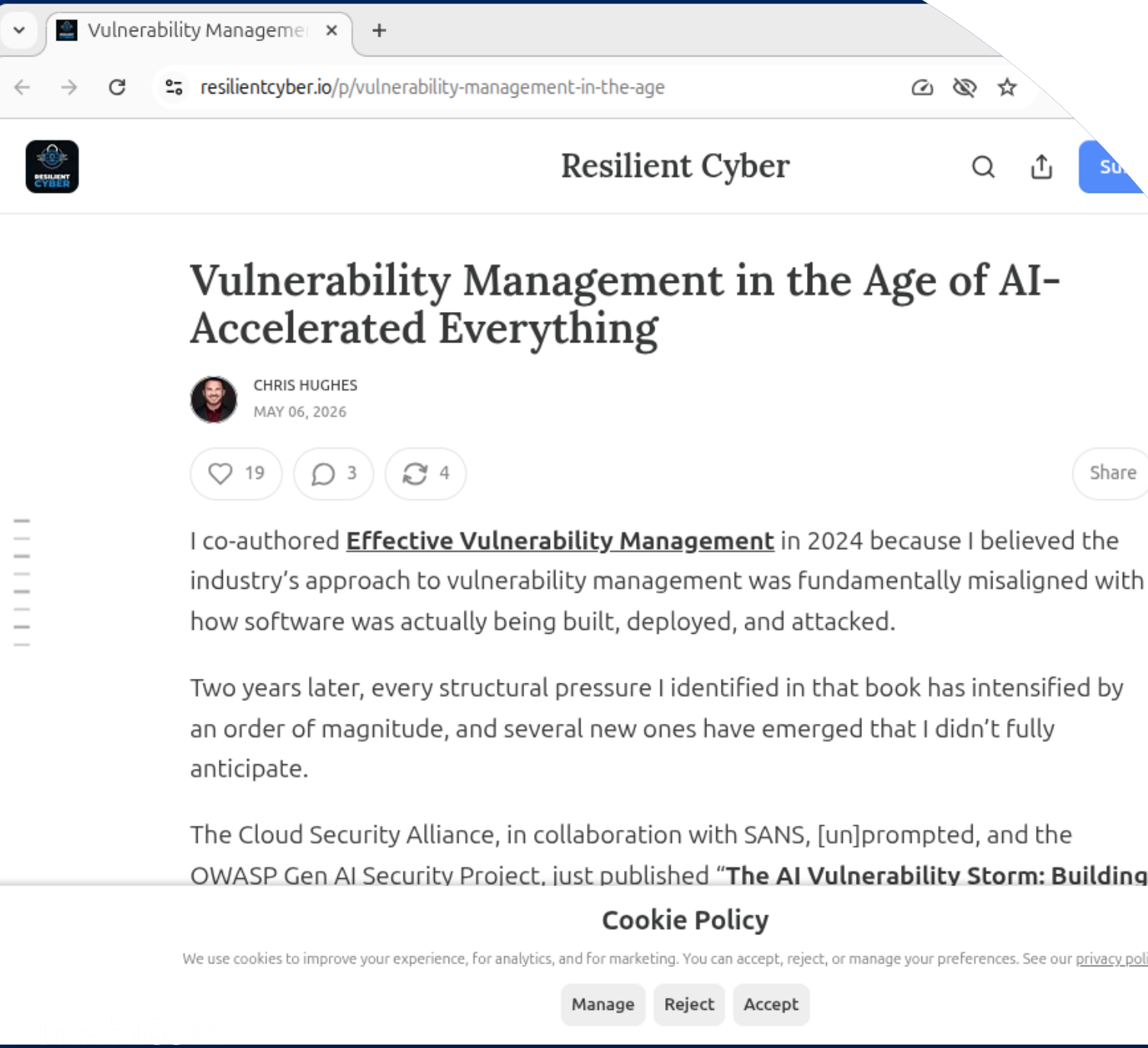
Ovatko haavat oikeita?

