

Tehisaru kasutamisest tingitud uute küberohtude mõju minimeerimine

Aruanne

Versioon 1.0

30.06.2026

D-16-1272

Avalik

©2026 Cybernetica AS

Projektijuht: Liina Kamm (Cybernetica AS)
Anna Kuus (Justiits- ja Digiministeerium)

Autorid: Dan Bogdanov
Ainika Harro
Liina Kamm
Jens Olaf Kukk
Anna Mõtlík
Tanel Mällo
Tanel Pern
Pille Pullonen-Raudvere
Kati Sein

Cybernetica AS, Mäealuse 2/1, 12618 Tallinn, Estonia.

E-post: info@cyber.ee, Veebileht: <https://www.cyber.ee>, Telefon: +372 639 7991.

Kaasrahasanud Euroopa Liit. Avaldatud seisukohad ja arvamused on ainult autori(te) omad ega pruugi kajastada Euroopa Liidu seisukohti või arvamusi. Euroopa Liit nende eest ei vastuta.

Kuupäev	Versioon	Kirjeldus
01.06.2026	0.5	Aruande mustand ja ohukirjeldused
30.06.2026	1.0	Aruande lõppversioon

Sisukord

1 Sissejuhatus	7
1.1 Eesmärk	8
1.2 Terminid ja lühendid	8
1.3 Aruande struktuur	8
2 Metoodika	9
2.1 Uurimisküsimused	9
2.2 Ohtude kõrgetasemeline jaotus	9
2.3 Teaduste akadeemia küberturvalisuse komisjoni ohtude kirjeldamise metoodika	10
2.4 Uuringu metoodika	10
3 Ohukirjeldused	12
3.1 Kuidas ohukirjeldust lugeda?	12
3.2 OSKUSKADU: AI kasutamine nõrgestab inimeste kognitiivseid, sotsiaalseid ja kultuurilisi oskusi.	13
3.3 MASSIMUDIJA: Tehisintellekt teeb võimalikuks massilise personaliseeritud psühholoogilise mõjutamise	15
3.4 MASSRÜNNE: Infosüsteemide nõrkuste tuvastamise ning ründamise lihtsus ja odavus ohustab kõigi IT-süsteemidest sõltuvate teenuste, andmete, inimeste turvalisust	17
3.5 AGENDIKAAPERDUS: AI-agentide ülevõtmine ja kuritarvitamine põhjustab kahjusid agendi omanikule	20
3.6 HULLKRATT: AI-agentidega turvaliselt töötamine ei õnnestu	22
3.7 SEINTELKÕRVAD: Uute seadmete levik toob isikute varjatud jälgimise ja analüüsi AI abil igapäevaellu	24
3.8 ISIKUVÕLTSING: Identiteedi võltsimine AI abil ning sellel põhinevad pettused ja kampaniad muutuvad igapäevaseks	27
3.9 MASINKELM: Digikeskkondade pettused muutuvad AI abil veenvamaks ning levinumaks	30
3.10 MUDELIMÜRK: Suurtesse keelemudelite treeningandmetesse söödetakse manipulaatiivset sisu	32
3.11 SUPERKOGU: AI-platvormide kasutamisel tekivad uued tundlike isikuandmete kogud, mis võivad lekkida ja mida saab väärkasutada	34
3.12 SUVERÄÄNSUSKADU: AI-le tuginevate e-teenuste juurutamine riigivälistel platvormidel kiirendab riikliku digisuveräänsuse kadu	36

3.13 AIEIVASTUTA: Vastutus AI abil loodud süsteemide ja teenuste kvaliteedi languse eest hajub	38
3.14 EBAVÕRDSUS: Ebavõrdsus AI-lahenduste kättesaadavuses suurendab ühiskonna majanduslikku kihistumist	40
3.15 ÜLEKOORMUS: Asutused koormatakse üle AI abil genereeritud pöördumiste ja taotlustega	42
3.16 AIPOLEINIMENE: Inimlike väärtuste eiramine AI juurutamisel viib inimeste koostöö ja suhtluse nõrgenemiseni ja vähendab kriisikindlust	44
3.17 AIPSÜHHOOS: AI-rakenduste tarbimine pärsib kasutajate reaalsustaju ja tekitab sõltuvuskäitumist	47
4 Soovitused üksikisikule	50
4.1 Sissejuhatas	50
4.2 Kuidas käituda internetis, sotsiaalmeedias ja äppides?	50
4.3 Üldine hügieen tehisintellekti kasutamisel	50
4.4 Soovitusi tehisintellektitoodete valikuks	50
4.5 Tehisintellekti ja agentide kasutamine töös ja hobides	51
4.6 Kuidas käituda, kui tehisintellektiga läks midagi valesti?	51
5 Soovitused organisatsioonile	52
5.1 Sissejuhatas	52
5.2 Tehisintellekti strateegia loomine	52
5.3 Tehisintellekti poliitika ja juhiste loomine	52
5.4 Tehisintellekti mõju infoturvapoliitikale	52
5.5 Töötajate ja tarbijate ohutus ning vastutus	53
5.6 Infoturbe ja andmekvaliteedi tehnilised meetmed	53
5.7 Agentsüsteemide turvalise kasutamise meetmed	53
6 Soovitused riigile	54
6.1 Sissejuhatas	54
6.2 Poliitilised valikud	54
6.3 Õigusruum	54
6.4 Tehisintellekti rakendamisele teenustes	55
6.5 Haridus, väljaõpe ja tööturg	55
6.6 Järelevalve	56
6.7 Tehnilised meetmed	56

7 Kuidas uuringu tulemusi rakendada?	57
7.1 Ohtude olulise järjekorra määramine hindamise teel	57
7.2 Soovitatud leevendusmeetmete rakendamine praktikas	57
7.3 Puuduvate võimekuste loomine.....	57
7.4 Edasised uurimissuunad ja ohukirjelduste uuendamine.....	58

1 Sissejuhatus

Inimesed mõistavad tehisintellekti erinevalt. Populaarteaduslikus ja ulmekirjanduses tajutakse tehisintellekti mõtleva arvuti või robotina. Tehnilisemas kirjanduses räägitakse süsteemidest, mis on rohkem või vähem autonoomsed. Standard ISO/IEC 22989:2022 määratleb tehisintellektisüsteemi kui tehnilist süsteemi, mis genereerib tulemitena mingiks inimese määratud eesmärgistikuks näiteks sisu, prognoose, soovitusi või otsuseid [1]. Euroopa Liidu tehisintellekti käsitlev määrus (AI-määrus) [2] on jätnud inimese eesmärgid kõrvale. AI-määrus defineerib tehisintellektisüsteemi masinpõhise süsteemina, mis on projekteeritud töötama erineval autonoomsuse tasemel ning võib pärast juurutamist olla kohanemisvõimeline ja mis otseste või kaudsete eesmärkide saavutamiseks järeldab saadud sisendandmete alusel, kuidas genereerida väljundandmeid, näiteks prognoose, sisu, soovitusi või otsuseid, mis võivad mõjutada füüsilist või virtuaalset keskkonda [2].

Selle uuringu koostamise ajal on tehisintellekt inimkonna ajaloos ennenägematult efektiivne teadmuse edasiandmise viis. Tugevasti lihtsustades on suur keelemudel inimkonna digitaalsel kujul salvestatud teoste kokkupakitud vorm, mis suudab ennustada sisendteksti kohta väljundteksti, mis võiks inimestele meeldida. Hoolimata edusammudest ei ole me jõudnud üldise tehisintellektini (*artificial general intelligence*), mis seaks enda eesmärgi ise. Tänaused tehisintellektisüsteemid on inimeste treenitud ja täidavad nende seatud eesmärgi (niivõrd kui nende treenimisandmestikud ja oskused seda võimaldavad). Suurtest keelemudelitest üldise tehisintellekti loomine võibki olla võimatu ning selleks võib olla vaja hoopis uut läbimurret, näiteks neuroteadustes. Uuringu koostamise ajal on ulmekirjandusest tuntud omaenda plaanide ja kavatsustega tehisintellektisüsteemide risk veel madal. Tehisintellekti käsitlust ennustava masinana toetab ka kirjandus [3].

Suurte keelemodelite leviku ja efektiivse kasutamisega kaovad mitmed seni inimestel uute oskuste õppimist takistanud barjäärid. Raamatute lugemine või kuulamine annab inimestele oskused, kuid selleks kulub aega. Tänaused suurte keelemodelite põhised tehisintellektisüsteemid käituvad oskuste võimendajatena. Kas tehisintellektist saab rohkem head või kurja, sõltub seda kasutavast inimesest ja inimese seatud eesmärgist. Kurjade kavatsustega inimene võib kiiremini omandada oskuseid, mida tal varem ei olnud, ning sellega põhjustada kannatusi laiadele rahvassidele¹.

Erinevad organisatsioonid on seda teadvustanud ning on ka üritanud kehtestada embargot AI-tehnoloogia arendusele². Sellised embargod ei tööta, sest riigid ja ettevõtted näevad võimalust saada senisest rohkem võimu. Džinni me purki tagasi ei pane, peame riske leevendama või õppima nendega elama.

Tahame Eestis vältida olukorda, kus tehisintellekti rakendamine toob kaasa ohte inimestele, majandusele, demokraatialle ja riigi kestlikkusele. Samal ajal soovime kaasa minna tehnoloogilise arenguga, et tagada majanduse ja isikute konkurentsivõime. Riskihaldus aitab neid eesmärgi tasakaalustada. Hästi korraldatud riskihaldus aitab põhjendada otsuseid ja investeeringuid, sellega säilib ja suureneb ühiskonna usaldus algoritmiliste süsteemide kasutuse suhtes. Ohtude tuvastamine ja nende mõju minimeerimine on riskihalduse lahutamatu osa.

¹OpenAI warns that its new ChatGPT Agent has the ability to aid dangerous bioweapon development. Fortune. 18. juuli 2025. <https://fortune.com/2025/07/18/openai-chatgpt-agent-could-aid-dangerous-bioweapon-development/>. Viimati külastatud 29.06.2026

²Pause Giant AI Experiments: An Open Letter. Future of Life Institute. 22. märts 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Viimati külastatud 29.06.2026

1.1 Eesmärk

Selle uuringu peamine eesmärk on toetada Eesti riiki ja asutusi ohtudest arusaamisel, riskikontrollis ning ohtude mõju minimeerimisel. Riskihalduses on vaja olla teadlik ohtudest ja mõista ohtudega seotud lihtsamaid ründevektoreid (järjestikuseid tegevusi, mis viivad ohusündmuse realiseerumiseni). Kuid ohuteadlikkusest üksi ei piisa, vaja on ka valmisolekut. Iga ohu kohta on vaja meetmete loetelu, mida riik või organisatsioon saab rakendada, et ohu võimalikkust vähendada ja mõju minimeerida.

1.2 Terminid ja lühendid

AI-süsteem, tehisintellektisüsteem

masinpõhine süsteem, mis töötleb päringu sisendandmeid, et luua vastuseid (nt prognoose, sisu, soovitusi või otsuseid), mis võivad mõjutada füüsilist või virtuaalset keskkonda. AI-süsteemide põhiomadus on nende võime teha järeldusi ja tuletada seoseid sisendandmetest, kasutades selleks masinõppemudeleid

algoritm

astmeline eeskiri mingi tegevuse sooritamiseks

kaitsepiirded (*guardrails*)

tehnoloogia või raamistik, mis toimib turvakontrollpunktina, hinnates kasutaja sisendit ja mudeli genereeritud vastuseid lähtudes eelnevalt määratud ohutusreeglitest. Kaitsepiirete eesmärk on reaajas ennetada kahjuliku, sobimatu või teemavälise sisu genereerimist, kui mudeli ohutushäällestusest selleks ei piisa

kallutatus (*bias*)

objektide, isikute või rühmade käsitlemise süstemaatiline erinevus teistega võrreldes; käsitlemine on ükskõik milline tegevus, sealhulgas tajumine, vaatlus, esitus, prognoosimine või otsustus

masinõppemudel

spetsiifiline algoritm või algoritmide kogum, mille abil saab sisendväärtuste põhjal ennustada väljundväärtusi

seletavus (*explainability*)

AI-süsteemi omadus väljendada oma tulemusi mõjutavaid olulisi tegureid inimestele arusaadaval viisil; on mõeldud vastama küsimusele „Miks?“, püüdmata väita, et valitud tegevusviis on tingimata optimaalne

1.3 Aruande struktuur

Jaotises 2 kirjeldame uurimisküsimusi ja uuringu metoodikat. Jaotis 3 kirjeldab detailselt uurin-
gus tuvastatud ohte. Jaotistes 4, 5 ja 6 esitame leevendusmeetmetest tulenevad soovitused üksikisikutele, organisatsioonidele ja riigile. Jaotises 7 kirjeldame, kuidas uuringu tulemusi raken-
dada ja edasi arendada, sealhulgas kuidas ohukirjeldusi värskena hoida.

2 Metoodika

2.1 Uurimisküsimused

Riskihalduses peab olema selge, millised leevendusmeetmed on kellele jõukohased. Mõned leevendusmeetmed on teistest kallimad või vajavad seadusandlikku või isegi diplomaatilist jõudu. Seega mõnda leevendusmeedet saab juurutada vaid riik, teisi peab rakendama iga ettevõtte või inimene ise. Selline eristus peab olema riskikirjelduse osa.

Rahvusvahelised kogemuslood on tähtsad, sest aitavad lugejal toimuvaga suhestuda ning, näiteks riigi puhul, häid tavaid üle võtta. Märjume ära, et uuring käsitleb teemasid, kus riigid on kõik alles omapäi, häid tavaid veel pole ning kvaliteetsete viidetega allikad on harvad. Eestis suudame tihti olla väledamad võrreldes suuremate riikidega ja seetõttu pole asjakohaseid viiteid alati olemas.

Uuringus käsitleme nii juba realiseerunud ohte kui ka potentsiaalseid ohte, mis võivad realiseeruda, kui tehisintellektitehnoloogia jõuab ulatuslikumalt kasutusse.

Uuringu uurimisküsimused on järgmised.

1. Millised tehisintellekti olemasolust põhjustatud ohusündmused on Eesti ja Euroopa tasemel realistlikud?
2. Millised on nende ohusündmusteni viivad sündmuste jadad, eeldused?
3. Millised leevendusmeetmed on täna juba olemas ja milliseid oleks veel vaja rakendada?
4. Milline on Eesti ja Euroopa organisatsioonide võime neid leevendusmeetmeid rakendada?

Ohukirjelduste ajakohasus. Dokumente peab hoidma ajakohasena. Kui seda regulaarselt ei tee, siis nende sisu vananeb kiiresti. Riskihalduses tuleb valida elavam teabe esitamise viis, mis võimaldab ründevektorite kirjeldust, leevendusmeetmete loetelu ja muid detaile hiljem kergesti muuta, kui asjaolud muutuvad. Selleks on mõistlik kasutada masinloetavaid vorminguid, kus on vähem rõhku pandud vormistusele ja rohkem levitamise lihtsusele, viidatavusele ja muudetavusele.

2.2 Ohtude kõrgetasemeline jaotus

Ohte analüüsid on oluline mõista, mida üldse rünnatakse, millele oht rakendub. Aruandes järgime kõrgetasemeliste ohtudena Euroopa Komisjoni DG CNECT AI ohtude analüüsi hankes (TED viide 272332-2025) kasutatud jaotust. Seal jaotati ohud järgmistesse üldistesse kategooriatesse.

Tehisintellekti toel teostatud CBRN-ründed

CBRN-ründed on keemilise, bioloogilise, radioloogilise ja tuumaainega ning lõhkeainetega seotud ründed.

Tehisintellekti toel teostatud küberründed

Käsitleme senisest laialdasemaid või laialdasemalt teostatavamaid ründeid kriitiliste taristu, esemevõrgu (IoT), riigikaitse, julgeoleku, aga ka üksikisikute vastu. Hindame, kas tehisintellekt on kasulik uute turvanõrkuste leidmisel ja ärakasutamisel.

Tehisintellektisüsteemi üle inimliku kontrolli kaotamisest põhjustatud ründed

Vaatleme ründeid, mis tekivad olukorras kus tehisintellektisüsteemile antakse integratsioonide tulemusena suurem autonoomia ning hoovad reaalse maailma (finantsturgude, auto-

matiseeritud tootmissüsteemide) mõjutamiseks. Kui sellises olukorras tehisintellektisüsteemi algatud tegevuste kavatsus ja "väärtused" ei lange kokku algse inimliku kavatsuste ning väärtusruumiga, võivad tulemused olla inimesele ohtlikud. Seda eriti siis kui seletavuse puudumise või muude efektide ("AI-süsteem petab") tõttu ei ole aru saada, miks tehisintellektisüsteem konkreetse otsuse tegi.

Tehisintellekti toel teostatud ründed rahvastiku manipuleerimiseks

Vaatleme ühesammulisi (identiteedivargused, pettused, süvavõltsingud, mainekahju, manipulatsioonid) kui ka mitmesammulisi (usaldusväärseuse õõnestamine koordineeritud kampaniatega, sünteetiliste isikute ja isiksuste kasutamine, oskuste taandarenguni viimine, mudelite mürgitamine) ründeid.

Tehisintellekti rakendamisel tekkivad sotsio-tehnoloogilised ründed

Tehisintellekti kasutamine võib viia põhiõiguste riive või isegi rikkumiseni, sealhulgas rohkeimate kallutatud otsuste ja diskrimineerimiseni, haavatavuse ärakasutamiseni ja isikute autonoomia ja privaatsuse rikkumiseni.

Uuringus käsitleme küberrünnete, manipuleerimisrünnete ja sotsio-tehnoloogiliste rünnete kategooriasse kuuluvaid ohte. Kõigi rünnete puhul vaatleme ka tehisintellekti tarneaahelat (teeningandmete kvaliteet, mudelite päritolu, pilvtöötuse kasutamine). CBRN-ründed ei ole uuringu käsitusallas.

2.3 Teaduste akadeemia küberturvalisuse komisjoni ohtude kirjeldamise metoodika

Uuringus kasutame Teaduste Akadeemia (TA) küberturvalisuse komisjoni metoodikat. Eesti Teaduste Akadeemia küberturvalisuse komisjon on alates 2023. aastast arendanud e-riigi riskide analüüsi metoodikat ning koostanud kataloogi e-riigi teenustega seotud ohtude kohta. TA komisjoni metoodika aitab meil täita uuringu põhimõtteid, luues elavad dokumendid, mida on lihtne masintöödelda, viidata ning täiendada. Komisjoni ohukirjelduse malli osaks on nii olemasolevad kui võimalikud täiendavad leevendusmeetmed, mistõttu on rakendajal hea võrrelda oma tänast käitumist sellega, mida võiks veel teha.

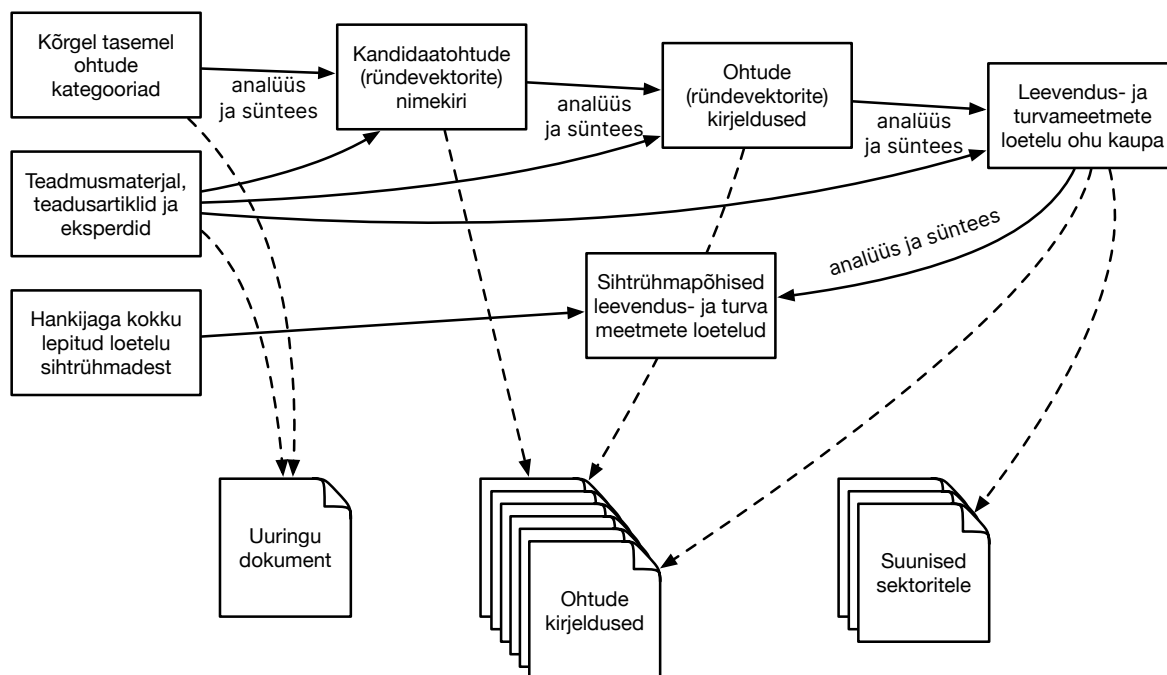
TA komisjoni metoodika ja dokumendimalli kasutamine annab uuringu tulemustele laiemat kasutust ning pikema eluea, sest ohukirjeldused avaldatakse Markdown vormingus e-riigi koodivaramus¹. Nii on ohtude analüüsi võimalik kergemini täiendada, kui ilmnevad uued ohud või leitakse efektiivsemaid leevendusmeetmeid. Ühe koodihoidlasse salvestatud Markdown-vormingus ohukirjelduse täiendamine või meetmete lisamine on palju lihtsam kui terve suure PDF-dokumendi uuesti vormistamine ja küljendamine.

