AI Horisontti



Avaimet *ovessa*

Tekoälyavustajat oppivat viikolla kirjoittamaan asiakasrekistereihin ja soittamaan yrityksiin käyttäjän puolesta. Samalla kaksi ohjelmointiavustajaa alkoi oletuksena tehdä aiempaa enemmän kysymättä.

TIETOTURVA
Tekoäly löytää aukkoja nopeammin kuin niitä ehditään käsitellä

HINNAT
OpenAI:n huippumalli halpeni, kahden Gemini-mallin hinta tuplaantuu

MARKKINA
Clauden kehittäjä Anthropic listautui torstaina pörssiin

PÄÄKIRJOITUS · VIIKKOKATSAUS

Kenen avaimilla avustaja kulkee

Uusi työntekijä saa ensimmäisenä aamunaan kulkukortin, joka avaa vain ne ovet, joiden takana hänen työnsä on. Kortti on harkittu, koska kukaan ei vielä tiedä, mihin tulokas pystyy. Tekoälyavustaja sai kuluneella viikolla kokonaisen avainnipun. ChatGPT oppi kirjoittamaan tietoja suoraan yritysten asiakasrekistereihin ja käyttämään terveydenhuollon ohjelmia ihmisen tavoin. Googlen Gemini taas alkoi soittaa yrityksiin käyttäjän puolesta. Avustaja ei enää jätä luonnosta ihmisen hyväksyttäväksi vaan tekee työn itse jonkun tunnuksilla ja näkee kaiken, mihin ne riittävät.

Osa avaimista jäi viikolla suoraan oveen. Kahden ohjelmointiavustajan oletusasetukset muuttuivat niin, että avustaja tekee aiempaa enemmän kysymättä, jolloin ratkaisevat päätökset jäävät helposti asetuksiin, joita kukaan ei ole tietoisesti valinnut. Samaan aikaan tutkimuslaitos Epoch AI antoi kuudelle tekoälymallille omia arkitöitään, eikä yksikään yltänyt tasolle, jota laitos vaatii työntekijöiltään. Osio ”Mitä tämä tarkoittaa käytännössä” kääntää tilanteen kolmeksi kysymykseksi, jotka jokaisen kannattaa esittää ennen kuin päästää avustajan asiakastietoihin: saako se vain lukea vai myös muuttaa, jääkö sen toimista jälki ja kuka tarkistaa sen työn.

Rahan puolella suunta riippui siitä, kenen malli taustalla pyörii. OpenAI:n uuden huippumallin perushinta laski viidenneksellä, kun taas Google kaksinkertaistaa kahden aiemman Gemini-mallinsa käyttöhinnan 1. joulukuuta. Verkkokaupan joulusesongin asiakaskeskustelut voivat siis maksaa joulukuussa kaksi kertaa marraskuun verran, mikäli palvelu nojaa jompaankumpaan niistä. Clauden kehittäjä Anthropic listautui torstaina pörssiin. Pörssiyhtiönä se raportoi jatkossa tuloistaan ja kuluistaan säännöllisesti, joten avainsepän omat kirjat aukeavat myös asiakkaiden luettaviksi.

Lukkojen puolella viikko oli raskas. Tekoäly löysi ohjelmistoista aukkoja nopeammin kuin niiden ylläpitäjät ehtivät niitä käsitellä. Laajasti käytetyn tietokantaohjelmiston uusi versio myöhästyi kahdella kuukaudella, koska kehittäjät joutuivat käymään läpi tekoälyn raportoimien haavoittuvuuksien tulvan. Markkina- ja turvallisuusosio kokoaa viikon päivityskehotukset, joiden perusteella pienyrityksen kannattaa kysyä verkkosivujensa ylläpitäjältä, ovatko korjaukset jo asennettu. Työkaluosion kiinnostavin uutinen vie puolestaan yhden avaimen takaisin omaan taskuun. Uusi avoin malli, jonka kuka tahansa voi ladata, toimii runsasmuistisella Macilla ilman pilvipalvelua, joten luottamuksellinen asiakirja ei lähde omalta koneelta. Mallin kyvyistä kerrotut luvut ovat toistaiseksi pääosin julkaisijan omia.

Osio ”Mitä seurata” nostaa esiin viikon suurimmat tiedeväitteet, jotka odottavat yhä tarkistajiaan. Matemaatikko kertoo löytäneensä yhdessä Clauden kanssa todistuksen alkulukuja koskevaan ongelmaan, joka on ollut ratkaisematta vuodesta 1859. Vahvistusta ei vielä ole, joten avainta ei ole kokeiltu lukkoon. Tekninen osa menee syvemmälle ja kertoo muun muassa turvatyökalusta, joka näytti lukitulta ovelta mutta jätti osan tarkistuksistaan hiljaa tekemättä. Avustajalle kuuluu sama kulkukortti kuin tulokkaalle.

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

Tekoälyavustaja siirtyi viikon mittaan päivä päivältä ehdottamisesta tekemiseen. Maanantaina se oppi käyttämään terveydenhuollon ohjelmia ihmisen tavoin, keskiviikkona kirjoittamaan tietoja yritysten asiakasrekistereihin ja lauantaina soittamaan yrityksiin käyttäjän puolesta. Muutos on tärkeä, koska samalla viikolla kahden ohjelmointiavustajan oletusasetukset muuttuivat niin, että avustaja tekee aiempaa enemmän kysymättä ja ratkaisevat päätökset jäävät helposti asetuksiin, joita kukaan ei ole tietoisesti valinnut. Markkinan kuvaa muutti torstaina Anthropicin listautuminen historian suurimmalla listautumisannilla, kun taas päivästä toiseen jatkui tietoturvan kierre, jossa tekoäly löysi ohjelmistoista aukkoja nopeammin kuin niiden ylläpitäjät ehtivät niitä käsitellä. Kohinaksi jäivät tällä erää toimittajien omat asiakastarinat ja suurimmat tiedeväitteet, kuten tekoälyn avulla syntyneiksi kerrotut todistukset fysiikassa ja matematiikassa, joille ei viikon aikana saatu riippumatonta vahvistusta.

---

1 **Avustaja alkoi toimia omin päin.** OpenAI toi ChatGPT:hen terveydenhuollon ohjelmistoja ihmisen tavoin käyttäviä avustajia ja myyntityön lisäosan, joka kirjoittaa tiedot suoraan yrityksen asiakasrekisteriin. Laajasti käytettyyn standardiin, jolla tekoälyavustaja kytketään muihin ohjelmiin, lisättiin samalla viikolla tapa käynnistää avustaja esimerkiksi uuden viestin tai järjestelmähälytyksen perusteella. Yrittäjälle ratkaisevaksi nousee, mitä avustaja näkee, saako se muuttaa tietoja vai vain lukea niitä ja kuka tarkistaa sen työn.

2 **Hinnat liikkuivat kahteen suuntaan.** OpenAI:n uuden huippumallin perushinta on viidenneksen edeltäjäänsä alempi ja täsmälleen sama kuin Anthropicin vastaavan mallin. Google puolestaan kaksinkertaistaa kahden aiemman Gemini-mallinsa käyttöhinnan 1. joulukuuta, kun mallien julkaisualennus päättyy. Tekoälyä hyödyntävän palvelun ostajan kannattaa kysyä toimittajaltaan, mitä mallia palvelu käyttää ja seuraako sen hinta mallien hintoja kumpaankin suuntaan.

3 **Anthropic pörssiyhtiöksi.** Clauden kehittäjä Anthropic listautui torstaina pörssiin arvostuksella, joka oli Latent Space -uutiskirjeen mukaan 1,5 biljoonaa dollaria. Osake päätti ensimmäisen kaupankäyntipäivänsä noin 17 % merkintähintaa korkeammalla. Pörssiyhtiönä Anthropic raportoi jatkossa tuloistaan ja kuluistaan säännöllisesti, mikä antaa sen asiakkaille aiempaa paremman kuvan toimittajan taloudellisesta pohjasta.

4 **Tietoturvan kierre kiihtyi.** Kyberturvallisuuskeskus kehotti keskiviikkona päivittämään Javan ja yleisen verkkopalvelinohjelmiston viipymättä, minkä jälkeen levisi varoitus kriittisestä aukosta Php-ohjelmointikielessä, jonka varassa toimii suuri osa verkkosivuista. Laajasti käytetyn tietokantaohjelmiston uusi versio viivästyi kahdella kuukaudella, koska sen kehittäjät joutuivat käsittelemään tekoälyn raportoimien haavoittuvuuksien tulvaa. Pienyrityksen kannattaa kysyä verkkosivujensa ylläpitäjältä, onko viikon korjaukset asennettu.

5 **Oikea työ on eri asia kuin koe.** Tutkimuslaitos Epoch AI:n testissä yksikään kuudesta tekoälymallista ei yltänyt laitoksen omasta arkityöstä poimituissa tehtävissä tasolle, jota laitos vaatii työntekijöiltään. Harvardin tutkijoiden havaintojen mukaan tekoälyn tuottaman koodin tarkastaminen on muuttunut ohjelmistotyön hitaimmaksi vaiheeksi. Tarjouksiin ja aikatauluihin kannattaa siksi varata aikaa tarkastamiselle silloinkin, kun luonnos syntyy minuuteissa.

6 **Hakuliikenne valuu tekoälyvastauksiin.** Revscope-yhtiön oman arvion mukaan Googlen tekoälyvastaukset ovat vähentäneet sivustoille ohjautuvaa hakuliikennettä 40 %. Samalla viikolla OpenAI sopi kahden ranskalaisen mediaryhmän kanssa niiden juttujen tuomisesta ChatGPT:hen ja valitsi ChatGPT:n mainoksille yksinoikeudellisen jälleenmyyjän 44 maahan. Oman sivuston tilanteen näkee Googlen ilmaisesta sivustotyökalusta, joka erottelee hakutuloksissa näkymisen ja klikkaukset.

7 **Kykenevä malli mahtuu omalle koneelle.** Thinking Machines -yhtiön uusi avoin malli eli malli, jonka kuka tahansa voi ladata ja ajaa itse, lukee tekstiä ja kuvia ja toimii runsasmuistisella Macilla ilman pilvipalvelua. Luottamuksellisen asiakirjan voi siis antaa kykenevän tekoälyn luettavaksi ilman, että teksti lähtee omalta koneelta. Mallin kyvyistä kerrotut luvut ovat toistaiseksi pääosin julkaisijan ja jakelijoiden omia.

01 Mitä tämä tarkoittaa käytännössä

Kiinteistönvälitystoimiston myyjä valmistautuu esittelyyn ja pyytää ChatGPT:tä kokoamaan tiedot ostajaehdokkaasta, laatimaan tapaamista varten muistion ja kirjaamaan kaiken jälkikäteen toimiston asiakasrekisteriin. Iltapäivällä toimiston puhelimeen soittaa kohtelias ääni, joka haluaa varata näyttöajan asiakkaan puolesta. Soittaja on tekoäly. Illalla verkkosivuja ylläpitävä alihankkija jättää ohjelmointiavustajan päivittämään varauslomaketta ja lähtee kotiin, koska avustaja on asetettu hyväksymään omat toimensa. Jokainen näistä tilanteista tuli kuluneella viikolla mahdolliseksi tai aiempaa helpommaksi. Yhteistä niille on se, että tekoäly tekee työn toisessa järjestelmässä eikä enää vain tarjoa ihmiselle luonnosta hyväksyttäväksi.

Avustaja, joka kirjoittaa asiakasrekisteriin tai liikkuu potilastietojärjestelmässä, toimii aina jonkun tunnuksilla ja näkee kaiken, mihin nuo tunnukset riittävät. Ennen käyttöönottoa kannattaa siksi selvittää kolme asiaa. Ensimmäinen on se, saako avustaja vain lukea tietoja vai myös muuttaa niitä. Toinen on se, jääkö sen jokaisesta toimesta merkintä tapahtumalokiin eli järjestelmän omaan kirjanpitoon, josta jälkikäteen näkee, kuka teki mitä ja milloin. Kolmas on se, kuka katsoo työn läpi ennen kuin se lähtee asiakkaalle tai muuttaa rekisteriä pysyvästi. OpenAI toi myyntityön lisäosan rinnalle yli 30 uutta tai laajennettua kumppaneiden liitännäistä eli lisäosaa, joiden kautta ChatGPT pääsee käsiksi muiden yhtiöiden järjestelmiin.

Julkistuksesta ei selviä, mihin tilauksiin lisäosat kuuluvat ja ovatko ne heti käytettävissä EU:ssa. Asiakasrekisterissä on lähes aina henkilötietoja, jolloin EU:n tietosuoja-asetuksen velvoitteet koskevat myös avustajan tekemää käsittelyä. Terveydenhuollossa vaatimukset ovat vielä tiukemmat, koska asetuksen 9 artikla luokittelee terveystiedot erityisiin henkilötietoryhmiin. OpenAI ei ole kertonut terveydenhuollon avustajiensa saatavuudesta Suomessa eikä potilastietojen käsittelyn ehdoista Euroopassa. Klinikan tai työterveysyrityksen ei kannata kokeilla palvelua oikeilla potilastiedoilla ennen kuin sopimus, tietojen säilytyspaikka ja käsittelyn ehdot on käyty läpi.

Puhelimeen vastaavalle henkilökunnalle muutos on toisenlainen. Google kertoi lauantaina, että sen vuonna 2018 ravintolavarauksia varten esitelty puhelinavustaja soittaa nyt Geminissä yrityksiin käyttäjän puolesta. Google on aiemmin kertonut, että palvelu ilmoittaa puhelun alussa olevansa automaattinen. Varaus kannattaa silti vahvistaa samalla tavalla kuin ihmisen kanssa tehty varaus. Saatavuudesta Suomessa ei ole tietoa. Aiemmin Google on tuonut puhelinominaisuutensa ensin Yhdysvaltoihin.

Ohjelmointiavustajissa sama kysymys ratkeaa asetusvalikossa. Anthropicin ohjelmointityökalu aloittaa nyt Amazonin, Googlen ja Microsoftin pilvipalvelujen kautta käytettynä automaattitilassa, ellei käyttäjä ole itse valinnut muuta. Automaattitilassa työkalu arvioi itse, mitkä toimet se voi tehdä ilman erillistä hyväksyntää. Toinen suosittu ohjelmointiavustaja, jonka koodi on avoimesti kaikkien nähtävissä, saa nyt oletuksena tehdä kaiken kysymättä silloin, kun sitä ajetaan automaattisesti eikä kukaan ole hyväksymässä toimenpiteitä. Kumpikin muutos nopeuttaa työtä ja siirtää samalla harkintaa ihmiseltä ohjelmalle. Verkkosivunsa tai sovelluksensa alihankkijalla teettävä yritys voi kysyä tekijältä, käyttääkö tämä tekoälyavustajaa ja millä asetuksilla. Kysymys on samaa sukua kuin se, kenellä on avaimet toimistoon.

Mitä enemmän avustaja tekee itse, sitä enemmän painaa se, mitä sen jäljiltä tarkistetaan. Tutkimuslaitos Epoch AI antoi viikolla kuudelle tekoälymallille samoja tehtäviä, joita sen omat työntekijät tekevät. Yksikään malli ei yltänyt tasolle, jolla laitos hyväksyisi ihmisen tekemän työn. Toinen esimerkki tuli ohjelmistomaailmasta. Laajasti käytetyn ohjelman, joka siirtää tietoa verkon yli lukemattomissa laitteissa ja palveluissa, pääkehittäjä laski tekoälyagenttien eli itsenäisesti monivaiheisia tehtäviä hoitavien tekoälyohjelmien kirjoittaneen ohjelman 185 000 koodiriviä uudelleen toisella ohjelmointikielellä runsaassa kolmessa viikossa. Samaan aikaan ihmiset tekivät alkuperäiseen ohjelmaan yli 300 muutosta, joista runsaat 200 oli virhekorjauksia. Yksikään agenttien tekemä kopio ei sisällä niitä. Harvardin tutkijoiden havaintojen mukaan tarkastaminen on nyt ohjelmistotyön hitain vaihe. Sama pätee teksteihin, raportteihin ja laskelmiin, joiden luonnoksen tekoäly kirjoittaa minuuteissa. Niiden lukemiseen ennen asiakkaalle lähtöä kuluu yhä ihmisen aikaa, joka kannattaa merkitä tarjoukseen ja aikatauluun. Luonnos halpenee vuosi vuodelta. Vastuu siitä pysyy sillä, jonka nimi on laskussa.

