TARTU ÜLIKOOL
Sotsiaalteaduste valdkond
Ühiskonnateaduste instituut
Ühiskonna ja infoprotsesside analüüsi õppekava

Heli Orav

Tehisintellekte puudutava läbipaistvuse käsitlemine Põhja- ja Baltimaade strateegiates ning Eesti avaliku sektori ekspertide arusaamad sellest

Magistritöö

Juhendaja:
Maris Männiste, MA

Tartu 2021

SISUKORD

SISSEJUHATUS

Tehisintellekt ühendab tehnoloogialiike, milles põimuvad andmed, algoritmid ja andmetöötlusvõimsus. Euroopa Liidu tehisintellekti Valge raamatu järgi (2020) peetakse tehisintellekti andmepõhise majanduse üheks kõige olulisemaks väljundiks. Enamus me kõik kasutame tehisintellekti või algoritmidel põhinevaid tooteid ja teenuseid. Tegemist on ühe olulise tehnoloogia valdkonnaga, mille rakendamisel on võimalus kasvatada ettevõtete loodavat lisandväärtust ja tõhustada avaliku sektori tööprotsesse (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020). Algorithm Watch´i raportis "Automating Society" (2020) on Eesti kohta välja toodud asjaolu, et viimastel aastatel on valitsus astunud mitmeid samme tehisintellekti rakendamise toetamiseks nii avaliku kui ka erasektori asutustes. Eesti avalikus sektoris ongi juba rakendatud mitmed tehisintellektid ehk kratid. Nad küll ei võta otsuseid veel iseseisvalt vastu, aga siiski on inimese abilised kiiremate ja paremate otsuste tegemisel (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020).

Algorithm Watch´i raporti järgi (2020) arutatakse Eestis automatiseeritud otsuste või tehisintellekti kasutamise üle, keskendudes praegu peamiselt nendelt tehnoloogiatelt saadavale kasule. Seda toetab Euroopa põhiõiguste ameti aruanne FRA (2020), mis toob välja, et valitsused ja ettevõtted kiirustavad omaks võtma tehisintellekti potentsiaalseid eeliseid. Samas on mitmel Euroopa Liidu liikmesriigil, kes kasutavad tehisintellekti julgeoleku valdkonnas ja sotsiaalmajanduslikes sektorites, olnud suuri raskusi tehnoloogia läbipaistvaks muutmisel (Euroopa Liidu Põhiõiguste Amet, 2020). Seega Euroopa Liit ja selle liikmesriigid peaksid tagama, et tulevastes ja praegustes tehisintellekti arendamise eeskirjades käsitletaks põhjalikke ja läbipaistvaid põhiõiguste mõjuhinnanguid. Sellele lisaks on oluline sõltumatute järelevalveasutuste tehtav järelevalve, et tagada vastutus, usaldusväärsus ja õiglus (Euroopa Liidu Põhiõiguste Amet, 2020). Algorithm Watch´i raportis "Automating Society" Soome (2020) kohta on rõhutatud, kui oluline on läbi viia uuringuid, et hinnata automatiseeritud süsteemide ehk täielikult või iseseisvalt otsuseid tegevate süsteemide kasutamisele võtmisega kaasnevaid probleeme. Oluline on tehisintellekti, algoritmide ja eetika üle peetav arutelu arenduste faasis ning

oluline on kaasata nendesse diskussioonidesse ka ettevõtted, kes süsteeme arendavad. Sellist dialoogi saab arendada avatust hindavas kultuuris, kus automatiseeritud süsteemide arendajad tunnevad, et neid ei rünnata, vaid et nad on käimasoleva ühiskondliku arutelu oluline osa (Algorithm Watch & Bertelsmann Stiftung, 2020).

Seega nagu iga uue tehnoloogia puhul, kaasnevad ka tehisintellekti kasutamisega nii uued võimalused kui ka uued riskid (Euroopa Liidu tehisintellekti Valge raamat, 2020). Eesti digiühiskonna 2030. aasta valdkonna arengukava tööversioonis on toodud välja proovikivi, milleks on asjaolu, et universaalsete ekspertide abil ei saa enam piisavalt lahendada komplekssseid riske, mida tekitab pilveandmetöötluse, tehisintellekti, krüptograafia, kvantarvutite, asjade interneti, liitreaalsuse, robootika jne võidukäik (Majandus- ja Kommunikatsiooniministeerium, 2021). Tehisintellektist võib olla abi kodanike turvalisuse kaitsel ja nende põhiõiguste tagamisel, siiski teevad ka kodanikele muret tehisintellekti võimalikud soovimatud tagajärjed või isegi selle potentsiaalne kuritahtlik kasutamine (Euroopa Komisjon, 2020). Ethan Fast ja Eric Horvitz (2017) on analüüsinud tekstikorpuseid, et mõista, kuidas on inimeste huvi, suundumused ja veendumused tehisintellekti kohta aja jooksul muutunud. Analüüsi võeti viimase 30-e aasta jooksul kirjutatud artiklid väljaandest *The New York Times*. Tulemustest selgus, et arutelu tehisintellekti üle on alates 2009. aastast järsult kasvanud ja olnud pigem optimistlikum kui pessimistlikum. Siiski on viimastel aastatel rohkem artiklites tõstatud konkreetseid probleeme nagu hirm selle ees, et tehisintellekti üle kaotatakse kontroll ning tõstatatud on ka mure tehisintellekti eetilise otsustamise puuduste pärast, mis võib kaasa tuua negatiivseid tagajärgi nagu oht inimelule (Fast & Horvitz, 2017). Selleks, et mõista, kust tehisintellektiga seonduvad probleemid tulevad ning leida neile lahendused, räägitakse tihti seda tüüpi süsteemide läbipaistvaks muutmisest (Kemper & Kolkman, 2019; Larsson & Heintz, 2020). Tehisintellekti läbipaistvuse juures peab silmas pidama, et ta on oma olemuselt nn „must kast“ (Larsson & Heintz, 2020; Ananny & Crawford, 2018; Pasquale, 2015), mille toimimist on raske seletada, kuna tihti ei tea ka arendajad, millisel viisil ja milliseid andmeid tehisintellekt iseseisva otsustamise ja tegutsemise jaoks kasutab. Autorid (Larsson & Heintz, 2020; Ananny & Crawford, 2018; Pasquale, 2015) toovad välja selle, et kuigi läbipaistvus on oluline, ei kiputa seda poliitikadokumentides defineerima ning jääb arusaamatuks, millisel moel läbipaistvust saavutada. Euroopa komisjon kutsus kokku kõrgetasemelise eksperdirühma, mis avaldas suunised usaldusväärse tehisintellekti kohta (Euroopa Komisjon, 2020). Ka Eesti tehisintellekti kasutuselevõtu ekspertrühma aruandes (2020) on sätestatud komisjoni poolt antud suunised, mille kohaselt peab Eestis krattide ehk tehisintellekti usaldusväärsuse saavutamiseks olema täidetud kolm tingimust: kratt peab 1) vastama seadusele, 2) olema kooskõlas eetika

põhimõtetega ja 3) olema töökindel. Nendele tingimustele vastamiseks on omakorda välja toodud seitse põhinõuet, mis kujundavad usaldusväärse krati kontseptsiooni ning üheks nendest on läbipaistvus. Kuid mida täpsemalt läbipaistvuse mõiste endast kujutab seda tegelikult aruanne välja ei too.

