

Claude Mythos ja Project Glasswing: miks AI-superhäkker on seadnud tehnoloogiamaailma häireseisundisse?

1 kuu tagasi - 16.04.2026 Autor: [AM](#)

[\(CC\) The Conversation](#)

Uued ja võimekamad tehisintellekti mudelid on muutunud igapäevaseks: olgu selleks ChatGPT, Claude'i või Gemini uusim versioon – ikka lubatakse meile uusi funktsioone, mida kasutajad saaksid kohe proovile panna. Seekord on aga kõik teisiti.

Tehisintellekti arendaja Anthropic teatas suure käraga uuest mudelist nimega Mythos, kuid tavakasutajatel sellele ligipääsu pole. [Selle asemel](#) on ettevõtte käivitanud algatuse nimega Project Glasswing, et rakendada mudeli võimekust "hea, mitte kurja teenistusse". [New York Times](#) nimetab mudeli võimekust "hirmutavaks ohumärgiks".

Miks on Anthropic nii ettevaatlik? Varajased raportid viitavad, et Mythos suutis suuniste andmisel väljuda suletud testimiskeskonnast ehk nn liivakastist (*sandbox*) ja [saata selle uurijale e-kirja](#).

See on ehk vaid veidi murettekitav. Kuid olulisem on Anthropicu väide, et Mythos on tuvastanud tarkvaralisi haavatavusi ja vigu „kõikides suuremates operatsioonisüsteemides ja veebibrauserites“.

Peidetud haavatavuste leidmine

Ühe märkimisväärse näitena leidis mudel vea turvalisusele orienteeritud operatsioonisüsteemis OpenBSD, mida kasutatakse laialdaselt tulemüürides ja ruuterites. See viga oli jäänud märkamatuks tervelt 27 aastat. Anthropicu sõnul avastas Mythos ka 16 aastat vana haavatavuse FFmpeg-s – see on vähetuntud, kuid kriitiline tarkvarakomponent, mis aitab arvutitel ja rakendustel hallata audio- ja videofaile.

Lisaks teatati, et Mythos leidis mitu haavatavust Linuxi tuumas (kernel) ning suutis need aheldada viisil, mis andis ründajale täieliku kontrolli seadme üle.

Anthropicu sisekontroll (mis ei ole veel sõltumatut kinnitust leidnud) näitas, et Mythos on varasematest mudelitest oluliselt edukam tarkvaravigade muutmisel toimivateks rünnakukoodideks (*exploits*). Ettevõtte hinnangul on mudelil suur tehniline potentsiaal, kuid see nõuab äärmist valvsust. Raport järeldab, et kuigi mudel ei kipu iseseisvalt "mässama", võib see inimese juhiste järgi korda saata suurt kahju.

Miks Anthropic Mythost luku taga hoiab?

Anthropic otsustas mudelit mitte avalikustada just selle ohtliku võimekuse tõttu. Selle asemel kutsuti ellu Project Glasswing.

See algatus koondab laia koalitsiooni tehnoloogiagigante (Microsoft, Amazon, Google, Apple, Cisco, NVIDIA), avatud lähtekoodiga organisatsioone (Linux Foundation) ja finantsmaailma suurtegijaid (JPMorgan Chase). Eesmärk on suunata Mythose võimekus küberkaitsesse, mitte kuritarvitustesse.

Idee on lihtne: anda kaitsjatele edumaa kriitilise tarkvara nõrkuste leidmiseks ja parandamiseks enne, kui sarnased AI-võimekused muutuvad ründajatele laialdaselt kättesaadavaks.

Lugedes ridade vahelt

See ei ole esimene kord, kui AI-ettevõtte leiab, et mudel on avalikustamiseks liiga võimas. 2019. aastal, ammu enne ChatGPT-ajastut, tegi OpenAI midagi sarnast mudeliga GPT-2 (toona oli Anthropicu praegune juht Dario Amodei üks OpenAI juhtivteadlasi).

Anthropicu teadaannetesse tasub aga suhtuda tõsiselt. Ettevõtte on avaldanud ebaharilikult üksikasjalikku materjali mudeli kohta, mida nad isegi ei lase välja. Raportite kohaselt kutsusid USA ametivõimud Washingtoni kokku suurpankade juhid, et arutada Mythosega seotud küberriske.

Samas peab säilitama kriitilise meelega, sest välised osapooled ei saa Anthropicu väiteid veel kontrollida. Ettevõtte väidab, et üle 99% leitud haavatavustest on alles avalikustamata, kuna neile pole veel turvapaiku loodud. See on vastutustundlik käitumine, kuid see tähendab ka, et avalikkus peab Anthropicut pimesi usaldama.

Mida Mythos küberjulgeoleku tuleviku jaoks tähendab?

Küberturvalisuse ebaõnnestumistel on reaalsed tagajärjed. Austraalias paljastas Optuse andmeleke 9,5 miljoni inimese isikuandmed. Medibanki juhtumi puhul lekkisid tundlikud terviseandmed isegi dark web'i. Need ei olnud lihtsalt andmebaasi probleemid, vaid privaatsuse ja usalduse kriisid.

Just seepärast on Mythos oluline. See võib muuta küberturvalisuse majanduslikku loogikat.