Pienen verkkokaupan asiakaspalveluikkuna, kuittien kirjausehdotuksia tekevä kirjanpitopalvelu ja tarjouksia luonnosteleva työkalu käyttävät usein taustalla jonkin suuren yhtiön kielimallia eli tekstiä ymmärtävää ja tuottavaa tekoälyä. Palveluntarjoaja maksaa mallin käytöstä sen mukaan, kuinka paljon tekstiä malli lukee ja kirjoittaa. Viikon hintauutiset osuivat ensin juuri tähän näkymättömään kerrokseen. OpenAI:n uuden huippumallin perushinta laski edeltäjästä viidenneksellä. Toistuvasti syötettävän taustatekstin, kuten saman ohjeistuksen tai asiakirjan, hinta putosi 60 %. Hinta on nyt sama kuin Anthropicin kilpailevalla huippumallilla, joka halpeni aiemmin samassa suhteessa. Yrittäjän omaan laskuun säästö ulottuu vain, mikäli palveluntarjoaja päättää jakaa sen eteenpäin.

Googlen malleja käyttäville suunta on päinvastainen. Kahden aiemman Gemini-mallin käyttöhinta kaksinkertaistuu 1. joulukuuta, kun niiden julkaisualennus päättyy. Joulusesongin asiakaskeskustelut voivat siis maksaa joulukuussa kaksi kertaa marraskuun verran, mikäli palvelu käyttää jompaakumpaa näistä malleista. Muutos koskee rajapintaa eli liitäntää, jonka kautta yrityksen oma ohjelma käyttää mallia ja josta laskutetaan käytön mukaan. Julkaistuista tiedoista ei käy ilmi, muuttuuko kuluttajien Gemini-tilausten hinta.

Hinnaston rinnalla laskuun vaikuttaa asetus, josta harva ostaja tietää. Googlen uudelle keskitason mallille voi säätää, kuinka pitkään se miettii ennen vastaamista. Ensimmäisessä riippumattomassa mittauksessa kevyimmällä asetuksella ajettu malli jäi selvästi tasosta, jolla sitä markkinoitiin. Tunnettu ohjelmistokehittäjä ja bloggaaja kokeili viikon lopulla yhtä Googlen uusista malleista korkeimmalla harkinta-asetuksella. Yhdessä piirrostehtävässä malli käytti noin kuusi kertaa enemmän tekstiä kuin pykälää alemmalla asetuksella. Tehtävän hinta nousi noin 4 sentistä noin 24 senttiin ja tulos parani selvästi. Yksi piirros ei ole mittaus. Saman tehtävän hinta voi silti vaihdella samalla mallilla kuusinkertaisesti sen mukaan, mihin asentoon harkinnan säädin on jätetty.

Ennen joulukuuta palvelun toimittajalta kannattaa kysyä kolme asiaa: mitä mallia palvelu käyttää, millä harkinta-asetuksella ja paljonko joulukuun käyttö maksaa uusilla hinnoilla. Halvempi malli tai kevyempi asetus voi riittää yksinkertaisiin tehtäviin, kuten toimitusajan kertomiseen. Vaihto kannattaa testata omilla asiakaskysymyksillä ennen käyttöönottoa. Omaa tekoälypalvelua myyvä yrittäjä sai viikolla lisäksi varoituksen. Tuotevastaaville määrittelydokumentteja kirjoittavan tekoälypalvelun perustaja kertoi järjestäytyneistä ryhmistä, jotka käyttävät ilmaiskokeiluja hyväkseen ja myyvät käyttöoikeutta eteenpäin, vaikka jokainen pyyntö maksaa palvelun tarjoajalle rahaa. Ilmaiskokeilu kannattaa siksi rajata käyttömäärällä eikä pelkillä päivillä. Mallin toimittajan hallintapaneeliin kannattaa lisäksi asettaa kuukausittainen kulukatto, joka pysäyttää poikkeavan käytön ennen kuin lasku ehtii kasvaa.

Näkyvyydessä raha alkaa kiertää yrittäjän ohi. Moni pieni yritys on hankkinut asiakkaansa vuosia saman kaavan avulla: hyödyllinen ohje tai vertailu nostaa sivun Googlen tuloksiin ja tuloksista tulee kävijöitä. Kaava katkeaa siinä kohdassa, jossa hakukone kirjoittaa vastauksen itse tulossivun yläreunaan. Revscope-yhtiön oman arvion mukaan Googlen tekoälyvastaukset ovat vähentäneet sivustoille ohjautuvaa hakuliikennettä 40 %. Ohjelmoijien suosimalla keskustelusivustolla yli 700 kommenttia keränneessä kirjoituksessa Berliiniin muuttaville suunnatun opassivuston ylläpitäjä kuvasi, miten tekoäly on tappamassa hänen sivustoaan. Sivusto on juuri sellaista käytännön ohjeistusta, jota tekoälyvastaukset tiivistävät. Oman sivuston tilanteen voi tarkistaa Googlen ilmaisella Search Console -sivustotyökalulla, joka näyttää erikseen, kuinka usein sivu näkyi hakutuloksissa ja kuinka usein sitä klikattiin. Näkymisten pysyessä ennallaan ja klikkausten laskiessa hakusivun vastaus on todennäköisesti alkanut hoitaa sen, minkä vuoksi sivulle ennen tultiin. Alttiinta on sisältö, jonka tekoäly voi tiivistää, kuten ohjeet, vertailut ja usein kysytyt kysymykset. Ajanvaraus, tarjouspyyntö ja ostos vaativat edelleen käynnin yrityksen omilla sivuilla.

Samaan aikaan tekoälyvastausten ympärille rakentuu maksullisia reittejä. OpenAI teki viikolla sopimukset luksuskonserni LVMH:n mediadivisioonan ja ranskalaisen mediatalon Le Groupe Figaron kanssa niiden journalismin tuomisesta ChatGPT:hen. Le Figaron jutut näkyvät tiivistelminä ja otteina, joissa on linkit alkuperäisiin juttuihin. Mainoksilleen yhtiö valitsi Aleph-nimisen yrityksen yksinoikeudelliseksi jälleenmyyjäksi 44 maahan, joissa mainostajat saavat paikallisen laskutuksen ja henkilökohtaista tukea. Yhtiön omassa esimerkissä painotuotteita myyvä VistaPrint tavoittaa ChatGPT:n kautta pienyrittäjiä, jotka valmistelevat avajaisia eivätkä vielä tiedä, mitä niihin pitäisi painattaa. Suomalaisen mainostajan kannattaa tarkistaa, kuuluuko Suomi näihin maihin. Suurella mediatalolla reitti tekoälyn vastauksiin kulkee sopimuksen kautta ja mainostajalla rahan kautta. Pienen opassivuston ylläpitäjällä on vain teksti, jonka tekoäly osaa jo kertoa omin sanoin.

02 Työkalut ja avustajat arjessa

OpenAI julkaisi tiistai-iltana kaksi uutta kielimallia: järeämmän vaativiin töihin ja kevyemmän arkikäyttöön. Ne tulivat ensimmäisinä ChatGPT:n työkäyttöön tarkoitettuun versioon, yhtiön ohjelmointiavustajaan ja rajapintaan. Julkaisutiedoista ei selviä, milloin mallit tulevat ChatGPT:n muihin tilauksiin. Yhtiön mukaan uudet mallit tuottavat ammattimaisempia työdokumentteja, ohjelmoivat paremmin ja käyttävät tietokoneen ohjelmia itse aiempaa pienemmin kustannuksin. Riippumaton arviointipalvelu mittasi molemmat jo julkistuspäivänä ja totesi ne yleisesti kyvykkäiksi. Tekoälyalan uutiskirje Latent Space arvioi, että kevyempi malli hoitaa todennäköisesti suurimman osan päivittäisestä työstä. Huippumallin varhaiset testaajat ovat puolestaan eri mieltä siitä, millaista sen kanssa on tehdä töitä.

Viikon aikana OpenAI julkaisi myös useita asiakastarinoita. Rahoitusteknologiayhtiö Revolutin kerrotaan lyhentäneen koodaukseen kuluvaa aikaa 30–60 % OpenAI:n mallien ja ohjelmointiavustajan avulla. Brittiläinen matkailualan yritys Holiday Extras käyttää avustajia, jotka toimittavat raportteja kenenkään pyytämättä ja käyvät työntekijän kanssa läpi tulevia palavereja työmatkan aikana. Japanilainen henkivakuutusyhtiö Sumitomo Life tutkii ChatGPT:n yritysversiolla muita vakuutusyhtiöitä ja valmistelee sillä tarjouksia. Talousalan tutkimustyökaluja kehittävän Hebbian mukaan OpenAI:n uusi malli kattoi ennakkotestissä lähteet paremmin ja viittasi niihin osuvammin. Kaikki neljä ovat toimittajan omaa markkinointiaineistoa eivätkä riippumattomia mittauksia. Sumitomo Lifen tarinan käyttökelpoisin havainto koskee ihmisiä eikä tekniikkaa:

johtajien varhainen oma käyttö auttoi työkalua leviämään koko henkilöstöön.

Uuden mallin lupauksen voi tarkistaa luotettavimmin omalla työllä. Edelliseltä viikolta kannattaa valita kolme tavallista tehtävää ja ajaa ne sekä vanhalla että uudella mallilla. Sopivia ovat esimerkiksi asiakaskirje, tarjouslaskelman tarkistus ja pitkän sopimuksen tiivistelmä. Yleinen mittaus ei kerro, tunteeko malli suomalaisen palkanlaskennan käsitteet tai osaako se kirjoittaa asiakkaalle oikealla äänensävyllä. Pyynnön muotoiluun saatiin viikolla havainnollinen esimerkki. Data-analyysityökalun, jonka koodi on kaikkien vapaasti käytettävissä, kehittäjä rakensi uuden karttalisäosan lähes kokonaan yhdellä pyynnöllä Anthropicin Claudelle. Pyyntö ei kuvannut lopputulosta tarkasti. Se osoitti avustajalle kaksi aiempaa karttalisäosaa ja yhden kokeiluun tehdyn verkkosivun ja pyysi yhdistämään niiden parhaat puolet. Sama keino toimii tekstien, taulukoiden ja raporttien kanssa, koska valmis esimerkki kertoo halutun muodon usein paremmin kuin sanallinen kuvaus.

Juristeille OpenAI avasi lakitekstien käsittelyyn erikoistuneen mallin, johon ChatGPT:n suuryritysversion asiakkaat voivat pyytää pääsyä. Yhtiö markkinoi mallia alansa parhaaksi oman vertailunsa perusteella. Julkistuksesta ei selviä, kattaako vertailu Suomen tai EU:n lainsäädäntöä. Suomalaisessa asianajotyössä vaitiolovelvollisuus ratkaisee usein jo etukäteen, mitä aineistoa tekoälylle ylipäätään voi antaa. Tutkijoille yhtiö lupasi maksuttoman pääsyn kehittyneimpiin malleihinsa ohjelmassa, joka kattaa 200 000 tutkijaa eri puolilla maailmaa. Oman organisaation ohjeet luottamuksellisen aineiston käsittelystä koskevat ilmaista palvelua samalla tavalla kuin maksullista.

Googlen uusi keskitason Gemini-malli sijoittuu kevyimpien ja kalleimpien mallien väliin. Latent Space kuvasi sitä kalleimpien kärkimallien haastajaksi, mikä on julkistuksen ympärille rakentunut kehys eikä vielä laajasti mitattu tulos. Kevyt harkinta-asetus vastaa nopeammin ja maksaa vähemmän. Pitkän sopimuksen läpikäyntiin se ei välttämättä riitä. Saatavuus Suomessa ja EU:ssa kannattaa tarkistaa Googlen omasta palvelusta ennen käyttöönottoa. Microsoft puolestaan julkisti yritysasiakkailleen suunnatussa tapahtumassa seitsemän uutta tekoälymallia, joita se markkinoi hyvällä hinnan ja suorituskyvyn suhteella. Väite on yhtiön oma.

Ohjelmointiavustajiin tuli viikon aikana useita turvallisuusmuutoksia. Ne koskevat myös yrityksiä, jotka eivät itse ohjelmoi mutta teettävät sivunsa ja sovelluksensa muilla. Anthropicin ohjelmointiavustaja kertoo nyt kansiota avattaessa, mihin kansion koukut on määritelty. Koukut ovat asetuksia, jotka ajavat komentoja automaattisesti avustajan työn lomassa. Verkosta ladatun tai työkaverilta saadun projektin mukana voi siis tulla komentoja, joita käyttäjä ei ole itse kirjoittanut. Saman avustajan tietoturvalisäosaan, joka tarkistaa avustajan tekemiä koodimuutoksia, tuli korjaus aukkoon, jonka vuoksi osa muutoksista jäi tarkistamatta, kun komento oli jaettu kahdelle riville. Työkalu ottaa nyt myös kirjautumisen yhteydessä käyttöön asetukset, jotka yrityksen ylläpitäjä on määrittänyt keskitetysti. OpenAI kuvasi blogissaan, miten sen ohjelmointiavustaja pitää ajamansa komennot käyttäjän asettamissa rajoissa tiedostojen ja verkkoyhteyksien osalta. Tunnettu ohjelmistokehittäjä ja bloggaaja esitteli yhdessä tekoälytutkimuslaboratorio Prime Radiantin kanssa mallin, jossa agentin ajama koodi eristetään hiekkalaatikkoon eli suljettuun tilaan, josta se ei pääse koskemaan muuhun koneeseen. Eristyksen tavoitteena on, ettei virhe tai kehoteinjektio eli tekoälylle luettavaksi annettuun tekstiin piilotettu vihamielinen ohje pääse tekemään vahinkoa. Avoin ohjelmointiavustaja, joka hyväksyy automaattiajoissa nyt kaiken kysymättä, aloittaa uuden käyttäjän ilmaisella mallilla Anthropicin maksullisen mallin sijaan. Sen tilinäkymässä on nyt linkki tietosuoja-asetuksiin, joissa käyttäjä päättää, miten palvelu saa käyttää hänen tietojaan.

Kulujen näkyvyys parani pienin askelin. Googlen ohjelmoijille tarkoitettu tekoälyavustaja näyttää nyt päivittäisen käytön lämpökarttana, josta näkee yhdellä silmäyksellä, minä päivinä tekstiä on kulunut eniten. Pieni avoin työkalu puolestaan kertoo, missä Anthropicin ohjelmointityökalun käytössä maksetaan tarpeettoman paljon. Anthropic antaa lisäksi Claudeen perustuvien sovellusten tekijöille kokeiluvaiheessa mahdollisuuden laskuttaa tekoälyn käyttö suoraan asiakkaan omalta Anthropic-tililtä. Asiakas voi asettaa sovellukselle kulukaton tai katkaista yhteyden kokonaan. Ominaisuus koskee vain asiakkaita, jotka maksavat tekoälyn käytöstä suoraan Anthropicille. Mallin käyttöönotosta kannattaa kysyä sovellustoimittajalta. Silloin tekoälyn kulut näkyvät omalla laskulla eivätkä piiloudu sovelluksen hintaan.

Omalla koneella ajettava tekoäly eteni koko viikon pienin askelin ja lauantaina isommalla. Thinking Machinesin avoin malli vie noin 20 gigatavua muistia, mikä tarkoittaa peruskonetta tehokkaampaa Macia: 16 gigatavun muistilla varustettu kannettava ei riitä. Helpoimman asennustavan tarjoava ilmainen ohjelma tukee mallia toistaiseksi vain Apple-koneilla. Windows-käyttäjälle sama onnistuu toistaiseksi vain teknisemmillä työkaluilla. Mallin lisenssi eli käyttöehdot kannattaa lukea ennen asiakastyötä, koska

avoimeksi kutsuttu malli ei aina salli vapaata kaupallista käyttöä. Julkaisutiedoissa lisenssiä ei mainita. Mallia voi kokeilla antamalla sille saman tehtävän, jonka nykyinen pilvipalvelu jo hoitaa, esimerkiksi tarjouspyynnön tiivistämisen tai ostolaskun tietojen poimimisen. Vasta rinnakkain vertailtuna selviää, riittääkö oman koneen malli juuri siihen työhön.

Viikon pienemmät askeleet kulkivat samaan suuntaan. Omalla koneella ajettavien mallien suosituissa ohjelmissa Mac-koneet oppivat testiversioissa käsittelemään useita pyyntöjä yhtä aikaa aiempaa nopeammin, mistä on hyötyä, kun samaa konetta käyttää useampi ihminen. Työkäytössä kannattaa odottaa vakaata versiota. Suosittuun selaimessa toimivaan avoimeen keskustelunäkymään tuli lisäsovellus, jonka avulla toimiston tehokkaalla koneella pyörivää mallia voi käyttää myös puhelimesta tai kotikoneelta ilman, että malli viedään pilveen. Kehittäjien mukaan koneen osoitteet, komennot ja salaiset avaimet eivät lähde koneelta. Avoimien mallien jakelualusta Hugging Face esitteli pienen puhemallin, joka tukee yli 40 kieltä ja mahtuu tavallisen kannettavan muistiin. Julkistuksesta ei selviä, onko suomi mukana. Laadussa pienet mallit jäävät yhä suurista pilvimalleista.