Käesoleva magistritöö eesmärgiks on välja selgitada, kuidas on läbipaistvus sõnastatud dokumentides ja kuidas mõistavad seda tehisintellekti kontekstis valdkonna eksperdid. Ühelt poolt käsitleb uurimus, kuidas on läbipaistvust tõlgendatud Põhja- ja Baltimaade tehisintellekti strateegiates. Esimeseks uurimisküsimuseks on käesolevas magistritöös seega: **Kuidas on tehisintellekti läbipaistvus sõnastatud Põhja- ja Baltimaade riiklikes strateegiates?** „Nordic Co-operation" ühenduse järgi on Põhja- ja Baltimaadel avalikus- ja erasektoris tihe riikidevaheline digitaalne koostöö, mille hulka kuulub ühine poliitiline dialoog ja digitaalsed innovatsioonialased ühised algatused. Ühenduse eesmärk on tugevdada piirkonna digitaalset rolli Euroopas ja kogu maailmas (Nordic Co-operation, 2021). Tehisintellekti rakenduste kasutuselevõtt on kõrge prioriteediga aruteluteemaks Põhjala-Balti koostöögrupis, mis koordineerib regiooni arengusuundasid (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020). Varasemalt on Stephen Cory Robinson (2020) uurinud Põhjamaid ning kuidas nende ühised väärtused usaldusest, läbipaistvusest ja avatusest kanduvad üle tehisintellekti riiklikesse dokumentidesse. Viidatud uuringus selgus, et on ilmne, et kõik Põhjamaade poliitilised juhised sisaldavad kultuurilisi väärtusi usaldusest ja läbipaistvusest. Lisaks kerkivad Põhjamaade riiklikes tehisintellekti poliitikates teemadena ka privaatsus, demokraatia, eetika ja autonoomia. Oluline on uurida läbipaistvust süviti, mida täpsemalt mõistetakse antud mõiste all ning leian, et oluline on ka lisaks Põhjamaadele uurida ka Baltimaid. Magistritöö tarbeks uurin seega tehisintellekti läbipaistvuse kajastamist Põhjamaade ja Balti regiooni tehisintellekti strateegiates kaasates järgmisi riike: Eesti, Soome, Rootsi, Norra, Island, Läti ja Leedu.

Tehisintellekti hõlmava arenduse algusfaasis peaks kõlapinda saama laiaulatuslikum arutelu tehisintellekti, algoritmide ja eetika üle ning oluline on kaasata arutellu ka tehisintellektide arendajad (Algorithm Watch & Bertelsmann Stiftung, 2020). Tihti kajastuvad infotehnoloogia vallas töötavate eetilised tõekspidamised ka näiteks nende kirjutatavas koodis (Ananny, 2016; O'Neil, 2016; Eubanks, 2018; Noble, 2018), mis võib saada tulevase tehisintellekti osaks. Oluline on mõista, mida peavad silmas eksperdid tehisintellekti läbipaistvuse all, kuna nende tõlgendused sellest võivad kajastuda tehisintellektide arendustes. Järgnevateks uurimisküsimusteks on seega: **Millised on arusaamad tehisintellektist infotehnoloogia vallas töötavate ekspertide seas?**

Millised on arusaamad tehisintellekti läbipaistvusest infotehnoloogia vallas töötavate ekspertide seas? Magistritöö tarbeks viiakse seega läbi poolstruktureeritud intervjuud just Eesti avaliku sektori infotehnoloogia vallas töötavate ekspertidega.

Magistritöö koosneb teoreetilisest raamistikust, kus esmalt selgitan erinevatele autoritele tuginedes tehisintellekti rollist andmestunud ühiskonnas. Teoreetilises raamistikus seletan lahti ka tehisintellekti olemuse ja kasutusalad ning kirjutan läbipaistvusest andmestumise ja tehisintellekti kontekstis. Uurimismetoodika osas kirjutan, milliseid uurimismeetodeid ma käesolevas magistritöös kasutasin, millele järgneb valimi ja andmeanalüüsi kirjeldus. Järgmises peatükis toon analüüsist lähtuvalt välja tulemused ning magistritöö lõpus on välja toodud järeldused ning kokkuvõte.

Tahan tänada oma juhendajat Maris Männistet, kes igati toetas ja aitas mind magistritöö valmimisel nõuannetega ning retsensenti väärtusliku sisendi eest. Samuti soovin tänada intervjueeritavaid, kes leidsid aega mu uurimuses osaleda ning andsid seeläbi panuse käesoleva uurimuse valmimiseks.

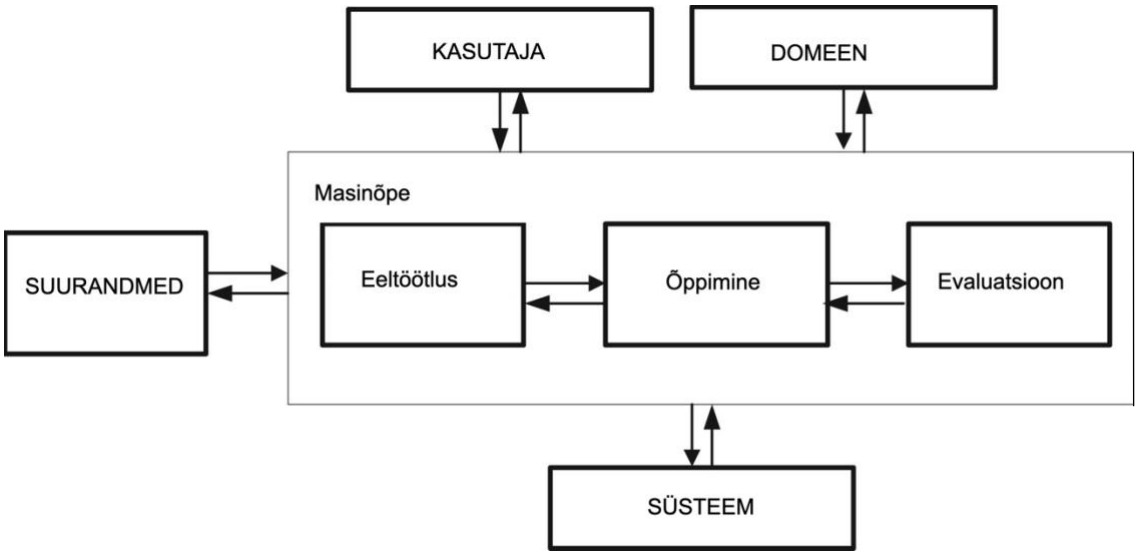
1. TEOREETILINE RAAMISTIK

Käesolevas peatükis seletan kirjandusele tuginedes lahti, kuidas on tehisintellekt seotud andmestunud ühiskonnaga. Samuti kirjutan tehisintellekti olemusest ning seejärel läbipaistvusest nii andmestumise kui tehisintellekti kontekstis.

