Anthropic Claude

Kirjeldus

Käesolev usaldusväärsuse hinnang keskendub Anthropic Claude pilvtöötlusteenuse riskide kirjeldamisele ning ei kohaldu toodetele, mida on võimalik kasutada asutuse sisestel ressurssidel.

Tehisintellekti (artificial intelligence, AI) all mõistame mistahes süsteemi, mis suudab sooritada ülesandeid pealtnäha inimintelligentsi kasutades. Majandus- ja Kommunikatsiooniministeeriumi kratikavad on näinud ette tehisintellekti ulatuslikku rakendamist avalikus sektoris. LLMide (*large language model*) ning difusioonipõhiste pildisünteesimudelite levik ja tavatarbijale kättesaadavamaks muutumine on põhjustanud selles valdkonnas arenguhüppe, sealhulgas AIaaS (artificial intelligence as a service, tehisintellekt teenusena) ärimudeli leviku.

Anthropic Claude pakub juturobotit, mille vahendusel on võimalik uusi ideid genereerida, koodist vigu leida, andmeid analüüsida ja visuaale luua.

Anthropic on asutatud 2021. aastal Dario Amodei ja Daniela Amodei poolt. 2020. aastal lahkusid nad koos viie töökaaslasega OpenAI firmast olles mures kasutatavate turvameetmete pärast. Nad lõid Anthropicu sest tahtsid luua usaldusväärseid, seletavaid ja juhitavaid tehisintellekti süsteeme. Nad panevad rõhku tehisintellekti edasisele uurimisele, eriti turvalisuse valdkonnas.[1]

Anthropicul on kontoreid üle maailma. Need asuvad Ameerika Ühendriikides (Boston, New York, San Francisco, Seattle, Washington), Suurbritannias (London), Iirimaal (Dublin), Šveitsis (Zülrich) ja Jaapanis (Tokyo).[2]

Anthropic on 14 rahastusvooruga kaasanud kokku 27,3 miljardit dollarit. Suurim rahastusring oli 13 miljardi dollari suurune F-seeria 2025. aasta septembris. 2024. aasta sügisel tugevdasid Anthropic ja AWS oma koostööd. AWS investeeris 4 miljardit dollarit ning kokku on nad panustanud 8 miljardit dollarit.[3] [4]

Anthropic on end süüdi tunnistanud autoriõigustega kaitstud sisu suuremahulises ebaseaduslikus allalaadimises ja kasutamises Claude mudeli treenimiseks[5].

Anthropic Claude volitatud töötlejate osas eraldi usaldusväärsuse hinnangut ei koostata ja usaldatakse teenusepakkujat.

[1] https://builtin.com/articles/anthropic
[2] https://www.anthropic.com/jobs
[3] https://www.anthropic.com/news/anthropic-amazon-trainium
[4] https://tracxn.com/d/companies/anthropic/SzoxXDMin-NK5tKB7ks8yHr6S9Mz68pjVCzFEcGFZ08/funding-and-investors#funding-rounds
[5] https://www.theguardian.com/technology/2025/sep/05/anthropic-settlement-ai-book-lawsuit

Toote võimalused

SaaS - Claude on juturobot, mida on võimalik kasutada viies erinevas stsenaariumis:

- Õppimine - Võimalik õppida ja saada uusi teadmisi läbi vestluse.
- Programmeerimine - Koodi genereerimine, vigade parandamine, koodi analüüsimine, koodi selgitamine ja uute keelte õppimine.
- Teadustöö - informatsiooni kogumine veebist ning vastuseks usaldusväärsed allikad.
- Analüüs - Andmete analüüsimine ja nende põhjal visuaalide loomine.
- Loomine - Ajurünnaku läbiviimine ja ideede edasiarendamine.

Claude Code võimaldab tehisintellekti abil aru saada koodibaasist ning genereerida uusi funktsionaalsusi. Samuti saab Claude Code ühendada käsureaga.[6]

Claude pakub võimalust ühendumiseks rakendusliidesega {API) kaudu. Rakendusliidesest on võimalik juurdepääs erinevatele mudelitele: Opus 4.1, Sonnet 4.5, Haiku 3.5.[7]

Claude võimaldab luua projekte, millega saab samateemalised vestlused koondada ühte gruppi. Projekti raames on tehisintellektil ülevaade kogu kontekstist ning järgnevad vastused põhinevad eelnevalt räägitule.

