Anthropicu juht nõuab kohustuslikke ohutusteste ja juurutuskeelde kõige võimsamatele tehisaru mudelitele

4 kuud tagasi - 26.06.2026 Autor: AM

(Sisuturundus)

Anthropicu tegevjuht Dario Amodei avaldas 10. juunil 2026 mahuka essee, milles nõuab, et riiklikud asutused saaksid seaduslikku volitust peatada tipptasemel tehisaru mudelite kasutuselevõtt, kui need ei läbi kohustuslikke kolmanda osapoole ohutusteste. Samal päeval avaldas Anthropic eraldi seadusandliku ettepaneku, mida ettevõte kavatseb ka rahaliselt toetada. See samm ei ole pelgalt ühte valdkonda puudutav avaldus, vaid otsene kursimuutus senisest lähenemisest, mille kohaselt on tehisaruvaldkonnas piisanud vabatahtlikest vastutusraamistikest.

Anthropic loobub vabatahtlikkuse põhimõttest

Letsdatascience andmetel ilmus essee "Policy on the AI Exponential" päev pärast seda, kui Anthropic oli turule toonud mudeli Claude Fable 5. Amodei kirjutab, et senised läbipaistvuspõhised meetmed, nagu ohutusprotseduuride ja testitulemuste avalikustamine, on osutunud ebapiisavaks, sest on tekkinud selge tõendusmaterjal, et tippmudeli tasemel tehisaru kujutab reaalset riski.

Essee põhisõnum on otsekohene. Ettevõte, mis on seni pooldanud vabatahtlikke avalikustamisstandardeid, esitab nüüd raami, milles valitsusel peaks olema seadusjõud peatada ohtlike mudelite kasutuselevõtt. See on sisuliselt senise positsiooni ümberpööramine.

Leedu loteriiplatvormi hindajad tunnevad sama loogikat ära

Loterijuguru toimetus, kes jälgib Leedu loteriiturgu, märkab Amodei nõutavas sõltumatu eelauditi põhimõttes tuttavat mõttemustrit. Amodei essees sõnastatud idee, et tehisaru mudelid peavad läbima enne kasutuselevõttu range tehnilise kontrolli, peegeldab loogikat, mis on Leedu mängijatele igapäevane. Amodei ütleb

oma essees:

“Frontier AI models, like airplanes, should be required to go through technical testing and auditing, and their release should be blocked or reversed as a threat to public safety if they do not meet high standards of safety.”

Toimetuse hinnangul kehtib põhimõte, et usaldusväärsus tuleb enne raha paigutamist tõendada, ka tarbijatele suunatud digiplatvormide puhul. Leedu kasutajad hindavad enne otsuse tegemist platvormide tehnilist turvalisust, läbipaistvust ja vastavust regulatiivsetele nõuetele, kaaludes hoolikalt, kas teenusepakkuja on usaldusväärne.

https://loterijuguru.lt/ jälgib ja analüüsib seda tarbijapoolset kontrolliprotsessi, kaardistades seaduslikult tegutsevaid loteriiplatvorme ning vaadeldes, milliste kriteeriumide alusel kasutajad usaldusväärseid teenusepakkujaid eristavad. See peegeldab laiemat suundumust Leedu digikeskkonnas, kus tarbijad ootavad kõigilt reguleeritud teenustelt selgeid standardeid, läbipaistvust ja vastutustundlikku tegutsemist.

Neljakategoorialine riskiraami ja lennundusparalleeli mehaanika

Ettepaneku sisuline raamistik tugineb neljale riskikategooriale, millele kohustuslikud testid peavad vastuse andma. Need on küberturvalisus, bioloogilised relvad, kontrolli kaotamine tehisaru süsteemide üle ning automatiseeritud teadus-arendustegevus, mis võib nimetatud riske kiirendada.

Testid viiks läbi kas valitsusasutus, mis on modelleeritud FAA eeskujul, või volitatud eraorganisatsioonid nn regulatiivse turu mudeli alusel. Sõltumata täideviijast peab valitsuse juurutuskeelu volitus jääma rangelt nelja määratletud riskikategooria piiridesse, kusjuures essee sisaldab selgesõnalisi kaitsemeetmeid poliitilise eelistamise vastu.

Illustratsioonina konkreetsest ohust toob Amodei välja Anthropicu enda mudeli Claude Mythos Preview, mille kohta kirjutab, et see on "seganud globaalse küberturvalisuse maastiku." Seegi fakt toetab väidet, et vabatahtlik aruandlus ei asenda sõltumatut tehnilist auditit.

Amodei ettepanekud on kriitikute hinnangul sektori ajaloo agressiivsemad

Axios kirjeldab ettepanekuid oluliselt rangemate regulatiivsete meetmetena võrreldes Valge Maja 2026. aasta juuni alguses allkirjastatud täitevkorraldusega, mis nägi ette vabatahtliku 30-päevase ülevaatuse. Politico omakorda nimetab Amodei raamistikku kõige agressiivsemaks regulatiivseks ettepanekuks, mida ükski suurettevõtte tehisarujuht avalikult toetanud on.

Kriitikuid, kes hoiatavad, et ranged reeglid võivad tugevdada praeguste turuliidrite positsiooni, adresseerib Amodei otse. Tema sõnul peavad kontrollmehhanismid kehtima nii valitsuste kui suurte tehisintelligentsi ettevõtete suhtes, sealhulgas Anthropicu enda suhtes.

Tööturu kaitse ettepanekud jätavad avatud küsimused

Paralleelselt regulatiivse raamistikuga sisaldab essee ka majanduslikke meetmeid tehisaruga seotud tööhõive mõjude leevendamiseks. Pakutavate instrumentide hulgas on palgakindlustus madalamapalgalisemale tööle siirduvate töötajate jaoks, tööjõukoolituse toetused, tööandjatele suunatud töötajate hoidmise maksusoodustused ning pikaajalise sissetulekutoena mõeldud "universaalsed kapitalikontod," mida rahastataks kapitalikasumimaksude või tehisarufirmadele kehtestatud tasude kaudu.

Ettepanekud on avalikud, kuid nende rakendamine on endiselt lahti. Küsimused, mille vastused kujundavad kogu raamistiku tegeliku toime, on sõnastamata. Kes täpselt otsustab, milline mudel ületab "tippmudeli" künnise, mis kohustuslikud testid käivitab? Millistel asutustel on pädevus auditeid läbi viia ja kuidas tagatakse nende sõltumatus? Need on küsimused, mis senini lahendust ootavad.

- Uudised
- Sisuturundus

Pilt