1.1 Tehisintellekt andmestunud ühiskonnas

Andmestumise termin annab märku sellest, et kõik inimese igapäevased tegevused on võimalik muuta andmeteks (Mejias & Couldry, 2019). Andmete kogumisel ei saa lahutada olulisi aspekte nagu väline infrastruktuur, mille kaudu andmeid kogutakse, töödeldakse ja säilitatakse. Teiselt poolt on oluline väärtuse loomine protsessidest, mis hõlmavad andmete rahaliseks väärtuseks muutumist, riiklikku kontrolli, kultuuriloomet ning kodanikuõiguste suurendamist (Mejias & Couldry, 2019). Andmestumise läbi kogutakse suures koguses igapäevaelu teavet, et muuta see arvutipõhiseks masinloetavateks andmeteks (Strauß, 2015). Andmestumisest räägitakse tihti kui suurandmete kogumisest. Suurandmete puhul peab silmas pidama, et tihti on tegemist üles kiidetud terminiga (Strauß, 2015; Gandomi & Haider, 2015). Suurandmete analüüsis on oluline seada fookus mitte ainult suurte arvudega kaasnevale maagiale, vaid ka uurimis- ja teabeprotsessile ja selle komponentidele, st mis teavet kasutatakse suurandmete jaoks, millistest allikatest, mis eesmärgil see koguti, mis on analüüsi eesmärk (mida tuleks uurida), kas tulemus on usutav või mitte, mis on (eba)usaldusväärsuse põhjused, millist lisateavet analüüs näitas, millised on tulemuste piirid ja sarnased küsimused (Strauß, 2015). Gandomi & Haider (2015) järgi on suurus ainult üks suurandmete mitmest mõõtmest. Suurandmete määratlemisel on sama olulised ka muud omadused, näiteks andmete genereerimise sagedus. Tuleks arvesse võtta ka erinevat tüüpi suurandmeid nagu teksti, heli, videot ja sotsiaalmeedia sisu (Gandomi & Haider, 2015). Seega kirjeldatakse suurandmeid tihti läbi kolme tunnuse: variatiivsus, suurus ja kiirus (*variety, volume, velocity*). Alvarez et al., (2021) järgi tähendab variatiivsus suurandmete puhul erinevaid andmetüüpe, mida on võimalik koguda. Suurus on seotud andmete kogusega, mida saab kindlaks määratud ajavahemikul koguda ja kiirus viitab andmete kogumise, salvestamise ja kasutamise kiirusele (Alvarez et al., 2021).

Asjade internetiks (*IoT*) nimetatud internetiühendusega seadmete kasutamise kasv on viimase kümne aasta jooksul suurenenud ning võimaldanud koguda aina suuremaid andmemahte (Alvarez et al., 2021). Asjade internetis suhtlevad paljud võrgu- ja töötlemisvõimalustega seadmed omavahel interneti kaudu lokaalselt või eemalt (Zeadally et al., 2019). Tuginedes masinõppe lähenemisviisidele kasutab tehisintellekt asjade interneti (*IoT*) või muude suurte andmeallikate kaudu saadud välist teavet sisendina reeglite ja mustrite kindlaks tegemiseks (Kaplan & Haenlein, 2019). Muhammad & Yan (2015) järgi on masinõppe üks põhieesmärke juhendada arvuteid andmete põhjal probleemi lahendama. Juba praegu on olemas arvukalt edukaid masinõppe rakendusi, näiteks klassifikaator, mida saab treenida e-kirjade pealt õppimiseks, et eristada rämpsposti; süsteemid, mis analüüsivad varasemaid müügiandmeid, et ennustada klientide ostukäitumist jne. Klassifitseerimine ja ennustamine on kaks andmeanalüüsi vormi, mida saab kasutada olulisi andmeklasse kirjeldavate mudelite loomiseks või andmete tulevaste suundumuste ennustamiseks. Klassifikatsioonitehnikaga on võimalik töödelda suuremat sorti ja mitmesuguseid andmeid. Klassifitseerimine on mudeli loomise protsess teatud siltidega treeningandmestikust. Saadud mudeli abil prognoositakse seejärel testjuhtumi silti, kus ennustavate tunnuste sisendväärtused on teada. Juhendatud klassifikatsioon on üks ülesannetest, mida intelligentsed tehnikad kõige sagedamini täidavad (Muhammad & Yan, 2015). Suurandmed pakuvad masinõppe algoritmide jaoks enneolematut rikkaliku teavet mudelite loomiseks (Zhou et al., 2017). Zhou et al., (2017) on koostanud raamistiku kirjeldamaks, kuidas masinõppes suurandmeid kasutatakse (vt Joonis 1).



Joonis 1. Masiõpe suurandmete pealt (*the framework of machine learning on big data*). Allikas: Zhou et al., 2017 (eesti keelde tõlgitud minu poolt).

Zhou et al., (2017) joonis kirjeldab, et suurandmed on masinõppes kirjeldatud läbi nelja komponendi, mille hulgas on suurandmed, kasutaja, domeen ja süsteem ning nende koostoimed toimuvad mõlemas suunas. Näiteks on suurandmed sisendiks masinõppele, mis genereerib väljundeid, mis omakorda muutuvad suurandmete osaks. Kasutaja võib suhelda masinõppega, pakkudes domeeniteadmisi, isiklikke eelistusi ja kasutatavuse tagasisidet ning võimendades õpiväljundeid otsuste tegemise parandamiseks. Domeen võib toimida nii teadmiste allikana masinõppe juhtimisel, kui ka õpitud mudelite rakendamise kontekstina. Süsteemi arhitektuur mõjutab seda, kuidas õppimisalgoritmid peaksid töötama ja kui tõhus on neid käivitada, ning seeläbi täiendatakse omakorda süsteemi arhitektuuri, et vastata masinõppe vajadustele.

Elish ja boyd (2018) järgi on tehisintellekti termini kasutamine välja kasvanud andmestumisest. Tehnoloogiaettevõtted, mida kunagi peeti andmestumise tehnoloogiate eestvedajateks, hakkasid oma jõupingutusi tehisintellektiks muutma (Elish & boyd, 2018). Sealhulgas ei ole andmestumise ja tehisintellekti taga olevad masinõppe ja andmeteaduse protsessid oma olemuselt uudsed ning neid on arendatud ja kasutatud aastakümneid. Näiteks andmestumise termini üles kiitmise tõttu hakati igasugusust andmeanalüütilist praktikat nimetama andmestumiseks, hoolimata selle tehnilisest keerukusest (Elish & boyd, 2018). Danah boyd ja Kate Crawford (2012) järgi toetuvad suurandmed tehnoloogiale, analüüsile ja metodoloogiale. Levinud on veendumus, et suured andmekogumid pakuvad metodoloogiliselt kõrgemat vormi intelligentsust ja teadmisi, mis loovad omakorda teadmisi, milleni jõudmine oli varem võimatu, tõe, objektiivsuse ja täpsusega (boyd & Crawford, 2012). Kas see veendumus peab andmestumise puhul paika? Strauß (2015) näeb andmestumist kui tehnoloogiat, mis on andnud vahendeid poliitiliste, majanduslike, sotsiaalsete ja keskkonnaalaste mustrite, suundumuste ja suhete tuvastamiseks, kuid toob sinna juurde, et seda tehakse ilma hüpoteeside või mudeliteta. Boyd ja Crawford (2012) on toonud välja, kuidas paljudes suurandmeid hõlmavates debattides jäetakse suuremahulistele numbritele tuginedes liiga kergekäeliselt kõrvale muud meetodid välja selgitamaks, miks inimesed teevad või kirjutavad miskit. Tuleb esitada keerukaid küsimusi suurandmete mudelite arusaadavuse kohta, enne kui need kinnistuvad praktikateks (boyd & Crawford, 2012). Andmestumise ideoloogia näitab, kui laialt on levinud uskumus, et objektiivselt kvantifitseerida ja jälgida saab igat tüüpi inimkäitumist ning sotsiaalsust kasutades *online* meedia tehnoloogiaid (Van Dijck, 2014). Tihti eeldatakse andmete ja inimeste vahel iseenesestmõistetavat suhet (Van Dijck, 2014). Boyd ja Crawford (2012) on arvamusel, et, suurandmete puhul on väited objektiivsusele ja täpsusele eksitavad. Suurandmed pakuvad humanitaarteadustele uusi võimalusi tõsta oma staatust kvantitatiivses teaduses ja