Hugging Face kertoi viikolla myös yhteistyöstä OpenRouterin kanssa. OpenRouter on välityspalvelu, jonka kautta sama sovellus voi käyttää kymmenien palveluntarjoajien malleja yhdellä tilillä. Välittäjän kautta kulkevat kysymykset ja asiakirjat päätyvät kolmannelle osapuolelle, jonka sijainti ja tietosuojakäytäntö vaihtelevat mallista toiseen. Asiakastietoja sisältäviä tehtäviä ei kannata ohjata välittäjän kautta ennen kuin on selvää, missä maassa malli ajetaan ja tallennetaanko syötteitä.

Avoimet työkalut tuovat arkeen myös uusia riskejä. Suosittuun avoimeen malliohjelmistoon lisättiin äänen kloonaus eli toisen ihmisen äänen jäljittely ääninäytteen perusteella. Sama kyky halventaa selostuksia ja opetusvideoita ja madaltaa äänihuijausten kynnystä. Läheisten tai työyhteisön kesken sovittu tunnussana on yhä yksinkertaisin keino varmistaa, että rahaa pyytävä soittaja on se, jolta hän äänestä kuulostaa. Avoimien mallien jakelupalvelun nousevien mallien listalla ladatuin oli tiistaina kolmannen osapuolen versio Alibaban kuvanluontimallista, josta on poistettu sisältörajoitukset. Sitä oli ladattu yli 160 000 kertaa. Viikon lopulla samalla listalla nousi vastaava rajoitukseton versio tekstiä ja kuvia lukevasta mallista lähes 130 000 latauksella. Yrityskäytössä vastuu tällaisella mallilla tuotetusta sisällöstä jää käyttäjälle. Epävirallisen version alkuperä ja käyttöehdot kannattaa selvittää ennen käyttöä.

Anthropic kertoi upottavansa useiden Claude-mallien tuottamaan tekstiin vesileiman eli tunnisteen, jonka avulla tekstin voi myöhemmin todeta tekoälyn kirjoittamaksi. Merkintä toimii Anthropicin erikseen luettelemilla alustoilla. Clauden avulla kirjoitettujen uutiskirjeiden, tarjousten ja verkkosivutekstien alkuperä voi siis olla jälkikäteen todennettavissa. EU:n tekoälyasetus edellyttää muutenkin, että tekoälyn tuottama sisältö merkitään koneellisesti tunnistettavaksi. Merkintä ei estä tekstin käyttöä.

03 Markkina, sääntely ja turvallisuus

Anthropicin osakkeen merkintähinta oli 83 dollaria. Kaupankäynnin avauksessa hinta nousi 111,50 dollariin ja laski päivän aikana 97,17 dollariin. Latent Space -uutiskirjeen mukaan listautumisanti oli historian suurin ja yhtiön arvoksi muodostui 1,5 biljoonaa dollaria. Tivin mukaan listautuminen tuo varhaisille sijoittajille suuria voittoja. Yksi niistä on tanskalainen hoitajien ja terveydenhuoltoalan eläkejärjestö, joka tekee listautumisessa miljoonien voiton. Rahan tarve näkyy tutkimuslaitos Epoch AI:n luvuista. Laitoksen mukaan Anthropic käytti vuoden 2026 kolmannella neljänneksellä laskentaan ja konesaleihin lähes 10 miljardia dollaria, mikä vastaa vuositasolla lähes 40 miljardia dollaria. Vuonna 2023 vastaava summa oli 433 miljoonaa dollaria. Anthropicin kulut nelinkertaistuvat vuodessa, kun OpenAI:n kulut kolminkertaistuvat. Juuri ennen listautumista tutkijoiden, oikeusoppineiden ja teknologiajohtajien liittouma julkaisi tiedotteen, jonka mukaan Anthropic ei ole turvallinen tekoälyyhtiö. Liittouma kyseenalaistaa yhtiön tavan käyttää turvallisuutta brändinsä kärkenä.

Muualla markkinassa luvut kasvoivat yhtä jyrkästi. Latent Spacen mukaan tekoälyhakupalvelu Perplexityn liikevaihdon vuositahti on noussut 1 miljardiin dollariin ja nelinkertaistunut vuoden 2026 aikana. Kiinalaisten Baidun ja Xiaomin toimitusjohtajat varoittivat samaan aikaan kasvavasta kuilusta tekoälyyn kohdistuvien odotusten ja sen tuottojen välillä. Tekoälysiruja valmistava Nvidia teki tulosraportissaan huipputuloksen. Tivin mukaan OP Pohjolan salkunhoitaja pitää tulosta moitteettomana mutta kutsuu osaketta putoavaksi puukoksi, koska sen arvostus on

nyt hankala kysymys. Yhtiön johdossa kuohuu silti: The Batch -uutiskirjeen mukaan Nvidian toimitusjohtaja ja yksi perustajista erosi edellisellä viikolla. Tivin välittämien Financial Timesin tietojen mukaan vakoilukohuun kaatunut johtaja hakee nyt Piilaakson liittolaistensa ja xAI:n perustajan tuella paluuta yhtiön johtoon. Samaan kohuun liittyy epäily, että vientikiellon alaisia tekoälysiruja olisi välitetty Kiinaan miljardien dollarien arvosta. Kohun keskellä oleva yhtiö kiistää laittomuudet ja on silti aloittanut asiasta sisäisen tutkinnan. Nvidian siruilla pyörii valtaosa maailman tekoälypalveluista.

Pohjoismaissa tekoälyn laskentakeskukset ja sähkö nousivat samalle pöydälle. Norja keskeytti väliaikaisesti uusien datakeskusten liittämisen sähköverkkoonsa, koska niiden kysyntä ylittää verkon kapasiteetin. Päätöksen taustalla on huoli energiapulasta ja sähkön hinnan noususta. Helsingin Sanomien mukaan Nokia suunnittelee miljardien eurojen investointia Ouluun kaavailtuun tekoälyn gigatehtaaseen eli valtavaan laskentakeskukseen, jossa malleja koulutetaan ja ajetaan. It-yhtiö Advania kertoi Microsoftin kanssa solmimastaan datakeskussopimuksesta, jonka arvoa ei kerrottu. Tasavallan presidentti puolusti datakeskuksia sanoin ”enemmän on enemmän” ja muistutti samassa yhteydessä ihmisiä säilyttämään oman ajattelukykynsä. Tivin uutisoiman järjestön mukaan datakeskusten sähkönkulutus ja päästöt kasvavat jyrkästi ennen kaikkea tekoälyn vuoksi. Järjestön mukaan Yhdysvallat ja Kiina kattavat uuden energiantarpeen pitkälti fossiilisilla polttoaineilla. It-alan etujärjestö Tietoala ry varoitti, ettei Suomi saa jäädä pelkäksi kansainvälisen tekoälyteollisuuden palvelinsaliksi. Järjestön mukaan alan liikevaihto kasvaa samaan aikaan, kun henkilöstön määrä vähenee.

Työmarkkinoilta tuli samansuuntaisia havaintoja eri lähteistä. Tivin uutisoimien ekonomistien mukaan tekoäly on vienyt 25 000 työpaikkaa kolmessa kuukaudessa, eniten tekoälylle alttiilla aloilla. Ohjelmoijien suosiman keskustelusivuston kuukausittaisen rekrytointiketjun ilmoitusmäärät pysyttelevät lähellä 12 vuoden pohjalukemia. Projektinhallintaohjelmia tekevän Linearin operatiivinen johtaja kertoi haastattelussa, että yli puolet tiketeistä eli työtehtävien kuvauksista kirjoittavat jo agentit.

Suomalaisesta tekoälykentästä saatiin viikolla kartta. AI Finland ja valtion pääomasijoitusyhtiö Tesi kartoittivat noin 100 tekoälyyn keskittyvää kasvuyritystä, jotka on perustettu vuonna 2020 tai myöhemmin. Useimmat niistä tekevät sovelluksia, kun taas sijoittajien tuomariston valitsemat kymmenen lupaavinta keskittyvät vaativaan teknologiaan aloilla, joilla Suomi on jo vahva. Tivin mukaan myös suomalaiset eläkeyhtiöt ovat innostuneet sijoittamaan kasvuyrityksiin. Brysselissä Euroopan komissio kokosi lähes 1 000 päättäjää, kasvuyritystä, teollisuuden edustajaa ja tutkijaa huippukokoukseen, jossa etsittiin keinoja vauhdittaa tekoälyn käyttöönottoa kymmenellä strategisella toimialalla ja julkisella sektorilla. Perjantaina komissio ja kahdeksan jäsenmaata, Suomi mukaan lukien, perustivat koulutuksen digitaalisen infrastruktuurin konsortion. Sen tavoitteena on koota voimavarat yhteisiin, turvallisiin ja yhteentoimiviin digitaalisiin palveluihin opetuksen tarpeisiin. Julkisiin hankintoihin tarjouksia tekeville Visma Aquilan teknologiajohtaja nosti esiin käytännön ongelman: tarjouspyynnöissä vaaditaan nykyään usein tekoälyä kertomatta, mihin sitä tarvitaan. Vaatimuksen sisältö kannattaa kysyä hankintayksiköltä ennen kuin tarjouksessa sitoudutaan mihinkään.

Varsinaisessa tekoälysääntelyssä viikko oli hiljainen. Kyberturvallisuuskeskuksen viikkokatsaus nosti esiin EU:n kyberkestävyysasetuksen, joka asettaa tietoturvavaatimuksia ohjelmistoa sisältäville ja verkkoon kytkeytyville tuotteille. Katsauksen tiivistelmä ei kerro, mitä asetuksen ajankohtaiset muutokset sisältävät. Ohjelmistoja tai verkkoon kytkeytyviä laitteita myyvän yrityksen kannattaa lukea katsaus kokonaan. EU asetti lisäksi venäläisille hakkereille pakotteita, joihin kuuluu varojen jäädyttämisiä ja matkustuskieltoja.

Tietoturvassa viikko kulki päivityskehotuksesta toiseen. Google Cloudin uhkaraportti nosti tiistaina esiin hyökkäykset, jotka etenevät ilman ihmisen ohjausta. Raportin mukaan hyökkäyskeinot eivät sinänsä ole uusia. Niiden nopeus ja laajuus ovat sen sijaan kasvaneet hurjaa tahtia. Keskiviikkona Kyberturvallisuuskeskus kehotti päivittämään Javan eli ohjelmointikielen ja ajoympäristön, jonka varassa toimii suuri osa yritysten taustajärjestelmistä. Oracle ja Javan avoimen version kehittäjät korjasivat 11 haavoittuvuutta, joista vakavimmat voivat antaa hyökkääjälle mahdollisuuden ajaa komentoja kohdejärjestelmässä verkon yli. Tivin mukaan kriittinen aukko löydettiin tekoälyn avulla. Samana päivänä F5 korjasi yleisestä verkkopalvelinohjelmistostaan kymmenen haavoittuvuutta, joista vakavin sallii koodin ajamisen palvelimella erityisesti muotoillun verkkoliikenteen avulla. Torstaina levisi varoitus Php-ohjelmointikielen kriittisestä haavoittuvuudesta, joka koskee kaikkia tuettuja versioita. Edellinen laaja Php-ongelma työllisti yrityksiä vain kuukausi sitten. Php:llä toimii suuri osa verkkosivuista, muun muassa WordPress-julkaisualusta. Avoimiin automaatiotyökaluihin, joilla yritykset rakentavat tekoälyä hyödyntäviä työnkulkuja, tuli viikon aikana useita

korjauksia. Omalla palvelimella niitä ajavan kannattaa lukea muutosloki ja päivittää.

Viikon lopulla näkyi, mitä tahti tekee korjaajille. Laajalti käytetyn avoimen tietokantaohjelmiston uusi versio julkaistiin kaksi kuukautta tavallista myöhemmin, koska kehittäjät joutuivat käsittelemään tekoälyn raportoimien haavoittuvuuksien tulvaa. Tivin uutisoiman senaattorin mukaan tekoäly löytää tietoturva-aukkoja ennätysvauhtia, mikä pakottaa yritykset hätäisiin päivityksiin, jotka kuormittavat järjestelmiä. Washington Sun -uutissivuston mukaan pankkien toimitusjohtajat ja Yhdysvaltain kansallinen turvallisuusvirasto NSA kertoivat kongressille, että tekoälyn avulla tehdyt ohjelmistokorjaukset ovat palvelukatkojen taustalla. Tivin mukaan Yhdysvaltain presidentti erottaa maan kyberjohtajan. Päätöksen uskotaan liittyvän viime viikkojen it-katkoksiin. Lähteistä ei selviä, ovatko katkokset samoja, joista kongressille kerrottiin. Tivin mukaan myös EU-komissioon hyökättiin tekoälyä hyödyntäen niin, että komission järjestelmät olivat sekaisin päiväkausia. OpenAI kertoi puolestaan työskentelevänsä Naton ja liittolaismaiden kyberviranomaisten kanssa, jotta luotetut puolustajat löytäisivät haavoittuvuuksia ja testaisivat korjauksia tekoälyn avulla.

Pienyrityksen suojakeinot ovat samat kuin ennen: ohjelmistopäivitykset ja kaksivaiheinen tunnistautuminen, jossa kirjautuminen vaatii salasanan lisäksi toisen varmistuksen, esimerkiksi puhelimeen tulevan koodin. Niiden lykkääminen on aiempaa riskialttiimpaa, koska tunnettu aukko löydetään nopeammin. Ulkoistetun verkkosivun tai verkkokaupan omistajan kannattaa kysyä palveluntarjoajalta, koskevatko Kyberturvallisuuskeskuksen 18.11. julkaisemat tiedotteet ja torstain Php-varoitus hänen järjestelmiään ja milloin korjaukset asennetaan. Kotikäyttäjille Kyberturvallisuuskeskuksen viikkokatsaus kertoi, miten rikolliset voivat peittää toimintaansa kotitalouksien internetyhteyksien avulla. Katsaus muistutti myös turvallisesta verkko-ostamisesta ennen Black Friday -alennuspäivää. Sunnuntaina Tivi kertoi, että suomalaishotellien asiakkaat ovat saaneet huijausviestejä, joilla kalastellaan maksu- ja kirjautumistietoja.

Yksityisyydessä viikon kaksi keskustelua koskivat sitä, mitä koneet ja puhelimet näkevät käyttäjästään. Ohjelmoijien suosimalla keskustelusivustolla vilkkaimmin kommentoitu blogikirjoitus väitti tiistaina, että Windows kerää tietoa siitä, mitä tekoälyavusteisia ohjelmointityökaluja koneella käytetään. Kyse on yksittäisen kehittäjän havainnosta, josta myös Tivi uutisoi. Lähteissä ei ole Microsoftin vastausta. Niistä ei myöskään selviä, mihin tietoja käytetään tai voiko keruun kytkeä pois. Yrityskoneiden ylläpitäjä voi tarkistaa ja säätää, millä tasolla koneet lähettävät diagnostiikkatietoja Microsoftille. Lauantaina viestisovellus Signal julkaisi kirjoituksen puhelinten käyttöjärjestelmiin rakennettujen tekoälyagenttien rajoittamisesta niin, etteivät ne pääse keräämään käyttäjän tietoja. Puhelimen avustaja, joka osaa lukea näytön, näkee myös sen, mitä viestisovelluksessa lukee. Signal on toiminut samaan suuntaan aiemminkin: vuonna 2025 sen tietokoneversio alkoi Windowsissa oletuksena estää Microsoftin ominaisuuden, joka tallentaa näytön sisällöstä jatkuvasti kuvia muistin tueksi. Puhelimen asetuksista kannattaa tarkistaa, mihin sovelluksiin ja tietoihin avustajalla on pääsy, varsinkin jos puhelimella hoidetaan asiakkaiden asioita.

OpenAI kertoi viikolla, että neljä riippumatonta tutkimusorganisaatiota, Apollo Research, METR, Transluce ja Goodfire, arvioivat sen mallien käyttäytymistä, suojauksia ja valvontajärjestelmiä yhtiön sisältä käsin. Julkistuksen kuvauksessa ei kerrota, saavatko arvioijat kertoa havainnoistaan julkisesti. Yhtiö julkaisi myös työkaluja ja ohjeita lasten seksuaalisen hyväksikäytön ehkäisemiseksi verkossa sekä kuvauksen siitä, miten se hallitsee kehittyneimpien malliensa riskejä. Kuvaus on yhtiön oma eikä ulkopuolinen arvio. Lauantaina OpenAI kertoi torjuvansa tislausta eli sitä, että joku kerää suuren määrän toisen yhtiön tekoälyn vastauksia ja kouluttaa niillä oman mallinsa. Yrityksen, joka suunnittelee kouluttavansa oman pienen mallin kaupallisen tekoälyn vastauksilla, kannattaa lukea palvelun käyttöehdot ennen hankkeen aloittamista.