Projektisiseselt saab määrata spetsiifilisi juhiseid, kuidas juturobot peaks küsimustele vastama.[8]

Anthropicu seadistused sisaldavad valikut, kus saab määrata, kas kliendiandmeid kasutatakse mudelite treenimiseks. Seda seadistust on võimalik iga hetk muuta. Kui luba on antud, siis kasutatakse treenimiseks uute ja jätkatud vestluste andmeid ning neid säilitatakse kuni viis aastat. Kui luba ei ole antud või vestlus kustutakse, hoitakse andmeid kuni 30 päeva.[9]

Soovitus: võimaldab optimeerida baasteenuste kulusid ning maandab tarkvara

Kasutusjuhud

Standardloetelu teenuse kasutusviisidest:

- ajurünnakute läbiviimine;
- ideede genereerimine;
- teabe (faktide) ja seletuste küsimine;
- analüüside sooritamine;
- allikate ja viidete leidmine;
- programmeerimine (koodi kirjutamine, vigade analüüsimine).

EL tehisintellektimääruse kohane riskitase

Anthropic Claude teenust vaatleme kui kogust üldotstarbelisi suuri keelemudeleid, mida kasutatakse juturoboti teenuse pakkumiseks läbi veebiliidese. Teenuse kasutamine vastavalt kirjeldatud töötlusjuhule liigitub minimaalse riskiga tehisintellektiks.

Anthropic Claude teenust on võimalik kasutada ka viisidel, mis liigituvad suure riskiga tehisintellektiks või keelatud tehisintellektiks. Sellised kasutusviisid aga ei pruugi olla kooskolas Anthropicu poolt seatud reeglitega kasutajale.

Soovitus: paika panna asutuses protsessid, et toodet ei kasutataks keelatud viisil

[6] https://claude.com/product/claude-code
[7] https://claude.com/platform/api
[8] https://www.anthropic.com/news/projects
[9] https://privacy.claude.com/en/articles/10023548-how-long-do-you-store-my-data

Rahaline mõju

Anthropic pakub erinevaid pakette. Individuaalseks kasutamiseks on võimalus valida kolme paketi vahel. Esimese paketi nimi on Free ehk teenust saab kasutada tasuta. Järgmine pakett on Pro ja selle eest tuleb maksta 17 dollarit kuus. Kolmas pakett on Max, mille eest tuleb maksta vähemalt 100 dollarit kuus.

Samuti pakutakse pakette tiimide jaoks. Pakett nimega Team sisaldab kaht võimalust. Esimene on Standard Seat ja selle eest peab maksma 25 dollarit kuus ning teine on Premium Seat ning see maksab 150 dollarit kuus.

Kõige ulatuslikum pakett tiimide jaoks on Enterprise ning selle osas tuleb ise küsida täpsemat pakkumist.

Anthropic'ul on eraldi hinnakiri kasutamaks API ligipääsu Claude mudelitele. Hinnad põhinevad kasutatud tookenite (token) arvul. Täpsem hinnakiri on leitav siit.[10]

Anthropic Claude rahaline mõju sõltub asutusele sellest, kuidas seda kasutatakse, millistes valdkondades ja kui suures mahus. Asutused peaksid hindama oma konkreetseid vajadusi ja eesmärke, et teha teadlikke otsuseid Anthropic Claude kasutamise osas.

Soovitus: kontrollida toodete hindu hinnakirja alusel

EL/NATO liikmesriigis hoitavad andmed

Anthropic andmekeskused asuvad üle maailma. Andmekeskused andmete talletamiseks asuvad Ameerika Ühendriikides. Andmete töötlemiseks kasutatakse servereid, mis asuvad Ameerika Ühendriikides, Euroopas, Aasias ja Austraalias. Anthropic kasutab mitmeid teisi teenuseid, mille andmekeskuste asukohad on leitavad siit.[11].

