Analyysi

Tekoälyn voimistamat kyberuhat: vaikutukset rahoitusvakauteen

Rahoitusvakaus | 07.07.2026 | Marko Buuri, Ilari Määttä, Heli Snellman

KIRJOITTAJAT



Marko Buuri
Neuvonantaja



Ilari Määttä
Ekonomisti



Heli Snellman
Vanhempi neuvonantaja

Ohjelmistovirheiden etsiminen muuttui keväällä 2026. Johtavat tekoälyalan toimijat toivat lyhyen ajan sisällä saataville kielimalleja, jotka löytävät vakavia haavoittuvuuksia aiempaa nopeammin ja kattavammin. Kehitys voimistaa rahoitusalan kriittiseen infrastruktuuriin kohdistuvia kybertoimintaympäristön riskejä ainakin lyhyellä aikavälillä. Euroopan järjestelmäriskikomitea julkaisi 7.7.2026 varoituksen kehittyneimmistä tekoälymalleista aiheutuvista systeemisistä kyberriskeistä.



Ilmiön tausta

Keväällä 2026 kiihtyi keskustelu siitä, kuinka kehittyneet tekoälymallit löytävät haavoittuvuuksia eri järjestelmistä ja lisäävät kyberuhkia.[1] Anthropicin Claude Mythos Preview -malli annettiin valikoidulle joukolle teknologia-, tietoturva- ja rahoitusalan toimijoita käytettäväksi ohjelmistojen turvallisuuden parantamiseen (Project Glasswing[2]). Myös OpenAI julkaisi vuoden alkupuoliskolla tietoturvaan erikoistuneita tekoälymalleja ja kumppaniohjelmia.[3] Tietoturvahuolet saivat Yhdysvaltojen hallinnon asettamaan ja sittemmin purkamaan käyttörajoituksia uusimmille malliversioille.[4]

Kyky löytää haavoittuvuuksia seuraa näiden tekoälymallien edistyneistä kyvyistä ymmärtää ja muokata monimutkaista ohjelmistokoodia. Kyvyn voi odottaa yleistyvän nopeasti myös muiden kuin yhdysvaltalaisten valmistajien kielimallipohjaisissa tekoälytuotteissa ja työkaluissa vuosien 2026–2027 aikana. Kiinan arvioidaan olevan nopea seuraaja tällä alueella.[5]

Kehityksellä on kaksi vastakkaista seurausta ohjelmistojen turvallisuudelle. Ohjelmistojen ja kriittisen infrastruktuurin turvallisuudesta vastaavat tahot saavat työkalun, jolla heikkoudet voidaan löytää ja korjata aiempaa nopeammin ja kattavammin. Työkalujen vaikutus kybertoimintaympäristön turvallisuuteen on siten selvästi myönteinen. Samalla vastaava kyky on väistämättä myös rikollisten ja valtiollisesti tuettujen toimijoiden ulottuvilla, ja on syytä olettaa, että nämä toimijat käyttävät työkalujaan samoihin tarkoituksiin kuin nytkin: geopoliittisten vastustajien vakoiluun, kriittiseen infrastruktuuriin tunkeutumiseen, tuhotöihin osana kansainvälisiä kriisejä sekä yritysten kiristämiseen ja muihin taloudellista etua tavoitteleviin toimiin. Lyhyellä aikavälillä hyöty voi kallistua väärinkäyttäjien puolelle, mikäli tärkeimpiä ohjelmistojamme ei voida turvata ajoissa.

Ohjelmistohaavoittuvuuksien mittakaava

Useimmat tuotteet ja palvelut sisältävät niin paljon ohjelmakoodia, ettei kaikkia haavoittuvuuksia ole pystytty tunnistamaan luotettavasti aiemmilla testaustyökaluilla. OpenAI:n julkisen koodin projekteilla tekemän analyysin mukaan karkeasti arvioiden noin prosentti koodimuutoksista tuo mukanaan jonkinlaisen virheen.[6] Osa näistä virheistä on tietoturvahaavoittuvuuksia. Tämän vuoksi jopa avoimen lähdekoodin projekteissa on edelleen vuosikymmeniä vanhoja vakavia haavoittuvuuksia.