Yhdysvalloissa lukion aseentunnistusjärjestelmä luuli Doritos-maissilastupussia aseeksi. Tekoälyn hälytyksen perusteella mustia lukiolaisia otettiin kiinni ja heidät tarkastettiin. Kansalaisoikeusjärjestö Lawyers' Committee for Civil Rights Under Law on nostanut oppilaiden puolesta kanteen. Järjestelmän virhe oli yksi väärä tunnistus, jonka seuraukset kantoivat nuoret.

04 Mitä seurata

Kuka mittaa tekoälyn kyvyt. Riippumaton tutkimusorganisaatio METR toteaa neljännesvuosikatsauksessaan, että tekoälyn tutkimus- ja kehitystyön kannalta olennaiset kyvyt paranevat nopeasti ja

niitä on yhä vaikeampi mitata. Mittaamisen tapoja haettiin viikolla monesta suunnasta. Epoch AI mittasi oikeita työtehtäviä ja alan vertailujärjestö MLCommons julkaisi ohjeet testiin, joka kuormittaa tekoälypalveluja verkon yli niin kuin asiakkaat niitä käyttävät. OpenAI mittasi, kuinka paljon ohjelmoivat agentit nopeuttavat yhtiön omaa tutkimusta. Tulos oli yhtiön mukaan epätasainen. Sunnuntaina yhtiö ehdotti yhteiseksi lähtökohdaksi mittaria, joka kertoo, kuinka hyvin malli osaa ennakoida sille näytettyä tekstiä. Suunnan näkee siitä, tulevatko uudet luvut riippumattomilta arvioijilta vai lähinnä mallien tekijöiltä. Jälkimmäisessä tapauksessa myyntipuheen ja todellisen hyödyn erottaminen jää yhä enemmän ostajalle.

Suuret väitteet odottavat tarkistajia. OpenAI kertoo mallinsa auttaneen viemään ehdotuksen siitä, miten hiukkasfysiikan perusteoria määritellään täsmällisesti, yksityiskohtaiseksi todistukseksi. Matemaatikko puolestaan kertoo löytäneensä yhdessä Clauden kanssa todistuksen Riemannin hypoteesille eli alkulukujen jakautumista koskevalle ongelmalle, joka on ollut ratkaisematta vuodesta 1859. Väitteen mukaan todistus on kirjoitettu myös muotoon, jonka tietokoneohjelma voi tarkistaa askel askeleelta. Stanfordin yliopiston matemaatikko taas julkisti toisen matemaatikon ehdottaman todistuksen kuuluisaan lukuteorian ongelmaan ja korosti, että työ on kokonaan ihmisen tekemä ilman tekoälyä. Yhtäkään kolmesta todistuksesta ei ole vielä vahvistettu. Ensimmäinen merkki on työn ilmestyminen esijulkaisuarkistoon, jonne tutkijat lataavat töitään ennen vertaisarviointia. Ratkaiseva on muiden tutkijoiden ja tarkistusohjelmien arvio.

Avustajat, jotka heräävät itse. Avustajien liitäntästandardiin lisätty tapahtumalaajennus on valinnainen. Sen vaikutus näkyy vasta, kun työkalujen tekijät ottavat sen käyttöön. Sunnuntaina standardin dokumentaation tukitaulukkoon lisättiin kaksi uutta ohjelmaa, jotka tukevat sen laajennuksia. Maanantaina Google ja Nvidia julkaisivat työkaluja yhteiseen standardiin, jolla avustajalle opetetaan tietty tehtävä, esimerkiksi yrityksen oman raporttipohjan käyttö. Kehityksen tunnistaa siitä, että avustajien asetuksiin ilmestyy käynnistimiä, joilla avustaja kytketään toisen palvelun tapahtumiin. Siinä vaiheessa olennainen kysymys on, mitä avustaja saa tehdä silloin, kun kukaan ei katso.

Anthropicin ensimmäinen osavuosikatsaus. Pörssiyhtiönä Anthropic julkaisee jatkossa säännöllisesti osavuosikatsauksia. Epoch AI:n mukaan yhtiön laskentakulut ovat vuositasolla lähes 40 miljardia dollaria ja nelinkertaistuvat vuosittain. Ensimmäisestä katsauksesta näkee, paljonko tuloja kulujen rinnalla on. Luvuista voi arvioida, kestääkö huippumallien nykyinen hintataso, jolle OpenAI ja Anthropic laskivat hintojaan samassa tahdissa.

Sähköverkko ja laskennan sijainti. Epoch AI:n mukaan kyky ajaa tekoälyagentteja kasvoi vuodessa 4,6-kertaiseksi, kun sirujen raaka laskentateho kasvoi samassa ajassa 2,6-kertaiseksi. Samasta laitteistosta saadaan siis agenttityöhön selvästi aiempaa enemmän irti. Laskennan sijainnin ratkaisee silti yhä useammin sähköverkko. Pohjoismaiden seurattavat merkit ovat Norjan liittämistauon kesto ja Nokian investointipäätös Oulun laskentakeskuksesta.

Kuka työn teki. Laajasti käytetyn avoimen ohjelman, jolla kielimalleja ajetaan omalla koneella, kehittäjät merkitsevät muutoksiinsa, mikä tekoäly auttoi niiden teossa. Sunnuntaina julkaistuista muutoksista kolmessa oli merkintä Anthropicin mallin tai ohjelmointityökalun avusta ja lauantaina yhdessä muutoksessa apuna oli kehittäjän omalla koneella toiminut malli. Ohjelmistoalalla puhutaan samaan aikaan matemaattisesta todistamisesta keinona erottaa toimiva koodi tekoälyn tuottamasta massasta. Kehityksen tunnistaa siitä, että tarjouksissa ja sopimuksissa aletaan kysyä tai luvata, kuka työn teki ja millä se on tarkistettu.

Nimettömät mallit vertailulistalla. Contra Labsin ylläpitämälle vertailulistalle ilmestyi kaksi nimetöntä mallia, jotka sijoittuivat Anthropicin Claude-mallin edelle. Nimettöminä vertailuihin tuodut mallit ovat aiemmin olleet suurten yhtiöiden tulevia julkaisuja, joita testataan ennen julkistusta. Mallien tekijöitä ei ole kerrottu.

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

05 Suuret toimijat

Viikon suurin mallijulkistus tuli OpenAI:lta 17.11., jolloin GPT-6.2 Astra ja GPT-6.2 Sol julkaistiin yhtä aikaa ChatGPT Workiin, Codexiin ja rajapintaan tunnisteilla `gpt-6.2-astra` ja `gpt-6.2-sol`. Astra on yhtiön huippumalli vaikeimpiin projekteihin ja Sol yleismalli arkityöhön. Julkistuksen painavin luku löytyi hinnastosta. Astra maksaa 8 dollaria syötteen ja 40 dollaria tulosteen miljoonalta tokenilta, mikä on 80 % GPT-6 Astran hinnasta ja täsmälleen sama kuin Anthropicin Claude Fable 5.2:n hinta. Myös Fable 5.2 hinnoiteltiin 80 prosenttiin edeltäjästään. Edelliset sukupolvet GPT-6 Astra ja Fable 5 maksoivat nekin samat 10 dollaria syötteeltä ja 50 dollaria tulosteelta. Kaksi kärkitaloa on siis asettunut samaan hintapisteeseen kahdesti peräkkäin.

Agenttikäytössä listahintaa suurempi muutos on välimuistista luetun syötteen hinta, joka laski 1 dollarista 0,40 dollariin miljoonalta tokenilta. Agenttisilmukka lähettää saman järjestelmäkehotteen, työkalukuvaukset ja kasvavan historian jokaisella kutsulla uudelleen. Pitkän istunnon laskussa välimuistitaksa painaa siksi enemmän kuin syötteen listahinta. Solin hinta ei käy ilmi julkistuksen lainauksista. OpenAI:n oma painotus näkyi Codex CLI:n versiossa 0.171.0, joka lisäsi molemmat mallit valikoimaan ja teki Solista oletuksen. Latent Spacen AINews-kooste päätyi samaan arvioon Solista todennäköisenä päätyöjuhtana. Astraa kooste kutsuu tämän hetken kakkoseksi huippumallien joukossa ja kuvaa varhaisten testaajien kokemuksia siitä työparina ristiriitaisiksi.

Google DeepMind julkaisi Gemini 4 Neonin seuraavana päivänä. Latent Space sijoittaa mallin Gemini 4 -sarjan keskitasolle Opus- ja Sol-luokan haastajaksi, mikä on uutiskirjeen kehys eikä mittaustulos. Sarjan hinnoittelu on mallia itseään vaikeampi hahmottaa. Ennen Neonia sarjassa olivat Hydrogen, Argon ja Carbon. Argonilla ja Carbonilla on sama hinta ja Hydrogen maksaa nyt kolme neljäsosaa siitä. Liitännäisen llm-gemini kehittäjän mukaan Argonin ja Carbonin hinta kaksinkertaistuu 1. joulukuuta julkaisualennuksen päättyessä, minkä jälkeen Hydrogen maksaa neljäsosan niiden hinnasta. Pelkkä kaksinkertaistuminen veisi suhteen kolmeen kahdeksasosaan eikä neljäsosaan. Luvut kannattaa siksi tarkistaa Googlen hinnastosta ennen budjetointia.

Listahinta kertoo silti vain osan tehtävän hinnasta. Liitännäisen pistejulkaisu 0.38.1 lisäsi Hydrogenille ja Neonille xhigh-päättelytason, jota Carbon ei tue. Kehittäjän vertailussa Hydrogen käytti high-tasolla 5 311 tulostetokenia ja maksoi 3,99 senttiä. Xhigh-tasolla sama malli käytti 31 477 tokenia ja maksoi 23,61 senttiä eli lähes kuusinkertaisesti. Tulokseksi saatu testikehotteen pelikaanipiirros kesti kehittäjän mukaan vertailun Neonin kanssa. Tehtäväkohtainen hinta määräytyy näin yhtä paljon päättelytasosta kuin hinnastosta.

Anthropicin viikkoa hallitsi listautuminen. Osake listautui tunnuksella ANT ja Latent Spacen mukaan anti hinnoiteltiin 83 dollariin osakkeelta. Uutiskirjeen otsikko kutsuu sitä historian suurimmaksi listautumiseksi 1,5 biljoonan dollarin arvostuksella. Ennen listautumista Business Wiren kautta jaettu tutkijoiden, oikeusoppineiden ja teknologiajohtajien kannanotto "Anthropic Is Not a Safe AI Company" kyseenalaisti yhtiön turvallisuusbrändin ja nousi Hacker Newsissä 148 pisteeseen. Epoch AI:n datakatsaus antaa mittakaavan sille, mihin rahaa menee. Anthropic käytti vuoden 2026 kolmannella neljänneksellä laskentaan ja infrastruktuuriin lähes 10 miljardia dollaria eli vuositasolla lähes 40 miljardia. Vuonna 2023 summa oli 433 miljoonaa dollaria. Epochin mukaan Anthropicin laskentakulut nelinkertaistuvat vuodessa ja OpenAI:n kolminkertaistuvat.

Rajapintaan tuli samalla viikolla kolme muutosta. Tekstivesileima on tuetuilla alustoilla Opus 4.6:n, Opus 4.5:n, Sonnet 4.5:n ja Haiku 4.5:n tuottamassa tekstissä. Merkki voi puuttua, jos pyyntö asettaa temperaturen, top_p:n tai top_k:n oletuksesta poikkeavaksi. Riippuvuus näytteenottoparametreista viittaa tokenivalintojen tilastolliseen ohjaukseen eikä metatietoon upotettuun merkkiin. Tunnistuksen kannalta seuraus on suora: puuttuva vesileima ei todista tekstiä ihmisen kirjoittamaksi. Betana tullut Linked billing (otsake `user-profiles-billing-2026-10-01`) antaa sovelluksen asiakkaan maksaa Claude-käytön oman organisaationsa kautta. Kehittäjä valitsee maksajan pyyntökohtaisesti `billing_policy`-kentällä sekä `/v1/messages`- että `/v1/chat/completions`-kutsuissa. Asiakas voi asettaa Claude Consolessa sovellukselle kulukaton tai purkaa linkityksen. Ominaisuus toimii vain organisaatioille, jotka maksavat API-käytöstä suoraan Anthropicille. Organisaation webhookeihin (beta) tuli lisäksi tapahtumat `external_key.failed` ja `external_key.recovered`. Ensimmäinen lähtee, kun avainpalvelu toistuvasti hylkää asiakkaan hallinnoiman salausavaimen. Jälkimmäinen kertoo avaimen toimivan taas.

Mallijulkistuksen ympärillä OpenAI:n viikko koostui pitkälti kuvauksista, joissa yhtiö kertoo omasta työstään omin sanoin. Piirisuunnittelukirjoituksen mukaan GPT-6 Astra optimoi Jalapeño-piirin virtapiirejä ja paransi RISC-V-ytimien suunnitelmia. Tiivistelmässä ei ole lukuja parannusten suuruudesta. Kirjoitus keräsi Hacker Newsissä 218 pistettä ja 165 kommenttia. Research acceleration: The jagged frontier mittaa, kuinka paljon koodausagentit nopeuttavat yhtiön omaa tekoälytutkimusta ja missä suoritus on yhä epätasaista. Rosoisen rajan käsite on peräisin Dell'Acquan ja kollegoiden kenttäkokeesta Navigating the Jagged Technological Frontier (Harvard Business School, 2023), jossa konsulttien työ parani mallin kyvykkyyden sisäpuolelle osuvissa tehtävissä ja heikkeni sen ulkopuolella. Managing frontier AI risk jakaa riskienhallinnan neljään kerrokseen, jotka ovat linjaus, hiekkalaatikkoeristys, estävät suojaukset ja valvonta. Adversariaalista tislausta käsittelevä kirjoitus yhdistää torjunnassa mallikoulutuksen, suojaukset, havainnoinnin ja sanktiot. Mallikoulutuksen mukanaolo viittaa siihen, että suojaus rakennetaan osin itse mallin käytökseen. Mitä koulutustason keinot käytännössä ovat, tiivistelmä ei kerro.

Ulkopuolista näkökulmaa tuovat upotetut arvioinnit, joiden ensimmäisessä vaiheessa Apollo Research, METR, Transluce ja Goodfire arvioivat OpenAI:n mallien linjausta, suojauksia ja valvontajärjestelmiä. Kokoonpano kattaa juonittelun ja harhautuksen mittaamisen, autonomisten kykyjen arvioinnin, tulkinta- ja valvontatyökalut sekä mekanistisen tulkittavuuden. Pääsyn laajuutta ja sitä, mitä tuloksista julkaistaan, kuvaus ei kerro. ChatGPT for Healthcare sai computer-use-agentit ja terveydenhuollon ohjelmistojen yli toimivat liitännäiset ilman arviointilukuja, tuettujen järjestelmien luetteloa tai tietoa EU-saatavuudesta. Avoimeksi jää myös, miten agentin toimet kirjataan auditointia varten käyttöliittymissä, joilla ei ole rajapintaa.

Matematiikasta tuli viikon mittaan väitteiden koealusta. OpenAI:n uuden tiedeblogin avausartikkeli esittelee ehdokkaan standardimallin ei-perturbatiiviseksi hilamääritelmäksi. Yhtiön mukaan GPT-6 Astra auttoi viemään idean yksityiskohtaiseksi todistukseksi. Kysymys on vanha: Nielsenin ja Ninomiyan no-go-teoreema (Nuclear Physics B, 1981) sulkee pois yksinkertaisimmat tavat sijoittaa kiraaliset fermionit hilalle ilman fermionien kahdentumista. Otsikko puhuu varovasti askeleesta ja ehdokkaasta. Ulkopuolista vertaisarviota ei ole raportoitu.

Loppuviikosta LLM-työkalujen kehittäjän blogi lainasi matemaatikkoa, joka väittää todistaneensa Clauden kanssa Riemannin hypoteesin ja yleistetyn Riemannin hypoteesin. Hänen mukaansa argumentit on formalisoitu ilman oletuksia. Lainauksessa ei mainita todistusavustajaa, julkaistua artefaktia eikä riippumatonta tarkistusta. Pian Latent Spacen AINews otsikoi Riemannin hypoteesin tulleen todistetuksi kahdesti toisistaan riippumatta. Samaan aikaan Stanfordin matemaatikko ilmoitti kiireellisesti, että Alessandro Masullon ehdottama todistus yleistetylle Riemannin hypoteesille (luonnos, 2026) on kokonaan ihmisen työtä ilman tekoälyä. Ilmoituksen mukaan toinen matemaatikko on käynyt todistusta läpi kuukausia ja luonnos julkaistiin suunniteltua aiemmin. Otsikko ja ilmoitus eivät puhu täsmälleen samasta väitteestä, koska toinen koskee hypoteesia ja toinen sen yleistettyä muotoa. Yhtäkään väitetyistä todistuksista ei ole vahvistettu. Formalisointiväite on silti niistä tarkistettavin. Julkaistun koneellisen todistuksen voi ajaa todistusavustajassa, kun perinteinen käsikirjoitus vaatii pitkän vertaisarvioinnin.