Anthropic'ul on lehekülg, kus on näha kõikide nende teenuste seis[12].

Samuti leidub lehekülg, kus on loetletud riigid ja regioonid, kus saab kasutada rakendusliidest ja Claude.ai[13].

Andmete asukoht sõltub seega teenuse infrastruktuurist ja Anthropic Claude serverite asukohtadest.

Soovitus: kontrollida, kus teenusepakkuja andmeid majutab

Teenusest lahkumise ja andmete ekspordi võimalus

Pilvandmeid kättesaamiseks tuleb esitada päring. Päringu tulemusel tagastatakse kliendi isikuandmed, mida Anthropic töötleb {teatud erandite ja tingimustega).[14]
Lisaks isikuandmetele on igal kasutajal võimalik alla laadida vestlusi. Enterprise paketi puhul saab ka alla laadida viimase 180 päeva logid.[15] [16]
Team ja Enterprise pakettide kasutajate jaoks on olemas funktsionaalsus Memory (mälu). Memory sisse lülitamisel jätab Claude meelde erinevate projektide info, kasutajate eelistused, läbivad mustrid ja

[10] https://www.anthropic.com/pricing
[11]https://privacy.anthropic.com/en/articles/7996890-where-are-your-servers-located-do-you-host-your-models-on-eu-servers
[12] https://status.anthropic.com/
[13] https://www.anthropic.com/supported-countries
[14] https://www.anthropic.com/legal/privacy
[15]https://support.claude.com/en/articles/9970975-how-to-access-audit-logs
[16]https://support.claude.com/en/articles/9450526-how-can-i-export-my-claude-data

konteksti. Seda mälu on võimalik eksportida kas seadistuste alt või otse vestluskanalist. Teiste juturobotite mälu on võimalik Claude'i importida. Kuna mälu uuendusi vaadatakse üle kord päevas, siis võib imporditud mälu kasutuselevõtt võtta aega kuni 24 tundi.[17]

Andmeid töödeldakse ja säilitatakse vastavalt teenuse kasutamisele ning vajadusele.
Andmekao vältimiseks peaks andmete omanik hindama Anthropic Claude pilvteenuses olevat andmete koosseisu ning nende ajaloolist väärtust, et hinnata, kas andmete eksportimine on vajalik ja otstarbekas.

Soovitus: kontrollida, kui kaua teenusepakkuja andmeid säilitab

Vastavus sertifikaatidele ja nõuetele (ISO, E-ITS jms)

AI tehnoloogia rakendamisel tuleb lisaks seadusele järgida ka küberturbe ja ühiskondliku ohutuse nõudeid[18].

Anthropic on saanud sertifikaadi ISO 27001:2022 ja ISO/IEC 42001:2023. Neid on auditeeritud American Institute of Certified Public Accountants {AICPA) SOC 2 Type 1 ja SOC 2 Type 2 auditites. On võimalik Anthropic seadistada nii, et see oleks vastavuses HIPAA (The Health Insurance Portability and Accountability Act) nõuetega.[19]

AI juurutamisel tuleks lähtuda juhtimissüsteemi rakendamisest, nt ISO/IEC 42001, mida saaks integreerida vajadusel ISO 9001 ja ISO/IEC 27001 standarditel põhineva juhtimissüsteemidega. Asutused saavad AI kasutuselevõtu korral lähtuda olemasolevatest infoturbe ja andmekaitse vastavusnõuetest.

Küberturvalisusel on oluline roll, et tagada AI süsteemide vastupidavus katsetele muuta nende kasutamist, käitumist, jõudlust või ohustada turvaomadusi pahatahtlike kolmandate osapoolte poolt, kes võivad süsteemi haavatavusi ära kasutada. Ründajad võivad võtta sihikule nt treeningandmed (andmemürgitus), treenitud mudelid (pahatahtlik rünne või kuuluvuse tuvastamise rünne) või kasutada ära AI süsteemi digitaalsete varade või selle aluseks oleva IKT infrastruktuuri haavatavusi. Riskidele vastava küberturvalisuse tagamiseks tuleb rakendada sobivaid ja tõhusaid meetmeid, võttes arvesse ka praegust tehnoloogia taset.
Euroopa Komisjon avalikustas 2021. aasta aprillis esimese AI-d reguleeriva raamistiku. Viidatud ettepanek on kantud riskipõhisest lähenemisviisist ehk AI süsteeme tuleb analüüsida ja klassifitseerida vastavalt sellele, millist ohtu need kasutajatele kujutavad. Samas ei tohi ära unustada ka kehtivat õigusraamistikku[20].

