Traditsioonilised IT-riskimudelid ei sobi tehisarule — viis spetsiaalset raamistikku võivad selle lünga täita

4 kuud tagasi - 24.06.2026 Autor: AM

(Sisuturundus)

Organisatsioonid, mis on kiirendanud tehisaru juurutamist oma põhiprotsessides, seisavad silmitsi valusa tõdemusega: kümnendite jooksul kujundatud riskijuhtimise raamistikud ei ole loodud tehisaru käitumismustrite, rikkeviisidega ega eetilist kaalutlust nõudvate otsustega toimetulekuks. Csoonline kirjutab, et sellele lüngale vastuseks on kujunenud viis eesmärgipärast raamistikku, mis katavad valdkondi alates juhtimisstruktuuridest kuni tehnilistest turvameetmetest.

Leedu kihlveoturg ja tehisaru haldusvajadus

LT Beti meeskond jälgib Leedu spordikihlvedude turgu sõltumatu vaatlejana ning märgib, et paljud selles sektoris tegutsevad operaatorid kasutavad juba praegu algoritmipõhiseid süsteeme koefitsientide kujundamiseks, pettuste tuvastamiseks ja kasutajakogemuse isikupärastamiseks. Tegemist on samalaadsete suure mõjuga automatiseeritud otsustusprotsessidega, mille jaoks on loodud ka artiklis käsitletud juhtimis- ja kontrolliraamistikud. LTBET Lietuva pakub meeskonnale vaatenurka turule, kus tehisintellektil põhinevad lahendused mängivad üha olulisemat rolli ning kus sellised süsteemid vajavad usaldusväärseks toimimiseks läbipaistvust, järjepidevat järelevalvet ja vastutustundlikku juhtimist.

“Tehisarulahendusi kasutavate kihlveoplatvormide otsustusprotsessid — kes saab pakkumise, milline koefitsient kuvatakse, milline tehing läheb kontrolli alla — on täpselt see kontekst, kus tehisaru haldusvead muutuvad reaalseks äri- ja regulatiivseks riskiks.”

Põhjus on struktuurne. Tavapärased IT-riskimudelid eeldavad süsteeme, mille käitumine on ette määratud ja kontrollitav; tehisaru mudelid aga muutuvad aja jooksul, võivad kuvada ootamatut käitumist ning langetavad otsuseid, mille mõju ulatub kaugemale üksikust tehingust.

ISO/IEC 42001 kui kõige põhjalikum alus

ISO/IEC 42001:2023 on ISO ja IEC koostöös detsembris 2023 avaldatud esimene rahvusvaheliselt tunnustatud ametlik tehisaruhalduse standard. See kohustab organisatsioone dokumenteerima, kuidas nad tehisarusüsteeme kavandavad, seiravad, valideerivad ja kontrollivad. Lisaks nõuab standard tehisaru mõjuhinnangute läbiviimist, mis katavad võimalikke juriidilisi, eetilisi ja ühiskondlikke tagajärgi.

Nicole Carignan, Darktrace'i julgeoleku- ja tehisarustrategia asepresident ning valdkonna CISO, iseloomustab seda standardit ühemõtteliselt:

“See pakub kõige tugevamat alust tehisaru riskijuhtimisprogrammi ülesehitamiseks, mitte aga üksikute tehisaururiskide eraldiseisvaks käsitlemiseks.”

Standard hõlmab juhtimisstruktuure, kolmandate osapoolte tarnijatest tuleneva järelevalve nõudeid, andmehaldust, läbipaistvuskohustusi ja elutsüklihaldust. See on vabatahtlik, kuid sertifitseeritav ning kehtib eri sektorite ja erineva suurusega organisatsioonide kohta. Siiski on sellel kaks praktilist piirangut, mida Carignan rõhutab: standard on ressursimahukas rakendada ning täistekst ei ole avalikult kättesaadav, mis muudab alustamise keeruliseks organisatsioonidele, kes on tehisaru halduskü simustes alles esimesi samme astumas.

NIST AI RMF ja küpsusepõhine lähenemine riskile

NIST AI RMF, avaldatud jaanuaris 2023, on üles ehitatud nelja omavahel seotud funktsiooni ümber. Govern loob sisemise kultuuri, poliitikad ja vastutuse; Map aitab mõista konteksti ja kaardistada tehisarusüsteemi võimalikud riskid; Measure hindab ja jälgib riske nii kvalitatiivsete kui kvantitatiivsete meetoditega; Manage tegeleb riskide prioritiseerimise ja reageerimisega, sealhulgas maandamise, ülekandmise või aktsepteerimisega. Eraldi Playbook annab iga funktsiooni jaoks praktilised rakendusjuhised.

