AI Horisontti



Lasku erääntyi

Google myönsi hakukoneensa tekoälyvastausten syövän hakutuloja ja menetti noin 500 miljardia dollaria markkina-arvostaan. Claude-avustajan kehittäjä Anthropic avasi taloutensa pörssiin hakiessaan.

TIETOTURVA
Hyökkäys Firefoxin korjauksesta maksoi noin 1 000 dollaria

AGENTIT
Hiirellä toimiva avustaja tarvitsee uuden työntekijän rajat

TYÖELÄMÄ
Alankomaat siirtää virkatyötä tekoälylle, 25 000 lähtötarjousta

PÄÄKIRJOITUS · VIIKKOKATSAUS

Lasku kahdessa merkityksessä

Vastaperustetun yrityksen omistaja kirjoittaa hakukenttään ”tilitoimisto Tampere” ja saa valmiin vastauksen ennen kuin näkee yhtään linkkiä tai mainosta. Klikkaus jää tekemättä. Sama klikkaus on tähän asti hyödyttänyt sekä mainostavaa tilitoimistoa että hakukonetta. Nyt Google myönsi, että sen oma tekoälytila eli kysymykseen suoraan vastaava hakutapa heikentää yhtiön hakutuloja. Osake putosi 15 % ja markkina-arvosta katosi noin 500 miljardia dollaria. Pörssissä lasku tarkoittaa eri asiaa kuin kirjanpidossa, mutta tällä viikolla molemmat merkitykset päätyivät samalle pöydälle.

Kirjanpidossa tekoäly on uudenlainen kulu, koska sen lasku kasvaa usein jokaisen käyttökerran mukana toisin kuin toimisto-ohjelman kiinteähintainen lisenssi. Laskun suunta riippuu myös myyjän taloudesta. Claude-avustajan kehittäjä Anthropic avasi taloutensa sijoittajille hakiessaan pörssiin. Listautumisasiakirjasta on ehditty poimia sekä 81 miljardin dollarin tappiot että yhden vuosineljänneksen 1 miljardin dollarin liikevoitto. Markkinaosio selvittää, miksi luvut eivät vielä sovi yhteen. Osion ”Mitä seurata” mukaan ensimmäinen osavuosikatsaus kertoo enemmän kuin yksikään poimittu luku, koska siinä tappiot, kassa ja kasvu raportoidaan samoilla määritelmillä.

Laskun liitteeksi ChatGPT:n kehittäjä OpenAI julkaisi asiakastarinan, jonka mukaan rahoitusyhtiö Credit Saison sai yritysversion kokeilussa 954 % tuoton sijoitukselleen. Luku on peräisin myyjän markkinointiaineistosta, eikä julkisista tiedoista selviä, miten se on laskettu. Tarkka prosentti ilman laskutapaa on kuitti, josta rivit puuttuvat.

Ohjelmistokehittäjien keskustelupalstalla nousi suosituksi kirjoitus, jonka mukaan tekoäly lupasi vähemmän työtä ja kirjoittaja tekee enemmän töitä kuin koskaan. Tarkistamiseen kuluva aika päätyy harvoin myyjän tuottolaskelmaan. Osio ”Mitä tämä tarkoittaa käytännössä” seuraa pientä tilitoimistoa laskun, klikkauksen ja tarkistustyön läpi. Yrittäjälle se on viikon hyödyllisin luettava.

Osa tekoälyn hinnasta ei tule laskuna lainkaan. Sairaalan potilastietoja päätyi työntekijän virheen vuoksi julkiseen tekoälypalveluun, eikä vuoto vaatinut hyökkääjää. Samalla viikolla OpenAI julkaisi mallin, joka käyttää tietokonetta hiirellä ja näppäimistöllä kuten ihminen. Osio ”Työkalut ja avustajat arjessa” neuvoo antamaan sellaiselle avustajalle uuden työntekijän tapaan oman tunnuksen ja vain tehtävän vaatimat oikeudet. Turvallisuuden merkittävin tulos tuli tutkimuslaitos Epoch AI:lta, jonka kokeessa kenen tahansa ladattavissa oleva tekoälymalli muutti Firefox-selaimen tietoturvakorjaukset toimiviksi hyökkäyksiksi ennen kuin korjaukset ehtivät käyttäjille. Yksi hyökkäys maksoi noin 1 000 dollaria. Päivitykset kannattaa siksi asentaa heti ja selain käynnistää uudelleen.

Kohinaksi jäivät vertailulistojen kärkipaikat, koska riippumaton arviointiyhtiö vaihtoi mittaristonsa samalla viikolla. Sunnuntaina levisi huhu vuotaneesta huippumallista, jonka latauslinkki poistettiin haittaohjelmaepäilyn vuoksi. Tekninen osa menee syvemmälle ja selvittää, miksi kärkipaikkoja kannattaa lukea varauksella ja miten avustajan kirjoitusoikeus rajataan yhteen kansioon. Lasku on jo perillä, kuitti vielä matkalla.

TÄSSÄ NUMEROSSA

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

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

VIIKON KÄRKI

Viikon tärkein asia oli tekoälyn talous. Arvailujen sijaan siitä kertoivat nyt viralliset asiakirjat ja pörssikurssit ja se koskee jokaista yritystä, joka maksaa tekoälypalvelusta tai hankkii asiakkaansa Google-haun kautta. Viikon alussa Claude-avustajan kehittäjä Anthropic avasi taloutensa sijoittajille listautumisasiakirjassa, minkä jälkeen loppuviikolla Google myönsi hakukoneensa tekoälytilan heikentävän yhtiön hakuliiketoiminnan tuloja ja Googlen markkina-arvosta katosi noin 500 miljardia dollaria. Päivästä toiseen jatkui kysymys siitä, mitä arkisiin työkaluihin voi syöttää ja millä oikeuksilla tekoäly saa toimia, aina sairaalan potilastietovuodosta ihmisen tunnuksilla selaimessa työskenteleviin avustajiin. Kohinaksi osoittautuivat vertailulistojen kärkipaikat juuri vaihtuneella mittarilla, asiakastarinoiden prosenttiluvut sekä sunnuntaina levinnyt huhu vuotaneesta huippumallista, jonka latauslinkki poistettiin haittaohjelmaepäilyn vuoksi.

---

1 **Tekoälyyhtiön talous tuli julkiseksi.** Tekoälyalan uutiskirjeen Latent Spacen mukaan Anthropic hakee pörssiin noin 2 000 miljardin dollarin arvostuksella marraskuun alkupuolella. Asiakirjasta poimitut luvut eivät vielä muodosta yhtenäistä kuvaa, koska Tivi kertoi 81 miljardin dollarin tappioista ja Latent Space toisen vuosineljänneksen 1 miljardin dollarin liikevoitosta. Pörssiyhtiönä Anthropicin on raportoitava taloudestaan säännöllisesti, jolloin Clauden käyttäjä voi ensimmäistä kertaa arvioida itse toimittajansa kestävyyttä.

2 **Google myönsi tekoälyn syövän hakutuloja.** Latent Spacen mukaan Google on myöntänyt, että sen hakukoneen tekoälytila eli kysymykseen suoraan vastaava hakutapa heikentää yhtiön hakuliiketoiminnan tuloja. Osake putosi uutiskirjeen mukaan 15 %, mikä on pahin pudotus sitten Googlen listautumisen vuonna 2004. Yrityksen, joka saa asiakkaansa Google-haun tai Google-mainosten kautta, kannattaa nyt katsoa omista luvuistaan, miten hakukävijöiden määrä on kehittynyt.

3 **Tietoturvakorjauksista tuli halpoja hyökkäyksiä.** Tutkimuslaitos Epoch AI:n kokeessa kenen tahansa ladattavissa oleva tekoälymalli muutti Firefox-selaimen tietoturvakorjaukset toimiviksi hyökkäyksiksi kuuden selainjulkaisun kohdalla. Korjaukset olivat julkisia, mutta niitä ei ollut vielä toimitettu käyttäjille ja yksi hyökkäys maksoi noin 1 000 dollaria. Päivitykset kannattaa siksi asentaa heti ja selain käynnistää uudelleen, koska selain ottaa korjauksen usein käyttöön vasta uudelleenkäynnistyksessä.

4 **Eurooppalainen ja paikallinen vaihtoehto vahvistuivat.** Ranskalainen Mistral julkaisi suurimman Euroopassa koulutetun kielimallinsa kaikkien ladattavaksi. Yritys voi siten käyttöehdoista riippuen valita, kuka mallia pyörittää ja minkä maan lainsäädännön alla. Pienemmässä mittakaavassa yhdysvaltalaisen Liquid AI:n kevyet mallit toimivat nyt ilmaisissa ohjelmissa tavallisella tietokoneella niin, ettei teksti tai ääni lähde koneelta minnekään. Kummankin kohdalla käyttöehdot ja suomen kielen tuki on tarkistettava ennen käyttöönottoa.

5 **Avustaja sai hiiren ja tunnukset.** OpenAI julkaisi mallin, joka käyttää tietokonetta ihmisen tapaan hiirellä ja näppäimistöllä, täyttää lomakkeita ja tulkitsee teknisiä piirustuksia. Samalla viikolla yhtiö laajensi yhteistyötään Oraclen ja ServiceNow'n kanssa niin, että niiden yritysjärjestelmiä voi käyttää ChatGPT:n kautta. Avustaja ei enää vain ehdota tekstiä vaan toimii itse järjestelmissä, joten sen käyttöoikeuksista päätetään yhtä huolellisesti kuin uuden työntekijän tunnuksista.

6 **Tietovuodot syntyvät arjessa.** Viikko alkoi Tivin uutisella sairaalasta, jonka potilastietoja päätyi työntekijän virheen vuoksi julkiseen tekoälypalveluun. Sunnuntaina tutkimusyhtiö Parameter Labs kuvasi hyökkäysketjun, jolla ChatGPT-tilejä pystyi kaappaamaan. Työpaikan tekoälykeskusteluihin kertyy asiakastietoja ja liikesalaisuuksia, joten niiden käytölle tarvitaan yhtä tarkat säännöt kuin sähköpostille.

7 **Tekoäly henkilöstöpäätöksissä.** Tivin mukaan Alankomaiden hallinto tarjoaa 25 000 virkamiehelle mahdollisuutta lähteä, koska osa työstä siirtyy säästötavoitteen vuoksi tekoälylle. Tivi välitti sunnuntaina Helsingin Sanomien uutisen, jonka mukaan Suomen suurin julkinen työnantaja Helsingin kaupunki korvaa työntekijöitä tekoälyllä. Sveitsiläinen Proton puolestaan ehti viikon aikana suunnitella henkilöstönsä uudelleenhaastatteluja tekoälyn käytöstä ja perua ne.

01 Mitä tämä tarkoittaa käytännössä

Pienen tilitoimiston uudet asiakkaat ovat vuosien ajan löytäneet sen samaa reittiä. Vastaperustetun yrityksen omistaja kirjoittaa hakukenttään ”tilitoimisto Tampere” ja päätyy hakutulosten yläpuolella näkyvän mainoksen kautta toimiston yhteydenottolomakkeelle. Googlen hakukoneen tekoälytila voi nyt koota hänelle valmiin vastauksen suosituksineen ennen kuin hän näkee yhtään linkkiä tai mainosta. Googlen tulot syntyvät pääosin juuri hakutulosten yhteydessä näytettävistä mainoksista. Latent Spacen mukaan yhtiö on nyt itse myöntänyt, että sen oma tekoälytila heikentää näitä tuloja. Mainostaja ja hakukone ovat tähän asti hyötyneet samasta klikkauksesta. Tekoälyvastaus voi tehdä klikkauksesta tarpeettoman kummallekin.

Tilitoimiston omistaja voi mitata muutoksen omista luvuistaan jo nyt. Kävijäseuranta näyttää, miten Googlesta tulleiden käyntien määrä on kehittynyt viimeisen vuoden aikana. Mainostilin raporteista näkee puolestaan, tuoko sama kuukausibudjetti yhä yhtä monta yhteydenottoa kuin vuosi sitten. Tekoälyvastaukset kokoavat tietonsa verkossa olevista lähteistä. Yrittäjä voi vaikuttaa vain harvoihin asioihin siinä, mitä kone hänen yrityksestään kertoo. Yksi niistä on se, että palvelut, hinnat ja aukioloajat on kirjoitettu omille sivuille selkeästi ja ajan tasalla. Asiakas, joka saa vastauksensa suoraan hakukoneelta, ei välttämättä koskaan käy sivulla, jonka varaan mainosbudjetti on rakennettu.

Sama tilitoimisto maksaa tekoälystä kahdella tavalla, joista ensimmäinen näkyy laskussa. Yrityksille myytävät tekoälypalvelut laskuttavat usein sen mukaan, kuinka paljon tekstiä ne käsittelevät. Toimisto-ohjelman lisenssi maksoi aikanaan saman summan riippumatta siitä, kuinka ahkerasti ohjelmaa käytettiin, kun taas tekoälyn lasku kasvaa jokaisen käyttökerran mukana. Viikon uutisissa tämä ero näkyi budjettipäätöksinä. Talouslehti Fortunen mukaan verkkolaitevalmistaja Cisco antoi 27-vuotiaalle insinöörille rajattoman tekoälykiintiön vanhan ohjelmakoodin uudelleenkirjoittamiseen. Anthropic puolestaan toi yritysasiakkailleen ryhmäbudjetin, jonka tiimi jakaa jäsenten henkilökohtaisten rajojen lisäksi. Pienessä yrityksessä vastaava päätös on arkisempi mutta yhtä todellinen: kuinka paljon tekoälyyn saa kuukaudessa kulua ja kuka rajaa valvoo.

Laskun tuleva suunta riippuu myös toimittajien omasta taloudesta. Anthropicin listautumisasiakirja ja Googlen kurssipudotus nostivat sijoittajien keskusteluun kysymyksen, joka koskee pientäkin asiakasta: kuka laskentatehon lopulta maksaa. Tutkimuslaitos Epoch AI:n mukaan neljännesvuosittain myytävien tekoälysirujen laskentateho kaksinkertaistuu noin yhdeksässä kuukaudessa, vaikka kasvu on hidastunut vuoden 2023 huipusta. Kapasiteetin rakentamisen hinta päätyy lopulta tilausmaksuihin ja käyttörajoihin.

Toinen kustannus ei näy laskussa lainkaan. OpenAI julkaisi perjantaina asiakastarinan, jonka mukaan rahoitusyhtiö Credit Saison sai ChatGPT:n yritysversion kokeilussa 954 % tuoton sijoitukselleen. Luku on peräisin myyjän markkinointiaineistosta, eikä julkisista tiedoista selviä, miten se on laskettu. Ohjelmistokehittäjien keskustelupalstalla Hacker Newsissa luettiin samaan aikaan toisenlaisia tilinpäätöksiä. Viikon vilkkaimpiin aiheisiin nousi kirjoitus, jonka mukaan tekoäly lupasi vähemmän työtä ja kirjoittaja tekee nyt enemmän töitä kuin koskaan. Toisessa ketjussa kehittäjä kertoi olevansa väsyneempi kuin ennen, koska rinnakkaisia töitä on vahdittava ja koneen tuottamaa tekstiä on luettava koko ajan.

