

AI Horisontti



Kaiva esiin *vanha lasku*

Tieto tekoälyn avulla löydetystä PHP-ohjelmointikielen aukoista vuoti ennen korjausta ja hyökkäysyritykset kasvoivat tunneissa miljooniin. Valtori sulki varotoimena valtion verkkosivuja.

MALLIT

Kiinalainen malli kärjen tuntumaan ja puoleen hintaan

SOPIMUKSET

Uusissa it-ehdoissa toimittajan vaihto on maksuton

AGENTIT

Verkkosivut alkoivat torjua käyttäjän puolesta asioivia avustajia

PÄÄKIRJOITUS • VIIKKOKATSAUS

Kaksi laskua ja yksi aukko

Keskiviikkoamuna kauppias avaa oman verkkokauppansa ja löytää etusivulta pelkän huoltoilmoituksen. Syy on PHP, ohjelmointikieli, jonka varassa toimii suuri osa verkkosivuista ja verkkokaupoista WordPress-sivustot mukaan lukien. Siitä oli löydetty tekoälyn avulla kaksi vakavaa aukkoa, joiden kautta ulkopuolinen voi ottaa palvelimen haltuunsa. Tieto aukoista vuoti ennen korjausta, hyökkäysyritykset kasvoivat tunneissa miljooniin ja Valtori sulki varotoimena valtion verkkosivuja. Kauppiaan tärkein asiakirja on sillä hetkellä vuosia vanha lasku, josta selviää, kuka sivuista huolehtii.

Laskun lähettäjältä kannattaa pyytää kirjallinen vahvistus siitä, että keskiviikkona 21.10. julkaistu korjaus on asennettu palvelimelle. Yhtä tärkeää on kysyä, tarkistettiinko palvelimelta merkit murrosta, sillä hyökkääjillä oli valmiit työkalut ennen kuin korjausta oli olemassa. Numeron käytännön osio käy kysymykset läpi ja kannattaa lukea ensimmäisenä, kun taas tekninen osa kertoo, miksi palvelimen ilmoittama versio ei yksin todista korjausta. Tapaus siirsi tekoälyn kyvyn etsiä ohjelmistovirheitä tutkimusraporteista pienen yrityksen huoltoilmoitukseksi ja alan johtavat yhtiöt odottavat samanlaisia löytöjä lisää.

Viikon toinen lasku näkyi hinnastoissa. Pekingiläinen Moonshot toi myyntiin mallin, jonka riippumaton vertailupalvelu sijoitti neljänneksi. Hinta on sama kuin OpenAI:n vastaavalla mallilla ja puolet Anthropicin vastaavan mallin hinnasta, mutta tasavertaisuus yhdysvaltalaisen kärkeimallien kanssa on toistaiseksi yhtiön omien testien varassa. Seurattava päivä on 4. marraskuuta,

johon mennessä malli luvataan vapaasti ladattavaksi, jolloin sitä voisi ajaa eurooppalaisilla palvelimilla ja tiedot pysyisivät EU:ssa. Työkaluosiossa sama toteutuu jo puheavustajissa, joiden tallenteet voivat jäädä EU:hun tai yrityksen omille palvelimille. Toimiiko avustaja suomeksi, sitä julkistus ei kerro.

Laskun lähettäjää on samalla aiempaa helpompi vaihtaa. Googlen pilvessä OpenAI:n ja Anthropicin mallit ovat nyt samalla hyllyllä ja samalla laskulla kuin yhtiön oma Gemini. Suomalaisten it-sopimusten uudet vakioehdot tekevät Tivin mukaan toimittajan vaihdosta maksutonta, kun sopimus viiittaa niihin. EU:n tekoälyasetus sai ensimmäisen käytännön ohjeensa eli standardin, joten työhakemuksia lajittelevaa tai luottopäätöksiä tukevaa järjestelmää ostava yritys voi nyt kysyä toimittajaltaan, noudattaako tämän laatujärjestelmä sitä.

Verkkosivustot ovat The Vergen mukaan alkaneet torjua käyttäjän puolesta asioivia tekoälyavustajia ja ChatGPT vie keskustelun aiheen mukaan valitut mainoksensa 26 uuteen maahan. Kauppiaan pitää päättää, päästääkö hän asiakkaan avustajan sisään ja haluaako tavoittaa tämän mainoksella kesken keskustelun tekoälyn kanssa. Tekoälyn loppulasku riippuu lisäksi sähköstä ja tehtaista yhtä lailla kuin malleista. Microsoftin ja tekoälypilviyhtiö Nscalen datakeskushanke näyttää kaatuneen riittämättömään sähköverkkoon ja taiwanilaisen siru- ja muistiyhtiön työntekijät menivät lakkoon ensimmäistä kertaa maan sirutehtaiden historiassa. Aukon löysi kone, mutta vanhan laskun joutuu kaivamaan esiin itse.

TÄSSÄ NUMEROSSA

KAIKILLE

Mitä tämä tarkoittaa käytännössä
Työkalut ja avustajat arjessa
Markkina, sääntely ja turvallisuus
Mitä seurata

TEKNINEN OSA

Suuret toimijat
Avoimet mallit ja ekosysteemi
Ajoalustat ja infra
Benchmarkit ja vertailut
Työskentelytavat ja agentit
Liitännäiset, MCP ja repot
Tietoturva ja haavoittuvuudet

VIIKON KÄRKI

Viikon tärkein asia olivat PHP-ohjelmointikielen kaksi vakavaa aukkoa. Ne löydettiin tekoälyä apuna käyttäen ja niiden vuoksi valtion it-palvelukeskus Valtori ja lukuisat muut ylläpitäjät sulkiivat keskiviikkona verkkopalvelujaan varotoimenä. Tapaus muutti viikon kuvaa, koska tekoälyn kyky etsiä ohjelmistovirheitä siirtyi tutkimusraporteista pienen yrityksen verkkokaupan huoltoilmoitukseksi ja koska alan johtavat yhtiöt odottavat samanlaisia löytöjä lisää. Päivästä toiseen jatkuivat myös tekoälyn hintojen lasku, vapaasti ladattavien mallien nousu kärjen tuntumaan ja palveluntarjoajan vaihtamisen helpottuminen, joka eteni samalla viikolla sekä suomalaisissa sopimusehdoissa että avustajien yhteisissä standardeissa. Kohinaksi osoittautuivat suuret luvut ja lupaukset, joiden takana oli pelkkä yhtiön oma ilmoitus. Esimerkki on tekoälyyhtiö xAI:n superälyä lupaava kumppaniverkosto, jonka jäsenistä ei kerrottu mitään.

1 Verkkosivujen aukko koskee pientäkin yritystä.
PHP on ohjelmointikieli, jonka varassa toimii suuri osa verkkosivuista ja verkkokaupoista WordPress-sivustot mukaan lukien. Tieto aukoista vuoti ennen korjauksen julkaisua, minkä jälkeen hyökkäysyritykset kasvoivat verkkoliikennettä välittävän Cloudflaren mukaan tunneissa miljooniin. Sivuston ylläpitäjältä tai palveluntarjoajalta kannattaa pyytää kirjallinen vahvistus siitä, että palvelimelle on asennettu keskiviikkona 21.10. julkaistu korjaus.

2 Kiinalainen malli kärjen tuntumaan ja puoleen hintaan. Pekingiläinen Moonshot AI toi myyntiin mallin, jonka riippumaton vertailupalvelu sijoitti yleislistallaan neljänneksi. Mallin käyttö maksaa saman verran kuin OpenAI:n vastaavan mallin ja puolet Anthropicin vastaavan mallin hinnasta. Yhtiö lupaa julkaista mallin vapaasti ladattavaksi 4. marraskuuta mennessä, jolloin sitä voisi ajaa myös eurooppalaisilla palvelimilla, mutta väite tasavertaisuudesta yhdysvaltalaisen kärkein kanssa perustuu yhtiön omiin testeihin.

3 Toimittajaa on aiempaa helpompi vaihtaa.
Googlen pilvipalvelun asiakas voi nyt käyttää saman tilin ja laskun kautta sekä OpenAI:n malleja että Anthropicin Claudea. Suomalaisten it-sopimusten uudet IT2026-ehdot tekevät Tivin mukaan toimittajan vaihtamisesta maksutonta niissä sopimuksissa, jotka viittaavat ehtoihin. Anthropic lahjoitti lisäksi avustajien työohjeiden muodon yhteiseen säätiöön, minkä ansiosta yhteen avustajaan laadittu ohje voi jatkossa toimia myös kilpailijan avustajassa.

4 EU:n tekoälyasetus sai ensimmäisen käytännön ohjeen. Euroopan komissio julkaisi viittauksen ensimmäiseen tekoälyasetusta tukevaan yhdenmukaistettuun standardiin. Se on EU:n tunnustama tekninen ohje, jota noudattamalla kehittäjä voi osoittaa täyttävänsä lain vaatimuksen. Standardi koskee laatu järjestelmää, jota laki vaatii esimerkiksi työhakemuksia lajittelevien ja luottopäätöksiä tukevien järjestelmien kehittäjiltä, joten tällaista järjestelmää ostava yritys voi nyt kysyä toimittajaltaan, noudattaako sen laatu järjestelmä uutta standardia.

5 Asiakkaan tiedot voivat jäädä talon sisään.

Mistral julkaisi työkalut puheavustajan rakentamiseen niin, että puhetallenteet voivat jäädä EU:hun tai yrityksen omille palvelimille. Ollama-yhtiön ilmainen ohjelma, jolla tekoälymalleja ajetaan omalla tietokoneella, oppi lajittelemaan asiakasviestejä ja lukemaan tekstiä ääneen. Kumpikaan julkistus ei kerro, toimiiko ratkaisu suomeksi.

6 Avustajat asettuvat asiakkaan ja kaupan väliin.

The Vergen mukaan verkkosivustot ovat alkaneet torjua tekoälyagentteja eli avustajia, jotka selaavat verkkoa ja asioivat käyttäjän puolesta. OpenAI puolestaan vie ChatGPT:n mainokset 26 uuteen maahan ja valitsee ne keskustelun aiheen perusteella. Yrittäjälle tästä seuraa kaksi päätöstä: päästääkö asiakkaan avustajan omaan verkkokauppaan ja haluaako tavoittaa asiakkaan mainoksella kesken hänen keskustelunsa tekoälyn kanssa.

7 Tekoälyn kasvu törmää sähköön ja siruihin.

Viikon aikana uutisoitiin kolmesta asiasta. Microsoftin ja tekoälypilvyhtiö Nscalen datakeskushanke kaatui riittämättömään sähköverkkoon, suljettuja ydinvoimaloita käynnistetään uudelleen datakeskusten tarpeisiin ja taiwanilaisen siru- ja muistiyhtiön työntekijät menivät lakkoon ensimmäistä kertaa maan sirutehtaiden historiassa. Tekoälypalvelujen hinta riippuu lopulta sähköasemista ja tehtaista yhtä lailla kuin mallien kehityksestä.

01 Mitä tämä tarkoittaa käytännössä

Jyväskyläläinen polkupyöräliike myy varaosia verkkokaupassa, jonka paikallinen freelancer rakensi muutama vuosi sitten. Kauppa toimii webhotellissa eli palvelimella, jota ylläpitää palveluntarjoaja ja josta liike maksaa laskun kerran vuodessa. Keskiviikkona kaupan etusivulla näkyy pelkkä huoltoilmoitus, eikä kauppias tiedä, mistä on kyse. Syy on PHP, ohjelmointikieli, jonka päälle suuri osa verkkokaupoista ja WordPress-sivustoista on rakennettu. Siitä paikattiin samana päivänä kaksi aukkoa, joiden kautta ulkopuolinen voi ottaa palvelimen haltuunsa internetin yli. Tieto aukoista vuoti ennen sovittua julkistusta, minkä jälkeen valmiita hyökkäys työkaluja kiersi verkossa jo tiistain ja keskiviikon välisenä yönä, ennen kuin korjausta oli

edes julkaistu. Verkkoliikennettä välittävän Cloudflaren mukaan hyökkäys yritykset kasvoivat tunneissa miljooniin. Moni ylläpitäjä sulki palvelunsa päivityksen ajaksi samalla tavalla kuin Valtori valtion verkkosivut.

Kyberturvallisuuskeskuksen mukaan aukot koskevat kaikkia julkaistuja PHP-versioita, minkä vuoksi sivuston iällä tai tekotavalla ei ole merkitystä. Ratkaisevaa on vain se, onko keskiviikkona 21.10. julkaistu korjaus jo asennettu palvelimelle. WordPressin tai sen lisäosien päivittäminen ei yksin riitä, koska aukko on itse ohjelmointikielessä eikä sivuston julkaisujärjestelmässä. Vastuu päivityksestä riippuu siitä, miten sivusto on aikanaan hankittu. Tavallisessa

webhotellissa PHP:n päivittää yleensä palveluntarjoaja omalla aikataulullaan. Erikseen vuokratulla palvelimella, jonka yritys tai sen verkkotoimisto on aikanaan pystyttänyt, päivitys jää yrityksen tai sen alihankkijan tehtäväksi eikä tapahdu itsestään.

Ylläpitäjälle kannattaa lähettää vielä tällä viikolla kolme kysymystä. Ensimmäinen kysymys on, onko PHP päivitetty ja minä päivänä. Toinen kysymys on, tarkistettiinko palvelimelta merkit murrosta päivitystä edeltävältä ajalta. Se painaa eniten, koska hyökkääjillä oli työkalut käsissään ennen kuin korjausta oli olemassa. Kolmas kysymys koskee tulevaisuutta: kuka päivityksistä vastaa ja kuinka nopeasti kriittiset korjaukset luvataan asentaa.

Kyberturvallisuuskeskuksen perjantainen viikkokatsaus, jonka kärjessä PHP-aukot ovat, kelpaa välitettäväksi ylläpitäjälle sellaisenaan.

Suljettu sivu ei itsessään kerro murrosta, koska moni ylläpitäjä sulki palvelunsa ennalta ehkäisevästi. Merkkejä murrosta voivat olla tuntemattomat ylläpitäjätunnukset, sivuille yllättäen ilmestynyt vieras sisältö tai viestit, joita asiakkaat ihmettelevät mutta joita yritys ei ole lähettänyt. Sivustolla olevat asiakkaiden henkilötiedot tekevät murrosta ilmoitettavan asian: EU:n tietosuojasetuksen mukaan siitä on pääsääntöisesti ilmoitettava tietosuojavaltuutetun toimistolle 72 tunnin kuluessa havaitsemisesta. Tietoturvaloukkauksesta voi ilmoittaa myös Kyberturvallisuuskeskukselle.

PHP-tapaus ei tullut tyhjältä. Tivi kertoi jo tiistaina, päivää ennen vuotoa, tekoälyn avulla tehdystä murtosarjasta, joka ulottui yli 100 suuryrityksen järjestelmiin. Lehden haastatteleman asiantuntijan mukaan tekoäly moninkertaistaa nyt yksittäisen hyökkääjän tuhovoiman. Keskiviikkona OpenAI käynnisti ohjelman, joka kokoaa tietoturvatutkijat ja ohjelmistojen ylläpitäjät korjaamaan aukkoja, joita tekoälyagentit eli itsenäisesti työskentelevät tekoälyohjelmat löytävät. Perjantaina Anthropicin tietoturvaohjaaja varoitti Tivin mukaan, että PHP-tapausta pahempaa on vielä tulossa. Uutiskirje Latent Space antoi vaiheelle nimen suuri koventaminen. Sillä se tarkoittaa sitä, että vanhoista ohjelmistoista etsitään ja tukitaan aukkoja tekoälyn avulla aiempaa laajemmin.

Yrittäjän arjessa muutos näkyy päivitysten tahtina. Ylläpitösopimus, jossa tietoturvapäivitykset asennetaan vuosihuollon yhteydessä, voi jättää sivuston auki viikoiksi tai kuukausiksi. Sopimuksesta kannattaa tarkistaa, kuuluvatko tietoturvapäivitykset siihen ja kuinka nopeasti ne luvataan asentaa.

