Texta

Kirjeldus

Käesolev usaldusväärsuse hinnang keskendub Texta pilvtöötlusteenuse riskide kirjeldamisele ning ei kohaldu Texta 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.

Texta on automaatne e-kirjade märgendaja (osades dokumentides nimega klassifitseerija). E-kirjade märgendamiseks loodud vahend (teenus), mis automaatselt tuvastab üldpostkasti laekunud kirjadest eeldefineeritud sarjadele ja märksõnadele vastavaid kirju ning märgib tuvastatud sarjad ja märksõnad e- kirja teemareale.
TEXTA OÜ on 2017. aastal loodud keeletehnoloogia ettevõte, mida juhivad selle asutajad Raul Sirel ja Silver Traat. Texta on esimene loomuliku keele töötlusele ja selle rakendamisele keskendunud keeletehnoloogia startup Eestis, mis on välja kasvanud STACCist. Ettevõtte tegevus on suunatud organiseerimata tekstiandmete probleemide lahendamisele andmepõhise tehisintellekti abil, Texta toodete tugevustena tuuakse mh välja võimekus hallata vastavusriske ning automatiseerida protsesse. TEXTA loodud Toolkit[1] on esimene tehisintellektil põhinev tarkvara - veebipõhise kasutajaliidesega tekstikaeveraamistik - mis on lisatud riigi koodivaramusse[2]. G2 portaali kliendid on hinnanud Texta.ai 4,2 punktiga viiest[3].

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

Töötlusjuhud

Klassifitseerimismudelite treenimise käigus sisestatakse treeningandmeid (pöördumiste (e-kirjade) näol) haldus- ja treeningkeskkonda.

Sisestatud andmete töötlemine mudelite treenimiseks, sh andmete koondamine, võrdlemine ning seoste ja mustrite tuvastamine.

Andmete (e-kirjade näol) sisestamine märgendamiseks ennustuskeskkonda.

Sisestatud andmete tekstianalüüs ja e-kirjade märgendamine, mille käigus on iga e- kiri töödeldud eraldi.

[1] https://docs.texta.ee/et/index.html
[2] https://docs.texta.ee/installation.html
[3] https://www.g2.com/products/texta-ai/reviews

Teenuse võimalused ja puudused

Texta pilvteenuseid majutatakse väljaspool Eesti territooriumi ning vajavad üldiselt toimimiseks püsivat Interneti ühendust. Ühenduseta Texta pilve pikaajalise katkestuse korral on vajalik kasutada alternatiivseid rakendusi.

Eesti riigiasutuste jaoks on oluline teabevahetuse korraldamine ka olukordades, kus välisühendused halvatakse kas pahatahtliku ründe tõttu või on sunnitud riik ennetava meetmena ise ühendused katkestama[4].

AI-süsteemidega peaks kaasnema automaatne DPIA, antud juhul puudub selgus, kus ikkagi treenitakse mudelit ning kas treenimisse on kaasatud avalikke pilvteenuseid {AWS). SMTP teenuse puhul ei pruugi olla tagatud edasisuunamise konfidentsiaalsus, käitajal tuleks tagada, et ei toimuks tagasilangust SMTPS teenuselt SMTP teenusele.

E-kirjade reaalajaliseks märgendamiseks on loodud TEXTA SMTP teenus, mida on võimalik kasutada igas SMTP protokolli toetavas meiliserveris. Selleks tehakse meiliserveris ümbersuunamine, mis suunab märgendamiseks valitud postkasti TEXTA SMTP teenusele, mis analüüsib ja liigitab ning märgendab sissetulnud e-kirja, lisab selle teemareale märgendi ning suunab kirja edasi valitud aadressile.

Pöördumiste märgendamise raames andmeid ei salvestata. TEXTA SMTP paigaldatakse eraldiseisva Dockeri ökosüsteemina, milles sisalduvad:

- TEXTA SMTP teenus,
- TEXTA mudelid,
- Apache Tika e-kirjade pärssimiseks.