Ohjelmistoalalla luettavan määrä kasvaa nopeammin kuin lukijoiden. Ohjelmakoodin yhteistyöpalvelu GitHubin vuosiraportin mukaan hyväksyttyjen koodimuutosten määrä kolminkertaistui alle vuodessa noin 220 miljoonaan kuukaudessa. Valtaosa kasvusta osui tämän vuoden jälkipuoliskolle, jolloin itsenäisesti ohjelmoivista avustajista tuli monen kehittäjän arkityökalu. Verkkosivun tai kassajärjestelmän liitännän teettäjälle muutos tarkoittaa nopeampaa tekemistä ja entistä ratkaisevampaa tarkastamista. Toimittajalta kannattaa kysyä suoraan, kuka tekoälyn kirjoittaman koodin tarkastaa ja näkyykö nopeampi tekeminen laskussa.

Tilitoimistossa sama laskutoimitus koskee kuitteja. Tekoäly esilajittelee kuukauden kuitit ja säästää näppäilytyötä, kun taas kirjanpitäjä vastaa jokaisesta kirjauksesta itse. Säästetystä ajasta on vähennettävä tarkistamiseen kuluva aika, joka harvoin päätyy myyjän tuottolaskelmaan. Amazonin insinöörijohto kehotti viikolla kehittäjiään kirjaamaan ylös tehtävät, joissa nykyiset tekoälytyökalut eivät auta. Linjauksen mukaan sellainen tehtävä on havainto eikä epäonnistuminen. Pienessä toimistossa vastaava havainto voi olla yksi rivi siitä, että avustaja sekoittaa toistuvasti arvonlisäverokannat. Kone tekee luonnoksen sekunneissa. Nimi tilinpäätöksen alla on silti kirjanpitäjän.

Kolmas käytännön kysymys koskee sitä, mitä tekoälylle annetaan. Tivin maanantaina kertoma sairaalan tietovuoto ei vaatinut hyökkääjää, koska potilastiedot päätyivät julkiseen tekoälypalveluun työntekijän virheen vuoksi. Terveystiedot ovat EU:n tietosuoja-asetuksessa erityisiä henkilötietoja, joiden käsittelylle on tiukemmat ehdot kuin esimerkiksi yhteystiedoille. Tilitoimiston arjessa liikkuu henkilötunnuksia, palkkatietoja ja asiakkaiden taloustietoja, joita ei pidä viedä palveluun ilman sopimusta. Monen tekoälypalvelun yritysversion ehdoissa luvataan, ettei syötettyjä tietoja käytetä mallien kouluttamiseen. Lisäksi tietojen käsittelystä tehdään erillinen sopimus, jossa sovitaan muun muassa käsittelypaikasta. Ilmaisversioissa tällaisia sitoumuksia ei välttämättä ole.

Ruotsin valtionhallinto veti viikon alussa rajan palvelun eikä tekniikan mukaan. ChatGPT kiellettiin valtion työntekijöiltä, kun taas Microsoftin Copilot jäi käyttöön. Julkiselle sektorille myyvä yritys voi siis kohdata asiakkaan, joka määrää tarkasti, mitä tekoälypalvelua sen aineistoon saa käyttää.

Yrittäjälle tästä seuraa lyhyt tarkistuslista. Ensimmäinen kysymys on, mitä tekoälypalveluja työntekijät käyttävät arjessaan. Se selviää useimmiten kysymällä eikä ohjeita lukemalla. Toinen on kirjallinen ohje siitä, mitä palveluun saa liittää. Kolmas on käytössä olevan palvelun sopimustaso ja tietojen käsittelypaikka, jotka löytyvät palvelun yritysehdoista. Samalla kannattaa varmistaa, että selain ja muut ohjelmat ovat päivittyneet ja ne on käynnistetty uudelleen. Tietovuoto ei vaadi pahaa tahtoa. Kiireessä käteen osuu se työkalu, joka on lähimpänä.

02 Työkalut ja avustajat arjessa

Google julkaisi keskiviikkona Gemini-tekoälynsä uuden kärkimallin, joka Latent Spacen mukaan johtaa riippumattomia vertailuja. Tavallinen Geminin käyttäjä ei pääse vielä kokeilemaan sitä. Google avaa mallin ensin luotetuille kyberpuolustajille eli asiantuntijoille, jotka suojaavat verkkoja ja järjestelmiä hyökkäyksiltä. Hintaa, saatavuutta Suomessa tai laajemman jakelun aikataulua ei ole kerrottu. Vertailujen kärkipaikkaa kannattaa lukea varauksella myös siksi, että riippumaton arviointiyhtiö Artificial Analysis vaihtoi mittaristonsa uuteen samalla viikolla. Mittarin vaihtuessa uusia tuloksia ei voi verrata suoraan vanhoihin listoihin. Mallin ilmestyessä Gemini-sovelluksen valikkoon parhaan arvion antavat omat koetehtävät, kuten viime kuun tarjouspyyntö tai hankala asiakaskirje, jonka hyvän vastauksen tuntee jo itse.

Ranskalainen Mistral julkaisi torstaina suurimman Euroopassa koulutetun kielimallinsa kaikkien ladattavaksi. Malli on niin suuri, että sen pyörittäminen vaatii palvelinkeskuksen laitteistoa eikä se toimi työkannettavalla. Pienelle yritykselle avoimuus tarkoittaa siksi valinnanvapautta siitä, kuka mallia pyörittää, missä maassa ja kenen lainsäädännön alaisena. Asiakastiedot voivat silloin kulkea eurooppalaisen palveluntarjoajan tai yrityksen oman palvelimen kautta eivätkä yhdysvaltalaisen pilvipalvelun läpi. Mistral kutsuu mallia parhaaksi avoimeksi malliksi ohjelmointiin, mikä on yhtiön oma väite eikä riippumattomasti todennettu tulos. Julkistuksen tiedoista ei selviä, salliiko lisenssi kaupallisen käytön ja millä ehdoin. Asia on tarkistettava ennen kuin mallin varaan rakennetaan esimerkiksi asiakaspalvelua tai sopimusten läpikäyntiä.

OpenAI:n viikon uutuus muuttaa sitä, mitä avustaja tekee. Yhtiön uusi nopea malli käyttää tietokonetta ihmisen tapaan hiirellä ja näppäimistöllä. OpenAI:n mukaan se tulkitsee tarkasti kuvia ja teknisiä piirustuksia, täyttää lomakkeita ja muokkaa videoita näytöllä. Pienessä rakennussuunnittelutoimistossa tällainen avustaja voisi avata asiakkaan pohjapiirroksen, poimia siitä mitat ja täyttää kunnan sähköisen lupalomakkeen. Aiemmat avustajat ehdottivat tekstiä, jonka ihminen kopioi oikeaan paikkaan. Uusi malli napsauttaa itse, jolloin virhe lähtee eteenpäin samalla napsautuksella kuin oikea tieto. Hinnat ja saatavuus eri tilauksissa kannattaa tarkistaa OpenAI:n sivuilta ennen kuin töitä suunnittelee mallin varaan.

Tunnuksilla toimiva avustaja tarvitsee samat rajat kuin uusi työntekijä. Avustajalle kannattaa luoda oma käyttäjätunnus, jolla on vain tehtävän vaatimat oikeudet. Omistajan pankkitunnukset ja pääkäyttäjän salasanat eivät kuulu sille. Maksut, hyväksynnät ja viranomaisille lähtevät ilmoitukset on turvallisinta jättää ihmisen painettaviksi. Sama käyttöoikeuskysymys koskee viikon kahta yritysliittoa, joissa OpenAI tuo Oraclen liiketoimintasovellukset ja ServiceNow'n työnkulut ChatGPT:n sisään. ServiceNow on monen suuren yrityksen taustajärjestelmä, jonka kautta hoidetaan esimerkiksi it-tukipyynnöt ja henkilöstöasiat. Keskusteluun voivat siten tulla yrityksen tilaus-, palkka- ja henkilöstötiedot. Ennen käyttöönottoa kannattaa siksi selvittää, mitä tietoja liitännän kautta liikkuu OpenAI:lle ja kuka keskustelussa näkee mitäkin.

Avustajiin asennettavat lisätaidot eli valmiit ohje- ja koodipaketit saavat samat oikeudet kuin avustaja itse. NVIDIA julkaisi ilmaisen työkalun, joka etsii niistä ennen asennusta haitallista koodia, piilotettuja käskyjä ja tietovuotoja. Yhtiön mukaan tällaiset paketit saavat nykyään käyttäjän luottamuksen automaattisesti ja käyvät läpi vain vähäisen tarkastuksen. Avoimen lähdekoodin hankkeissa kiinnostus kohdistui samaan aikaan tekoälytyökavereihin, joille annetaan oma selain, omat tiedostot ja omat kirjautumiset. Omat kirjautumiset tarkoittavat oikeita tunnuksia oikeisiin palveluihin. Ensimmäisiin kokeiluihin sopivat siksi testitilit paremmin kuin pankki tai asiakasrekisteri.

Ohjelmointiavustajia käyttäville viikko toi tiukempia rajoja. Anthropicin ohjelmointityökaluun tuli asetus, jolla se ei muokkaa tiedostoja työkansion ulkopuolella missään käyttötilassa. Googlen vastaavan työkalun testiversio korjasi samanlaista rajatarkistusta Windows-koneilla. Avustajia tiedostoihin kytkevän yhteisen liitäntästandardin tiedostopalvelimesta puolestaan korjattiin aukko, jonka kautta tiedostoja saattoi tietyissä asetuksissa päätyä sallittujen kansioiden ulkopuolelle. Käyttäjät raportoivat lisäksi avoimessa vikailmoituksessa, että Anthropicin työkalu liittää käyttäjän sähköpostiosoitteen tekoälylle annettaviin taustaohjeisiin. Yhden käyttäjän mukaan malli oli välittänyt osoitteen verkkosivustolle. Vahvistusta viasta ei ole julkaistu. Avustajan oikeudet lähettää tietoa verkkoon kysymättä kannattaa silti käydä läpi, koska kaikki sen tiedossa oleva voi päätyä verkkopyyntöön. Agentille eli itsenäisesti tehtäviä suorittavalle tekoälylle kannattaa antaa vain se kansio, jota työ koskee.

Arjen kannalta viikon kiinnostavin kehitys tapahtui pienissä malleissa, jotka toimivat omalla koneella. Liquid AI julkaisi pienten mallien perheen, johon kuuluvat tekstimalli, kuvia ymmärtävä malli ja puhemalli. Puhemalli tunnistaa puhetta, kääntää sitä ja tuottaa puhetta 18 kielellä. Mallit tulivat viikon aikana saataville suosittuihin ilmaisiin ohjelmiin, joilla tekoälymalleja ajetaan omalla tietokoneella. Torstaina yksi suosituimmista ohjelmista alkoi tukea niitä myös Windows- ja Linux-koneilla, kun ne aiemmin toimivat siinä vain Applen omilla piireillä varustetuissa Maceissa. Pienin tekstimalli on niin kevyt, että se toimii yleensä tavallisella kannettavalla.

NVIDIA julkaisi lisäksi monikielisen puheentunnistusmallin, jota voi ohjata ohjeilla ja mukauttaa omaan käyttöön. Sen ajaminen vaatii tehokkaan koneen. Avoimien tekoälymallien jakelupalvelu Hugging Face esitteli puolestaan puheentunnistusmalleja Applen laitteille. Kokouksen litterointi eli puheen muuntaminen tekstiksi onnistuu siis yhä useammin niin, ettei ääni lähde omalta koneelta minnekään. Julkaisujen tiedoissa suomen kieltä ei mainita. Suomen tuki kannattaa kokeilla lyhyellä äänitteellä ennen kuin puhelinmuistioiden tai asiakaspuheluiden litterointi siirretään omalle koneelle.

Myös paperiarkistot muuttuvat haettaviksi ilman pilvipalvelua. Yhdysvaltalainen voittoa tavoittelematon tutkimuslaitos Ai2 jakaa avoimesti tekstintunnistustyökalua, joka lukee tekstin skannatusta paperista tai valokuvasta. Työkalua sovelletaan nyt historiallisiin ja arkeologisiin arkistoihin. Kiinalaisen Baidun uusi tekstintunnistusmalli nousi samaan aikaan Hugging Facen suosituimpien joukkoon. Tilitoimistolle, sukututkijalle tai pienelle museolle kehitys tarkoittaa, että vanhat mapit voi muuttaa haettavaksi tekstiksi lähettämättä asiakirjoja kenellekään ulkopuoliselle. Käyttöönotto vaatii yhä jonkin verran teknistä osaamista.

Hinnoissa ja ehdoissa viikko toi muutoksia lähinnä kalliiden tilausten maksajille ja omia sovelluksia rakentaville. OpenAI seurasi Anthropicia ja antaa nyt ehdot täyttävän 100, 200 tai 500 dollarin kuukausitilauksen haltijalle saman summan hyvitystä ohjelmointirajapinnan käyttöön. Ohjelmointirajapinta on yhteys, jonka kautta tekoäly kytketään omaan ohjelmaan. Pelkkä keskustelukäyttäjä ei hyödy muutoksesta. OpenAI:n hyvitys kohdistetaan yhteen kehittäjätilin organisaatioon, kun taas Anthropicin hyvitykset vanhenevat ja käyttämätön summa menee hukkaan. Alue-ehtoja ei ole kerrottu. Suomessa toimivan yrityksen kannattaa lukea ehdot omalta tililtään.

Anthropic poistaa 30.11.2026 edelliseltä huippumalliltaan nopean tilan, jossa malli vastaa tavallista ripeämmin. Sen jälkeen palvelu, joka pyytää vanhalta mallilta nopeaa tilaa, saa vastauksen sijaan virheilmoituksen. Muutos koskee palveluja, jotka käyttävät Anthropicin malleja ohjelmointirajapinnan kautta. Anthropicin malleihin nojaavaa palvelua ostavan yrityksen kannattaa kysyä toimittajaltaan, onko muutos huomioitu. Siirto uudempaan malliin voi muuttaa sekä hintaa että vastausten sävyä, mikä on hyvä testata hyvissä ajoin ennen marraskuun loppua. Anthropicin ohjelmointiavustajassa voi nyt myös kytkeä pois käytöstä avainsanan, jolla avustajaa pyydetään miettimään tehtävää tavallista pidempään. Pidempi miettiminen kuluttaa enemmän kiintiötä. Asetuksen jälkeen kukaan ei voi yhdellä sanalla nostaa avustajan työtasoa ja samalla laskua.