objektiivsetes uurimismeetodites, muutes sotsiaalseid ruume kvantifitseeritavaks. Tegelikkuses on aga suurandmetega töötamine endiselt subjektiivne. Suurandmete läbi kvantifitseerimine ei pruugi tingimata objektiivsele tõele lähemal olla, eriti sotsiaalmeedia sõnumeid uurides (boyd & Crawford, 2012). Couldry (2020) on arutlenud hetkel sotsiaalse kriitika vähesuse üle, mida hõlmab endas andmestumine oma võimalike tagajärgedega sotsiaalse teadmise ja sotsiaalse maailma kohta. Autori (Couldry, 2020) arvates kipub kriitika liiga palju tuginema pigem loodusteaduslikele teooriatele ehk ANT/STS teooriatele (*Actor Network Theory and Science and Technology Studies*). Tema hinnangul aga ANT/STS teooriad ei selgita piisavalt, kuidas mõjutab andmestumine sotsiaalset korda. Kriitilise sotsiaalteaduse väljakutse tänapäeval on mõista andmestumise protsesside vormi ja dünaamikat ning nende tagajärgi laiemale sotsiaalsele korrale, mis iseloomustavad kaasaegset sotsiaalset maailma USA-st Hiinani (Couldry, 2020). Kokkuvõtteks saab öelda, et andmestumise tulemusena on kogutud ja pidevalt kogunemas suures mahus tehisintellekti arendamiseks vajalikke andmeid ning selline andmestumine toetab ka suuremat tehisintellekti kasutusele võttu.

1.2 Tehisintellekti olemus ja kasutusalad

Eesti tehisintellekti strateegias (2020) on tõdetud, et tehisintellekti mõiste defineerimine on keeruline. Andreas Kaplan ja Michael Haenlein (2019) on määratlenud tehisintellekti, kui süsteemi võimet väliseid andmeid korrektselt tõlgendada, nendest andmetest õppida ja kasutada konkreetsete ülesannete ning eesmärkide saavutamiseks. Tehisintellekt annab masinatele võime näha, kuulda, maitsta, nuusutada, puudutada, rääkida, kõndida, lennata ja õppida (Marr, 2019). Andreas Kaplan ja Michael Haenlein'i järgi on tehisintellekti sisendiks näiteks suurandmetena kogutud andmed, mille läbi üritab tehisintellekt tuvastada masinõppe meetodil reegleid ja mustreid. Masinõpe on tehisintellekti oluline osa, kuid on laiem kui tehisintellekt, kuna see hõlmab ka süsteemi võimet andmeid tajuda (nt keele või pildi tuvastamine) või erinevaid objektide juhtida, teisaldada ja manipuleerida, olgu selleks robot või mõni muu seade (Kaplan & Haenlein, 2019). Siiski võib kirjanduses masinõpe tähendada ka tehisintellekti. Cheng et al. (2021), järgi on tehisintellektil olnud ja jääb ka edaspidi keskne roll lugematutes elamise ja vabaduse aspektides. Nimelt kutsub tehisintellekt esile transformatsiooni, mis ei piirdu ainult tehniliste valdkondadega. Tehisintellekt on oma olemuselt tõeliselt sotsiotehniline nähtus, mis mõjutab tervishoidu, haridust, kaubandust, majandust, õigust, rääkimata igapäevaelust (Cheng et al., 2021).

Andreas Kaplan ja Michael Haenlein (2019) järgi saab tehisintellekti defineerida läbi evolutsioonietappide (**kitsas**, **lai/üldine** ja **super tehisintellekt**). Tänapäeval on kasutuses **kitsad tehisintellektid** (*artificial narrow intelligence*) ehk rakendused, mis kasutavad tehisintellekti ainult konkreetsete ülesande jaoks. Tulevikus võime näha tehisintellekti teist põlvkonda, mida nimetatakse **laiaks/üldiseks tehisintellektiks** (*artificial general intelligence*), mis suudab probleeme iseseisvalt põhjendada, kavandada ja lahendada selliste ülesannete jaoks, milleks neid kunagi isegi ette nähtud polnud. Kolmanda põlvkonna tehisintellekt on aga kauges tulevikus, mida kutsutakse **super tehisintellektiks** (*artificial super intelligence*), mis tähendab tõeliselt eneseteadlikuid süsteeme (Kaplan & Haenlein, 2019). Ray Kurzwell (2005) kirjeldab samuti **kitsast** ja **laia/üldist** tehisintellekti. **Kitsas tehisintellekt** tähendab süsteemide loomist, mis suudavad täita konkreetseid funktsioone, mis varem nõudsid inimintellekti rakendamist. Kitsas tehisintellekt ei ole aga veel saavutanud midagi, mis sarnaneks **tugeva tehisintellektiga**, mis oleks võrdväärne inimese intelligentsusega. Tugeva tehisintellekti revolutsioon hõlmaks inimese aju pöördprojekteerimist, mis tähendab inimese intelligentsuse mõistmist ja seejärel saadud teadmiste ühendamist üha võimsamate arvutitega (Kurzweil, 2005).

Lisaks läbi evolutsioonietappide on Andreas Kaplan ja Michael Haenlein (2019) tehisintellekti kirjeldanud ka läbi erinevat tüüpi tehisintellekti süsteemide ning on klassifitseerinud kolm peamist tehisintellekti gruppi: **analüütiline** (*Analytical AI*), **inimesest inspireeritud** (*Human-Inspired AI*), **humaniseeritud** (*Humanized A*I). **Analüütilise** tehisintellekti süsteemid loovad kognitiivse kujutise maailmast (Kaplan & Haenlein, 2019). Varasema kogemuse õppimise pealt suudavad nad omakorda luua informatsiooni tuleviku otsuste tarbeks. **Inimesest inspireeritud** tehisintellekti süsteem sisaldab elemente kognitiivsusest ning samuti emotsionaalsest intelligentsusest (Kaplan & Haenlein, 2019). Sellised süsteemid oskavad lisaks tunnetuslikele elementidele mõista ka inimeste emotsioone ja võtta neid arvesse otsuste tegemisel. **Humaniseeritud** tehisintellektil leidub aga omadusi igat tüüpi pädevustest (tunnetuslikest, emotsionaalsest ja sotsiaalsest intelligentsist). Selliseid süsteeme pole veel saadaval, mis oleksid eneseteadlikud ning omakorda ka teadlikud oma suhtlusest teistega (Kaplan & Haenlein, 2019). Kevin Warwick (2011) on jaganud tehisintellekti kaheks: **klassikaline** ja **moderne tehisintellekt**. **Klassikaline tehisintellekt** oma olemuselt üritab väljapoolt lähenedes kopeerida inimaju ja inimese intelligentsust (Warwick, 2011). Siiski ei ole klassikaline lähenemine tehisintellekti arendamisel parim lahendus ümbritseva situatsiooni tundma õppimisel ning varasemalt õpitud kogemuste võrdlemisel. **Moderne tehisintellekt** ei kopeeri üks-ühele inimese intelligentsust vaid läheneb intelligentsusele fundamentaalselt (Warwick, 2011), seeläbi on võimalus tunnetada ümbritsevat

maailma ning uutes situatsioonides varasemale kogemusele tuginedes teha parimaid otsuseid. Siiani on toimunud kindlasti edasiminek süsteemide arendamisel, mis suudavad tajuda ning järgi teha inimestele omaseid tegevusi. Sellise tehisintellekti loomine jääb aga tulevikku, mis suudab läbinisti fundamentaalselt tunnetada maailma (Kaplan & Haenlein, 2019). Warwick (2011) kirjutab, et suures mahus uuringuid on tehtud selleks, et tehisintellekt suudaks probleeme lahendada, planeerida ja otsuseid vastu võtta, kuid siiski pole keskendutud nii palju ümbritseva maailma sensoorse tunnetamise arendamisesse (Warwick, 2011). Tehisintellekti tüüp, mida kasutatakse praegu, on enamasti kõige rohkem oma olemuselt Kaplain ja Haenlein järgi (2019) analüütiline, näidetena saab tuua süsteeme, mida kasutatakse pettuste tuvastamiseks finantsteenustes, pildituvastuses või isejuhtivates autodes.