TEXTA SMTP teenuse poolt kasutatavad mudelid luuakse/treenitakse TEXTA Toolkit keskkonnas, mida on võimalik paigaldada kas märgendajat käitava asutuse taristule või kasutada teenustaja taristul. Mudelite konfigureerimine, uuendamine produktsioonikeskkonnas toimub läbi skripti, mis pöördub TEXTA Toolkit API poole ning salvestab vajalikud mudelid failisüsteemi. Mudelite ümbertreenimine toimub kord aastas. Treenimine toimub asutuse infrastruktuuril, seega on välistatud andmete liikumine asutusest väljapoole.

Kuigi lahendus toetab hübriidformaati, eelistatakse siiski alati reeglipõhiseid mudeleid.

Texta Toolkit dokumentatsioon sisaldab juhiseid reeglipõhiste ja statistiliste mudelite loomise ja treenimise kohta[5].

Reeglipõhiseid mudeleid saab luua Regex Tagger rakenduses läbi GUI. Regex tagger ehk mustripõhine märgendaja kasutab tekstide mustripõhiseks märgendamiseks regulaaravaldisi. Regex Tagger põhineb Pythoni regexi moodulil ning seetõttu tuleb mustrite loomisel kasutada Pythoni regulaaravaldiste süntaksit[6].

Statistilised mudelid luuakse ja treenitakse Tagger rakenduses, info on leitav Texta Toolkit dokumentatsioonist. Tagger ehk märgendaja on ükskeelne binaarne tekstiklassifitseerija dokumendi märgendi ennustamiseks. Treeningandmestikuna saab kasutada nii eelnevalt salvestatud Texta Toolkit otsingut kui ka otse päringut Elasticsearchi (JSON objekt). Texta Toolkit taggerite treenimiseks kasutatakse teeki scikit-learn, mudelitest on valikus logistiline regressioon ja tugivektormasinad (support vector machines, SVM). Texta Toolkit jagab treeningandmed automaatselt treening- ja testandmestikuks (osakaaludega 80- 20) ning kasutab ruudustikotsingut koos k-kordse ristvalideerimisega, leidmaks mudelitele parimad hüperparameetrid.

SVM kasutatakse taggeri sees tunnuste selekteerimiseks, ebaoluliste tunnuste selekteerimiseks ning mudeli tihendamiseks. Tunnustena kasutatakse

[4] https://www.aki.ee/sites/default/files/ringkirjad/andmetootlusest_avalikes_pilveteenustes_0.pdf

[5] https://docs.texta.ee/et/index.html

[6] https://docs.texta.ee/et/tagger.html

nii sõnu kui ka tähemärgipõhiseid mitmikuid (n-gramme)[7].

AI süsteemi andmekaitsealastele nõuete tagamiseks peab võtma arvesse IKÜMi artikli 5 lõikes 1 kehtestatud isikuandmete töötlemise põhimõtteid, mille täitmise eest vastutab ja peab olema võimeline nõuete täitmist tõendama vastutav töötleja (IKÜM artikkel 5(2)).

USA-s tegutsevatele isikutele ja asutustele ja/või USA-s asuvatesse andmekeskustesse isikuandmete edastamine ei ole üldjuhul lubatud, sest Euroopa Liidu ja Ühendriikide vahel ei ole alates 2020. aasta juulikuust[8], mil Euroopa Kohus tunnistas kehtetuks *Privacy Shield'i* nimelise andmekaitseraamistiku, toimivat andmekaitsealast koostööd ning andmete edastust. Sellega seoses ei ole andmesubjektidele USA-s toimuva andmetöötluse osas tagatud samaväärsed õigused (ei pruugi olla tagatud turvameetmed, võidakse teostada andmekorjet või edastada andmeid kolmandatele isikutele, nt järelevalveasutused, vt ka *CLOUD Act*[9]), nagu kehtivad Euroopa Liidus toimuva andmetöötluse suhtes. Varasemad lepped EU ja USA vahel on korduvalt õigustühiseks tunnistatud. Töö uue vastava andmekaitseraamistiku loomise nimel käib siiski aktiivselt ja selle heakskiitmist võib prognoosida 2024. aasta jooksul. 2023 aastal vastu võetud *Data Privacy Frameworki* kohaldami se osas töö käib ja selle funktsionaalsust hinnatakse perioodiliselt (kontrollitakse, kas kõik asjakohased meetmed on rakendatud ning toimivad praktikas, et kaitsta isikuandmete edastamist EU-USA vahel).[10]