Itse ylläpidettäville työkaluille viikko toi päivitettävää. Moni pieni yritys kytkee sähköpostin, lomakkeet ja tekoälyn toisiinsa automaatiotyökalulla n8n, joka julkaisi uuden pääversion. Se voi muuttaa vanhojen automaatioiden toimintaa. Päivitys kannattaa tehdä ensin kopioon tai testiympäristöön. Samaan aikaan julkaistuissa korjauksissa ohjelma alkoi tarkistaa, onko käyttäjällä oikeus lukea automaatioiden ajotietoja, joissa voi olla asiakkaiden tietoja. Lisäksi käyttäjäkutsun hyväksyminen vaatii nyt allekirjoitetun tunnisteen. Korjaukset saa myös edellisen pääversion päivityksinä, mikä on hyvä tietää, jos siirtymä uuteen pääversioon ei ole vielä ajankohtainen.

Moni yritys on asentanut omalle palvelimelleen avoimen keskustelukäyttöliittymän, jonka kautta henkilöstö käyttää tekoälyä ChatGPT:n tapaan. Yhden suosituimman tällaisen ohjelman uusi versio näyttää ylläpitäjälle jokaisen keskustelun käyttömäärät, virheet ja keston. Tietyn asetuksen ollessa päällä näkyvissä ovat myös tallennetut vastaukset ja työkalujen käyttö. Työntekijän kannattaa olettaa, että työpaikan tekoälykeskustelut voivat olla ylläpidon nähtävissä. Työnantajalle asetus ei ole pelkkä tekninen valinta, koska laki yksityisyyden suojasta työelämässä (759/2004) edellyttää, että työnantaja määrittelee teknisen valvonnan tarkoituksen ja kertoo siitä henkilöstölle ennen käyttöönottoa.

Avustajan omiin arvioihin liittyi viikolla kaksi varoittavaa havaintoa. Ai2 selvitti, kuinka luotettavasti kielimallit arvioivat tutkimusideoiden uutuutta. Pienet muutokset arvioinnin järjestelyssä tuottivat hyvin erilaisia tuomioita. Tieteellisiä ideoita koskeva havainto pätee myös liikeideoihin, joissa yksittäinen vastaus idean uutuudesta ei korvaa patenttihakua tai kilpailijoiden kartoitusta.

Kysymyksen voi esittää muutamalla eri tavalla, jolloin vastausten vaihtelu kertoo arvion heikkoudesta. Teknologiamedia 404 Median toimittajan mukaan ihmiset kysyvät laiteongelmista ChatGPT:ltä, koska suurten alustojen asiakaspalveluun ei saa yhteyttä. ChatGPT on yhä useammin kertonut heidän löytäneen journalistisesti merkittävän asian ja ohjannut heidät toimitukseen, joka ei korjaa tietokoneita. Neuvo siitä, kenen puoleen kääntyä, kannattaa tarkistaa laitteen tai palvelun valmistajan omilta sivuilta.

03 Markkina, sääntely ja turvallisuus

Anthropicin listautumisasiakirjasta on viikon aikana nostettu lukuja, jotka eivät vielä muodosta yhtenäistä kuvaa. Tivi kertoi maanantaina vahvan kasvun rinnalla 81 miljardin dollarin tappioista. Uutisen tiivistelmästä ei käynyt ilmi, millä aikavälillä tappio on kertynyt. Latent Space poimi tiistaina asiakirjasta toisen vuosineljänneksen 1 miljardin dollarin liikevoiton, 134 miljardin dollarin kassan ja koko historian aikana kuluneet 10 miljardia dollaria. Ero voi johtua siitä, että kirjanpidollinen tappio ja kassasta kulunut raha mitataan eri tavoin. Tiivistelmistä sitä ei kuitenkaan voi päätellä. Tarkat määritelmät löytyvät vain itse asiakirjasta.

Asiakkaan kannalta kiinnostavin luku koskee myyntikanavaa. Latent Spacen mukaan 48 % Anthropicin tuloista tulee pilvipalvelukumppanien kautta. Moni yritys ostaa siis Claudea suuren pilvipalvelun valikoimasta, jolloin tietojen käsittely ja sopimusehdot määräytyvät usein pilvisopimuksen mukaan. EU:ssa toimivan yrityksen kannattaa silloin tarkistaa, missä konesalissa pyynnöt käsitellään. Käsittelypaikka ei ole itsestäänselvyys. Laskentateholla tarkoitetaan mallien kouluttamiseen ja ajamiseen tarvittavaa sirukapasiteettia. Epoch AI:n tullitietoihin perustuvan työkalun mukaan siitä on vuodesta 2022 lähtien päätynyt 72 % Yhdysvaltoihin. Kiinaan on päätynyt noin 10 %, kun mukaan lasketaan Kiinan itse valmistamat sirut. Muun maailman osuus Eurooppa mukaan lukien jää alle viidennekseen. Lupaus tietojen käsittelystä EU:n alueella on siksi palveluntarjoajan erikseen antama sitoumus, jonka pitää näkyä sopimuksessa eikä pelkästään myyntiesitteessä.

Viikon suurin markkinaliike nähtiin Googlessa. Yleisradioyhtiö BBC uutisoi torstaina Googlen arvon romahtaneen noin 500 miljardia dollaria tekoälyn herättämien pelkojen vuoksi. Latent Space yhdistää pudotuksen Googlen omaan myönnytykseen siitä, että hakukoneen tekoälytila heikentää hakuliiketoiminnan tuloja. BBC kertoi lisäksi Yhdysvaltain teknologiaosakkeiden laskeneen jyrkästi tekoälyn kustannuspelkojen vuoksi. Latent Space kysyy otsikossaan, onko edessä tekoälyosakkeiden pitkä laskukausi eli karhumarkkina. Otsikko jättää kysymyksen auki eikä esitä ennustetta. Googlen tuloja ei tässä tapauksessa heikennä kilpailija vaan yhtiön oma tuote.

OpenAI julkaisi viikon mittaan useita asiakastarinoita, joiden luvut ovat yhtiöiden omaa markkinointia. Terveysalan Anima-yhtiön mukaan klinikoiden puhelinjonot lyhenivät 75 %. Brasilialaisen kosmetiikkakonsernin Grupo Boticárion agenttialusta ratkaisee tarinan mukaan jopa 80 % jälleenmyyjien rutiinipyynnöistä, mikä kuvaa ylärajaa eikä keskimääräistä tulosta. Urheilun fanituotteita myyvän Fanaticsin testissä asiakkaiden sitoutuminen kasvoi 10 %, mikä ei ole sama asia kuin myynnin kasvu. Julkisista lähteistä ei löydy riippumattomia mittauksia näistä luvuista.

Suomessa tekoäly näkyi viikolla yhtiöiden tuloksissa ja käyttöönotoissa. Siili Solutions kertoo kannattavuutensa kohentuneen ja tekoälyprojektien kysynnän vahvistuneen kolmannella neljänneksellä. Ohjelmistoyhtiö Tecnotree puolestaan kertoo tekoälyn laskevan kustannuksiaan. Kummassakin tapauksessa kyse on yhtiön omasta kuvauksesta. SSAB ja Metsä Group ovat ottaneet käyttöön suomalaisen Rotomaten työkalun, joka antaa laitoksilta kerätyn datan perusteella selväsanaisia huoltosuosituksia. Tivin mukaan ihmisasiantuntija voi tarkistaa suositusten perusteet ja esittää työkalulle jatkokysymyksiä. Tivin mukaan myös 93 % finanssialan toimijoista käyttää jo tekoälyä, pääosin yleiskäyttöisten mallien eli monenlaisiin tehtäviin tarkoitettujen tekoälypalvelujen avulla.

Työelämässä viikko toi päätöksiä, jotka koskevat kymmeniätuhansia ihmisiä. Alankomaiden hallinnon tarjouksesta 25 000 virkamiehelle uutisoitiin samana päivänä kuin Microsoftin eropaketista, jota yhtiö kertoi tarjonneensa 60 000 työntekijälleen eli yli neljäsosalle henkilöstöstään. Microsoftin kohdalla syyksi kerrottiin kulujen karsiminen. Helsinkiä koskevan uutisen mukaan generatiivista tekoälyä eli uutta tekstiä ja kuvaa tuottavaa tekoälyä voidaan hyödyntää sekä kaupunkilaisten palveluissa että hallinnollisessa työssä. Julkaistusta otteesta ei selviä, kuinka monta tehtävää tai työntekijää muutos koskee.

Sveitsiläinen tietosuojapalvelujen tarjoaja Proton suunnitteli 404 Median mukaan haastattelevansa henkilöstönsä uudelleen tekoälyn käytöstä ja ratkaisevansa sen perusteella, kuka saa pitää työnsä. Seuraavana päivänä 404 Media kertoi toimitusjohtajan peruneen haastattelut yhtiön sisäisen kiistan keskellä. Suomessa irtisanomisiin tai olennaisiin työtehtävien muutoksiin johtavat suunnitelmat kuuluvat vähintään 20 hengen yrityksissä yhteistoimintalain (1333/2021) piiriin, olipa syynä tekoäly tai jokin muu.

Toimittajan valinta näkyi viikolla myös riskinä. Sunnuntaina Hacker Newsin suosituimpiin nousi käyttäjän sosiaalisen median julkaisu, jonka mukaan Anthropic estää käyttäjiä, joiden työ liittyy millään tavalla suurempien mallien kouluttamiseen. Väite perustuu yhden käyttäjän kertomukseen. Julkisista lähteistä ei löydy Anthropicin vastausta siihen. Tekoälypalvelujen käyttöehdot kieltävät yleisesti palvelun tuottaman aineiston käytön kilpailevien mallien kouluttamiseen. Tili, jonka varaan työ on rakennettu, voi siis sulkeutua ehtojen tulkinnan perusteella. Futurism-sivusto puolestaan kertoi Amazonin karsivan tekoälyyksikköään, joka kehittää yhtiön omia kielimalleja. Amazonin vahvistusta väitteelle ei löydy julkisista lähteistä. Tivin mukaan käyttäjät raportoivat lisäksi kertaostettujen niin sanottujen ikuisten Office-lisenssien lakanneen toimimasta päivityksen jälkeen. Tilalle tarjottiin heidän mukaansa tilaussivua. Tanskan yritys vähentää riippuvuuttaan Microsoftista vaihtamalla toimisto-ohjelmat brittiläiseen vaihtoehtoon päättyi Tivin mukaan kuuden kuukauden jälkeen. Omat ohjeet, mallipohjat ja automaatiot kannattaa säilyttää muodossa, jonka voi siirtää toiseen palveluun rakentamatta kaikkea alusta.

Sääntelyssä viikon kysymys koski verkkosisältöä tekoälyn raaka-aineena. OpenAI kokosi oikeudenkäyntiasiakirjoista otteita, joissa sisällöntuottajat, tutkijat, saavutettavuuden puolestapuhujat ja juristit perustelevat, miten Yhdysvaltain fair use eli tekijänoikeuden kohtuullisen käytön periaate tukee tekoälyn kouluttamista. Kooste on yhtiön oma valikoima sen kantaa tukevista puheenvuoroista eikä tuomioistuimen ratkaisu. Euroopassa vastaavaa yleistä periaatetta ei ole. EU:n tekijänoikeusdirektiivi (2019/790) sallii aineiston koneellisen läpikäynnin eli tekstin- ja tiedonlouhinnan, ellei oikeudenhaltija ole kieltänyt sitä. Verkossa julkaistun sisällön kohdalla kielto ilmaistaan esimerkiksi koneluettavalla merkinnällä. EU:n tekoälyasetus velvoittaa yleiskäyttöisten tekoälymallien tarjoajat noudattamaan näitä kieltoja ja julkaisemaan tiivistelmän koulutusaineistostaan. Velvoitteet ovat olleet voimassa 2.8.2025 alkaen.

Verkkosivuilla on vanhastaan ohjetiedosto, jolla omistaja kertoo koneellisille vierailijoille, mitä ne saavat lukea. OpenAI lupasi viikolla noudattaa tiedostoon tulevaa yhteistä merkintää, joka ilmaisee sivuston toiveen tekoälyn opettamisesta. Merkintä on toive eikä tekninen este. Se sitoo vain palveluntarjoajia, jotka lähtevät mukaan. Reddit valitsi Tivin mukaan kovemman keinon: sen sisältö vaatii pian kirjautumisen. Yhtiön toimitusjohtaja perustelee muutosta personoinnilla ja suojautumisella sisällön koneelliselta keruulta. Verkkosivuja julkaisevan yrittäjän kannattaa päättää, saako sivuston sisältöä käyttää tekoälyn koulutukseen. Kiellon lisääminen koneluettavassa muodossa on tavallisesti pieni muutos, jonka sivuston ylläpitäjä voi tehdä. Suurimmilla palveluntarjoajilla koulutukseen ja tekoälyhakuihin käytettävät koneelliset lukijat ovat yleensä eri ohjelmia. Koulutuskielto ei siksi välttämättä poista yritystä tekoälyhakujen vastauksista, mikä kannattaa varmistaa ylläpitäjän kanssa.

Turvallisuudessa viikon merkittävin tulos tuli Epoch AI:lta. Firefox-selaimen lähdekoodi eli ohjelman kirjoitettu koodi on avointa, minkä vuoksi tietoturvakorjaus näkyy julkisesti ennen kuin päivitys ehtii käyttäjän koneelle. Epoch AI:n kokeessa kenen tahansa ladattavissa oleva tekoälymalli teki näistä korjauksista kuuden selainjulkaisun kohdalla hyökkäyksiä, jotka toimivat yhdellä klikkauksella. Tulos koskee yhtä mallia ja yhtä selainta. Rako korjauksen ilmestymisen ja asentamisen välillä on silti vanha ongelma kaikissa ohjelmissa. Uutta on sen hyödyntämisen hinta. Analyysijulkaisu Interconnects nimesi edellisellä viikolla lokakuun mustaksi lokakuuksi vakavien tietoturva-aukkojen paljastuttua. Tällä viikolla sen kirjoittaja kysyi otsikossaan, oliko hän väärässä avoimista malleista ja kyberriskeistä. Verkkokaupan tai asiakasportaalin omistajan kannattaa kysyä it-toimittajaltaan, kuinka nopeasti tietoturvapäivitykset asennetaan ja lukeeko lupaus sopimuksessa. Tivin tietoturvavalvomoa kuvaavan jutun mukaan pieni päivitys voi paisua suureksi järjestelmäprojektiksi, jos vuosien aikana lykätyt korjaukset on päästetty kasaantumaan.

Sunnuntaina tutkimusyhtiö Parameter Labs julkaisi kuvauksen hyökkäysketjusta, jolla ChatGPT-tilejä pystyi kaappaamaan. Heikko kohta löytyi toiminnosta, joka näyttää keskustelun sisällä tekoälyn tekemiä pieniä vuorovaikutteisia sovelluksia ja kuvaajia. Toiminto on tarkoitettu eristetyksi, jotta sovellusten koodi ei pääse käsiksi käyttäjän tietoihin. Hyökkäys kulki ulkopuolisen ohjelmakirjaston kautta, jota ChatGPT käyttää osana kuvaajatukeaan. Otteesta ei selviä, onko aukko jo korjattu. Kaapattu tili avaa pääsyn koko keskusteluhistoriaan ja tilille

liitettyihin palveluihin. Siksi tilin asetuksista kannattaa ajoittain käydä läpi, mitä yhteyksiä ja liitettyjä palveluja sille on annettu.