13. märts 2024 vastuvõetud AI määrus sätestab tingimused, millistel juhtudel on AI süsteemi kasutamine keelatud, nt kui see põhjustab kahju inimese elule, tervisele või põhjustab diskrimineerimist (artikkel 5)[21]. AI määrus kirjeldab ka, millised on riskitasemed AI süsteemide osas ning reguleerib ka AI süsteemide kasutamist ja määratleb rikkumisest teavitamise võimalused (nt järelevalveasutuse).

Kui AI süsteem või seda opereeriv isik kuulub määrus(t)e kohaldamisalasse, tuleb

[17] https://support.claude.com/en/articles/12123587-importing-and-exporting-your-memory-from-Claude
[18] https://www.ria.ee/sites/default/files/documents/2024-03/Tehisintellekti-masinoppe-tehnoloogia-riskide-uuring-2024.pdf
[19] https://support.anthropic.com/en/articles/10015870-what-certifications-has-anthropic-obtained
[20] https://www.ria.ee/sites/default/files/documents/2024-03/Tehisintellekti-masinoppe-tehnoloogia-riskide-uuring-2024.pdf
[21] https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

hinnata, millised konkreetsed nõuded määrus(te)st tulenevad GDPR-i kohaselt on oluline hinnata, kas kvalifitseerutakse näiteks isikuandmete vastutavaks või volitatud töötlejaks, AI määruse puhul aga näiteks AI-süsteemi teenustajaks (”provider”) või juurutajaks (”deployer”). Määruste kohaseid rolle on veelgi ja ka need on mõistlik üle vaadata. Konkreetsete nõuete tuvastamiseks on vaja teada ka seda, mis on andmetöötluse ja AI kasutamise eesmärk, millised andmetöötlusprotsessid süsteemis toimuvad, millised andmed ja kelle vahel liiguvad ning millist AI-süsteemi või komponenti (sh selle riskitase) süsteemis kasutatakse[22].

Standardite järgimine panustab toodete või teenuste ohutuse, kvaliteedi ja töökindluse tagamisse, samuti võivad need parandada ja tõhustada ettevõtte süsteeme või protsesse. AI süsteemide erinevate elutsüklite puhul rakendatavate standardite kohta on võimalik lugeda ENISA publikatsioonist heade küberturvalisuse praktikate kohta AI süsteemide puhul[23].

Soovitus: kontrollida olemasolevate sertifikaatide vastavust KüTSile

Erinõuded

Nõuded teenustajale. Üldotstarbeline tehisintellektimudel peab järgima AI määruse artiklis 50 kirjeldatud läbipaistvuskohustusi.

Tehisaru poolt loodud sisu, näiteks pildid, heli või videofailid, tuleb sellistena selgelt märgistada, et selle sisu edasistel kasutajatel oleks võimalik tuvastada, et see on loodud tehisaru abil.

Nõuded kasutusele võtjale. Juhul kui teenusesse sisestatakse isikuandmeid, tuleb järgida isikuandmete kaitse reegleid, mh hinnata andmekaitse mõjuhinnangu vajalikkust ja vajadusel see läbi viia. Samuti tuleb arvestada muude kohalduvate õigusaktidega (näiteks avaliku teabe seadus).

Anthropicul on kasutuspoliitika, mis nimetab arvukad juhud, mille puhul ei tohi Anthropicu mudeleid kasutada.