Julkisiin haavoittuvuuskantoihin kirjattiin vuonna 2025 noin 49 000 uutta haavoittuvuutta.[7] Valtaosa näistä ei ole vakavia. Vakavimpia ovat haavoittuvuudet, joita hyökkääjät ehtivät käyttää

ennen kuin niihin on saatavilla korjaus (ns. nollapäivähaavoittuvuus ja -hyökkäys). Googlen uhkatiedusteluyksikkö havaitsi 90 tällaisen haavoittuvuuden aktiivista hyväksikäyttöä vuonna 2025; vuonna 2024 vastaava luku oli 78.[8] Hyökkääjät kohdistivat lähes puolet vuoden 2025 tapauksista yritysten käyttämiin alustateknologioihin, erityisesti verkko- ja tietoturvalaitteisiin. Syy on looginen: yhden tällaisen laitteen murtaminen avaa pääsyn laajaan järjestelmäkokonaisuuteen. Esimerkiksi verkon reunalaitteiden tehokas tietoturvavalvonta[9] edellyttää erilaista teknologiaa kuin työntekijöiden päätelaitteiden, eikä vastaavaa valvontakykyä useimmissa reunalaitteissa ole [10], mikä on osaltaan myötävaikuttanut ilmiöön.

Haavoittuvuuksia hyödyntävät valtiolliset toimijat ja ammattimaiset rikollisryhmät. Tähän asti haavoittuvuustutkinta ja niitä hyödyntävät kohdennetut hyökkäykset korkean tietoturvatason yrityksiin ovat vaatineet tekijältä paljon osaamista ja resursseja sekä kohdistuneet rajattuun toimijajoukkoon. Nyt tekoälytyökalut auttavat haavoittuvuuksien tunnistamisessa ja hyökkäysten automatisoinnissa.

Tekoäly tehostaa esimerkiksi syötteiden käsittelyyn liittyvien ohjelmistovirheiden löytämistä. Tällainen haavoittuvuus perustuu siihen, että ohjelmisto hallitsee puutteellisesti tilanteita, joissa käyttäjä voi tarkoituksella manipuloida ohjelmiston toimintaa omilla viesteillään tai muilla syötteillään, kuten tiedostoilla. Manipulointi voi esimerkiksi harhauttaa verkkokaupan tulostamaan käyttäjälle asiakastietokantansa sisällön.

Tekoälytyökalujen hyödyntäminen näkyy jo löydettyjen haavoittuvuuksien määrässä. Anthropicin toukokuussa 2026 julkaistun väliraportin mukaan Mythos Preview -malli oli löytänyt ensimmäisen kuukauden aikana yhdessä noin 50 kumppanin kanssa yli 10 000 vakavaa, aiemmin tuntematonta haavoittuvuutta kriittisen infrastruktuurin käyttämistä ohjelmistoista.[11] OpenAI puolestaan kertoi sen nykyisiä palveluja edeltäneen mallin myötävaikuttaneen yli 3 000 haavoittuvuuden korjaamiseen.[12] Konkreettisena ja julkisesti raportoituna esimerkkinä Mozilla korjasi huhtikuussa 2026 yhden kuukauden aikana yli 400 Firefox-selaimen tietoturvavirhettä, joka on noin kaksikymmenkertainen määrä tavanomaiseen kuukauteen verrattuna.[13] Löydösten määrä on kasvanut niin suureksi, että ohjelmistoturvallisuuden pullonkaula on siirtynyt: haavoittuvuuksien löytäminen ei ole enää hidastava tekijä, vaan löydösten varmentaminen, ilmoittaminen ja korjaaminen.

Tekoälypalveluiden käyttöä kyberhyökkäyksissä pyritään estämään rakentamalla niihin pidäkkeitä (safeguards), jolloin niiden käyttö kyberturvan osalta on rajatumpaa kuin suljettuihin kumppaniohjelmiin osallistuvilla yrityksillä ja tutkijoilla. Näin on esimerkiksi kesäkuussa 2026 julkaistun Claude Fable 5 -malliversion tapauksessa, jossa valmistaja pyrkii aktiivisesti

rajoittamaan mahdollista kaksoiskäyttöä kyberturvallisuuden, biologian ja kemian alueilla[14]. Tämä tarkoittaa sitä, että palvelu analysoi aktiivisesti käyttäjän pyyntöjä ja mallin vastauksia, ja riskitason ylittyessä keskeyttää tämän mallin käytön.