Samaan siivoukseen sopii kolme pienempää toimea. Google paikkasi Chrome-selaimesta kymmeniä haavoittuvuuksia, joista ainakin yksi voi Tivin mukaan antaa hyökkääjälle pääsyn laitteeseen. Korjaus tulee voimaan vasta, kun selain käynnistetään uudelleen. Intialaiselta eMudhralta, joka myöntää sähköisiä todistuksia eli varmenteita, vuoti kahdeksan varmenteiden allekirjoittamiseen käytetyn varmenteen salaiset avaimet. Niillä tehty väärennös voisi saada huijaussivun näyttämään selaimessa aidolta lukkokuvausta myöten, minkä vuoksi puhelimen, tietokoneen ja selaimen päivitykset kannattaa asentaa heti niiden ilmestyttyä. Palvelinten omat luettelot luotetuista varmenteista eivät päivity selaimen mukana, joten tieto kuuluu välittää myös yrityksen palvelimista huolehtivalle. Kolmas toimi on kaksivaiheinen tunnistautuminen eli kirjautumisen vahvistaminen salasanan lisäksi esimerkiksi puhelimella. Se kannattaa ottaa käyttöön ainakin sähköpostissa, kirjanpito-ohjelmassa ja niiden tekoälypalvelujen tileillä, joihin on kertynyt yrityksen asiakirjoja ja keskusteluja.

Monen pienen yrityksen kotisivu on tehty vuosia sitten verkkotoimistossa, jonka yhteystiedot löytyvät enää vanhasta laskusta. Sen laskun kaivaminen esiin voi olla viikon tärkein tietoturvatyö.

Toinen viikon käytännön muutos ei koske kuukausimaksulla ChatGPT:tä, Geminiä tai Copilotia käyttävää. Se koskee yrityksiä, joiden asiakaspalvelubotti, tarjouslaskuri tai sopimusten tiivistäjä on rakennettu jonkin tekoälymallin varaan ja joiden lasku kasvaa jokaisen käsitellyn viestin myötä. Tällainen käyttö laskutetaan sanapaloina, joiksi malli pilkkoo lukemansa ja kirjoittamansa tekstin. Moonshot pyytää uudesta mallistaan 2 dollaria jokaisesta miljoonasta mallille syötetystä sanapalasta ja 10 dollaria jokaisesta miljoonasta sanapalasta, jonka malli kirjoittaa vastaukseksi. Miljoona sanapalaa vastaa satoja sivuja tekstiä. Hinta on sama kuin OpenAI:n vastaavalla mallilla ja puolet Anthropicin vastaavan mallin hinnasta, minkä vuoksi mallin valinnalla on satojen tai tuhansien eurojen kuukausilaskussa jo merkitystä.

Hintapaineen taustalla on muutakin kuin yksi uusi malli. DeepSeek julkaisi vapaasti käytettäväksi ohjelmiston, jolla se pyörittää omia mallejaan palvelimillaan suuressa mittakaavassa. Latent Space kuvaa ohjelmiston tuottavan noin 80 prosentin katteen, mikä on uutiskirjeen tieto eikä riippumattomasti vahvistettu luku. Paikkansa pitäessään luku kertoo, että käyttöhinnoissa on vielä laskuvaraa ja että tehokas tapa tarjota malleja on nyt minkä tahansa palveluntarjoajan käytettävissä Euroopassakin.

Halpa hinta ei ratkaise sitä, minne tiedot kulkevat. Moonshot on pekingiläinen yhtiö. DeepSeekin oma palvelu on puolestaan herättänyt EU:n tietosuojaviranomaisissa huolta käyttäjätietojen käsittelystä Kiinassa. Asiakkaiden henkilötietoja ei kannata lähettää kiinalaisen yhtiön palveluun ennen kuin on selvitetty, missä maassa tiedot käsitellään ja millä perusteella ne saa EU:n tietosuoja-asetuksen mukaan siirtää sinne. Luvattu vapaa julkaisu muuttaisi aselman, koska ladattavaa mallia voisi tarjota myös eurooppalainen pilvipalvelu, jolloin tiedot pysyisivät EU:ssa ja sopimuskumppani olisi tuttu.

Samalla viikolla mallin vaihtaminen helpottui monesta suunnasta. Googlen pilvettä jo käyttävä yritys voi ottaa OpenAI:n mallit ja Anthropicin Clauden käyttöön saman tilin ja laskun kautta. Clauden palvelua pyörittää silti Anthropic

itse, minkä vuoksi kannattaa selvittää, kumman yhtiön ehdot koskevat yrityksen tietoja. Suomalaisissa it-sopimuksissa yleisesti käytetyn sopimus pohjan uudet IT2026-ehdot tekevät Tivin mukaan toimittajan vaihtamisesta maksutonta ja ulottavat pelisäännöt myös tekoälyyn. Ehdot sitovat vain silloin, kun sopimus viittaa niihin. Seuraavassa tekoälypalvelun hankinnassa kannattaa siksi kysyä, käyttääkö toimittaja uusia ehtoja ja mitä sopimus sanoo yrityksen omien tietojen siirtämisestä pois palvelusta. EU:n datasäädös vie samaan suuntaan: pilvipalvelun vaihtamisesta ei saa periä vaihtomaksuja 12. tammikuuta 2027 alkaen.

Palvelun rakentajalta kannattaa kysyä, minkä yhtiön malli palvelussa pyörii, voiko sen vaihtaa ja missä tiedot käsitellään. Vaihtamisen vapaudesta on hyötyä vasta, kun joku yrityksessä tietää nämä kolme vastausta.

02 Työkalut ja avustajat arjessa

Mistral julkaisi työkalupaketin, jolla yritys voi rakentaa puheavustajan, joka kuuntelee, puhuu ja hoitaa tehtäviä yli 50 kielellä. Avustaja kirjoittaa puheen tekstiksi, vastaa puhumalla eri tunnesävyin ja osaa käynnistää toimintoja muissa järjestelmissä, jolloin puhelinvastaaja voisi esimerkiksi varata ajan kalenteriin. Paketin voi ajaa Mistralin EU:ssa sijaitsevassa palvelussa tai yrityksen omilla palvelimilla. Fysioterapia-aseman ajanvarauspuhelussa potilas kertoo usein vaivastaan, jolloin puhettalenne on terveystä koskevaa arkaluonteista henkilötietoa. EU:ssa tai omalla palvelimella tapahtuva käsittely poistaa tietosuojan hankalimman kysymyksen eli tietojen siirron Euroopan ulkopuolelle. Palveluntarjoajan kanssa tarvitaan silti sopimus tietojen käsittelystä. EU:n tekoälyasetus edellyttää lisäksi, että soittajalle kerrotaan hänen puhuvan tekoälyn kanssa, ellei se ole muutenkin ilmeistä. Kaikki tieto perustuu toistaiseksi Mistralin omaan julkistukseen. Ennen sitoutumista työkalua kannattaa kokeilla oikeilla puheluilla, joissa on murretta, taustamelua ja päällekkäin puhumista.

Iso osa yrityksen rutiineista on pieniä päätöksiä: kuuluuko sähköposti laskutukselle vai asiakaspalvelulle, onko viesti reklamaatio ja pitääkö siihen vastata vielä tänään. OpenAI julkaisi torstaina ohjelmistokehittäjille palvelun tällaisiin nopeisiin, rajattuihin ja edullisiin päätöksiin. Palvelua käytetään rajapinnan kautta: yrityksen oma ohjelma lähettää kysymyksen ja saa vastauksen ilman, että kukaan istuu ruudun ääressä. Julkistuksen tiivistelmä ei kertonut palvelun hintoja.

Samana vuorokautena sama ajatus tuli saataville myös omalle koneelle. Ollama-yhtiön ilmainen ohjelma, jolla tekoälymalleja ajetaan omalla tietokoneella, sai päätöksiin erikoistuneen malliperheen neljässä eri koossa. Julkaisun esimerkissä mallille annetaan asiakkaan viesti kahdesti veloitetusta kortista ja kysytään, onko kyse laskutusasiasta. Malli vastaa kyllä tai ei ja kertoo samalla todennäköisyyden, joka kuvaa sen varmuutta. Epävarmat tapaukset voi siten ohjata ihmiselle ja selvät suoraan oikeaan jonoon. Mallien pienimmät versiot ovat niin kevyitä, että ne toimivat tavallisella kannettavalla tietokoneella. Asiakkaan viesti ei tällöin kulje minkään ulkopuolisen palvelun kautta, mikä keventää tietosuojaselvityksiä.

OpenAI julkaisi tiistaina asiakastarinan yhdysvaltalaisesta franchise-ravintoloitsijasta. Tarinan mukaan kassan täsmäytykseen kuluva aika lyheni 94 % ja ruokalistan hintojen päivittämiseen kuluva aika 95 %. Tarina on OpenAI:n markkinointia ja sen luvut ovat yrittäjän ja yhtiön omia. Suomalaiseen yritykseen siitä siirtyy lähinnä havainto tehtävien laadusta. Kassan täsmäytys ja hintojen päivitys ovat toistuvia taulukkotöitä, joiden tuloksen yrittäjä pystyy tarkistamaan, koska hän tietää tarkalleen, miltä oikea tulos näyttää.

Sama Ollaman ohjelma oppi myös puhumaan. Applen Mac-koneilla se lukee tekstin tai tekoälyn vastauksen ääneen sitä mukaa kuin vastaus syntyy, jolloin pitkän vastauksen kuuntelun voi aloittaa ennen kuin vastaus on valmis. Saman vuorokauden aikana puhe tuli myös avoimeen ohjelmistoon, jonka päällä moni omalla koneella toimiva tekoälysovellus

pyörii. Siihen lisätty puhemalli on niin pieni, ettei ääneen lukeminen vaadi tehokasta näytönohjainta. Kumpikaan julkistus ei kerro, puhuvatko äänet suomea. Suomenkielisessä asiakaskäytössä tulos kannattaa kuunnella itse ennen kuin sen varaan rakentaa mitään.

Kiinnostus omalla laitteistolla ajettaviin malleihin näkyi myös latausluvuissa. Latent Space arvioi tiistaina, etteivät DeepSeekin ja MiniMaxin uudet kevyet ja edulliset malliversiot ole vielä laajassa käytössä. Itävallan tiede- ja teknologiainstituutin tutkimusryhmä oli kuitenkin pakannut DeepSeekin kevytmallista version omalle laitteistolle ja loppuviikosta se oli tekoälymalleja jakavan Hugging Facen nousulistalla yli 400 000 latauksella. Pakattu malli toimii ilman yhteyttä DeepSeekin omaan palveluun, jolloin käsiteltävä aineisto ei kulje yhtiölle. Ladattavan mallin lähteellä on silti väliä. Tietoturvayhtiö ProjectDiscovery varoitti keskiviikkona, että verkosta ladattavat muokatut mallit, joista on poistettu kyky kieltäytyä haitallisista pyynnöistä, voivat altistaa käyttäjänsä murrolle. Tuntemattoman tahon muokkaama malli on tietoturvan kannalta samassa asemassa kuin mikä tahansa tuntemattomasta lähteestä ladattu ohjelma.

Ohjelmointiin tekoälyä käyttäville viikko toi useita korjauksia, joista osa koskee sitä, mitä työkalu tekee käyttäjän huomaamatta. Maanantaina Anthropic korjasi ohjelmointityökalunsa turvallisuuslisäosan. Sen koodintarkastukset olivat jääneet hiljaa tekemättä, jos käyttäjä oli käynnistänyt työkalun asetuksella, joka rajoittaa ylimääräistä verkkoliikennettä tai käyttötietojen lähettämistä yhtiölle. Tietosuoja varten tehty valinta kytki siis samalla pois turvallisuustarkastuksen. Nyt lisäosa ilmoittaa asiasta enintään kerran viikossa ja tarkastaa myös tiedostot, joissa koodi ja sitä selittävä teksti ovat samassa asiakirjassa. Lauantain päivitys poisti saman ilmiön toisen puolen: käyttötietojen keruun estäneet eivät enää jää paitsi taustalla käynnissä olevien tehtävien seurannasta, joka toimii nyt myös työkalua Amazonin, Googlen tai Microsoftin pilven kautta käyttäville. Samassa päivityksessä korjattiin virhe, jossa automaattinen siivous saattoi poistaa tuoreita tiedostoja, kuten avustajalle liitettyjä kuvia, jos säilytysajaksi oli asetettu hyvin pitkä aika. Perjantaina työkalusta paikattiin useita kohtia, joissa sen eristetty tila eli suojattu alue, jonka sisällä tekoäly saa ajaa komentoja, ei pitänyt kaikkea sisällään. Lauantaina kehittäjien keskustelupalstalla Hacker Newsissa jaettiin lisäksi yksittäisen käyttäjän somejulkaisu, jonka mukaan sama työkalu lähettää käyttäjän sähköpostiosoitteen verkkoon tekemiensä pyyntöjen mukana. Keskustelu jäi muutaman kommentin mittaiseksi, eikä väitteelle ole esitetty riippumatonta vahvistusta.

Muiden yhtiöiden ohjelmointityökaluissa muutokset liittyivät käyttöönottoon. OpenAI:n ohjelmointityökalu sai ominaisuuksia, joilla yritys voi liittää sen omiin kirjautumista vaativiin järjestelmiinsä. Googlen komentorivillä toimiva ohjelmointiavustaja oppi tuomaan Anthropicin työkaluun tehtyjä automaattiosääntöjä eli komentoja, jotka ajetaan tiettyjen tapahtumien yhteydessä. Google kertoo kuitenkin itse, ettei kyseistä työkalua ole viritetty yhtiön uusimmille Gemini-malleille, joita varten se suosittelee toista työkaluaan. Yrittäjän, jonka sovellusta tai verkkosivua tekee alihankkija tekoälyavustajien kanssa, kannattaa kysyä kaksi asiaa: onko työkalut päivitetty ja näkyvätkö turvallisuustarkastusten tulokset jossakin. Hiljaa pois päältä ollut tarkastus näyttää ulkopuolelta täsmälleen samalta kuin tarkastus, joka ei löytänyt mitään.

Pitkäikäisin viikon työkalumuutoksista voi olla työhjepaketien yleistymisen. Työhjepaketti on kansio, johon on koottu kirjallinen ohje, mallipohjat ja tarvittaessa pieni ohjelma. Avustaja ottaa paketin käyttöön silloin, kun tehtävä tulee vastaan, esimerkiksi kun sen pitää koota tarjous yrityksen omaan pohjaan. Anthropic kehitti muodon alun perin Claudelle ja lahjoitti sen viikolla Linux Foundationin alaiselle Agentic AI Foundationille, jonka se perusti viime vuonna yhdessä Blockin ja OpenAI:n kanssa. Samassa säätiossa on jo yhteinen määrittely, jolla avustajat liitetään muihin ohjelmiin, kuten kalenteriin tai asiakasrekisteriin.

Viikonloppuna leviämisestä näkyi ensimmäisiä merkkejä. Amazonin pilvipalvelun virallinen työkalupaketti ohjelmointiavustajille sisältää työhjepaketteja, joiden kerrotaan toimivan Anthropicin, OpenAI:n ja usean muun yhtiön ohjelmointiavustajissa. Avoimeen keskustelusovellukseen, jonka yritys voi asentaa omalle palvelimelleen, tuli asetus, jolla paketit ladataan vasta silloin, kun malli tarvitsee niitä. Muutos pienentää jokaisen viestin mukana lähetettävää tekstiä ja sitä kautta laskua, koska tekoälypalvelut laskuttavat käsitellyn tekstin määrän mukaan. Samaan päivitykseen tuli keskusteluun merkintä kohdasta, jossa vanhemmat viestit on tiivistetty. Pitkässä keskustelussa avustaja säilyttää alun viestit lopulta vain tiivistelmänä ja merkinnästä käyttäjä näkee, mistä kohdasta alkaen yksityiskohdat ovat saattaneet kadota. Lahjoitus ja nämä ensimmäiset käyttöönotot eivät vielä tarkoita, että jokainen avustaja osaisi lukea paketteja. Omiin työhjeisiin käytetty aika sitoo kuitenkin yhteen toimittajaan vähemmän kuin vielä viikko sitten.