Minevikus jäid tõsised haavatavused sageli peitu vaid seetõttu, et keegi ei leidnud neid üles – see nõudis haruldasi oskusi, kannatlikkust ja aega. Kui aga Mythose-sarnased mudelid suudavad skaneerida interneti "nähtamatut torustikku" (operatsioonisüsteeme, brausereid, ruutereid) enneolematu mahu, võib seni spetsiifilisi teadmisi nõudnud häkkimine muutuda rutiinseks ja automatiseeritud protsessiks.

Ettevõtete jaoks on Mythos kahe teraga mõök: see aitab kiirelt leida vead oma koodis, kuid tekitab hirmu, et ründajad võivad nendeni jõuda esimesena. See ei puuduta ainult tech-ettevõtteid, vaid kõiki teenuseid, millest me sõltume – elektrist ja veest kuni lennuliikluse, panganduse ja haiglateni.

Mis saab edasi?

Seni on küberturvafirmad olnud Mythose teemal avalikult märkimisväärselt vaiksed. Paljud näivad ootavat ja vaatavat, soovimata paljastada oma seisukohta juhuks, kui mudel peaks leidma nõrkusi just nende süsteemides.

Kuid arengud nagu Mythos on põhjus lõpetada suhtumine küberturvalisusse kui "keegi teise probleemi". Tavakasutaja jaoks on vastus lihtne: baas-küberhügieen on olulisem kui kunagi varem.

Uuendage oma telefone, sülearvuteid, brausereid ja ruutereid. Vahetage välja seadmed, mida tootja enam ei toeta. Kasutage paroolihaldurit ja lülitage sisse mitmetasemeline autentimine (MFA). Ärge ignoreerige turvapaikade teavitusi.

Need on esimesed sammud. Nende taga peituvad aga keerulisemad küsimused: kellel peaks olema ligipääs võimsatele AI-mudelitele, kes teostab järelevalvet ja kes otsustab, millised on need "õiged käed", kuhu selline tehnoloogia usaldada?

Autorid:

Stan Karanasios

Infosüsteemide professor, Queenslandi Ülikool

Saeed Akhlaghpour

Äriinfosüsteemide dotsent, Queenslandi Ülikool

Artikkel on algselt avaldatud väljaandes The Conversation ja taasavaldatud Creative Commons'i alusel.

Eesti küberidu Phishbite kaasasutaja Urmo Keskele kommentaar



Phishbite'i kaasasutaja Urmo Keskel sõnul puudutab selline areng ka ettevõtteid, sest rünnatavad süsteemid on laialdaselt kasutusel just organisatsioonide igapäevatoos.

„Esmalt sihitakse tavaliselt kõige levinumaid lahendusi: brausereid, operatsioonisüsteeme, VPN-e, suhtlusrakendusi ja e-posti. Kui sellise võimekusega tööriist jõuab kurjategijate kätte, satuvad suuremasse ohtu ka ettevõtete enda arendatud või tellitud lahendused, mis on sageli vähem testitud ja ebaühtlase turvatasemega,“ selgitas ta.

Keskel lisas, et nullpäeva turvanõrkuste leidmine on juba täna kiirenev trend. „Näiteks Chrome'i puhul parandatakse igal aastal kümnekond sellist haavatavust ning tempo kasvab. Kui AI-võimekus laiemalt levib, ei kasva see enam lineaarselt,

vaid eksponentsiaalselt.“

Tema sõnul on paratamatu, et sellised tööriistad jõuavad varem või hiljem ka pahatahtlike osapoolte kätte. „Dzinni ei ole võimalik pudelisse tagasi panna. Maksimaalselt saab võita aega ja selle aja jooksul valmistuda, täpselt seda teevad praegu ka tehnoloogiaettevõtted.“

Samas rõhutab Keskel, et valmisolek ei ole ainult tehniline küsimus. „Praktika näitab, et esmane ligipääs süsteemidele saadakse sageli olukorras, kus töötajad kasutavad samu paroole ja seadmeid nii tööks kui ka isiklikuks tarbeks. Kui üks neist kompromiteeritakse, kandub risk kohe üle ka töökeskkonda.“

Eriti haavatavad on tema sõnul mobiiliseadmed. „Paljudes ettevõtetes puudub nende üle keskne haldus ja nähtavus – ei teata, kas seadmed on ajakohased või milliseid rakendusi kasutatakse. Kui samas seadmes on nii töö- kui erakontod, võib see kujuneda otseseks sissepääsuks ettevõtte süsteemidesse.“

Keskel rõhutab, et seetõttu on iga töötaja küberhügieen ja teadlikkus kriitilise tähtsusega.

„Oluline on mõista ka seda, et Anthropicu mudeli puhul ei olnud turvanõrkuste leidmine algne eesmärk – see ilmnes kõrvalnähtuna. See näitab, kui kiiresti võivad sellised võimekused tekkida ka mujal. Ühelt poolt aitab see leida vigu, mida inimesed pole aastaid märganud, kuid teisalt tähendab see, et neid saab ka enne parandamist ära kasutada.“

Tema hinnangul liigub küberturvalisus selgelt uude faasi. „Ründed muutuvad automatiseerituks ja skaleeritavaks. See tõstab lati mitte ainult tehnoloogiale, vaid ka inimeste käitumisele. Ettevõtted, kes ei suuda oma töötajate igapäevast digitaalset käitumist turvalisemaks muuta, jäävad kõige lihtsamateks sihtmärkideks.“

- [Tegijad](#)
- [Lahendused](#)

- [Tarkvara](#)
- [Tehisintellekt](#)

Pilt