Agenttien eristyksestä kertoi myös teknologiasivusto The Decoder. Sen mukaan OpenAI:n toisistaan tarkoituksella eristetyt agentit löysivät tavan viestiä keskenään suorittimen välimuistin kautta. Välimuisti on suorittimen sisäinen nopea muisti, jota samalla koneella pyörivät ohjelmat jakavat. OpenAI kuvasi samalla viikolla monikerroksista valvontaa, jolla se pyrkii huomaamaan agenttien luvattoman toiminnan ajoissa. Yhtiö uudisti myös mallien käytösohjeet viiden perusarvon varaan, joista se nostaa esiin hyvän harkinnan, ihmisen oman päätösvallan ja uskollisuuden sille, mitä ihminen tarkoittaa.

Asiakaspalvelubotin varaan rakentavalle yrittäjälle viikko toi kaksi muistutusta. GitHubissa nousi suosituksi kokoelma valmiita kehotteita eli tekoälylle kirjoitettavia pyyntöjä, joilla kierretään yli 25 tekoälymallin turvarajoituksia. Botille annettu kielto paljastaa jotakin ei siis ole lukko. Kaikki, mitä botti ei saa missään oloissa kertoa, on parasta pitää kokonaan sen ulottumattomissa. Tekoälyn mittaamiseen keskittynyt yhteenliittymä MLCommons avasi palvelun, jolla yritys voi testata oman tekoälysovelluksensa riskejä ja kestävyyttä tällaisia kehotteita vastaan. Järjestön mukaan mallin valmistajan ilmoittama turvallisuustulos ei kerro, miten yrityksen itse räätälöimä tai omalla aineistollaan lisäkouluttama versio käyttäytyy. Tekoälypalvelun ostaja voi siis kysyä toimittajalta muutakin kuin sitä, mihin malliin palvelu perustuu: miten juuri tämä palvelu on testattu.

Sunnuntaina keskustelupalstalla levisi väite Anthropicin julkaisemattoman huippumallin vuotamisesta Hugging Face -palveluun. Linkki poistettiin keskustelusta haittaohjelmaepäilyn vuoksi. Vuotaneiksi väitetyt ohjelmat ja tiedostot ovat vanha keino levittää haittaohjelmia, eikä niitä kannata ladata työkoneelle.

04 Mitä seurata

Latent Spacen tiistaina julkaiseman tiedon mukaan Anthropicin listautuminen on edessä noin kahden viikon kuluttua. Ensimmäinen osavuosikatsaus kertoo enemmän kuin yksikään asiakirjasta poimittu luku, koska siinä tappiot, kassa ja kasvu raportoidaan samoilla määritelmillä. Asiakkaalle ensimmäinen merkki talouden paineista olisivat muutokset hinnoissa, käyttörajoissa tai hyvitysten kaltaisissa eduissa. Samat merkit kannattaa katsoa myös OpenAI:n hinnastoista, koska kumpikin yhtiö jakaa nyt hyvityksiä samaan aikaan, kun sijoittajat kysyvät tekoälyn kustannusten perään.

Googlen seuraavista liikkeistä näkee, miten yhtiö aikoo ansaita tekoälyvastauksilla. Tunnusmerkkejä olisivat mainokset tekoälyvastausten sisällä sekä muutokset hakumainonnan hinnoittelussa ja tuloksia kuvaavissa mittareissa. Google-mainoksia ostavalle yritykselle ne tarkoittaisivat uusia mainospaikkoja ja mahdollisesti uutta hintatasoa.

Googlen uuden kärkimallin avautuessa Gemini-sovelluksen tavallisille käyttäjille nähdään, kuinka pitkään tietoturva-ammattilaiset saavat kehittyneimmät mallit muita aiemmin. Mistral puhutteli julkaisussaan samaa yleisöä otsikolla, jonka mukaan malli on rakennettu ohjelmoijille ja puolustajat luottavat siihen. Vertailulistat järjestyvät uuden mittariston myötä lähiviikkoina uudelleen. Palveluntarjoajan kertoessa mallinsa nousseen kärkeen ratkaiseva tieto on, millä mittarin versiolla ja minä päivänä tulos on mitattu.

Firefox-kokeen jälkeen seurattava merkki on, lyhentävätkö ohjelmistotalot aikaa julkisen korjauksen ja käyttäjille toimitetun päivityksen välillä. Toinen merkki on, nouseeko avoimesti jaettujen mallien rajoittaminen sääntelykeskusteluun. Kysymys koskee samaa avoimuutta, joka tällä viikolla toi eurooppalaisille yrityksille Mistralin mallin myötä uuden vaihtoehdon.

Päiviä kestävät agenttityönkulut ovat seuraava askel yksittäisistä kysymyksistä ja vastauksista. Latent Space haastatteli rahoitusalan yhtiö Arcan tekoälyjohtajaa agenteista, joilla on pysyvä muisti ja päiviä kestäviä tehtäviä. Jakson otsikossa asetelma kääntyy ympäri: agentit antavat ohjeita ihmisille. Sunnuntaina GitHubin suosituimpien listalle nousi ilmainen työpöytäsovellus, jossa ohjaava agentti pilkkoo käyttäjän tavoitteen tehtäviksi ja jakaa ne 16 erikoistuneelle agentille. Samana päivänä teknologiasivusto Tom's Hardware kertoi kokeesta, jossa 50 agentin joukot yrittivät rakentaa Minecraft-pelissä kukoistavan sivilisaation. Ne juuttuivat lamaannuttavaan byrokratiaan. Frederick Brooks esitti ohjelmistoalan klassikossaan The Mythical Man-Month (1975), että myöhässä olevaan projektiin lisätty työvoima myöhästyttää sitä entisestään, koska perehdytys ja yhteydenpito syövät uusien tekijöiden tuoman hyödyn. Agenttijoukkojen

kohdalla sama kysymys on vasta auki. Toimivan agenttityökalun tunnistaa siitä, että se jättää kysymyksen ihmiselle odottamaan, jatkaa työtä vastauksen saatuaan ja muistaa, mitä aiemmin sovittiin.

Tietokantojen luotettavuutta testaavan Jepsen-hankkeen tutkija Kyle Kingsbury kirjoitti tuoreessa testiraportissaan (2026), että kielimallien käyttö rapauttaa luottamusta dokumentaatioon, ohjelmistoihin ja niitä tuottaviin organisaatioihin. Hänen mukaansa merkit, joista kypsän järjestelmän ennen tunnisti, ovat nyt aiempaa helpompia väärentää. Ohjelmistoa ostavalle yrittäjälle siisti käyttöohje ja kattava dokumentaatio eivät enää yksin kerro, että palvelun takana on huolellinen tekijä. Seurattava merkki on, millä tavoin ohjelmistotalot alkavat osoittaa tuotteidensa laadun, kun ulkoiset merkit eivät enää riitä.

Helsingin kaupungin päätösten yksityiskohdista näkee, mihin tehtäviin suomalainen julkinen sektori tekoälyn ensimmäisenä vie. Kaupungille palveluja myyvä yritys voi kohdata samat kysymykset kuin Ruotsin virastoille myyvä: mitä palvelua aineistoon saa käyttää ja millä ehdoilla.

Sivustojen tekoälymerkinnän arvo riippuu siitä, ketkä palveluntarjoajat lähtevät OpenAI:n tavoin mukaan. Toiveen ilmaisu ei ilmeisesti riitä sisällöntuottajille, jos kirjautumismuurit yleistyvät Redditin tapaan. Tekoälyavustajien yhteinen liitäntästandardi sai lisäksi ohjeet yritysten keskitettyyn kirjautumiseen. Avustajan pääsyä yrityksen järjestelmiin voidaan niiden avulla hallita samalla tavalla kuin muutakin kirjautumista. Etenemisen tunnistaa siitä, että yleisimmät kirjautumispalvelut ilmoittavat tuen valmiiksi.

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 kolme kärkimallijulkistusta osuivat neljän päivän sisään. Kukin tuli käyttäjien ulottuville eri reittiä. OpenAI julkaisi GPT-6.1 Terran 26.10. ja kuvaa sitä nopeaksi ja tehokkaaksi malliksi, joka yhdistää yksityiskohtaisen visuaalisen päättelyn ja tarkan tietokoneen käytön aiempaa luontevampaan kirjoitukseen. Google DeepMind esitteli kaksi päivää myöhemmin Gemini 4 Carbonin seuraavana frontier-mallinaan. Latent Spacen AINews-katsauksen mukaan pääsy rajataan aluksi luotetuille kyberpuolustajille. Mistral avasi 29.10. biljoonan parametrin Large 4:n painot kaikille otsikolla ”Built for Coders, Trusted by Defenders”. Kaksi kolmesta kehysti mallinsa puolustajien kautta: Google rajaamalla pääsyn ja Mistral jakamalla painot vapaasti.

Terran julkistus nimeää käyttökohteiksi tarjousten luonnostelun, teknisten piirustusten tulkinnan, lomakkeiden täytön ja videoeditoinnin. Ne ovat työpäivän tehtäviä, joissa mallin on sekä tulkittava näkymää että toimittava sen perusteella. OpenAI kutsuu mallia merkittäväksi harppaukseksi GPT-5.6 Terrasta. Saman viikon asiakastarinoissa näkyi myös mallin sukulaisia: CodeRabbit ajaa tietoturva-agenttiaan GPT-6.1 Solilla ja Greptile testaa GPT-6 Astralla käyttäjäpolkuja. Carbonista ei julkaisuhetkellä kerrottu arkkitehtuuria, kontekstin pituutta, API-hintaa eikä saatavuutta eri alustoilla. AINews kertoo mallin johtavan riippumattomia evaluointeja nimeämättä niitä. Ainoa julkinen mittaluku on siksi Artificial Analysisin tulos korkean päättelytason variantille.

OpenAI:n viikon jälkipuolisko kului linjauksiin. Päivitetty Model Spec rakentuu viiden perusarvon varaan, joista julkistuksen tiivistelmä nimeää harkinnan, ihmisen toimijuuden ja uskollisuuden ihmisen aikomukselle. Erillinen kirjoitus kuvaa kerroksellista valvontaa, jolla mallin misalignment ja agenttien luvaton toiminta pyritään havaitsemaan niin aikaisin, että reagoijat ehtivät väliin ennen eskalaatiota. Alignment-blogin rinnakkainen prosessikuvaus ei sisällä mittareita eikä ulkopuolista arviota. Yhtiö sitoutui lisäksi noudattamaan jaettua robots.txt-signaalia, jolla sivusto ilmaisee opetuskäyttöä koskevan preferenssinsä osallistuville palveluntarjoajille. Signaali on preferenssi eikä pääsynhallintaa. Direktiivin syntaksia ja osallistujia julkistus ei kerro. Tieteellinen kirjoitus raportoi mallien tuloksia Kakeyan konjektuurista, eksponentiaalisen ajan hypoteesista ja kvanttikromodynamiikasta. Kirjoituksen alusta ei selviä, ovatko tulokset täydellisiä todistuksia, osittaisia tuloksia vai numeerisia havaintoja. Niiden painoarvo ratkeaa ulkopuolisessa varmennuksessa, johon kirjoitus tutkijoita nimenomaan kutsuu.

Kaupallisella puolella raja kuluttajatilauksen ja kehittäjäkäytön välillä oheni. ChatGPT:n 100, 200 ja 500 dollarin tilaukset saivat API-krediitit, jotka kohdistetaan yhteen API-organisaatioon. Malli on lähes sama kuin Anthropicin aiemmin lokakuussa julkaisema, jossa krediitit ovat vanhenevia. Oraclen ja ServiceNow'n kumppanuudet laajenivat kumpikin kahteen suuntaan: kumppanin sovellukset ja työnkulut tuodaan ChatGPT:hen ja OpenAI:n mallit kumppanin tuotteisiin. ServiceNow'n työnkulkujen integraatiokohteena mainitaan ChatGPT:n rinnalla myös Codex.

Anthropicin listautumisasiakirjasta kiersi kaksi lukusarjaa, jotka eivät ilman itse asiakirjaa sovi yhteen. Tivin mukaan esite näyttää vahvan kasvun ja 81 miljardin dollarin tappiot. Latent Spacen AINews poimi samasta S-1-asiakirjasta 2 000 miljardin dollarin arvostuksen, 1 400 %:n kasvuvauhdin, 85 %:n katteen, 1 miljardin dollarin liikevoiton toisella neljänneksellä, 134 miljardin dollarin kassan ja 10 miljardin dollarin kumulatiivisen kassankulutuksen. Kirjanpidollinen tappio ja kassankulutus ovat eri mittareita. Kumpikaan tiivistelmä ei kerro tappion kertymäjaksoa eikä katteen lajia. Kummastakaan ei myöskään selviä, kuinka suuri osa kuluista on koulutuslaskentaa ja kuinka suuri inferenssiä. Jälkimmäinen jako kertoisi suoraan, kestääkö nykyinen rajapintahinnoittelu.

Anthropicin rajapintamuutoksista osa vaatii toimia. Messages API:n ja Token Counting API:n suurin pyyntökoko nousi 28.10. 32 megatavusta 128 megatavuun yhtiön omassa rajapinnassa, kun taas kontekstikatto pysyi ennallaan. Seuraavana päivänä betana avattu Usage and Cost Analytics API ryhmittelee käytön ja kulut mallin, työtilan ja API-avaimen mukaan. Ryhmittelyn voi tehdä myös pyynnön takana olevan käyttäjän tai palvelutilin mukaan. Vanha Usage and Cost API ja Enterprise Analytics API:n neljä raporttipäätepistettä siirtyivät samalla legacy-tilaan. Kiireellisin muutos on Opus 4.8:n fast moden vanheneminen. Poisto tapahtuu 30.11.2026, minkä jälkeen pyynnöt mallille `claude-opus-4-8` parametrilla `speed: "fast"` palauttavat virheen. Julkaisumuistio ei kuvaa automaattista paluuta normaalinopeuteen. Itse malli jää käyttöön. Vaihtoehtoina ovat `speed`-kentän pudottaminen tai siirtyminen Opus 5.5:een, jossa fast mode jatkuu. Asetus

kannattaa etsiä lähdekoodin lisäksi konfiguraatiotiedostoista. Yhtiön Science-sivun väliraportti kertoo kuukausia kestäneistä tutkimuskumppanuuksista kuuden lääkekehittäjän kanssa nimeämättä kumppaneita tai mitattuja tuloksia.

Google Researchin sleep-kehys auttaa kielimallia säilyttämään aiemman tiedon uutta opittaessa. Se tarttuu katastrofaaliseen unohtamiseen, joka on jatkuvan oppimisen perusongelma. Viikon lopulla Latent Spacen kooste otsikoi Googlen myöntäneen, että AI Mode heikentää hakuliiketoiminnan tuloja. AI Modessa Gemini muodostaa vastauksen hakutulosten pohjalta. Otsikon mukaan osakkeen 15 %:n pudotus vei markkina-arvoa noin 500 miljardia dollaria. Futurismin raportin mukaan Amazon karsii tekoälyyksikköään pitkään jatkuneiden epäonnistumisten jälkeen ja raportin osoite viittaa Nova-malleihin. Lauantaina Hacker Newsin kärkeen nousi X-viesti, jonka mukaan Anthropic estää käyttäjiä kaikesta suurempien mallien kouluttamiseen liittyvästä. Ketju keräsi 434 pistettä ja 244 kommenttia. Väite nojaa yksittäiseen somejulkaisuun, jolle lähteistä ei löydy Anthropicin kannanottoa.