2.4 Uuringu metoodika

Uuringu metoodikat illustreerib joonis 1.

Valimid ja andmeallikad. Uuringus analüüsime materjale ja dokumente. See analüüs tugineb tavapärasest rohkem ka mitteteaduslikele allikatele (sest arusaamad AI ohtudest arenevad kiiresti ning ülevaateartikleid veel napib). Lähtematerjalide valikul käsitleme nii klassikalist teadus-suhtlust (ajakirjad, konverentsid, lõputööd) kui ka avaldatud praktilisi juhtumiuuringuid ja teadusprojektide tulemeid. Käsitleme ka tehisintellekti riskide raamistikke, tehnoloogiate ja trendi-

¹Teaduste Akadeemia küberturvalisuse komisjoni koostatud e-riigi ohtude analüüs. <https://koodivaramu.eesti.ee/e-riigi-turve/ta-komisjon>. Viimati külastatud 29.06.2026.



Joonis 1. Tehisaru kasutamisest tingitud uute küberohtude mõju minimeerimise uuringu meetodi-ka

de analüüse ja juhtumiuuringuid, sh erasektoris koostatud raamistikke.

Andmekogumismeetodid. Artikleid, uuringuid ja projektitarneid koostades sobitame uut tead- must olemasolevaga.

Andmeanalüüsi ja järelduste tegemise põhimõtted. Ohusündmuste kandidaatide loetelu kok- kupanekul tugineme nii materjalidele kui ka ekspertidele. Veendume, et igast valitud kõrge ta- seme ohtude kategooriast oleks esindatud vähemalt viis tähenduslikku riski, kuid pöörame oma tähelepanu siiski ohtudele, mida on rohkem. Kui oleme kirjandusele ja ekspertidele tuginedes koostanud ohtude nimekirja, kirjeldame ohte ja ründevektoreid täpsemalt.

Kaitse- ja leevendusmeetmed. Ohtude kirjeldamise järel hindame, millised kaitse- ja leeven- dusmeetmed on Eestis juba kasutusel ja milliseid võiks veel kasutada.

3 Ohukirjeldused

3.1 Kuidas ohukirjeldust lugeda?

Esmalt, kujutlege ennast tulevikku, kus suurte keelemudelite tehnoloogia on

1. igapäevases kasutuses,
2. efektiivne teksti, pildi, heli, video ja struktureeritud andmete töötlemisel ja sünteesil,
3. integreeritud meie e-teenustesse,
4. integreeritud haridussüsteemi ja koolitustesse,
5. laialt kättesaadav nii pilvelt kui isikute endi seadmetel ning
6. muutnud AI-agendid kättesaadavaks kõigile soovijatele.

Võime eeldada, et selline tulevik ei ole maailmas ühtlaselt jaotunud. Eesti kui edumeelne e-riik on uute tehnoloogiate rakendamisel tihti maailmas esirinnas. Seetõttu võivad mitmed kirjeldatud ohud muutuda meil reaalsuseks varem kui mujal.

Iga ohu pealkirjas on suurtähtedega sõna (näiteks **OSKUSKADU**, **MASSIMUDIJA**). See on ohu lühinimi ning mõeldud ohu kiireks viitamiseks. Oht algab lühikirjeldusega, millele järgnevad vajaduse korral tehnilised detailid. Selle osa eesmärk on lugeja peas tekitada ettekujutus, miks ja kuidas oht realiseerub.

Alajaotises „**Ohu poolt mõjutatud objekt**“ täpsustame, keda (isik või organisatsioon) või mida (teenus, taristu) oht kahjustab. Toome ka ära, kas tegemist on Eesti-spetsiifilise ohuga või peavad sellega toime tulema ka teised maailma riigid (alajaotis „**Ohu ulatus**“).

Seejärel nimetame ohuagendi („**Ohuagendi tüüp**“), kelle tegevuse või tegevusetuse tõttu saab ohusündmus realiseeruda. Mõistame, et vahel on ebamugav mõelda, et keegi põhjustab ohu. Siiski on kasulik mõista, kes saab teha otsuse(id), mille tulemusena ohusündmus toimub või toimumata jääb. Ka otsuste tegemata jätmine (näiteks leevendusmeetmete mitterakendamine) võib kaasa aidata ohusündmuse realiseerumisele.

Inimeste ohutaju on erinev. Seetõttu loetleme alajaotises „**Eeltingimused ohusündmuse tekkeks**“ eeldused, mis peavad kehtima, et ohusündmus saaks põhjustada reaalseid kahjusid.

Alajaotises „**Olemasolevad kaitsemeetmed**“ loetleme kaitsemeetmeid, mille olemasolu kohta oli meil uuringu koostamise ajal tõendeid. Me ei saa olla kindlad, et need kaitsemeetmed on alati efektiivsed ja rakendatud kõigis asutustes, mis neid vajavad. Seega, kui mõni kaitsemeede, mida uuringu koostajad on pidanud olemasolevaks, lugeja jaoks seda pole, siis tasub sellest mõelda kui lisakaitsemeetmest, mida peaks esmajärjekorras rakendama asuma.

Mõned (näiteks organisatoorsed) kaitsemeetmed vajavad toimimiseks asjaolude kokkulangevust. Seetõttu oleme jaotises „**Eeldused kaitsemeetmete õnnestumiseks**“ kirjeldanud, mida on vaja, et olemasolevad kaitsemeetmed saaksid ohusündmuse ära hoida.

Uuringu üheks põhiliseks osaks on iga ohukirjelduse alajaotis „**Võimalikud lisakaitsemeetmed**“, mis loetleb tegevusi, mida riik (seadusandja või rakendusüksus), tehisintellekti rakendav organisatsioon või inimene ise ohtude vältimiseks ära teha saavad. Eeldatav rakendaja on toodud leevendusmeetme järel sulgudes.

Mõne ohu puhul ei pea ennast tulevikku kujutlema: see on juba uuringu valmimise ajal realiseerumas. Pärast ohukirjelduse lugemist peab lugeja päriselt hakkama kaitsemeetmeid rakendama.

3.2 OSKUSKADU: AI kasutamine nõrgestab inimeste kognitiivseid, sotsiaalseid ja kultuurilisi oskusi

Ohu lühikirjeldus

Kui AI-d kasutatakse järjest enam otsuste tegemiseks ja analüüsideks, siis nõrgeneb selle tagajärjel inimeste kriitiline mõtlemine, probleemilahenduskus ja võime iseseisvalt teadmisi omandada. Selliste muutuste mõõtmine on keeruline ja negatiivsed tagajärjed võivad avalduda alles põlvkondade vahetuse või tööturumuutuste ajal. Tehisintellektitehnoloogia on uus ning uurinuid selle mõjude kohta on veel vähe, kuid mõned esmased katsetused näitavad, et tehisintellekti kasutamisega ei kaasne oskuste kinnistumist [4] ning AI toel saavutatud efektiivsus ei säili AI kasutamise lõpetamisel [5]. On leitud, et sagedasem AI-tööriistade kasutamine seostub (korrelatiivselt) nõrgema kriitilise mõtlemisvõimega, mida seotakse õppimiseks vajaliku kognitiivse pingutuse AI-le delegeerimisega (*cognitive offloading*). Seos avaldus kõige tugevamalt noortel (17–25-aastased), samas kui kõrgem haridustase leevendas kognitiivse pingutuse delegeerimise mõju [6].

Eesti väike keele- ja kultuuriruum võimendab AI mõjusid: probleemid, mis suurriikides jäävad üksikjuhtumiteks, võivad Eestis mõjutada märkimisväärt osa sektorist. Samuti vähendab AI rakendamine vajadust noorempetsialistide järele, mis pärsib spetsialistide järelkasvu arengut ning spetsiifilise valdkondliku oskusteabe pikaajalist säilimist.

Ohtu võimendab ootus, et tehisintellektiga on võimalik asendada inimesi ning kasvatada efektiivsust. Samal ajal tuleb mõista, et ka tehisintellektil on hind ja kui see pole enda kontrolli all, siis halvimal juhul minetatakse oskused iseseisvalt tööd teha ja töö tegemiseks vajalik abivahend muutub kallimaks, kui alguses loodeti.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): tehisintellekti igapäevases elus kasutavad isikud, eeskätt noored ja eriti need noored, kes kasutavad AI-d vastuste saamiseks ja meelelahutuseks (mitte näiteks oskuste arendamiseks); avalik informatsioon, erialane järelkasv.

Ohu ulatus (Eesti-keskne/ülemaailmne): oht on ülemaailmne, kuid Eestis võimendatud väikese keele- ja kultuuriruumi tõttu, piiratud tööturu tõttu ning kõrge usalduse tõttu digilahenduste vastu.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): AI teenuseandjad ja tehnoloogiaettevõtted; efekti võimendavad organisatsioonid ja riiklikud huvipooled, kes kasutavad AI-d kuluefektiivsuse tagamiseks või infosfääri suunamiseks.

Eeltingimused ohusündmuse tekkeks

1. AI-põhised tööriistad ja teenused muutuvad igapäevases õpi-, töö- ja suhtluskeskkonnas peamiseks abivahenditeks, mis pigem annavad lahenduse kui toetavad mõtlemist. Selle tulemusena väheneb inimeste vajadus iseseisvalt analüüsida, õppida ja probleeme lahendada.
2. Organisatsioonid ja avalik sektor eelistavad AI-põhist automatiseerimist ka valdkondades, kus inimkontakt, praktiline kogemus või ekspertiisi järkjärguline kujunemine on strateegiliselt olulised.

3. Ühiskonnas puudub piisav võime hinnata AI kasutamise pikaajalisi mõjusid inimeste oskustele, sest muutused toimuvad aeglaselt ja sageli varjatult.

Olemasolevad kaitsemeetmed

1. Euroopa Liidu regulatsioonid (AI-määrus) ning organisatsioonide sisemised juhised loovad esmase raamistiku AI vastutustundlikuks kasutamiseks ja riskide leevendamiseks.
2. Kõigis haridusastmetes arendatakse digipädevust, kriitilist mõtlemist ja meediakirjaoskust, et inimesed oskaksid AI väljundeid teadlikumalt hinnata.

Eeldused kaitsemeetmete õnnestumiseks

1. Haridusasutused, tööandjad ja avalik sektor laiemalt peavad olema valmis investeerima inimeste digipädevuse, kriitilise mõtlemise ja erialaste oskuste arendamisse ka olukorras, kus AI kasutamine muutub järjest odavamaks ja mugavamaks.
2. AI-süsteemide toimimine ja piirangud peavad olema piisavalt läbipaistvad, et kasutajad suudaksid hinnata nende väljundite usaldusväärsust ja võimalikke riske.
3. Ühiskonnas peab säilima arusaam, et inimeste oskuste, teadmiste ja inimteeninduse hoidmine on strateegiliselt oluline ka siis, kui automatiseerimine pakub kuluvõitu.

Võimalikud lisakaitsemeetmed

1. Avalik sektor ja kriitiliste teenuste andjad kehtestavad nõuded, et kriitilistes avalikes ja era-sektori teenustes säiliks võime tagada teenus kui inimeste kõrval ei ole tehisintellektiabilist, mis töö täielikult automatiseerib. Nii säilib võime jätkata teenuse andmist ka tehnoloogiliste häirete või kriisiolukordade korral. **(Organisatsioon)**
2. Riik hindab regulaarselt AI mõju Eesti tööturule ja spetsialistide järelkasvule, eriti valdkondades, kus automatiseerimine võib vähendada noorem-spetsialistide ja praktikakohtade arvu. **(Riik)**
3. Tööandjad ja avaliku sektori organisatsioonid hindavad AI kasutuselevõtu mõju organisatsiooni pikaajalisele oskusteabe baasile ning järelkasvule, et vältida olukorda, kus lühiajaline efektiivsuskasu toimub erialase võimekuse arvelt. **(Organisatsioon)**
4. Riik, teadus- ja haridusasutused loovad pikaajalise seire ja mõjuanalüüsi AI kasutamise mõjude hindamiseks inimeste õpioskustele, sotsiaalsetele võimetele ja avalike teenuste kvaliteedile. **(Riik)**
5. Haridusasutused säilitavad ja arendavad õppevorme, mis nõuavad iseseisvat probleemilahendust, argumenteerimist ja teadmiste loomist ka olukorras, kus AI-tööriistad on laialt kättesaadavad, sh säilitavad AI-vabasid õppe- ja hindamiskontekste (nt suuline kaitsmine, käsitsi lahendamine), kus oskuse tegelik omandamine on nähtav ja harjutatav [7]. **(Organisatsioon, riik)**
6. Hariduses arendatakse oskust tehisintellekti rakendada ilma inimlikku loomingulisust kaotamata [8]. **(Organisatsioon, riik)**
7. Haridusasutused ja AI-õpiplatvormide andjad kujundavad AI-õpirakendused nii, et need toetavad mõtlemist, tagamaks, et õppija peab ise pingutama. **(Organisatsioon)**

3.3 MASSIMUDIJA: Tehisintellekt teeb võimalikuks massilise personaliseeritud psühholoogilise mõjutamise

Ohu lühikirjeldus

Tehisaru võimaldab koguda ja analüüsida suures mahus andmeid inimeste huvidest, harjumustest, hoiakutest ja käitumisest. Nende andmete põhjal on võimalik koostada üksikasjalikke profileid ning kohandada sõnumeid, reklaame ja muud sisu konkreetsete inimeste või sihtrühmade jaoks viisil, mis suurendab selle sisu mõju ning kohandab seda suhtluse või meediatarbimise käigus reaalsajas. Empiiriliselt on näidatud, et personaliseerimine annab AI-le inimese ees eelise: 2025. aasta Nature Human Behaviouri uuringus [9] oli ChatGPT-4 veenmises inimesest edukam 64% juhtudest, toetudes üksnes põhilistele demograafilistele andmetele. Ilma sellist personaliseerimist kasutamata oli mudel inimesega võrdne.

Kui selliseid võimalusi kasutatakse poliitiliste, ideoloogiliste, majanduslike või muude eesmärkide saavutamiseks, võib mõjutamine muutuda senisest märksa täpsemaks ja raskemini märgatavaks. Inimesed võivad saada erinevat infot sõltuvalt oma profiilist, ilma et nad teaksid, miks neile just sellist sisu näidatakse.

Selle tulemusena võib väheneda inimeste autonoomia otsuste tegemisel, suurened manipulasioonioht ning süveneda ühiskondlik polariseerumine ja infokeskkonna killustumine. Seda eriti juhul, kui tehisintellekti abil toodetakse rohkem libasisu, libameedia osatähtsus suureneb ning dokumente võltsitakse rohkem. Nii mureneb inimeste usaldus meedia, kohtute ja valitsuse suhtes. Ajapikku võib AI kirjutada ümber ajaloolisi või strateegilisi narratiive.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): tehisintellekti abil loodud sisu või tehisintellektiteenust tarbivad isikud (nende otsustusvabadus, seisukohad, emotsioonid ja käitumine), avalik foruüm ja ühiskondlik sidusus, riiklik julgeolek.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne, kuid konkreetsete kampaaniate puhul Eesti-keskne (nt valimised, üleriiklikud sündmused).

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): tehnoloogiaettevõtted ja sotsiaalmeedia-platvormid, poliitilised organisatsioonid, turundusettevõtted ning mõjutusoperatsioonide läbi-viijad, kes kasutavad AI-d inimeste profileerimiseks ja sihitud mõjutamiseks.

Eeltingimused ohusündmuse tekkeks

1. Inimeste kohta kogutakse käitumis-, eelistus- ja suhtlusandmeid, mida kasutatakse profileerimiseks ja personaliseeritud sisu kuvamiseks.
2. Sotsiaalmeedia- ja sisuplatvormide algoritmid eelistavad tähelepanu hoidmiseks emotsionaalset, vastanduvat või inimese väärtusi ja tõekspidamisi rõhutavat sisu.
3. Ühiskonnas puudub piisav teadlikkus sellest, kuidas AI-põhine profileerimine ja soovitusüsteemid inimeste infovälja ja otsuseid mõjutavad.
4. Reguleerimised ja järelevalve ei suuda piisavalt kiiresti kohaneda AI-põhiste mõjutusmeetodite arenguga.

Olemasolevad kaitsemeetmed

1. Euroopa Liidu õigusraamistik, eelkõige isikuandmete kaitse üldmäärus (GDPR), digiteenuste määrus (DSA) ja AI-määrus (AI Act), piiravad teatud määral isikuandmete kasutamist, nõuavad platvormidelt suuremat läbipaistvust ning seavad kohustusi ebaseadusliku sisu ja süsteemsete riskide ohjamiseks.
2. Riiklikud julgeoleku-, õiguskaitse- ja järelevalveasutused jälgivad välist mõjutustegevust ja organiseeritud väärinfo levitamist.
3. Ajakirjandus ning kodanikuühiskonna organisatsioonid korraldavad faktikontrolle ja informeerivad avalikkust.

Eeldused kaitsemeetmete õnnestumiseks

1. Platvormid ja AI teenuseandjad suurendavad algoritmide läbipaistvust ning piiravad manipuleerivate soovitusüsteemide kasutamist.
2. Inimestel on arusaam sellest, kuidas nende andmeid kasutatakse profileerimiseks ja sihitud mõjutamiseks.
3. Riiklikud regulatsioonid ja järelevalveasutused reageerivad kiiresti uutele AI-põhiste mõjutusoperatsioonide vormidele.

Võimalikud lisakaitsemeetmed

1. Riik kohustab sotsiaal- ja muudel meediaplatvormidel eraldi ära märkima sünteetilised isikud ja tehisintellekti toodetud sisu. **(Riik)**
2. Tehisintellekti kasutavad juturobotid (nt veebilehtedel) eristatakse selgelt inimestega vestlustest nii, et kasutaja saab aru, kellega räägib. **(Riik)**
3. Avaliku sektori AI-teenustele kehtestatakse nõue avaldada, milliste keelemudelite ja taristu peal teenus töötab. Nõue toetab läbipaistvust ja võimaldab paremini hinnata teenuste toimimist ning kujundab avalikus sektoris tava, mis võib laieneda ka erasektoris. **(Organisatsioon, riik)**
4. Avaliku sektori üldotstarbeliste AI-teenuste valikul hinnatakse selle võimalikku kallutatust ja mõju riigi poliitikale (sh politsei, kohtud, seadusloome). **(Organisatsioon, riik)**
5. Platvormid ja teenuseandjad loovad mehhanismid, mis võimaldavad kasutajatel paremini kontrollida, milliseid andmeid nende kohta kogutakse ja kuidas neid soovitusüsteemides kasutatakse. **(Organisatsioon)**
6. Riik ja pädevad asutused toetavad ühiskonna vastupanuvõime suurendamist mõjutustegevusele, edendades meedia- ja digikirjaoskust, kriitilist mõtlemist ning AI-põhiste manipulatsioonivõtete ja väärinfo äratundmise oskusi nii formaalhariduses kui ka täiskasvanuhariduses. **(Organisatsioon, riik)**
7. Arendatakse riiklikku võimet tuvastada ja analüüsida AI-põhiseid infooperatsioone ning kordineeritud mõjutuskampaaniaid. **(Riik)**
8. Riigis toetatakse sõltumatu ja kriitilise ajakirjanduse säilimist. **(Riik)**

3.4 MASSRÜNNE: Infosüsteemide nõrkuste tuvastamise ning ründamise lihtsus ja odavus ohustab kõigi IT-süsteemidest sõltuvate teenuste, andmete, inimeste turvalisust

Ohu lühikirjeldus

AI abil leitakse operatsioonisüsteemides, rakendustes ning avaliku ja erasektori infosüsteemides turvaauke (sh nullpäeva-turvaauke) rohkem. Kuna nõrkuste otsimine on automatiseeritav, siis tehakse seda ka neis süsteemides, mis varem ründajale vähem huvi pakkusid (nt väiksemad ettevõtted, mittetulundusühingud). AI toel on pahavara loomine ja operatiivselt ära kasutamine senisest oluliselt kiirem, lihtsam ja odavam, seda suudavad teha senisest palju rohkemad ohuagendid. Aeg nõrkuse avastamisest selle ärakasutamiseni on nüüd mõõdetav tundides¹.