Haenlein'i ja Kaplan'i (2019) järgi arendatakse ja kasutatakse tehisintellekti süsteeme laias laastus teadust edendavates asutustes nagu ülikoolides, era- ja avalikus sektoris. Just ülikoolides on toimunud kõige olulisemad edusammud tehisintellekti süsteemide arendamises. Mõiste tehisintellekt (*artificial intelligence*) loodi Darthmouth'i kolledžis 1956. aastal töötoas, mille korraldas arvutiteadlane John McCarthy. Peale selle, et ülikoolis saavad alguse mitmed ideed tehisintellektide arendamiseks, on näiteks analüütilised tehisintellekti rakendused juba hakanud muutma ka õppejõudude rolli (Kaplan & Haenlein, 2019). Georgia Tech kasutab õpilaste küsimustele vastamiseks tehisintellektipõhist virtuaalset õpetajaassistenti. Süsteem on niivõrd tähelepanuväärne, et paljud õpilased saavad tehisintellektist teada alles pärast seda, kui seda neile öeldakse. Vaadates erasektorit on tehisintellekt seda mõjutanud nii, et saab rääkida tervete tööstuste ümberkujundamisest, seda eriti teenindussfääris (Kaplan & Haenlein, 2019). Analüütilist tehisintellekti kasutatakse personalijuhtimises, näiteks elulookirjelduste klassifitseerimisel sobivuse osas. Finantsteenuste sektoris on toimunud finantstehnoloogia tõus, mille kaudu toimub vara juhtimine ja finantstehingute analüüs masinõppemudeli abil. Inimesest inspireeritud tehisintellekt võimaldab Walmarti-sugustel ettevõtetel tuvastada õnnetud ja pettunud kliendid kassajärjekorras rakendades inimestele näotuvastustehnikaid ning võimaldades seeläbi olukorra muutmist näiteks uute kassade avamise teel (Kaplan & Haenlein, 2019). Ülikoolides ja erasektoris kasutatakse tehisintellekti selleks, et muuta erinevate ülesannete tegemist efektiivsemaks ning sarnaselt toimitakse ka avalikus sektoris. Näiteks Ameerika Ühendriikides Jacksonville'i linnas kasutatakse analüütilist tehisintellekti, mis muudab tänavavalgustite heledust vastavalt kogutud andmetele inimeste liikumise kohta. Ameerika Ühendriikide armee kasutab inimesest inspireeritud tehisintellekti värbamiseks tajudes tulevaste sõdurite emotsioone läbi veebikaamera. Tehisintellekti kasutatakse värbamisprotsessis küsimustele vastamiseks, kvalifikatsiooni üle

vaatamiseks ning edasiseks kandidaatide määramiseks tegelikele värbajatele (Kaplan & Haenlein, 2019). Kokkuvõtteks saab öelda, et olenemata sellest, et täielikult ümbritsevat maailma tunnetav tehisintellekt on veel arendusjärgus, siis on tehisintellektidel märkimisväärne kasutusala nii teaduslikes asutustes, era- ning avalikus sektoris.

1.3 Läbipaistvus andmestunud ühiskonnas

Andmestumisega kaasnevad erinevad probleemid. Mitmete autorite arvates on andmestumisel ohud privaatsusele, esineb algoritmilist diskrimineerimist, eetilisi probleeme ja läbipaistvuse puudumist (Redden, 2018; Van Dijck, 2014; boyd & Crawford, 2012). Valitsustel on näiteks kodanike kohta ulatuslikud andmekogumid, mis on sageli ajaloolised, personaalsed ja pidevalt uuenevad, muutes need andmeanalüütika seisukohast väärtuslikuks ning privaatsuse, turvalisuse ja õiguste seisukohast väga riskantseks (Redden, 2018). Strauß (2015) järgi soodustab andmestumine üksikisiku andmete levikut ja jälgimist ning tema arvates muutuvad tänu andmestumisele tõenäoliselt pakilisemaks muu hulgas järgmised küsimused: kuidas suurandmeid piisava anonüümsusega kasutada, kuidas toimida automatiseeritud otsustega ning kuidas parandada läbipaistvust ja vastutust. Andmestumisega kaasnevad privaatsuse, turvalisuse ja autonoomia tõhusa kaitsmise kontseptsioonide jaoks täiesti uued väljakutsed (Strauß, 2015). Redden (2018) on toonud andmestumisega kaasnevate probleemide lahenduseks läbipaistvuse suurendamise, kuid see peab olema kõrvutatud vastutusele võtmisega.

Larsson (2017) on välja toonud ühe andmestumise tagajärjena ka informatsiooni asümmeetria inimeste ja andmeid koguvate süsteemide ja ettevõtete vahel. Asümmeetriat saaks vähendada, kui informeerida ettevõtete ja teenuste tarbijaid rohkem teemal, kuidas nende andmeid kogutakse, analüüsitakse ning kellega täiendavalt jagatakse (Larsson, 2017). Üheks viisiks oleks suurendada andmetoimingute läbipaistvust; teine oleks õiguslike ja struktuuriliste kaitsemeetmete ümberkujundamine, et paremini kaitsta nõrgemaid osapooli (Larsson, 2017). Seega toob andmestumine kaasa vajaduse uute regulatsioonide järgi (Mejias & Couldry, 2019). Meijas ja Couldry (2019) sõnul puudub hetkel mõistmine, kellele andmed kuuluvad ning millised õigused erinevatel osapooltel on. Ühelt poolt peavad regulatsioonid arvestama andmeid tootvat seadet omava isiku huve. Teiselt poolt peavad regulatsioonid arvestama selle infrastruktuuri omanike huve, mille kaudu andmed liiguvad ja kogutakse (sotsiaalse kvantifitseerimise sektor). Küll aga on Euroopa Liidu reguleerivad asutused hakanud teatud regulatsioonide kaudu nagu isikuandmete

kaitse üldmäärus (lühend eesti keeles IKÜM, inglise keeles GDPR)[1] sellesse suhtesse sekkuma, et kaitsta inimese minimaalseid õigusi (Mejias & Couldry, 2019). Kokkuvõtteks saab öelda, et andmestumise toob endaga kaasa erinevaid probleeme, mille lahendusena nähakse vastavate regulatsioonide loomist. Andmestumisega kaasnevatest probleemidest ilmnes selge vajadus läbipaistvuse ja vastutuse tagamise järele.