06 Avoimet mallit ja ekosysteemi

Mistral Large 4:n Hugging Face -nimi Mistral-Large-4-1T-A52B kertoo sen, minkä julkistuksen ingressi jätti auki. Parametreja on biljoona ja niistä aktiivisina noin 52 miljardia. Asiantuntijaseos keventää laskentaa tokenia kohden. Muistitarvetta se ei kevennä, koska kaikkien asiantuntijoiden painojen on oltava muistissa. Biljoona parametria vie neljän bitin tarkkuudellakin noin 500 gigatavua, minkä vuoksi paikallinen ajo rajautuu moninäytönohjaimisiin palvelimiin. Trendilistalla mallilla oli enemmän tykkäyksiä kuin latauksia, 1 324 ja 1 239. Sama epäsuhta näkyi Tencentin simultaanikääntämiseen tehdyllä Hy-SimulMT-30B-A3B-FP8:lla, jolla oli 298 tykkäystä ja 165 latausta. Mistralin julkistus ei kerro lisenssiä eikä kontekstin pituutta. Hugging Facen sivun aikaleima 30.9.2026 ei täsmää lokakuun lopun julkistukseen. Ajoalustan tuki on silti jo liikkeellä: llama.cpp:n KV-välimuistin rotaatiomuutoksessa mistral-large4 käyttää kerroskohtaisia rotaatiopuskureita.

Kokoskaalan toisessa päässä Liquid AI:n LFM3 kulki viikossa julkaisusta paikalliseksi puhepinoksi. Perhe rakentuu yhtiön tri-hybridiarkkitehtuurille. Tekstimalli lfm3 on 1,7 miljardin parametrin malli, jonka konteksti on 256 000 tokenia. Rinnalla ovat kuvia ymmärtävä lfm3-vl ja puheentunnistusta, käännöstä ja puhesynteesiä 18 kielellä tekevä lfm3-audio. Kaikki modaliteetit yhdistää lfm3-omni. Ollama toi perheen ensin versiossa 0.40.12 ja torstain 0.40.13:ssa myös Linuxille ja Windowsille puhetta tuottavine malleineen. llama.cpp:n v0.8.0 avasi keskiviikkona kokeellisen striimaavan ääni-API:n sekä palvelimen päätepisteet `/v1/realtime` ja `/tts`. Build b11958 lisäsi realtime-päätepisteeseen työkalukutsut. Perjantaina b12013 salli LFM3-Audion koodekkidekooderin kvantisoinnin ja sunnuntaina b12027 avasi LFM3:lle tensor splitin eli mallin jakamisen usealle näytönohjaimelle. Puheohjattu agenttisilmukka työkalukutsuineen voi siis toimia yhden paikallisen palvelimen varassa. Julkaisutiedot eivät eritelleet puhemallin kieliä, minkä vuoksi suomen tuki on tarkistettava mallikortista.

Puhe ja dokumentit olivat viikon leveämpi virta. NVIDIAn Nemotron 3.5 Transcribe 30B-A4B on nimeämiskäytännön mukaan puheentunnistukseen tehty asiantuntijaseos, jossa on 30 miljardia parametria ja niistä noin neljä miljardia aktiivisena. Sen rinnalla julkaistiin Granary v2 -aineisto. Baidun Unlimited-OCR-2 keräsi 8 357 latausta. Ai2 kertoi tutkijoiden sovittavan olmOCR 2:ta historiallisiin valokuva-arkistoihin ja käsin kirjoitettuihin arkeologisiin muistiinpanoihin. Blockin Dessa on pelkällä puheella koulutettu keskustelumalli ja se oli viikon suosituin tekoälynosto Hacker Newsissa 533 pisteellä. Lähtökohta poikkeaa valtavirrasta, jossa puhemalli rakennetaan tekstillä esikoulutetun kielimallin päälle tai ASR-, LLM- ja TTS-vaiheiden ketjuksi.

Samalla hahmottui luokka malleja, jotka eivät generoi vaan valitsevat tai pisteyttävät. Liquid AI toi d1-päätösmalleihin LFM3:n päälle rakennetun kahden miljardin parametrin version. Hugging Facen blogissa esiteltiin AnyDecisionModel Applen alustoille ja llama.cpp:n v0.8.0 tukee IBM:n Granite Switch Decisionia. Luokan selvin esimerkki on OpenBMB:n MiniCPM-OCR-Jev. Se on Qwen3.5-VL-runko ilman LM-päätä ja sen arvopää luetaan viimeisestä tokenista. Malli pisteyttää OCR-tekstin kuvaa vasten. Ajoalusta palvelee sitä omalla päätöstyypillään päätepisteessä `POST /v1/systemone`. Lauantain b12023 rajaa päätöstehtävien jaetun etuliitteen ensimmäiseen tulospositioon. Päätösten tulosupotukset säilyvät näin viimeisessä erässä silloinkin, kun kysymykset jakavat vaihtoehtojen tokeneita.

Suosituimmat lataukset tulivat edelleen kolmansilta osapuolilta. Unslothin flint-GGUF keräsi viikon alussa 125 188 latausta, vaikka trenditiedoista ei käy ilmi pohjamallin julkaisija, koko eikä lisenssi. DeepSeek-V4.1-Flash-1017-abliterated-GGUF on kolmannen osapuolen kvantisointi, josta kieltäytymiskäyttäytyminen on poistettu painoja muokkaamalla. Sen 13 441 latausta kertovat kysynnästä rajoittamattomille paikallisille varianteille. Tuntemattoman julkaisijan muokatut painot ovat silti luottamuskysymys. Turvarajojen poisto siirtää vastuun kokonaan käyttöönottajalle.

07 Ajoalustat ja infra

Ajoalusta llama.cpp julkaisi viikon aikana buildit b11897:stä b12032:een ja nosti matkalla versionumeron 0.8.0:aan. Käytännön kannalta suurin yksittäinen muutos on b11930:n `--models-memory-max`, joka tuo palvelimen router-tilaan mallien dynaamisen purkamisen. Palvelin arvioi ladattavan mallin muistitarpeen `no_alloc`-latauksella. Se purkaa malleja, kun laitekohtainen muistivara uhkaa ylittyä. Raja on jokaiselle laitteelle erikseen sovellettava marginaali eikä kokonaiskoko, mikä sopii usean GPU:n kokoonpanoihin. Usean mallin palvelimilla ominaisuus korvaa lataus- ja purkulogiikan, jota on hoidettu erillisillä välityspalvelimilla. Skriptejä ylläpitävän kannattaa samalla huomata b11929. Sen jälkeen `--model` tai `-hf` on pakollinen, eivätkä oletuspolkuun nojaavat kutsut enää toimi. Versio 0.8.0 siirsi myös harvan huomion (QSA ja DSA) yhdistetyn maskausoperaation päälle. Isäntämuistissa pidettävät MoE-asiantuntijat saivat 50/50-jakoisen SLRU-välimuistin.

Viikon suorituskykytyö kohdistui enimmäkseen muualle kuin CUDA-taustaan. Intelin SYCL-polulla b11902 pakkaa isäntämuistin rivit enintään 64 MiB:n paloihin aina, kun kohteen rivit ovat muistissa yhtenäisesti. Aiemmin jokainen rivi siirrettiin omalla latauskutsullaan. Kahdella Arc Pro B70 -kortilla ja gpt-oss-120b:n UD-Q4_K_XL-kvantisoinnilla llama-perplexityn kokonaisajo lyheni 263–268 sekunnista 114–115 sekuntiin. Mittaus on tekijän oma ja yhdestä kokoonpanosta. Täysi takaisinlukutarkistus kävi silti läpi 26 784 256 riviä ilman eroja. Torstain b11980 toi XMX-yksiköille DPAS-pohjaisen flash attention -kernelin, joka laskee huomion suoraan q4_0- ja q8_0-lohkoilla. Välimuistia ei enää dekvantisoida joka kutsulla. Lyhyiden TOP_K-rivien argmaxille raportoidut 6,95–18,51-kertaiset nopeutukset taas ovat mikrobenchmarkista eivätkä kerro tokeninopeudesta.

Qualcommin Adreno-näytönohjaimille suunnattu OpenCL-tausta sai viikossa iq2-kvanttien polut ja q2_0:n alkuversion. Mukaan tulivat myös nopea Walsh-Hadamard-muunnos ja sunnuntaina iq3-tuki ja Qualcommin kehittäjät ovat muutoksissa kanssatekijöinä. Saman yhtiön Hexagon-NPU:n taustassa b12030 korjasi vian, joka koski v75-sukupolvea ja uudempia. Niissä 2D-DMA-kuvaaja tallentaa rivikoon ja askelpituudet 24 bittiin, minkä vuoksi vähintään 16 megatavun askelpituudet katkesivat ja siirrot osuivat vääriin osoitteisiin. Apple Siliconilla b11927 porttasi CUDA-taustan fuusioidun top-k-maskioperaation Metaliin. Harvaa huomiota käyttävät mallit saavat näin saman operaation Mac-koneilla. Metal-polulla nähtiin myös regressio. Unaarikernelit kaatuivat vanhemmilla työkaluketjuilla, mihin Ollaman julkaisuehdokas 0.40.13-rc1 toi väliaikaisen kiertotien ennen llama.cpp:n korjausta b11983:ssa. Vulkan-taustan b11933 korjasi tarkistuksen, joka hylkäsi matriisikertolaskut, joiden syötteet erosivat ne3-ulottuvuudessa. Ne ajettiin aiemmin CPU:lla painojen kopiolla.

Muistin käsittelyssä viikko jätti kesken asioita, jotka kannattaa mitata. Perjantain b12007 pitää KV-välimuistin rotaatiovakiot GPU:lla ja tukee kerroskohtaisia rotaatiopuskureita. Usean laitteen rotaatiotyö dokumentoitiin keskeneräiseksi ja rotaatiopuskurien testit poistettiin. Samana päivänä b12006 tiukensi palautuksia. Rekurrentti muisti ja DSV4-välimuisti hyväksyivät aiemmin palautuksen `n_rs_seq`-rajaan asti, vaikka viimeinen ubatch ei ollut kirjoittanut kyseistä tilannevedosta. Nyt `seq_rm` torjuu sitä syvemmät pyynnöt. Lauantaina b12025 korjasi huomiotensorien rotaation kerroksissa, joissa KV-välimuisti on jaettu. Useampi build muuttaa myös tulosteita tavalla, joka näkyy versioiden välisissä vertailuissa. Build b12021 skaalaa Gemma 4:n kuvat ilman letterbox-täytettä ja b12019 poistaa Qwen3-Coderin think-lohkon tulosteesta asetuksella `reasoning_format none`. Responses-rajapinnalle kirjoitetut asiakkaat voivat b12022:sta alkaen välittää lisäparametreja kentässä `extra_params`.

Inferenssipalvelin vLLM:n 0.33.0 kulki viikon julkaisuehdokkaina rc0:sta rc3:een ja valmistui sunnuntaina. Ehdokkaiden näkyvät muutokset olivat kapeita. Rc2 ohittaa ROCm-ympäristössä AITER:n viritetyn GEMM:n gfx950-kohteessa, kun M*N on vähintään 2^31. Muut olivat CI-korjauksia. Lopullisen julkaisun kuvaus koostuu pelkästä

otsikosta, minkä vuoksi muutosloki on luettava ennen tuotantopäivitystä.

Automaatioalusta n8n julkaisi 27.10. sekä version 2.46.0 että pääversion 3.0.0. Molempien julkaisutiedot alkavat samalla korjauslistalla. Pääversion rikkovat muutokset eivät näy listan alussa, minkä vuoksi siirtymäohje on luettava ennen päivitystä. Viikon lopussa korjauksia sai neljä rinnakkaista linjaa: 3.0, 2.46, 2.45 ja 1.123. Kakkoslinjan julkaisut parantavat v3-siirtymäraporttia merkinnöillä poistetuista solmuista ja ratkaisuohjeilla. Agenttien kannalta kiinnostavimmat uudistukset koskevat valtuutusta. OAuth2-tunnuksille voi määrittää omat scopet ja julkinen API ilmoittaa OAuth-resurssihaasteen jokaisessa 401-vastauksessa. Kuvauksen perusteella kyse on samasta resurssihaastemallista, jota MCP:n valtuutusmäärittely käyttää.

Pitkien striimien kestävyys oli viikon hiljainen yhteinen teema. LangGraph 1.2.18 ottaa DeltaChannelissa täyden snapshotin 100 päivityksen välein 1 000:n sijaan. Pitkien threadien luku nopeutuu tallennustilan kustannuksella ja vanhan käytöksen saa takaisin parametrilla `snapshot_frequency=1000`. Perjantain SDK 0.4.11 yrittää epäonnistuneita uudelleenyhdistyksiä kasvavalla viiveellä myös kutsuissa `runs.wait`, `join` ja `cancel(wait=True)`. Se säilyttää SSE-datan ja kursorin tilan pätkien yli ja varoittaa, kun striimi yhdistyy uudelleen ilman `Last-Event-ID`:tä. Sama vikaluokka korjattiin viikon lopulla Gemini CLI:n yöversiossa, jossa CodeAssistServer tyhjentää nyt odottavan SSE-puskurin striimin päättyessä. Aiemmin viimeinen tapahtuma saattoi jäädä käsittelemättä. Clinen CLI puolestaan yrittää uudelleen vuoroa, jonka yhteys katkeaa mallin ehdittyä striimata vain päättelyosuuden.

Konesaliluokassa NVIDIA kuvasi CUDA 13.4:n muistirajoitetun GPU-jaon Kubernetesissa. Lähtökohta on, että moni tekoälykuorma tarvitsee vain osan näytönohjaimesta ja kokonainen laite jokaiselle jättää kapasiteettia käyttämättä. Mission Control 2.5 tuo Kubernetes-ympäristöihin automaattisen palautumisen tuhansien GPU:iden koulutusajoissa. Dynamo-kirjoitus rakentaa validoidun Kubernetes-ajoympäristön hajautetulle inferenssille. Kaikki ovat valmistajan omia tuotekirjoituksia. Mittakaavan kapasiteetin kasvulle antaa Epoch AI:n datajulkaisu (2026). Sen mukaan kvartaaleittain myydyn tekoälysirukapasiteetin 2,5-kertainen vuosikasvu vastaa noin yhdeksän kuukauden tuplaantumisaikaa.