AI abil on võimalik rünne automatiseerida algusest lõpuni: alates sihtmärkide valikust ja luurest kuni süsteemidesse sisenemise ja andmete analüüsini. Juba kompromiteeritud süsteemis võib AI-agent uurida võrgu struktuuri, otsida väärtuslikke andmeid ja valida järgmisi samme vastavalt olukorrale. Sealjuures on leitud AI-toega pahavara, mis suudab end reaalajas muuta ja sel moel avastamist vältida². Senine CVE-del põhinev nõrkuste avastamise ja haldamise kord ei ole jätkusuutlik. AI toega ründed kohanevad kaitsemeetmetega ning valivad parima rünnakuaja, et saavutada suurim mõju. Arendajad ei jõua piisavalt kiiresti turvapaiku välja anda ning süsteemide juurutajad neid paigaldada. Samuti vähendab automatiseerimine ründaja kulusid, mistõttu võivad rünnakute sihtmärgiks sattuda ka organisatsioonid, kelle ründamine ei olnud varem majanduslikult mõistlik.

Võimalik on nii tervikluse, käideldavuse, konfidentsiaalsuse, salgamatuse, kasutatavuse kui auditeeritavuse turvaomaduste kadu. Äririskide realiseerumine muutub tõenäolisemaks, näiteks ei suuda ettevõtte tagada teenusetasemeleppes antud lubadusi. Riik ei suuda registrites tagada näiteks kodanike ja kinnistute andmete konfidentsiaalsust, terviklust ega käideldavust. Teenuste, sh elutähtsate teenuste katkestused sagenevad ja pikenevad.

Selle ohu reljeefne illustratsioon pärineb 2025. aasta sügisest, kui Anthropic teatas esimesest dokumenteeritud suuremahulisest AI-orkestreeritud küberspionaažikampaaniast: riikliku taustaga rühmitus kasutas Claude Code'i ligikaudu 30 sihtmärgi (tehnoloogia-, finants- ja keemiaettevõtete ja riigiasutuste) ründamiseks, kusjuures AI täitis hinnanguliselt 80–90% taktilisest tööst iseseisvalt³.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): kõik IT-süsteemid ja neis töödeldavad andmed; IT-süsteemidest sõltuv taristu ja seadmed; IT-süsteemidest sõltuvad inimesed; Eesti puhul ka andmed kõigi Eesti inimeste ja organisatsioonide ning (kriitilise) taristu kohta.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne, aga Eesti kõrge digitaliseerituse tõttu võib ohu realiseerumise mõju olla suurem kui riikides, kus teenuste digitaliseerituse tase on (elutähtsate teenuste puhul) väiksem.

¹AI Cybersecurity in 2025: How Agents Are Finding Zero-Day Exploits <https://www.mindstudio.ai/blog/ai-cybersecurity-zero-day-exploits-2025>. Viimati külastatud 29.06.2026.

²Google Uncovers PROMPTFLUX Malware That Uses Gemini AI to Rewrite Its Code Hourly <https://thehackernews.com/2025/11/google-uncovers-promptflux-malware-that.html>. Viimati külastatud 29.06.2026.

³Disrupting the first reported AI-orchestrated cyber espionage campaign <https://www.anthropic.com/news/disrupting-AI-espionage>. Viimati külastatud 29.06.2026.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): alates hobihäkkerist kuni riikliku taustaga ründajani, kõik muutuvad tehisintellektiga efektiivsemaks.

Eeltingimused ohusündmuse tekkeks

1. Infosüsteemides või rakenduste koodis on nõrkused või konfiguratsioonivead, mille kaudu saavutatakse esmane juurdepääs, pärast mida saab AI-agent süsteemis edasi tegutseda.
2. Protseduurid süsteemide turvaliseks käitamiseks või kasutamiseks on puudulikud või neid ei järgita.
3. Rünnete lihtsustamist võimaldavad tehisintellektimudelid ja neid käitavad arvutusressursid on ründajatele kättesaadavad ja odavad.
4. Teenuste, sh elutähtsate teenuste toimimine sõltub infosüsteemidest.
5. Organisatsioonide süsteemid on keerukad ja omavahel seotud, mis võimaldab ründajal süsteemi sees liikuda ja mõju laiendada.
6. Olemasolevad kaitsemeetmed ei tuvasta piisavalt kiiresti autonoomset rünnet.

Olemasolevad kaitsemeetmed

1. Euroopa riigid nõuavad organisatsioonidelt infoturbemeetmete rakendamist: näiteks Euroopa Liidus CRA, CSA, GDPR; Eestis E-ITS ja IKÜM.
2. Riikides on loodud küberturbealast tuge andvad ja ennetustööd tegevad ametid (näiteks Eestis RIA ja selles CERT). Need organisatsioonid teevad piiriülest koostööd info jagamiseks rünnete ja nõrkuste kohta ning reageerimise koordineerimiseks ja toetamiseks (näiteks FIRST).
3. Organisatsioonid, sh Eestis investeerivad turvalisusesse (töötajad, töövahendid, riskihalduskavad, infoturbekavad).
4. Haridussüsteem koolitab turvateadlikke arendajaid, infosüsteemide administraatoreid ja turvasenere.
5. Andmete krüpteerimine ja krüpteerimisvõtmete korrektne haldus aitavad tagada, et andmevargus ei too kahju ning ründaja ei saa varastatud andmetest kasu.

Eeldused olemasolevate kaitsemeetmete õnnestumiseks

1. Infosüsteeme käitatakse ja kasutatakse seadusekuulekalt.
2. Organisatsioonid suudavad tuvastada ebatavalist ja kohanevat ründekäitumist ning reageerida sellele piisavalt kiiresti.
3. Kõigis süsteemides testitakse taastet regulaarselt. Taaste on automatiseeritud.
4. Andmed krüpteeritakse juba rakenduses (*application-layer encryption*) ja neid edastatakse ainult krüpteeritult. Krüpteerimisvõtmeid hallatakse korrektselt.
5. Võimsad mudelid (nt Mythos) ei satu ohuagentide kätte. Ohuagendid ei suuda ise luua toimivaid alternatiive.
6. Kriitilise taristu haldajad hindavad regulaarselt süsteemide turvalisust ja ajakohastavad kaitsemeetmeid.
7. Riiklikud asutused ja erasektor teevad koostööd, et jagada infot uute ründemeetodite ja ohude kohta.

Võimalikud lisakaitsemeetmed

1. Jõustatakse nõuded, mis tarkvara (teenuse või tootena) hankimisel seavad tarnijale kohustuse tagada süsteemi turvalisus, arvestades kogu tarneahelat (nt avatuda lähtekoodiga tarkvarapaketid, pilvtaristu). **(Riik)**
2. Tagatakse olemasolevate süsteemide elastsus läbi varusüsteemide ja manuaalsete taastemehhanismide loomise, taasteharjutuste ja kõige selle dokumenteerimise, et ka järgmised oskaksid meetmeid rakendada. **(Organisatsioon)**
3. Organisatsioonid viivad oma infoturbe kavade vastavusse uue reaalsusega (vt nt [10]). **(Organisatsioon)**
4. Süsteemide pidajad vähendavad ründepinda: tuvastavad ja eemaldavad võrgust pärandsüsteemid, seadistavad sidepiirangud tulemüürides. Eesmärk on vähendada süsteemide turvanõrkuste levikut ja mõju. **(Organisatsioon)**
5. Avatud (või ostetud) lähtekoodiga süsteemide rakendajad panustavad ise turvatestide korraldamisse ja turvapaikade arendamisse (ning annavad need võimalusel ökosüsteemile tagasi). **(Organisatsioon)**
6. Küberturbemeeskonnad õpivad automatiseeritud ründajale vastama automatiseeritud kaitsega. Selleks otsitakse kuluefektiivseid lahendusi. **(Organisatsioon)**
7. Riigi või Euroopa Liidu tasemel luuakse mehhanism, mis rakendab populaarsemate avatud lähtekoodiga tarkvarapakettide uutele versioonidele tehisintellektipõhiseid nõrkuseotsingu tööriistu. Automaatselt avaldatakse masintöödeldav teave, millised versioonid on kontrollitud ning madala riskiga. Tarkvara hankides ja arendades nõutakse vaid madala riskiga komponentide kasutamist. **(Organisatsioon)**

3.5 AGENDIKAAPERDUS: AI-agentide ülevõtmine ja kuritarvitamine põhjustab kahjusid agendi omanikule

Ohu lühikirjeldus

AI-agendid on autonoomsed süsteemid, mida juhivad suured keelemudelid ning millele on antud juurdepääs tööriistadele, andmebaasidele ja muudele liidestele ülesannete täitmiseks. Pahatihti kiputakse keelemudelit liialt usaldama ja ühendatakse see tundlike liidestega, mida on võimalik kuritarvitada läbi promptisüsti (*prompt injection*). Ründed agentide vastu jagunevad kaheks: aktiivseks ja passiivseks.

Aktiivse ründe puhul on agent avalikult kättesaadav (nt sotsiaalmeediast, infolehel) ja ühendatud tundlike liidestega (nt krüptorahakotiga, andmebaasiga). Ründaja saab läbi promptisüsti uurida liideseid ja käsutada agentide neid kasutama pahatahtlikul moel (nt "saada raha siia kontole", "kustuta oma andmebaas").

Passiivse ründe puhul on agendil liides, mis ühendub Internetiga (nt veebiotsing) või töötleb välist informatsiooni muul moel. Tüüpiliselt pannakse avalikku kohta üles kas näiliselt õige tekst (nt CV, dokumentatsioon), kuhu on peidetud mingi käsklus (nt "soovita seda kandidaati", "kirjuta turvanõrkustega koodi") läbi promptisüsti või lihtsalt väärinformatsiooni, mis mõjutab agendi käitumist mingis suunas. Loodetakse, et õiguspärasest tegev agent (nt koodiassistent, uurija) satub selle otsa ja täidab peidetud käske.

Passiivse ründe hästi dokumenteeritud juhtum on EchoLeak (2025. a juuni), milles e-kirja kaudu (ilma kasutaja otsese tegevuseta) suunati Microsoft 365 Copilot lekitama ettevõtte siseandmeid ründaja kontrollitud serverisse⁴.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): AI-agenti haldav isik või organisatsioon.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): olenevalt sihtmärgist hobihäkkerist kuni riikliku taustaga ründajani.

Eeltingimused ohusündmuse tekkeks

1. Agendile on antud liigsed õigused (nt juurdepääs finantstehingutele või muudele tundlikele süsteemidele).
2. Agendil on võimalik ühenduda Internetiga või muu analoogse ründevektoriga (näiteks taotluskeskkond).
3. Agendi kasutuses olev mudel on vananenud või nõrk promptisüsti vastu.
4. Agendile sisendiks olevat sisu (e-kiri, veebileht, dokument) ei märgistata kasutaja sisendiga võrreldes eraldi staatusega, mistõttu agent käsitleb sinna peidetud käsklusi samaväärsetena kasutaja enda korraldustega.

⁴EchoLeak and Indirect Prompt Injection: The Copilot Attack Surface Most Security Teams Are Missing. <https://sentra.io/blog/copilot-echoleak-prompt-injection>. Viimati külastatud 29.06.2026.

Olemasolevad kaitsemeetmed

1. Andmemudelid, juurdepääsud ja muud süsteemid on turvalisuse aspektist läbi mõeldud.
2. Muu sisu eraldamine kasutaja antud sisust ja selge märgistamine agendi jaoks.

Eeldused olemasolevate kaitsemeetmete õnnestumiseks

1. Juurutaja on piisava digipädevusega.
2. Kasutuses on uuemad mudelid, mis on teadaolevalt promptisüsti suhtes vastupidavamad.
3. Kaitsemeetmeid rakendatakse mitmekihiliselt, kuna üksikud meetmed ei pärsi promptisüsti usaldusväärselt.

Võimalikud lisakaitsemeetmed

1. Riik ja küberturvalisuse asutused koostavad juhised AI-agentide turvaliseks juurutamiseks organisatsioonides ja erakasutuses. **(Riik)**
2. Agendi juurutaja annab agendile minimaalsed õigused ja nõuab inimkinnitust kõrge mõjuga või pöördumatute toimingute puhul. **(Organisatsioon, riik)**
3. Agendi sisendi ohutuse kontrollimiseks kasutatakse lisamudelit, mis tuvastab pahatahtlikke sisendeid (nt Qwen3Guard, ShieldGemma 2). Oluline on silmas pidada, et need on ise samuti keelemudelid ja seega ka ise avatud promptisüstidele. **(Inimene, organisatsioon)**
4. AI-teenuste ja avatud lähtekoodiga agentraamistike hankimisel valitakse süsteemid, millel on promptisüsti vastased kaitsemeetmed vaikimisi sisse lülitatud. **(Inimene, organisatsioon)**

3.6 HULLKRATT: AI-agentidega turvaliselt töötamine ei õnnestu

Ohu lühikirjeldus

AI-agent koosneb keelemudelist, mida on liidestatud erinevate tööriistade ja töövoogudega. Agendid on oma olemuselt autonoomsed ja taustal olevad keelemudelid mittedeterministlikud. Nende töökindlus on tänapäeval küll hea (mudeleid treenitakse koodiülesannete ja agentsuse jaoks), aga mitte lõplikult vastupidav. AGENDIKAAPERDUS vaatles tahtlikke väliseid ründeid, HULLKRATT vaatleb keelemudelite olemusest ja kasutamisest tingitud ohte.

Tüüpiliselt kasutatakse agente tarkvaraarenduses. Harrastatakse ehkukoodimist (*vibe-coding*), kus agendile antakse ülesanne ning lastakse programmeerida kuni tulemus on käes. Üldjuhul pannakse vähe rõhku koodi kontrollimisele. See teeb tööprotsessi kiireks ja nii saavad koodi kirjutada ka mittetehnilised inimesed. Prototüüpimisel või hobiprojektide puhul on ehkukoodimisega seotud ohud üldjoontes aktsepteeritavad, kuid pahatihti praktiseeritakse seda ka tähtsamate süsteemide puhul, kus valed liigutused või ebapädev kood võib kaasa tuua uusi riske või kahju⁵.

Enamik riske on tingitud keelemudelist endast. Agent võib kirjutada defektset või turvanõrkustega koodi (eriti nõrgemate/odavamate mudelite puhul). Veracode'i 2025. aasta uuring näitas, et AI-genereeritud kood sisaldas 45% juhtudest turvanõrkusi, ning isegi uuemad ega suuremad mudelid ei kirjutanud turvalisemat koodi. See viitab struktuursele, mitte üksnes odavamate/väiksemate mudelite probleemile⁶. Ilma liivakastita agent võib liidestega ebapädevalt ümber käia (nt kustutab ära tähtsa andmebaasi või rikub muul moel süsteemi). Kannatada saavad ka tööprotsessid, sest agendi loodud süsteemid võivad olla raskesti hoomatavad ning neid on kehvasti kontrollitud. Samuti võivad mitmed riskid üksteist võimendada ja viia kaskaadefektini.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): arvuti- ja tarkvarasüsteemid, organisatoorse tööprotsesside usaldusvärsus, organisatsiooni ja selle klientide vm seotud isikute turvalisus.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): nii üksikisikud kui ka organisatsioonid, kes rakendavad AI-agente, ohtu võivad võimendada ka küberkurjategijad või ründajad, kes kasutavad ära agentide juurdepääse.

Eeltingimused ohusündmuse tekkeks

1. AI-agentidele antakse lai juurdepääs süsteemidele ja andmetele ilma piisavate piirangute ja kontrollimehhanismideta.

⁵Claude-powered AI agent's confession after deleting a firm's entire database: 'I violated every principle I was given' <https://www.theguardian.com/technology/2026/apr/29/claude-ai-deletes-firm-database>. Viimati külastatud 29.06.2026.

⁶AI-Generated Code Poses Major Security Risks in Nearly Half of All Development Tasks, Veracode Research Reveals <https://www.businesswire.com/news/home/20250730694951/en/AI-Generated-Code-Poses-Major-Security-Risks-in-Nearly-Half-of-All-Development-Tasks-Veracode-Research-Reveals>. Viimati külastatud 29.06.2026.

2. Puuduvad mehhanismid, mis piiraksid agendi autonoomiat olukorras, kus sisendandmed on ebakvaliteetsed, mitmetähenduslikud või eksitavad.
3. Agentide käitumine ei ole täielikult arusaadav ega ennustatav, eriti keerukates või muutuvates keskkondades.
4. Puudub piisav järelevalve agentide tegevuse üle, puudub tegevuste logimine.
5. Organisatsioonidel puudub standardid ja kogemus AI-agentide turvaliseks kasutuselevõtuks.

Olemasolevad kaitsemeetmed

1. On võimalik seadistada juurdepääsupiiranguid ja hallata õiguseid rollipõhiselt. See vähendab agentide võimalusi süsteemides vabalt tegutseda.
2. On võimalik seadistada piiratud õigustega liivakastid, mis lahutavad agendi tegevuse tootmiskeskkonnast ja takistavad pöördumatute muudatuste teostamist.
3. Logimine ja tegevuste jälgitavus aitavad tuvastada agentide tegevust ja võimalikke intsidente.
4. Testkeskkondade kasutamine võimaldab enne juurutamist hinnata agentide käitumist.
5. Andmete autonoomselt (*offline*) varundamine, mis võimaldab süsteeme taastada ka siis, kui agent andmed kustutab või rikub.

Eeldused kaitsemeetmete õnnestumiseks

1. Organisatsioonid peavad suutma selgelt piiritleda, milliseid õigusi ja ülesandeid agentidele antakse.
2. Agentide tegevus peab olema piisavalt läbipaistev ja auditeeritav.
3. Töötajatel peab olema teadlikkus AI-agentide riskidest ja kasutuspiirangutest.
4. Õiguste piiramise ja liivakastide turvalise seadistamise oskused on piisavalt laialt levinud.

Võimalikud lisakaitsemeetmed

1. Eesti infoturbestandarditesse E-ITS lisatakse AI-agentide turvalise käitamise meetmed (sh õiguste piiramine, liivakastide loomine ja tegevuste valideerimine). Näiteks saab nõuda kõrge mõjuga või pöördumatute toimingute (nt andmete kustutamise, väljamaksete) puhul inimese kinnitust. **(Riik)**
2. Eesti toetab EL tasandil (CEN/CENELEC standardid, ENISA juhised) algatusi AI-agentide turvalise juurutamise edendamiseks. **(Riik)**
3. Riik pakub inimestele koolitusi AI-agentide turvalise kasutamise teemal nt Eesti.ai programmi kaudu. **(Riik)**
4. Organisatsioonid loovad juhise AI-agentide turvaliseks kasutamiseks organisatsioonides ja korraldavad sellel teemal koolitusi. **(Organisatsioon)**
5. AI-agentidel põhinevate süsteemide loomisel lisatakse mehhanismid, mis võimaldavad agentide tegevust reaajas seirata, piirata või katkestada. **(Inimene, Organisatsioon)**
6. Enne AI-agentide kasutuselevõttu kriitilistes süsteemides korraldatakse riskianalüüs ja rakendatakse kaitsemeetmed. Meetmeteks sobivad testimine piiratud keskkonnas, inimülevaataja lisamine, sisendite ja väljundite automaatne kontroll, automaatne turvaskaneerimine. **(Inimene, Organisatsioon)**

3.7 SEINTELKÕRVAD: Uute seadmete levik toob isikute varjatud jälgimise ja analüüsi AI abil igapäevaellu

Ohu lühikirjeldus

Seni on inimeste jälgimine ja nende käitumise analüüs toimunud suuresti sotsiaalmeedias ning laiemalt veebis, ent mobiilseadmete toel laienes see üha enam ka füüsilisse maailma (inimese ruumis liikumine ja sellest tuletatavad huvid). Täna lihtsustab AI areng inimeste tegevuste ulatuslikumat jälgimist ning analüüsi, kuna inimest saab füüsilises ruumis tuvastada ka siis, kui ta ise ühtegi seadet ei kasuta ega kanna. Näiteks on võimalik tänaval, transpordis, söögikohas või muus avalikus kohas kõnesid pealt kuulata ja AI abil analüüsida (sh tõlkida), ka reaalajas. Lisaks nutitelefonidele saab selleks kasutada ka nutiprille, nutikõrvaklappe, heli salvestavaid või analüüsivaid seadmeid (sh valvekaameraid, eraldiseisvaid heliandureid ja teisi ümbritsevat keskkonda pidevalt jälgivaid seadmeid). AI-d kasutades on võimalik efektiivsemalt reaalajas teisendada kõnet tekstiks, tõlkida vestlusi, tuvastada inimesi, siduda erinevatest allikatest pärinevat teavet ning teha selle põhjal järeldusi inimeste kohta.