Teenust on võimalik kasutada ka kõrge riskiga kasutusjuhtudeks. Sellised kasutusjuhud tekivad, kui käsitletakse teemasid: juriidika, tervishoid, kindlustus, rahandus, töövõime, eluase, akadeemia ja ajakirjandus. Nende teemade käsitlemiseks on Anthropic määranud kaks reeglit, mida tuleb järgida[24]:

- Otsustusprotsessi tuleks kaasata oma ala professionaal, kes saab nõu anda ja selgust tuua AI poolt antud vastusele.
- Kui tehisintellekti poolt genereeritud sisuks on vajadus edastada see järgmistele isikutele või tarbijatele, tuleb neile teada anda, et sisu on loodud AI abiga.

Anthropic on välja toonud, et võib leiduda erandeid seoses kasutuspoliitika rakendamisega. Sellisteks võimalikeks klientideks on valitsusasutused, kellega koostöös võib Anthropic nende jaoks muuta kasutamise reegleid, kui need on vastavuses selle asutuse ülesande ja seadustega.[25]

Soovitus: kontrollida, milliste kasutusjuhtude puhul on mudeli kasutamine keelatud

[22] https://www.ria.ee/sites/default/files/documents/2024-03/Tehisintellekti-masinoppe-tehnoloogia-riskide-uuring-2024.pdf

[23] https://www.ria.ee/sites/default/files/documents/2024-03/Tehisintellekti-masinoppe-tehnoloogia-riskide-uuring-2024.pdf

[24] https://www.anthropic.com/legal/aup

[25] https://support.claude.com/en/articles/9528712-exceptions-to-our-usage-policy

Andmekaitsetingimused

Andmetöötlust reguleerivad peamiselt teenuseleping (Commercial Terms of Service) ja andmetöötluslisa (Data Processing Addendum.

Teenuselepingu kohaselt ei või Anthropic kasutada kliendisisu mudelite treenimiseks.

Teenustesse sisestatud andmete (Teenuseleping ei täpsusta, et isikuandmete, kuid andmetöötluslisa seda teeb) töötlemine toimub vastavalt andmetöötluslisale. Teenuseleping sisaldab üldist konfidentsiaalsusklauslit.

Anthropic võib kasutada kliendi nime ja logo, et avalikult klienti kliendina välja tuua. Klient võib sellest eraldi keelduda[26].

Anthropic ei müü kliendi isikuandmeid. Ühtlasi ei säilita, kasuta ega avalda Anthropic isikuandmeid laiemalt kui konkreetne pooltevaheline suhe ning muul eesmärgil kui andmetöötluslisas on kokku lepitud. Kui seadusega ei ole lubatud, ei kombineeri Anthropic kliendi isikuandmeid muude isikuandmetega, mida ta saab teistelt isikutelt või saab otsesest suhtlusest andmesubjektiga.

Anthropic kasutab alltöötlejaid. Alltöötlejate nimekiri on leitav veebilehelt: https://www.anthropic.com/subprocessors.

Turvaintsidendi toimumisel teavitab Anthropic klienti 48 tunni jooksul alates intsidendist teada saamisest.

Teenuselepingu lõppemisel annab Anthropic kliendi soovil 30 päeva jooksul kliendile tema andmete koopiad ja kustutab andmed, kui puudub alus pikemaks säilitamiseks[27].

Lisaks on olemas privaatsuspoliitika (Privacy Policy), kuid see reguleerib andmetöötlust piiratud ulatuses (kohaldub olukorras, kus Anthropic Claude on vastutav töötleja)[28] ning lisaks on tarbijatele kohalduvad teenusetingimused (https://www.anthropic.com/legal/consumer-terms).

Erinevate teenuste osas on lisaks olemas teenuse-spetsiifilised tingimused, mis on leitavad veebilehelt: https://www.anthropic.com/legal/service-specific-terms. Mh paistab, et mingitel juhtudel võib privaatsuspoliitika siiski kohalduda ka juhul, kui ülejäänud teenuseosutamise vaatest ei ole Anthropic vastutavaks töötlejaks (nt Claude for Work (Team Plan; Enterprise Plan puhul on privaatsuspoliitika kohaldumisele viidatud).

Isikuandmete kaitse üldmäärus peab oluliseks kaitsta füüsilisi isikuid isikuandmete automatiseeritud töötlemise puhul. Lisaks eelnevale peab tehisaru arendamisel, rakendamisel ja kasutamisel arvestama ka muude nõuetega, näiteks intellektuaalomandi õigusega.