Yksityisyyden puolella opettavaisin tapaus oli ilmainen liitântä, jolla Claude-avustaja pääsee lukemaan käyttäjän henkilökohtaisia WhatsApp-viestejä ilman WhatsAppin yritysversiota tai omaa palvelinta. Tekijä kuvaa ratkaisuaan

itse harmaaksi alueeksi ja rajaa sen yhden käyttäjän käyttöön. Viestihistoriassa on myös perheenjäsenen, asiakkaiden ja kollegoiden kirjoittamaa tekstiä. Kukaan heistä ei ole antanut viestejään tekoälypalvelun käsiteltäviksi. WhatsApp on aiemmin sulkenut tilejä, joita on käytetty epävirallisten sovellusten kautta, minkä vuoksi kokeilun hintana voi olla koko tili. Yrittäjälle, joka hoitaa asiakasviestintää WhatsAppissa, kytkös tarkoittaa asiakkaiden henkilötietojen siirtämistä palveluun, jonka käsittelystä yritys vastaa EU:n tietosuoja-asetuksen nojalla.

Työpaikoilla sama kysymys koskee tilin valintaa. Uber ottaa ChatGPT:n työpaikkaversioon käyttöön yli 30 000 työntekijälleen. Yrityksille myytävissä versioissa käyttäjien aineistoa ei oletuksena käytetä mallien kouluttamiseen. Henkilökohtaisella tilillä keskustelut voivat päättyä mallien kehittämiseen, ellei käyttäjä ole kieltänyt sitä asetuksista. Kuvissa yksityisyyttä uhkaa kehittäjien suosioon noussut ilmainen työkalu, joka lupaa palauttaa mosaiikilla peitettyjä

kohtia videoista ja kuvista kotikoneen näytönohjaimella. Kasvot, rekisterikilvet ja asiakirjojen tiedot kannattaa siksi peittää umpinaisella peitolla sumennuksen tai pikselöinnin sijaan.

Tekoälyn arvioihin luottamisesta viikko toi kaksi muistutusta. Mallien vertailuun keskittyvän sivuston kirjoituksen mukaan kaikilla tekoälymalleilla on sama ”kiistanalainen” näkemys ihmisistä. Kirjoitus keräsi Hacker Newsissa yli 570 kommenttia. Yksittäisenäkin kokeiluna tulos muistuttaa, että ChatGPT:n vastauksen tarkistaminen Geminiltä ei ole sama asia kuin riippumaton toinen näkemys. Yhdysvaltalainen tutkimuslaitos Ai2 puolestaan julkaisi testin, jonka mukaan tekoälyt perustelevat päätöksiään eri tavoin sen mukaan, mitä väestöryhmää päätös koskee. Erot näkyvät myös silloin, kun lopullinen valinta näyttää tasapuoliselta. Työhakemuksia tai lainahakemuksia tekoälyn avulla seulovan kannattaa lukea perustelut eikä pelkkää lopputulosta.

03 Markkina, sääntely ja turvallisuus

Pilvimarkkinan viikon suurin liike oli se, että Google myy nyt kilpailijansa OpenAI:n kärkeille ja ohjelmointiavustajan työntekijöitä omassa valikoimassaan Geminin vieressä. Kaksi päivää myöhemmin Anthropic avasi palvelun, jossa Googlen pilven asiakas ottaa Clauden käyttöön omalla Google-tilillään ja maksaa sen Googlen kauppapaikan kautta. Aiemmin Googlen pilvessä tarjottua Claudea pyöritti Google itse. Uudessa palvelussa sitä pyörittää Anthropic, vaikka käyttöoikeuksia hallitaan Googlen työkaluilla. Microsoftin pilvessä OpenAI:n mallit ovat olleet tarjolla jo vuosia. Mallin valinta ei siksi enää sido pilvipalvelun valintaan yhtä tiukasti kuin ennen. Kumpikaan julkistus ei kerro, millä alueilla palvelut ovat saatavilla tai käsitelläänkö tiedot EU:ssa. Eurooppalaisen ostajan kannattaa selvittää tämä ennen kuin palveluun syötetään henkilötietoja. Mistral puolestaan kertoi kehittävänsä Nvidian kanssa tekoälyn nojaavia työntekijöitä teollisuuden suunnittelu- ja insinööritöihin. Hintoja, aikataulua tai asiakkaita tiedotteessa ei mainittu.

Mallien kehittäjät lähestyvät samalla asiakkaidensa toimialoja. OpenAI ostaa yhtiön Orchestra Bion ja OpenAI:n mukaan kaupalla tuodaan tekoälyä lääkeyhtiöiden tutkimus- ja kehitystyön arkeen. Loppuviikosta yhtiö kertoi ajavansa tuhansia tekoälyagentteja biologisen tutkimusaineiston parissa uusien johtolankojen löytämiseksi. Tiistaina OpenAI esitteli myös ChatGPT:n parantuneita taitoja harvinaisemmissa ohjelmointikielissä, kuten vuosikymmeniä

vanhassa COBOLissa, jolla on kirjoitettu monen pankin ja vakuutusyhtiön taustajärjestelmät. Mittaukset ovat yhtiön omia. Hacker Newsissa yli 500 kommenttia keräsi kysymys siitä, miksi OpenAI ja Anthropic myyvät mallejaan muille sen sijaan, että perustaisivat niiden varaan itse 100 yritystä.

Hintakilpailu kiristyi viikon lopulla. Moonshot myy perushinnan lisäksi kahdesta kolmeen kertaa nopeampaa versiota kaksinkertaisella hinnalla, jolloin vastausnopeudesta tulee erikseen hinnoiteltu tuote. Myös Anthropicin tehokkaimmasta mallista on Googlen kautta tarjolla nopea tila. Latent Space etsikoi, että vapaasti ladattavat mallit ovat enää noin kuukauden jäljessä niistä, joita voi käyttää vain valmistajan omassa palvelussa. Saman jutun alaotsikko kuittasi väitteen vitsillä, jonka mukaan näin kertovat maksetut hehkuttajat. Avoimia malleja seuraava uutiskirje Interconnects totesi tiistaina länsimaisten avointen mallien ottavan asemiaan takaisin. Samassa katsauksessa se nosti esiin uudet sääntelytoimet sekä Yhdysvalloissa että Kiinassa. Sunnuntaina sama uutiskirje omisti viikon podcastinsa Moonshotin mallille ja sille, mitä se kutsuu uudeksi kybertodellisuudeksi.

Mainonta on siirtymässä tekoälyavustajiin. OpenAI vie ChatGPT:n mainokset 26 uuteen maahan ja yhtiön mukaan mainosalustaa on käyttänyt jo yli 100 000 mainostajaa. OpenAI:n kuvauksen mukaan mainosjärjestelmä tulkitsee keskustelun tarkoituksen ja valitsee mainoksen sen

perusteella, mitä käyttäjä on vertailemassa tai ostamassa. Lupaus mainosten pitämisestä erillään vastauksista on OpenAI:n oma eikä ulkopuolisen tarkastuksen tulos. Julkistuksesta ei selviä, onko Suomi uusien maiden joukossa. Samana päivänä Euroopan komissio avasi muodollisen menettelyn Booking.comia vastaan digipalvelusäädöksen eli suuria verkkoalustoja velvoittavan EU-lain nojalla. Komissio selvittää, onko Booking.com torjunut riittävästi huijausilmoituksia ja tietojenkalastelua ja onko hakutuloksissa korkeammalle maksettuja kohteita merkitty selvästi mainoksiksi. Päivää aiemmin komissio julkaisi saman lain pohjalta selvityksen verkon talouspetoksista. Sen mukaan niitä esiintyy kaikenlaisissa verkon välityspalveluissa suurimmat alustat ja hakukoneet mukaan lukien. Selvitys ei sellaisenaan velvoita alustoja uusiin toimiin.

Kaupan ja asiakkaan suhteeseen tulee samaan aikaan uusi osapuoli. The Vergen mukaan verkkosivustot ovat alkaneet torjua avustajia, jotka avaavat sivuja, täyttävät lomakkeita ja hoitavat varauksia käyttäjän puolesta. Ajanvarauspalvelu, joka torjuu agentit, voi menettää asiakkaita, jotka haluavat hoitaa varauksen avustajan kautta. Avoimesti agentit sisään päästävä sivusto taas ottaa vastaan liikennettä, jonka takana olevaa ihmistä se ei näe. Moni sivusto toimii palveluntarjoajan valmiilla asetuksilla, jotka voivat torjua tekoälyliikennettä ilman ylläpitäjän erillistä päätöstä. Oman verkkopalvelun asetukset kannattaa siksi tarkistaa, jotta linja on oma eikä oletus. Tekoälymalleja arvioiva yhtiö Vals.ai testasi samaan aikaan, kuinka pitkälle avustajat menevät yrittäessään saada hyvityksen, jota ei ole mahdollista saada. Kauppiaan silmissä avustajan liioittelu tai painostus on asiakkaan omaa puhetta.

Sääntelyn viikko alkoi maanantaina EU:n ensimmäisestä tekoälyasetusta tukevasta yhdenmukaistetusta standardista. Se koskee laatujärjestelmää eli niitä kirjattuja toimintatapoja, joilla suuririskisen tekoälyjärjestelmän kehittäjä huolehtii tuotteensa laadusta ja turvallisuudesta koko sen elinkaaren ajan. Asetuksen 17 artikla velvoittaa kehittäjän ylläpitämään tällaista järjestelmää, johon kirjataan esimerkiksi testaus, datan hallinta ja vakavista häiriöistä ilmoittaminen viranomaisille. Standardin noudattamisesta seuraa vaatimustenmukaisuusolettama: lähtökohtaisesti katsotaan, että standardin kattama lain vaatimus on täytetty. Tekoälyä ostavaa tai käyttävää yritystä laatujärjestelmävelvoite ei koske. Ostaja voi silti kysyä toimittajaltaan, noudattaako tämän laatujärjestelmä uutta eurooppalaista standardia, jonka tunnus on EN 18286, vai aikooko toimittaja osoittaa vaatimustenmukaisuuden muulla tavalla. Standardien viivästyminen on ollut yksi perusteista, joilla suuririskisiä

järjestelmiä koskevien velvoitteiden voimaantuloa on EU:ssa esitetty lykättäväksi. Ensimmäinen standardi kattaa velvoitteista vain yhden.

Samana päivänä Yhdysvalloissa puitiin toisenlaista mallia. Tom's Hardwaren mukaan Anthropicin ehdottama turvallisuuslaki tulisi sovellettavaksi vasta, kun tekoälymalli uhkaa aiheuttaa vähintään 50 ihmisen kuoleman tai 1 miljardin dollarin vahingot. Lehti tulkitsee ehdotuksen tekevän Yhdysvaltain kauppaministeriöstä käytännössä ainoan viranomaisen, joka voisi pysäyttää tekoälymallin. Samansuuruinen vahinkoraja on Kalifornian vuonna 2025 säättämässä tekoälylaissa. EU:ssa velvoitteet määräytyvät sen sijaan käyttötarkoituksen mukaan: työnhakijoita lajittevalta järjestelmältä vaaditaan laatujärjestelmä ennen markkinoille tuloa riippumatta siitä, onko se koskaan aiheuttanut kenellekään vahinkoa. Yhdysvaltain linjaukset koskevat niitä malleja, joiden varaan suuri osa Suomessakin käytetyistä palveluista rakentuu. Torstaina ja lauantaina sääntelyrintamalla oli hiljaista.

Tietoturvan viikkoa hallitsi PHP-tapaus. Toinen aukkojen kahdesta löytäjästä oli OpenAI:n ohjelma, joka etsii tekoälyn avulla haavoittuvuuksia laajasti käytetyistä ohjelmistoista. Myös tietoturvayhtiö Wiz kertoo löytäneensä tekoälyn avulla kriittisen aukon, jonka kautta ulkopuolinen voi ajaa omaa koodiaan sivustoa pyörittävällä palvelimella. Perussuomalaisten verkkosivujen katkon taustalla on puoluesihteerin mukaan todennäköisesti juuri PHP-aukko. Tivin analyysi kysyi, pitikö verkkopalvelut sulkea paniikissa. Lehden vastaus oli jo otsikossa: vaihtoehtona olisi voinut olla uusi Vastaamo eli tapaus, jollainen nähtiin vuonna 2020, kun murtautuja vei psykoterapiakeskuksen potilastiedot ja kiristi niillä potilaita.

Puolustajat liikkuivat samaan aikaan. OpenAI julkaisi suunnitelman, jossa huippumalleja annetaan tietoturvan puolustajien käyttöön ja julkisen sektorin tietoturvaa uudistetaan. PHP-säätiö kertoi, miten se vahvistaa kielen tietoturvaa vuodon jälkeen. Anthropicin tuore tutkimus käsittelee sitä, miten yhtiön tekoälyllä voi karsia ohjelmista kohtia, joiden kautta niihin voi murtautua. Nvidia puolestaan muistutti, että agentti voi suorittaa tehtävänsä loppuun ja silti ylittää matkalla tietoturvarajan. OpenAI kuvasi lauantaina, miten se eristää tehokkaimpien malliensa koulutuksen ja testauksen internetistä irrotettuihin ympäristöihin ja koettelee suojausjärjestelmillä tarkoituksellisia murtoyrityksiä. Kuvaus on yhtiön oma eikä ulkopuolisen tarkastajan.

Tivin mukaan myös suomalaisen pelikaupan asiakastietoihin on murtauduttu. Tapauksesta on ilmoitettu viranomaisille, mutta yhteyttä PHP-aukkoon ei ole kerrottu. Murtojen

jälkeen liikkeelle lähtee usein huijausviestejä kaupan nimissä, minkä vuoksi salasanan vaihtoa pyytävään viestiin kannattaa suhtautua epäillen ja kirjautua palveluun suoraan eikä viestin linkin kautta.

Tekoälyn kasvu törmää yhä selvemmin fyysisiin rajoihin. Tivin mukaan Microsoftin ja Nscalen yhteinen datakeskushanke näyttää kaatuneen, koska sähköverkko ei riitä. Toisessa jutussaan lehti toteaa, etteivät Suomen nykyiset valokuituyhteydet sellaisinaan riitä datakeskusten tarpeisiin. Keskiviikkona Tivi kertoi suljettujen ydinvoimaloiden uudelleenkäynnistyksistä datakeskusten sähköntarpeen kattamiseksi. Torstaina Taiwanissa miljardiluokan siru- ja muistiyhtiön työntekijät menivät lakkoon ensimmäistä kertaa maan sirutehtaiden historiassa. Muistipiirit ovat tekoälypalvelinten kriittinen osa, minkä vuoksi pitkittyvä lakko voisi näkyä palvelinten ja lopulta pilvipalvelujen hinnoissa.