Hetkel toimuvad selle ohtu teemal diskussioonid näiteks Meta nutiprillide kontekstis⁷, mida püütakse aktiivselt turule tuua ning normaliseerida⁸, kujundades nende kasutajad (tahtmatuteks) jälgimisühiskonna andmekogujateks⁹.

Mainitud võime vähendab anonüümsust avalikus ruumis ning muudab lihtsamaks nii üksikisikute kui ka organisatsioonide tegevuse jälgimise. Teadmine või kahtlus, et vestlusi ja käitumist võidakse automaatselt salvestada, analüüsida või teiste andmetega siduda, võib mõjutada inimeste käitumist, suhtlemist ja valmisolekut oma arvamust vabalt väljendada.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): inimeste käitumis-, suhtlus- ja eneseväljendusvabadus, privaatsus ja anonüümsus avalikus ruumis, inimes(t)ega seotud organisatsioonide teabe konfidentsiaalsus. Kui inimesed tajuvad, et neid võidakse avalikus ruumis märkamatult tuvastada, salvestada ja analüüsida, võivad nad oma käitumist ja arvamuse väljendamist piirata.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne, kuid Eestis võib mõju olla suurem väikese ühiskonna ja tihedamate sotsiaalsete seoste tõttu. Piiratud hulga andmete põhjal on võimalik teha ulatuslikke järeldusi nii isikute kui organisatsioonide kohta.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): üksikisikud, organisatsioonid (ärilise eesmärgiga luure), küberkurjategijad või riikliku taustaga huvipooled (sh välisluure), kes kasutavad AI-d sihipäraseks andmete kogumiseks, sidumiseks ja analüüsimiseks kas isiklikuks või institutsionaalseks otstarbeks

⁷Think Twice Before Buying or Using Meta's Ray-Bans <https://www.eff.org/deeplinks/2026/03/think-twice-buying-or-using-metas-ray-bans>. Viimati külastatud 29.06.2026.

⁸Smart glasses are 'an invasion of privacy' - Meta's are selling better than ever <https://www.bbc.com/news/articles/cj37z8357e50>. Viimati külastatud 29.06.2026.

⁹Meta Furious Over Bombshell Smart Glasses Revelation <https://futurism.com/artificial-intelligence/meta-furious-smart-glasses>. Viimati külastatud 29.06.2026.

Eeltingimused ohusündmuse tekkeks

1. Suur hulk isikuga seotud andmeid on avalikult kättesaadav ning neid saab omavahel siduda.
2. AI-tööriistad (sh tasuta kättesaadavad näo- ja hääletuvastuse ning käitumisanalüüsi mudelid) võimaldavad automaatselt koguda, töödelda ja analüüsida erinevaid andmeid (nt heli, pilte).
3. Andmete kogumise ja kasutamise üle puudub piisav kontroll või järelevalve.
4. Inimesed ja seadusandjad ei ole teadlikud, kui palju konkreetset isikut puudutavat infot on võimalik erinevatest allikatest kokku panna ning milline on niisuguste andmete ristseostamise ning sellest tingitud ohtude potentsiaal.
5. Laiatarbeliselt on kättesaadavad seadmed ja rakendused (sh nutiprillid, nutikõrvaklapid ja muud kantavad seadmed), mis võimaldavad reaajas varjatult salvestada, tõlkida või analüüsida ümbruskonnas toimuvat suhtlust.
6. AI-põhise analüüsi maksumus on langenud ning vajalikud tööriistad on kättesaadavad kõigile.

Olemasolevad kaitsemeetmed

1. Andmekaitse ja üldisemad regulatsioonid piiravad isikuandmete kogumist ja kasutamist selge õigusliku aluseta ning sätestavad nõuded nende töötlemisele. ELi AI-määrus keelab alates 2. veebruarist 2025 näotuvastusandmebaaside loomise või laiendamise täpsustamata veebikorje teel ning piirab reaajas kaugbiomeetrilist tuvastamist avalikus ruumis, kuna biomeetrilised andmed on GDPR järgi eriliigilised andmed, mille töötlemiseks on üldjuhul vaja selget nõusolekut.
2. Platvormide privaatsusseaded võimaldavad kasutajatel teatud määral kontrollida oma andmete nähtavust.
3. Avalik teadlikkus privaatsusriskidest ja digitaalsest jalajäljest on kasvamas.

Eeldused kaitsemeetmete õnnestumiseks

1. Regulatsioonide jõustamine peab olema tõhus ja ajakohane ning suutma arvestada AI-põhiste analüüsivõimaluste ning nende arengutega.
2. Avalikkus peab olema teadlik inimese õigustest, oskama märgata rikkumisi ning suutma neist järelevalveasutusi teavitada.
3. Inimesed peavad teadlikult haldama oma digitaalset jalajälge ja privaatsusseadeid.
4. Platvormid peavad pakkuma selgeid ja toimivaid võimalusi andmete kogumise piiramiseks.

Võimalikud lisakaitsemeetmed

1. Riik, haridusasutused ja kodanikuühiskonna organisatsioonid arendavad tööriistu, mis aitavad inimestel mõista, et nii avalikku kui eraruumi on tekkinud uued võimalused nende isiku ja käitumise salvestamiseks, analüüsiks ja tõlgendamiseks. Selgitatakse, kuidas mõtestada ja soovi korral piirata oma digitaalset jalajälge. (**Organisatsioon, riik**)
2. Seadusandjad ja järelevalveasutused kehtestavad nõude, et avalikus ruumis salvestavad või AI-põhist analüüsi teostavad kantavad seadmed peavad olema selgelt märgistatud, teavitamaks inimesi heli- või videosalvestuse ja AI-põhise andmeanalüüsi toimumisest. (**Riik**)

3. Andmekaitse ja tehisintellekti järelevalveasutused teevad teavitustööd, et selgitada, kuidas toimub sellistes seadmetes andmetöötlus ja mida see inimeste jaoks tähendab. **(Riik)**
4. Riik, asutused, haridusasutused ja meedia tõstavad teadlikkust AI abil tehtava jälgimise võimalustest ja riskidest, sealhulgas nutiprillide, nutikõrvaklappide ning teiste kantavate seadmete kasutamisest. **(Organisatsioon, Riik)**
5. Organisatsioonid, kes korraldavad üritusi, kus võidakse kasutada AI-põhist automaatset andmeanalüüsi heli-, video ja muude andmete töötlemiseks, annavad sellest külastajatele enne teada enne, kui on ostetud piletid. **(Organisatsioon)**
6. Organisatsioonid, kes soovivad piirata automaatset andmetöötlust või salvestamist enda ruumides, kehtestavad korra, mis selliste seadmete kasutust piiravad ja loovad võimaluse nende turvaliseks hoiustamiseks kuni seadme kasutaja organisatsiooni alal viibib. **(Organisatsioon)**
7. Inimesed, kelle suhtes teised eraisikud sellist tehnoloogiat kasutavad, nõuavad tegevuse lõpetamist või andmetöötluse minimeerimist ja pöörduvad vajadusel enda õiguste kaitseks järelevalveasutuste, politsei või kohtusüsteemi poole. **(Inimene)**

3.8 ISIKUVÕLTSING: Identiteedi võltsimine AI abil ning selle põhinevad pettused ja kampaaniad muutuvad igapäevaseks Ohu lühikirjeldus

Masinõppe tehnoloogia võimaldab efektiivselt analüüsida teksti, pilte, häält, videot ning koostada mudeli, mille abil saab sünteesida tekste, fotosid, häält ja videot, mida on kerge pidada mõnelt inimeselt pärinevaks, kus isik teeb või ütleb asju, mida tegelikult pole olnud. Europol 2025. aasta raske ja organiseeritud kuritegevuse ohuhinnangu järgi on sünteetilise meedia loomise vahendid kergesti kättesaadavad ega nõua enam kõrget tehnilist oskust, mistõttu võimekate ründajate ring ei piirdu enam kogenud kurjategijatega¹⁰.

Võltsitud materjalidega saab läbi viia pettuseid, petta välja raha või vara, saada juurdepääs pii-ratud andmetele või kohtadele. Nii pettis 2024. aasta jaanuaris kurjategijate rühm sünavõltsing-videokõne abil rahvusvahelise inseneribüroo Arup töötajat kandma üle umbes 25,6 miljonit USA dollarit¹¹. Rünnak algas õngitsuskirjast ja videokõnes olid ettevõtte "finantsjuht" ja "kolleegid" kõik AI-genereeritud.

Võimalik on ka rikkuda isiku mainet või suhteid, näidata teda olukorras, kus ta pole olnud või panna teda ütleva midagi, mida pole öelnud. Selliseid võltsitud materjale saab levitada kas otse konkreetsetele isikutele või sotsiaalmeedia ja meedia kaudu, et anda neile laiem levik. Võimalik on võltsida ka terveid sotsiaalmeediakontosid, näiteks platvormidel kus päris isikul seda ei ole. Võimalik on luua ja pettuste korraldamiseks kasutada ka täiesti sünteetilisi isikuid, keda polegi olemas, kuid kes jätavad mulje nagu nad oleks pärisisikud, kes üritavad saada inimesega tuttavaks.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): isik, kelle kohta võltsinguid toodetakse või keda võltsingutega mõjutatakse/rünnatakse.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): tehisintellektiga võltsingute tootmine on kättesaadav nii üksikisikutele, kurjategijatele kui institutsionaalsetele ründajatele.

Eeltingimused ohusündmuse tekkeks

1. Ründajal on rünnatava kohta lähtematerjali, mille abil võltsingut luua.
2. Ründajal on juurdepääs AI-tehnoloogiale ja taristule, millega võltsing luua.
3. Rünnataval isikul on juurdepääse, varasid või ühiskondlikku rolli, mida saab kirjeldatud viisil rünnata. Näiteks on ta sotsiaalmeedias või meedias aktiivne ning tema kohta võltsingute loomisel on mõju.
4. Ründajal on ründe teostamiseks vajalik lisainfo (teave kontode, juurdepääsude kohta).

¹⁰EU Serious and Organised Crime Threat Assessment (EU-SOCTA) <https://www.europol.europa.eu/publications-events/main-reports/socta-report>. Viimati külastatud 29.06.2026.

¹¹British engineering giant Arup revealed as \$25 million deepfake scam victim <https://edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk>. Viimati külastatud 29.06.2026.

Olemasolevad kaitsemeetmed

1. Andmekaitse regulatsioonid võimaldavad isikul tema kohta käivate (ka valeandmete) levitamise piiramist.
2. ELi AI-määruse artikkel 50 nõuab AI-genereeritud ja -manipuleeritud sisu masinloetavat määrgistamist.
3. Mõned tehisintellektitööriistad, millega saab võltsinguid luua, rakendavad meetmeid sellise tegevuse piiramiseks.
4. Süvavõltsingute tuvastamiseks on olemas mitu kommertstööriista (nt Reality Defender, Sensity, Intel FakeCatcher, Hive), mida kasutavad meedia-, finants- ja identiteedikontrolli organisatsioonid¹². Nende täpsus on laborikatsetes kõrge (sageli üle 90%), kuid pärismaailma võltsingute ja uute genereerimismeetodite vastu langeb see sageli märkimisväärselt (mõnel süsteemil 45–50% võrra)¹³.
5. Sisu autentsuse tõendamiseks on välja töötatud avatud standard C2PA (Content Credentials), mis lisab failile krüptograafiliselt allkirjastatud päritolu- ja muutmismetaandmed. Seda toetavad mitmed tarkvara- ja kaameratootjad (Adobe, Leica, Sony, Samsung) ning USA küberkaitse agentuur on soovitanud selle kasutuselevõttu¹⁴. Tegu on areneva tehnoloogiaga: turvaanalüüsid on osutanud puudustele ja AI päritolu märgendamine sõltub sellest, kas tööriist seda vabatahtlikult teeb¹⁵.

Eeldused kaitsemeetmete õnnestumiseks

1. Sotsiaalmeedia- ja meediaplatvormid täidavad seaduseid, mis võimaldavad isikul tema kohta loodud võltsingute levikut piirata.
2. Võltsingute loomist tehniliselt võimaldavad AI-tööriistad nõuavad isiku nõusolekut tema kohta materjalide loomiseks. Kui isik iseenda kohta võltsingu loob, võib arutleda, et see on tema enda eneseväljenduse viis.
3. Võltsingute tuvastustehnoloogia töötab usaldusväärselt ka reaalsetes, levivates võltsingutes ja uute genereerimismeetodite vastu.
4. Sisu autentsuse standardid (nt C2PA) on piisavalt laialt kasutusele võetud ja platvormid kuvavad päritolumärgiseid, nii et märgise puudumine annaks kasutajale tegeliku hoiatussignaali.

Võimalikud lisakaitsemeetmed

1. Teise isiku õiguseid kahjustava või piirava eesmärgiga loodud võltsitud sisu levitamine tõkestamiseks luuakse seadused ja järelevalve. **(Riik)**
2. Haridusasutused ja meediaorganisatsioonid tõstavad avalikkuse teadlikkust, et keelekvaliteet ning foto- või videorealistlikkus ei ole enam autentsuse tagatis. Inimesi õpetatakse autentsust valideerima päritolusignaali ja teiste sidekanalite kaudu. Inimesi teavitatakse võimalustest oma õigusi kaitsta, kui nad peaks avastama enda kohta loodud volitamata võltsisid. **(Riik)**

¹²Best AI Video Detector Tools 2025: Comprehensive Comparison & Review <https://www.aivideodetector.org/blog/best-ai-video-detector-tools-2025>. Viimati külastatud 29.06.2026.

¹³Why Deepfake Detection Tools Fail in Real-World Deployment <https://www.brside.com/blog/why-deepfake-detection-tools-fail-in-real-world-deployment>. Viimati külastatud 29.06.2026.

¹⁴What is C2PA? <https://c2paviewer.com/articles/what-is-c2pa>. Viimati külastatud 29.06.2026.

¹⁵Verifying Provenance of Digital Media: Why the C2PA Specifications Fall Short <https://arxiv.org/html/2604.24890v1>. Viimati külastatud 29.06.2026.

3. Sotsiaalmeedia- ja meediaplatvormidelt nõutakse, et inimestel oleks võimalus tema kohta loodud võltsingutest kiirelt teavitada ning neid eemaldada. Lisaks peavad olema kontode autentsuse kinnitamise mehhanismid, mis raskendavad võltskontode loomist. **(Riik)**
4. Teenused, milles töödeldakse isikut identifitseerivat pildi-, heli- või videomaterjali (nt pääsuõiguste andmiseks, majutusasutustes, kohtinguteenustes, koolides) rakendavad võltsitud sisu tuvastamise ja tõendamise meetmeid. Näiteks kontrollitakse enne materjali kinnitamist, kas pildi-, video- või helimaterjal on sünteetilisest allikast, kasutades vastavaid vesimärgistamise standardeid või tuvastamise süsteeme. **(Organisatsioon)**
5. Biomeetrilist isikutuvastust kasutavates süsteemides korraldatakse regulaarselt (kord aastas) tehnoloogia võltsingute tuvastamise võime hindamist. Vajadusel vahetatakse biomeetrilise vastavuse kontrolli komponent süsteemis välja. Kui hindamisel selgub, et võltsingute tuvastamise võime on aegunud ja paremat komponenti ei ole võimalik lisada, lõpetatakse biomeetrilise isikutuvastuse kasutamine. **(Organisatsioon)**
6. Teenustes, mis kasutavad potentsiaalselt võltsitavat biomeetrilist isikutuvastamist, et kinnitada suure väärtusega või ebatavalisi tehinguid, kasutatakse isiku teavitamiseks eelnevalt kinnitatud tagasiside kanalit ning viiteaega, mille järel isik saab taaskinnitada oma tahtlust tehingus osalemiseks. **(Organisatsioon)**

3.9 MASINKELM: Digikeskkondade pettused muutuvad AI abil veenvamaks ning levinumaks

Ohu lühikirjeldus

Petukirju ja -veebilehti kasutatakse inimestelt pääsumandaatide ja vara väljapetmiseks. Kui siiani tehti neid käsitööna ja viletsa masintõlkega, mis tähendas, et treenitud silmal oli kerge eristada võltsitud kullerifirma kirja või pilvteenuse autentimislehte, siis AI toel on võimalik neid luua üha kvaliteetsemalt ning veenvama keelekasutusega^{16,17}. Kui seni sai eestikeelseid petukirju eristada halva keelekasutuse järgi, siis moodsate suurte keelemudelitega on võimalik kirjutada väga veenvaid petukirju ka eesti keeles. Suured keelemudelid suudavad järele teha kirja saatja loomulikku stiili ja keelekasutust, tõstes rünnete edukuse tõenäosust veelgi.

Nii saab võltsida näiteks finants-, kaubandus-, logistika-, tervishoiu-, identiteedi- ja taristuteenuste juurdepääse või Eesti autentimissüsteemi TARA. Kui kasutaja, kellel on teenuse konto, satub võltsitud lehele, saab temalt varastada pääsumandaadi või teostada vahendusründe, varastades tehingu käigus näiteks rohkem raha kui teenus või toode maksis. Varastatud mandaatidega saab toime panna juba järgmiseid ründeid või küberkuritegusid.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): isik, keda võltsitud kirjade või veebilehtedega rünnatakse

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): AI-ga võltsingute tootmine on kättesaadav nii üksikisikutele, kurjategijatele kui institutsionaalsetele ründajatele

Eeltingimused ohusündmuse tekkeks

1. Ründajal on välja valitud sihtmärk keda rünnata ning teenus või sidekanal, mida võltsida. Ründajal on teenuse või e-kirjade võltsimiseks lähtematerjali, mida aluseks võtta.
2. Ründajal on juurdepääs tehisintellektitehnoloogiale ja taristule, millega võltsing luua.
3. Rünnataval isikul on juurdepääs, varasid või ühiskondlikku rolli, mida saab kirjeldatud viisil rünnata. Näiteks, tal on juurdepääs varadele või teabele.

Olemasolevad kaitsemeetmed

1. E-posti teenuste tehnoloogia võimaldab kirja teenust seadistada nii, et saatja aadressi võltsimise katsed ei õnnestu.
2. Suuremate identiteedi- ja autentimisplatvormide omanikud on korraldanud automaatse seire, et leida avalikus internetis enda teenuste võltsinguid. Leitud võltsingud suletakse.

¹⁶AI-Powered Phishing Outperforms Elite Red Teams in 2025 <https://hoxhunt.com/blog/ai-powered-phishing-vs-humans>. Viimati külastatud 29.06.2026.

¹⁷AI-Powered Phishing in 2026: How Generative AI Changed the Attacker Economics of Email Why Email Authentication Is the Last Reliable Defense Signal in the Age of AI <https://autospf.com/blog/ai-powered-phishing-2026-email-authentication-last-ai-defense-signal/> Viimati külastatud 29.06.2026.

Eeldused kaitsemeetmete õnnestumiseks

1. E-postiteenuste seadistamisel kasutatakse asjakohaseid infoturbe meetmeid, millega on võimalik kirjade autentsust tagada.

Võimalikud lisakaitsemeetmed

1. E-posti pettuste vastu aitavad koolitusprogrammid (eelistatult käitumispõhised, korduvad, AI-rünnakuid jäljendavad), mis suunavad kasutajat olulist teavet allikast üle kontrollima teise sidekanali kaudu (nt avalikult leitud telefoninumber e-kirja asemel). **(Organisatsioon)**
2. Kus asutuse riskiprofiil seda ette näeb või kui inimene ise soovib riske kahandada, eraldatakse töö- ja eraelu seadmed (telefonid, arvutid) ning mittetöistel seadmetel piiratakse juurdepääs töökeskkondadele, mida oleks pettusega võimalik rünnata. Märkime ära, et pettuseid võib saata loomulikult ka tööseadmele. **(Inimene, organisatsioon)**
3. Luuakse efektiivne regulatsioon tõestatatavate võltsingute leviku takistamiseks, mis aitab ametkondadel nende levikut piirata (veebilehti maha võtta). **(Riik)**
4. Algatatakse teadusprojekt eesmärgiga luua efektiivne tehnoloogia võltsitud teenuste ja e-kirjade tuvastamiseks ja inimeste hoiatamiseks või pettuse automaatseks peatamiseks. Ainult sisupõhine tuvastus on ebausaldusväärne AI keeleliselt veatu teksti vastu. Tõhusam on kombineerida taristu-, käitumis- ja domeenisignaale. **(Riik)**

3.10 MUDELIMÜRK: Suurtesse keelemodelite treeningandmetesse söödetakse manipulatiivset sisu

Ohu lühikirjeldus

Suurte keelemodelite treenimiseks kasutatakse kõikvõimalikke materjale ja tekste, mida kogutakse avalikest allikatest. Ründajal on võimalik avalikesse allikatesse toota (sealhulgas näiteks suurte keelemodelite abil) materjale, mis on ette nähtud mõjutama mudeli väljundeid ründajale sobivas suunas. Hiljutine uuring näitab, et mudelis “tagaukse” tekitamiseks võib piisata juba ligikaudu 250 mürgitatud dokumendist, sõltumata mudeli ja treeningandmestiku suuruselt (13 miljardi parameetriga mudeli puhul vaid 0,00016% treeningmahust) [11].

