Claude Mythos paljastas üle 10 000 kriitilise turvaaugu

2 tundi tagasi - 24.05.2026 Autor: AM

Ühel päeval jalutab majja üliintelligentne lukuabi robot, kes skaneerib hoone vaid mõne sekundiga ja teatab: "Teie uksed on tegelikult lahti. Leidsin just tuhandeid viise, kuidas siia sisse murda." Just taolise äratuse osaliseks sai ülemaailmne tehnoloogiakogukond, kui tehisintellekti arendaja Anthropic lasi küberruumi lahti oma uusima loomingu – Claude Mythos AI.

Eelmisel kuul käivitatud kaitseotstarbeline algatus nimega Project Glasswing on andnud üsna uskumatu tulemuse. Vaid loetud nädalatega suutis see tehisaru üles leida üle 10 000 kõrge või kriitilise tähtsusega turvaaugu tarkvarades, mis hoiavad püsti meie digitaalset maailma. See pole lihtsalt järjekordne tarkvarauuendus, vaid väga suur muutus selles, kuidas me oma digivara kaitseme. Kas tehisintellekt on muutunud küberkurjategijate jaoks halvimaks unenäoks?

Digitaalne verekoer ei maga enam kunagi

Project Glasswingi eesmärk on lihtne, kuid mastaapne: turvata kriitilist globaalset tarkvara-infrastruktuuri. Anthropic andis umbes 50 hoolikalt valitud partnerile varajase ja eksklusiivse ligipääsu mudelile Claude Mythos Preview. Tegu on n-ö digitaalse verekoeraga, kes suudab autonoomselt nuusutada välja nõrkusi laialt levinud tarkvarakoodis, olles pahalastest alati sammu võrra ees.

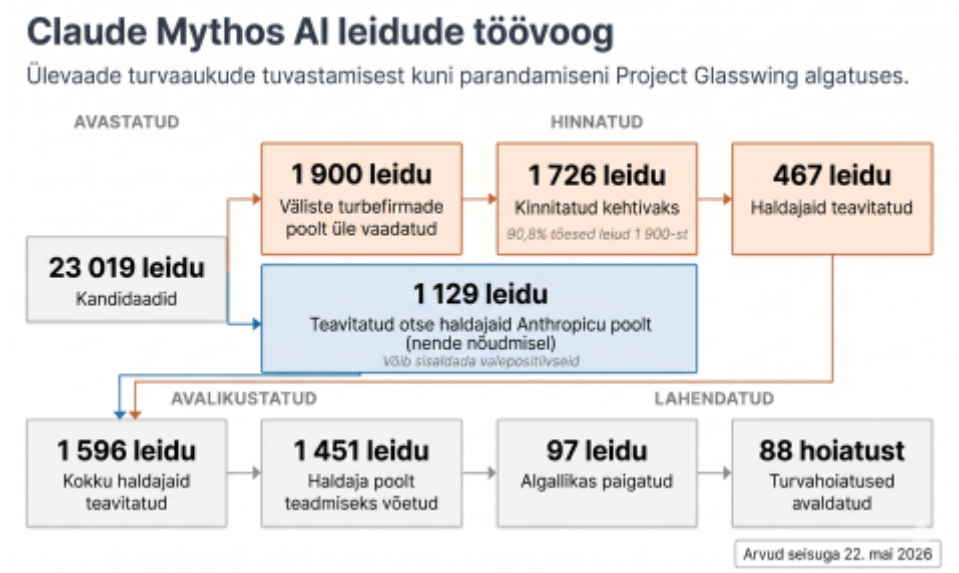
Tulemused räägivad enda eest. Ainuüksi avatud lähtekoodiga (*open-source*) projektides tuvastati 6202 potentsiaalset haavatavust. Pärast põhjalikumat analüüsi kinnitati, et neist lausa 1726 on reaalsed ja töötavad turvaaugud, millest 1094 klassifitseeriti kõrge või kriitilise raskusastmega ohuks. See on olukord, kus tehisintellekt töötleb andmeid ja leiab seoseid koodiridade vahel kiirusega, mis käiks inimmõistusele üle jõu.

Üks eredamaid näiteid on kriitiline viga krüptograafiateegis WolfSSL (CVE-2026-5194, CVSS skooriga 9.1). Keerulise tehnilise žargooni taga peitub stsenaarium, mis on võrreldav meistriklassi dokumendivõltsijaga: see turvaauk oleks lubanud küberpättidel luua valesertifikaate ja esineda legitiimse teenusena. Tavalisele internetikasutajale tähendab see olukorda, kus sa logid enda arvates sisse oma

kodupanka, kuid tegelikult annad oma andmed ja raha otse kurjategijale.

Tänaseks on need ja paljud teised vead (kokku 97 parandust ja 88 ametlikku hoiatust) juba likvideeritud.

Image



Parandamine on raskem kui leidmine

Kui vigade leidmine käib tehisintellekti abil nüüd lennates, siis nende parandamine nõuab endiselt inimeste aega ja vaeva.