Soovitus: kontrollida, kas teenust treenitakse kliendi andmetega

EL tehisintellekti määruse kohane riskitase

Teenuse kasutamine vastavalt kirjeldatud tootlusjuhtudele klassifitseerub võib klassifitseeruda nii minimaalse kui ka piiratud riskiga tehisintellektiks.

Tootja väitel on klassifitseerija arendamisel tagatud, et treenitud liigitamismudelid ei sisalda isikuandmeid üheski etapis. Samuti ei tee klassifitseerija klassifitseerimismudelite treenimisel mistahes toiminguid andmesubjekti isiku tuvastamiseks ega isiku identiteedi või muude eraeluliste asjaolude kohta järelduste tegemiseks.

Ennustuskeskkonnas töötleb klassifitseerija iga pöördumist eraldi ning ei koonda, võrdle ega seosta andmeid erinevate pöördumiste vahel ega tee selle pinnalt järeldusi. Klassifitseerija ei analüüsi andmeid andmesubjekti isiku tuvastamise eesmärgil ega tee isiku identiteedi kohta mistahes järeldusi.

Andmekaitsenõuetes kohustab tootja kasutusele võtjat rakendama meetmeid tehisintellektil ja masinõppel põhinevaid süsteeme ohustavatest rünnetest tulenevate ohtude maandamiseks, ent täpsemaid juhiseid selleks ei anna.

Andmekaitsenõuded mainivad ka märgendaja {klassifitseerija) disaini ja arhitektuuri sisse ehitatud turvameetmeid, ent ei täpsusta neid. Texta ToolKit on avatud lähtekoodiga[11].

E-kirjade subjekti rida võib sisaldada isikuandmeid, kuid saatjatele ei saa öelda, et nad sinna midagi tundlikku ei paneks.

Skoor: kontrollida, milliseid meetmeid rakendatakse

[7] https://docs.texta.ee/et/regex_tagger.html

[8] EKo 16.07.2020, C-311/18 – Data Protection Commissioner vs Facebook Ireland Ltd, Maximillian Schrems

[9] https://en.wikipedia.org/wiki/CLOUD_Act

[10] https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/eu-us-data-transfers_et

[11] https://docs.texta.ee/et/index.html

Rahaline mõju

Loodud lahendus põhineb avatud lähtekoodiga vabavaralisel TEXTA Toolkitil, mis on litsentseeritud GPLv3 alusel. Seetõttu ei kaasne TEXTA Toolkiti ega TEXTA SMTP lahenduse kasutamisega litsentsikulusid.

Produktsioonikõlbuliku lahenduse väljatöötamiseks kulus 580 tundi täistööaja ekvivalendis. 2024. aastal oli mudelite treenimise ja produktsioonis kasutatavate mudelite uuendamise poolaastatasu 1500 eurot ilma käibemaksuta.

Texta 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 Texta kasutamise osas.

Soovitus: kontrollida hinnakirja ning kaasnevaid kulusid

EL/NATO liikmesriigis hoitavad andmed

E-kirjade märgendaja kasutatavad alusmudelid on majutatud Texta 0Ü taristul. Kuigi Texta ei avalikusta, milliste andmekeskuste teenuseid kasutatakse, on siiski 4 aastat tagasi avaldatud Texta töökuulutuses mh mainitud Amazon AWS, mis võib kaudselt viidata selle taristu kasutusele.