08 Benchmarkit ja vertailut

Artificial Analysis julkisti Intelligence Indexin viidennen version keskiviikkona. Saman viikon aikana sen sivuille kertyi lukuja viidelle uudelle mallille. GPT-6.1 Terra sai Max-tasolla 48 ja Kimi K3.1 Pro 55. Gemini 4 Carbonin korkea päättelytaso sai 59, Mistral Large 4:n Medium-asetus 23 ja DeepSeek V4.1 Flash 1017:n päättelytön variantti 19. Tulossivujen otsikoista ei selviä, millä indeksiversiolla kukin luku on laskettu. Rinnakkain luettuna Mistralin 23 näyttää heikolta Terran päättelyttömän tason 33 vieressä. Vertailu kantaa kuitenkin vain, mikäli molemmat on laskettu samalla versiolla. Indeksi on lisäksi yleisen kyvykkyyden koostemittari. Mistralin väitettä parhaasta avoimesta koodausmallista se ei suoraan testaa.

Terran kuusi päättelytasoa piirtävät käyrän: Non-reasoning 33, Low 36, Medium 41, High 44, Xhigh 45 ja Max 48. Ääripäiden ero on 15 pistettä. High-tasolta ylöspäin lisää tulee enää neljä pistettä, joista kolme vasta Max-tasolla. Indeksi ei kerro, paljonko lisäpisteet maksavat viiveenä ja tokeneina. Kehittäjäbloggaajan nostamassa lainauksessa mallitalon työntekijä muistuttaa, että suorituskyky on monidimensionaalista eikä matematiikassa paras malli ole välttämättä paras vitseissä. Lainauksen mukaan suuri osa vaihtelusta selittyy silti mallin koolla. Isompi malli on luotettavampi, kalliimpi tokenia kohden ja muistaa paremmin harvinaisia faktoja.

Interconnects tarttui indeksin uudistukseen kirjoituksella Artificial Analysis v5 and the benchmarker's dilemma. Sitä kuvataan katsaukseksi vertailualaan ja siihen, mitä siltä puuttuu. MLCommons lähestyi tulosten pelattavuutta palvelun rakenteella. AILuminate Safety 1.2.1 ja Jailbreak 1.1 tulivat saataville uudessa Candela-palvelussa, jossa harjoitusajoja saa tehdä rajattomasti, kun taas virallinen versioitu tulos arvioidaan erillisellä kehotejoukolla. Järjestön mukaan oma arvioija ja persentiilipisteytys on suunniteltu kestämään benchmaxxingia ja tulosten poimintaa. Rinnakkainen kirjoitus perustelee, miksi turvallisuuspisteet eivät ole kompositionaalisia. Hienosäätö siirtää turvallisuutta ennalta arvaamattomasti ja moni vertailu on liian kohinainen havaitsemaan muutosta.

Viikon muissa arviointiasetelmissa korostui kysymys siitä, kuka mittaa. Elorianin houkutuslaskentatesti paljastaa yhtiön mukaan frontier-malleista kuilun näkemisen ja päättelyn

välillä. Samaan aukkoon yhtiö kehittää omaa Glimpse-tuotettaan, joka ajattelee kuvilla. Ai2:n tutkimuksessa Old Ideas, Novel Problems (2026) pienet muutokset arviointiasetelmaan tuottivat kielimallituomareilta hyvin erilaisia tuomioita tutkimusideoiden uutuudesta. Havainto koskee jokaista putkea, jossa malli suodattaa ideoita ennen ihmisen arviota. Microsoft Research (2026) kuvaa Copilotin arviointia synteettisellä yritysdatalla, koska oikeaa dataa ei yhtiön mukaan saa käyttää arviointiin missään olosuhteissa. Menetelmän arvo riippuu siitä, kuinka hyvin generaattori toistaa oikean organisaation sotkun.

Hacker Newsissa nousi kahdesti saman viikon aikana väite, jonka mukaan viisi avointa Qwen3.8-27B-mallia yltää yhdessä Fable 5:n tasolle. Keskiviikon GVS5H-repo (185 pistettä) kertoi tason syntyvän orkestrointikehyksellä. Lauantain Memdoor (385 pistettä, 188 kommenttia) nimeää vertailun: viiden mallin tiimi päihittää sen mukaan Fable 5:n LiveCodeBench Hardissa. Kumpikin tulos on tekijän oma, eikä kumpaakaan ole toistettu riippumattomasti. Lähteistä ei selviä Fable 5:n julkaisija. Samana lauantaina 176 pistettä keräsi kirjoitus, jonka mukaan edellisen sukupolven GPT-5.6 Sol pelaa Factorion läpi rauhanomaisessa tilassa.

09 Työskentelytavat ja agentit

Agentti-istunto alkoi viikon julkaisuissa muistuttaa palvelinprosessia enemmän kuin keskustelua. Kehittäjä sulkee työpöytäsovelluksen kesken agentin ajon ja avaa saman tehtävän komentoriviltä. Clinen aiemmissa versioissa tästä seurasi toinen ajo, jota ei voinut pysäyttää, kun taas alkuperäinen jatkoi tavoittamattomissa ja kirjoitti transkriptiotaan vanhentuneeseen tilannevedokseen. Viikon alussa julkaistuissa SDK 0.2.0:ssa ja CLI 3.0.74:ssä avaaminen liittyy elävään istuntoon eikä käynnistä uutta. Istunnon takana oleva hub kirjaa kaikki asiakkaat, jotka ovat rekisteröineet istunnolle kyvykkyyden, myös kesken ajon liittyneet. Pyynnöt ohjataan ensimmäiselle yhä yhdistetylle asiakkaalle. Pyyntö kaatuu nopeasti toimintaohjeen sisältävään virheeseen, jos asiakkaita ei ole jäljellä.

Sama ajatus näkyi muissa työkaluissa eri muodoissa. Claude Coden 2.1.311:ssä self-hosted-runner sai valitsimet `--hold-idle-file` ja `--max-hold-min`. Niillä tyhjäkäynnillä oleva istunto pidetään auki niin kauan kuin isäntä raportoi istunnon jättämiä prosesseja käynnissä oleviksi, esimerkiksi agentin käynnistämän kehityspalvelimen tai testiajon. Anthropicin alustatiimi kuvasi Latent Spacen haastattelussa hallinnoitujen agenttien periaatteen, jossa päättelevä malli eli aivot erotetaan kädistä eli ympäristöstä, jossa työkalut ajetaan. Perjantain 2.1.314 tiivistää VS Codessa ja Claude Desktopin Code-välilehdellä yli 200 000 tokenin keskustelun, kun se jää joutokäynnille. MCP-määrittelyyn tuli tukimatriisi Tasks-ominaisuudelle eli pitkäkestoisille asynkronisille tehtäville ja Claude Code ja Claude.ai merkittiin tukijoiksi. Arcan tekoälyjohtaja kuvasi Latent Spacen haastattelussa rahoitustyötä, jossa työnkulut kestävät päiviä. Agentti tarvitsee siinä pysyvän muistin ja oman ajokehyksen. Haastattelun otsikko Agents Prompt People kääntää tavallisen suunnan: agentti kehottaa ihmistä.

Ajojen pidetessä ihmisen kontrollipiste siirtyy luvan antamisesta lopputuloksen katselmointiin. Clinen laajennuksen 4.1.26 pilvitehtävissä Plan/Act-vaihto ja auto-approve-palkki piilotetaan kokonaan. Työpöytäsovellus 0.0.49 rakentaa tarkistuksen valmiiden vuorojen varaan. Istunnon otsakkeen Changes-painike avaa paneelin, jossa ovat viimeisimmän vuoron tai koko istunnon muuttamat tiedostot diffeineen. Virallisen code-review-liitännäisen versio 1.0.1 tulostaa katselmoinnin oletuksena terminaaliin ja kirjoittaa rivikohtaiset kommentit vain `--comment`-lipulla. Gemini CLI:n yöversio liittää ACP-lupapyyntöihin MCP-palvelimen ja työkalun metatiedot. Editori voi näin näyttää, mikä palvelin ja mikä työkalu lupaa pyytää. OpenAI:n asiakastarinoissa CodeRabbitin tietoturva-agentti jättää löydökset kehittäjän arvioitaviksi ja Greptile laajentaa katselmoinnin diffistä käyttäjäpolkujen testaamiseen. Molemmat ovat toimittajan omia tapauskuvauksia ilman riippumatonta mittausta.

Tarkimman kuvan katselmoinnista antoi MCP-palvelinrepon PR #5322, joka korjasi tiedoston siirron eri tiedostojärjestelmien välillä, joissa `rename()` palauttaa EXDEV-virheen. Commitit on merkitty `Co-Authored-By: Claude Opus 5.5` -rivillä ja ylläpitäjän allekirjoituksella. Jokainen katselmointikierros löysi hienovaraisemman reunatapauksen. Ensin kopio muutettiin luotavaksi tilassa 0600, jotta yksityinen tiedosto ei ole hetkeäkään muiden luettavissa. Myöhemmin todettiin, että paikallaan tehty muokkaus säilyttää laitteen ja inoden, minkä vuoksi koon ja aikaleimojen on täsmättävä nanosekunnin tarkkuudella. Viimeisellä kierroksella huomattiin, että `fs.futimes` tulkitsee negatiivisen luvun hetkeksi ”nyt”, jolloin ennen vuotta 1970 olevat aikaleimat katosivat. Projektissa llama.cpp avustus merkitään Assisted-by-trailerilla. Viikon commiteissa avustajina mainittiin Claude Opus 5.5, Codex

ja paikallisesti ajettuja malleja laitteistoineen, esimerkiksi MacBookilla (M5 Max, 128 GB) ajettu Qwen3.8-Flash-Next-GGUF. MCP-määrittelyn commiteissa Claude on kanssatekijä ja osassa on linkki istuntoon, jossa muutos tehtiin. Kyle Kingsbury kirjoittaa Jepsenin ParticleDB 1.1.4 -tuloksissa (2026), että kielimallien käyttö rapauttaa luottamusta dokumentaatioon, ohjelmistoihin ja niitä tuottaviin organisaatioihin. Hänen mukaansa kypsän järjestelmän merkit ovat nyt helpompia väärentää. Hän kysyy, miten ylläpitäjät voivat varmistaa oikeellisuuden ja osoittaa laadun käyttäjille.

Kustannusten ja päättelyn hallinta siirtyi samaan aikaan ylläpidon käsiin. Claude Coden 2.1.316 tuo ympäristömuuttujan `CLAUDE_CODE_SUBAGENT_EFFORT`, joka asettaa aliagenttien effort-tason silloin, kun mikään muu ei sitä nimeä. Aiemmin aliagentit perivät istunnon tason. Nyt oletukseksi voi asettaa työnjaon, jossa raskas pääistunto suunnittelee ja kevyemmät aliagentit tekevät haut ja rutiinimuokkaukset. `CLAUDE_CODE_DISABLE_ULTRATHINK` estää nostamasta päättelyn tasoa yhdellä avainsanalla. Claude-sovellusten gatewayhin tuli ryhmän yhteinen kuukausibudjetti `pooled_budget`-kentässä. Gemini CLI luokittelee projektin kulukaton ylityksen nyt lopulliseksi kiintiövirheeksi. Hacker Newsissa 543 pistettä kerännyt gist väittää Claude Coden lähettävän jokaisen taustalla ajettavan agentin keskusteluhistorian säännöllisesti mallille. Mekanismille ei julkisista lähteistä löydy Anthropicin selitystä.

Eristetyn ympäristön ja editorin yhteys parani toisaalla. Gemini CLI:n v0.67.0-preview.0 siltaa IDE-companionin konttihiekkalaatikoihin exec-stdion kautta eikä verkkoyhteydellä. Hiekkalaatikossa ajettu agentti säilyttää näin yhteyden editorin kontekstiin. JetBrainsin thinkrail on kevyt agenttinen IDE pi-koodausagentin ympärille. Työ eristetään siinä erillisiin työtiloihin ja agentti saa pysyvää projektitietoa elävistä spekseistä ja uudelleenkäytettävistä taidoista. Spec-vetoisesta työtavasta kiinnostuneelle se on kokeilun arvoinen. OpenAI:n Codex CLI:stä ilmestyi viikon aikana vakaat 0.165.0 ja 0.166.0 sekä useita alfoja aina 0.167.0-alpha.2:een asti. Julkaisujen kuvauksena oli koko viikon pelkkä versionumero.

Ihmispuolen keskustelu kulki kahta raidetta. GitHubin Octoverse 2026 -lukujen mukaan yhdistettyjen pull requestien määrä kolminkertaistui alle vuodessa 220 miljoonaan kuukaudessa. Kehittäjäbloggaaja tulkitsee kasvun agenttipohjaisen kehityksen vaikutukseksi. Samalla viikolla hän julisti lokakuun 2026 vibe-pelikehityksen käännekohdaksi, jonka ajurina hän mainitsee Claude Opus 5.5:n. Fortunen jutussa Cisco antaa 27-vuotiaalle insinöörille rajattomasti tokeneita vanhan koodin uudelleenkirjoittamiseen ja ketju keräsi 351 kommenttia. Toisella raiteella kirjoitus ”AI promised less work. I've never worked more” keräsi 675 pistettä ja 659 kommenttia. SigNozin uutiskirje väittää tekoälyn rapauttavan SRE-taitoja ja martinfowler.com kysyy, miksi hyvät kehittäjät suhtautuvat agenttiseen kehitykseen varauksella. Y Combinatorin kokeilu agenttiparvista Minecraftissa ja Mindustryssä nousi perjantaina 382 pisteeseen. Sunnuntaina Tom's Hardware raportoi Minecraftiin päästetyistä parvista, joiden 50 agentin tiimit eivät rakentaneet sivilisaatiota vaan ajautuivat lamauttavaan byrokratiaan. Lähteistä ei selviä, onko kyse samasta kokeilusta. Torstaina 119 pistettä keräsi video, jossa Linuxin tiedostotyökalut kirjoitetaan C:llä käsin ilman tekoälyä 52 minuutissa.

10 Liitännäiset, MCP ja repot

MCP:n viikko kului protokollan hallintoon enemmän kuin itse protokollaan. Release Cycle Updates -ehdotus SEP-3398 kirjattiin hyväksytyksi ja yhdistettiin 27.10. Se määrittelee julkaisusyklin prosessin, julkaisupäivät, jäädytyksen äänestysmääräajat ja julkaisupäällikön käyttöoikeudet. Kaksi päivää myöhemmin SEP-3392 (Extension Lifecycle) merkittiin Finaliksi. Beta-vaiheen laajennusten on jatkossa ilmoitettava aiottu määränpäänsä, joko Stable tai Core ja määränpää näkyy laajennussivulla. PR #3364 selventää, että laajennuksista neuvotellaan laajasti myös 2026-07-28-revisiota vanhemmissa versioissa initialize-vaiheessa. SDK:ita kannustetaan kuljettamaan `capabilities.extensions`-kenttää ilman velvoitetta. Yritysten valtuutukselle tuli tukitaulukko (#3306), joka yhdistää tunnistuspalvelun ja valtuutuspalvelimen roolit. Okta, PingFederate ja Descope dokumentoivat molemmat roolit. Oktan XAA on yleisesti saatavilla Identity Engine 2026.08.1:ssä, kun taas WorkOS ja Scalekit ovat rajatussa varhaisessa käytössä. Tilat perustuvat toimittajien julkaisumuistioihin 30.8.2026 mennessä. Yhteisön toimintasäännöt kieltävät nyt kolmannen osapuolen tekoälymuistiinpanijat ja kokousbotit yhteisön kokouksissa. Tietoturvan kiinnostusryhmä organisoitiin uudelleen reitittäväksi ryhmäksi.