Soovitus: kontrollida, milliseid alltöötlejaid teenusepakkuja kasutab

Turvameetmete rakendamine

Rakendatud turvameetmed on andmete krüpteerimine (nt AES- 256 jõude- ja TLSl.2+ liikvel andmete jaoks), lõppseadmete kaitsesüsteemid, pääsukontrolli meetmed, ettevõtte personali haldus, turvalogimine ja seire ning võrgu segmentimine kliendiandmete keskkondade eraldamiseks[29]. Turvalisust juhitakse infoturbekava järgi, mida auditeerib väline osapool. Ka

[26]https://www.anthropic.com/legal/commercial-terms

[27]https://www.anthropic.com/legal/data-processing-addendum

[28]https://www.anthropic.com/legal/privacy

[29]https://trust.anthropic.com/

kliendiandmeid töötlevate süsteemide läbistustestimisi teeb väline osapool. Tarnijatega seotud riske hallatakse.

Avalikest allikatest ei selgu, kelle käes on krüptovõtmed - need võivad olla Anthropicu või tema pilvepakkuja valduses. Antrhopic ei näi pakkuvat võimalust võtta teenust kliendi enda võtmetega.

Meetmed mudelite ohutuse tagamiseks. Claude Code turvalisuse lehekülje järgi on vaikimisi rangelt ainult lugemisõigused.

Rakendatud kontrollimeetmed: lubatavate ja riskantsete käskude loendid (allowlisting, blacklisting), sisendi saneerimine, prompti/päringu kontekstiteadlik analüüsimine, kinnituse küsimine võrgupöördumistele, kahtlasena tuvastatud käskudele ja veebist andmete tõmbamiseks, isoleeritud kontekstiakna kasutamine, API võtmete krüpteerimine.

Skriptide käitamiseks ja tööriistade käivitamiseks eelistatakse kasutada virtuaalmasinaid, seda eriti väliste veebiteenustega suhtlemisel. Kliendil on võimalik kasutada enda mudelikonteksti protokolli (model context protocol, MCP). Anthropic soovitab kasutada arenduskonteinereid (devcontainer) ja regulaarselt vaadata üle õiguste seadistused.

On kasutaja vastutus anda AI agendile vaid need õigused, mida see vajab tegutsemiseks; samuti on kasutaja kohustatud genereeritud koodi ja käsud enne vastu võtmist ohutuse seisukohast läbi vaatama[30].

Tootja juhised AI kaitsepiirete (guardrails) kasutamiseks vt[31] ja sellele järgnevad lehed.

RiskRubic.ai[32] järgi põruvad Anthropicu Claude mudelid läbipaistvuse kategoorias, küll aga saavad kõrgeid hinnanguid ülejäänud viies kategoorias.

AI-spetsiifilised nõrkused. On näidatud, et promptisüstiga (indirect prompt injection) on võimalik manipuleerida Claude AI API eraldama kasutajatele andmeid ja edastama seda ründaja kontole[33].

Anthropic on välja andnud "vastutustundliku mastabeerimise poliitika"(RSP, Responsible Scaling Policy)[34], mis sätestab AI kasutamisega seotud riskide vahendamise protokollid ja protsessid . Selles poliitikas defineeritakse uus raamistik nimega AI turvalisuse tasemed"(ASL, AI Safety Levels), mis kirjeldab vajalikke turvameetmeid erineva taseme tehisintellekti mudelite jaoks.

Anthropic ise on kasutusele võtnud ASL 3 taseme turvameetmed.

Soovitus: kontrollida, kus ja kelle käes asuvad krüptovõtmed

Riskid

AI-ga seotud riskid

Andmete töötlemise osas pole võimalik veenduda, kuidas ja milliseid andmeid töödeldakse ning mis otsuseid AI teha võib andmete pinnalt. Andmete lekkimise oht (konfidentsiaalsed- ja isikuandmed).
Kindlad protseduurid, milliseid andmeid antakse ning millised on reeglid töötlemisel. Tehakse konkreetsele pakkujale/tehnoloogiale turvaanalüüs seoste/tagauste tuvastamiseks ja võetakse tootepõhiselt kasutusele lisaturvameetmeid.