Suomessa Elisa etenee suunnitelmassaan rakentaa yhdessä Nvidian ja Nokian kanssa Euroopan ensimmäinen

tekoälypohjainen matkapuhelinverkko. Nvidian ja Nokian hankkeissa tällä tarkoitetaan yleensä verkkoa, jonka tukiasemien laskenta tapahtuu samoilla piireillä, joilla voidaan ajaa myös tekoälypalveluja. Tivin mukaan Nvidia on nostanut Elisan listalle, jolle mahtuu vain kuusi yhtiötä. Asiakkaan liittymässä muutos ei vielä näy. Suomalaiset verkkokaupat aloittavat Business Finlandin miljoonarahoituksella yhteistyön, jonka tavoite on tehdä Suomesta tekoälyä hyödyntävän kaupan edelläkävijä ja tuoda alalle jopa 1 miljardin euron lisäliikevaihto. Miljardi on hankkeen tavoite eikä ennuste. Tekoälyyhdistys AI Finlandin syyskuun jäsenkyselyssä 83 prosenttia organisaatioista kertoi lisänneensä tekoälyn käyttöä varsinaisessa työssä edellisen neljänneksen aikana. Kysely kuvaa suomalaisen tekoälykentän aktiivisinta osaa eikä keskivertoista suomalaista yritystä. Samalla sivustolla esitetyille maailmanlaajuisille luvuille, joiden mukaan 84 prosenttia johtajista uskoo tekoälyn muuttavan toimialaansa ja 93 prosenttia organisaatioista kamppailee käytön laajentamisessa, ei kerrottu lähdeä.

04 Mitä seurata

Moonshotin mallin julkaisu 4. marraskuuta. Ratkaisevaa on, millä lisenssillä eli käyttöehdoilla malli julkaistaan, salliiiko lisenssi kaupallisen käytön ja ilmestyykö malli eurooppalaisten pilvipalvelujen hinnastoihin. Vasta silloin lähes kärkitason mallia voi käyttää niin, että tiedot pysyvät EU:ssa ilman omaa konesalia. Toinen mittari on se, vahvistavatko riippumattomat vertailut yhtiön oman väitteen tasavertaisuudesta yhdysvaltalaisen kärkimallien kanssa.

Korjausten tahti. OpenAI:n korjausohjelma, sen suunnitelma antaa huippumalleja puolustajien käyttöön ja PHP-säätöön lupaama tietoturvan vahvistaminen ennakoivat samanlaisia löytöjä lisää. Sunnuntain vastaisena yönä avoin hanke, joka ylläpitää avustajien yhteisen liitännäsmäärittelyn esimerkkiliitännöjä, kirjasi neljän minuutin sisällä tietoturvapäivitykset yhdeksään ohjelmakirjastoon tunnettujen aukkojen vuoksi. Muutoksiin on merkitty ylläpitäjän rinnalle tekijäksi Anthropicin tekoälymalli. Kehityksen tunnistaa siitä, että suosittujen ohjelmistojen kiireellisiä päivityksiä tulee entistä tiheämmin. Toinen merkki on se, että tekoälyn löytämät haavoittuvuudet muuttuvat Kyberturvallisuuskeskuksen katsauksissa erikoisuudesta arkiuutiseksi.

EU:n seuraavat standardit. Tekoälyasetuksen muista vaatimuksista, kuten riskienhallinnasta, datan laadusta, tarkkuudesta ja kyberturvallisuudesta, valmistellaan omia

standardejaan. Etenemisen tunnistaa uusista viittauksista EU:n virallisessa lehdessä ja siitä, että toimittajat alkavat viitata standardeihin tarjouksissaan ja sopimuksissaan. Ai2:n reiluudesta osuu samaan kohtaan, koska rekrytoinnissa ja luotonannossa käytettävät järjestelmät ovat asetuksen tiukimmin säänneltyä aluetta. Muutoksen merkki on se, että toimittajat raportoivat tasapuolisuudesta myös perustelujen eivätkä vain lopputulosten osalta.

Muistipiirilakon kesto. Taiwanin lakon vaikutus tekoälypalvelujen hintoihin ratkeaa sen mukaan, kuinka pitkään työtaistelu kestää. Ensimmäiset merkit näkyisivät palvelinlaitteiden hinnoissa ja toimitusajoissa.

Avustajien mainokset Suomeen. Seurattavaa on, milloin ChatGPT:n mainokset tulevat Suomeen ja miten ne merkitään. Avoinna on myös, kohdistuuko komission katse avustajien mainoksiin samalla tavalla kuin varauspalvelun maksettuun näkyvyyteen.

Agentit ja verkkosivut. Verkkopalvelut voivat rakentaa agenteille omia tunnistettuja sisäänkäyntejä tai jatkaa estämistä. Asiakkaalle ero näkyy siinä, toimiiko avustajan kautta tehty varaus vai pitääkö se tehdä itse. Samaan aikaan kehittäjien suosioon nousi avoin agentti, joka toimii käyttäjän omalla koneella ja käyttää hänen tiedostojaan sekä sovelluksia, joihin hän on jo kirjautunut.

Asiakkaat ja tekoälyllä tehty sisältö. Pelikauppa Steamin käyttäjät vaativat suodatinta, jolla tekoälyllä tehdyt pelit voi piilottaa. Vaatimus nousi, kun uusien pelien esittelytapahtuma täyttyi nopeasti tekoälyn avulla tehdyistä peleistä. Samalla viikolla esiteltiin pieni avoin malli, joka on koulutettu kiinalaisen Alibaban mallin pohjalta kirjoittamaan ilman tekoälytekstin tuttuja maneeereja. Sen laatuvaatteet ovat tekijän omia. Viikonloppuna suosiota sai Together AI:n tekemä työohjepaketti verkkosivujen suunnitteluun, jonka luvataan kieltäytyvän näyttämästä tekoälyn tekemältä. Merkki kehityksen jatkumisesta on se, että kauppapaikat alkavat merkitä tai suodattaa tekoälysisältöä. Muutos koskisi myös yrityksiä, jotka myyvät tekoälyllä tuotettuja kuvia, tekstejä tai tuotteita.

Suomea puhuva ääni omalla koneella. Kaksi toisistaan riippumatonta hanketta toi puheen tuottamisen omalle

koneelle saman vuorokauden aikana. Suomalaiselle käyttäjälle merkittävä hetki on se, jolloin paikallisten työkalujen äänivalikoimiin ilmestyy luontevasti suomea puhuva ääni.

Näkyvä vai piilotettu päättely. Maanantaina kehittäjien keskustelupalstan suosituin tekoälyaihe oli Together AI:n tutkimus agenteista, jotka ajattelevat mallin sisäisinä lukuarvoina ja kirjoittavat tekstiä vasta toimiessaan. Monet nykyiset avustajat näyttävät käyttäjälle tekstimuotoisen päättelynsä tai tiivistelmän siitä ja sen avulla voi jäljittää, miten vastaukseen päädyttiin. Lähestymistavan yleistyminen kaventaisi tätä näkymää. Suunnan tunnistaa avustajista, jotka mainostavat nopeutta ja näyttävät samalla vähemmän ajatteluaan.

OSA 2 • TEKNISILLE LUKIJOILLE

Tekninen osa

Teknisille lukijoille: mallit, alustat, työkalut, vertailut ja tietoturva yksityiskohtineen. Osa 1 kertoi viikon olennaisen arkikielellä.

TÄSSÄ OSASSA

- | | | | |
|----|-------------------------------|----|-------------------------------|
| 05 | Suuret toimijat | 09 | Työskentelytavat ja agentit |
| 06 | Avoimet mallit ja ekosysteemi | 10 | Liitännäiset, MCP ja repot |
| 07 | Ajoalustat ja infra | 11 | Tietoturva ja haavoittuvuudet |
| 08 | Benchmarkit ja vertailut | | |

05 Suuret toimijat

Google Cloudista tuli viikon aikana hankintakanava, jonka kautta kolmen kilpailevan laboratorion kärkeimallit päätyvät samalle laskulle. Keskiviikkona OpenAI kertoi, että sen frontier-mallit ja Codex-työnkulut tulevat Google Cloudin Model Gardeniin Geminin rinnalle. Ne hankitaan Google Cloud Marketplacen kautta. Perjantaina Anthropic avasi jakelun Claude Models on Google Cloud (Anthropic managed). Siinä Anthropicin itse operoimaa Claude API:a käytetään Google Cloud -projektin kautta IAM-pääsynhallinnalla ja Marketplace-laskutuksella. Vertex AI:n aiemmassa Claude-jakelussa inferenssiä ajaa Google. Uudessa jakelussa operaattori on Anthropic, mikä näkyy jo siinä, että Opus 5.5:n fast mode on tarjolla ensimmäisestä päivästä. Kummastakaan tiedotteesta ei selviä, millä alueilla dataa käsitellään. OpenAI:n tiedotteesta puuttuu lisäksi tieto siitä, mitkä mallit ovat mukana ja tarjotaanko Codex hallinnoina palveluna vai integraationa. Tiimille, jonka hankinnat kulkevat yhden pilven sopimuksen kautta, järjestely tuo Geminin, OpenAI:n ja Clauden saman IAM-politiikan piiriin. EU-rajaus on silti varmistettava ennen tuotantokäyttöä.

Anthropicin keskiviikkona julkaisema API-versio 2026-10-21 vaatii kirjastojen käyttäjiltä enemmän huomiota kuin päivämääräleima antaa ymmärtää. Sen mukana tulivat uudet pääversiot kaikista asiakaskirjastoista, muun muassa Python SDK 2.0.0, TypeScript SDK 1.0.0, Go SDK 2.0.0, Java SDK 3.0.0 ja C# SDK 13.0.0 sekä ant CLI 2.0.0. Uudet versiot lähettävät uuden API-version oletuksena. Niissä on lisäksi rikkovia muutoksia, jotka eivät riipu API-versiosta. Pelkkä riippuvuuden päivitys voi siksi rikkoa koodin, vaikka API-version kiinnittäisi vanhaan. Muutokset on koottu dokumentteihin Migrating to API version 2026-10-21 ja SDK and CLI changes. Kirjastojen lisenssi vaihtui samalla MIT:stä Apache 2.0:aan. Apache 2.0 sisältää nimenomaisen patenttilisenssin ja edellyttää uudelleenjakelussa muokattujen tiedostojen merkitsemistä sekä mahdollisen NOTICE-tiedoston säilyttämistä. Vaihto koskee käytännössä niitä, jotka toimittavat kirjastot oman tuotteensa mukana tai ylläpitävät niistä haaroja. Enterprise-organisaatioille inference hooks poistui beetasta ja Claude Tag API avautui betana. Claude Tag API:n päätepiisteet lukevat ja muuttavat Slack-kanava- ja työtilakohtaisia asetuksia Admin API -avaimella, jonka laajuus on read:claude_tag tai write:claude_tag. Asetuksia voi näin hallita koodina ja tarkistaa skriptillä.

OpenAI:n viikko rakentui tietoturvan ympärille tavalla, joka kytkeytyy suoraan viikon suurimpaan haavoittuvuustapaukseen. Patch the Planet -hanke kokoaa tutkijat, ylläpitäjät ja häiriönhallinnan koordinoituun julkistukseen niille kriittisille haavoittuvuuksille, joita agentit löytävät avoimesta koodista. Latent Spacen AINews-koosteen mukaan hanke oli toinen kahdesta tahosta, jotka löysivät PHP2Shell-aukot PHP:n ytimeistä GPT-5.6 Sol -mallilla. Democratic governance of frontier AI -sarjan toinen osa hahmottelee mallin, jossa huippumallit annetaan kyberpuolustajien käyttöön vahvempien suojausten kanssa. Cloudfaren asiakastarinassa Daybreak tutkii haavoittuvuuksia ja ehdottaa korjauksia ihmisen katselmoitaviksi. Kummassakaan julkaisussa ei ole mitattuja tuloksia, kuten löydösten määrää tai väärin hälytysten osuutta. Lauantaina yhtiö kuvasi frontier-mallien koulutusympäristön eristyksen vahvistamista offline-ympäristöillä ja suurella laskennalla tehtävällä red-teamingillä. Eristyksen tasoa tai red-teamingin kohteita kirjoitus ei avaa.

Muut OpenAI:n julkistukset hajosivat laajalle. Yhtiö ostaa Orchestra Bion biotieteiden tutkimus- ja kehitystyöhön ja kertoi ajavansa tuhansia agentteja biologisen aineiston parissa uusien tutkimusjohtolankojen löytämiseksi. ChatGPT:n mainokset laajenivat 26 uuteen maahan samana päivänä, jona yhtiö julkaisi teknisen kuvauksen mainospinostaan. Kuvauksessa kielimalli tulkitsee keskustelun aikomuksen ja mainokset pidetään erillään vastauksista. Erottelu on yhtiön oma kuvaus, jota ulkopuolinen taho ei ole aineiston perusteella arvioinut.

Metan Muse Wonder näkyi viikon uutiskirjeissä, mutta aineistossa ei ole mallista omaa teknistä julkaisua. Latent Space luonnehti 13.10. Muse Wonderia ja Wonder Fullia Claude 5.5 -tason frontier-malleiksi ja piti Meta Superintelligence Labsin nousua frontier-laboratorioksi valmiina. The Batch käsitteli 23.10. Muse Wonderia ja Spark Openia. Kyse on uutiskirjeiden arvioista eikä riippumattomasta mittauksesta.

Laboratorioiden turvallisuuskäytäntöjä tarkasteltiin viikolla myös ulkopuolelta. METR julkaisi tiistaina katselmoinnin Anthropicin elokuun 2026 riskiraportin osiosta Misalignment in high-stakes settings. Torstaina se julkaisi raportin OpenAI:n sisäisten agenttivalvontajärjestelmien red team -testauksesta. Iso-Britannian AI Security Institute julkaisi kirjoituksen kohdistusvirheistä omassa välikohtauksessaan.

Kaikista kolmesta aineistossa on vain otsikot tai lyhyet tiivistelmät, joiden perusteella tuloksia ei voi arvioida. Poliitiikan puolella Tom's Hardware raportoi Anthropicin ehdottamasta tekoälyturvallisuuslaista, joka aktivoituisi vasta 50 kuoleman tai 1 miljardin dollarin vahinkojen kynnyksellä.

Kynnys on sama, jolla Kalifornian SB 53 (2025) määrittelee katastrofaalisen riskin kärkimallien kehittäjille. Julkaisun tulkinnan mukaan ehdotus tekisi Yhdysvaltain kauppaministeriöstä käytännössä ainoan viranomaisen, jolla on toimivalta pysäyttää malli.

06 Avoimet mallit ja ekosysteemi

Viikon puhutuin avoin malli ei ole vielä avoin. Artificial Analysis sijoitti perjantaina Moonshot AI:n Kimi K3.1 Pron Intelligence Indexinsä neljänneksi ja kertoi sen nousevan avoimien painojen mallien kärkeen. Futuurimuoto osoittautui olennaiseksi. Moonshot toi mallin 23.10. ainoastaan Kimi API:iin ja lupaa painot erikseen: "Full model weights will be released by Nov 4, 2026". Latent Spacen otsikon merkintä 104BA-2.8T kertoo asiantuntijaseoksesta, jonka 2,8 biljoonasta parametrasta 104 miljardia on aktiivisena tokenia kohden. Painot vievät nelibittisinä noin 1,4 teratavua ja kahdeksanbittisinä noin 2,8 teratavua. Julkaistunakin malli on siis palvelinsalin kiihdytinsolmujen kokoluokkaa. Se ei mahdu työasemaan tai yksittäiseen suurmuistiseen Maciin, vaikka tokenikohtainen laskenta pysyy noin 100 miljardin tiheän mallin tasolla.