Toteuttajalle kiireellisimmät muutokset koskivat dokumentaation esimerkkikoodia, jota on voitu kopioida tuotantoon. Embedded resource -esimerkissä `annotations` oli resource-kentän sisällä, vaikka EmbeddedResource-määrittelyssä se on kentän sisarus. Virhe oli draft-, 2025-11-25- ja 2025-06-18-sivuilla, eikä esimerkistä kopioitu rakenne validoitunut skeemaa vasten. Pythonin client credentials -esimerkit välittivät parametrin `scopes=`, joka aiheuttaa paketissa mcp 2.3.0 TypeErrorin, koska oikea nimi on `scope=`. Puuttuva `issuer=` antaa MCPDeprecationWarningin ja muuttuu pakolliseksi versiossa 3.0. Asiakastutoriaalin TypeScript-versio luki vain ensimmäisen sisältölohkon. Kotlin-versio kutsui `first().text()`-metodia lohkolle, joka ei välttämättä ole tekstiä. Ruby-versio lähetti yhden tool_resultin jatkokutsua kohden, vaikka kaikki tulokset kuuluvat samaan viestiin. Lauantaina korjattiin vielä Java- ja Kotlin-tutoriaalit, joiden esimerkit eivät toimineet sellaisinaan. Rustin tutoriaali rakennettiin samana päivänä uudelleen rmcp 3:lle.

Taitojen ja liitännäisten jakelu sai viikossa samoja mekanismeja, joita ohjelmistopaketeilta on totuttu odottamaan. Anthropic siirsi Claude Coden liitännäisten jakelun claude-plugins-official-repoon niin, että olemassa olevat tunnisteet toimivat edelleen. Kymmenelle liitännäiselle annettiin eksplisiittiset versiot, koska ilman versiota jokainen repon commit näkyi uutena versiona ja tuotti käyttäjille tyhjiä päivityksiä. Runwayn runway-api-liitännäinen kiinnitettiin git-subdir-lähteenä ensin commitiin fe79eab0 ja samana iltana commitiin 6e98d3f3. Kolmannen osapuolen päivitys kulkee siis kokoelman oman muutoksen kautta eikä suoraan julkaisijan päähaarasta. MCPJam Inspector 3.17.0 tarkistaa MCP:n yli jaettujen taitojen koon ja tiivisteen, verifioi frontmatterin ja nimeää taidot palvelinkohtaiseen nimiavaruuteen.

Expo julkaisi viralliset agenttitaidot Expo- ja EAS-projektien rakentamiseen, julkaisuun, päivittämiseen ja vianetsintään. Claude Codeen ja Codexiin ne asennetaan liitännäisenä, jolloin päivitykset tulevat automaattisesti ja Expon dokumentaatio sekä CLI pysyvät totuuden lähteenä. Expolla agentin avulla rakentavalle ne ovat kokeilun arvoiset. NVIDIA/SkillSpector skannaa Claude Coden, Codex CLI:n, Gemini CLI:n ja MCP:n taitoja ennen asennusta. Se etsii haavoittuvuuksia, prompt injection -yrityksiä, tietojen ulosvientiä ja toimitusketjuriskejä. Taustalla on tutkimusaineisto, josta analysoitiin 31 132 taidon osajoukko. Staattinen skannaus ei kata ajonaikaista käytöstä. Kolmannen osapuolen taitojen asennusputkessa se on silti halpa ensimmäinen suodatin.

Työpöydän agenttitiimeistä tuli trendilistan oma luokka. Nightly-labsin openbot ajaa pysyviä tekoälytyökavereita paikallisesti ja liittää ACP:n kautta muun muassa Grok CLI:n, OpenCoden ja Cline CLI:n. CopilotKitin samanniminen OpenBot antaa AG-UI-agenteille oman selaimen, tiedostot ja rajatut työkalut. Se tekee päätöksen ennen jokaista toimea ja kirjauksen sen jälkeen. Sunnuntaina listalle nousi Orkas, MIT-lisensoitu paikallinen sovellus macOS:lle ja Windowsille, jossa Commander-agentti koordinoi 16 sisäänrakennettua erikoisagenttia sekä Claude Codea, Codexia ja OpenCodea. Eristystä ja auditointia miettivälle tiimille CopilotKitin OpenBot on luokasta kokeilun arvoisin. Claude-Code-Game-Studios muuttaa Claude Code -istunnon pelistudion hierarkiaa jäljitteleväksi joukoksi, jossa on 49 agenttia ja saman kuvauksen mukaan joko 72 tai 74 taitoa. Agenttien määrä ei todista laatua.

Viikon uteliaimmat repot käänsivät kysymyksen mallista ja koneesta toisin päin. Nesformer upottaa Qwen3.5-4B:n painoihin toimivan emulaattorin, joka jäljittelee Nintendo Entertainment Systemin 6502-suoritinta. Mallin sisällä voi ajaa NES-ROMeja, kuten Super Mario Brosia, joskin hyvin hitaasti. Tekijä kertoi Hacker Newsissa käyttämänsä Opus 5.5 -kehotteet. Ensimmäinen antaa mallille luettavaksi qwenputer-kokeilun artikkelin ja lähdekoodin. Sitten se pyytää selvittämään, voisiko samalla tekniikalla emuloida NES:n suorittimen ja muistin. Repon lukemisen arvo on kehotteissa, joissa tutkimuskysymys annetaan agentille lähteineen avoimena selvitystehtävänä. Bootgpt kulkee toiseen suuntaan: koko inferenssimoottori mahtuu pääkäynnistystietueen 510 tavuun. Painot ja sanasto ovat sitä seuraavissa sektoreissa. Kone tuottaa loputtomasti Linuxin käynnistysviestejä laitteistolle, jota ei ole olemassa. Tinyworldformer ajaa pelejä pienellä maailmamallilla Raspberry Pi Pico 2 W:llä. Kaikki kolme ovat rajakokeiluja eivätkä työkaluja.

11 Tietoturva ja haavoittuvuudet

Tekoälytyökalujen julkaisuissa ei ollut CVE-tunnisteella merkittyjä korjauksia. Yksi GitHubin tietoturvaneuvonta ja joukko valtuutus- ja rajakorjauksia vaativat silti toimia. Viikon havainnot osuivat rajoihin, joiden varaan agenttien eristys rakennetaan.

Välimuisti peitekanavana. The Decoderin raportin mukaan OpenAI:n eristetyt agentit löysivät viestintäkanavan suorittimen välimuistin kautta. Hacker Newsissa ketju keräsi 347 pistettä ja 245 kommenttia. Kontit ja virtuaalikoneet eristävät tiedostojärjestelmän ja verkon. Jaettu viimeisen tason välimuisti jää kuitenkin yhteiseksi resurssiksi, jonka tilaa voi käyttää peitekanavana. Vaatimus, ettei samalla isännällä ajettavien agenttien pidä voida viestiä keskenään, edellyttää erillisiä isäntiä tai välimuistin osiointia.

Kirjoitusrajat ja vuotoreitit. Claude Code 2.1.314 lisäsi asetuksen `permissions.blockWritesOutsideWorkingDirectories`. Sen avulla Write, Edit ja NotebookEdit kieltäytyvät työhakemistojen ulkopuolisista poluista kaikissa käyttöoikeustiloissa. Bash ei kuulu lueteltuihin työkaluihin, minkä vuoksi asetus ei muutoslokin perusteella ulotu kuoren kautta tehtyyn kirjoittamiseen. Gemini CLI:n yöversio teki samaan aikaan `isWithinRoot`-tarkistuksesta Windowsissa kirjainkoosta riippumattoman. Claude Coden 2.1.312:ta edeltävissä versioissa VS Code -laajennuksesta kirjautuneen käyttäjän istunto jätti hallinnoidut asetukset ja organisaation käytännöt hakematta loppuistunnon ajaksi. Keskitetysti asetetut rajoitukset eivät siis olleet näissä istunnoissa voimassa. Tunnistukseen riittää versionumero ja VS Code -kirjautumisella alkaneiden istuntojen läpikäynti. Sama versio estää ylikirjoittamasta suurta CLAUDE.md:tä tai sääntötiedostoa, josta malli on nähnyt vain alun. Alihakemistojen ohjetiedostojen aiemmat muutokset kannattaa käydä läpi versionhallinnasta. Gemini CLI:n v0.67.0-yöversio noudattaa 31.10. alkaen varjorepossa tiedostoa `.git/info/exclude` ja asetusta `core.excludesFile`. Varjorepo on erillinen git-repo, johon työkalu tallentaa tilannevedoksia palautusta varten. Aiemmissa versioissa vedoksiin saattoivat päätyä tiedostot, jotka oli poissuljettu vain näillä mekanismeilla, esimerkiksi paikalliset salaisuudet. Korjaus on toistaiseksi vain yöversiossa. GitHub-issue #81138 raportoi Claude Coden lisäävän käyttäjän sähköpostiosoitteen kehotteeseen. Redditissä käyttäjä kertoo mallin lähettäneen osoitteen verkkosivustoille. Kaikki kontekstissa oleva voi päätyä työkalukutsun parametriksi, kuten hakukyselyyn tai URL:ään. Verkkotyökalujen sallitut kohteet voi rajata käyttöoikeussäännöillä.

MCP-palvelimet. Filesystem-palvelimessa asiakkaan ilmoittama Root voi olla symlink, jonka kohde vaihdetaan Rootien latauksen jälkeen. Funktio `resolveUnicodeEquivalentPath` aloitti puuttuvan polun läpikäynnin sallitun hakemiston realpathista tarkistamatta sitä. Tällöin `write_file` tai `create_directory` uudelleenohjatun aliaksen alle päätyi kaikkien sallittujen hakemistojen ulkopuolelle. Korjaus tarkistaa aloituskohdan realpathin samoin kuin jokaisen sen alla olevan osan. Se yhdistettiin v2-haaraan 27.10. ja julkaistiin 28.10. Altistuvia ovat kokoonpanot, joissa sallittu hakemisto annetaan symlinkkinä ja symlinkin kohdetta voi muuttaa palvelimen ollessa käynnissä. Korjatun npm-paketin versionumeroa commit-otsikot eivät kerro. Git-palvelimen neuvonta GHSA-w766-5x43-whvw korjattiin samassa julkaisussa. Korjaus sisältyy Python-palvelimien CalVer-versioon 2026.10.28. Neuvonnan sisältö ja vakavuus on luettava GitHubin tietoturvasivulta.

n8n:n ajotietojen oikeudet. Versioissa 2.46.0 ja 3.0.0 AI Builderin Assistant tarkistaa `execution:read`-oikeuden ennen ajotietojen lukemista. Samoissa versioissa Assistantin mallikutsut kulkevat konfiguroidun AI-kuljetuksen kautta, mikä on olennaista ympäristöissä, joissa mallikutsujen on kuljettava määritellyn yhdyskäytävän läpi. Työnkulkukohtaisten reittien vastaava tarkistus tuli 2.46-linjaan versiossa 2.46.2 (#42445) ja 2.45-linjaan perjantain versiossa 2.45.7 (#42446). Perjantain 2.46.3 ei mainitse korjausta, joka oli jo 2.46.2:ssa. 3.0-linjan julkaisutiedoissa vastaavaa merkintää ei näy. Ykköslinjan 1.123.90 vaatii käyttäjäkutsun hyväksymiseen allekirjoitetun tunnisteen (#42292). Versio 1.123.89 korjasi MCP Server Trigger -solmun muistivuodon, jossa käyttämättömät istunnot jäivät vapauttamatta. Julkisesti tavoitettavassa MCP-päätepisteessä vuoto on resurssien uuvuttamisen riski. Ajotietoihin tallentuvat solmujen syötteet ja tulosteet, joissa voi olla asiakasdataa. Altistuvia ovat monen käyttäjän instanssit, joiden versio on vanhempi kuin oman linjan korjausversio.

ChatGPT-tilien kaappaus. Kehittäjäbloggaaja nosti sunnuntaina esiin Parameter Labsin kuvauksen hyökkäyksestä, jolla ChatGPT-tilejä kaapattiin. Kohteena oli artefaktimallin toteutus, jota OpenAI on kutsunut Canvasiksi ja jonka uusin versio näyttää kulkevan nimellä HTML previews. Niissä interaktiiviset HTML-sovellukset näytetään hiekkalaatikoidussa iframessa. Iframen on pysyttävä tiukasti eristettynä, jotta epäluotettava koodi ei pääse OpenAI:n rajapintoihin tai yksityiseen dataan. Kuvauksen mukaan monimutkaisen hyökkäysketjun osana on MathLive-kirjasto, jota OpenAI käyttää kuvaajatuessa. Julkaistusta tiivistelmästä ei selviä ketjun loppuosa eikä korjauksen ajankohta.

Avoimet painot ja korjausikkuna. Epoch AI:n julkaisussa (2026) avoimesti jaettu GLM 5.3 muutti julkaistut mutta jakeluun ehtimättömät Firefox-korjaukset toimiviksi yhden klikkauksen exploiteiksi kuuden julkaisun kohdalla.

Kappalehinta oli noin 1 000 dollaria. Hyökkäysikkuna on korjauksen julkisuuden ja jakelun välinen aika. Tiivistelmä ei kerro haavoittuvuuksia eikä ihmisen osuutta. Siitä ei myöskään selviä, ylittivätkö exploitit selaimen hiekkalaatikon. Ylläpitäjän mittari on oma viive ylävirran korjauksesta tuotantoon, jota omat paketit ja pitkään tuetut haarat voivat pidentää. Interconnects kysyi samalla viikolla, oliko se väärässä avoimista malleista ja kyberriskeistä. Kirjoitus kytkeytyy Black October -termiin, joka syntyi PHP2Shell- ja GitLab-paljastusten ryöpyssä. UK AI Security Institute arvioi, pystyvätkö koodausagentit rakentamaan biologisia tekoälytyökaluja uudelleen pelkän tutkimusartikkelin perusteella. Toisessa kirjoituksessaan se käsittelee esikoulutusdatan suodattamista avoimen painon biologisissa malleissa. Suodatus pysyy voimassa myös silloin, kun painot ovat vapaasti jaossa ja jälkikäteen lisätyt suojaukset voi hienosäätämällä poistaa. Kyberturvallisuuskeskuksen viikkokatsaus 44/2026 nostaa Brotli- ja Node.js-haavoittuvuudet organisaatioiden riskeiksi ja korjaavat versiot on haettava katsauksesta.