[30] https://docs.claude.com/en/docs/claude-code/security
[31] https://docs.claude.com/en/docs/test-and-evaluate/strengthen-guardrails/reduce-hallucinations
[32] https://riskrubric.ai
[33] https://www.securityweek.com/claude-ai-apis-can-be-abused-for-data-exfiltration/
[34] https://www-cdn.anthropic.com/872c653b2d050ld6ab44cfEpdf

AI-ga seotud riskid	Andmete töötlemisel ei saa tagada, et järgitakse GDPR-i. AI võib valimatult koguda andmeid, mille õiguslikku alust töötlemiseks pole. AI võib andmeid muuta selliselt, et tekib kahju asutusele. Kuna töödeldakse näiteks pöördumiste sisu on pahatahtlikul muutmisel suur mõju IT-abile ja kasutajale. **Andmed anonümiseerida või muuta isikustamata kujule. Majutada enda juures. AI kasutuselevõtu korral tuleb veenduda, et asutuse osas antud info väljund ei oleks asutuse mainet kahjustav.**
	Ründajad kasutavad ära AI turvanõrkusi, et saada ligi kasutajate andmetele. AI mitte otstarbelisest kasutamisest võib tekkida turvanõrkus. AI-d kasutatakse küberrünnakuteks, et pääseda mööda turvanõuetest ja seeläbi ära kasutada süsteeme. **Erinevate tarkvarade kasutuselevõtmine, mis kaitseb kahjurprogrammide eest. AI süsteemid ei tohiks määratletud tingimustel viia olukorrani, kus inimelu, tervis, vara või keskkond on ohus.**
	AI teadlikkuse kasv ning õppimisvõime muudab AI-d autonoomsemaks ning AI võib võtta vastu otsuseid, mis ei ole kooskõlas tellija nõuetega. AI võib teha otsuseid iseseisvalt, mis tekitab täiendavat turvariski. Andmete lekke ning andmete väärkasutamise oht suureneb. **AI määruse vastuvõtmine EU poolt, täiendavate kriteeriumite loomine ja järgimine, mis alustel tohib AI-d kasutada.**
	Intellektuaalse omandi osas rikkumine, kui pole selge, kellele vastutus kuulub. Kiired regulatiivsed muudatused võivad kaasa tuua keelde AI kasutuselevõtu osas või liigselt piirata turgu, mistõttu võivad olemasolevad lahendused olla keelatud ning vastuolus seadusega. **Pidev arengute järgimine, et käia kaasas kehtiva seadusandlusega ning anda ka sisendeid seaduse muudatuste osas (õigusloomesse panustada riigi tasandil).**
	AI on soodsam kui inimtööjõud. AI otsuseid on vaja kontrollida ja on vaja inimressurssi, et õpetada AI-d. Ainult AI-st sõltumine tekitab täiendavat riski. Kuna AI areneb, siis on vaja kvalifitseeritud tööjõudu, kes suudaks AI-d arendada ja kontrollida. Ilma kvalifitseeritud tööjõuta ei suudeta kasutada AI kogu potentsiaali ning võidakse teha vigu. **Tööprotsesside seadmine selliselt, et ainult AI otsuste pinnalt ei tehta otsuseid, ning neid otsuseid valideerib viimasena inimressurss. Oskusjõud valideerida ja jälgida AI kvaliteedimõõdikuid ning vajadusel neid peen häälestada.**
	AI sooritatud tegevuste ning langetatud otsuste eest vastutab AI treener, kes teda õpetas. Kui AI hakkab asutuse mainet kahjustama ning konfidentsiaalseid andmeid jagama, siis vastutab töötaja, kes AI-d õpetas. **Personali ressurss, et palgata AI-d tundvaid töötajaid. Vastutustundlik projekteerimine, arendus ja kasutuselevõtt, juurutajatele selge teave süsteemi vastutustundliku kasutamise kohta, juurutajate ja lõppkasutajate vastutustundlik otsuste tegemine, riskide selgitused ja dokumenteerimine, mis põhinevad juhtumite empiirilistel tõenditel.**
	Raske on tuvastada või parandada AI vigu või nõrkusi, mis ei ole hästi defineeritud või üheselt mõistetavad. Puudub vastav kompetents, kes suudaks vigu kiirelt märgata ning neid ennetada. **Personali ressurss, et palgata AI-d tundvaid töötajaid. Lisada inimkontroll, kui AI ei suuda ise vigu tuvastada või parandada. Teostada testimisi ning monitooringut. Domeenisisene testimine, reaalajas jälgimine ja võimalus sulgeda, muuta või lasta inimesel sekkuda süsteemidesse, mis erinevad kavandatud või oodatud funktsionaalsusest.**