API-listahinta on 2 dollaria miljoonalta syötetokenilta ja 10 dollaria miljoonalta tulostetokenilta. Datasetten luojaan blogin vertailussa hinta vastaa GPT-6.1 Solia ja on puolet Claude Opus 5.5:n hinnasta. Highspeed-versio on kahdesta kolmeen kertaa nopeampi kaksinkertaisella hinnalla. Poikkeuksena ovat välimuistiosumina laskutettavat syötetokenit, jotka maksavat saman kuin perusversiossa. Agenttisilmukassa, jossa pitkä etuliite toistuu kutsusta toiseen, lisähinta kohdistuu siten lähinnä tulosteeseen. Moonshotin omissa vertailuissa malli on suunnilleen tasoissa Opus 5.5:n ja GPT-6 Astran kanssa ja voittaa ne muutamassa testissä. Latent Spacen AINews otsikoi sen Fable 5.1 -luokan avoimeksi malliksi, joka kaventaa suljettujen mallien etumatkan kuukauteen. Saman koosteen alaotsikko ("We have Fable at home... or so we are being told by paid skills") ottaa väitteeseen ironisen etäisyyden. Artificial Analysisin koostesijoitusta lukuun ottamatta riippumattomia mittauksia ei vielä ole. Lisenssistäkään ei ole tietoa, vaikka juuri lisenssi ratkaisee, voivatko eurooppalaiset palveluntarjoajat ajaa mallia kaupallisesti. Aiemmat K2-sarjan mallinsa Moonshot julkaisi muokatulla MIT-lisenssillä. Interconnects omisti sunnuntain podcastinsa Kimi K3.1 Prolle, uudelle kybertodellisuudelle ja Metan Muse-malleille.

Toinen viikon kuvio muodostui kolmesta kiinalaisesta kevyen hintaluokan mallista, jotka ilmestyivät

päivämääräpääätteineen muutaman päivän välein. DeepSeek V4.1 Flash 1017 sai Artificial Analysisin Intelligence Indexissä 45 ja Ant Groupin inclusionAI-tiimin Ling 3.1 Flash 1019 sai 44. MiniMax-M3.1-Flashin lukema oli 42. Latent Spacen arvion mukaan yksikään niistä ei ollut alkuvuodesta laajassa käytössä. Johdannaisten tahti kertoo silti kysynnästä. DeepSeekin mallista ilmestyi muutamassa päivässä dealignain FP8-muotoinen UNCENSORED-versio, josta kieltäytymiskäytös on riisuttu ja jolla oli perjantaina 4 141 latausta. ISTA-DASLabin GSQ-RCO-merkitty GGUF-kvantisointi julkaistiin samana päivänä ja oli sunnuntaihin mennessä ladattu 412 392 kertaa. Kolmannen osapuolen johdannaisissa alkuperä ja muutokset on tarkistettava ennen kuin malli päätyy tuotantoon. Inceptionin Mercury 3 jatkoi samalla viikolla diffuusiopohjaista sarjaa, jossa vastaus tuotetaan tarkentamalla useita tokeneita rinnakkain.

Interconnectsin avoimien artefaktien katsaus #25 katsoi tilannetta toisesta suunnasta. Sen alaotsikko Western open models strike back viittaa länsimaisten avoimien painojen kiristämiseen kiinalaisten julkaisijoiden etumatkaa vastaan. Samassa katsauksessa käsitellään Yhdysvaltojen ja Kiinan uusia sääntelytoimia, jotka voivat vaikuttaa siihen, mitä painoja jatkossa voi jakaa ja käyttää.

Länsimaisista julkaisuista käytännöllisin oli Recursiven Flint, joka tuli keskiviikkona Ollamaan. Se on Apache 2.0 -lisensoitu 40 miljardin parametrin MoE-malli, jonka aktiivinen osuus on noin 2,5 miljardia parametria tokenia kohden. Kontekstia on 262 000 tokenia. Kohteena on repotason agenttityö eli vieraan koodin ymmärtäminen, bugien jäljittäminen, ominaisuuksien toteuttaminen ja testien lisääminen. Recursive ilmoittaa omissa arvioinneissaan 54,9 % Terminal-Bench 2.1:ssä ja 74,2 % SWE-bench Verifiedissä. Pieni aktiivinen osuus nopeuttaa dekodeausta. Muistitarpeen ratkaisee silti koko painomassa, joka on nelibittisenä karkeasti 20 Gt. Torstain Ollama-julkaisuehdokas korjasi Flintin vastaukset tilassa think:false niin, että templaatti sulkee ajatteluosion ja malli vastaa suoraan. Hugging Facen trending-listalle nousi samaan aikaan RecursiveSI/flint. Lähteistä ei käy ilmi, onko kyse samasta mallista. Microsoft Researchin samalla viikolla

esittelemällä visualisointikielellä Flint-interactive ei ole yhteyttä kumpaankaan.

Pienten mallien puolella NVIDIAn Nemotron 3.5 Bolt E4B on Gemma 4 -pohjainen näkö-kielimalli nopeisiin paikallisiin keskusteluihin, kuvien ymmärtämiseen ja työkalujen käyttöön 128 000 tokenin kontekstilla. Mallin mukana toimitetaan oma DFlash-luonnosmalli spekulatiivista dekodausta varten. Saman perheen upotusmalli Nemotron-3.5-Embed sai llama.cpp:ssä tokenisointikorjauksen (b11833). Vanhemmilla buildeilla lasketut upotukset eivät siksi välttämättä ole vertailukelpoisia korjatulla versiolla laskettujen kanssa. Hakuindeksi kannattaa rakentaa uudelleen. Qwen3.8 vakiintui viikon aikana johdannaisten pohjaksi. NVIDIA julkaisi siitä paikalliseen ajoon FP4-kvantisoidun kuva-tekstimallin Qwen3.8-Flash-Next-FP4-Local. Saman perheen malleja nimettiin myös llama.cpp:n commit-viesteissä avustajiksi. Tencentin Hy-Memory-30B-A3B on nimeämiskäytännön perusteella 30 miljardin parametrin MoE-malli noin kolmen miljardin aktiivisella osuudella. Nanbeige4.3-33B nousi sunnuntaina listalle 269 tykkäyksellä ennen kuin yhtään latausta oli kirjautunut. Kummankaan mallikortin sisältöä aineistossa ei ole.

Tutkimuksen puolella suurimman yhteisönoston sai Together AI:n tutkijoiden julkaisu Hybrid Reasoning Agents (2026, arXiv 2610.18861), joka keräsi Hacker Newsissa 499 pistettä. Agentti päättelee latenttiavaruudessa jatkuvina piilotiloina ja palaa tekstiin vasta toimiessaan, kuten työkalukutsuissa ja vastauksissa. Linja jatkaa työtä Hao ym. (2024), *Training Large Language Models to Reason in a Continuous Latent Space* (Coconut). Toimintarajapinta pysyy tekstinä, mikä säilyttää yhteensopivuuden nykyisten työkalukutsuformaattien kanssa. Latentti päättely ei kuitenkaan jätä luettavaa ajatusketjua. Ajatusketjun valvontaan nojaavat turvallisuuskäytännöt eivät siksi sovellu siihen sellaisinaan. Koodi on repossa [togethercomputer/HybridAgent](#).

Kapeammalla alalla Ai2:n OlmoEarth v1.3 julkaisi maapallon havainnointiin tarkoitetun perusmallinsa valmiiksi lasketut globaalit upotukset ja selainpohjaisen Embeddings Explorerin. Luokittelu, muutosten havaitseminen ja samankaltaisten alueiden haku onnistuvat näin kevyellä mallilla ilman satelliittiaikasarjojen käsittelyä. Huipputasen luonnehdinta on Ai2:n oma.

07 Ajoalustat ja infra

Viikon suurimman yhteisönoston sai DeepSeekin julkaisema DeepInfer, korkean suorituskyvyn päättelykehys DeepSeekin malleille. Hacker Newsissa se keräsi 2 291 pistettä ja 878 kommenttia, mikä on noin kymmenkertainen määrä seuraavaksi suurimpaan nostoon verrattuna. Latent Spacen AINews otsikoi julkaisun luvuilla 155 biljoonaa tokenia vuorokaudessa klusteria kohden, 22 biljoonaa dollaria päivätuloa noin 80 %:n katteella ja 81 päivän takaisinmaksuaika. Lähteestä ei selviä, ovatko luvut DeepSeekin omia vai uutiskirjeen johtamia. Riippumatonta todennusta niille ei ole. DeepSeek kuvasi päättelyjärjestelmäänsä ja sen teoreettista katetta jo keväällä 2025 ja avasi tuolloin yksittäisiä komponentteja. Nyt saatavilla on kokonainen kehys, jota ei ole vielä julkaistuissa vertailuissa ajettu vLLM:ää tai SGLangia vastaan DeepSeekin malleilla.

Arkisempi infratyö keskittyi saumoihin, joissa yhden toimittajan agenttityökalu puhuu toisen ajoalustan kanssa. Tyypillinen asetelma on agenttityökalu, joka käyttää Anthropicin Messages-muotoa ja on ohjattu välityspalvelimen kautta paikalliseen llama.cpp-palvelimeen. Tiistain build b11756 paljasti, että palvelimen /v1/messages-

muunnos kopioi vain rajatun joukon parametreja ja jätti id_slot-kentän pois. Välityspalvelin, joka kiinnitti istunnon KV-välimuistia pitävään slottiin, putosi siksi hiljaa automaattiseen valintaan. Hybridi- ja rekurrenttimalleilla jokainen kierros laski koko kehotteen uudelleen. Keskiviikon b11795 muuttaa päätepisteen välittämään pyynnön läpi ja hyväksyy llama.cpp:n omat näytteistysparametrit extra_params-kentässä, jolloin Messages-asiakas voi säätää paikallista näytteistystä ilman välikerrosta. Ollama 0.40.7 korjasi OpenAI-yhteensopivan Responses-rajapinnan. Keskustelun jatkaminen ei enää kaadu, kun malli vaihtuu ajattelevan ja ajattelemattoman välillä tai historiassa on toisen palveluntarjoajan kompaktointielementtejä. Striimauksessa korjattiin tilanne, jossa työkalukutsut saattoivat tulla ennen avoimia päättelyelementtejä. Esimerkkinä tällaisesta asiakkaasta mainitaan verkkohakua käyttävä Codex. n8n 2.45.0 alkoi toistaa AI Agent -solmussa Anthropicin ajattelukierrokset työkaluargumentteineen seuraavissa pyynnöissä, kuten rajapinta edellyttää. Lauantaina sauma korjattiin asiakaspäästä. Claude Code 2.1.310 yrittää uudelleen ilman effort-2025-11-24-betaa tai kesken keskustelun lähetettyjä system-viestejä, jos yhdyskäytävä tai yhteensopiva palvelin hylkää ne. Yhteistä

tapauksille on, että pyyntö meni läpi ja vain jokin sen osa katosi matkalla.

Spekulatiivinen dekodaus siirtyi viikon aikana erikseen valittavasta lisästä osaksi mallijulkaisua. Nemotron 3.5 Bolt E4B toimitettiin oman DFlash-luonnosmallinsa kanssa. Ollama 0.40.10 avasi MLX-taustassa Gemma 4:n piilotilat ja raa'at logitit luonnosmallien käyttöön. Projektissa llama.cpp torstain b11822 korjasi DFlash-luonnosmallien kaatumisen qwen4exp-kohdemallilla, joka julkaisi hyper-connection-residuaalin eri muodossa kuin luonnosmalli odottaa. Perjantain b11834 laajensi todennäköisyyspohjaisen näytteistuksen DFlash2-luonnoksiin ja muutoksen yhteiskirjoittajana oli NVIDIA:n kehittäjä. Lauantain b11856 lisäsi testin, joka varmistaa hylkäysotannon tuottavan saman jakauman kuin kohdemalli yksinään. Intelin XMX-yksiköille tuli monisarakkeinen mul_mat_vec_q (b11821), joka kohdistuu juuri luonnosten verifiointiin kahdesta kahdeksaan tokenia kerrallaan.

Toinen toistuva kuvio oli hiljaa väärin tulosten muuttaminen näkyviksi virheiksi. Metal-taustan fuusioitu moe_reduce-ydin (b11735) saattoi kierrättää syötteen muistin ulostuloksi, mistä seurasi harvinaisia epädeterministisiä vääriä tuloksia MoE-malleilla. SYCL-taustan CONV_2D (b11815) luki syötteen aina yhtenäisenä tensorina, vaikka hyväksyi minkä tahansa asettelun. Testiajon virheet putosivat korjauksen jälkeen 16:sta nollaan. Lauantain b11858 kieltää -sm tensor -jakotilan arkkitehtuureilta, joiden MoE-amax-reduktio kattaa koko matmul-rivin. Aiemmin kukin laite sai vain oman viipaleensa osittaisen maksimin ja tulos oli hiljaa väärä. Keskiviikon b11800 yhdisti harvan huomion muunnelmien top-k-maskin yhdeksi fuusioiduksi ggml_top_k_mask-operaatioksi, jonka qwen4exp, glm5-next, glm5.3-next, DSA-mallit ja DeepSeek V4:n CSA nyt jakavat. Perjantain b11843 korjasi qwen4exp:n valinnan, jossa yhdistetty maski palautti usean streamin ajossa väärän muodon ja kaatui jo graafin varausvaiheessa. Sunnuntain b11883 on näistä käyttäjälle painavin. Kvantisointityökalu laski neliulotteisen tensorin rivien määrän ilman neljättä ulottuvuutta, jolloin vain osa viipaleista kvantisoiitiin ja loput rivit pudotettiin. Samassa korjauksessa MXFP4_MOE-polku lakkasi tulkitsemasta kolmi- tai useampiulotteisia tavallisia tensoreita asiantuntijatensoreiksi. Tällaisia ovat esimerkiksi BailingMoE3:n MLA-projektiot ja neliulotteiset konvoluutioytimet. Vanhemmilla buildeilla tehdyt kvantisoinnit malleista, joissa tällaisia tensoreita on, kannattaa tuottaa uudelleen.

Rautapuolella viikko kallistui pois yhden kiihdytinvalmistajan oletuksesta. Hugging Facen blogissa ilmestyi kahdessa päivässä kolme kirjoitusta: HF-mallien

ajaminen AWS:n Trainiumilla TorchNeuronin kautta, Googlen TPU:iden natiivi tuki PyTorchissa TorchTPU:n myötä ja Applen MPS-tausta PyTorch 2.15:ssä. Perjantaina blogi julkaisi hakusivun, josta näkee, toimiiko malli Trainiumilla torch-neuronx:n kautta. Projektin llama.cpp taustatyö kulki samaan suuntaan. Qualcommin Hexagon sai työkaluketjun v0.8, säikeistetyt uudelleenpakkauksen ja Qualcommin insinöörin kanssa tehdyn SGN-operaation. AMD:n HIP-käännökseen tuli korjaus amdgcnspirv-kohteelle, jolla ennen korjausta 66 tiedostoa 145:stä ei kääntynyt. Vulkan sai src1-tarkkuuden tuen ylivuotojen käsittelyineen. NVIDIA puolestaan esitteli Python-pohjaisella CUDA Tilella kirjoitettuja Mamba-2-ymityä sekä kahdeksan DGX Spark -työaseman klusterin, jossa yhtenäistä muistia on noin teratavu. TensorRT 11.1:n Global Performance Tuner virittää käännöksen automaattisesti tarkkuuden huomioiden sitomatta sitä yhteen build-reittiin. Sunnuntaina llama.cpp:n CUDA-taustaan tuli kaksi fuusiota. Tiheän FFN-kerroksen gate- ja up-matmulit ajetaan kehoitteen käsittelyssä yhtenä MMQ-käynnistyksenä GLU:n kanssa (b11879) ja Qwenin MRoPE yhdistetään set-rows-operaatioon (b11878).

Paikallinen puhe nousi viikon aikana kahden yleisen ajoalustan ominaisuudeksi. Mistralin Voxtral 3 aloitti tiistaina puheagenttien SDK:na, joka kattaa reaaliaikaisen puheen yli 50 kielellä, transkription, tunnesävyä tuottavan synteessin ja työkalujen käytön. Sitä voi ajaa EU:ssa isännöitynä palveluna tai omana käyttöönottona. Tarkkuusväitteet ovat valmistajan omia. IBM:n Granite Speech 5.0 lupaa otsikossaan aiempaa tarkempaa ja nopeampaa litterointia, käännöstä ja attribuointia. Ollama 0.40.9 toi Apple Siliconille puhesynteesin Gradiumin Phonon-mallilla. Komento `ollama run phonon "Hello from Ollama."` toistaa tekstin tai chat-mallin vastauksen ääneen sitä mukaa kuin ääntä syntyy. Sovellukset striimaavat ääntä OpenAI-yhteensopivan `/v1/audio/speech`-päätepisteen kautta, jolloin olemassa olevan TTS-asiakkaan siirto paikalliseen on lähinnä osoitteen ja äänen nimen vaihto. Julkaisuehdokas rc3 siirtyi Phononin GA-tarkistuspisteisiin ja hylkää vanhat esijulkaisuversiot, jotka on tuotava uudelleen. Lauantain 0.40.11 lisäsi Phonon-äänikatalogin macOS:lle. Samana päivänä llama.cpp:n b11850 toi Soprano-80M- ja Soprano-1.1-80M-äänigeneraattorit, joissa Qwen3-rungon perässä on pelkästään generoivana mmproj-tiedostona ajettava Vocos-dekooderi.