Epäluotettavat artefaktit ja pienemmät korjaukset. Trendilistalle nousi jailbreaks-kokoelma, jossa on 31 kehotetta yli 25 mallin turvarajojen kiertämiseen. Mukana on järjestelmäkehotteita, persoonalukituksia, AGENTS.md-tiedostoja ja Ollaman Modelfileja. AGENTS.md ja Modelfilen SYSTEM-lohko ovat ohjeita kantavia tiedostoja. Kolmannen osapuolen tiedoston tuominen repoon on vieraan järjestelmäkehotteen hyväksymistä ja tiedostot kannattaa katselmoida pull requesteissa kuten koodi. Sunnuntaina Hacker Newsiin nousi väite Anthropicin Mythos-luokan mallin vuotamisesta Hugging Faceen. Ketjun linkki poistettiin haittaohjelmaepäilyjen vuoksi, eikä väitettä ole varmennettu. Ajoalustan llama.cpp build b11935 sieppaa RPC-palvelimen `recv_msg`-funktiossa `std::length_error`-poikkeuksen. Poikkeus syntyy, kun vastaanotettu kokokenttä ylittää varattavan puskurin rajan, jolloin virheellinen viesti on voinut kaataa prosessin. Projektin ohjeistus varoittaa edelleen ajamasta RPC-palvelinta luottamattomassa verkossa. LangGraphin checkpoint 4.3.1 hylkää NUL-tavut namespace-nimissä ennen tallennuskerrosta. Se tuli perjantaina myös checkpoint-sqlite 3.1.3:n mukana. Päivitys on olennainen sovelluksissa, jotka muodostavat nimiavaruuksia käyttäjäsyötteestä.

Lähteet

Introducing GPT-6.1 Terra · OpenAI

GPT-6.1 Sol helps CodeRabbit strengthen security across complex codebases · OpenAI

Greptile takes bug detection beyond code review with GPT-6 Astra · OpenAI

Gemini 4 Carbon: Our next frontier model · Google DeepMind

[AINews] Gemini 4 Carbon tops leaderboards, launches for cyber defenders · Latent Space

Le Chonk: Built for Coders, Trusted by Defenders. · Mistral

Welcome Le Chonk: Mistral Large 4 on Hugging Face · Hugging Face -blogi

mistralai/Mistral-Large-4-1T-A52B · HF trending -mallit

An updated Model Spec for more capable and beneficial AI · OpenAI

How we detect frontier model misalignment · OpenAI

How we detect frontier model misalignment · OpenAI Alignment Research Blog

Making it easier for websites to express AI training preferences · OpenAI

Working with academic communities to build scientific understanding · OpenAI

API credits benefit for eligible ChatGPT subscriptions · Simon Willison

OpenAI expands partnership with Oracle to bring its apps to ChatGPT · OpenAI

OpenAI and ServiceNow expand partnership across enterprise workflows · OpenAI

Claude-tekoälyn kehittäjä on menossa pörssiin: Anthropic paljasti samalla 81 miljardin dollarin jättitappiot · Tivi

[AINews] Anthropic files for $2T IPO in 2 weeks: 1400% growth rate, 48% of revenue through cloud partners, 85% margin, $1B op profit in Q2, $134B cash in bank and $10B lifetime burn · Latent Space

Claude Platform release notes: October 28, 2026 · Anthropic API release notes

Claude Platform release notes: October 29, 2026 · Anthropic API release notes

Claude Platform release notes, October 30, 2026 · Anthropic API release notes

Early progress on accelerating drug development with Claude · Anthropic news

The sleep framework: A new way for LLMs to retain knowledge in continual learning · Google Research

[AINews] Google admits AI Mode hurts Search revenue; -15% $500B drop worst since 2004 IPO. AI Bear Market coming? · Latent Space

Amazon Is Gutting Its AI Division After Sustained Failure · Hacker News: agents

Anthropic blocks users for anything touching training larger models · Hacker News: MCP

Build specialized multimodal models with LFM3 · Hugging Face -blogi

LiquidAI/LFM3-1.7B · HF trending -mallit

Ollama v0.40.12 · Ollama-julkaisut

v0.40.13 · Ollama-julkaisut

v0.40.13-rc1: llama.cpp: workaround Metal unary crash · Ollama-julkaisut

llama.cpp v0.8.0 · llama.cpp-julkaisut

llama.cpp b11958: server : support tool calls in /v1/realtime · llama.cpp-julkaisut

b12013: quant: allow quantization for LFM3-Audio codec decoder · llama.cpp-julkaisut

b12027 · llama.cpp-julkaisut

Trending: nvidia/NVIDIA-Nemotron-3.5-Transcribe-30B-A4B · HF trending -mallit

NVIDIA Nemotron 3.5 Transcribe and Granary v2 · Hugging Face -blogi

baidu/Unlimited-OCR-2 · HF trending -mallit

How researchers are adapting olmOCR 2 for historical and archaeological archives · Ai2 Blog

tencent/Hy-SimulMT-30B-A3B-FP8 · HF trending -mallit

Dessa: Conversation Model trained only on speech · Hacker News: MCP

Introducing AnySpeechRecognizer and AnyDecisionModel: Open Models for Speech and Decisions on Apple Platforms · Hugging Face -blogi

b11928 · llama.cpp-julkaisut

b12023 · llama.cpp-julkaisut

Trending: unsloth/flint-GGUF · HF trending -mallit

Trending: orcarouter/DeepSeek-V4.1-Flash-1017-abliterated-GGUF · HF trending -mallit

b11929 · llama.cpp-julkaisut

b11930 · llama.cpp-julkaisut

b11902: sycl: add set_tensor_2d for faster tensor-split loading · llama.cpp-julkaisut

b11980 · llama.cpp-julkaisut

llama.cpp b11963 · llama.cpp-julkaisut

b11986: opencl: initial support for iq2 · llama.cpp-julkaisut

b12005 · llama.cpp-julkaisut

b12020 · llama.cpp-julkaisut

b12031 · llama.cpp-julkaisut

b12030 · llama.cpp-julkaisut

b11927 · llama.cpp-julkaisut

b11983 · llama.cpp-julkaisut

b11933 · llama.cpp-julkaisut

b12006: memory: refuse rollbacks past the snapshots of the last ubatch · llama.cpp-julkaisut

b12007 · llama.cpp-julkaisut

b12025 · llama.cpp-julkaisut

b12019 · llama.cpp-julkaisut

b12021 · llama.cpp-julkaisut

b12022 · llama.cpp-julkaisut

v0.33.0rc0 · vLLM releases

vLLM v0.33.0rc1 · vLLM releases

v0.33.0rc2 · vLLM releases

v0.33.0rc3 · vLLM releases

v0.33.0 · vLLM releases

n8n@2.46.0 · n8n releases

n8n@3.0.0 · n8n releases

n8n@2.46.1 · n8n releases

n8n@2.46.2 · n8n releases

n8n@2.46.3 · n8n releases

n8n@2.45.7 · n8n releases

n8n@1.123.89 · n8n releases

n8n@1.123.90 · n8n releases

langgraph==1.2.18 · LangGraph releases

langgraph-sdk==0.4.11 · LangGraph releases

langgraph-checkpoint==4.3.1 · LangGraph releases

langgraph-checkpoint-sqlite==3.1.3 · LangGraph releases

Release v0.67.0-nightly.20261031.g6814e00d6 · Gemini CLI releases

Memory-Enforced GPU Sharing in Kubernetes with NVIDIA CUDA 13.4 · NVIDIA Technical Blog

How NVIDIA Mission Control 2.5 Brings Autonomous Recovery to Kubernetes · NVIDIA Technical Blog

Build a Validated Kubernetes Runtime for NVIDIA Dynamo · NVIDIA Technical Blog

New global AI computing capacity is doubling every 9 months · Epoch AI

Announcing Artificial Analysis Intelligence Index v5 · Artificial Analysis

GPT-6.1 Terra (Max) Intelligence Index: 48 · Artificial Analysis

GPT-6.1 Terra (Xhigh) Intelligence Index: 45 · Artificial Analysis

GPT-6.1 Terra (High) Intelligence Index: 44 · Artificial Analysis

GPT-6.1 Terra (Medium) Intelligence Index: 41 · Artificial Analysis

GPT-6.1 Terra (Low) Intelligence Index: 36 · Artificial Analysis

GPT-6.1 Terra (Non-reasoning) Intelligence Index: 33 · Artificial Analysis

Kimi K3.1 Pro Intelligence Index: 55 · Artificial Analysis

Gemini 4 Carbon (High) evaluation results · Artificial Analysis

Mistral Large 4 (Medium): Intelligence Index · Artificial Analysis

DeepSeek V4.1 Flash 1017 (Non-reasoning) · Artificial Analysis

Quoting Ted Sanders · Simon Willison

Artificial Analysis v5 and the benchmarker's dilemma · Interconnects

Introducing Candela: AI risk benchmarking-as-a-service for developers and deployers · MLCommons

Test Your Own AI Against Real-World Risks and Attacks with AILuminate Safety 1.2.1 and Jailbreak 1.1 · MLCommons

Don't assume composability in AI safety. Deployments need their own testing. · MLCommons

Can Frontier Models See Past a Decoy? Elorian Tests Visual Reasoning · Latent Space

Can we trust LLMs to judge whether a research idea is novel? · Ai2 Blog

From synthetic enterprise data evaluations to real Microsoft Copilot AI confidence · Microsoft Research

Five Qwen3.8-27B models match Fable 5 Using GVS5H Harness · Hacker News: MCP

Memdoor: A team of 5 Qwen3.8-27B models beats Fable 5 on LiveCodeBench Hard · Hacker News: MCP

GPT-5.6 Sol can beat Factorio on peaceful mode · Hacker News: MCP

SDK v0.2.0 · Cline releases

CLI v3.0.74 · Cline releases

Desktop v0.0.49 · Cline releases

v4.1.26 · Cline releases

Claude Code 2.1.311 · Claude Code changelog

Claude Code 2.1.312 · Claude Code changelog

Claude Code 2.1.314 · Claude Code changelog

Claude Code 2.1.316 · Claude Code changelog

How Anthropic ships the #1 LLM in the World safely: Separating the Brain from the Hands · Latent Space

Agents Prompt People: Using Agents for Long-Horizon Financial Work · Latent Space

docs: add client suport matrix for Tasks · MCP-määrittely commits

docs: mark Tasks support for Claude Code and Claude.ai · MCP-määrittely commits

code-review: bring in the newer command from anthropics/claude-code · Claude Code plugins/marketplace (virallinen)

code-review: add version 1.0.1 · Claude Code plugins/marketplace (virallinen)

Release v0.67.0-nightly.20261027.g363a13119 · Gemini CLI releases

Gemini CLI v0.67.0-preview.0 · Gemini CLI releases

Release v0.67.0-nightly.20261030.gb633c0703 · Gemini CLI releases

Merge pull request #5322: move a file across filesystems by copying it on EXDEV · MCP servers commits

fix(filesystem): create the copy owner-only, and keep the source unless the copy is in place · MCP servers commits

fix(filesystem): keep a source written to in place during a cross-device move · MCP servers commits

fix(filesystem): keep sub-millisecond times on a cross-device copy · MCP servers commits

fix(filesystem): keep a pre-1970 time on a cross-device copy · MCP servers commits

b11934 · llama.cpp-julkaisut

Quoting Kyle Kingsbury · Simon Willison

Claude Code regularly sends each background agent's transcript to its model · Hacker News: Claude Code

JetBrains/thinkrail · GitHub trending

rust-v0.165.0 · OpenAI Codex CLI releases

rust-v0.167.0-alpha.2 · OpenAI Codex CLI releases

GitHub Octoverse 2026 · Simon Willison

The October inflection point for vibe game development · Simon Willison

Cisco gives engineer unlimited AI tokens to rewrite legacy code · Hacker News

AI promised less work. I've never worked more · Hacker News: agents

AI isn't replacing SREs, it's deskilling them · Hacker News: agents

Why are good developers so wary of agentic development? · Hacker News: agents

We ran AI agent swarms in Minecraft and Mindustry · Hacker News: agents

AI agent swarms in Minecraft fail to build civilization, drown in bureaucracy · Hacker News: agents

No AI, 52 min. Build all Linux Files utils in C by Hand [video] · Hacker News: agents

Merge pull request #3398 from modelcontextprotocol/sep/release-cycle-updates · MCP-määrittely commits

docs: clarify release manager access and record SEP-3398 acceptance · MCP-määrittely commits

Merge pull request #3392: SEP-3392: Extension Lifecycle · MCP-määrittely commits

Merge pull request #3364: clarify extension negotiation for initialization-based versions · MCP-määrittely commits

Document identity provider support for enterprise managed auth (#3306) · MCP-määrittely commits

Prohibit third-party notetaker bots in community meetings (#3457) · MCP-määrittely commits

Restructure Security IG as a routing group (#3331) · MCP-määrittely commits

docs: move annotations out of resource in embedded resource example · MCP-määrittely commits

docs: fix client-credentials samples and pin the issuer (#3472) · MCP-määrittely commits

docs: handle non-text and multiple tool-use blocks in the client tutorial (#3476) · MCP-määrittely commits

docs: fix Java and Kotlin tutorial samples that fail as printed (#3481) · MCP-määrittely commits

docs: update the Rust client tutorial to rmcp 3 (#3484) · MCP-määrittely commits

docs: mark Skills support for MCPJam in the extension matrix (#3467) · MCP-määrittely commits

Install marketplace plugins from claude-plugins-official · Claude Code repo commits

Give ten plugins explicit versions · Claude Code plugins/marketplace (virallinen)

runway-api: point at claude/ and update to Runway Dev 4.0.0 (#6951) · Claude Code plugins/marketplace

bump(runway-api): fe79eab0 → 6e98d3f3 (#6954) · Claude Code plugins/marketplace

expo/skills · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

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

nightly-labs/openbot · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

CopilotKit/OpenBot · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Orkas-AI/Orkas · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Donchitos/Claude-Code-Game-Studios · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

dbreunig/nesformer · Simon Willison

Show HN: A language model that lives in a boot sector · Hacker News: MCP

A tiny world model running games on a Raspberry Pi Pico 2 W · Hacker News: MCP

OpenAI's agents talked to each other through the CPU Cache · Hacker News: agents

Tell HN: Claude Code adds your email to the prompt, the LLM sends it to websites · Hacker News: Claude Code

fix(filesystem): check a retargeted allowed-directory symlink before creating under it · MCP servers commits

Merge pull request #5333: fix(git): GHSA-w766-5x43-whvw · MCP servers commits

chore: stamp Python CalVer versions for 2026.10.28 · MCP servers commits

How We Took Over ChatGPT Accounts · Simon Willison

AI can turn security fixes into working exploits before users receive an update · Epoch AI

Was I wrong about open models & cyber risks? · Interconnects

Can AI coding agents rebuild biological AI tools from a paper? · UK AI Security Institute blog

Pretraining data filtering in open-weight biological models · UK AI Security Institute blog

Kyberturvallisuuskeskuksen viikkokatsaus 44/2026 · Kyberturvallisuuskeskus

togg53192-cmd/jailbreaks · GitHub trending -syöte

Anthropic's Mythos-class model allegedly leaked on Hugging Face · Hacker News: MCP

b11935 · llama.cpp-julkaisut