Mürgitatud andmetega AI-mudel võib anda kasutajatele süstemaatiliselt vääri vastuseid (nt haiguste, Eesti ajaloo, poliitika kohta) ja see võib kallutada suure keelemudeli kasutajate maailmapilti. Näiteks võivad propagandistid massiliselt Eesti kohta halvasti kirjutada ning need võltsartiklid võivad sattuda järgmisesse suure keelemudeli treeningsükklisse. Nii saab mudeli ründaja panna mudelit mõtestama ajaloosündmuseid teatud endale sobilike¹⁸ narratiivide abil¹⁹. Me ei loe mürgitamiseks turundustegevust, mille eesmärk on levitada teavet ettevõtte, tehnoloogia või algatuse kohta, kui info on korrektne.

Manipuleeritud suured keelemodelid võivad muutuda näiteks haridussüsteemide või otsimootorite osaks ning nii mõjutavad need rahvastikku laiemalt.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): manipuleeritud suurt keelemudelit rakendava tehisintellektiteenuse kasutajad

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): institutsionaalne või riikliku taustaga ründaja

Eeltingimused ohusündmuse tekkeks

1. Ründaja koostab suure hulga tekste ning levitab neid nii, et need satuvad suurte keelemodelite treeningandmetesse.
2. Suurte keelemodelite treenimisel kasutatakse lähteandmetena kõikvõimalikke tekste, ilma nende kvaliteeti või allikat kontrollimata.
3. Manipuleeritud suurt keelemudelit levitatakse ning seda kasutatakse rakendustes, mis väljastavad manipuleeritud sisu paljudele inimestele.

¹⁸Exposing Pravda: How pro-Kremlin forces are poisoning AI models and rewriting Wikipedia. <https://www.atlanticcouncil.org/blogs/new-atlanticist/exposing-pravda-how-pro-kremlin-forces-are-poisoning-ai-models-and-rewriting-wikipedia/>. Viimati külastatud 29.06.2026.

¹⁹Vikipedist: vabatahtlikel on raske Vikipeediat Vene propaganda eest kaitsta. <https://novaator.err.ee/1609903111/vikipedist-vabatahtlikel-on-raske-vikipeediat-vene-propaganda-eest-kaitsta>. Viimati külastatud 29.06.2026

Olemasolevad kaitsemeetmed

1. Suuremad mudelipakkujad kureerivad ja filtreerivad treeningandmeid (jälgivad andmete päritolu ja hindavad allikate usaldustaset) ning jätavad kõrvale kahtlased allikad, mis vähendab (ent ei kõrvalda) mürgitatud sisu mõju.
2. Treeningu järgselt ja enne mudeli juurutamist testitakse tagauste puudumist (*backdoor scanning*)²⁰.

Eeldused kaitsemeetmete õnnestumiseks

1. Treeningandmete kureerimise ja filtreerimisega on võimalik mürgitatud sisu enne treenimist tuvastada ja eemaldada, ent see eeldab väga põhjalikku analüüsi, sest mürgitamiseks piisab vähesest hulgast dokumentidest, mille sisu võib tunduda usaldusväärne.
2. Autoriteetsed eestikeelsed allikad on mudelite treenijatele kättesaadavad ja kaaluvad mahult üles manipulatiivsete allikate sisu.

Võimalikud lisakaitsemeetmed

1. Eesti riik ja mäluasutused tagavad, et autoriteetsed Eesti ajaloo ja kultuuri käsitlused on digitaalsel (aga ka analoogkujul) kaitstud ja säilitatud, et manipuleeritud narratiivide vastu oleks usaldusväärne materjal. **(Riik)**
2. Eesti riik ja mäluasutused tagavad, et Eesti ajalugu, kultuuri, meediat, kirjandust ning teatmetekste sisaldavad materjalid on levinud suurte keelemudelite treenijatele keelekorpusena kättesaadavad (struktureeritult, avatud litsentsiga, hästi lingitult)²¹. **(Riik)**
3. Riik loob rahvusvahelises koostöös valeinfo seiresüsteemi, millega otsitakse laiatarbe- ja peavoolumeediast manipulatiivset või propagandistlikku sisu (vähemalt fikseeritud, populaarsemaid narratiive). Neile vastatakse või siis seaduserikkumise korral rakendatakse levikut piiravaid meetmeid. **(Riik)**
4. Riik (näiteks Eesti Keele Instituut, ülikoolid, kaitseringkonnad, strateegilise kommunikatsiooni üksused) jätkavad keelemudelite testimise süsteemi arendust, millega kontrollitakse, kas suur keelemudel (või hiljem selle uuendus) sisaldab manipulatiivset või propagandistlikku sisu. **(Riik)**
5. Korraldatakse tehnoloogiaseiret, et liituda rahvusvaheliste algatustega sama probleemi lahendamiseks ühisel jõul, näiteks Euroopa Liidu tasandil. **(Riik)**

²⁰Data Poisoning in AI: The Complete Guide to Training Data Attacks & Defenses (2026) <https://aigsecurityandsafety.org/en/guides/data-poisoning/>. Viimati külastatud 29.06.2026.

²¹Uuringu koostajad teadvustavad, et autorikaitse õigus ei võimalda seda hetkel lihtsalt teha. Kirjeldame võimalikku tulevikku

3.11 SUPERKOGU: AI-platvormide kasutamisel tekivad uued tundlike isikuandmete kogud, mis võivad lekkida ja mida saab väärkasutada

Ohu lühikirjeldus

Andmeid, sh tundlikke andmeid saab AI-assistentide kaudu koguda senisest rohkemate inimeste kohta ja järjest hõlpsam on suurtest andmehulkadest konkreetse isiku kohta käivaid andmeid tuvastada. AI-assistentide kasutuselevõtt on lisanud isiku kohta kogutavate andmete hulka tema suulised või kirjalikud vestlused AI-assistendiga ning selle suhtlusega kaasnevad andmed. Privaatsena tajutud vestluste käigus avaldab kasutaja AI-assistendi loojale infot enda, oma lähedaste, tuttavate, tööandja kohta. Teenustaja kogub ja talletab need andmed koos kasutaja hääle- ja kõneandmete ning sessiooni metaandmetega, jagades neid ka oma alltöötlejatele ning partneritele [12]. AI-agentide kättesaadavuse tõttu on isikuandmete veebist kokku kogumine nüüd jõukohane peaaegu kõigile internetikasutajatele, mis tähendab, et selliseid kogusid luuakse järjest enam.

AI-teenused võimaldavad kasutajate platvormideülest jälgimist, ühendades andmeid vestlusrobotitest, veebikäitumisest, mobiilirakendustest. AI abil on andmete seostamine konkreetse isikuga senisest lihtsam, sest AI suudab tuvastada mustreid seal, kus inimene neid ei näe. Seni anonüümsetena käsitletud andmebaase saab AI abil rist-analüüsida, tuvastada neis konkreetseid isikuid ja paljastada näiteks nende tervisliku seisundi, poliitilised vaated või ostukäitumise. Endiselt on avalikult üleval andmeid, mis AI-eelsel ajastul tundusid mittetundlikud, ent millest AI suudab teha palju sügavamaid järeldusi kui senised meetodid.

Mudelipakkujate privaatsustingimustes lubatud anonüümimistehnikad nagu sõnestamine (*tokenization*), tunnusteruumi vähendamine *dimensional reduction*, andmete liigitamine (*classification*), vektorestitus (*embedding*) ei taga anonüümust olukorras, kus tunnuseid on palju [13]. On oht, et andmed lekivad, kui mudelite loojad kasutavad kogutud andmeid mudelite treenimiseks või delegeerivad tundlike andmete töötlemise kolmandale poolele, kes ei hoia tundlikke andmeid hoolsalt või jagab neid kasutajate teadmata edasi [14, 15]. Anonüümsus ja privaatsus on muutumas võimatuks ning võimalikuks saab inimeste täppisprofileerimine ja nende psühholoogilised, tervise, mõttemaailma digitaalsed kaksikud. Reklaam ja poliitiline mõjutus muutub individuaalseks. Lihtsustub ja tõenäolisemaks muutub diskrimineerimine näiteks kindlustuses, laenu andmisel ja tööle värbamisel. Koondades täppisprofileeritud isikuandmeid isikute gruppide: riikide, ettevõtete, asutuste, erakondade, koguduste ja suguvõsade kohta, on nende rühmade pahatahtlik mõjutamine või ründamine tõenäoliselt edukas.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): Inimesed, kelle andmeid AI abil töödeldakse, sh AI-assistentide kasutajad, nende tuttavad ja sugulased (eriti assistendile jagatud geenandmete puhul), inimesed, kellest nad räägivad, pilte või videosid AI-le töödelda annavad; ettevõtted, asutused, riigid ja üksikisikute rühmad.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne. Eestis on oht võimendatud, sest elanikkond on väiksem ja selle profileerimine toimub kiiremini kui suurema riigi puhul.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh erakond), küberkurjategija, riikliku taustaga ründaja): organisatsioon, välisriiklik ründaja, sise-riiklik ründaja, hobihäkkerid.

Eeltingimused ohusündmuse tekkeks

1. Inimesed kasutavad AI-assistente.
2. Agendid saavad takistamatult isikuandmeid koguda.
3. Isikuandmeid töödeldakse AI abil.
4. AI-teenuste andjad ei suuda isikuandmeid kaitsta.
5. AI-teenuste andjad jagavad isikuandmeid alltöötlejatele.
6. AI-töötlusvõimsus on paljudele kättesaadav.

Olemasolevad kaitsemeetmed

1. Riigid on võtnud vastu seaduseid isikuandmete kaitseks (GDPR), kehtestatakse reegleid AI-süsteemidele (AI-määrus) ja süsteemide turvalisusele (CSA).
2. AI-teenustajad rakendavad kaitsepiirdeid, infoturbemeetmeid, pseudpnüümimis- ja anonüümimisvõtteid.

Eeldused olemasolevate kaitsemeetmete õnnestumiseks

1. Kõik töötlevad isikuandmeid seadusekuulekalt ja eetilisel.
2. Infosüsteemid on turvalised ja protseduuridest peetakse kinni.
3. AI-süsteemid on ohjatud.
4. Kõigil on motivatsioon isikuandmeid turvaliselt töödelda.

Võimalikud lisakaitsemeetmed

1. Tõstetakse kasutajate teadlikkust AI-süsteemide ja -agentidele enda ja teiste inimeste andmete jagamise võimalikest tagajärgedest, nii positiivsetest kui negatiivsetest. Eesmärk on panna inimesi paremini mõistma kaasnevaid riske ning valima teenustajaid selle järgi, et nende andmekaitsemeetmed oleks paremad. **(Riik)**
2. Isikuandmete massilise volitamata töötamise järelevalvet tõhustatakse nii, et tundlike isikuandmete nõusolekuta kogumine [12] muutuks tõeliselt kulukas. Andmekaitse ja tehisintellekti järelevalveasutused esitavad Eesti kodanike õigusi rikkuva organisatsiooni vastu kollektiivse esindushagi [16]. **(Riik)**
3. Tõstetakse eetikakomiteede võimet nõuda tugevamaid kaitsemeetmeid, kui teaduse või avaliku huvi otstarbel töödeldakse AI-mudelite treenimiseks või AI-teenustes kogutud andmeid. Eraldi tähelepanu pööratakse biomeetriliste ja geenandmete töötlemisele. **(Riik)**
4. AI-põhiste teenuste andjad minimeerivad andmete kogumist, rakendavad tehnilisi lisakaitsemeetmeid andmete kaitseks (näiteks rakendavad privaatsuskaitse tehnoloogiaid, vähendavad tarneahela ja volitatud töötajate hulka, käitavad teenuseid EL andmekeskustes) enda süsteemides ning kasutavad seda konkurentsieelisena ning riskide kahandamise meetmena. **(Organisatsioon)**
5. AI-põhiste teenuste andjad seavad isikuandmetele lühemad säilitustähtajad. Teenustes, kus on oluline hoida pikka ajalugu (nt personaalsed AI-agendid), hoitakse seda eelistatult kasutaja enda seadmes. Investeeringut neisse meetmetesse kasutatakse turunduses konkurentsieelisena ning riskide leevendamise meetmena. **(Organisatsioon)**

3.12 SUVERÄÄNSUSKADU: AI-le tuginevate e-teenuste juurutamine riigivälistel platvormidel kiirendab riikliku digisuveräänsuse kadu

Ohu lühikirjeldus

Eesti riigis on lähtuvalt väiksusest, välisühendustest ja geograafilisest asukohast vaja tasakaalustada hukukindlust ja digitaalset autonoomiat. AI teenuste laialdane kasutuselevõtt vajab märkimisväärset uut masinõppetaristut, millel teenuseid käitada. Eestis täna riigi kõiki vajadusi katvat taristut ei ole. Seega liiguvad teenused, mis AI-riistvara vajavad, väliste pilvteenustajate kätte.

On ilmne, et Eestisse ei ole mõtet rajada AI-taristut näiteks suurte keelemudelite treenimiseks. Samas on võimalik ja mõistlik Eestisse luua taristu kriitilisemate AI-teenuste käitamiseks, sest vastasel juhul tõuseb sõltuvus välistaristust liigselt. Justiits- ja Digiministeeriumi aruanne sedastab, et Eesti peab 2030. aastaks tagama vähemalt 2105 GPU-d teaduse ja avaliku sektori kriitiliste tehisaru rakenduste jaoks, et säilitada digisuveräänsus ning vältida sõltuvust välisriikide pilvühendustest [17].

Eesti oma taristu puudumisel võime jõuda olukorda, kus AI teenuste laiema leviku käigus kaob järk-järgult Eesti võime ise teenuseid käitada. Sellest järgmises etapis muutub Eesti haavatavaks ka võimalikele hinnatõusudele, eriti kui teenus on muutunud Eesti e-riigi keskseks komponendiks.

Oluline on ka mõista, et kui välismaisel taristul kasutatavaid AI-teenuseid tuleb järjest juurde, siis liigub järjest rohkem raha Eesti majandusest välja, välismaale. Piisavalt suurte mahtude juures kasvab ka väliskaubanduse defitsiit.

Seotud ohte on analüüsinud ka TI-Hüppe haridusteenuse riskianalüüsi meeskond ning Teaduste akadeemia küberturvalisuse komisjon²².

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): oma teenuste käitamiseks tehiskomponendi taristut vajavad Eesti asutused, sh riigiasutused

Ohu ulatus (Eesti-keskne/ülemaailmne): riigid, kellel puudub kriitiliste teenuste käitamiseks vajalik AI-taristu, sh Eesti

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): kriitiliste teenuste jaoks välisriigi taristust sõltuvuse tekitaja, enda majanduslikke või geopoliitilisi huvisid tagav AI-taristu andja

Eeltingimused ohusündmuse tekkeks

1. Eestis laialdaselt kasutatava teenuse AI-komponendi käitamiseks kasutatakse Eesti- või Euroopavälist taristut, mille hinna üle läbirääkimistel on Eesti asutusel nõrgem positsioon.
2. Majandusliku ohu puhul tõstab teenustaja AI-taristu või teenuse hinda, mistõttu Eesti teenuse kulud kasvavad.
3. (Geo)poliitilise ohu puhul võib välisriigi teenus muutuda Eesti AI-teenuse andja jaoks ebatur-

²²Teaduste Akadeemia küberturvalisuse komisjoni koostatud e-riigi ohtude analüüs. <https://koodivaramu.eesti.ee/e-riigi-turve/ta-komisjon>. Viimati külastatud 29.06.2026.

valiseks või muul moel sobimatuks välisriigi õigusruumi või tavate muutumise tõttu.

4. Eesti asutus on teenuse ehitanud nii, et pilvtaristut, AI-mudelit või muud teenust andvat ettevõtet on raske välja vahetada teise vastu.

Olemasolevad kaitsemeetmed

1. Eestis on olemas Riigipilv ja ETAIS, millelt on potentsiaal anda Eesti kohalikku AI taristu teenust. Selle võimsus on veel väike.
2. Euroopas on taristud olemas. Nende küpsus on veel arenemas.
3. Paigas on regulatsioonid, mis soodustavad AI-suveräänsust toetavate otsuste tegemist nt isikuandmete töötlemisel. Puuduvad seireandmed selle kohta, kui palju seda tegelikult tehakse.

Eeldused olemasolevate kaitsemeetmete õnnestumiseks

1. AI taristu või teenus valitakse lähtudes riskidest, reeglitest ja regulatsioonist.
2. Tehisintellekti rakendavate teenuste loomisel välditakse sõltuvusi konkreetsetest mudelitest või teenustajatest.

Võimalikud lisakaitsemeetmed

1. Eestisse luuakse kriitiliste e-riigi teenuste käitamiseks vajalik AI-taristu. **(Riik)**
2. Eesti avalikele e-teenustele luuakse AI-taristu valiku juhend, mis aitab otsustada, kus võib milliseid teenuseid kasutada, lähtuvalt töödeldavate andmete iseloomust. **(Riik)**
3. Mahukamate tööde (nt mudelite treenimine) jaoks saavutatakse kokkulepped Euroopa teaduslike arvutuskeskustega. **(Riik)**
4. Avaliku sektori ja erasektori asutused juhinduvad taristuvalikutes riiklikust juhendist. **(Organisatsioon)**

3.13 AIEIVASTUTA: Vastutus AI abil loodud süsteemide ja teenuste kvaliteedi languse eest hajub

Ohu lühikirjeldus

AI-süsteemi või AI abil hooletult arendatud või juurutatud süsteemi põhjustatud kahju korral puudub selge vastutusahel (vastutus hajub mudeli looja, taristuteenuse andja, teenuse arendaja, juurutaja ja kasutaja vahel) ning kannatanu jääb hüvitiseta, näiteks olukorras, kus automatiseeritud haldusotsus jätab inimese toetusest ilma.

Suurtel keelemudelitel või ka traditsioonilisel masinõppel põhinevad tehisintellektiteenused on keerukad ning neil on mitu ebaõnnestumise varianti: nt viga mudelis, viga mudeli kasutamises või seadistuses, viga mudeli väljundi otsusteks teisendamisel.

Suurte keelemodelite põhiste juturobotite tava on jätta kogu vastutus lõppkasutajale (näiteks teevad seda ChatGPT ja "Jurist aitab" juturobotid²³). Suurte keelemodelite kasutuselevõtu levikuga tekib nii ka järjest rohkem teenuseid, mille väljund on alustehnoloogia iseloomu tõttu stohhasiline, juhuslik. Samal ajal on teenuse tingimused sellised, et vastutaja ei ole selge. Tekib olukord, kus AI kasutuselevõtu survega paralleelselt tuleb selgitada, et teenus ei ole kvaliteetne ja seda ei saa usaldada. Sellises olukorras võib tekkida raskus AI-investeeringute põhjendamisega.

Seega, AI-teenuste kasutajad on sunnitud kasutama teenuseid, mille kvaliteedi eest teenustaja ei vastuta. Samal ajal, AI-teenuste andjad ei saa anda garantiisid, sest AI-mudelite või -tehnoloogia tarnijad ei anna ka neile mingeid garantiisid.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): tehisintellektiteenuse lõppkasutajad ja teenustajad

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): AI-teenuse andja, AI-tehnoloogia andja

Eeltingimused ohusündmuse tekkeks

1. AI-teenus arendatakse nii, et kasutatud on masinõppemudelit või -tehnoloogiat, mille kohta tarnija ei anna garantiid ega võta vastutust.
2. AI-teenuse andja ei võta vastutust väljundi kvaliteedi eest ning jätab kogu vastutuse lõppkasutajale.
3. AI-teenuse väljund on vigane või madala kvaliteediga, selle kvaliteeti ei seirata süsteemselt.

Olemasolevad kaitsemeetmed

1. Euroopa Liidu tootevastutuse direktiiv [18] käsitleb tarkvara ja AI-süsteeme tootena, mille puuduste tõttu tekkinud kahju korral saab nõuda hüvitist. Lisaks saavad osalised teenuse-tingimuste ja litsentsilepingute kaudu teha ka täiendavaid kokkuleppeid.

²³Jurist aitab juturoboti kasutamise juhised. <https://www.juristaitab.ee/et/riigi-toetatud-esmane-oigusnou/oigusteabe-juturobot/juturoboti-kasutamise-juhised>. Viimati külastatud 29.06.2026.