Vaikka pidäkkeet estävät mallien hyödyntämistä esimerkiksi kyberhyökkäyksissä, muodostavat ne yksittäisen vikapisteen, jonka ohittamisella voisi olla vakavia vaikutuksia. Esimerkiksi Anthropicin tapauksessa riskiä rajoittaa se, että valmistaja valvoo palveluidensa käyttötapoja ja on aiemmin tunnistanut ja estänyt väärinkäyttöä[15]. Kuitenkaan skenaariota, jossa pidäkkeet kierretään (nk. prompt jailbreak), ei voi poissulkea, sillä väärinkäyttöä on tapahtunut aikaisemmissakin malleissa ja palveluissa[16].

Ilmiön vaikutukset rahoitusvakauteen

Rahoitusala kuuluu toimialoihin, joihin kohdistuu eniten kyberhyökkäyksiä.[17] Kyberhyökkäys voi uhata rahoitusvakautta erityisesti silloin, kun se kohdistuu finanssimarkkinoiden infrastruktuureihin, keskuspankkeihin ja muihin sellaisiin järjestelmän kannalta merkittäviin yrityksiin tai kriittisiin palveluntarjoajiin, joita ei voi nopeasti korvata. Koska systeemisesti merkittävät infrastruktuurit ovat tiiviisti kytkeytyneitä toisiinsa, niihin kohdistuva hyökkäys voi levitä sekä kotimaisille että kansainvälisille markkinoille.

Lyhyellä aikavälillä tekoäly lähinnä voimistaa ja muuttaa olemassa olevia rahoitusvakausriskejä. Hyökkääjän kyvyssä muutos näkyy ennen kaikkea neljällä tavalla: monivaiheisten ja kompleksisten hyökkäysten toteuttaminen helpottuu, hyökkäysten nopeus ja skaalautuvuus kasvavat, osaamiskynnys madaltuu tuntuvasti, ja vanhan koodikannan riski (tekninen velka) korostuu, kun aiemmin piiloon jääneet haavoittuvuudet tulevat helpommin löydetyiksi. Nämä kykymuutokset heijastuvat rahoitusvakauteen kolmen näkökulman kautta:

1. Samanaikaisuus ja mittakaava

   Tähän saakka haavoittuvuudet ovat voineet maata syvällä koodissa vuosia, ja hyökkäykset ovat olleet työläitä valmistella. Kun työkalu käy läpi valtavat koodimäärät ja rakentaa hyökkäyksen nopeasti, samanaikaiset hyökkäykset niin suuriin kuin pieniin toimijoihin eri puolilla maailmaa muuttuvat mahdollisiksi.

   Monet sektorin toimijoista tukeutuvat samoihin johtaviin alustateknologioihin ja perusohjelmistoihin, joten yhdestä laajalti käytetystä tuotteesta löytyvä haavoittuvuus on lähtökohtaisesti hyödynnettävissä yhtä aikaa useissa sen asennuksissa. Tekoälyä voi hyödyntää tällaisten hyökkäysten automatisointiin. Lyhyellä aikavälillä tämä uhkaa koko digitaalisen toimintaympäristön eheyttä: jos tuotteiden ja palveluiden valmistajat eivät ehdi korjata ohjelmistojaan ennen kuin rikolliset murtavat ne, syntyy aikaikkuna, jonka aikana laajat

häiriöt ovat mahdollisia. Kriittiselle infrastruktuurille tämä on olennainen riski. Rahoitusvakauden kannalta on tärkeää, että sektorin suurten toimijoiden ohella myös keskisuuret ja pienet yritykset onnistuvat pitämään järjestelmänsä turvassa.