Perjantaina kaksi julkaisua osui saman tuoteluokan kohdalle muutaman tunnin välein. OpenAI:n Decisions API lupaa nopeita, kohdennettuja ja edullisia päätöksiä mihin tahansa

sovellukseen kertomatta mallia, hinnoittelua tai vastausformaattia. Ollama toi vLLM Semantic Routerin d3-päätösmallit kahden, neljän, yhdeksän ja 27 miljardin parametrin kokoina. Ne ottavat syötteeksi tekstiä, rakenteista tilaa tai kuvan ja luokittelevat, vastaavat kyllä tai ei tai pisteyttävät vaihtoehdot todennäköisyyksineen -ask-lipulla annetun kysymyksen mukaan. Perplexityn Qwen3.8-

pohjainen 27 miljardin parametrin pplx-decider tuli Ollamaan jo keskiviikkona. Reitittimissä, suojakaiteissa ja automaatiiovirtojen haaroissa vapaamuotoisen vastauksen jäsentämisen voi näin korvata todennäköisyyden kynnysarvolla. Kalibroinnista kumpikaan julkaisu ei kerro. Kynnysarvot on siksi asetettava omalla validointiaineistolla.

08 Benchmarkit ja vertailut

Riippumattomien indeksien liikkeet tulivat viikolla Artificial Analysisilta. Kolmen Flash-mallin lukemat 42–45 asettuvat kapealle vyöhykkeelle. Kimi K3.1 Pro sijoittui indeksin neljänneksi, mikä tekisi siitä painojen julkaisun jälkeen avoimien mallien kärjen. Indeksi on koostemittari, joka ei kerro mallin toiminnasta suomeksi tai pitkissä agenttiaoissa. Kevyiden mallien varsinainen myyntivaltti eli hinta-suoritusuhde on katsottava mallisivuilta. Lukuja kannattaa verrata vain saman indeksiversion sisällä, koska Artificial Analysis on uudistanut indeksin koostumusta useaan kertaan.

Viikon uudet vertailut siirsivät mittaamisen lopputuloksesta käyttäytymiseen. Vals.ai:n arviossa avustajat tavoittelivat hyvitystä, jota ei voi saada. Mahdottomassa tehtävässä onnistumista ei ole tarjolla. Esiin tulee siksi se, keksiikö agentti perusteluja, painostaako se vastapuolta vai luovuttaako se. Ai2:n FairStress mittaa reiluutta perusteluista eikä päätöksistä. Se paljastaa väestöryhmien välisiä eroja siinä, miten mallit puolustavat valintojaan, silloinkin kun päätösjakaumat näyttävät tasapuolisilta. Auditointi, joka katsoo vain päätöksiä, voi siis päästää läpi mallin, jonka perustelut kohtelevat ryhmiä eri tavoin. OpenAI:n alignment-blogi ehdotti monitoroitavuuden arviointiin g-mean²-mittarin rinnalle pareittain sovitettuja vertailuja ja valvontatoimien jälkeen jäävää riskiä. OpenAI esitteli lisäksi arvioinnit ohjelmointityölle kielillä, joista opetusdataa on vähän, kuten SQLite-työlle ja COBOL-migraatiolle.

IBM Researchin ALTK-Evolve nousi julkaisijan mukaan τ^2 -benchin asiakaspalvelutehtävien kärkeen ilman huippumallia, koska agentti hyödyntää opittua muistia. Kärkisija on julkaisuhetken tilanne ja julkaisijan oma

ilmoitus. Parannus tulee kuitenkin mallin ulkopuolelta, samasta suunnasta kuin headroom-repon lupaus 60–95 %:n tokensäästöstä JSON-aineistolle.

Valmistajien omia lukuja kertyi viikolla poikkeuksellisen paljon. Laminarin RL:llä koulutettu flow-1 yltää yhtiön mukaan agenttijälkien virheiden tunnistuksessa GPT-6-solin tasolle 23 kertaa halvemmalla. Halpa jälkianalyysi mahdollistaisi Laminarin mukaan jokaisen tuotantojäljen läpikäynnin otannan sijaan. Terminal X:n GPT-6.1 Solilla toimiva arviointiagentti osuu OpenAI:n asiakastarinassa 88,6 %:sti yksiin kokeneiden sijoittajien arvioiden kanssa. Dharma-AI väittää sanaston erikoistamisen leikkaavan pienten mallien päättelykustannuksia 77 %. NVIDIA:n mukaan Kumo Relational päihittää ohjatun koneoppimisen relaatiotiedatalla ilman koulutusta. Yhtäkään näistä luvuista ei ole toistanut riippumaton taho.

Hacker Newsin vilkkain vertailukeskustelu (604 pistettä, 574 kommenttia) syntyi modelcompare.ai:n väitteestä, että jokaisella mallilla on sama kiistanalainen näkemys ihmisistä. Kysymysten muotoilu ja testattu mallijoukko eivät käy ilmi lähteestä. Havainto osuu silti käytäntöön, jossa eri laboratorioden malleja käytetään toistensa tarkistajina tai arvioijina. Samansuuntaiset vinoumat tekevät tällaisesta äänestyksestä vähemmän riippumattoman kuin mallien määrä antaa ymmärtää. Paikkatiedossa Ai2:n PLANTEUR laajensi Ranskan peltolohkoihin perustuvan PASTIS-vertailun (Sainte Fare Garnot ja Landrieu, 2021) kuuteen trooppiseen alueeseen. Se mittaa, yleistyvätkö mallit maatalousjärjestelmiin, jotka poikkeavat eurooppalaisista.

09 Työskentelytavat ja agentit

Kolmen suuren koodausagentin viikon muutokset kulkivat samaan suuntaan: autonomia lakkaa olemasta koko istunnon kattava tila ja muuttuu tehtävä- ja kutsukohtaiseksi säädöksi. Claude Code 2.1.304 lisäsi modien tool.check-vastaukseen

kentän decider: "person", jolloin pyynnön voi hyväksyä vain lupadialogi tai isäntäsovelluksen callback. Ympäristössä, jossa ketään ei voi kysyä, pyyntö hylätään modin antamalla perusteella. Tapahtuman uusi

`e.isReadOnly` kertoo, vain lukeeko kutsu. Yhdessä ne mahdollistavat politiikan, jossa lukevat kutsut kulkevat automaattisesti läpi ja valitut kirjoittavat kutsut vaativat aina ihmisen riippumatta siitä, mitä automaattinen tila päättäisi. Opastettu `/auto-mode-setup` johtaa auto moden rajat projektista ja viimeaikaisista istunnoista eikä vain käsin ylläpidetyistä sallittujen komentojen listoista. Se toimii nyt myös Bedrockissa, Vertexissä, Foundryssa ja gatewayn kautta.

Codex CLI 0.164.0 ilmestyi perjantaina vakaana versiona yhdeksän julkaisutiedoitusta tulleen alfan jälkeen. Komennot `/model` ja `/permissions` ottavat nyt argumentit suoraan, jolloin malli, päättelyn määrä, nopeus ja käyttöoikeusprofiili vaihtuvat yhdellä komennolla kesken istunnon. Tutkiva vaihe voi näin kulkea väljemmillä oikeuksilla ja tiedostoja muuttava vaihe tiukemmilla. Uusi `/presets` tallentaa uudelleenkäytettävät kehotteet ja slash-komennot kymmeneen numeroituun paikkaan. Gemini CLI:n yöversiot veivät IDE-kumppanin `exec-stdion` kautta konttihiikkalaatikkoon, jolloin eristetty agentti näkee edelleen editorin tilan. Hyväksytyt liput muistetaan, mikä vähentää toistuvia kyselyjä. Ei-interaktiivisessa tilassa päättämissignaali käsitellään ja työkalujen suoritus estetään peruutuksen jälkeen, mikä koskee CI-putkia, joissa agenttia ajetaan ilman ihmistä. Työkalun oma README kertoo samalla, että ilmaistason ja Google One -käyttäjille Gemini CLI korvattiin Antigravity CLI:llä 18.6.2026. README:n mukaan sitä ei ole optimoitu Gemini 4 Argonille, jota Google ohjaa käyttämään Antigravityn kautta.

Orkestroinnissa viikon kiinnostavin havainto oli, että malli kirjoittaa koordinaatiologiikan itse. Datasetten luoja testasi Claude Code for webin Ultracode-tilaa, jossa Claude Code generoi JavaScriptiä, joka määrittelee ja koordinoi aliagenttien verkon monimutkaisen ongelman ratkaisemiseksi. Tuloksena syntyi `sparqlite`, Rustilla kirjoitettu SQLite-laajennus, joka toteuttaa SPARQL 1.1:n. Orkestrointia ei ole rakennettu työkaluun kiinteästi, koska malli kirjoittaa sen tehtäväkohtaisesti koodiksi. Latent Spacen mukaan Anthropic esitteli AI Engineer New Yorkissa Dynamic Workflows -toimintatavan, joka lainaa Unixin ja Linuxin opeista keinoja koordinoita agentteja suurissa projekteissa niin, ettei kehittäjä huku. Google Research julkaisi Harness2:n, joka rakentaa agenttivaljaat rekursiivisesti. Show HN -nosto `lights-on-software-factory` tarjoaa pohjan Claude Codella ajettavalle ohjelmistotehtäälle. Tekijän oman mittauksen mukaan orkestroija vähensi tokenikustannusta noin 80 % paljaaseen työnkulkuun verrattuna. Käsityön puoli näkyi samaan aikaan Hacker Newsin `tmux-oppaassa` (169 pistettä), joka käsittelee

rinnakkaisten agentti-istuntojen hallintaa `send-keys`-komennolla ja menetelmän kattoa. Modalin toimitusjohtaja kuvasi Latent Spacen haastattelussa agenttien hiikkalaatikoiden kysynnän kasvua. Eristetty ajoympäristö, jossa agentti suorittaa koodia, on hänen mukaansa muuttumassa omaksi laskentatuotteekseen.

Pienessä mittakaavassa sama muutos näkyy kehotteissa. Battledrones-pelin 14. versiossa kehote määritteli vain tavoitteen ja reunaehdot: neljän pelaajan kartta, nelinkertainen kiertosymmetria, enintään neljä tekoälypelaajaa, toimivuus WebSocket-moninpelissä ja tyhjien paikkojen täyttö tekoälyllä. Opus 5.5 lisäsi Compass Rose -kartan ja muokkasi verkkokoodin tukemaan neljää pelaajaa jakolinkkien ja QR-koodien kautta yhdellä kertaa. Kehote kuvasi rajoitteet ja varakäytöksen eikä toteutusta.

Toinen puoli kysymystä on se, mitä ihmiselle jää. Latent Space kuvasi Claude Modsin Attention Interfacen varhaiseksi muodoksi, jossa agentin valjaat käännetään katsomaan koneen sijaan ihmiseen. Painopiste siirtyy siihen, miten agentti pyytää ja kuluttaa ihmisen huomiota. Yhteisöstä tuli jo ensimmäisiä modeja, kuten tilauksen käyttöikkunaa suojaava `agent-usage-guard`. Sphinx (249 pistettä) avaa jokaisen tuotantoon pushauksen yhteydessä kyselyn siitä, ymmärtääkö pushaaja koodin toiminnan. Se mittaa ihmisen ymmärrystä muutoksesta eikä muutosta itseään. Riskinä on, että kysely klikataan läpi kuin evästebanneri. Ask HN -ketju nosti esiin Metan insinöörin kirjoituksen, jonka mukaan koodin kirjoittaminen tuntuu nyt mustalta laatikolta: järjestelmän yleisrakenne on hallussa ja yksityiskohdat jäävät hämärään. Ohjelmoijan essee `My obituary for programming` nimesi menetyksen tarkemmin. Ohjelmoinnissa maksettiin visaisten pienten logiikkapulmien ratkomisesta omalla tavalla ja omalla aikataululla. Juuri pulma on se, minkä koodausagentit vievät.

Avoimen perusinfran versiohistoria dokumentoi muutoksen arjen. Projektin `llama.cpp commit-viesteissä` Assisted-by-merkinnät nimesivät viikon aikana muun muassa OpenCoden, Codexin ja paikallisesti ajettuja avoimia malleja. Lauantain MoE-amax-korjauksessa avustajana oli MacBookilla (M5 Max, 128 Gt) `llama.cpp`:llä ajettu Qwen3.8-Flash-Next-GGUF. Sunnuntain muistivälimuistikorjauksessa avustajana oli Mac Studiolla (M2 Ultra, 192 Gt) ajettu DeepSeek-V4-Flash-Vision-Exp-GGUF. Inferenssimoottorin korjauksia syntyy siis osin malleilla, joita sama moottori ajaa. MCP-referenssipalvelinten commiteissa Co-Authored-By-rivi nimeää Claude Opus 5.5:n. Saman repon `pull requestissa` Copilotin katselmointikierros löysi ennen yhdistämistä tarpeen sarjallistaa git-työkalukutsut. Kirjoittava ja

katselmoiva työkalu voivat siis olla eri toimittajilta. Merkintä ei silti korvaa katselmointia. Torstain rerank-muutoksessa avustajana oli Opencode/Qwen3.8 27B ja ihmisen

katselmoinnissa löytyivät silti kuvien hiljainen pudotus ja väärät virhekoodit.

10 Liitännäiset, MCP ja repot

Anthropic lahjoitti torstaina Agent Skills -määrittelyn Linux Foundationin alaiselle Agentic AI Foundationille (AAIF). Säätiön Anthropic perusti viime vuonna Blockin ja OpenAI:n kanssa ja sille se oli jo aiemmin luovuttanut MCP:n. Skill on hakemisto, jonka SKILL.md-tiedoston YAML-otsakkeesta kontekstiin ladataan vain nimi ja kuvaus. Varsinainen ohje, skriptit ja resurssit luetaan vasta, kun agentti päättää käyttää skilliä. MCP kuvaa saman säätiön alla, mihin työkaluihin agentti pääsee käsiksi. Skillit kuvaavat, miten agentti tekee tietyn tehtävän. Formaattia on kutsuttu avoimeksi standardiksi jo aiemmin. Hallinnoinnin siirto tekee siitä myös muodollisesti toimittajaneutraalin.