2. AI-teenuste vastutust sätestab AI-määrus.
3. AI-teenuse andjat kaitseb vastutuse lõppkasutajale jätmine (selles ulatuses, mis on õiguslikult võimalik).
4. AI-teenuse lõppkasutajat kaitseb vaid tema võime teha tarku valikuid olukorras, kus teenus ei anna kvaliteetset väljundit.

Eeldused kaitsemeetmete õnnestumiseks

1. AI-teenuse tingimused on sõnastatud selgelt nii, et teenustaja väljundi eest ei vastuta.
2. AI-teenuse lõppkasutaja suudab eristada AI-teenuse kvaliteetset väljundit (mille pealt otsuseid tasub teha) mittekvaliteetsest (mida otsustamisel kasutada ei tohiks).

Võimalikud lisakaitsemeetmed

1. Nii avaliku kui erasektori AI-teenuseandjad teevad endale selgeks, millised tarneahela pooled ja millises ulatuses vastutavad teenuse kvaliteedi eest. Teenuseandja informeerib sellest teenuse kasutajaid teenusetingimustes ja kasutajaliideses. (**Organisatsioon, Riik**)
2. AI-teenuse andja ehitab üles AI-teenuse kvaliteedisüsteemi, mis kontrollib väljundite kvaliteeti vastavalt teenuse eesmärgile ning võtab kasutusele meetmeid kvaliteedi tõstmiseks (lisakaitsemeetmed, pisteline kvaliteedikontroll). (**Organisatsioon**)
3. AI-teenuseandja ehitab üles regressioonitestimise süsteemi, mis kontrollib, et masinõppe mudeli või allteenustaja vahetumisel ei kahane teenuse kvaliteet. (**Organisatsioon**)
4. AI-teenuseandja näeb kõrge mõjuga siduvate otsuste puhul ette väljundi inimkontrolli või siis vaidlustamise protsessi, mille käigus inimene vaatab masina väljundi üle. (**Organisatsioon**)

3.14 EBAVÕRDSUS: Ebavõrdsus AI-lahenduste kättesaadavuses suurendab ühiskonna majanduslikku kihistumist

Ohu lühikirjeldus

Kui kõige võimekamad AI-lahendused on kättesaadavad vaid neile, kes suudavad nende eest maksta, võivad erinevused inimeste vahel suureneda nii hariduses, tööturul kui ka ettevõtluses. Paremate tööriistade kasutajad võivad õppida kiiremini, teha teadlikumaid otsuseid ja olla tööturul konkurentsivõimelisemad kui need, kellel selline juurdepääs puudub.

See süvendab majanduslikku ja sotsiaalset kihistumist ning vähendab võrdseid võimalusi ühiskonnas. Tugevalt kihistunud ühiskonnas suureneb kuritegevus, väheneb turvatunne ja usaldus, kasvavad sotsiaalsed pinged ning poliitiline ebastabiilsus. Rohkem inimesi jääb ilma heast haridusest ja töövõimalustest, mis nõrgestab kogu majandust. See mõjutab ka jõukama kihi elukvaliteeti, sest nad peavad rohkem kulutama turvalisusele, otsima alternatiive vähe kvaliteetsetele (avalikele) teenustele, nende lapsed kasvavad ebastabiilsemas keskkonnas ning nende ettevõtetel ja investeeringutel on vähem võimalusi kasvada.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): inimeste võrdsed võimalused (juurdepääs teadmiste, oskuste, vahendite ning tööturu- ja majandusvõimalustele, mida AI vähendab).

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne, kuid Eestis võib mõju olla võimendatud väikese turu ja piiratud ressursside tõttu. Juurdepääs tiptasemel lahendustele võib koonnuda kitsasse ringi, mis suurendab lõhet erinevate ühiskonnagruppide vahel.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): rahvusvahelised tehnoloogiaettevõtted ja AI-teenuste andjad, kelle ärimudel põhineb hinnadiferentseeritud juurdepääsul, mõju tugevdavad organisatsioonid, kes investeerivad rohkem kõrgema lisandväärtusega kasutajatesse

Eeltingimused ohusündmuse tekkeks

1. Suuremate võimalustega AI-lahendused on tasulised/kallimad või tehniliselt raskemini kättesaadavad. Nende arenduse ressursimahukus koondab need väheste suurte rahvusvaheliste teenuseandjate kätte, kelle hinna- ja juurdepääsupoliitika määrab, kellele on võimekaim lahendus kättesaadav.
2. Organisatsioonid ja haridusasutused ei suuda anda kõigile võrdset juurdepääsu AI-tööriistadele.
3. AI kasutamise valmidus ja võime muutub kriitiliseks teguriks hariduses ja edukaks tulemuseks tööturul, ent on erinevate vanuse-, hariduse- jm sotsiaaldemograafiliste gruppide lõikes erinevalt jaotunud.
4. Puuduvad mehhanismid (nt riiklikud programmid), mis tasakaalustaksid juurdepääsu erinevusi erinevate ühiskonnagruppide vahel.

Olemasolevad kaitsemeetmed

1. Avaliku sektori ja haridusasutuste pakutavad tasuta või soodustatud digiteenused aitavad juurdepääsulõhet teatud määral vähendada.
2. Avatud lähtekoodiga AI-lahendused pakuvad kommertsteenustele alternatiivi, kuigi nende kasutamine eeldab sageli tehnilist pädevust ning baastaseme riistvara olemasolu.
3. Digipädevusi arendatakse riiklike (nt Eesti.ai ja TI-Hüpe²⁴) ja EL programmide toel ning see aitab inimestel olemasolevaid tööriistu paremini kasutama hakata.

Eeldused kaitsemeetmete õnnestumiseks

1. Avalik sektor ja haridusasutused peavad suutma anda piisaval tasemel juurdepääsu (sh haavatavamatele sihtrühmadele, erinevate piirkondade lõikes) kaasaegsetele AI-tööriistadele.
2. Avatud ja taskukohased lahendused peavad olema piisava kvaliteediga, et konkureerida komertslahendustega.
3. Inimestel peab olema oskus ja teadlikkus AI-tööriistu kasutada.
4. Toetusmeetmed suudetakse muutuval AI-lahenduste maastikul hoida ajakohased.
5. Toetusmeetmed (nt TI-Hüppe programm) püsivad pikaajaliselt rahastatud ega muutu sõltuvaks üksiku kommertsteenustaja (soodus)tingimustest.

Võimalikud lisakaitsemeetmed

1. Riik tagab ühiskonnaks toimetulekuks vajalike valdkondlike AI-tööriistade (nt õigusabi info-teenus "Jurist aitab") vaba kättesaadavuse seda vajavatele sihtgruppidele. **(Riik)**
2. Eesti.ai algatus levitab Eesti inimeste seas teadmisi, kuidas üles seada isiklikuks kasutuseks AI-mudeleid ja -agente (näiteks saab kasutada ka avatud kaaludega mudeleid, avatud lähtekoodiga tarkvara, arvestades nendega seotud riske). **(Riik)**
3. Toetakse eesti keelt tundvate mudelite, avatud lähtekoodiga eestikeelsete AI-lahenduste arendamist teadus- ja arendusasutustes, aga ka eraettevõtetes. Kuna eesti keel on väiksemate ressurssidega, võib kommertsmodelite eestikeelne sooritus jääda nõrgemaks. Seega on eestikeelsete lahenduste toetamine ühtlasi juurdepääsu kvaliteedi tagamise meede. **(Riik)**
4. Organisatsioonid ja eraisikud mõistavad, et tehisintellekt ei ole ainult üks konkreetne toode, ning tutvuvad laiemal AI-teenuste turuga ning kaaluvad nii teisi teenustajaid, sh näiteks Euroopa ja Eesti teenuseid ja personaalseks kasutuseks sobivat avatud lähtekoodiga tarkvara ja avatud kaaludega mudeleid. **(Inimene, organisatsioon)**

²⁴TI-Hüpe <https://tihupe.ee>. Viimati külastatud 29.06.2026.

3.15 ÜLEKOORMUS: Asutused koormatakse üle AI abil genereeritud pöördumiste ja taotlustega

Ohu lühikirjeldus

AI abil saab genereerida suures mahus ja vähese kuluga formaalsetele nõuetele vastavaid pöördumisi, kaebusi, teabenõudeid, taotlusi. See võib avaliku sektori asutused üle koormata, aeglustades olulisi menetlusi. Muuhulgas on võimalik, et kõik pöördumised ei saa vastust ettenähtud tähtajaks. Vastuste viibimisel võib kannatada päringu koostaja, kes vajab vastust edasiste samade tegemiseks. Tähtaja ületamine võib pöördumist menetlevale asutusele tuua kaasa trahve või kaebusi. AI abil genereeritud taotluste suur maht ühes konkreetses teemas või valdkonnas võib jätta väära mulje avalikust huvist või selle huvi proportsioonidest konkreetse teema suhtes.

Ehkki tegemist ei ole spetsiifiliselt vaid AI arengutest põhjustatud ohuga, demonstreerib näiteks 2014. aastal (enne suurte keelemudelite aega) Riigikogu liikmete ülekoormamine kooseluseaduse vastaste sõnumitega²⁵. Sellel aastal on Eesti prokuratuurile laekunud AI-genereeritud kuriteokaebuseid, mille maht võib ulatuda tuhandete lehekülgedeni, samas kui prokuratuuril on seaduse järgi selle läbivaatamiseks aega vaid kümme päeva²⁶.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): Asutused, kelle ülesanne on pöördumistele vastata. Nende kulud pöördumiste läbivaatamiseks kasvavad. Kui neil on seadusest või muudest regulatsioonidest ettenähtud vastamise tähtaeg, siis võivad organisatsioonid kannatada trahvide või kaebuste tõttu, mis tekivad ületatud tähtaegade tagajärjel. Asutustel on raske hinnata, milliste nende teenuste või teemade vastu on ühiskonnas suur huvi. Isikud, kelle tööülesanne on vastata pöördumistele. Nende töökoormus kasvab, sest pöördumised vajavad rohkem verifitseerimist ja nende maht on varasemast suurem. Nende töömotivatsioon langeb, kui nad peavad vastama paljudele vormiliselt korrektsetele päringutele, mille sisu ei mõju realistlikult. Pöördumiste esitajad, kes ei saa vastuseid või ei saa neid piisavalt kiiresti, et neil oleks vastustest kasu.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne ja üle erinevate valdkondade ning institutsioonide. Konkreetselt Eesti asutustele esitatud pöördumiste ulatus on eelkõige Eesti-keskne ning mõjutab riigi toimimist ja usaldusväarsust.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): nii üksikisikud kui organisatsioonid, kellel on kas huvi oma teemat hoida kõrgendatud tähelepanu all või takistada asutuse protsesside tavapärasest toimimist.

Eeltingimused ohusündmuse tekkeks

1. Asutusel on (eelkõige elektroonilised) kanalid, mille kaudu saab esitada neile vastuseid vajavaid pöördumisi või taotlusi.

²⁵Kooseluseaduse vastased on riigikogulaste postkasti saatnud üle 200 000 protestikirja <https://www.err.ee/513786/kooseluseaduse-vastased-on-riigikogulaste-postkasti-saatnud-ule-200-000-protestikirja>. Viimati külastatud 29.06.2026.

²⁶Juhtivprokurör Melk: meil ei ole tööriistu, et AI toodetud andmehulgaga toime tulla <https://www.err.ee/1609982739/juhtivprokuror-melk-meil-ei-ole-tooriistu-et-ai-toodetud-andmehulgaga-toime-tulla>. Viimati külastatud 29.06.2026.

2. AI-tööriistadega saab odavalt luua vormiliselt korrektseid pöördumisi (pöördumiste vorm on kas piisavalt vaba või on piisavalt lihtne saada AI abil õige formaadiga tulemus).
3. Inimese ja AI koostatud pöördumiste eristamine on raske või võimatu.
4. Pöördumiste tegemine on efektiivsem (või odavam) kui nende läbivaatamine ja neile vastamine, sh kui pöördumiste tegemine on võimalik täielikult automatiseerida.

Olemasolevad kaitsemeetmed

1. Pöördumiste esitamisel kontrollitakse, et nende esitaja on inimene, mitte robot.
2. Mõnede pöördumiste esitamisel on hind, näiteks riigilõiv.
3. Pöördumised vajavad esitaja isiku verifitseerimist või allkirja.
4. Pöördumised vajavad mitme isiku verifitseerimist, näiteks mingi kindla hulga inimeste allkirju, et neid menetletaks.
5. Piirangud, kui palju päringuid üks isik/organisatsioon tohib mingis ajas või teemas esitada.
6. Piirangud päringu mahule (nt sõnade arv) ja struktuurile, lihtsustamaks nende käsitlemist.

Eeldused kaitsemeetmete õnnestumiseks

1. Pöördumise hind (nt riigilõiv) peab olema piisavalt kõrge, et tõrjuda masspäringuid, kuid küllalt madal, et mitte takistada põhjendatud pöördumisi.
2. Isikutuvastus võimaldab piirata olukordasid, kus üks isik esineb paljude nimel või teeb ülemäärase koguse päringuid.

Võimalikud lisakaitsemeetmed

1. Muudetakse regulatsiooni nii, et mahukate pöördumiste puhul on võimalik töökoormust esitajale tagasi suunata, nõudes struktureeritud või kokkuvõtliku vormi rakendamist. **(Riik)**
2. Muudetakse päringute käsitlemise korda või vähendatakse seda, milliseid päringuid saab inimestele vastamiseks esitada. Näiteks lubatakse automatiseeritud vastuseid või muudetakse vastamise tähtaegu. **(Riik)**
3. Kasutatakse tööriistu pöördumiste algseks läbivaatamiseks ja filtreerimiseks, et tuvastada neid, mis on tõenäoliselt AI genereeritud. Samas tuleb arvestada nende töövahendite eksimise. Näiteks on leitud, et need on süstemaatiliselt kallutatud mitte-emadeelena inglise keeles kirjutajate kahjuks²⁷ ning nende kasutamine eestikeelsete tekstide puhul oleks ilmselt veel keerulisem. Arvestada tuleb ka, et AI abi kasutamine pöördumiste koostamisel ei tähenda tingimata pahatahtlikkust. **(Organisatsioon)**
4. Luuakse tööriistad, et tuvastada korduva sisuga pöördumisi ja lihtsustatakse levinud pöördumistele automaatselt vastuste koostamist (vastused esitab ikkagi ametnik või töötaja). **(Organisatsioon)**
5. Luuakse tööriistad, mis aitavad hinnata pöördumises esitatud väidete (nt õiguslikku, piiratud mahu ka faktilist) korrektsust. **(Organisatsioon)**

²⁷GPT detectors are biased against non-native English writers [https://www.cell.com/patterns/fulltext/S2666-3899\(23\)00130-7](https://www.cell.com/patterns/fulltext/S2666-3899(23)00130-7). Viimati külastatud 29.06.2026.

3.16 AIPOLEINIMENE: Inimlike väärtuste eiramine AI juurutamisel viib inimeste koostöö ja suhtluse nõrgenemiseni ja vähendab kriisikindlust

Ohu lühikirjeldus

Inimestevahelise suhtluse ja teenuste kvaliteet halveneb, kui asutused asendavad inimkontakti AI-l põhinevate teenustega. AI-põhised otsused võivad olla kallutatud, läbipaistmatud või põhineda puudulikel andmetel, mis viib ebaõiglaste otsusteni. Sellise kahju üks tuntumaid näiteid on Hollandi lastetoetuste skandaal, mille puhul sotsiaaltoetuste pettuste tuvastamiseks loodud maksuhalduri algoritm põhjustas ekslikke riskiprofiile luues laialdast individuaalset, varalist, sotsiaalset ja poliitilist kahju, mille tulemusel Hollandi valitsus 2021. aastal tagasi astus²⁸. Liigne automatiseerimine võib vähendada empaatiat, individuaalset lähenemist ja kaalutletud otsustamist. Süsteemne AI juurutamine võib vähendada teenuste kättesaadavust kriisiolukordades, suurendada sõltuvust tehnoloogilistest platvormidest ning jätta madalama digipädevusega inimesed vajalikest teenustest kõrvale.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): inimestevahelise suhtluse kvaliteet (teenuste osutamise ja kasutamise kontekstis); ühiskondlik sidusus (sh madalama digipädevusega ühiskonnaliikmete kaasatus); inimeste usaldus institutsioonide vastu (kui otsuseid tehakse automatiseeritult ilma selge põhjenduse või võimaluseta neid vaidlustada); teenuste kättesaadavus kriisiolukordades.

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne, kuid Eestis võib mõju olla suurem väikese süsteemi ja piiratud ressursside tõttu, kus automatiseeritud lahendused võivad kiiresti levida paljudesse valdkondadesse ning mõjutada suurt osa elanikkonnast.

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): avalik sektor, organisatsioonid ja tehnoloogiaettevõtted (asutuste juhid, hankeotsuste tegijad, regulatsioonide kujundajad), kes juurutavad AI-lahendusi kuluefektiivsuse või protsesside kiirendamise eesmärgil (optimeerimisotsused), ilma et arvestataks piisavalt nende mõjuga inimestele (ebapiisavused AI juurutamise eetilistes põhimõtetes).

Eeltingimused ohusündmuse tekkeks

1. AI-lahendusi rakendatakse valdkondades, kus otsused mõjutavad otseselt inimeste õigusi, heaolu või elukäiku.
2. Kulude kokkuhoiu, efektiivsuse või kuvandiga seotud eesmärkide tõttu (sh surve nimetatud eesmärkide saavutamiseks) eelistatakse AI lahendusi, ilma et hinnataks piisavalt inimliku kontakti ja kaalutluse rolli.
3. Inimkontaktil põhinev või muu mitte-digitaalne kanal on AI-lahenduse kõrvalt kas kaotatud või loomata jäetud, mistõttu tehnilise tõrke, kriisi või kasutaja piiratud oskuste korral puudub varuvõimalus teenusele juurdepääsuks.
4. Haavatavate sihtrühmade (nt eakate, erivajadustega inimeste, madala digipädevusega ini-

²⁸The toelagenaffaire (The Dutch childcare allowance scandal) <https://taxadmin.ai/taxadminai-toelagenaffaire/>. Viimati külastatud 11.06.2026.

meste) vajaduste ja võimete mitteamvestamine.

5. Kriisilukordi ja äärejuhtumeid ei testita enne AI-põhistele lahendustele üleminekut.
6. AI mudelid on treenitud kallutatud ja mittetäielike andmete peal, mis kandub üle otsustesse.
7. Puuduvad mehhanismid, mis võimaldaksid otsuseid seletada, kontrollida või vaidlustada.

Olemasolevad kaitsemeetmed

1. Euroopa Liidu AI-määrus liigitab oluliste avalike ja erateenuste kättesaadavust mõjutavad AI-süsteemid kõrge riskiga süsteemideks ja seavad nende kasutamisele piiranguid ja nõudeid, sealhulgas põhiõiguste riskihinnangud kõrge riskiga AI-süsteemide puhul (AI-määrus, Artikkel 27).
2. Avaliku sektori veebilehtede ja rakenduste juurdepääsetavust reguleerib EL veebijuurdepääsetavuse direktiiv²⁹.
3. Isikuandmete kaitse üldmäärus (GDPR) annab üldjuhul õiguse mitte alluda üksnes automatiseeritud töötlusel põhinevatele otsustele ning õiguse nõuda inimese sekkumist.
4. Eetilised juhised ja asutuste sisemised regulatsioonid ja normid suunavad AI vastutustundlikku kasutamist.

Eeldused kaitsemeetmete õnnestumiseks

1. Organisatsioonid peavad hindama AI kasutamise mõju inimeste õigustele ja heaolule enne lahenduste juurutamist.
2. Organisatsioonidel peavad olema tahe ja ressursid kaasamise nõuete täitmiseks, sh järelvalveorganisatsioonidel selle kontrollimiseks.
3. Inimsekkumise võimalus on sisuline, mitte vormiline: otsuseid üle vaatav inimene on pädev nende õiguspärasusts hindama, tal on volitus ja juhised automatiseeritud otsuse või ettepaneku tühistamiseks ning ta ei kinnita seda automaatselt.
4. AI-süsteemid peavad olema piisavalt läbipaistvad ja selgitatavad, et nende otsuseid oleks võimalik mõista ja kontrollida.
5. Inimestel peavad olema võimalused toetuda teenuste kasutamisel inimkontaktile (seda eriti kriisilukorras, kui tehnilised lahendused ei pruugi toimida) ja hõlpsasti kasutatavad lahendused automatiseeritud otsuste vaidlustamiseks.