Muilta suurilta toimijoilta viikko toi julkistuksia, joista tiedetään toistaiseksi lähinnä otsikot. Microsoft julkaisi Ignite 2026 -tapahtumassa seitsemän uutta mallia, joiden tiedot nojaavat Microsoftin omaan julkistukseen. Julkistus keräsi Hacker Newsissä 223 pistettä ja 149 kommenttia. Google DeepMindin Genie 4 -maailmamalli tuli saataville ja The Batchin mukaan World Labs avasi Atlasin kehittäjille. Kummastakaan lähteissä ei ole teknisiä tietoja, hinnoittelua tai saatavuusalueita. Google Research kuvasi, miten vuonna 2018 esitelty Duplex-puheagentti soittaa nyt Geminissä yrityksille käyttäjän puolesta. Latent Spacen koosteen mukaan Perplexityn liikevaihdon vuositahti saavutti 1 miljardin dollarin ja nelinkertaistui vuoden 2026 aikana. Cloudflare kutsuu itseään neolabiksi ja rakentaa System One -malleja, joiden tehtävä on tehdä päätöksiä ja hakea dataa millisekunneissa. Nimi viittaa Daniel Kahnemanin teoksessa Thinking, Fast and Slow (2011) kuvattuun nopeaan ajatteluun.

06 Avoimet mallit ja ekosysteemi

Thinking Machinesin Inkling 2 Mini oli viikon merkittävin avoin julkaisu. Mallissa on 26 miljardia parametria, joista aktiivisena on 5,4 miljardia. Arkkitehtuurin erikoisuus ovat silmukoidut kerrokset: samoja kerroksia ajetaan useaan kertaan, jolloin laskentaa on tokenia kohden enemmän kuin parametrimäärä antaa olettaa. Toteutuksen llama.cpp:hen

tuoneen muutoksen (#32694) historiassa esiintyvät loop-graafi ja attribuutti is_varloop, joka nimettiin myöhemmin muotoon varying. Nimistä päätellen silmukoinnin määrä voi vaihdella. Malli käsittelee tekstiä ja kuvia sekä tukee työkalukutsuja ja ajattelutilaa. Ollaman mukaan se mahtuu Macilla noin 20 gigatavuun. Ollaman julkaisutiedotteessa mallin SWE-Bench Verified -tulokseksi ilmoitetaan 79,2 %, kun Artificial Analysisin Intelligence Index antaa sille 20 pistettä. Lisenssiä julkaisuotteet eivät mainitse.

Ekosysteemi otti mallin vastaan vuorokaudessa. Hugging Face julkaisi esittelyblogin 20.11. ja Unslothin GGUF-kvantisoinnit ilmestyivät samana päivänä. Tuki tuli llama.cpp:hen buildissa b12600 saman päivän iltana ja Ollamaan versiossa 0.40.17 seuraavana aamuna. Lataustilastot kertovat, miten mallia käytännössä ajetaan. Alkuperäisellä repolla on 1 883 tykkäystä ja 7 944 latausta. Unslothin GGUF-versiota on ladattu 238 152 kertaa eli noin 30-kertaisesti. Sunnuntain trendilistalla oli lisäksi AtomicChatin 20.11. illalla julkaisema nimensä mukaan sensuroimaton GGUF-johdannainen, jota oli ladattu jo 129 693 kertaa.

Sama kuvio näkyi viikon alussa Alibaban Qwen-Image 2.1 -kuvamallin kohdalla. Sen kolmannen osapuolen kvantisoitu johdannainen, josta on riisuttu sisältörajoitukset, keräsi 160 748 latausta, kun listan seuraavaksi ladatuin malli jäi noin 3 000 lataukseen. Trendilistan latausvolyymi kasautuu siis kolmansien osapuolten kvantisoituihin ja rajoituksista riisuttuihin johdannaisiin. Niiden lisenssi ja tekijän tausta kannattaa tarkistaa ennen tuotantokäyttöä. Toisessa ääripäässä ovat mallit, joilla on tykkäyksiä ilman latauksia: Applen tekstistä videota tuottava FastFrames-L (188 tykkäystä), Applen syvyysestimointimalli DepthPro2 (254), Inworldin voice-isolator (210) ja Sentinel7/h3 (193), jolla ei ole edes pipeline-tunnistetta. Tykkäys ei mittaa käyttöä. Nvidian Nemotron-3.5-Bolt-E4B-4GB on päinvastainen tapaus. Lokakuun alussa päivätty malli on kerännyt 172 088 latausta, mikä kertoo käytöstä eikä julkaisukohusta. E4B on muissa malliperheissä merkinnyt efektiivistä noin neljän miljardin parametrin kokoa ja pääte 4GB viittaa muistijalanjälkeen, joka mahtuu kuluttajanäytönohjaimeen.

Suurten mallien ajamiseen tuli kaksi avausta Hugging Facelta. Sen ja OpenRouterin yhteistyön otsikko lupaa tehdä open-weight-mallien ajamisesta helppoa. Molemmilla on jo oma reitityskerroksensa: Hugging Facen Inference Providers ohjaa Hubin mallien kutsuja ulkoisille tarjoajille ja OpenRouter kokoaa kymmeniä tarjoajia OpenAI-yhteensopivan rajapinnan taakse yhteen laskutukseen. Avoimiksi jäävät tarjoajan valinta, hinnoittelu ja mahdollisuus rajata kutsut EU:ssa toimiviin tarjoajiin. Toinen kirjoitus ajaa 180 miljardin parametrin asiantuntijaseosmallia Transformers-kirjastolla 32 gigatavun RAM-muistilla. Osoitteen perusteella tekniikka on asiantuntijoiden striimaus, jossa muistissa pidetään vain kulloinkin reititetyt asiantuntijat. Käyttökelpoisuuden ratkaiseva token-nopeus ei käy ilmi otsikosta.

Pienemmistä erikoismalleista Hugging Facen blogissa esitelty Rumble-2-Flash on avoin 0,5 miljardin parametrin puhemalli yli 40 kielelle. Otsikosta eivät selviä lisenssi eikä se, onko kyse puheentunnistuksesta vai -synteesistä. Show HN -ketjussa 289 pistettä kerännyt dettato litteroi ääntä MIDI-muotoon. Tekijän mukaan avoimin painoin julkaistu malli voittaa 1,3 miljardin parametrin muscriptorin kaikilla testijoukoilla ja on noin 60 kertaa pienempi. Tekijä kertoo myös, ettei mallia ole koulutettu tekijänoikeuden alaisilla äänitteillä. Koulutusaineiston alkuperä on musiikkimallien keskeinen oikeudellinen kysymys. Molemmat väitteet ovat tekijän omia.

Vähemmälle huomiolle jäivät Resolve.ai:n tuotantohäiriöiden selvittämiseen tarkoitettu SRE-1 ja RadixArkin DaisyRose-27B, jonka otsikko lupaa jälkikoulutuksen tehokkuuden uutta kärkeä. Molemmissa on 27 miljardia parametria ja väitteet ovat julkaisijoiden omia. Tutkimuspuolelta eniten huomiota sai arXiv-paperi 2611.21885, jonka otsikon mukaan kielimallit voivat kirjoittaa omat painonsa uudelleen (466 pistettä, 140 kommenttia). Asetelma liittyy jatkuvaan oppimiseen eli siihen, miten malli voisi päivittää tietonsa käytön aikana ilman erillistä uudelleenkoulutusta. Ai2 toi Asta-alustaansa Theorizerin, joka kytkee tieteellisen kirjallisuuden ehdotettuihin selityksiin tutkijan tarkasteltaviksi ja testattaviksi. Taylor Geospatialin kanssa aloitettu kumppanuus lupaa paikkatietotekoälyyn avoimia malleja, yhteisiä vertailuja ja parempia satelliittikuva-aineistoja.

07 Ajoalustat ja infra

Viikon aikana llama.cpp:n buildinumero eteni b12431:stä b12634:ään. Muutosten massasta erottuu ryhmä korjauksia, joita yhdistää hiljaisuus: ohjelma ei kaadu vaan tuottaa väärää tulosta. PowerPC:n tinyBLAS_Q0-ydin salli tiilitetyn

polun leveydellä `nc = 4`, vaikka ydin vaatii kahdeksalla jaollisen leveyden. Seurauksena oli virheellisiä matriisikertolaskuja, jotka näkyivät sotkuisena tekstinä esimerkiksi Flint Q4_K_M- ja Gemma 3 27B Instruct -malleilla ennen buildia b12608. CDNA3-piireillä NVFP4-skaalat puolitettiin kahteen kertaan ennen buildia b12632. WebGPU:n flash attention -varjostimet indeksoivat maskia ilman pääulottuvuutta, jolloin maskit, joilla `ne[2] > 1`, antoivat vääriä tuloksia ennen buildia b12441. Vulkanin top-k valitsi tasapelit säikeiden ajoituksen mukaan, kunnes b12435 teki valinnasta deterministisen samoin kuin CUDA:ssa ja Metalissa. Qwen3-Next-arkkitehtuurista puuttui output projectionin painoskaalaus buildiin b12474 asti. Qwen3-Next-malleja GGUF-muodossa ajavan kannattaa päivittää ja verrata tulosteita aiempiin.

Kaatuvat virheet sentään ilmoittavat itsestään. Mallit, joilla on vähän tensoreita kerrosta kohden, kuten cohere2moe (443 tensoria, 49 kerrosta), ylittivät graafikontekstin, kun MoE-välimuisti yhdistettiin kvantisoituun KV-välimuistiin K/V-rotaatiolla. Ajo pysähtyi `GGML_ASSERT(obj_new)`-tarkistukseen, koska MoE-välimuistin slottihaku jäi solmubudjetin ulkopuolelle ennen buildia b12569. Samaan aikaan arkkitehtuurikohtaisia listoja siirrettiin mallien luokkiin virtuaalimetodeiksi. Graafin solmubudjetti tulee nyt metodista `llama_model::graph_nodes_per_tensor()`, jonka oletus on kahdeksan solmua tensoria kohden (b12572).

Toinen jatkuva lanka on palvelin käännöskerroksena. Yksi paikallinen llama-server tarjoaa OpenAI:n Chat Completions- ja Responses-rajapinnat sekä Anthropicin Messages-rajapinnan. Mallien keskustelupohjat eivät silti noudata standardirajapintojen oletuksia. Viikon alussa b12431 alkoi palauttaa 400-virheen ja tuettujen tasojen luettelon, jos keskustelupohja ei hyväksy pyynnön `reasoning_effort`-tasoa. Aiemmin esimerkiksi taso `max` Qwen3.8:lla eteni pohjaan asti ja palasi 500-virheenä kaikista kolmesta rajapinnasta. Sunnuntaina b12624 jatkoi samaa työtä toisesta päästä välittämällä arvon `reasoning_effort = none`, jos malli tukee sitä. Saman yön b12634 korjasi työkalukutsut malleilta, jotka eivät käytä aloitusmerkkiä. Llama 3.x tuottaa kutsun pelkkänä `{"name": ...}`-JSONina, minkä vuoksi `tool_choice=auto` -tilassa kielioppia ei syntynyt lainkaan. Nyt laiska kielioppi laukeaa kutsuolion alkua vastaavasta kuviosta. NVIDIAn teknisen blogin kirjoitus työkalukutsujäsentimien rakentamisesta Rustilla agenttien avulla kuvaa samaa ongelmaa yleisemmin: kutsu voi saapua paloina, argumentti voi katketa ja mallikohtainen merkintä voi vuotaa vastaukseen.

Kolmas suunta vie ajoalustaa tekstin ulkopuolelle. Multimodaalinen mtmd-kerros sai tuen Minimax-Music3-musiikkimallille sekä drama3-mallin äänikloonauksen, jossa Fish S2 -koodekin enkooderi tuottaa referenssiklipin koodit. Äänikloonaus kulkee näin samassa C/C++-pinossa kuin tekstimallit ilman Python-ajoympäristöä. Kokoro-puhesynteesi sai G2P-pohjaansa sanakirjan ja Qwen3-ASR:n kielitunnistetuloste siirtyi ajatteluosioon. Metalille tuli konvoluutioytimiä, joista M5:n tensoriytimiä käyttävä `kernel_conv_2d_mm_tensor` (b12543) korjaa tilanteen, jossa uuden sirun tensoriytimet jäivät konvoluutiossa käyttämättä. Laitteistokirjo levisi samalla. OpenCL-taustaosa sai Adreno 850 -grafiikkapiirille cooperative matrix -GEMM-ytimet kuudelle kvanttityypille (Q4_0, Q5_0, Q8_0, Q4_K, Q5_K ja Q6_K) buildissa b12629 ja HIP-taustaosa RDNA1:lle oman MMVQ-parametritaulukon. Julkaisumatriisissa ovat nyt CUDA 13.5, ROCm 10.0 ja Ubuntu 26.04 s390x zDNN-kiihdytyksellä.

Ollaman 0.40.17 rakentui viikon mittaan kolmessa vaiheessa. Ensimmäinen julkaisuehdokas toi MLX-ajuriin spekulatiivisen dekoodauksen eräajetun verifioinnin, jonka ansiosta luonnostelu pysyy päällä myös rinnakkaisessa generoinnissa. Toinen ehdokas toi MLX-taustaosaan rinnakkaisen generoinnin, jolloin Apple Silicon -kone voi palvella useaa pyyntöä yhtä aikaa. Lopullinen versio toi Inkling 2 Minin MLX-ajurin kautta, minkä vuoksi tuki rajautuu toistaiseksi Apple siliconiin ja PC-tuki on luvattu myöhemmäksi. Ollaman toinen moottori kulkee jäljessä. Toisen ehdokkaan mukana tullut llama.cpp oli buildissa b12257, kun upstream oli perjantaina jo buildissa b12590. Ero on runsaat 300 buildia. Viikon MoE-korjaukset ja äänikloonaus eivät siksi vielä kulje Ollaman mukana. Paikallisia malleja Claude Coden kanssa ajaville versio toi viivästetyn työkalujen latauksen: `ollama launch claude` lataa MCP-työkalut tarpeen mukaan eikä kerralla, mikä jättää paikallisen mallin rajalliseen kontekstiin enemmän tilaa varsinaiselle työlle.

Agenttiautomaation kehyksissä korjaukset osuivat samaan kohtaan eli tilaan, joka kulkee rinnakkaisten työkalukutsujen välillä. Automaatioalusta n8n:n AI Agent -solmu säilyttää versioista 2.49.0 ja 3.3.0 alkaen mallin työkalukutsuja edeltävän tekstin, pitää jokaisen syötealkion työkalukutsuhistorian erillään ja tallentaa vuoron kaikki työkalukierrokset keskustelumuistiin. Sama erottelu tuotiin ylläpitolinjaan versiossa 2.48.5. Aiemmissa versioissa monivaiheisen agentin muisti saattoi siis jäädä vajaaksi ja usean alkion ajossa historiat sekoittua keskenään. Samassa versiossa muuttuivat lausekehiekkalaatikko ja tunnistetietojen käyttöoikeustarkistukset, joita ei ole

merkitty tietoturvakorjauksiksi. Torstaiaamuna n8n julkaisi 12 minuutin sisällä kolme korjausversiota kolmeen haaraan ilman muutoskuvauksia ja perjantaina kolme lisää. Ylläpitolinja 1.x eteni viikossa versiosta 1.123.97 versioon 1.123.100. LangGraph 1.2.23 korjaa saman ongelmaluokan toisessa kehyksessä. Send-komennolla jaetut työkalukutsut saavat nyt saman tilan kuin reunan kautta kulkevat ja `wrap_tool_call`-kääreen tilaohitukset päätyvät työkalulle asti. Versio 1.2.24 korjaa `REMOVE_ALL_MESSAGES`-merkinnän käsittelyn sekä `DeltaChannel`in None-tarkistuspisteet.

Open WebUI v0.13.4:n Tether-sovellus liittää käyttäjän oman koneen instanssiin. Koneen Ollama-, LM Studio- ja OpenAI-yhteensopivat mallit sekä MCP- ja OpenAPI-työkalupalvelimet näkyvät vain kyseiselle käyttäjälle. Osoitteet, komennot ja avaimet pysyvät koneella. Työkalukutsut kulkevat kuitenkin yhdistetyn instanssin kautta, mikä tekee instanssin ylläpidosta osan luottamusketjua. Kapasiteetin mittakaavasta kertoo Epoch AI:n arvio, jonka mukaan agenttisen päättelyn kapasiteetti kasvoi vuodessa 4,6-kertaiseksi ja kahdeksanbittinen laskentakapasiteetti 2,6-kertaiseksi. NVIDIAn teknisessä blogissa käsiteltiin samaan aikaan harvan huomion yhteissuunnittelua pitkän kontekstin päättelyyn sekä Dynamo-alustaa vahvistusoppimisen rollout-pullonkaulojen purkamiseen.