Anthropicin väliraportti havainnollistaa pullonkaulaa konkreettisesti: vaikka haavoittuvuuksia löytyy nopeasti, yhden vakavan virheen korjaaminen kestää keskimäärin noin kaksi viikkoa.[18] Tuotteiden kaupalliset valmistajat ja avoimen lähdekoodin vapaaehtoiset ylläpitäjät voivat kuormittua löydösten määrästä. Kun haavoittuvuuksia löytyy nopeammin kuin niitä ehditään korjata, tunnistettu korjausvelka kasvaa. Systeemisen riskin mahdollisuus korostuu juuri tässä aikaikkunassa.

2. Keskittymä ja korvattavuus

Monet rahoitusalan palvelut ovat keskittyneet pienelle joukolle kansainvälisiä palveluntarjoajia. Kun vakavia haavoittuvuuksia löydetään tekoälyn avulla aiempaa suurempia määriä, kuormitus kohdistuu juuri näiden toimittajien rajalliseen korjaus- ja tukikapasiteettiin. Vaikka valmistuneet korjaukset ovat lähtökohtaisesti kaikkien asiakkaiden saatavilla samanaikaisesti, keskeinen kysymys on, pystyvätkö palveluntarjoajat tukemaan eri alueilla ja eri kokoluokissa olevien asiakkaidensa korjausten käyttöönottoa yhtä aikaa. Koska keskittyneitä palveluntarjoajia ei voi nopeasti korvata, niiden kapasiteetista voi käytännössä syntyä kilpailua. Tässä kilpailussa parhaiten pärjäävät ne toimijat, joilla on varaa maksaa eniten tai riittävästi omaa osaamista. Korjausresursseista joutuvat kilpailemaan myös eri toimialat keskenään. Viimeisenä korjauksia odottavat toimijat eivät kuitenkaan ole irrallisia muusta järjestelmästä. Rahoitusvakauden kannalta riskinä on, että häiriö yhdessäkin sektorin toimijassa heijastuu vastapuoliin ja palveluihin. Keskittymä, korvaamattomuus ja keskinäiskytkennät voivat siten yhdessä muuttaa paikallisen ongelman järjestelmänlaajuiseksi.

3. Luottamus

Laajamittaiset samanaikaiset tekoälyteknologialla voimistetut hyökkäykset horjuttaisivat luottamusta pankkisektoriin ja muihin rahoitusalan toimijoihin, vaikka asiakasvarat pysyisivät turvassa.

Luottamusta heikentävät myös ihmisiin suoraan kohdistuvat hyökkäykset. Kehittyneet mallit mahdollistavat syvävääreennösten, valeinformaation ja kalasteluviestien laajamittaisen tuotannon, jolloin hyökkäysten ei tarvitse enää kohdistua valikoituihin uhreihin, vaan ne voidaan suunnata laajasti digitaalisten palveluiden käyttäjiin. Manipulaation tavoitteena voi olla yksilön tai yrityksen huijaaminen tai esimerkiksi informaatiolle herkkien osakemarkkinoiden heilautus. Tällaisten häiriöiden torjunta edellyttää uudenlaisia työkaluja ja kuluttajien tekoälyosaamisen vahvistamista.

Koska kyseessä on globaali ilmiö, tiedonkulun, toimenpiteiden ja sääntelyn kansainvälinen koordinointi on avainasemassa näiden kysymysten ratkaisemisessa. Tiedonkulussa korostuu kyky jakaa nopeasti tietoa uusista haavoittuvuuksista, hyökkäysmenetelmistä ja torjuntakeinoista

viranomaisten, finanssisektorin toimijoiden ja muiden kriittisten toimialojen välillä. Mitä nopeammin tieto uusista uhkista leviää, sitä paremmin voidaan ehkäistä yksittäisten teknisten ongelmien muuttuminen laajoiksi systeemisiksi häiriöiksi.