Ram Varadarajan, Acalvio tegevjuht, soovitab just seda raamistikku lähtepunktina, kuna see on üles ehitatud küpsuse, mitte läbikukkumist/läbimist hindavate auditite loogikale.

“Veelgi olulisem on see, et see sunnib läbi viima kolm vestlust, mis peavad aset leidma kõigepealt: kes omab tehisaruriske, milline tehisaru tegelikult töötab, ja kes saab kannatada, kui midagi läheb valesti.”

Forrester on samas kritiseerinud raamistikku liiga kirjeldava ja vähe ettekirjutava olemuse pärast, mis tähendab, et see näitab, mida mõõta, kuid ei ütle alati täpselt, kuidas seda teha.

ENISA FAICP ja ISO/IEC 23894 — kaks erinevat kihti

ENISA avaldas juunis 2023 tehisaru küberturvalisuse tavade raamistiku (FAICP), mis on korraldatud kolme progressiivse kihi ümber. Esimene kiht hõlmab tavapärastest tarkvarasüsteemidest päritud põhilisi IKT-küberturvalisuse tavasid. Teine kiht käsitleb tehisaruspetsiifilisi riske, sealhulgas vastandusrünnakuid, mudelite manipuleerimist, andmevoo terviklust ja tarneahela turvalisust. Kolmas kiht pakub sektoriüleseid juhiseid reguleeritud valdkondadele nagu energia, tervishoid ja telekommunikatsioon.

FAICP-i tihe seotus EL tehisaruseaduse ja NIS2 direktiiviga, mis on EL peamine küberturvalisuse õigusakt, tähendab, et EL regulaatorid käsitlevad seda raamistikku lähtealusena kõigile EL-is äritegevusega tegelevatele organisatsioonidele.

ISO/IEC 23894:2023, avaldatud veebruaris 2023, erineb sellest selgelt. Tegemist on ainult juhendava standardiga, mis ei ole sertifitseeritav, ja mis laiendab ISO 31000 metoodikat tehisaruspetsiifilistele riskidele, sealhulgas algoritmilisele erapoolikusele, mudeli triivimisele, ettearvamatu käitumise võimalusele ja otsustusprotsesside läbipaistmatusele. ISO iseloomustab seda standardit ISO 31000 ja ISO/IEC 42001 täiendusena.

Google SAIF ja küsimus, kuidas viie raamistiku vahel valida

Google Secure AI Framework (SAIF), käivitatud 2023. aastal, on teistest raamistikest inseneriteaduslikuma suunitlusega. See keskendub konkreetsetele ohuliikidele, nagu andmemürgitamine, prompti süstimine ja mudelite manipuleerimine, ning hõlmab andmekäitlust, tehisaru all olevat infrastruktuuri, mudeleid endid, kasutajale suunatud rakendusi ja kontrollimisprotsesse.

Carignan toob viite raamistiku koostoimel esile positiivse aspekti:

“Nende raamistike vahel on kattuvusi, kuid see kattuvus on kasulik. See tugevdab põhitavasid, mida organisatsioonid peavad õigesti tegema: juhtimine, andmete terviklus, turvalisus, vastutus, järelevalve, testimine ja pidev täiustamine.”

Varadarajan prognoosib, et kahe kuni kolme aasta jooksul kujuneb olukord, kus EL tehisaruseadus seab õigusliku aluspõhja ning NIST AI RMF pakub selle täitmiseks tegevusliku töövihiku, järgides sama rada, mida Euroopa andmekaitseõigus on teinud üleilmselt kohaldatavaks standardiks ettevõtetele sõltumata nende asukohast.

“Kahe kuni kolme aasta jooksul oodake, et domineerima hakkab kaks raamistikku: EL tehisaruseadus, mis määrab õigusliku miinimumi, ja NIST AI RMF, mis pakub tegevusliku plaani selle täitmiseks.”

David Brumley, Bugcrowd tehisaru- ja teadusjuht, hoiatab aga, et raamistikule ülemäära keskendumine kannab oma riski:

“Tehisaru kasutuselevõtt ei oota täiuslikku haldust, ja need, kes keskenduvad riskijuhtimisraamistikule, võivad tahtmatult luua oma organisatsioonis vari-tehisaru probleemi.”

See hoiatus kõlab eriti asjakohaselt kiiresti arenevatel turgudel, kus tehisarulahenduste kasutuselevõtt käib kiiremini kui juhtimisstruktuuride kujundamine. Varadarajani konsolideerumisennustus ja Brumley varjusüsteemide oht koos meenutavad, et tehisaru haldus ei ole ühekordne nimekiri, mille lõpus saab linnukese lisada, vaid pidevalt arenev operatiivne ülesanne.

- Uudised
- Sisuturundus

Pilt