Andmeid on võimalik hoida AWS-i erinevates regioonides. Siiski ei ole võimalik olla 100% veendunud, et kõik kliendi andmed (sh metaandmed) asuvad igal ajahetkel Euroopas. GDPR-i kohaselt võib andmeid edastada väljapoole Euroopa Liitu kui on rakendatud asjakohased kaitsemeetmed (artikkel 46[12]).

Andmete asukoht sõltub seega teenuse infrastruktuurist ja Texta serverite asukohtadest.

Soovitus: kontrollida, kus andmeid majutatakse

Teenusest lahkumise ja andmete ekspordi võimalus

Kuivõrd andmete töötlemisel märgendajaga (sh nii klassifitseerimismudeli treenimise raames kui ka pöördumiste (e-kirjade) märgendamise raames) tuginevad vastutavad töötlejad õigusliku alusena seadusest tuleneva kohustuse täitmisele, siis andmete ülekandmise õigus ei kohaldu. Andmesubjekti õigus nõuda andmete kustutamist tagatakse võimalusega põhjendamatu viivituseta pöördumatult kustutada märgendaja treeningandmete haldus- ja treeningkeskkonnas salvestatud andmed.

Andmeid töödeldakse ja säilitatakse vastavalt teenuse kasutamisele ning vajadusele.

Andmekao vältimiseks peaks andmete omanik hindama Texta pilvteenuses olevat andmete koosseisu ning nende ajaloolist väärtust, et hinnata, kas andmete eksportimine on vajalik ja otstarbekas.

Konkreetsed käideldavusaspektid sõltuvad sõlmitud kokkulepetest.

Soovitus: kontrollida, kas andmeid kustutatakse peale teenuse lõpetamist

[12] https://eur-lex.europa.eu/legal-content/ET/TXT/HTML/?uri=CELEX:32016R0679&qid=1689851996164

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

AI tehnoloogia rakendamisel tuleb lisaks seadusele järgida ka küberturbe ja ühiskondliku ohutuse nõudeid[13].

Avalikes allikates ei leidu infot teenustaja või Texta teenuste sertifikaatide kohta.

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[14].

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)[15]. 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 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[16].

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

[13] https://www.ria.ee/sites/default/files/documents/2024-03/Tehisintellekti-masinoppe-tehnoloogia-riskide-uuring-2024.pdf
[14] https://www.ria.ee/sites/default/files/documents/2024-03/Tehisintellekti-masinoppe-tehnoloogia-riskide-uuring-2024.pdf
[15] https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf
[16] https://www.ria.ee/sites/default/files/documents/2024-03/Tehisintellekti-masinoppe-tehnoloogia-riskide-uuring-2024.pdf

publikatsioonist heade küberturvalisuse praktikate kohta AI süsteemide puhul[17].

Soovitus: kontrollida, kas teenusepakkujal on kehtivad sertifikaadid

Andmekaitse meetmete rakendamine

Puudub info, kas Texta arendatud märgendajaga seonduva andmetöötluse (majutus, mudeli treenimine) osas on sõlmitud kehtiv andmetöötlusleping või mingil muul kujul kehtivad andmekaitset reguleerivad dokumendid.

Lepingus sisalduvad sätted reguleerivad isikuandmete töötlemist vaid lepingu täitmise käigus.

E-kirjade klassifitseerijaga isikuandmete töötlemise andmekaitsenõuete kohaselt kasutatakse mudeli treenimiseks isikuandmeid (ei ole välistatud eriliiki isikuandmete töötlemine), kuid treenitud mudelid ei sisalda isikuandmeid. Treenimine toimub haldus- ja treeningkeskkonnas.

Treenimise osas on märgitud, et „klassifitseerija ei tee klassifitseerimismudelite treenimisel mistahes toiminguid andmesubjekti isiku tuvastamiseks ega isiku identiteedi või muude eraeluliste asjaolude kohta järelduste tegemiseks." Vajab täpsustamist, et kas juhul, kui selgeid isikuandmeid ei ale, ei tee märgendaja täiendavaid pingutusi isiku tuvastamiseks või on mõeldud midagi muud.