Leviäminen näkyi viikon muissa julkaisuissa. Open WebUI 0.12.4 toi sunnuntaina asetuksen, jolla chatissa valitut tai malliin liitetyt skillit vain listataan mallille ladattaviksi tarvittaessa. Niiden koko ohjetta ei enää lähetetä jokaisen viestin mukana. Asetus (ENABLE_SKILL_LAZY_LOADING) on oletuksena pois päältä ja vaatii sisäänrakennetun Skills-työkalun. AWS:n virallinen agent-toolkit-for-aws kokoa MCP-palvelimia, skillejä ja liitännäisiä, jotka toimivat Claude Codessa, Codexissa, Cursorissa, Devinissä, Kirossa ja fx:ssä. Käyttöönotto alkaa komennolla `aws configure agent-toolkit`. Together AI:n Hallmark on suunnitteluskil Claude Codelle, Codexille, Cursorille ja OpenCodelle. Se valitsee 48 teeman katalogista kolme ja rakentaa sivun kävijän tehtävän ympärille. Lopuksi se tarkistaa tuloksen oikeassa selaimessa koneen tekemän suunnittelun tunnusmerkkejä vastaan, kuten jokaisen otsikon yläpuolelle lisättyä pientä yläotsikkoa. Kokoelma wshobson/agents tarjoaa 94 liitännäistä, jotka toimivat seitsemässä eri agenttityökalussa. Rajat ohenevat myös asetusten tasolla. Gemini CLI tuo Claude Coden koukkuasetuksia omikseen ja lauantain yöversio korjasi siirron aikakatkaisun yksikön. Ennen korjausta sekunteina annettu arvo tulkittiin millisekunneiksi. Siirretyn koukun aikakatkaisu kutistui siis tuhannesosaan tarkoitettusta. Kokoelma agentic-awesome-skills lupaa yli 2 724 SKILL.md-ohjekirjaa, mutta määrä ei kerro laadusta. Jokainen asennettu SKILL.md on agentille annettu ohje, minkä vuoksi tiedostot kannattaa lukea ennen asennusta.

MCP:n hallinto eteni samaan suuntaan. Määrittelyyn yhdistetty SEP #3440 nimesi Triggers and Events -

työryhmän Tasks and Events -työryhmäksi ja siirsi sille Agents-työryhmän Tasks-laajuuden, vetäjät, resurssit ja työkohteet. Agents-työryhmästä tuli Interest Group. Pitkäkestoisten tehtävien määrittely etenee siis määrittelymuutoksia tuottavassa ryhmässä ja agenttiteema keskusteleavassa. Perjantaina yhden ylläpitäjän GitHub-tunnus vaihdettiin ja Anthropic-kytkös poistettiin ylläpitäjä-, työryhmä- ja roadmap-viittauksista. Ylläpitäjä jatkaa rooleissaan.

Referenssipalvelinten v2-työ jatkui koko viikon. Maanantaina Python-palvelimet fetch, git ja time siirtyivät SDK:n matalan tason Server-luokasta korkean tason MCPServeriin. Siinä työkalut ovat tyypitettyjä funktioita ja syöteskeema johdetaan tyyppivihjeistä. Tuntematon työkalunimi palauttaa nyt JSON-RPC-virheen -32602. Fetchin muut kuin URL-parametrit ovat tiukkoja, joten merkkijonoa ei enää muunneta kokonaisluvuksi tai totuusarvoksi. Git-palvelimeen lisättiin lukko, koska MCPServer ajaa synkroniset työkalut työsäikeessä ja päällekkäiset kutsut olisivat voineet lomittaa tarkistuksen ja muutoksen jaetussa indeksissä. Python-paketit vaativat `mcp>=2.3` ja TypeScript-pakettien Node-alaraja nousi 24:ään. Lauantai-iltana repossa tehtiin jo v2.1.0-julkaisun auditointikorjauksia. TypeScript SDK nousi versioon 2.4.0 ja Node-adapteri versioon 2.1.2 ja riippuvuuksia päivitettiin useiden tietoturvatiedotteiden vuoksi. Palvelimia ajavan tarkistuslista on lyhyt: Node 24 -ajoympäristö, Python-puolella `mcp>=2.3`, asiakkaan -32602-käsittely ja tiukemmat parametrityyppit. Asiakaspuolella Codex CLI 0.164.0 toi yritysten MCP-yhteyksille tunnistautumisen Codex-tilin oikeuksilla tai erillisellä kirjautumisella. Tuki kattaa myös OAuth-asiakkaat, jotka vaativat client secretin. SSO:n takana olevat sisäiset MCP-palvelimet tulevat näin käyttöön ilman omaa välityskerrosta. Gemini CLI:n yöversio lisäsi MCP:n OAuth-kulkuun valinnaisen tuen yksityisen verkon päätepisteille.

Trending-listojen repoista kaksi tarttuu samaan ongelmaan eli siihen, mitä agentin kontekstiin päätyy. headroomlabs-ai/headroom tiivistää työkalujen tulosteet, lokit, RAG-palat ja keskusteluhistorian paikallisesti ennen kuin ne menevät mallille. Se lupaa koodausagenteille 20 % ja JSON-aineistolle 60–95 % vähemmän tokeneita samoilla

vastauksilla. Väite samoista vastauksista on tekijän oma. Tiivistys, joka pudottaa sisältöä, voi viedä mukanaan juuri sen lokirivin, jota agentti olisi tarvinnut. Kokeilun arvoinen repo on JSON-raskaissa ketjuissa, kun tuloksia verrataan omilla tehtävillä tiivistyksen kanssa ja ilman. Graphify-Labs/graphify muuttaa koodikannan dokumentteineen, SQL-skeemoineen ja konfiguraatioineen kyseltäväksi tietämysgraafiksi deterministisellä AST-jäsennyksellä ilman vektorivarastoa. Se on uskottava vaihtoehto upotushaulle keskikokoisten repojen navigoinnissa.

Sunnuntain listalla nousi makecindy/cindy, avoimen lähdekoodin agentti, joka kokoaa useita valjaita, malleja ja työkaluja yhteen. Ensimmäiset tuetut valjaat ovat Claude

Code ja Codex ja oma valjas on tekijän mukaan työn alla. Agentti toimii paikallisesti käyttäjän oikeilla tiedostoilla ja kirjautuneilla sovelluksilla. Juuri tämä laaja käyttöoikeus tekee siitä kokeilukohteen erilliselle koneelle tai käyttäjätillille eikä työkoneelle. Yhteisön nostoista suurimman huomion sai WorldTUI (597 pistettä). Siinä 360 miljoonan parametrin malli on koulutettu ennustamaan, miltä päätteen tuloste näyttäisi, ajamatta yhtään komentoa. Kapean alan maailmamalli on niin pieni, että kokeilu on kenen tahansa toistettavissa. Listojen loppupää oli kohinaa. Siihen kuuluivat Node.js:n kaltaiset vakioikävijät, kiinalaisen astrologian laskentamoottori 20,4 miljardin merkin tulkinta-aineistoinen ja e-kirjasovellusten kirjojen latauslinkkejä kokoava repo.

11 Tietoturva ja haavoittuvuudet

PHP2Shell on kahden etäkoodinsuoritusaukon (RCE) kokonaisuus PHP:n ytimessä. Kyberturvallisuuskeskuksen haavoittuvuustiedotteen 2026-37 mukaan aukot koskevat kaikkia julkaistuja PHP-versioita ja mahdollistavat mielivaltaisten kommentojen etäsuorituksen sekä palvelimen täyden haltuunoton. Löytöön käytettiin tekoälyä. Latent Spacen AINewsin mukaan OpenAI:n Patch the Planet oli toinen kahdesta löytäjästä ja käytti GPT-5.6 Sol -mallia. Wiz julkaisi 21.10. oman kirjoituksensa otsikolla PHP2Shell: AI uncovers critical RCE in PHP ja avasi tapaukselle sivuston php2shell.com. Lähteet eivät suoraan kerro, onko Wiz AINewsin nimeämätön toinen löytäjä. Koordinoitu prosessi kesti kaksi kuukautta ja petti viimeisenä vuorokautena.

PHP-säätiön selvityksen mukaan molemmat CVE-tunnisteet alkoivat kiertää foorumeilla ja sosiaalisessa mediassa tiistaina 20.10. noin kello 15 UTC. Niiden mukana kiersivät väärä julkaisuaika ja väärä CVSS-pistemäärä 10.0. Säätiön mukaan todellinen pistemäärä on alempi. Tivin mukaan julkistus aikaistui, koska yksi taho julkaisi päivityksensä etuajassa. Toimivia hyödyntämiskoodeja on kiertänyt tiistain ja keskiviikon välisestä yöstä asti. Korjaukset julkaistiin keskiviikkona 21.10. koko 8.x-haaralle. Cloudflaren havaintojen mukaan hyödyntämisyritykset nousivat tunneissa miljooniin. Suomessa Valtori ja moni muu organisaatio sulki verkkopalvelujaan. AINewsin otsikon 70 prosenttia verkosta vastaa PHP:n yleisesti mitattua osuutta niistä sivustoista, joiden palvelinpuolen kieli tunnetaan. Luku ei kerro, kuinka moni asennus oli käytännössä hyödyttävissä. Anthropicin tietoturvaohjaaja varoitti Tivin haastattelussa, että pahempaa on vielä tulossa. Flask-verkkokehityksen tekijä tiivistä tunnelman lauseeseen: "As fun

as AI is, we built so much bad software over the last 30 years, and now the universe is punishing us for our sins."

Altis on jokainen PHP 8.x -asennus, jota ei ole päivitetty 21.10. julkaistuihin korjausversioihin. Elinkaarensa päähän tullee 7.x-haaroille virallisia korjauksia ei julkaista. Komento `php -v` ei yksin riitä tunnistamiseen kahdesta syystä. Se kertoo komentorivin PHP-version, joka voi poiketa PHP-FPM:n tai `mod_php`:n käyttämästä. Lisäksi Debianin, Ubuntun ja Red Hatin kaltaiset jakelut takaportaavat korjauksia muuttamatta upstream-versionumeroa. Tila tarkistetaan siksi jakelun tietoturvatiedotteesta ja paketin muutoslokista. Päivityksen jälkeen PHP-FPM-poolit ja `mod_php`:tä käyttävät Apache-prosessit on käynnistettävä uudelleen, koska vanhat prosessit jatkavat muuten haavoittuvalla binäärillä. Kontti-imagit rakennetaan päivitetyn pohjaimagen päälle. Välimuistissa oleva vanha tagi tai lukittu digest pitää haavoittuvan version tuotannossa, vaikka isäntäkone olisi päivitetty. Ylläpitopaneeleissa, joissa on rinnakkain useita PHP-versioita, jokainen versio päivitetään erikseen. CVE-tunnisteita ja korjausversioiden numeroita aineiston lähteissä ei ole. Ne löytyvät `php.net`in muutoslokista ja jakelujen tiedotteista. Lähteet eivät kuvaa hyökkäysten tunnusmerkkejä. Murron mahdollisuus tiistai-illan ja päivityksen väliseltä ajalta selvitetään siksi lokeista ja tiedostojen eheyden tarkistuksella.

Samana keskiviikkona varmenneketju horjui toisaalla. Varmentaja eMudhran kahdeksan välivarmenteen yksityiset avaimet ovat vaarantuneet. Varmenteet ketjuuntuvat juuriin, joihin Mozillan, Microsoftin, Applen ja Chromen juuriohjelmat luottavat. Ohjelmistot, jotka paketoivat oman

luotettujen varmenteiden listansa, on päivitettävä. Linux-jakeluissa tämä tarkoittaa ca-certificates-pakettia. Python-maailmassa päivitetty listat ovat versioissa certifi 2026.10.19, pip 26.2.2 ja uv 0.13.4. Asia koskee suoraan tekoälytyökaluja, koska httpx- ja requests-kirjastoja käyttävät Python-SDK:t ja agenttityökalut nojaavat oletuksena certifiin. Kiinnitetyt riippuvuudet eivät päivitty itsestään, minkä MCP:n referenssipalvelimet havainnollistivat. Fetch-palvelimen lukitustiedostoon certifi 2026.10.19 päivitettiin vasta lauantaina tiedotteen GHSA-6q48-6c7w-9gw2 perusteella. Samassa auditoinnissa git-palvelimen gitpython nostettiin versioon 3.2.1 haavoittuvuuden CVE-2026-114145 vuoksi. Asennetun certifin version näkee komennolla `pip show certifi` tai lukitustiedostosta.

Agenttityökalujen omat suojaukset saivat viikolla useita korjauksia. Claude Code 2.1.309 korjaa muutoslokin otteen mukaan ainakin viisi hiekkalaatikon ja käyttöoikeuksien aukkoa. Hiekkalaatikoidut Bash-komennot ajoivat macOS:llä ja Linuxilla shellin käynnistystiedostot hiekkalaatikon ulkopuolella ennen jokaista komentoa. Applen alustalla denyWrite-glob-säännöt eivät osuneet polun todelliseen sijaintiin, kun kansioon päästiin symbolisen linkin kautta. Painavin kohta koskee MCP:tä. Hiekkalaatikoidut komennot ja automaattisesti hyväksytyt muokkaukset pystyivät muuttamaan `--mcp-config`-valitsimella annettua konfiguraatiota, joka määrää, mitä palvelinprosesseja istunto käynnistää. Lisäksi myöhempi käyttöoikeuskysely saattoi yhdistää uudelleen käytöstä poistetun MCP-palvelimen. Bash-oikeustarkistus luotti muuttujaan, jonka asetti komento, jonka uudelleenohjauksen avaaminen voi epäonnistua. Vakavuusluokituksia muutoslokeissa ei ole. Asennettu versio tarkistetaan komennolla `claude --version`. Lauantain 2.1.310 korjasi taustasiivouksen, joka poisti tuoreita tiedostoja, kuten liitettyjä kuvia, kun `cleanupPeriodDays` oli asetettu hyvin suureksi. Gemini CLI 0.65.0 pakottaa työtilan asetukset vain luku-tilaan epäluotetuissa kansioissa, mikä kaventaa riskiä siitä, että kloonattu repo muuttaa agentin asetuksia. Sama versio validoi OAuth-takaisinkutsun iss-parametrin RFC 9207:n metatietoja vasten, mikä torjuu mix-up-hyökkäyksiä. Gemini CLI:n lauantain yöversio lopettaa jatkovirran käsittelyn ja automaattisen jatkoehotteen, kun käyttäjä peruu pyynnön. Ennen korjausta peruttu pyyntö saattoi jatkaa käsittelyä taustalla. Codex CLI 0.164.0 mainitsee tunnistetietojen peittämisen, mutta käytettävissä oleva ote katkeaa ennen tarkempaa kuvausta. Lauantaina Hacker Newsiin nousi X-palvelun viestiin perustuva väite, jonka mukaan Claude Code lisää käyttäjän sähköpostiosoitteen lähteviin HTTP-pyyntöihin (21 pistettä,

neljä kommenttia). Aineistossa ei ole vahvistusta väitteelle eikä kuvausta siitä, mihin pyyntöihin osoite päätyisi.

Viikon opettavaisin sarja tuli Anthropicin virallisen liitännäiskokoelman security-guidance-liitännäisestä, joka ajaa LLM-katselmointeja Stop-, commit- ja push-vaiheissa. Maanantaina siihen yhdistettiin kolme korjausta. Kaikissa niissä kontrolli oli epäonnistunut avoimeksi ilman merkkiä käyttäjälle. Kirjautuneessa istunnossa, joka oli käynnistetty muuttujalla `CLAUDE_CODE_DISABLE_NONESSENTIAL_TRAFFIC_DISABLE_TELEMETRY` tai `DO_NOT_TRACK`, Claude Code ei välittänyt kirjautumista koukuille. Kaikki LLM-katselmoinnit ohitettiin siksi hiljaa. Toisessa viassa käynnistin päätyi macOS:n vakiotulkkiin, jonka versio 3.9 ei lataa agentti-SDK:ta, jos koneen ainoa uudempi tulkki oli python3.14. Kolmannessa NotebookEdit-solujen muokkaukset eivät koskaan tuottaneet kuviotarkistuksen varoitusta. Tiistain muutos teki näkyväksi rajauksen: sisäkkäiset repot, kuten submoduulit ja vendoroidut checkoutit, jäävät tarkastuksen ulkopuolelle. Torstain versiossa 2.0.16 korjattiin Windows-tapaus, jossa yksikin OpenSSL:lle kelpaamaton varmenne järjestelmän säilössä esti TLS-kontekstin rakentamisen ja kaatoi jokaisen katselmointikutsun. Ohitetut katselmoinnit raportoivat nyt tilan `no_creds` ja ohitetut varmenteet kentän `tls_store_skipped`. Yksityisyyttä korostavat asetukset vaikuttivat samaan aikaan toiseenkin ominaisuuteen. Claude Code 2.1.310 laajensi Monitor-työkalun ja liitännäisten monitorit ympäristöihin, joissa telemetria tai ei-välttämätön liikenne on kytketty pois. Hiljaa ohittuva kontrolli näyttää käyttäjälle samalta kuin kontrolli, joka ei löytänyt mitään.