1.4 Läbipaistvus tehisintellekti kontekstis

Erinevad osapooled mõistavad tehisintellekti kontekstis läbipaistvust tihti erinevalt. Arendaja jaoks on läbipaistvuse eesmärk mõista algoritmi toimimist ja saada aimu, miks ta nii töötab (Cheng et al., 2021). Selle jaoks, kes omab ja ka avalikustab algoritmi, on läbipaistvuse eesmärk muuta süsteem tarbijatele turvaliseks ja mugavaks kasutamiseks. Kasutaja jaoks tähendab läbipaistvus aru saada, mida tehisintellekti süsteem teeb ja miks (Cheng et al., 2021). Läbipaistvusest räägitakse tihti tehisintellekti süsteemide juures vastutusele võtmise kontekstis (Cheng et al., 2021; Pepito et al., 2019; Larsson & Heintz, 2020; Spielkamp & Loi, ia). Pepito et al., (2019) järgi seisab lähitulevikus ees raske probleem vastutussüsteemi ülesehitamises, mis soodustab innovatsiooni ja on kasulik, samal ajal pakkudes piisavat ja õiglast hüvitist tehisintellektist kahju saanud inimestele. Vastutusele võetud inimesed peaksid olema võimelised mõistma tehisintellektiga kaasnevate riskide ulatust ning vastutust ja neil peaksid olema erinevaid viise riskide maandamiseks (Pepito et al., 2019). Spielkamp & Loi (ia) uurisid eetilisi suuniseid (N=16) Algortihm Watch'i ülemaailmsete tehisintellekti juhendite hulgast. Vastutusele võtmist kirjeldatakse üldiselt kui soovitavat eesmärki või nõuet, kuid suunistes jääb arusaamatuks mis see on ja mida see täpsemalt kaasa toob (Spielkamp & Loi, ia). Autorid (Spielkamp & Loi, ia) aga näevad läbipaistvusel tähtsat rolli vastutusele võtmise saavutamisel, eristades seal juures otsest ja kaudset avalikku vastutust. Otsest avalikku vastutust on võimalik saavutada läbi avaliku läbipaistvuse, mis tähendab, et tehisintellekti süsteemi toimimine on arusaadav ja läbipaistev laiemale avalikkusele (Spielkamp & Loi, ia). Kaudset avalikku vastutust on võimalik saavutada läbi läbipaistvuse tagamise audiitoritele, kes annavad omakorda avaliku hinnangu süsteemi läbipaistvusele (Spielkamp & Loi, ia).

[1]GDPR- *General Data Protection Regulation 2016/679*

Cheng et al., (2021) on kirjeldanud sotsiaalselt vastutustundliku tehisintellekti kontseptsiooni, mis viitab inimese väärtuspõhisele protsessile, mis koosneb põhimõtetest nagu õiglus, läbipaistvus, vastutus, usaldusväärsus ja ohutus, privaatsus ja turvalisus ning kaasatus. Nende põhimõtete saavutamise vahendiks on sotsiaalselt vastutustundliku tehisintellekti algoritmide disainimine. Eesmärgiks aga jagatud väärtuste loomine ning sotsiaalsetele ootustele vastamine, milleks on tehisintellekti intelligentsuse ning sellest saadava kasu suurendamine (Cheng et al., 2021). Tehisintellekti läbipaistvusel on väga oluline roll üldises püüdluses arendada usaldusväärsemat tehisintellekti (Larsson & Heintz, 2020). Läbipaistvus, õiglus ja ohutus on aluseks usalduse tekkimisele tehisintellekti suhtes (Cheng et al., 2021). Tehisintellekti puhul on läbipaistvus ülemaailmselt üheks eetiliseks printsiibiks (Jobin et al., 2019). Jobin et al., (2019) on läbi viinud uurimuse tehisintellekti eetilisi suuniseid hõlmavate dokumentide pealt (N=84) eesmärgiga kaardistada olemasolevate tehisintellekti puudutavate eetiliste põhimõtete ja juhiste globaalset maastikku. Läbipaistvus on uuringu tulemustele tuginedes praeguses juhistes kõige levinum põhimõte, mida kajastatakse 73-s allikas (Jobin et al., 2019). Jobin et al., (2019) temaatiline analüüs paljastab aga olulisi erinevusi läbipaistvuse tõlgenduse, põhjenduse, rakendusvaldkonna ja saavutamisviisi osas. Uuringust selgub, et viited läbipaistvusele hõlmavad tihti ka jõupingutusi selgituse, tõlgendatavuse või muude suhtlemis- ja avalikustamistoimingute suurendamiseks (Jobin et al., 2019). Agudo ja Matute (2021) on rõhutanud selliste algatuste tähtsust nagu Euroopa Komisjoni usaldusväärse tehisintellekti eetikajuhised, mille eesmärk on suurendada tehisintellekti usaldusväärsust. Siiski usuvad autorid (Agudo & Matute, 2021), et usaldusväärne tehisintellekt on ainult osa lahendusest, sest suurenenud usaldus tehisintellekti vastu võib suurendada ka potentsiaalseid ohte, kuna pole veel piisavalt uuringuid, et mõista, kuidas algoritmide veenmine mõjutab inimeste otsuseid. Läbipaistvus on oluline, kuna inimesed võivad usaldada algoritmi pimesi ning nad ei tohiks seda teha (Agudo & Matute, 2021). Agudo ja Matute (2021) viisid läbi uuringu, kus soovisid aru saada, kas soovitusalgoritmid suudavad inimesi veenda, keda hääletada või millist partnerit valida kohtingule minemiseks. Viidi läbi neli katset, kus inimestele öeldi, et algoritm hindab nende isiksust ning soovitab neile vastavalt sellele sobiva kandidaadi ja partneri. Tegelikkuses seda algoritm ei teinud ning sobiva pildi valisid välja uuringu läbiviijad. Leiti, et algoritmi läbi veenmine oli võimalik ja erinevad veenmisstiilid (nt selgesõnaline, varjatud) olid tõhusamad sõltuvalt inimese otsuse kontekstist (nt poliitiline ja kohting). Uuringuga tõdeti, et inimesed on nõus algoritmide ettepanekuid oma otsuse tegemisel arvestama, teadmata, kuidas algoritm toimib (Agudo & Matute, 2021).