08 Benchmarkit ja vertailut

Artificial Analysisin Intelligence Indexissä viikon kärjessä oli GPT-6.2 Astra (Max) 62 pisteellä. Sol (Max) sai 57, Gemini 4 Neon (Low) 44 ja Alibaban Qwen4 Max 43. Neonin kohdalla luku riippuu asetuksesta poikkeuksellisen paljon, sillä Latent Space raportoi mallille 61 pistettä kertomatta päättelytasoa. Väli on 17 pistettä. Hinta-laatuvertailu kannattaa siksi tehdä samalla asetuksella, jolla mallia aiotaan ajaa. Pienemmistä malleista arvion saivat JoyAI-VL-Preview (28), Inkling 2 Mini (20), K2 Horizon 32B (14) sekä nimen perusteella Liquid AI:n laitteessa ajettavaan sarjaan kuuluvat LFM3-1.7B (6) ja LFM3-VL-2B (2). Inkling 2 Minin indeksiluku ja julkaisuviestinnän 79,2 % SWE-Bench Verifiedissä mittaavat eri asioita. Ensimmäinen on usean testin riippumaton yhdistelmä ja jälkimmäinen yksittäinen ohjelmointitesti. Astra for Law 6.2 sijoittui OpenAI:n mukaan ensimmäiseksi 24 oikeudenalan julkistusvertailussa, joka on OpenAI:n oma. Artificial Analysis lisäsi mallin arvioituihin ilmoittamatta indeksilukua.

Indeksien vastapainoksi Epoch AI testasi Claude Fable 5.2:ta, GPT-6.2 Astraa ja neljää muuta mallia omilla oikeilla työtehtävillään. Yksikään ei täyttänyt vaatimustasoa, jonka Epoch asettaa omille työntekijöilleen. Rajan alle jäivät myös ne kaksi mallia, jotka johtavat Epochin omaa ECI-kyvykkyysindeksiä. Tehtävät on otettu organisaation omasta työstä eikä rakennettu mittaamista varten. Hacker Newsissä kiersi samalla kaksi epävirallista listaa. Contra Labsin frontierdesign-listalle ilmestyi kaksi nimetöntä stealth-mallia Claude Fable 5.2:n edelle (94 pistettä). Henkilökohtaisella GitHub Pages -sivustolla julkaistu Models Engineering Index nosti kärkeen Gemini 4 Neonin ja keräsi 372 pistettä ja 244 kommenttia. Kummankaan menetelmä ei käy ilmi nostoista.

Vertailun yksikkö siirtyi viikon aikana mallista järjestelmään. Runtan FrontierHarness Eval v1.1 (219 pistettä) kysyy, mikä agenttiympäristö eli harness toimii parhaiten kunkin mallin kanssa, koska sama malli tuottaa eri ympäristöissä eri tuloksia. Hugging Facen trendilistan tykätyimpien kärkeen nousi yhteisön harness-hub/Qwen3.8-27B-pi-adapted. Julkaisijan nimi viittaa kehyssovituksiin, vaikka muutokset eivät käy ilmi listauksesta. Loppuviikosta Hugging Face esitteli harness adapters -konseptin, joka kääntää tavallisen järjestyksen: kehyksen virittämisen sijaan malli sovitetaan kehykseen suorituskyvyn ja kustannusten parantamiseksi. LMArena laajensi areenansa agentteihin Agent Arena Indexillä, jonka menetelmä ei käy ilmi julkistuksen otsikosta.

Sama siirtymä näkyy MLCommonsin vertailuissa. MLPerf Endpoints v1.0 mittaa eläviä API-päätepisteitä tuotantoa vastaavalla kuormalla, jossa keskusteluhistoria kasvaa vuorosta toiseen ja tuotos sekoittaa päättelyä ja työkalukutsuja. MLCommonsin perustelu on, että koko Pareto-rintama kertoo enemmän kuin yksittäinen huippuarvo, koska sama palvelu voi olla nopea yhdelle käyttäjälle ja hidastua jyrkästi kuorman kasvaessa. Nvidia kertoo johtavansa vertailua Vera Rubin NVL72:lla. Ensimmäisellä kierroksella aiempaa vertailukohtaa ei ole. MLPerf Training v6.1 -kierrokseen tulivat agenttisen vahvistusoppimisen ja generatiivisen suosittelun testit. Suurimmat klusterit olivat 10 240 GPU:ta ja 8 192 TPU:ta. Monisolmuisia kokoonpanoja oli 76 % lähetyksistä. MLCommonsin mukaan suurimmat suorituskykyhypyt

tulivat kohdennetuista ohjelmisto-optimoinneista ja rinnakkaistusratkaisuista. Nvidian väite Vera Rubin NVL72:n jopa kaksinkertaisesta suorituskyvystä Blackwell Ultraan nähden koskee parasta tapausta valituissa testeissä. Testikohtaiset luvut ja muiden toimittajien tulokset ovat MLCommonsin tulostaulukoissa.

Mittaajat itse kirjoittivat viikolla mittaamisen rajoista. METR julkaisi muistion alignment-arviointien telleistä eli vihjeistä, joista malli voi päätellä olevansa testattavana ja käyttäytyä testissä toisin kuin käytössä. METR:n vuoden viimeisen neljänneksen kykykatsauksen otsikon mukaan tekoälyn tutkimus- ja kehitystyön kannalta olennaiset kyvyt paranevat nopeasti ja käyvät yhä vaikeammiksi mitata.

METR:n tunnetuin mittari on tehtävien aikahorisontti, joka kertoo, kuinka pitkistä ihmisen työtehtävistä malli suoriutuu tietyllä onnistumistodennäköisyydellä (Kwa ym., 2025, Measuring AI Ability to Complete Long Tasks). Sunnuntaina OpenAI julkaisi kirjoituksen, jonka mukaan perpleksiteetti voisi auttaa kehittäjiä löytämään yhteisen tavan arvioida uusia kykyjä. Nimi viittaa kielimallin ennustustarkkuutta kuvaavaan mittariin eikä samannimiseen hakuyhtiöön. Samalla viikolla OpenAI julkaisi varhaisen lääkekehityksen tehtäviä mittaavan MedChemBenchin, joka kattaa molekyylirakenteiden tulkinnan, päättelyn kokeellisesta näytöstä ja yhdisteiden valinnan. Testin on rakentanut mallintoimittaja itse.

09 Työskentelytavat ja agentit

Kolmen suuren agenttityökalun muutoslokit kertoivat viikolla samaa: agenttityö venyy pitkäkestoiseksi, rinnakkaiseksi ja taustalla jatkuvaksi. Työkalut rakentavat sitä varten hallintanäkymiä. Codex CLI 0.171.0 teki TUI-työtilasta oletuksen. Pääkeskustelun rinnalla näkyvät haettavat dokumenttiesikatselut, kuvien esikatselut ja aliagenttien transkriptit, jolloin aliagentin työtä voi seurata eikä vain lukea sen palauttamaa lopputulosta. Komento `codex exec --remote` käynnistää tai jatkaa ei-interaktiivisen istunnon olemassa olevalla sovelluspalvelimella. Claude Code 2.1.329 kokoaa `/tasks`-näkymään istunnon ajastetut tehtävät, `/loop`-heräämiset, tavoitteen ja keskustelun haarat. Minkä tahansa niistä voi pysäyttää x-näppäimellä vahvistuksen jälkeen. Gemini CLI:n `/stats heatmap` näyttää päivittäisen tokenikulutuksen.

Kulutuksen näkyvyys kiinnosti muuallakin. Microsoft Researchin TokenEconomizer lähtee siitä, että jokainen istunto lähettää mallille ohjeet, työkalukuvaukset ja historian, vaikka kehittäjät palaavat toistuvasti saman työn pariin. Menetelmä oppii aiemmista istunnoista keventääkseen seuraavia. Säästön suuruutta tiivistelmä ei kerro. Nvidian SkillEvaluator ajaa saman tehtävän sekä Agent Skills -taidon kanssa että ilman sitä ja mittaa, muuttaako taito lopputulosta. Se on kokeilun arvoinen tiimeille, jotka ylläpitävät useita taitoja.

Claude Coden viikon viisi versiota (2.1.326–2.1.330) tekivät näkyviksi rajoituksia, jotka olivat aiemmin olleet hiljaisia. Käynnistys varoittaa nyt `--allowedTools`-, `--disallowedTools`- tai `--tools`-arvosta, joka ei näytä täydeltä säännöltä. Väärin muotoiltu kieltosääntö voi muuten jättää työkalun estämättä ilman, että kukaan huomaa. Näkymä `/status` listaa rikkinäiset agentti-, skill-, komento- ja output style -tiedostot, saavuttamattoman hiekkalaatikon, ohitetut käyttöoikeusasetukset ja hook-matcherit, jotka eivät koskaan osu. Sama näkymä ja `claude doctor` huomauttavat OpenTelemetry-viejästä, jolta puuttuu päätepiste ja jonka data todennäköisesti katoaisi. Ei-interaktiivisissa isännissä, kuten Claude Desktopissa, keskusteluun tulee varoitus, jos hallittujen asetusten hookit eivät lataudu. MCP-työkaluja kutsuvat hookit saivat valinnan `"onFailure": "block"`, jolloin epäonnistuva hook estää toiminnon. Komento `claude mcp add-json <name> -` lukee palvelimen JSON-määrittelyn vakiosyötteestä, jolloin tokenit eivät päädy komentoriville.

Kolmas lanka koskee sitä, missä agentin rajat asetetaan. Clinen CLI 3.0.75 käynnistää kertaluontoisen ajon, jossa kehote annetaan argumenttina tai putkitetaan stdiniin, yolo-tilassa eikä act-tilassa, koska kukaan ei ole hyväksymässä työkalukutsuja. Act-tila säilyy, jos automaattinen hyväksyntä on kytketty pois, tilalippu annetaan, plan-tila on tallennettu tai kehote on /team-kehote. Tiimityökaluja tarvitseva ajo vaatii --act-lipun, koska yolo-tila ei sisällä niitä. Lisäksi --timeout ja Ctrl+C pysäyttävät nyt myös ensimmäisen vuoron. Skriptissä tai CI-putkessa hyväksyntädialogi ei enää rajaa mitään. Rajat asettavat kontti, tiedostojärjestelmän oikeudet ja tunnukset. Clinen työpöytäsovellus ja laajennus vaihtoivat samalla oletusmallin: Cline-palveluntarjoaja käynnistyy ensimmäisellä ilmaisella mallilla eikä Claude Sonnet 5.5:llä, jos tallennettua valintaa ei ole.

Claude Code 2.1.328 toi auto-tilan Opus 4.6:lle ja Sonnet 4.6:lle Amazon Bedrockissa, Google Cloudin Agent

Platformissa, Microsoft Foundryssa ja Claude apps gatewayssa. Interaktiiviset istunnot, joille ei ole asetettu permission-tilaa, käynnistyvät nyt siinä. Näiden alustojen kautta Claude Codea ajavan organisaation kannattaa siksi määrittää tila hallituissa asetuksissa eksplisiittisesti. Versio 2.1.330 toi kirjautumisen Enterprise Admin Consolessa hallinnoituihin organisaatioihin, jolloin reititys pilvipalveluntarjoajalle tai yhdyskäytävälle määräytyy keskitetysti eikä jokaisen kehittäjän ympäristömuuttujista.

Hiekkalaatikoista esitettiin viikolla kaksi mallia. OpenAI:n kirjoitus Codexin hiekkalaatikosta kuvaa, miten agentti ajaa komennot käyttäjän koneella ja valvoo käyttäjän määrittämiä tiedostojärjestelmä- ja verkkorajoja macOS:llä, Linuxilla ja Windowsilla. LLM-työkalun tekijän ja Prime Radiant -laboratorion ulkoisen hiekkalaatikon mallissa agentti toimii hiekkalaatikon ulkopuolella ja sen ajama koodi erillisessä, tähän varatussa hiekkalaatikossa. Lähtökohtana on paikallisten mallien kokeilu ilman pelkoa siitä, että virhe tai kehoteinjektio johtaa vahingollisiin toimiin. Hacker Newsissä 126 pistettä kerännyt kirjoitus kuvasi tavan ajaa ohjelmointiagentteja antamatta niille kehittäjän omia tunnuksia. Kaupallisessa päässä StarSling (YC P25) antaa jokaiselle CI-työlle oman lyhytikäisen hiekkalaatikon Zen 5 -koneella NVMe-levyineen. Rinnakkaisuutta ei rajoiteta, jotta agenttien PR-ryöppy ajetaan kerralla.

Viikon eniten keskusteltu työtapakysymys ei koskenut tuotantoa vaan tarkistamista. Claude Code -tiimin kirjoitus verifioinnista eli agentin työn tarkistamisesta keräsi 295 pistettä ja 233 kommenttia. Harvardin tutkijoiden havainto, jonka mukaan koodausagentit siirsivät pullonkaulan koodikatselmointiin, keräsi 327 pistettä ja 462 kommenttia. Ask HN -ketju kysyi, miksi tuottavuuslupaus ei näy tyhjentyneinä virhelistoina. Curlin pääkehittäjän laskelma antoi kysymykselle mittakaavan. Sinä aikana, kun agentit käänsivät curlin 185 000 riviä C-koodia 340 000 riviksi Rustia, projektiin yhdistettiin runsaassa kolmessa viikossa yli 300 committia. Niistä runsaat 200 oli virhekorjauksia ja lisäksi korjattiin muutama haavoittuvuus. Seuraavan kuukauden aikana virhekorjauksia kertyi toiset 200. Yksikään Rust-klooni ei sisällä niitä.

Viikonloppuna nostettu esimerkki näyttää Rust-työn toisesta päästä. Kehittäjä, joka aloitti mod_wsgi:n 20 vuotta sitten, rakentaa tekoälyavusteisesti Rustilla Muru-palvelinta ASGI- ja WSGI-sovellusten ajamiseen kehityksessä ja tuotannossa. Hän julkaisee työstä poikkeuksellisen yksityiskohtaista selostusta osana kehitysprosessiaan. Koko selostus on syntynyt viimeisen kolmen viikon aikana. Datasetten kehittäjä noudattaa omaa kurinalaisuuttaan pienemmässä mittakaavassa: uutta liitännäisominaisuutta ei julkaista ilman liitännäistä, joka käyttää sitä. Datasette 1.0a43:n striimaavien vientimuotojen rajapinta todennettiin rakentamalla sen päälle datasette-arrows, jonka Apache Arrow -muodon ehdotti ideakumppanina toiminut Claude Fable 5.2. Lisäosa datasette-maplibre syntyi pääosin yhdellä kehotteella, joka ei määritellyt toteutusta vaan ohjasi mallin lukemaan kaksi olemassa olevaa lisäosaa ja paikallisen OpenFreeMap-demon. Määrittely korvautui toimivilla viitetoteutuksilla. Latent Spacen katsaus esittää Leanin ja tekoälyn yhdistelmän vastavoimana slopille eli määrältään paisuvalle ja laadultaan epävarmalle generoidulle sisällölle. Ohjelman oikeellisuuden todistaminen on perinteisesti vaatinut erikoisosaajia. Tekoälyn automatisoidessa siitä osan todistamisen hinta suhteessa testaamiseen muuttuu.

Avustuksen kirjaaminen vakiintui samalla viikolla arkiseksi. Viikon llama.cpp-muutosten Assisted-by-riveillä esiintyivät Codex, Grok Bot, Claude Code ja Claude Opus 5.5. Muodossa "Claude (Opus 5.5 Medium)" rivi kertoi myös päättelytason. Kahdessa muutoksessa avustajana oli DeepSeek-V4.1-Flash GGUF-muodossa kehittäjän Mac Studiolla (M5 Ultra, 256 Gt), jolloin ajonaikaa kehitettiin mallilla, jota se itse ajaa. MCP-referenssipalvelinten testimuutoksessa Claude Fable 5.2 ja MCP-määrittelyn dokumentaatiomuutoksessa Claude Opus 5.5 on merkitty toiseksi tekijäksi.