AI-ga seotud riskid	AI-st sõltumise korral lähtutakse ainult AI toest ja selle puudumisel protsessid pidurduvad. Tekitab loovuse puudujääke ning vähenevad inimkompetentsid. Riski põhjustab asjaolu, et AI sooritab tegevusi ning teeb otsuseid teisiti kui inimene. **Alternatiivide omamine, et ei sõltuks ainult AI-st ja oleks võimalik vajadusel ka valideerida AI poolt tehtavat. Kvaliteedimõõdikute paika panemine ja jälgimine, mis võimaldab järjepidevalt mõõta AI sisulist toimimist ning anda alust tema peenhäälestamiseks.**
	Andmete kustutamises ei saa veenduda, kuna konto üleminekul on võimalik kontol olev andmestik kolmandale isikule edasi anda (nt juhile). Ei saa kontrollida, kas andmed on kustutatud. **Sätestada asutusesisene protseduur andmete kustutamise osas.**
	Kasutajate tehtud vead ja eksimused, mille läbi antakse AI-le töötlemiseks konfidentsiaalset teavet ning isikuandmeid. Sisestatakse andmeid, mis kahjustavad asutuse mainet ning tekitavad turvanõrkusi (võrgujoonised, riskid, protsesside kirjeldused jms). **Rakendada asutusesisesed protsessid ja poliitikad, mille kaudu kasutajal pole võimalik vastavaid andmeid sisestada. Koolitada kasutajaid ning koostada vastavaid juhendeid, kuidas AI-ga tööd teha. Privaatsuse tagamiseks tuleb rakendada asjakohast andmehaldust (nt pääsuprotokolle) ja koostada andmekaitsealane mõjuhinnang.**

Kokkuvõte

Anthropic Claude on rakendanud infoturbe ja andmekaitse alaseid meetmeid, millega saab lugeda Anthropic Claude teenus usaldusväärseks. Anthropic Claude andmete majutamise võimalus on Ameerika Ühendriikides, mistõttu pole andmekaitsenõuete täitmine tagatud.

Anthropic Claude pakub klientidele võimalusi, kuidas muuta kasutajate töö produktiivsemaks ja kulutõhusamaks. Anthropic Claude klientidel on võimalus seadistada Anthropic Claude pilvteenuse kasutamine turvaliseks. Samas on vajalik rakendada täiendavaid töökorralduslikke meetmeid ning luua asutusesiseseid protsesse, millega on tagatud turvaline AI kasutamine.

Lisaks tuleb tagada, et asutuse teenuste toimepidevus ei sõltuks ainult Anthropic Claude teenuse katkematust tööst.

Vajalik on rakendada täiendavaid meetmeid paralleelselt Anthropic Claude kasutamisega, mh koolitada kasutajaid, arendada alternatiivseid töömeetodeid, kontrollida Anthropic Claude kätte saadavaid andmeid jms.

Arvestada tuleks, et AI kasutamise reguleerimine EU tasandil on alles algusfaasis ning puuduvad ka regulatsioonid kohalikul tasandil, mis annaksid selgeid suuniseid AI kasutamiseks.

Asutus peab hindama, milliste kasutusjuhtude korral on Anthropic Claude sobiv kasutamiseks. Arvesse tuleks võtta AI kasutamisega ja pilvtoodetega seonduvaid riske ning Anthropic Claude poolt rakendatud meetmeid turvalisuse tagamiseks.