Ananny & Crawford (2018) on välja toonud täieliku läbipaistvuse saavutamisel esinevad piirid, millest üheks on asjaolu, et läbipaistvusel on tehnilised piirangud. Süsteemi vastutusele võtmiseks võib osutuda vajalikuks juurdepääs süsteemi koodile, kuid lihtsalt koodi nägemine on ebapiisav, kuna süsteemi üles ehitajad ei suuda tihti ka ise keeruka süsteemi toimimist selgitada (Ananny & Crawford, 2018). Haenlein & Kaplan (2019) on arvamusel, et kuigi tehisintellekt on oma olemuselt objektiivne ja ilma eelarvamusteta, siis ei tähenda see, et tehisintellektil põhinevaid süsteemid ei saaks olla kallutatud. Hoopiski iga kallutatus, mis on olemas tehisintellekti arendusperioodil algandmetes, on hilisemalt olemas ka tehisintellektis endas (Haenlein & Kaplan, 2019). Kemper & Kolkman (2019) järgi ei tähenda süsteemi sisse nägemine ilmtingimata selle käitumise või päritolu mõistmist. Algoritmide läbipaistvuse saab saavutada ainult huvitatud kriitilise huvigrupi kaudu ja isegi siis on läbipaistvuse saavutamisel selgelt piirid (Kemper & Kolkman, 2019). Nimelt oma olemuselt seotakse algoritmid koodi sees tavaliselt koos sadade teiste algoritmidega, et luua algoritmilised süsteemid. Kriitilisi huvigruppe huvitavad peamiselt nende algoritmiliste süsteemide töö, mitte konkreetsed algoritmid (paljud neist on üsna healoomulised ja protseduurilised). Nende ökoloogiate lahti haakimine aga osutub nende topoloogilise keerukuse tõttu sageli võimatuks (Kemper & Kolkman, 2019). Autorite (Kemper & Kolkman, 2019) arvates, ei saa algoritmiliste mudelite tagajärgi ja eeliseid välja tuua mudeli taga olevast koodist, isegi kui see on muudetud täiesti läbipaistvaks. Läbipaistvus eeldab kavandatud koodide ja juhiste terviklikku mudelit, kuid algoritmiline olemus raskendab sellist visiooni. Isegi kui algoritmiline mudel muudetakse täiesti läbipaistvaks, ei saa järeldada kõiki selle võimalikke mõjusid ja toimimisi. Algoritmi keerukus oma autonoomsete võimalustega toovad tingimata kaasa selle, et osa selle potentsiaalsusest jääbki suletuks (Kemper & Kolkman, 2019). Ka Larsson ja Heinz (2020) on arvamusel, et läbipaistvuse tõlgendamine algoritmide juures on raske, kuna andmete ja algoritmide suhe, mis on masinõppeprotsessi esilekerkivatest omadustest, on koodi ülevaatamisel tõenäoliselt tuvastamatu. Kaasaaegne masinõpe, millel tehisintellekti arendamine suuresti baseerub, on oma olemuselt nö must kast (Haenlein & Kaplan, 2019). Kuigi selliste süsteemide väljundi kvaliteedi (nt õigesti tuvastatud piltide osakaalu) hindamine on lihtne, jääb selleks kasutatav protsess suures osas läbipaistmatuks. Selline läbipaistmatus võib olla tahtlik (nt kui ettevõte soovib säilitada algoritmi saladust), samuti võib see tingitud olla tehnilisest kirjaoskamatusest või on seotud rakenduse ulatusega (nt juhul, kui tegemist on paljude süsteemi arendamises osalenud programmeerijatega). Ananny & Crawford (2018) on väitnud, et platvormide ja andmesüsteemide konkreetses kontekstis, algoritmilise süsteemi ühe osa - näiteks algoritmi või isegi selle aluseks olevate andmete - nähtavaks tegemine ei ole sama, mis süsteemi vastutusele võtmine. Läbipaistvus üksi ei saa luua vastutusvõimelisi süsteeme (Ananny &

Crawford, 2018). Kokkuvõtteks võib öelda, et läbipaistvust peetakse küll tehisintellekti kontekstis oluliseks, aga läbipaistvuse saavutamisel on mitmed piirangud.

1.4.1 Läbipaistvuse suurendamine tehisintellekti kontekstis

Larsson (2017) on seisukohal, et reguleerivad asutused peavad välja töötama kriitilisema perspektiivi ja parema arusaama sellest, kuidas juhtida andmepõhiseid ja algoritmiga juhitavaid protsesse ning andmete analüüse. Asutused peaksid parandama oma teadvustamise võimet ära tunda tarbijaid, kes vajavad kaitset ning püüdma pürgida läbipaistvuse poole ning mõista, kuidas uus tehnoloogia töötab ja milliseid eeskirju peaks tagama, et edasised arengud toimuksid kasutajate huvides (Larsson, 2017). Kemper & Kolkman (2019) järgi on akadeemiline kogukond hakanud välja töötama suuniseid, mis hõlmavad algoritmiliste mudelite vastutustundlikku juhtimist, analüüsi ning nende kasutamist. Autorid (Kemper ja Kolkman, 2019) leiavad, et läbipaistvus on nende juhiste võtmetegur, kuid siiski on läbipaistvusel piirid. Rohkem on vaja praktikas kasutatavate algoritmiliste mudelite empiirilisi uuringuid; eelkõige tuleb hinnata tingimusi, milles läbipaistvusmeetmed tegelikult positiivset mõju avaldavad, soodustades sealjuures produktiivseid suhteid kasutajaskonnaga, tunnistades samas ka sellise suhte vajalikke piire (Kemper & Kolkman, 2019). Lähitulevikus on tehisintellektisüsteemid üha enam osa meie igapäevasest elust ning see tekitab küsimuse, kas tehisintellekti kontekstis on vaja reguleerimist ja kui on, siis millisel kujul (Haenlein & Kaplan, 2019). Haenlein & Kaplan (2019) on välja toonud kolm perspektiivi, kuidas vaadelda tehisintellekti reguleerimist. Üheks neist on **mikrotasandi vaatenurk**, kus tehisintellekti enda reguleerimise asemel on ilmselt parim viis välja töötada üldtunnustatud nõuded tehisintellekti algoritmide arendamise ja testimise kohta, vältimaks, et masinõppe andmestike kallutatus, mis on olemas tehisintellekti arendusperioodil algandmetes, on hilisemalt tehisintellektis endas. Teiseks Haenlein & Kaplan'i (2019) poolt välja pakutud regulatsioonidele lähenemise vaatenurgaks on **mesotasandi perspektiiv** austusest tööhõive vastu. Näiteks võidakse ettevõtetelt nõuda teatud protsendi automaatika abil kokku hoitud raha kulutamist töötajate koolitamiseks uutele töökohtadele, mida ei saa automatiseerida. Viimaseks autorite Haenlein & Kaplan (2019) poolt välja pakutud regulatsioonidele lähenemise vaatenurgaks on **makrotasandi perspektiiv**, mille läbi töötada välja ülemaailmsed regulatsioonid. Lisaks regulatsioonidele on oluline kasutajate teadlikkuse tõstmine. Cory Robinson (2020) on välja toonud asjaolu, et riigid peavad prioriteediks seadma oma kodanike koolitamise tehnoloogiate, sealhulgas tehisintellekti osas. Kui kodanikud ei mõista süsteemide toimimist, ei saa nad nende

süsteemide toimimiseks vajalike isikuandmete kasutamisega nõuetekohaselt nõustuda (Robinson, 2020).

2. UURIMISMETOODIKA

Järgnevalt toon välja magistritöös kasutatud uurimismeetodid ning valimi moodustamise põhimõtted. Käesoleva peatüki lõpetan uurija reflektsiooniga.

2.1 Uurimismeetodid

Oma magistritöös kasutasin kvalitatiivseid andmete kogumise ja analüüsi meetodeid. Andmete kogumise ja analüüsi meetodina viisin läbi magistritöös kvalitatiivse kontentanalüüsi, mille eesmärgiks oli mõista, kuidas oli tehisintellekti läbipaistvus sõnastatud Põhja- ja Baltimaade riiklikes strateegiates. Teisena viisin läbi andmete kogumiseks poolstruktureeritud süvaintervjuud eesmärgiga uurida Eesti avaliku sektori infotehnoloogia vallas töötavate ekspertide arusaamu tehisintellektist ja selle läbipaistvusest. Intervjuude transkriptsioonide analüüsimiseks kasutasin kvalitatiivset sisuanalüüsi.