Lisaks võib märgendaja tegeleda profileerimisega, kuivõrd on välja toodud, et töötlemise käigus võib muuhulgas toimuda andmete koondamine, võrdlemine ning seoste ja mustrite tuvastamine. Samas on märgitud, et ennustuskeskkonnas vaatab märgendaja iga pöördumist eraldi. Täpsustada tuleb, kas mingil hetkel võib siiski toimuda profileerimine

Seejuures annab eelviidatu mõningal määral aimu märgendaja lahenduse vastavusest andmekaitsenõuetele, kuid mitmed aspektid sõltuvad siiski märgendajat kasutavast asutusest. Samuti puuduvad viited andmekaitselepingule (GDPR artikkel 28 (3) kohane leping).

Texta kasutab tõenäoliselt alltöötlejaid, nt majutuse jaoks, kuid nimekirja neist ei ole Texta avalikustanud.

Märgendaja kasutamisele seatud nõuded on loetletud e-kirjade klassifitseerijaga isikuandmete töötlemisele kohalduvad andmekaitsenõuded:

- Vastutava töötlejatena peab asutus määratlema õigusliku aluse asjakohaseks andmete töötlemiseks.
- Peab olema (võimalusel ka märgendaja arendamisel) välistatud võimalus kasutada treenitud klassifitseerimismudeleid muudel eesmärkidel kui asutustele edastatud pöördumiste lahendamisel ja nendele vastamisel pöördumiste märgendamiseks ning igal juhul ei tohi andmetega treenitud klassifitseerimismudeleid kasutada konkreetsete füüsiliste isikute kohta mistahes järelduste tegemiseks.
- Peab olema tagatud, et klassifitseerimismudelite treenimiseks kasutatavad pöördumised (hõlmates neis sisalduvaid andmeid) ei oleks vananenud (nt asutuste pädevuse või struktuuri muutumise korral muudatusele eelnenud pöördumiste kasutamine) ega valed (nt asutuse töötajale edastatud e-kirjad, mis sisaldavad isiklikke sõnumeid, mitte pöördumist asutuse poole, asutuste siseselt omavahelised sõnumid vms).

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

- Tuleb mudelite treenimiseks andmete töötlemise raames välistada andmete väärkasutamise võimalus (nt võimalus kasutada treeningandmeid muudel eesmärkidel või teistel eesmärkidel kasutatavate mudelite treenimiseks) ning tagada, et töötlemine oleks andmesubjektidele läbipaistev.

- Vastutava töötlejana peab asutus avaldama teabe andmete töötlemise kohta.

i) klassifitseerimismudelite treenimise raames ja

ii) pöördumiste märgendamise raames. Teave tuleb andmesubjektidele avaldada enne konkreetse töötlemise alustamist.

- Klassifitseerija süsteemis peab olema tagatud pöördumiste ja nendes sisalduvate andmete pöördumatu kustutamise võimalus parast asutuse poolt säilitamise tähtaja möödumist.

- Asutus, kelle haldusalas teenuse serverid asuvad, peab tagama nii andmete kui süsteemi tervikluse ja jätkusuutlikkuse.

- Vastutav töötleja (märgendajat kasutav asutus) peab olema võimeline tõendama andmekaitsenõuete täitmist, selleks on vajalik logida klassifitseerija tehtud andmetöötlustoimingud (treeningandmete saamine, muutmine, vaatamine, edastamine, kustutamine; treenimise ja treeningandmete seas jm).

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, kas teenusepakkuja kasutab alltöötlejaid

Turvameetmete rakendamine

Juhul kui nii iga-aastane mudelite treenimine kui igapäevane e-kirjade märgendamine toimuvad mõlemad asutuse {kliendi) infrastruktuuril, ei liigu e-kirjad asutusest väljapoole. Sellisel juhul peab meetmed nii andmete kui paigaldatud süsteemi konfidentsiaalsuse, tervikluse, käideldavuse ja kerksuse (vastupidavuse/jätkusuutlikkuse) tagamiseks rakendama märgendajat kasutav asutus {vastutav töötleja ja võimalik{ud) volitatud töötleja{d)), mitte tootja.