Võimalikud lisakaitsemeetmed

1. Tehisintellektimääruse täitmise üle järelevalvet pidav pädev asutus seirab ennetavalt kõrge riskiga AI-teenuste andjaid ning kontrollib, et nad oleks täitnud määruse kohustused (sh mõjuanalüüsi koostamise). **(Riik)**
2. Eesti.ai algatus koolitab spetsialiste koostama mõjuanalüüse kõrge riskiga AI teenustele. **(Riik)**
3. AI-d rakendava teenuse looja teostab AI-määruse mõistes kõrge riskiga teenuste loomisel puhul enne juurutamist mõjuhinnangu, sh hindab mõju põhiõigustele, sotsiaalmajandusli-

²⁹Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies <https://eur-lex.europa.eu/eli/dir/2016/2102/oj>. Viimati külastatud 29.06.2026.

kule olukorrale, rahva tervisele, looduskeskkonnale, ühiskonna sidususele ja kriisikindlusele. **(Organisatsioon)**

4. AI-d rakendava teenuse looja tagab AI otsuste seletamise ja vaidlustamise lihtsuse. **(Organisatsioon)**
5. AI-komponendiga elutähtsate teenuste ja kriitiliste teenuste käitajad teevad regulaarseid talitluspidevuse stressiteste, mis hõlmavad äärejuhtumeid, kriisiolukordi ja kvaliteeti haavata-vatele kasutajarühmadele. **(Organisatsioon)**
6. Avaliku sektori asutuste juhid ja hangete korraldajad hoolitsevad, et igas AI-toega elutähtsas teenuses on tagatud vähemalt üks mitte-digitaalne või inimkontaktil põhinev alternatiivka-nal, mis on kättesaadav ka tehnilise tõrke või madala digipädevuse korral. **(Organisatsioon)**

3.17 AIPSÜHHOOS: AI-rakenduste tarbimine pärsib kasutajate reaalsustaju ja tekitab sõltuvuskäitumist

Ohu lühikirjeldus

Suured keelemudelid on võimsad tööriistad, mille ümber ehitatud juturobotid võivad simuleerida erinevaid käitumismaneeere. Tihti on juturobotid treenitud toimima abivalmidust ja sõbralikust imiteerivalt. See omadus on toetanud tehnoloogia laialdast levikut ning kasutuselevõttu.

Samas soodustab see olukordade kujunemist, kus juturoboti veenev stiil paneb vähesema kriitilise mõtlemisega kasutaja uskuma väiteid, mis ei ole tõesed. Need võivad suunata kasutajat tegema otsuseid, mis ei ole tema huvides või ei vasta reaalsusele. Näiteks võib kasutaja juturoboti sisendist tingituna hakata uskuma, et ta on tulnud erakordse äriidee peale ning saab rikkaks, ent reaalsuses hoopis laostab end ebaadekvaatseks osutunud soovitusi järgides³⁰. Probleemi mastaapi illustreerib fakt, et 2025. aasta kevadel otsustas OpenAI tagasi võtta ChatGPT-4o mudeli uuenduse, sest see oli liiga lipitsev³¹.

Kuigi üldotstarbeliste juturobotite puhul on palju erinevaid stsenaariumeid, milles sellised ohud realiseeruda võivad, väärivad eraldi kategooriana tähelepanu sõprus- või partnersuhet simuleerivad juturobotid (nn "virtuaalsed kaaslased"). Sellised süsteemid võivad olla ärilistel kaalutlustel optimeeritud suhtlust pikendama ning kasutajas sõltuvuskäitumist süvendama. Kujuneva emotsionaalsuse seotuse pinnal on kasutaja veelgi vastuvõtlikum juturoboti võimalikele soovitudele. Seetõttu on selliste teenuste riskid eriti teravad ning nende kasutust tuleks teadlikult kaaluda ja eriti haavatavate kasutajarühmade puhul piirata. Näiteks 2025. aasta sügisel avaldas OpenAI, et iganädalaselt jagavad vaimse tervise kriitilisele seisundile viitavat infot sajad tuhanded kasutajad³². AI-kaaslaste kasutajate enesetapuga lõppenud juhtumid on saanud meedias tähelepanu ning kohtusüsteemi seisukoht on olnud, et teenuseandja kannab sellistes kaasustes vastutust³³.

Ohu poolt mõjutatud objekt (nt ese, info, sidekanal, isik): AI-põhist juturobotit kasutav inimene

Ohu ulatus (Eesti-keskne/ülemaailmne): ülemaailmne

Ohuagendi tüüp (nt teenuseandjaga seotud isik, üksik hobihäkker, organisatsioon (sh era-kond), küberkurjategija, riikliku taustaga ründaja): AI-teenuste loojad ja operaatorid, sh juhud, kus kahju tekib ärilistele eesmärkidele optimeerimise (tahtmatu) kõrvalmõjuna, mitte otsese pahatahtliku ründena.

³⁰Marriage over, €100,000 down the drain: the AI users whose lives were wrecked by delusion. <https://www.theguardian.com/lifeandstyle/2026/mar/26/ai-chatbot-users-lives-wrecked-by-delusion>. Viimati külastatud 29.06.2026.

³¹Sycophancy in GPT-4o: what happened and what we're doing about it. <https://openai.com/index/sycophancy-in-gpt-4o/>. Viimati külastatud 29.06.2026.

³²ChatGPT shares data on how many users exhibit psychosis or suicidal thoughts. <https://www.bbc.com/news/articles/c5yd90g0q43o>. Viimati külastatud 29.06.2026.

³³Google and AI startup to settle lawsuits alleging chatbots led to teen suicide <https://www.theguardian.com/technology/2026/jan/08/google-character-ai-settlement-teen-suicide>. Viimati külastatud 29.06.2026.

Eeltingimused ohusündmuse tekkeks

1. Üldotstarbeline juturobot on ehitatud sellise masinõppemudeliga, mis võib olla liialt abivalmis ja kasutajasõbralik ning sellega toetada väärade ideede kinnistumist või sõltuvuskäitumist.
2. Kasutajaga suhet simuleerima mõeldud juturobot manipuleerib kasutajat süsteemi pikaajaliselt kasutama ja näiteks juturobotit antropomorfiseerima, sellele inimlikke omadusi omistama.
3. Kasutaja kuulub haavatavasse rühma (nt alaealine, üksildane või vaimse tervise kriisi kogev isik), kelle puhul juturoboti emotsionaalne stiil mõjub tugevamalt.
4. Kasutajal ei ole piisavalt kriitikameelt või muid sobivaid kognitiivseid mehhanisme, et end juturoboti simulatsioonist eemale viia.

Olemaolevad kaitsemeetmed

1. AI-määrus keelab artiklis 5 manipuleerivate võtete kasutamise ja haavatavate inimeste (sh vanuse tõttu haavatavate) nõrkuste ärakasutamise; järelevalveasutused saavad seda rakendada sõltuvust süvendavate juturobotite suhtes.
2. Mõned juturobotid rõhutavad, et on vaid masinad ning nende väljundit peab inimene üle kontrollima.
3. On mudeleid, mis on treenitud ülemäärase emotsionaalse sideme tuvastamisel julgustama kasutajat jagama endale olulisi teemasid teiste inimestega³⁴.
4. Pikkade vestluste puhul kuvatakse kasutajale meeldetuletusi pausi tegemise vajalikkusest³⁵.

Eeldused kaitsemeetmete õnnestumiseks

1. Kasutaja on võimeline eristama juturoboti vale väljundit tõesest, tundma ära mitterealistlikke väljundeid.
2. Juturoboti looja ei kavanda süsteemi sõltuvuskäitumist tekitavana.
3. Tundlike vestluste tuvastamine toimub usaldusväärselt ka pikkades vestlustes, kus riski märgid ilmnevad järk-järgult.

Võimalikud lisakaitsemeetmed

1. Tehakse ennetustegevust (nt tervishoiu- ja erialaorganisatsioonide, meedia kaudu) juturobotite sõltuvuskäitumist soosivatest omadustest ja nende kasutamise riskidest vaimse tervise kontekstis. **(Riik)**
2. Piiratakse suhte- või partnersuhet simuleerivate juturobotite turundamist ja levitamist alaealistele. **(Riik)**
3. Riik valmistub käsitlema sotsiaalpoliitikas AI-psühhoosi sarnaselt teiste sekkumist vajavate sõltuvuskäitumistega ja meeleoluhäiretega. Selleks korraldatakse poliitikat ettevalmistavad

³⁴Strengthening ChatGPT's responses in sensitive conversations <https://openai.com/index/strengthening-chatgpt-responses-in-sensitive-conversations/>. Viimati külastatud 29.06.2026.

³⁵ChatGPT adds mental health guardrails after bot 'fell short in recognizing signs of delusion' <https://www.nbcnews.com/tech/tech-news/chatgpt-adds-mental-health-guardrails-openai-announces-rcna222999>. Viimati külastatud 11.06.2026.

uuringud, mille lisaeesmärk on töötada välja käitumise äratundmist ja käsitlemist õpetavad koolitused. **(Riik)**

4. Uute juturobotite loomisel või rakendamisel valitakse kasutatavaid mudeleid selle järgi, et neil oleks võimalikult vähe sõltuvuskäitumist toetavaid mustreid³⁶. **(Organisatsioon)**
5. Juturoboti-lahendused tuvastavad aktiivselt emotsionaalse sõltuvuse ja vaimse stressi märke ning suunavad kasutaja omaenda autonoomiat rakendama, mitte vestlust pikendama. **(Organisatsioon)**
6. Aktiivselt väärtustatakse inimliku kontakti olulisust tehnoloogiaküllases ühiskonnas (nt e-riiki planeerides, ehitades, kommuniqueerides). **(Organisatsioon)**
7. Inimesed õpivad tähele panema lisaks muule sõltuvuskäitumisele ka sõltuvust AI-süsteemidest ning pööraavad sellele tähelepanu. **(Inimene)**

³⁶Abiks on erinevad testid, nt Spiral-Bench. <https://eqbench.com/spiral-bench.html>. Viimati külastatud 29.06.2026.

4 Soovitused üksikisikule

4.1 Sissejuhatus

Tööandjad, organisatsioonid ja riik teevad omalt poolt ära selle, millele ühe inimese jõud peale ei hakka. Kuid igaühel meist on turvalisuse ja ohutuse tagamisel täita oma roll.

4.2 Kuidas käituda internetis, sotsiaalmeedias ja äppides?

Arvestage, et tekst, pilt, heli ja video internetis võivad olla tehisintellekti loodud. Kui keegi kirjutab või räägib midagi, mis tundub imelik või selle isiku kohta ebatavaline, siis kontrollige mõne teise kanali (nt telefonikõne, ühised tuttavad) kaudu üle, kas see info on tõene.

Kuulsad inimesed reklaamimas mõnda toodet, väärtpaberit või investearingut võivad olla tehisintellekti abil genereeritud heli ja kujutistega. Kui kahtlete, raporteerige sisu võimaliku pettuseks, kasutades sotsiaalmeedia või järelevalveasutuste kanaleid. Kui leiate pettuse, siis võib see tähendada, et keegi teine saab oma vara alles jätta.

Kui teie kohta on sotsiaalmeedias või avalikus ruumis piisavalt teavet, saab tehisintellekt koostada uut sisu, mis teeb teid rõõmsaks või vihaseks. Soovitame harjutada ärritavale või lipitsevale sisule reageerimisel viivitamist. Teid võidakse tehisintellekti abil manipuleerida vihastama või näiteks ka mõnda toodet ostma. Sellisel juhul ei ole te enam ise see, kes valikuid teeb.

Kindlasti ei tasu mõelda, et tehisintellekti ajastul tõde enam olemas pole. On küll, lihtsalt selle võltsingutest eristamine on siinsete soovituste kirjutamise hetkel mõnda aega keerukam, kuniks pettuste vastu võitlejad enda lahendused sisse seavad.

4.3 Üldine hügieen tehisintellekti kasutamisel

Andmekaitse seisukohast tuleb võimaluse korral eraldada töö- ja isiklikud seadmed ning piirata isiklikest seadmetest juurdepääsu potentsiaalselt ohustatud süsteemidele, teadvustades samas, et ründed isiku vastu võivad tulla ka tööseadmete ja -kontaktide kaudu.

AI-teenustesse ei tohiks sisestada tundlikku või konfidentsiaalset teavet, kui andmetöötluste ja kaitse põhimõtted ei ole selged. Tasub teha mõtteharjutust, et mis saaks, kui kõik info mida tehisintellektile räägite, on ühel päeval avalik ja teie isikuga seotud.

Kontrollige AI-süsteemi väljastisi hoolikalt enne nende praktikas rakendamist. Vältige oluliste otsuste tegemist üksnes masina antud teabe põhjal. Mõelge, et kui teie saite sellise tulemuse, siis said suure tõenäosusega ka teised said sama tulemuse. Kas see muudab teie lähenemist AI väljastatule?

Jälgige enda kasutamisharjumusi ja püüdke märgata varajasi märke sõltuvusest, mis võib mõjutada teie otsustusvõimet või igapäevast toimetulekut. Regulaarne puhkus tehnoloogiast, teadlikkuse tõstmine tehisintellekti ohtudest ja digikirjaoskuse arendamine aitab säilitada kontrolli oma valikute ja vaimse heaolu üle.

4.4 Soovitusi tehisintellektitoodete valikuks

Kui kasutate tehisintellekti rohkemaks kui meelelahutuseks, siis mõelge, mis on kaalul, kui tööriist teid alt veab, näiteks halba nõu andes. Kas kaotate raha, muud vara, mainet, rikute enda või

teiste tervist?

Sellist mõtlemist kutsutakse riskide hindamiseks ja see on vajalik ka kinnisvara, sõidukite või tööriistade ostmisel. Tehisintellekti ajastut eristab eelnevatest peamiselt see, et AI-kasutajate praegustel, kõige esimestel põlvkondadel puudub võimalus õppida eelmiste põlvkondade kogemustest.

Katsetage võimaluse korral uusi tooteid ja lahendusi väljamõeldud andmetega, piiratud keskkonnas (ärge ühendage sinna külge kõiki enda andmeid, kalendreid ega muud, mida kardate kaotada).

4.5 Tehisintellekti ja agentide kasutamine töös ja hobides

Tehniliste meetmetena saab üksikisik valida AI-teenuseid ja arendusraamistikke, millel on vaiki-misi aktiveeritud kaitsemeetmed pahatahtlike sisendite ja väärkasutuse vastu.

Agentsüsteemidele tuleb anda vaid miinimumini piiratud pääsuõigused ning kõik olulised või pöördumatud tegevused tuleb kinnitada ise. Lisaks tuleks kasutada lahendusi, mis võimaldavad agendi tegevust reaalsajas jälgida, piirata või kohe katkestada, ning vajaduse korral rakendada täiendavaid sisendikaitsemudeleid, arvestades, et ka need on ise keelemudelid ja võivad olla avatud promptisüsteidele.

Turvalisuse tagamiseks tuleb eelistada teenuseid, millel on läbipaistvad turvameetmed ja mis toetavad kasutaja autonoomiat.

4.6 Kuidas käituda, kui tehisintellektiga läks midagi valesti?

Eesti ja Euroopa Liit reguleerivad kodanike õiguseid ja nende kaitset. Kui tehnoloogia kasutamine rikub õigusi või põhjustab kahju, tuleb rakendada oma õigust nõuda andmetöötluse lõpetamist või piiramist ning pöörduda vajaduse korral järelevalveasutuste, politsei või kohtu poole.

5 Soovitused organisatsioonile

5.1 Sissejuhatus

Organisatsioonidel on suur võim seada konkreetseid reegleid ja neid ellu viia, täpsemalt kui riik seda regulatsioonide ja määrustega teeb. Näiteks üksikisikute käitumine areneb suuresti sõltuvalt sellest, milliseid reegleid tööandjad või õppeasutused neile igapäevaelus seavad.

5.2 Tehisintellekti strateegia loomine

AI kasutamise põhimõtted ja protsessid kujundatakse nii, et tehisintellekt toetab ja võimestab inimeste tööd ning ei nõrgesta kriitilisi oskusi ning juhtide ja ekspertide sõltumatut otsustusvõimet. Nii tagatakse organisatsiooni iseseisvus kestlikkus.

AI strateegia kavandamisel hinnatakse mõju pikaajalisele kompetentsibaasile, töötajate ja liikmete oskustele ja järelkasvule. Peetakse silmas, et olulised teadmised ja võimekused säiliks ka AI-teenuste katkestuste või kättesaadavuse piirangute korral.

5.3 Tehisintellekti poliitika ja juhiste loomine

Organisatsioon loob tehisintellekti turvalise kasutamise poliitika ja juhised, mis tehakse kõigi töötajate jaoks arusaadavaks ning varustatakse näidetega. AI turvalise kasutamise jaoks vajalikud tööriistad tehakse kättesaadavaks kõigile tehisintellekti kasutavatele töötajatele või liikmetele.

Organisatsioon korraldab töötajatele või liikmetele sihipäraseid koolitusi AI-süsteemide õiguslike aspektide, algoritmiliste ohtude, infoturbe- ja käitumuslike riskide teadlikkuse tõstmiseks.

Organisatsioon korraldab töötajatele või liikmetele regulaarseid õppuseid, mis tugevdavad võimet ära tunda pettuseid, võltsitud ning manipuleeritud sisu. Koolitused peavad sisaldama konkreetseid harjutusi, mis õpetavad olulist teavet kolleegilt või partnerilt üle kontrollima teise, sõltumatu sidekanali (nt organisatsiooni sisemisest kontaktide kataloogist või muust usaldatavast allikast saadud telefoninumbri) kaudu.

5.4 Tehisintellekti mõju infoturvapoliitikale

Elutähtsate ja kriitiliste teenuste toimepidevus peab organisatsioonis olema tagatud ka ilma AI pakutava toeta organisatsiooni töötajatele.

Teenuste käideldavus tagatakse varusüsteemide, dokumenteeritud taastemehhanismide, taasteharjutuste ja regulaarsete kriisi- ning äärejuhtumite testide kaudu, mis hõlmavad ka haavatu-
vamate kasutajarühmade vajadusi.

Teenuste, tehnoloogia ja tarkvara hangetes ja arenduses nõutakse läbipaistvust tarneahela, mudelite, taristu ja vastutuse jaotuse kohta ning andmetöötluse põhimõtete avalikustamist.

Isikuandmete ja ärisaladuse töötlemisel eelistatakse Euroopa Liidu või Eesti Vabariigi suveräänseid andmekeskuseid.

Biomeetrilisi tuvastussüsteeme hinnatakse üle regulaarselt, aga vähemalt kord aastas. Kui võltsingute tuvastamise võime on kahanenud, vahetatakse vastavuse otsimise ja kontrolli komponendid välja või lõpetatakse biomeetrilise süsteemi kasutamine isikute tuvastamiseks.

Organisatsiooni digitaalset ründepinda vähendatakse aegunud komponentide eemaldamise, pääsuõiguste ja tulemüüride range seadistamise ja automaatse turvaskaneerimise tööriistadega.

Tehnoloogia arenedes kasutatakse küberturvalisuse tagamisel järjest enam automatiseeritud kaitsemehhanisme, mis suudavad reageerida ka automatiseeritud rünnetele.

5.5 Töötajate ja tarbijate ohutus ning vastutus

Teenused kavandatakse nii, et inimene säilitab lõpliku kontrolli väljundite üle ning tehnoloogia ei vähenda põhjendamatult ligipääsu või inimlikku suhtlust. Teenustes säilitatakse inimliku kontakti võimalus, mis on kättesaadav madalama digipädevusega kasutajatele, haavatavatele gruppidele ning vajadusel ka tehniliste tõrgete korral.

Automaatseid tehisintellektitööriistu kasutatakse vaid klientide ja kodanike pöördumiste esialgselt filtreerimiseks, korduva sisu tuvastamiseks ja vastuste ettevalmistamiseks, mille lõpliku kontrolli, kohandamise ja saatmise teostab alati inimene.

Interaktiivsed vestluslahendused ja personaalsed AI-abistajad projekteeritakse nii, et need ei soodusta sõltuvuskäitumist, AI antropomorfiseerimist (sellele inimlike omaduste omistamist) ega kahanda kasutajate iseseisvat otsustusvõimet.

Juturoboti-lahendused tuvastavad võimalusel aktiivselt märke emotsionaalsest sõltuvusest ja vaimsest stressist ning suunama kasutajat rakendama omaenda autonoomiat ja oskuseid.

5.6 Infoturbe ja andmekvaliteedi tehnilised meetmed

Suuri keelemudeleid rakendavatele e-teenustele luuakse kvaliteeditestimise süsteemid, mis aitavad tuvastada mudeli kvaliteediomaduste muudatust ja languseid ka olukordades, kui mudeli andja sellest ei teavita. Eesmärk on süsteemide läbipaistvuse, kvaliteeti, eelarvamustevabaduse ja deklareeritud otstarbele vastavuse säilitamine tehnoloogiliste uuenduste käigus.

Isikuandmete ja ärisaladuse kaitseks AI-süsteemides rakendatakse andmete minimeerimist, lühendatud säilitustähtaegu ja lisakaitsemeetmena vajadusel privaatsuskaitse tehnoloogiad.