10 Liitännäiset, MCP ja repot

MCP-määrittelyn viikko kulki kahteen suuntaan: ydin karsiutui ja laajennukset saivat prosessin. Yleiskatsaussivut lakkasivat esittämästä samplingia, rootsia ja loggingia nykyisinä ominaisuuksina. Ne on merkitty vanhentuneiksi 2026-07-28-versiossa ja →

MCP-määrittelyn viikko kulki kahteen suuntaan: ydin karsiutui ja laajennukset saivat prosessin. Yleiskatsaussivut lakkasivat esittämästä samplingia, rootsia ja loggingia nykyisinä ominaisuuksina. Ne on merkitty vanhentuneiksi 2026-07-28-versiossa (SEP-2577), jonka niitä käyttävän toteutuksen ylläpitäjän kannattaa käydä läpi ennen uutta kehitystyötä. Laajennusten elinkaarta kuvaava SEP-3392 on

Final-tilassa. Kuka tahansa Maintainer voi perustaa `experimental-ext-`-etuliitteisen repon ja Core Maintainerit voivat kirjatulla päätöksellä ohittaa Beta-vaiheen. Samalla viikolla lopulliseksi eteni Events-laajennus (SEP-3415, `io.modelcontextprotocol/events`), jonka määrittely elää omassa ext-events-repossaan eikä päärepossa.

Laajennuksella MCP-asiakas tilaa ylävirran järjestelmän tapahtumia ja agentti reagoi niihin ilman käyttäjää. Esimerkkeinä SEP mainitsee Slack-viestin, repon pushin ja PagerDuty-häiriön. Palvelin listaa tapahtumatyypit `events/list`-kutsulla. Jokaisella tyypillä on nimi, tilauksen argumenttien `inputSchema`, `payloadSchema` ja tuetut toimitustavat. Toimitustapoja on kolme: kysely (`events/poll`), pitkäikäinen `events/stream`-yhteys ja webhook (`events/subscribe`). Mikään niistä ei ole pakollinen. Asiakkaan on siis tarkistettava, mitä palvelin tukee. MCP-palvelin muuttuu laajennuksen myötä työkalujen tarjoajasta myös agentin käynnistäjäksi.

Hallinnossa painopiste siirtyi työryhmiin. Rahoituspalvelujen kiinnostusryhmän (FSIG) peruskirjaan ehdotettiin muutosta, jonka mukaan ryhmä tunnistaa ongelmia, jakaa olemassa olevia toteutuksia ja ohjaa ratkaisemattomat tarpeet työryhmille oman ratkaisukehityksen sijaan. Prototyypit sallitaan vaatimusten validointiin. GitHub Discussions rajattiin kokousmuistioihin ja tiedotteisiin, joita voi avata vain kirjoitusoikeudella. Ideat, ominaisuuspyynnöt ja kysymykset ohjataan Discordiin, jonka kutsulinkki vaihtui sunnuntaina osoitteeseen discord.gg/modelcontextprotocol. MCP Inspector sai kuormitustestauksen, jota ajetaan web-asiakkaan Load test -paneelista, CLI:n `--load`-lipulla tai mcpdo:n `tools/call --load` -komennolla yhteisellä moottorilla. Uusi Active Sessions -näkymä näyttää taustan avoimet istunnot ja sallii niiden päättämisen. Laajennusten asiakasmatriisiin lisättiin sunnuntaina Archestra.AI:n Skills-tuki ja mcp-use, joka tukee sekä Appsia että Skillsiä.

Agent Skills -spesifikaatio näkyi viikon trendilistoilla usean toimittajan yhteisenä pakkausmuotona. Googlen stitch-skills kokoaa taitoja Stitch-suunnittelutyökalun MCP-palvelimelle ja listaa yhteensopiviksi Codexin, Antigravityn, Gemini CLI:n, Claude Coden, Cursorin ja OpenCoden. Muille kuin Stitchin käyttäjille se on lähinnä malli MCP-palvelimen ja taitojen paketoinnista yhdessä. Repo kajisho5/ffmpeg-skill antaa agentille paikallisen videoeditorin, jossa on 42 FFmpeg-työkalua, koneluettava sopimus ja MCP-palvelin. Työkalut kattavat leikkauksen, tekstityksen, rajauksen esimerkiksi 9:16-muotoon ja hiljaisten kohtien poiston. Agentti tarkistaa lopuksi oman tuotoksensa. Pilveä tai API-avaimia ei tarvita ja toteutus nojaa Pythonin standardikirjastoon. Repo on kokeilun arvoinen lyhytvideoita tekeville.

Kaksikäyttöisen elementalsouls/Claude-BugHunterin luvut kertovat, mitä taitopaketin sisällöstä kannattaa kysyä. Paketissa on kuvauksen mukaan 82 taitoa (README:ssä 83), 15 slash-komentoa ja 681 julkistetuista raporteista koottua kaavaa 24 haavoittuvuusluokassa. Kaavoista 433 on yksilöidysti lähteistetty ja 248 jää ilman yksilöityä lähdettä. Paketti kuuluu vain luvalliseen testaukseen ja sen agentin kontekstiin tuomat ohjeet kannattaa käydä läpi ennen käyttöä. Repo mvanhorn/last30days-skill painottaa Redditin, X:n, YouTuben, Hacker Newsin ja Polymarketin lähteitä äänten, tykkäysten ja rahallisten vetojen mukaan, mikä sopii trendien kartoitukseen paremmin kuin faktojen tarkistukseen. Repo mjmirza/app-store-compliance kokoaa 117 App Storen ja Google Playn hylkäysperustetta korjauksineen sekä 95 seurattua määräaikaa, muun muassa EU:n tekoälyasetuksen, DSA:n ja COPPA:n. Tarkistuslistana se on hyödyllinen. Määräajat kannattaa silti varmistaa alkuperäislähteistä, koska repo on yksityisen ylläpitäjän koonti.

Paikallisesti ajettavista työkaluista kaksi on kokeilun arvoisia sellaisenaan. Firecrawl Local on yksi binääri, joka muuntaa verkkosivut markdowniksi, toimii agentin MCP-palvelimena ja tarjoaa Firecrawl API:n localhostissa. Se sopii tilanteisiin, joissa agentin verkkohaku ei saa kulkea ulkopuolisen pilven kautta. Datasette Lite kirjoitettiin uudelleen service workereiden ja Pyodiden varaan. Selaimessa ajettava versio on nyt lähes erottamaton palvelinsovelluksesta. Osoitteen lite.datasette.io parametreilla voi avata SQLite-tietokannan, JSON- tai CSV-tiedoston tai SQL-skriptin ilman palvelinta. Menetelmien opetteluun sopii jingyaogong/minimind, joka lupaa 64 miljoonan parametrin kielimallin koulutuksen tyhjästä kahdessa tunnissa noin 3 juanin laskentakuluilla. Väite on projektin oma. Repo toteuttaa natiivilla PyTorchilla koko ketjun esikoulutuksesta SFT:hen, LoRAan, DPO:hon, GRPO:hon, agenttiseen vahvistusoppimiseen ja tislaukseen.

Alfavaiheen Tracer-Cloud/opensre lukee epäonnistuneen CI-ajon, jäljittää syyn ja työntää korjauksen katselmoitavaksi omalla infralla ja omilla malleilla. Projekti varoittaa itse rajapintojen ja integraatioiden muuttuvan. Kokeilu sopii erilliseen testirepoon eikä vielä tuotantoputkeen. Appwrite markkinoi itseään MCP- ja agenttilähtöisenä taustapalveluna, jota koodausagentti voi operoida elävänä projektina. Agentille kannattaa antaa oma kehitysprojekti

eikä tuotantoa. Kokoelma punkpeye/awesome-mcp-servers toimii hakemistona eikä laadun takeena. Harvinaista ensikäden tietoa tarjosi Laravel-hallintapaneelikirjaston Backpackin kirjoitus, jonka mukaan llms.txt-tiedosto ja MCP-palvelin eivät liikuttaneet sen mittareita. Ketju jäi 31 pisteeseen. Kirjoitus on lukemisen arvoinen sille, joka harkitsee MCP-palvelinta näkyvyyden lisäämiseksi.

11 Tietoturva ja haavoittuvuudet

Tekoälyn löytämät haavoittuvuudet näkyivät viikolla ylläpitäjien aikatauluissa. PostgreSQL 19 julkaistiin tavanomaisen syyskuun sijaan marraskuussa. LLM-työkalujen kehittäjän blogin mukaan viive johtui pääosin tekoälyn raportoimien haavoittuvuuksien ryöpystä, jota blogi kutsuu käynnissä olevaksi haavoittuvuusapokalypsiksi. Tivin mukaan myös Javan kriittinen aukko löydettiin tekoälyn avulla. Kyberturvallisuuskeskuksen tiedote 2026-38 koskee Oraclen ja OpenJDK:n päivityksiä 11 haavoittuvuuteen tuetuissa Java-versioissa. Vakavimmat mahdollistavat komentojen etäsuorituksen ja Kyberturvallisuuskeskus kehottaa päivittämään välittömästi. CVE-tunnisteet ja korjatut versiot on tarkistettava Oraclen ja OpenJDK:n tiedotteista. Pankkien toimitusjohtajat ja NSA kertoivat kongressille tekoälyvetoisten paikkausten olevan katkosten takana. Teknisiä tietoja kuulemisesta ei lähteissä ole.

Mallipalvelinten ja Open WebUI:n kaltaisten käyttöliittymien edessä yleisen NGINX:n tiedote 2026-39 koskee kymmentä haavoittuvuutta. Vakavin mahdollistaa koodin suorittamisen tietyllä tavalla muotoilluilla HTTP/2-paketeilla. Altistuksen tunnistaa versiosta (`nginx -v`) ja siitä, onko konfiguraatiossa `http2 on;` tai `listen`-direktiivissä `http2`-parametri. Väliaikainen kiertotie on poistaa HTTP/2 käytöstä, kunnes korjattu versio on asennettu. Php:n kriittinen haavoittuvuus koskee Tivin mukaan kaikkia tuettuja versioita ja korjaava päivitys oli määrä julkaista 19.11. Käytössä olevan version näkee komennolla `php -v`. Jakelupaketteja käyttävien kannattaa seurata jakelun tiedotteita, koska korjaus voi tulla jakeluun upstream-julkaisua myöhemmin.

MCP:n referenssipalvelinten repoon `modelcontextprotocol/servers` yhdistettiin 18.11. v2-haaraan korjaukset kahdeksaan GitHubin tietoturvatiedotteeseen. Kuusi niistä koskee git-palvelinta `mcp-server-git` (GHSA-h2g3-rwxm-xfm3, GHSA-9vjp-x34c-xg6v, GHSA-87qg-cmw8-jc5m, GHSA-pm7p-vphh-2cg7, GHSA-8wv7-r664-7594 ja GHSA-r2q8-f657-4r56). Filesystem-palvelinta koskee GHSA-rq2p-q7c4-6g27 ja everything-palvelinta GHSA-38j4-j742-j7r3. Haavoittuvuuksien luonnetta commit-viestit eivät kuvaa. Yksityiskohdat löytyvät tiedotteista tunnisteiden perusteella. Python-paketit leimattiin heti korjausten jälkeen CalVer-versioon 2026.11.18, joka on ensimmäinen korjaukset sisältävä git-palvelimen versio. Sitä vanhemmat asennukset eivät sisällä korjauksia. Filesystem- ja everything-palvelimen npm-pakettien tarkat versiot näkyvät niiden muutoslokeista.

Anthropicin virallisen security-guidance-liitännäisen viikko oli pitkä korjaussarja, jota yhdistää sama piirre: turvatyökalu ei kaatunut vaan jätti työnsä tekemättä kertomatta siitä kenellekään. Vakavin aukko koski git-tarkistusta. Komento, jossa `git \` jatkui seuraavalla rivillä sanalla `commit` tai `push`, käynnisti hookin. Komentokuviot sallivat sanojen väliin kuitenkin vain tyhjämerkkejä, minkä vuoksi mitään ei katselmoitu. Korjaus #7104 hyväksyy rivinjatkon kenoviivalla (Bash) ja gravis-merkillä (PowerShell). Korjaamattoman version läpi menneitä committeja ja pusheja voi etsiä agentti-istuntojen komentohistoriasta hakemalla git-rivejä, jotka päättyvät rivinjatkomerkkiin.

Toinen korjaus koskee salaisuuksia. Tekstitiedosto, jonka nimi tai kansio viittaa salaisuuksiin, kuten `api_keys.txt`, `tokens.txt` tai `secrets/prod.md`, luettiin aiemmin kokonaan katselmoivan mallin kehotteeseen, jos se oli nimetty uudelleen koodimaiseksi. Nyt tällainen tiedosto jätetään katselmoinnin ulkopuolelle. Samoin käsitellään tiedosto, jonka tuoreet commitit ovat nimenneet uudelleen tällaisesta nimestä. Kolmas korjaus koskee Windowsia, jossa pyenv-winin python-nimet päättyvät eräajotiedostoon. Julkaistussa versiossa 3.1.1 cmd.exe tulkitsi tulkin tarkistuksen merkit `>=` uudelleenohjaukseksi. Jokainen hook jätti projektiin tyhjän `(3`-nimisen tiedoston ja tulkki hylättiin. Koneella, jolla ei ollut muuta Pythonia, yksikään turvahook ei ajautunut. Oireen tunnistaa projektissa olevista tyhjistä `(3`-tiedostoista. Neljäs korjaus ohittaa muut kuin istunnon oman repon, jos niiden konfiguraatio muuttaa sitä, mitä git ajaa. Esimerkiksi `core.fsmonitor` ja `core.hooksPath` voivat saada tavallisen git-komennon käynnistämään repon määrittämän ohjelman. Korjaukset sisältyvät versioon 2.0.21.

Samaan sarjaan kuuluvat viat, joissa katselmointi ei toiminut lainkaan. Windowsin App Installer -alias sai jokaisen Edit- ja Write-toiminnon odottamaan 600 sekuntia, koska hookin käynnistin ajoi aliaksen joka kerta. Python.orgin Python ilman CA-varmenteita kaatoi macOS:llä jokaisen katselmoinnin varmenteen tarkistukseen. Suuri tuonti, 5 800 tiedostoa ja 4,6 miljoonaa riviä, vei 6,4 gigatavua muistia ennen kuin katselmointi ohitettiin tiedostomäärän vuoksi. Nyt git-tuloste luetaan virtana. Liitännäisen mukautettujen sääntöjen globeissa toimivat lisäksi aaltosulkeet, joiden laajennusten määrä on rajattu sekä globikohtaisesti että kaikkien sääntöjen yli. Rajan merkityksen näytti samalla viikolla glob-kirjasto. Versio 14.0.6 korjaa GHSA-qxr8-q827-836c:n, jossa ignore-kuvioilta puuttui sama `braceExpandMax`-raja kuin varsinaisilta kuvioilta. Rajaton aaltosulkulaajennus voi kuluttaa resursseja, jos ignore-kuviot tulevat epäluotettavasta lähteestä. Asennetun version näkee komennolla `npm ls glob`.

Agenttityökalut tiukensivat samalla sitä, mitä kloonattu repo voi tehdä. Clinen CLI 3.0.75 ei enää lue käynnistyshakemiston `.env`- eikä `bunfig.toml`-tiedostoa. Aiemmin tutkittavaksi kloonattu repo saattoi näiden kautta vaikuttaa CLI:n ympäristöön. Bunissa `bunfig.toml` voi esimerkiksi määrittää preload-skriptejä, jotka ladataan ennen ohjelmaa. Yolo-oletuksen kanssa CI:ssä ajettava kertaluontoinen ajo hyväksyy kaiken. Putkissa, joissa kehote sisältää luottamatonta tekstiä kuten issue-kuvauksia, automaattisen hyväksynnän asetus ja ajoympäristön rajaus kannattaa siksi tarkistaa. Claude Coden repossa @claude-maininnat eivät enää käynnisty forkeista tulevissa pull requesteissa (#111198). Forkin pull request on luottamatonta syötettä, joka voi sisältää kehoteinjektion. Agentilla taas on usein pääsy repon salaisuuksiin. Versio 2.1.329 näyttää luottamusdialogissa, missä kansion hookit on määritelty. Kloonatun repon projektitasoiset hookit ovat reitti, jota pitkin vieras koodi pääsee ajoon luottamuksen myöntämisen jälkeen.

Paikallisessa päättelyssä llama.cpp korjasi buildissa b12604 llama-grammarin kokonaislukuylivuodon. Virheelliset arvot tarkistetaan nyt jäsennyksessä ja jäsennys keskeytetään hallitusti. Riski koskee erityisesti llama-serveriä, joka hyväksyy pyynnöissä käyttäjän antaman GBNF-kieliopin, koska kieliopin syöttävä asiakas pääsee ajamaan juuri sitä jäsennyskoodia, jossa ylivuoto oli. CVE-tunnistetta tai hyödynnettävyysarviota julkaisutiedoissa ei ole. Kielioppeja luottamattomilta asiakkailta ottavat palvelimet kannattaa päivittää vähintään buildiin b12604. Ollaman julkaisuehdokkaan mukana kulkenut llama.cpp oli buildissa b12257. Lähteistä ei selviä, koskeeko ylivuoto Ollaman kielioppipolkua. Buildi b12580 toi lipun `--ui-mcp-proxy-allow-private-urls`, jonka nimestä seuraa, että web-käyttöliittymän MCP-välitys ei oletuksena ota yhteyttä yksityisiin osoitteisiin. Rajaus suojaa SSRF-tyyppisiltä hyökkäyksiltä. Lippua ei kannata käyttää palvelimella, joka on muiden käytettävissä verkon kautta.