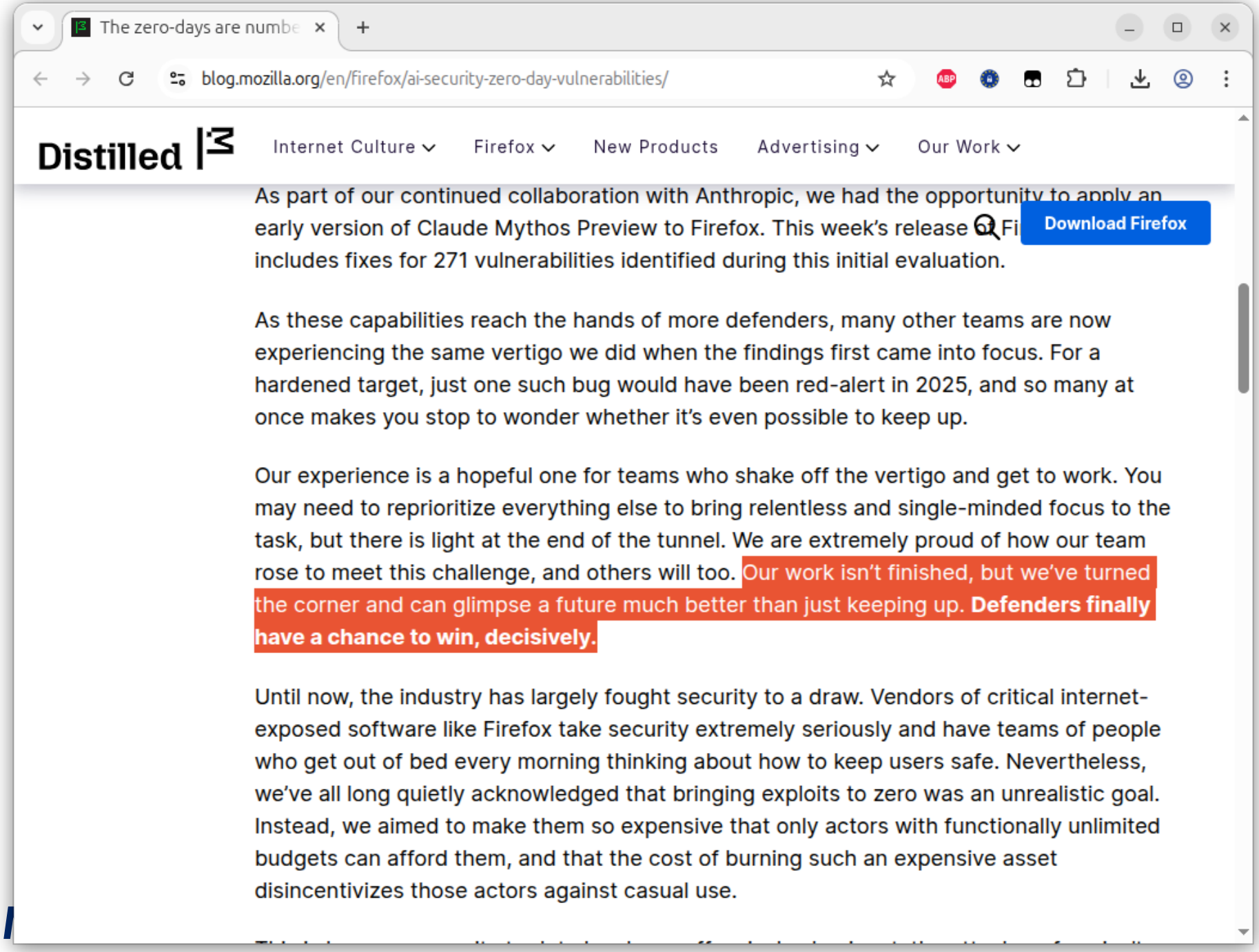

Tracking CVEs Attributed
vulncheck.com/blog/anthropic-glasswing-cves
VulnCheck
Products
Government
Resources
Community
Company
Partners
April 15, 2026
Tracking CVEs Attributed to Anthropic Researchers and Project Glasswing
Patrick Garrity
in/patrickmgarrity/
Anthropic's Project Glasswing has generated significant attention—but very little concrete data. One question keeps coming up: what exactly did it find, disclose, and receive CVEs for? We've fielded this question repeatedly, so I did the work of tracking down publicly disclosed CVEs credited to the Anthropic research team at this time.
Key Takeaways
75 CVEs mention “Anthropic”
40 are actually credited to Anthropic researchers
Only 1 is explicitly attributed to Glasswing
10 are from external collaboration programs (Calif.io / MADBugs)
Taken together, this suggests that while Anthropic researchers are actively contributing to vulnerability discovery and appears to be promising, the publicly attributable impact of Glasswing itself remains limited so far.


The zero-days are numbe
blog.mozilla.org/en/firefox/ai-security-zero-day-vulnerabilities/
Distilled
Internet Culture
Firefox
New Products
Advertising
Our Work
Download Firefox
Since February, the Firefox team has been working around the clock using frontier AI models to find and fix latent security vulnerabilities in the browser. We wrote previously about our collaboration with Anthropic to scan Firefox with Opus 4.6, which led to fixes for 22 security-sensitive bugs in Firefox 148.