> „Haavatavuste leidmise suhteline lihtsus võrreldes nende parandamise keerukusega on küberjulgeoleku jaoks suur väljakutse," tõdes Anthropic pressiteates. „selle väljakutse edukas ületamine muudab meie tarkvara varasemast palju turvalisemaks."

Seda suundumust kinnitavad ka teised tehnoloogiahiiud. Microsoft nendib, et tehisintellekti abiga leitakse üha enam turvaauke ning seetõttu ootavad nad, et igakuiste tarkvarapaikade maht kasvab lähiajal märkimisväärselt. Oracle on aga äsja üle läinud igakuisele paikamistsüklile, et kriitilistele turvaprobleemidele kiiremini reageerida.

Autonoomne küberturvalisuse platvorm XBOW ei hoia Mythos Preview'd kiites sõnu kokku, nimetades seda "suureks edasiminekuks", mis on "märkimisväärselt parem kui varasemad mudelid haavatavuste kandidaatide leidmisel" ning "osav lähtekoodi analüüsimisel turvalisuse vaatenurgast".

Rohkem kui lihtsalt veaparandaja

Mythos Preview võimekus ei piirdu aga vaid koodi lugemisega. Ühel juhul kasutas Project Glasswingi partnerist pank seda tehisintellekti mudelit, et avastada ja peatada 1,5 miljoni dollari suurune petturlik pangaülekanne. Küberkurjategija oli sisse murdnud kliendi e-posti kontole ja teinud võltsitud telefonikõnesid, kuid AI suutis anomaalia tuvastada enne, kui raha pöördumatult kadus.

Kuna sellise võimekusega mudelid võivad lähitulevikus muutuda laiemalt kättesaadavaks, kutsub Anthropic tarkvaraarendajaid üles tempot tõstma.

„Võrgukaitsjad peaksid lühendama oma tarkvarapaikade testimise ja juurutamise ajakava," märkis Anthropic teates, „see hõlmab selliseid samme nagu võrkude vaikekonfiguratsioonide tugevdamine, mitmeastmelise autentimise jõustamine ning põhjalike logide pidamine avastamiseks ja reageerimiseks."

Tulevik on kaitsjate poolel?

Anthropic teatas ühtlasi küberverifitseerimise programmi (Cyber Verification Program) käivitamisest. See võimaldab küberturvalisuse spetsialistidel kasutada nende mudeleid ilma tavapäraste tehisintellekti piiranguteta – seda kõike legitiimsetel eesmärkidel, nagu haavatavuste uurimine ja süsteemide vastupidavuse testimine (red teaming).

Sarnase sammu on astunud ka OpenAI oma Daybreak+ i programmiga, kasutades GPT-5.5-Cyber mudelit.

Laiema avalikkuse ette need ülivõimsad mudelid aga niipea ei jõua, sest kardetakse nende väärkasutamist kurjategijate poolt mastaapidel, milleks maailm pole veel valmis.

„Glasswing aitab süsteemselt kõige olulisematel küberkaitsjatel saavutada asümmeetrilise eelise," toodi välja Anthropicu teadaandes, „siiski on tungiv vajadus, et võimalikult paljud organisatsioonid tugevdaksid oma küberkaitset. Loodame, et meie laiemalt kättesaadavad mudelid ning uued tööriistad, ressursid ja uuringud, mida me nendega koos pakume, toetavad neid organisatsioone oma küberjulgeoleku seisu parandamisel."

Claude Mythos Preview (Project Glasswing) - mis see on?

Kuna Anthropicu uus mudel toimib reaalselt revolutsioonilise "tootena" küberturvalisuse maastikul, toome teieni lühiülevaate selle tugevustest ja nõrkustest.

PLUSSID

- Ülikiire ja autonoomne: Suudab analüüsida tohutus koguses lähtekoodi ja tuvastada vigu murdosa ajaga võrreldes inimekspertidega.
- Proaktiivne kaitse: Leiab haavatavused (nagu 10 000+ avastatud kriitilist viga) enne, kui küberkurjategijad jõuavad neid ära kasutada.
- Mitmekülgsus: Lisaks koodiveale suudab reaalajas analüüsida sotsiaalset manipuleerimist ja peatada reaalseid finantspettusi (nt 1,5 miljoni dollari suuruse pangaülekande peatamine).
- Annab kaitsjatele eelise: Loob turvaekspertidele nn "asümmeetrilise eelise", olles pahalastest sammu võrra ees.

MIINUSED

- Kättesaadavus: Avalikkusele ja väiksematele ettevõtetele pole mudel praegu kättesaadav suurte väärkasutuse riskide tõttu.
- Tekitab tohutu töölaadungi: Avastatud tuhanded vead loovad arendajatele hiiglasliku surve ja vajaduse lühendada tarkvarapaikade testimise tsükleid.
- Nõuab usaldust: Nõuab tarkvarapakkujatelt ja süsteemiadministraatoritelt äärmist operatiivsust ja uusi kaitsemeetmeid (nagu range MFA ja logimine).

- Uudised
- Tarkvara
- Turvalisus

Pilt


Claude
Mythos
Anthropic
Vulnerability analysis