Kriittisen infrastruktuurin omistajat avainroolissa

Samat työkalut antavat myös järjestelmien ylläpitäjille mahdollisuuden löytää ja korjata heikkoudet aiempaa nopeammin, ja uusien järjestelmien kehityksessä voidaan välttää sekä vanhoja että uusia haavoittuvuuksia. Kun nämä kyvyt ovat aikanaan kaikkien ohjelmistokehittäjien käytössä osana ohjelmistojen kehitystä ja ylläpitoa, uusien vakavien haavoittuvuuksien pääsy tuotantoon vähenee todennäköisesti tuntuvasti ja voi muuttua harvinaiseksi. Pidemmällä aikavälillä lopputulos voi siten olla aiempaa vahvemmat järjestelmät. Tämä ei kuitenkaan tapahdu itsestään: vanha koodikanta, korjausten käyttöönoton hitaus ja toimijoiden eritasoiset resurssit jättävät pitkäksi aikaa siirtymävaiheen, jonka aikana hyökkääjillä voi olla etu. Siirtymä edellyttää riittäviä investointeja, sillä tekoälykehitys vain kiihtyy ja muut teknologiat, kuten kvanttilaskenta, voimistavat uhkia entisestään.

Kyberturvallisuuskeskus painottaa ennakoivan varautumisen merkitystä.[19] Painopiste siirtyy haavoittuvuuksien löytämisestä niiden korjaamiseen ja häiriöiden rajaamiseen. Kun löydösten määrä kasvaa nopeammin kuin niitä ehditään korjata, keskeistä on kyky priorisoida[20], ottaa korjaukset käyttöön ripeästi ja rajata yksittäisen haavoittuvuuden vaikutukset.

Kaikkia haavoittuvuuksia ei kuitenkaan voida poistaa ennen niiden mahdollista hyväksikäyttöä. Tämän vuoksi rahoitusvakauden kannalta ratkaisevaa on myös operatiivinen resilienssi, eli kyky ylläpitää kriittisiä palveluja häiriöiden aikana sekä palauttaa ne nopeasti toimintakuntoon. Varajärjestelmät, vaihtoehtoiset toimintatavat, palautumiskyvykkyys ja säännöllinen harjoittelu korostuvat tilanteessa, jossa hyökkäysten nopeus ja mittakaava kasvavat.

Toimijoiden oman varautumisen lisäksi Eurooppa tarvitsee kansainvälistä yhteistyötä kyberuhkia torjuttaessa. Tekoälykehitys on globaalia, mutta huippumallien saatavuus oli ainakin kesällä 2026 rajoitettua. Niitä kehittävien yhtiöiden kumppanuusohjelmat ovat suurelta osin yhdysvaltalaisvetoisia, eikä eurooppalainen yritys välttämättä pääse testaamaan tehokkainta yhdysvaltalaista tai kiinalaista mallia johtuen kumppanuusohjelman säännöistä tai mallien vientirajoituksista. Ja vaikka pääsisikin, se ei ehkä halua viedä omaa koodiaan toisella mantereella sijaitsevaan testausympäristöön eikä toisen mantereen toimija välttämättä luovuta tuotettaan eurooppalaisen yrityksen tietojärjestelmiin. Teknologian turvallinen testaus edellyttäisi neutraalia ympäristöä.

Euroopan riippuvuus Euroopan ulkopuolisista järjestelmätoimittajista on suurta, ja sitä tulisi

pyrkiä vähentämään. Teknologinen kehitys ei ole ainoastaan kilpailukykytekijä, vaan myös edellytys strategisen autonomian kannalta. Nykytilanteessa varautuminen nojaa pitkälti kansainvälisen yhteistyön sujuvuuteen. Perustutkimus, osaaminen ja kasvuyritykset ovat Euroopassa huippuluokkaa, mutta niiden kaupallistaminen ja skaalaaminen viedään muualle[21]. Ratkaisuja, kuten yhteisiä pääomamarkkinoita, on yritetty edistää jo pitkään, mutta kehitettävää riittää edelleen[22]. Tekoälyteknologia kehittyy nopeasti ja sillä on jo nyt voimakkaat vaikutukset, mikä kannustaa entistä rohkeampiin ratkaisuihin.