Copy Fail — CVE-2026-31
copy.fail
Copy Fail
CVE-2026-31431
Exploit
Mitigation
FAQ
Timeline
Contact
How is this different from Dirty Cow?
Does it require /usr/bin/su?
Is this remotely exploitable?
What does the patch do?
Will you release the full PoC?
Was this AI-found?
AI-assisted. The starting insight — that splice() hands page-cache pages into the crypto subsystem and that scatterlist page provenance might be an under-explored bug class — came from human research by Taeyang Lee at Xint.
From there, Xint Code scaled the audit across the entire crypto/ subsystem in roughly an hour. Copy Fail was the highest-severity finding in the run.
Where's the full technical write-up?

Mitä seuraavaksi?




The zero-days are numbe
blog.mozilla.org/en/firefox/ai-security-zero-day-vulnerabilities/
Distilled
Internet Culture
Firefox
New Products
Advertising
Our Work
Download Firefox
As part of our continued collaboration with Anthropic, we had the opportunity to apply an early version of Claude Mythos Preview to Firefox. This week's release of Fi includes fixes for 271 vulnerabilities identified during this initial evaluation.
As these capabilities reach the hands of more defenders, many other teams are now experiencing the same vertigo we did when the findings first came into focus. For a hardened target, just one such bug would have been red-alert in 2025, and so many at once makes you stop to wonder whether it's even possible to keep up.
Our experience is a hopeful one for teams who shake off the vertigo and get to work. You may need to reprioritize everything else to bring relentless and single-minded focus to the task, but there is light at the end of the tunnel. We are extremely proud of how our team rose to meet this challenge, and others will too. Our work isn't finished, but we've turned the corner and can glimpse a future much better than just keeping up. Defenders finally have a chance to win, decisively.
Until now, the industry has largely fought security to a draw. Vendors of critical internet-exposed software like Firefox take security extremely seriously and have teams of people who get out of bed every morning thinking about how to keep users safe. Nevertheless, we've all long quietly acknowledged that bringing exploits to zero was an unrealistic goal. Instead, we aimed to make them so expensive that only actors with functionally unlimited budgets can afford them, and that the cost of burning such an expensive asset disincentivizes those actors against casual use.

Retaining defensive adva

ncsc.gov.uk/blogs/retaining-defensive-advantage-in-the-age-of-frontier-ai-cyber-capabi...

Retaining defensive advantage in the age of frontier AI cyber capabilities

As AI accelerates vulnerability discovery, organisations must raise their security baselines to safeguard their cyber security.

Download & print article PDF

Share

WRITTEN BY

Richard Horne

Chief Executive Officer (CEO)

PUBLISHED

Back to top

Executive Summary

What happened:

- AI, as demonstrated by Anthropic's Mythos, has significantly increased the likelihood of attackers discovering new vulnerabilities, creating new exploits, and using them in complex automated attacks at scale.
- While AI also increases the speed to develop patches, and reduces defects in new software, the burden on defenders, by comparison, increases due to the inherent limitations of patching. The attackers gain asymmetric benefits.

How is this different from the status quo?

- In the near term, security organizations will likely be overwhelmed by the need to apply patches and respond to AI-discovered vulnerabilities, exploits, and autonomous attacks.

What to do now to deal with the current risk spike?

- Adjust risk calculations and re-orient security program resources for increasing volume of patches, decreasing time to patch, and more-persistent complex attacks.
- Focus on the basics and harden your environment further. Segmentation, egress filtering, multifactor authentication, and defense-in-depth/breadth all increase the difficulty for attackers.

What do we believe will happen next?

- The storm of vulnerability disclosures from Project Glasswing is the first of many large waves of AI-discovered vulnerabilities that may occur in rapid sequence.
- The capabilities seen in Mythos will quickly become more widely available, dramatically increasing the number and frequency of complex, novel attacks organizations will face.

What else should start now to be ready for the next waves?

- Prioritize robust dependency management to reduce vulnerabilities in third-party and open-source components.
- Consistently enforce automated security assessments in your development processes, including using LLM-powered agents to find vulnerabilities before the attackers.
- Introduce AI agents to the cyber workforce across the board enabling defenders to match attackers' speed and begin closing the gap.
- Re-evaluate your risk tolerance to operational downtime caused by vulnerability remediation to account for shorter adversary timelines.

Epäiletkö tietoturvaloukkausta?

Jos teihin on kohdistunut tai epäilette teihin kohdistuneen tietoturvaloukkauksen, olkaa yhteydessä Traficomin Kyberturvallisuuskeskukseen.

- Sähköinen lomake www.kyberturvallisuuskeskus.fi/fi/ilmoita
- Sähköposti: cert@traficom.fi
- Puhelin: 0295 345 630 (arkisin klo 9-15)

Muissa asioissa voitte olla meihin yhteydessä osoitteessa kyberturvallisuuskeskus@traficom.fi.

Kyberturvallisuuskeskuksen eri toimintojen ja hankkeiden yhteystiedot löydät keskitetysti osoitteesta www.kyberturvallisuuskeskus.fi/fi/ota-yhteytta/yhteystiedot.