Muuhulgas peaksid nii märgendaja kasutuselevõtja kui võimalikud volitatud töötlejad rakendama meetmeid järgnevaks:

- takistada volitamata juurdepääs treeningandmetele ning nendega manipuleerimise võimalus (sh volitamata andmete sisestamise, tutvumise, muutmise, kopeerimise ja kustutamise võimalus);
- takistada klassifitseerija kasutamist andmesidevahendite abil volitamata isikute poolt;
- volitatud kasutajatele tagada juurdepääs üksnes volituste piires;
- andmetega tehtud toimingute logimine;
- tagada süsteemi toimimine ja ilmnevatest käiduvigadest teavitamine;
- tagada võimalus vajadusel andmeid ja süsteemi taastada;
- välistada ja takistada andmete moonutamine süsteemirikete tagajärjel;
- võtta tarvitusele meetmed rünnetest tulenevate ohtude maandamiseks, sh tehisintellektil ja masinõppel põhinevatele süsteemidele iseloomulikke ründeid arvestades.

Treenimiskeskkond Texta Toolkit toetab kasutajate identiteedihaldust {autentimine kas kasutajanimi- parool või kolmanda osapoole {CloudFoundry) abil[18].[19] Märgendajat käitava asutuse ülesanne on välja vahetada Texta Toolkiti

[18] https://docs.texta.ee/authentication.html

[19] https://docs.texta.ee/installation.html

administraatori parool ning korraldada kasutajate volitamine.

Soovitus: kontrollida, kas teenusepakkuja rakendab turvameetmeid

Erinõuded

Nõuded teenustajale. Piiratud riskiga süsteemide puhul tuleb tagada süsteemi läbipaistvus.

Nouded kasutusele votjale. Lahenduse kasutamisel tuleb tagada märgendaja poolt andmete töötlemise logimine. Eelkõige peaksid logid võimaldama tuvastada/jälgida:

- treeningandmete sisestamist, muutmist, vaatamist, edastamist ja kustutamist märgendaja treeningandmete haldus- ja treeningkeskkonnas;
- liigitusmudelite versioonide seost treeningandmetega (võimaldaks vajadusel tuvastada, millised mudelite versioonid on treenitud manipuleeritud treeningandmetega);
- pöördumiste märgendamise toiminguid märgendaja ennustuskeskkonnas;
- süsteemi väärkasutuse või sissetungikatseid.

Terviklus tuleb tagada süsteemi käitajal (märgendajat kasutaval asutusel).

Märgendaja tehtud töötlustoimingute logide tõendusväärtuse tagamiseks soovitab tootja süsteemi kasutajal varustada need digitaalse ajatempliga. Logikirjete säilitamise tähtajad ning reeglid logide kontrollimiseks ja nendega tutvumiseks tuleb kehtestada igal asutusel vastavalt organisatsiooni eripäradele.

Soovitus: kontrollida, kas logid vastavad tegelikkusele

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.**
	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-ga seotud riskid	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-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

Texta OÜ on rakendanud infoturbe ja andmekaitse alaseid meetmeid, millega saab lugeda Texta teenuse usaldusväärseks. Samuti on Textal andmete majutamise võimalus Euroopa Liidus, millega on formaalselt tagatud andmekaitsenõuete täitmine.

Texta pakub klientidele võimalusi, kuidas muuta kasutajate töö produktiivsemaks ja kulutõhusamaks. 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 Texta teenuse katkematust tööst.

Vajalik on rakendada täiendavaid meetmeid paralleelselt Texta kasutamisega, mh koolitada kasutajaid, arendada alternatiivseid töömeetodeid, kontrollida Textale 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 Texta sobiv kasutamiseks. Arvesse tuleks võtta AI kasutamisega ja pilvtoodetega seonduvaid riske ning Texta poolt rakendatud meetmeid turvalisuse tagamiseks.