Kyberriskit uhkaavat rahoitusvakautta

Kehittyneet kielimallit löytävät ohjelmistokoodeista haavoittuvuuksia aiempaa nopeammin ja kattavammin. Tämä voi ainakin lyhyellä aikavälillä lisätä rahoitusalan kyberriskejä ja uhata rahoitusvakautta. Yhtäältä uudet tekoälytyökalut auttavat puolustajia tunnistamaan ja korjaamaan vakavia virheitä aiempaa tehokkaammin. Toisaalta ne edesauttavat rikollisten ja valtiollisten toimijoiden mahdollisuuksia toteuttaa laajoja ja samanaikaisia hyökkäyksiä rahoitusalan yrityksiin ja kriittiseen infrastruktuuriin. Sekä rahoitusalan toimijoiden että viranomaisten on vahvistettava varautumista, nopeutettava korjausten käyttöönottoa ja tiivistettävä yhteistyötä kansallisesti ja kansainvälisesti.

Viitteet

1. Euroopan järjestelmäriskikomitea varoitti 7.7.2026, että kehittyneimmät tekoälymallit voivat koetella rahoitusjärjestelmän kyberresilienssiä. Ks. tiedote: Frontier AI models could strain cyber resilience in the financial system, ESRB warns. ↑

2. Ks. https://www.anthropic.com/glasswing. ↑

3. Scaling Trusted Access for Cyber with GPT-5.5 and GPT-5.5-Cyber | OpenAI. ↑

4. Ks. esim. https://www.anthropic.com/news/fable-mythos-access ja Redeploying Claude Fable 5 \ Anthropic ↑

5. https://hai.stanford.edu/ai-index/2026-ai-index-report. ↑

6. Ks. https://openai.com/index/introducing-aardvark/. ↑

7. Ks. https://www.first.org/blog/20251229-Vulnerability-Forecast-Review . ↑

8. Ks. https://cloud.google.com/blog/topics/threat-intelligence/2025-zero-day-review. ↑

9. Ks. https://kyberturvallisuuskeskus.fi/fi/uutiset/riskialttiit-verkon-reunalaitteet-aktiivisten-murtoyritysten-kohteena. ↑

10. Ks. https://cloud.google.com/blog/topics/threat-intelligence/2025-zero-day-review. ↑

11. Ks. https://www.anthropic.com/research/glasswing-initial-update. ↑

12. Ks. https://openai.com/index/scaling-trusted-access-for-cyber-defense/. ↑

13. Ks. https://hacks.mozilla.org/2026/05/behind-the-scenes-hardening-firefox/. ↑

14. Claude Fable 5 on rajoitettu malliversio Mythos 5:stä. https://www.anthropic.com/news/claude-fable-5-mythos-5. ↑

15. Ks. https://www.anthropic.com/news/AI-enabled-cyber-threats-mitre-attack. ↑

16. Ks. https://www.anthropic.com/research/many-shot-jailbreaking. ↑

17. Ks. https://www.imf.org/-/media/files/publications/gfsr/2024/april/english/text.pdf. ↑

18. Ks. https://www.anthropic.com/research/glasswing-initial-update. ↑

19. Ks. https://traficom.fi/fi/uutiset/kyberturvallisuuden-ensimmainen-vuosipuolisko-2026-kyberuhkien-torjuminen-vaatii-vahvempaa-varautumista-ja-resursointia. ↑

20. Esimerkiksi UK NCSC korostaa sellaisten järjestelmien priorisointia, jotka ovat suoraan internetistä tai muuten organisaation ulkopuolelta saavutettavissa ja siten alttiimpia hyökkäyksille: Preparing for a ‘vulnerability patch wave’ | National Cyber Security Centre Tämä tukee riskiperusteista lähestymistapaa, jossa haavoittuvuuksien volyymin lisäksi ↑

tarkastellaan myös niiden hyväksikäytettävyyttä, kohdejärjestelmän kriittisyyttä ja mahdollista vaikutusta toiminnan jatkuvuuteen.

21. Ks. https://www.ecb.europa.eu/press/fie/box/html/ecb.fiebox202605_04.en.html. ↑

22. Ks. https://commission.europa.eu/topics/competitiveness/draghi-report_en. ↑

Tässä artikkelissa esitetyt mielipiteet ovat kirjoittajien omia eivätkä välttämättä edusta Suomen Pankin näkemystä.

Asiasanat

kyberriskit, rahoitusvakaus, tekoäly