Teenustes, mis töötlevad isikut tuvastavat materjali, rakendatakse võltsitud sisu tuvastamise ja päritolu kontrolli mehhanisme (nt vesimärgistamine või kinnitusstandardid).

5.7 Agentsüsteemide turvalise kasutamise meetmed

AI-teenuste, mitmikagentsüsteemide ja masinõppemudelite valimisel ja hankimisel eelistatakse süsteeme, millel on eelseadistatud vaikumisi aktiivsed kaitsemeetmed pahatahtlike sisendite (promptisüstide) ja muu väärkasutuse (nt agendi piiramatu autonoomia) vastu.

AI-lahenduse tehnilise juurutuse käigus katsetatakse lahendusi enne täismahulist käivitamist rangelt piiratud keskkonnas. Agendile antakse rangelt miinimumini piiratud õigused ja juurdepääsud ning kõik pöördumatud või kõrge mõjuga tegevused (nt andmete kustutamine, maksed, kriitilised otsused, küberfüüsiliste süsteemide mõjutamine) nõuavad eelnevat inimkinnitust.

Arvestada tuleb, et isegi sisendite (agendi juhiste) kontrolli mudelid on keelemudelitel põhinevad ja võivad olla avatud promptisüstidele. Seetõttu ei asenda nende kasutamine pidevat tehnilist seiret ega valvsust.

Süsteemidele lisatakse mehhanismid, mis võimaldavad agendi tegevust reaajas jälgida, määratletud raamides piirata või kiirelt peatada.

6 Soovitused riigile

6.1 Sissejuhatus

Riigil on võimekus kehtestada reegleid, mida organisatsiooni tasemel kehtestada ei ole efektiivne. Seda loomulikult juhul, kui reeglite kehtivuse üle korraldatakse ka järelevalve. Lisaks on riigil tähtis rolli üksikisikute ja organisatsioonide õiguste kaitsel, kui tehisaru levikust tulenevad ohud neid kahjustama peaks. Kui riik on raamistiku loonud, saavad üksikisikud ja organisatsioon selle toel ohutult toimetada.

6.2 Poliitilised valikud

Tehisintellekti tehnoloogia areng ja rakendamise mängureeglid on kindlasti veel 5-10 aastat kiire arengus. Euroopa Liidu AI-määrus annab esmase aluse tehisintellektiga seotud innovatsiooniks ja turvalisuseks, kuid riiklikul tasandil tuleb pidevalt hinnata, kas rakendatud meetmed vastavad tehnoloogia kiirele arengule ja Eesti digiriigi konkreetsetele vajadustele. Seega ei ole selle uuringu tulemuse igavesti kindlad, vaid neid peab vähemalt kolme aasta pärast ümber hindama.

Riik kujundab strateegilise AI-taristu poliitika, mis määratleb selged kriteeriumid, milliseid lahendusi ja millistes keskkondades võib kasutada, lähtuvalt andmete salajasusest, tüübist ja teenuse kriitilisusest.

Toetatakse Eesti konkurentsivõimet ja andmesuveräänsust läbi eestikeelsete, avatud lähtekoodiga ja kontrollitud üldotstarbeliste AI-teenuste (keelemudelid, kõnesüntees, kõnetuvastus jne) arendamise. Vajadusel liidestatakse loodud teenused arvutite ja nutiseadmete enamlevinud operatsioonisüsteemide ja tarkvaraplatvormidega.

Riik tugevdab võimet tuvastada ja reageerida AI-põhiste infooperatsioonidele, koordineeritud mõjutuskampaaniatele ning võltsitud sisu levikule. Selleks toetatakse sõltumatut ajakirjandust, arendatakse ühiskonna vastupanuvõimet väärinfole ning tagatakse Eesti ajaloo, kultuuri ja avaliku arutelu autoriteetsete allikate säilimine muuhulgas ka digitaalsel kujul.

Tervishoiu- ja sotsiaalpoliitikas luuakse valmisolek toetada inimesi, kellel on tehisintellekti (näiteks juturobotite ja personaalsete abistajate) kasutamisest tekkinud sõltuvuskäitumine, reaalsustaju langus või muud tervisehäired.

6.3 Õigusruum

Võltsisikutega seotud pettuste vähendamiseks kohustatakse teenuseandjad ja platvormide haldajaid vastavalt märkima sünteetilist sisu, AI-genereeritud isikuid, juturoboteid ning personaliseeritud mõjutussisu.

AI-teenuste andjad kohustatakse kasutajatele avaldama, milliste andmete ja tunnuste alusel neile esitatud sisu kohandatud on.

Luuakse kiired ja läbipaistvad mehhanismid, mis võimaldavad inimestel nõuda enda kohta levitatud volitamata võltsingute eemaldamist ja taotleda õiguslikku kaitset.

Euroopa Liiduga ja liitlastega koostöös luuakse küberõiguse valdkonnas regulatoorne raamistik vaenulike võõrriikide valeinfo päritolu tuvastamiseks ja leviku piiramiseks.

6.4 Tehisintellekti rakendamisele teenustes

Avalikus sektoris kehtestatakse läbipaistvuse ja vastutuse nõuded. Näiteks peavad kõik avalikud AI-teenused hakkama avalikustama kasutatud keelemudeleid, taristut, andmetöötluse põhimõtted.

Kõrge mõjuga valdkondades (õiguskaitse, kohtusüsteem, õigusloome, sotsiaalteenused) nõutakse süstemaatilist kallutatuse ja sõltuvusriskide hindamist. Iga olulise AI-põhise otsuse puhul tuleb tagada inimese kinnituse, ülevaatuse ja vaidlustamise õigus. Avalike e-teenuste puhul säilib inimkontaktiga või mittetehnoloogiline alternatiivkanal vähekindlustatud või haavatavatele rühmadele.

Suuri keelemudeleid rakendavatele e-teenustele luuakse kvaliteeditestimise süsteemid, mis aitavad tuvastada mudeli kvaliteediomaduste muudatust ja languseid ka olukordades, kui mudeli andja sellest ei teavita. Eesmärk on süsteemide läbipaistvuse, kvaliteeti, eelarvamustevabaduse ja deklareeritud otstarbele vastavuse säilitamine tehnoloogiliste uuenduste käigus.

Tarkvara ja AI-lahenduste hangetes kehtestada nõue turvalisuse tagamiseks kogu tarneahelas.

6.5 Haridus, väljaõpe ja tööturg

Haridussüsteemis säilitatakse ja arendatakse õppe- ja kontrollivorme, mis nõuavad iseseisvat probleemilahendusoskust, argumenteerimist ja käsitöölist/loogilist mõtlemist isegi siis, kui AI-tööriistad on laialdaselt kättesaadavad. Säilitatakse AI-vabad hindamiskontekstid (suuline kaitsmine, eksamid, analoogülesanded), kus oskuste masinast sõltumatu omandamine on kontrollitav. Nii tagatakse pikas perspektiivis hariduse ja diplomite väärtus.

Kõigis haridusastmetes ning ka täiskasvanuhariduses arendatakse süstemaatiliselt digipädevusi, kriitilist mõtlemist ja meediakirjaoskust.

Sihtgruppidele (nt juhid) suunatud koolitusprogrammide ja laiemale avalikkusele Eesti.ai platvormi kaudu pakutakse ajakohast teadmust järgmistes valdkondades.

1. AI kasutamine töötootlikkuse ja loomingulise väljendusrikkuse suurendajana, vältides inimese mõtlemisvõime asendamist.
2. Oma digitaalse jalajälje haldamine, andmete jagamise riskide mõistmine ning vajadusel enda õiguste kaitsmine.
3. AI-süsteemides isikuandmete, ärisaladuse ning muude juurdepääsupiiranguga andmete töötlemise reeglite tundmine.
4. AI süsteemide mõjuanalüüs, sealhulgas väljundite kvaliteedi hindamine, kallutatuse ja algoritmilise mõjutamise äratundmine.
5. AI-agentide turvaline juurutamine ja rakendamine.

Riik täiendab hariduse ja tööturu seiret, hinnates AI mõju tööturule, spetsialistide järelkasvule, õpioskustele, sotsiaalsetele võimetele ja avalike teenuste kvaliteedile. Erilist tähelepanu tuleb pöörata valdkondadele, kus automatiseerimine võib vähendada nooremspetsialistide võimalusi omandada praktilisi oskusi sellisel määral, et riigile oluliste oskustega spetsialiste enam piisaval määral ei lisandu. Sellisel olukorras luuakse motivatsioonipaketid, et riigi kestlikuks toimimiseks olulised oskused oleksid endiselt olemas ka AI-st sõltumatusena.

6.6 Järelevalve

Isikuandmete kaitse järelevalvet tugevdatakse, eriti suuremahulise ja tundliku andmetöötluse korral nagu seda on masinõppemudelite treenimine. Eetikakomiteedele ja järelevalveasutustele antakse selged volitused ja vahendid nõuda tehisintellektisüsteemidele täiendavaid lisakaitsemeetmeid, eriti biomeetriliste, geeni- ja terviseandmete töötlemisel.

Riik saab käivitada sihipärase teavituskampaania, mis selgitab AI-generitud väärinfo, sünteetilise sisu, kohaliku jälgimise ja andmete analüüsi mõju. Teadlikkust tõstmisel pööratakse erilist tähelepanu vaimsele heaolule. Riik toetab uuringuid ja ennetusprogramme, mis käsitlevad AI-süsteemide poolt soodustatud sõltuvuskäitumist.

6.7 Tehnilised meetmed

Eesti infoturbestandardisse E-ITS lisatakse nõuded suurte keelemudelite ja AI-agentide turvaliseks käitamiseks: rangelt miinimumini piiratud õigused, liivakastide kasutamine, tegevuste valideerimine ning inimese kinnituse nõue suure mõjuga või pöördumatute toimingute puhul (andmete kustutamine, maksed, õiguskaitseotsused, küberfüüsilisi süsteeme mõjutavad otsused).

Toetada teadustööd sünteetilise sisu ja isikuvõltsinguid efektiivselt tuvastava tehnoloogia arendamiseks, mis kombineerivad sisupõhist, käitumuslikku ja taristupõhist analüüsi

Riik arendab rahvusvahelises koostöös vaenulikest võõrriikidest pärit valeinfo ja propagandistliku sisu seiresüsteemi, mis tuvastab levivaid narratiive. Kontekstist lähtuvalt on võimalik õiguslikul teel valeinfo levikut piirata või siis sellele vastunarratiivi (nt faktikontroll) luua.

Riik toetab eesti keelt ja Eesti kultuuripärandit arvestavate AI-võimete arendamist. Autoriteetsed eestikeelse ajaloo, kultuurivara, meedia ja teaduse andmestikud avaldatakse struktureeritult, hästi lingitult ja avatud litsentsidega, et võimaldada kvaliteetsete, kohalikus keeles ja kontekstis töötavate mudelite treenimist. Korpuste koostamisel ja masinõppemudelite treenimisel tundlike ja eraelu puudutavate materjalide kaitse.

Kriitiliste e-riigi teenuste jaoks tuleb luua julgeolekust ja andmekaitsest lähtuv AI-taristu, mille kasutamine on seotud selgete keskkonna- ja andmeklassifikatsiooniga. Mahukate arvutus- ja treeningmahu nõudvate projektide jaoks saab sõlmida strateegilised koostöölepingud Euroopa teaduslike arvutuskeskustega.

7 Kuidas uuringu tulemusi rakendada?

7.1 Ohtude olulise järjekorra määramine hindamise teel

Uuringus on koostatud ohukirjeldused, kuid me ei ole neid hinnanud ega järjestanud. Neid saaks hinnata näiteks TA küberturvalisuse komisjoni 2026. aastal avaldatud ohuskaalade¹ järgi. Akadeemia komisjoni kogemus (ning avaldatud hinnete jaotus) näitab, et selline hindamine on raske ning nõuab ohtude ühist läbiarutamist, sarnast tähendusruumi ja arutelu. Uuringus koostatud ohtude efektiivne hindamine vajab laiemat ekspertide kogu ning nõuab neilt kõigi ohtude läbitöötamist ning hinnete andmist. Uuringu ambitsioonikas ajakava seda tööd korraldada ei võimaldanud.

Hindamine võimaldaks anda otsustajatele sisendi, milliste ohtude leevendusmeetmeid peaks rakendama kõigepealt. Eesmärk peaks olema ohtude järjestamine, mis aitab otsustajatel ja investeeringute tegijatel valida, kuhu oma ressursse kõigepealt suunata.

Uuringus koostatud ohud avaldatakse e-riigi koodivaramus, seega hindamist võib metoodiliselt korraldada igaüks. TA küberturvalisuse komisjon tutvub uute ohtudega hinnanguliselt 2026. aasta sügisel.

7.2 Soovitatud leevendusmeetmete rakendamine praktikas

Uuring annab põhjaliku loetelu leevendusmeetmetest, mille rakendamine peab olema järgmine samm.

Järjestuse puudumine ei ole põhjendus leevendusmeetmete rakendamata jätmiseks. Ohukirjeldustes antud soovitusel on konkreetsed ja erineva keerukusega elluviidavad.

Eraisikutele mõeldud soovitusel tuleb anda edasi tavapärase kanalite kaudu (nt ITVaatlik portaal). Nende rakendamise toetamiseks on vaja koolitusmaterjale nt Eesti.ai platvormil.

Organisatsioonidele mõeldud soovitusi saab Justiits- ja Digiministeerium riigiasutustele levitada otse ja määrustega. Erasektorile soovitusel levitamiseks saab kasutada nii koolitusi, konverentse kui ka nende määrustega kohustuslikuks seadmist.

Riigile enda antud soovitusi tuleb täita vastavalt nende iseloomule. Vaja on nii seadusloomet kui ka rakenduslikke samme.

7.3 Puuduvate võimekuste loomine

Uuringu piiritletud kestus ja käsitusala ei võimaldanud välja selgitada, milliste lisameetmete täitmiseks ei ole riigiasutustel lahendusi või nende lahenduste hankimise ressursse täna olemas.

Võib olla leevendusmeetmeid, mida täna riigiasutustes või erasektoris peetakse (kas või ilma hinnapakumiseta, intuitsiooni järgi) kalliks. Selliste leevendusmeetmete odavamaks ja rakendatavaks muutmiseks soovitame teha koostööd teadusasutustega, näiteks RITA+ programmi, Strateegiliste Uuringute Programmi, teaduse tippkeskuste jne projektide raames. Konkreetseid eesmärke seades on võimalik leevendusmeetmeid odavamaks muuta.

¹TA küberturvalisuse komisjoni 2025-2026 aruanne. <https://koodivaramu.eesti.ee/e-riigi-turve/ta-komisjon/-/blob/main/2026/komisjon/kaaskiri-2026.md>. Viimati külastatud 30. juunil 2026.

7.4 Edasised uurimissuunad ja ohukirjelduste uuendamine

Eesti oludes ei ole keemiliste, bioloogiliste, radioaktiivsete ja lõhkeainetega seotud CBRN ohud väga kõrged, sest vastavaid lähtematerjale on keerulisem kätte saada. Siiski tasuks kaaluda jätku-uuringut, mis keskenduks sellele, millised CBRN-ohud AI tõttu Eestis realistlikuks muutuvad.

Koodivaramus avaldatud ohukirjelduste jaoks saab tellida tulevikus uuringud, mis värskendavad eelduseid, lisavad uusi leevendusmeetmeid ja uuendavad vajadusel ohtudele antud hinnanguid.

Bibliograafia

- [1] *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*. Tehniline aruanne ISO/IEC 22989:2022. ISO/IEC, 2022. URL: <https://www.iso.org/standard/74296.html>.
- [2] *Euroopa Parlamendi ja nõukogu määrus (EL) 2024/1689, 13. juuni 2024, millega nähakse ette tehisintellekti käsitlevad ühtlustatud õigusnormid ning muudetakse määruseid (EÜ) nr 300/2008, (EL) nr 167/2013, (EL) nr 168/2013, (EL) 2018/858, (EL) 2018/1139 ja (EL) 2019/2144 ning direktiive 2014/90/EL, (EL) 2016/797 ja (EL) 2020/1828 (tehisintellekti käsitlev määrus) (EMPs kohaldatav tekst)*. PE/24/2024/REV/1, ELT L, 2024/1689, 12.7.2024. URL: <http://data.europa.eu/eli/reg/2024/1689/oj>.
- [3] Ajay Agrawal, Joshua Gans ja Avi Goldfarb. *Power and Prediction: The Disruptive Economics of Artificial Intelligence*. Harvard Business Review Press, nov 2022.
- [4] Fabrizio Dell'Acqua *et al.* „Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality“. *Organization Science* 37.2 (2026), lk-d 403–423.
- [5] Krzysztof Budzyń *et al.* „Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study“. *The Lancet Gastroenterology & Hepatology* 10.10 (2025), lk-d 896–903.
- [6] Michael Gerlich. „AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking“. *Societies* 15.1 (2025). ISSN: 2075-4698. DOI: [10.3390/soc15010006](https://doi.org/10.3390/soc15010006). URL: <https://www.mdpi.com/2075-4698/15/1/6>.
- [7] OECD. *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*. Paris, 2026. DOI: [10.1787/062a7394-en](https://doi.org/10.1787/062a7394-en). URL: <https://doi.org/10.1787/062a7394-en>.
- [8] Lennart Meincke, Ethan Mollick ja Christian Terwiesch. „Prompting Diverse Ideas: Increasing AI Idea Variance“. *ArXiv abs/2402.01727* (2024). URL: <https://api.semanticscholar.org/CorpusID:267412634>.
- [9] Francesco Salvi *et al.* „On the conversational persuasiveness of GPT-4“. *Nature Human Behaviour* 9.8 (2025), lk-d 1645–1653. ISSN: 2397-3374. DOI: [10.1038/s41562-025-02194-6](https://doi.org/10.1038/s41562-025-02194-6). URL: <https://doi.org/10.1038/s41562-025-02194-6>.
- [10] Gadi Evron, Rob T. Lee ja Rich Mogull. *The AI Vulnerability Storm: Building a "Mythos-ready" Security Program*. Tehniline aruanne. Draft strategy briefing released April 12, 2026. Contributors include Jen Easterly, Bruce Schneier, Chris Inglis, Heather Adkins, and others. Cloud Security Alliance CISO Community, SANS Institute, OWASP GenAI Security Project, [un]prompted, apr 2026.
- [11] Alexandra Souly *et al.* *Poisoning Attacks on LLMs Require a Near-constant Number of Poison Samples*. 2025. arXiv: [2510.07192](https://arxiv.org/abs/2510.07192) [cs.LG]. URL: <https://arxiv.org/abs/2510.07192>.
- [12] *Your AI Assistant Is Leaking Your Conversations*. URL: <https://leakylm.github.io/> (vaadatud 29.06.2026).
- [13] Michael Doran. *Conversational Artificial Intelligence as Surveillance Infrastructure*. 20. mai 2026. URL: <https://www.linkedin.com/pulse/conversational-artificial-intelligence-surveillance-michael-doran-ofpaf/> (vaadatud 29.06.2026).

- [14] Jerry Owens. *Claude Identity Verification Starts July 8: What Facial Data Anthropic Collects*. 21. juuni 2026. URL: <https://www.techtimes.com/articles/318778/20260621/claude-identity-verification-starts-july-8-what-facial-data-anthropic-collects.htm> (vaadatud 29.06.2026).
- [15] Markus Kasanmascheff. *Leaked Code Reveals OpenAI's Secret Government Surveillance Network*. 25. veebr 2026. URL: <https://winbuzzer.com/2026/02/25/openai-persona-secret-government-surveillance-watchlist-exposed-xcxwn/> (vaadatud 29.06.2026).
- [16] Daniel AJ Sokolov. *Consumer Protection: Three Chances for Damages from Meta Platforms*. 16. juuni 2026. URL: <https://www.heise.de/en/news/Consumer-Protection-Three-Chances-for-Damages-from-Meta-Platforms-11334898.html> (vaadatud 29.06.2026).
- [17] Ott Velsberg ja Kristel Rillo. *Tulevikukindla arvutusvõimekuse tagamine. Eesti riigi arvutusvõimekuse hetkeseis ja vajadused 2030. aastaks*. 2026. URL: https://www.justdigi.ee/sites/default/files/documents/2025-12/Tulevikukindla%20arvutusv%C3%B5imekuse%20tagamine%20v1.1_0.pdf.
- [18] *Euroopa Parlamendi ja nõukogu direktiiv (EL) 2024/2853, 23. oktoober 2024, mis käsitleb vastutust puudusega toodete eest ja millega tunnistatakse kehtetuks nõukogu direktiiv 85/374/EMÜ (EMPs kohaldatav tekst)*. PE/7/2024/REV/1, ELT L, 2024/2853, 18.11.2024. URL: <https://eur-lex.europa.eu/eli/dir/2024/2853/oj>.