Paikallisia palvelimia ajaville sunnuntain llama.cpp-buildit toivat kaksi resurssien ehtymistä koskevaa korjausta. Build b11882 rajoittaa kieliopin rekursiosyvyuden pinon ylivuodon estämiseksi. Build b11884 korjaa DRY-näytteistyksen sekvenssikatkaisijan, joka saattoi kuluttaa suorittimen ja muistin loppuun. Palvelimen pyynnöissä kielioppi ja DRY-katkaisijat ovat asiakkaan annettavissa. Verkkoon avoimen llama-serverin ylläpitäjälle päivitys on siksi kiireellinen. Buildin numero näkyy komennolla `llama-server --version`. Automaatiopuolella n8n 2.45.0 alkoi soveltaa projektirajausta suorituksiin, joita haetaan jäsentason API-avaimilla (#40887). Korjauksen muotoilusta voi päätellä, että jäsenen avain saattoi aiemmin nähdä muiden projektien suoritukset. Monen projektin instansseissa API-avainten käyttö kannattaa tarkistaa lokeista. LangGraph CLI 0.4.37 lopetti aliprosessin stdin-syötteen tulostamisen verbose-tilassa. Aiempien versioiden CI-lokeja säilyttävien kannattaa tarkistaa, päätyikö niihin salaisuuksia.

Laajemmin agenttien ajonaikaiseen valvontaan NVIDIA esitteli Sentryn ja OpenShell-referenssiarkkitehtuurin. Yhtiön mukaan agentti voi suorittaa tehtävänsä ja silti ylittää tietoturvarajan. OpenShell kuvaa, miten tunteja kestäviä agenttikuormia suojataan, kun agentit käyttävät tiedostoja, työkaluja ja internetiä ja käynnistävät aliagenteja. Anthropicin tutkimus Using Claude to reduce the attack

surface of code käsittelee mallin käyttöä hyökkäyspinnan pienentämiseen. ProjectDiscoveryn tutkimus How ablated models can get you pwned (244 pistettä) käsittelee malleja, joiden painoista kieltäytymissuunta on poistettu. Hyökkäysmekanismi ei näy aineiston otteesta. Sama mallityyppi ilmestyi viikolla DeepSeek V4.1 Flashista muutamassa päivässä perusmallin julkaisun jälkeen.

Lähteet

OpenAI models are coming to Google Cloud · OpenAI

Claude Platform release notes, October 22, 2026 · Anthropic API release notes

Claude Platform release notes, October 21, 2026 · Anthropic API release notes

Claude Platform release notes: October 20, 2026 · Anthropic API release notes

Claude Platform release notes, October 20, 2026 · Anthropic API release notes

Patch the Planet: global response coordination as agents uncover critical vulnerabilities · OpenAI

AINews: OpenAI's Patch the Planet program is 1 of 2 parties to discover PHP2Shell vulnerabilities in core PHP affecting 70% of the web, using GPT-5.6 SoI; coordinated response over 2 months · Latent Space

Democratic governance of frontier AI, part II: Empowering cyber defenders · OpenAI

Cloudflare links vulnerability discovery to defense with OpenAI · OpenAI

Securing our frontier model training environment · OpenAI

OpenAI acquires Orchestra Bio · OpenAI

Using thousands of AI agents to uncover new leads in biological data · Hacker News: agents

ChatGPT Ads expands to 26 new countries as over 100,000 advertisers have used the platform · OpenAI

How we built ChatGPT's novel ad stack for user trust and relevance · OpenAI

Meta Muse Wonder & Wonder Full: MSL's Claude 5.5-tier Frontier Model · Latent Space

Meta's Muse Wonder and Spark Open, Educational Use of ChatGPT, Reflection's Partnership with Nvidia · The Batch (deeplearning.ai)

Review of the "Misalignment in high-stakes settings" section in Anthropic's Risk Report (August 2026) · METR

Red-Teaming OpenAI's Internal Agent Monitoring Systems · METR

Investigating alignment failures in AISI's incident · UK AI Security Institute blog

Anthropic's proposed AI safety bill only kicks in if 50 deaths or \$1B in damages · Hacker News: agents

Introducing Kimi K3.1 Pro · Simon Willison

[AINews] Kimi 3.1 Pro 1048A-2.8T: Fable-5.1-Class Open Weights narrows Closed-AI lead to 1 month... and Pro Highspeed API with 2-3× speed · Latent Space

Kimi K3.1 Pro launches at 4th on the Intelligence Index and will lead open weights models · Artificial Analysis

Podcast: Kimi K3.1 Pro, the new cyber reality, and Muse musings · Interconnects

DeepSeek V4.1 Flash 1017 (Max): Intelligence Index 45 · Artificial Analysis

Ling 3.1 Flash 1019: Intelligence Index 44 · Artificial Analysis

MiniMax-M3.1-Flash: Intelligence Index 42 · Artificial Analysis

[AINews] not much happened today · Latent Space

Trending: dealignai/DeepSeek-V4.1-Flash-1017-UNCENSORED-FP8 · HF trending -mallit

Trending: ISTA-DASLab/DeepSeek-V4.1-Flash-1017-GSQ-RCO-GGUF · HF trending -mallit

Latest open artifacts (#25): DeepSeek V4.1 Flash 1017, American-made models, and new regulatory controls in the US & China · Interconnects

Ollama v0.40.8 (Recursive Flint) · Ollama-julkaisut

v0.40.8: mlx: add Flint model support · Ollama-julkaisut

v0.40.9-rc2: flint: fix replies with thinking disabled · Ollama-julkaisut

Trending: RecursiveSI/flint · HF trending -mallit

Flint-interactive: A language for creating interactive data visualizations · Microsoft Research

v0.40.10 · Ollama-julkaisut

v0.40.10-rc0: Update gemma renderers · Ollama-julkaisut

b11833: vocab: fix Nemotron-3.5-Embed tokenization · llama.cpp-julkaisut

Trending: nvidia/Qwen3.8-Flash-Next-FP4-Local · HF trending -mallit

Trending: tencent/Hy-Memory-30B-A3B · HF trending -mallit

Trending: Nanbeige/Nanbeige4.3-33B · HF trending -mallit

Hybrid Reasoning Agents: LLM agents that think in latent space, act through text · Hacker News: agents

OlmoEarth v1.3: State-of-the-art embeddings for the whole planet · Ai2 Blog

DeepInfer: High-performance inference framework for DeepSeek models · Hacker News: MCP

[AINews] DeepSeek DeepInfer: Open Source SOTA Inference running 155T toks/day/cluster making \$22m/day at ~80% margin for a 81 day payback · Latent Space

b11756 · llama.cpp-julkaisut

b11795 · llama.cpp-julkaisut

v0.40.7 · Ollama-julkaisut

n8n@2.45.0 · n8n releases

Claude Code 2.1.310 · Claude Code changelog

b11822 · llama.cpp-julkaisut

b11834: Extend probabilistic sampling for DFlash2 drafts · llama.cpp-julkaisut

b11856 · llama.cpp-julkaisut

b11821 · llama.cpp-julkaisut

b11735: metal: fix moe_reduce fusion race when dst aliases its inputs · llama.cpp-julkaisut

b11815 · llama.cpp-julkaisut

b11858 · llama.cpp-julkaisut

b11800: llama: create new fused op for QSA mask · llama.cpp-julkaisut

b11843: qwen4exp: build the QSA selection without the kq_mask · llama.cpp-julkaisut

b11883 · llama.cpp-julkaisut

TorchNeuron: running HF models on AWS Trainium is easier than ever · Hugging Face -blogi

Hugging Face ❤️ TorchTPU: TPUs, Natively · Hugging Face -blogi

MPS in PyTorch 2.15 · Hugging Face -blogi

Does Your Model Run on AWS Trainium? Look It Up · Hugging Face -blogi

b11751 · llama.cpp-julkaisut

b11790 · llama.cpp-julkaisut

b11849 · llama.cpp-julkaisut

b11853 · llama.cpp-julkaisut

b11854 · llama.cpp-julkaisut

Accelerating Mamba-2 with High-Performance Forward Pass Kernels in NVIDIA CUDA Tile · NVIDIA Technical Blog

Running Frontier-Class AI Models Across 8 NVIDIA DGX Spark Nodes · NVIDIA Technical Blog

Scaling NVIDIA TensorRT Performance Optimization Beyond a Single Build Route with the Global Performance Tuner, Part 1 · NVIDIA Technical Blog

b11878 · llama.cpp-julkaisut

b11879 · llama.cpp-julkaisut

Voxtral 3 · Mistral

Granite Speech 5.0: Transcribe, Translate, and Attribute more Accurately and Faster · Hugging Face -blogi

v0.40.9 · Ollama-julkaisut

v0.40.9-rc3: Support GA Gadium Phonon checkpoints · Ollama-julkaisut

v0.40.11 · Ollama-julkaisut

v0.40.11-rc0: voice: add voice catalog for Phonon (#19242) · Ollama-julkaisut

b11850 · llama.cpp-julkaisut

Build more responsive apps with the Decisions API · OpenAI

How Far Will AI Assistants Go for a Refund They Can't Get? · Hacker News: agents

FairStress: A deeper evaluation of fairness in LLM decision-making · Ai2 Blog

Beyond g-mean²: complementary metrics for monitorability · OpenAI Alignment Research Blog

Advancing coding in low-resource programming languages · OpenAI

ALTK-Evolve: How learned memory tops the τ^2 -bench leaderboard on customer service tasks without a frontier model · Hugging Face -blogi

Launch HN: Laminar (YC S24): RL-trained model that finds errors in agent traces · Hacker News: MCP

Terminal X matches senior investor judgment 88.6% with GPT-6.1 · OpenAI

How Specializing the Vocabulary of SLMs Can Cut Inference Cost by 77% in Real Applications · Hugging Face -blogi

NVIDIA Kumo Relational Outperforms Supervised Machine Learning on Relational Data, No Training Required · Hugging Face -blogi

Every AI model has the same "controversial" take on humans · Hacker News: MCP

A crop-mapping benchmark for the tropics · Ai2 Blog

Claude Code 2.1.304 · Claude Code changelog

0.164.0 · OpenAI Codex CLI releases

0.164.0-alpha.12 · OpenAI Codex CLI releases

Release v0.67.0-nightly.20261022.g09d81bf11 · Gemini CLI releases

Release v0.67.0-nightly.20261023.ga713410ab · Gemini CLI releases

google-gemini/gemini-cli · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

sparqlite: SPARQL 1.1 inside SQLite, written in Rust · Simon Willison

How Anthropic's Dynamic Workflows Coordinate Claude Agents · Latent Space

Harness2 by Google Research: Recursive Agent Harnessing · Hacker News: agents

Show HN: Lights-on software factory boilerplate for Claude Code · Hacker News: Claude Code

Tmux and Claude Code: Sessions, Send-Keys, and the Ceiling · Hacker News: Claude Code

Why AI Compute Comes in Rectangles: Erik Bernhardsson, Modal CEO · Latent Space

battledrones v14: Compass Rose, a four player map · Simon Willison

Claude Mods: Flipping the Harness Inside Out · Latent Space

Show HN: agent-usage-guard, a Claude Code mod to protect your usage window · Hacker News: Claude Code

Show HN: Sphinx, A hook that quizzes you on your AI's code before shipping · Hacker News: agents

Ask HN: How to deal with shipping AI-written code you don't understand? · Hacker News: agents

My obituary for programming · Simon Willison

b11818 · llama.cpp-julkaisut

b11877 · llama.cpp-julkaisut

b11881 · llama.cpp-julkaisut

fix(python): serialize git tool calls, delegate malformed names, require mcp 2.3 · MCP servers commits

Donating Agent Skills to the Agentic AI Foundation · Simon Willison

v0.12.4 · Open WebUI releases

aws/agent-toolkit-for-aws · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Nutlope/hallmark · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

wshobson/agents · GitHub trending -syöte

Release v0.67.0-nightly.20261024.gcd1b45c2d · Gemini CLI releases

sickn33/agentic-awesome-skills · GitHub trending -syöte

Merge pull request #3440: Form the Tasks and Events Working Group and convert the Agents Working Group to an Interest Group · MCP-määrittely commits

Merge pull request #3451 from modelcontextprotocol/pja/update-github-handle · MCP-määrittely commits

Merge pull request #5175: Python servers: build fetch, git and time on MCPServer · MCP servers commits

feat(fetch): build on MCPServer; an unknown tool is a -32602 error · MCP servers commits

fix: raise the TypeScript servers' Node floor to 24 · MCP servers commits

Merge pull request #5245 from modelcontextprotocol/v2/chore/5227-ts-sdk-2.4.0 · MCP servers commits

Merge pull request #5247 from modelcontextprotocol/v2/chore/5243-release-audit · MCP servers commits

Release v0.67.0-nightly.20261021.gd00859d6b · Gemini CLI releases

headroomlabs-ai/headroom · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Graphify-Labs/graphify · GitHub trending -syöte

makecindy/cindy · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

WorldTUI: I trained a 360M model to pretend to be a terminal · Hacker News: MCP

nodejs/node · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Renhuai123/ziwei-doushu · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

jbiaojerry/ebook-treasure-chest · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Kriittisiä haavoittuvuusia PHP:ssa · Kyberturvallisuuskeskus

PHP2Shell: AI uncovers critical RCE in PHP · Hacker News: agents

Today's PHP Security Release, and How We're Strengthening PHP Security · Simon Willison

Yhden yhtiön sooloilu sulki verkkopalvelut Suomessa ja maailmalla: "Avasi hyökkääjille aikaikkunan" · Tivi

Jättimäinen hyökkäysaalto iski nettiin: Määrä räjähti miljooniin tunneissa · Tivi

Valtori ajaa valtion verkkosivuja alas · Tivi

Anthropicin tietoturvajohdaja varoittaa: "Vielä paljon pahempaa on tulossa" · Tivi

Quoting Armin Ronacher · Simon Willison

eMudhra Intermediate CA Compromise · Simon Willison

Merge pull request #5244 from modelcontextprotocol/v2/chore/5229-fetch-certifi · MCP servers commits

fix(deps): bump gitpython to 3.2.1 in src/git/uv.lock for CVE-2026-114145 · MCP servers commits

Claude Code 2.1.309 · Claude Code changelog

Release v0.65.0 · Gemini CLI releases

Claude Code slyly adding users' emails to outbound HTTP · Hacker News: Claude Code

security-guidance: say when the LLM reviews can't run · Claude Code plugins/marketplace (virallinen)

security-guidance: run pattern checks on NotebookEdit cells · Claude Code plugins/marketplace (virallinen)

security-guidance: find Python releases newer than the launcher's list · Claude Code plugins/marketplace (virallinen)

security-guidance: stay out of repositories nested in the work tree · Claude Code plugins/marketplace (virallinen)

security-guidance: reviews run when the Windows certificate store has an unreadable entry · Claude Code plugins/marketplace (virallinen)

security-guidance: bump version to 2.0.16 · Claude Code plugins/marketplace (virallinen)

b11882 · llama.cpp-julkaisut

b11884 · llama.cpp-julkaisut

langgraph-cli=0.4.37 · LangGraph releases

Inside NVIDIA Sentry: Secure AI Agents at Runtime · NVIDIA Technical Blog

How NVIDIA OpenShell Reference Architecture Secures and Governs AI Agent Workloads at Scale · NVIDIA Technical Blog

Using Claude to reduce the attack surface of code · Hacker News: Claude Code

How abilitated models can get you pwned · Hacker News: MCP