Virtualisointikerroksessa CVE-2026-99312 koskee Linuxin KVM:ää x86-arkkitehtuurilla. Korjaus nollaa MMU-kontekstin ehdoitta, kun KVM emuloi INIT-signaalia. Asia koskee KVM-isäntiä, myös niitä, joilla agenttien koodia eristetään Firecrackerin kaltaisiin mikrovirtuaalikoneisiin. KVM:n käytön näkee komennolla `lsmod | grep kvm` ja korjauksen tilan jakelun ytimen tietoturvatiedotteista. Datasette 1.0a43 korjasi yksityisten tallennettujen kyselyjen paljastumisen, jossa kyselyn otsikko ja SQL olivat nähtävissä. Vakaaseen versioon 0.65.6 siirretty vähäisempi korjaus koskee tietokantaa, joka on rekisteröity kutsulla `add_database(db, name=..., route=...)` nimestä poikkeavalla reitillä. Se oli ladattavissa .db-tiedostona reitin osoitteesta myös käyttäjille, joilta pääsy oli evätty.

Käyttöjärjestelmätasolla yksittäisen kehittäjän analyysi Windows 11:n telemetriasta (694 pistettä, 568 kommenttia) väittää käyttöjärjestelmän keräävän tietoa koneella käytetyistä ohjelmointiagenteista ja kehitystyökaluista. Lähtevää diagnostiikkadataa voi tarkastella Windowsin Diagnostic Data Viewerilla. Hallituissa ympäristöissä keruun tasoa säädetään ryhmäkäytännöillä tai laitehallinnan kautta. Signal kirjoitti mobiilikäyttöjärjestelmien tekoälyagenttien tiedonkeruun rajoittamisesta (564 pistettä, 332 kommenttia). Kirjoitus jatkaa linjaa, jolla Signal Desktop otti vuonna 2025 Windowsissa oletuksena käyttöön näyttösuojauksen Microsoft Recallia vastaan.

Lähteet

Introducing GPT-6.2 Astra and Sol · OpenAI

GPT-6.2 Astra and GPT-6.2 Sol · Simon Willison

AINews: GPT-6.2 Astra, the #2 Frontier Model is better and 20%-600% cheaper, but has mixed reviews; 6.2 Sol likely main workhorse · Latent Space

0.171.0 · OpenAI Codex CLI releases

Introducing Gemini 4 Neon · Google DeepMind

[AINews] Gemini 4 Neon: GDM's Opus/Sol Contender · Latent Space

llm-gemini 0.38 · Simon Willison

llm-gemini 0.38.1 · Simon Willison

[AINews] Anthropic successfully completes largest IPO in history, valued at $1.5T · Latent Space

"Anthropic Is Not a Safe AI Company": Coalition Challenges Anthropic Safety · Hacker News: agents

Anthropic's compute spending is increasing by 4x per year, outpacing OpenAI's 3x · Epoch AI

Claude Platform release notes, November 16, 2026: Text watermark · Anthropic API release notes

Claude Platform release notes, November 17, 2026: Linked billing · Anthropic API release notes

Claude Platform release notes, November 16, 2026: Encryption key webhook events · Anthropic API release notes

Advancing chip design with OpenAI models · OpenAI

Research acceleration: The jagged frontier · OpenAI

Managing frontier AI risk: From training to deployment · OpenAI

Combatting adversarial distillation · OpenAI

Our first phase of embedded assessments · OpenAI

More ways to move care forward with ChatGPT for Healthcare · OpenAI

Our new blog for science · OpenAI

A step toward defining the Standard Model on a lattice · OpenAI

Quoting Levent Alpöge · Simon Willison

[AINews] Riemann Hypothesis proved...twice independently · Latent Space

Quoting Ravi Vakil · Simon Willison

Seven new models at scalable price-performance · Hacker News: MCP

Genie 4 is now available · Google DeepMind

Jensen Huang's Nvidia legacy, Claude Fable 5.2 and Anthropic's IPO, World Labs Releases Atlas to Developers · The Batch (deeplearning.ai)

How Google Duplex evolved to also call businesses for you in Gemini · Google Research

[AINews] Perplexity hits a $1B revenue run rate, up 4x in 2026 · Latent Space

Cloudflare Joins the Jevolution and Labels Itself a Neolab · Latent Space

Welcome Inkling 2 Mini, Cutting-edge Performance for your Local Setup · Hugging Face -blogi

Trending: thinkingmachines/Inkling-2-Mini · HF trending -mallit

Trending: unsloth/Inkling-2-Mini-GGUF · HF trending -mallit

Trending: AtomicChat/Inkling-2-Mini-Uncensored-GGUF · HF trending -mallit

New language model evaluation results available Inkling 2 Mini Intelligence Index: 20 · Artificial Analysis

v0.40.17 · Ollama-julkaisut

b12600 · llama.cpp-julkaisut

chatrlabs/Qwen-Image-2.1-Uncensored-Omni-Quantized · HF trending -mallit

apple/FastFrames-L · HF trending -mallit

Trending: apple/DepthPro2 · HF trending -mallit

Trending: inworld/voice-isolator · HF trending -mallit

Trending: Sentinel7/h3 · HF trending -mallit

Trending: nvidia/NVIDIA-Nemotron-3.5-Bolt-E4B-4GB · HF trending -mallit

Hugging Face and OpenRouter join forces to make open-weight models easy to run · Hugging Face -blogi

180B Parameters, 32GB of RAM: Pushing MoE Inference to the Limit with Transformers · Hugging Face -blogi

Rumble-2-Flash: an open 0.5B speech model for 40+ languages · Hugging Face -blogi

Show HN: I made an audio-to-MIDI model 60x more efficient · Hacker News: MCP

SRE-1: 27B model for Production Investigations · Hacker News: MCP

DaisyRose-27B: Pushing the Frontier of Post-Training Efficiency · Hugging Face -blogi

Language Models Can Rewrite Their Own Weights · Hacker News: MCP

Theorizer comes to Asta · Ai2 Blog

Ai2 and Taylor Geospatial partner to make geospatial AI easier to compare, run, and build on · Ai2 Blog

b12608 · llama.cpp-julkaisut

b12632 · llama.cpp-julkaisut

b12441 · llama.cpp-julkaisut

b12435 · llama.cpp-julkaisut

llama.cpp b12474 · llama.cpp-julkaisut

b12569 · llama.cpp-julkaisut

b12572 · llama.cpp-julkaisut

b12431 · llama.cpp-julkaisut

b12624 · llama.cpp-julkaisut

b12634: chat : trigger lazy grammar on bare JSON tool calls (#32497) · llama.cpp-julkaisut

Build Tool-Calling Parsers in Rust with AI Agents · NVIDIA Technical Blog

llama.cpp b12479: mtmd: support Minimax-Music3 · llama.cpp-julkaisut

b12590: mtmd: drama3 voice cloning (#32586) · llama.cpp-julkaisut

llama.cpp b12478 · llama.cpp-julkaisut

b12597 · llama.cpp-julkaisut

b12543 · llama.cpp-julkaisut

b12629 · llama.cpp-julkaisut

b12599 · llama.cpp-julkaisut

b12550: ggml : add backend scheduler tests (#27258) · llama.cpp-julkaisut

v0.40.17-rc1: mlxrunner: batch speculative verification (#19663) · Ollama-julkaisut

v0.40.17-rc2 · Ollama-julkaisut

n8n@2.49.0 · n8n releases

n8n@3.3.0 · n8n releases

n8n@2.48.5 · n8n releases

n8n@2.49.2 · n8n releases

n8n@2.48.6 · n8n releases

n8n@1.123.99 · n8n releases

n8n@2.49.3 · n8n releases

n8n@2.48.7 · n8n releases

n8n@1.123.97 · n8n releases

n8n@1.123.100 · n8n releases

langgraph==1.2.23 · LangGraph releases

langgraph==1.2.24 · LangGraph releases

Open WebUI v0.13.4 · Open WebUI releases

Agentic inference capacity grew 4.6× in a year, versus 2.6× for compute capacity · Epoch AI

Co-Designing Sparse Attention for Efficient Long-Context Inference · NVIDIA Technical Blog

Reducing Rollout Bottlenecks in Reinforcement Learning with NVIDIA Dynamo · NVIDIA Technical Blog

GPT-6.2 Astra (Max) · Artificial Analysis

GPT-6.2 Sol (Max) · Artificial Analysis

New language model evaluation results available Gemini 4 Neon (Low) Intelligence Index: 44 · Artificial Analysis

Qwen4 Max · Artificial Analysis

JoyAI-VL-Preview · Artificial Analysis

K2 Horizon 32B · Artificial Analysis

LFM3-1.7B · Artificial Analysis

LFM3-VL-2B · Artificial Analysis

Astra for Law 6.2 (Max) · Artificial Analysis

Astra for Law leads in legal intelligence · OpenAI

How good are Fable 5.2 and GPT-6.2 Astra at real work? · Epoch AI

Two stealth models just appeared on the leaderboard, both ahead of Fable 5.2 · Hacker News: MCP

Gemini 4 Neon is the new #1 on the Models Engineering Index · Hacker News: MCP

What is the best harness for each model? · Hacker News

harness-hub/Qwen3.8-27B-pi-adapted · HF trending -mallit

Introducing harness adapters: adapt your favorite model to your harness to improve performance and cost · Hugging Face -blogi

Introducing the Agent Arena Index · LMArena

Testing AI Exactly How It's Used: MLPerf Endpoints v1.0 · MLCommons

NVIDIA Vera Rubin NVL72 Leads Agentic Inference Performance on MLPerf Endpoints v1.0 · NVIDIA Technical Blog

MLCommons Releases MLPerf Training v6.1 Results, Spotlighting Recent Advances in AI Training · MLCommons

MLPerf Training v6.1: Bigger Clusters, Broader Hardware, and the Rise of Software-Driven Gains · MLCommons

NVIDIA Vera Rubin NVL72 Achieves up to 2x Blackwell Ultra Performance in MLPerf Training v6.1 · NVIDIA Technical Blog

Measuring alignment eval tells · METR

AI R&D-relevant model capabilities are improving rapidly and are getting harder to measure · METR

Perplexity and Intelligence · OpenAI

Can AI help scientists identify promising drug candidates? · OpenAI

Release v0.69.0 · Gemini CLI releases

TokenEconomizer: Learning from Experience to Make AI Coding Agents More Efficient · Microsoft Research

NVIDIA/SkillEvaluator · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Claude Code 2.1.326 · Claude Code changelog

Claude Code 2.1.327 · Claude Code changelog

Claude Code 2.1.328 · Claude Code changelog

Claude Code 2.1.329 · Claude Code changelog

Claude Code 2.1.330 · Claude Code changelog

CLI v3.0.75 · Cline releases

Desktop v0.0.50 · Cline releases

v4.1.27 · Cline releases

Inside the Codex sandbox · OpenAI

External sandboxes for coding agents · Simon Willison

Running coding agents without handing them my credentials · Hacker News

Launch HN: StarSling (YC P25): Self-improving CI for agent-volume repos · Hacker News: agents

How the Claude Code team thinks about verification · Hacker News: Claude Code

Harvard researchers find AI coding agents shifted the bottleneck to code review · Hacker News: agents

Ask HN: If AI tools are so good, why isn't your bug backlog empty? · Hacker News

Quoting Daniel Stenberg · Simon Willison

Muru · Simon Willison

datasette-arrows 0.1a0 · Simon Willison

datasette-maplibre 0.1a0 · Simon Willison

The War Against Slop and an Emerging Weapon to Fight it: Formal Verification · Latent Space

llama.cpp b12475 · llama.cpp-julkaisut

b12508: cuda : fuse NORM + MUL (+ ADD) (#32454) · llama.cpp-julkaisut

b12542 · llama.cpp-julkaisut

b12546 · llama.cpp-julkaisut

b12606 · llama.cpp-julkaisut

b12628 · llama.cpp-julkaisut

test(git): keep a GHSA-8wv7-r664-7594 test independent of the system git configuration · MCP servers commits

docs: add Archestra.AI and mcp-use to the extension client matrix (#3574) · MCP-määrittely commits

docs: stop presenting deprecated features as current on the spec overview pages (#3552) · MCP-määrittely commits

docs: align the extensions overview with SEP-3392 (#3565) · MCP-määrittely commits

sep(SEP-3415): Events Extension (#3415) · MCP-määrittely commits

docs(SEP-3415): register the Events extension (#3568) · MCP-määrittely commits

docs: align FSIG charter with problem-first, WG-led governance (#3572) · MCP-määrittely commits

docs(community): say who can open GitHub Discussions and what they are for (#3573) · MCP-määrittely commits

docs: update Discord invite link (#3577) · MCP-määrittely commits

docs(inspector): add load testing and Active Sessions docs (#3503) · MCP-määrittely commits

google-labs-code/stitch-skills · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

kajisho5/ffmpeg-skill · GitHub trending

elementalsouls/Claude-BugHunter · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

mvanhorn/last30days-skill · GitHub trending

mjmirza/app-store-compliance · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

firecrawl/firecrawl · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Running the full server-side Datasette Python app in the browser with service workers and Pyodide · Simon Willison

jingyaogong/minimind · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Tracer-Cloud/opensre · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

appwrite/appwrite · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

punkpeye/awesome-mcp-servers · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Llms.txt and MCP didn't move the needle · Hacker News: MCP

PostgreSQL 19 Released · Simon Willison

Moderni maailma nojaa Javaan: Nyt siitä löydettiin vakava haavoittuvuus · Tivi

Kriittisiä haavoittuvuuksia OpenJDK:ssa · Kyberturvallisuuskeskus

Bank CEOs, NSA Tell Congress That AI-Driven Patches Are Behind Outages · Hacker News: agents

Kriittinen haavoittuvuus NGINX-tuotteiden HTTP/2-toteutuksissa · Kyberturvallisuuskeskus

Varoitus leviää maailmalla: Php:ssä jälleen kriittinen haavoittuvuus · Tivi

Merge pull request #5581: fix(git): GHSA-h2g3-rwxm-xfm3 · MCP servers commits

Merge pull request #5582: fix(git): GHSA-9vjp-x34c-xg6v · MCP servers commits

Merge pull request #5585: fix(git): GHSA-87qg-cmw8-jc5m · MCP servers commits

Merge pull request #5586: fix(git): GHSA-pm7p-vphh-2cg7 · MCP servers commits

Merge pull request #5587: fix(git): GHSA-8wv7-r664-7594 · MCP servers commits

Merge pull request #5590: fix(git): GHSA-r2q8-f657-4r56 · MCP servers commits

Merge pull request #5584: fix(filesystem): GHSA-rq2p-q7c4-6g27 · MCP servers commits

Merge pull request #5583: fix(everything): GHSA-38j4-j742-j7r3 · MCP servers commits

Merge pull request #5591: chore: stamp Python CalVer versions for 2026.11.18 · MCP servers commits

Merge pull request #5537: chore: version packages · MCP servers commits

security-guidance: match git commit and push continued on the next line · Claude Code plugins/marketplace (virallinen)

Merge pull request #7104 from anthropics/mhegazy/sg-line-continuation · Claude Code plugins/marketplace (virallinen)

security-guidance: leave renamed key lists out of the review · Claude Code plugins/marketplace (virallinen)

security-guidance: keep the interpreter check working through a .bat shim · Claude Code plugins/marketplace (virallinen)

security-guidance: skip other repositories whose settings change what git runs · Claude Code plugins/marketplace (virallinen)

Merge pull request #7143: security-guidance: bump version to 2.0.21 · Claude Code plugins/marketplace (virallinen)

security-guidance: edit hook no longer waits on App Installer's python alias · Claude Code plugins/marketplace (virallinen)

security-guidance: use the CA file macOS ships when Python has none · Claude Code plugins/marketplace (virallinen)

Merge pull request #7140: commit and push reviews no longer hold a large diff in memory · Claude Code plugins/marketplace (virallinen)

security-guidance: custom rule globs can use braces · Claude Code plugins/marketplace (virallinen)

build(deps-dev): bump glob from 14.0.4 to 14.0.6 (#3558) · MCP-määrittely commits

Don't run @claude mentions on pull requests from forks (#111198) · Claude Code repo commits

b12604 · llama.cpp-julkaisut

b12580 · llama.cpp-julkaisut

KVM: x86: Unconditionally reset MMU context when emulating INIT (CVE-2026-99312) · Hacker News: MCP

datasette 1.0a43 · Simon Willison

datasette 0.65.6 · Simon Willison

Windows 11 is collecting data on which coding agents and developer tools you use · Hacker News

Restricting AI agents in mobile operating systems platforms from capturing data · Hacker News: agents