2.1.1 Kontentanalüüs

Kvalitatiivse andmekogumis- ja andmeanalüüsimeetodina viisin magistritöös läbi dokumendianalüüsi kasutades kontentanalüüsi meetodit. Bowen (2009) on kirjutanud, et dokumendianalüüs on süstemaatiline protseduur nii trükitud kui ka elektroonilise materjali läbi vaatamiseks või hindamiseks. Dokumendianalüüs hõlmab endas esmast üle vaatamist, põhjalikku lugemist ja seejärel tõlgendamist (Bowen, 2009). Lagerspetz'i (2017) järgi ei ole tekstid ja dokumendid ainult uurimisallikad, vaid need on ka ise ühiskondliku elu tähtis osa. Dokumendid ja tekstid ei peegelda ainult tegelikkust, vaid ka loovad ja muudavad seda, näiteks kehtestades norme või arusaamu faktidest, andes või kinnitades nähtuste nimetusi ja definitsioone (Lagerspetz, 2017). Leian, et ka tehisintellekti riiklikutes strateegiates sõnastatud läbipaistvus aitas aru saada, kuidas ühiskonnas antud nähtust mõistetakse.

Dokumendianalüüsiks valisin käesolevas magistritöös kontentanalüüsi. Lagerspetz'i (2017) järgi peetakse kontentanalüüsi all silmas teksti omaduste mõõtmist arvuliselt ning analüüsi üks peamisi eeliseid on süstemaatilisus. Veronika Kalmuse (2015) järgi kujutab standardiseeritud

kontentanalüüs endast teatavat silda kvantitatiivsete ja kvalitatiivsete meetodite vahel. Kontentanalüüsi lähtekohaks on arusaam, et mõne teema, mõtte, sõna või mõiste esinemissagedus peegeldab selle tähtsust analüüsitavas materjalis (Lagerspetz, 2017). Minu magistritöös võimaldaski kontentanalüüsi meetod läbipaistvusega seotud teema või mõtte esinemissagedust tehisintellekti riiklikes strateegiates loendada. Lagerspetz'i (2017) järgi saab esinemissagedust võrrelda materjali eri osades, ajavahemikel või seostes, kus sõnad, teemad või muu selline esinevad (Lagerspetz, 2017). Kontentanalüüsis püütakse selgitada materjalis ilmnevaid seisukohti ja nende sagedust ning leida ja seletada erinevusi materjali eri osade vahel (Lagerspetz, 2017). See oli oluline ka minu magistritöö kontekstis, kus läbipaistvusega seotud teema või mõtte esinemissageduse loendamisel võimaldas kontentanalüüs mul neid analüüsitud strateegiate vahel võrrelda. Seda peab ka Veronika Kalmus (2015) üheks kontentanalüüsi eeliseks, kuna kontentanalüüsi tulemusel määratakse huvipakkuvate teksti omaduste absoluutne ja suhteline esinemissagedus, siis võimaldab meetod erinevaid tekstikogumeid täpsetel alustel võrrelda. Kontentanalüüsi üheks abivahendiks on kodeerimisjuhis, kuhu on märgitud muutujad ja nende väärtused (Lagerspetz, 2017). Lagerspetz'i (2017) järgi on aga kontentanalüüsi puhul tavaline, et pärast esialgset materjaliga tutvumist kujuneb täpsem arusaam sellest, milliseid muutujad ja muutujate väärtusi on otstarbekas kasutada. See oli oluline ka minu magistritöös, kuna võimaldas kodeerimisjuhist pärast materjaliga tutvumist täiendada.

Veronika Kalmus (2015) on aga standardiseeritud kontentanalüüsi kriitikana välja toonud, et tekstianalüüsi standardiseerituse ja usaldusväärsuse kasvades kahaneb analüüsi sisukus ja sügavus. Standardiseeritud ja seetõttu kodeerija vaatevinklist jäikade analüüsikategooriate ja skaalade kasutamine võib viia uuritavate tekstiliste nähtuste lihtsustamise ja fragmenteerimiseni ning algupärastest tekstidest kaugenemiseni (Kalmus, 2015). Lagerspetz (2017) toob välja, et kontentanalüüs puudutab peamiselt kommunikatsiooni nähtavaid, kergesti leitavaid omadusi ning latentsete omaduste kodeerimisel võib aga analüüsi usaldatavus jääda madalaks. Arvesse võttes kriitikat leian, et kontentanalüüs oli minu magistritöö jaoks siiski sobilik meetod, kuna selle süstemaatilisuse läbi avanes võimalus läbipaistvusega seotud esinevaid teemasid tehisintellekti strateegiates loendada ning analüüsitud strateegiate vahel võrrelda.

2.1.2 Poolstruktureeritud intervjuu

Teiseks andmekogumise meetodiks kasutasin poolstruktureeritud intervjuud ning kogutud materjali analüüsisin kvalitatiivse sisuanalüüsi läbi. Lagerspetz'i (2017) järgi mõeldakse

poolstruktureeritu all seda, et küsimused esitatakse küll kindlas järjekorras, kuid neile vastatakse vabalt. Termin süvaintervjuu viitab taotlusele jõuda vastaja mõtete ja tunnete autentsele, sügavamale tasandile. Valisin andmete kogumiseks poolstruktureeritud intervjuu, kuna see võimaldab uurida vaateid ja arvamusi ning lubab intervjueeritavatel oma vastuseid laiendada (Gray, 2004). Antud asjaolu on vajalik, kui uurimuse eesmärk on uurida vastajate subjektiivseid arvmusi erinevatele mõistetele ja sündmustele. Selline lähenemine võib lubada ka intervjuu suunamise uutele radadele, mis küll algselt ei olnud intervjuu osad, kuid aitavad saavutada uurimistöö eesmärke (Gray, 2004). See oli oluline ka minu töö kontekstis, sest võimaldas uurimisküsimustest lähtuvad suunata intervjuud vajaliku tähenduste, kogemuste ja arvamuste kätte saamiseks. Gray (2004) järgi on intervjuu läbi viimisel aga oht, et intervjueerija on kas teadlikult või alateadlikult kallutatud ning suunab intervjueeritavat teemale kindlas suunas vastama. Selleks, et kallutatust vältida peab kõik intervjuud läbi viima standardiseeritud kujul, lähtuma intervjuu kavast ning intervjuu läbi viimisel taotlema neutraalset rolli (Gray, 2004). Intervjuu kavast lähtumine muidugi ei tähenda seda, et kõik intervjuud peaksid olema identsed ning vahepeal peab intervjueerija kalduma kavast kõrvale, et pakkuda juhendamist ja selgitamist (Gray, 2004). Ka minu uuringus tuli ette selgitada rohkemalt mõnda intervjuu kavas olevat küsimust, et intervjueeritaval oleks minu küsitu paremini mõistetav. Leian, et poolstruktureeritud intervjuu meetod pakkus mulle piisavalt vahendeid, et ette valmistada intervjuukava, milles intervjuudel lähtuda. Teiselt poolt aga oli mul piisavalt palju vabadust, et suunata intervjueeritavaid lähtuvalt uurimisprobleemist vastama, mõistmaks nende kogemusi, tähendusi ning arusaamu.

Andmeanalüüsimeetodiks kasutasin intervjuude transkriptsioonide analüüsimiseks kvalitatiivset sisuanalüüsi, millele on omased kindlad lähtepunktid. Kindlatest reeglitest saab välja tuua kaks lähenemist – induktiivne kategooriate arendamine ning deduktiivne kategooriate moodustamine (Mayring, 2000). Lagerspetz (2017) nimetab neid vastavalt teooriapõhiseks analüüsiks ning empiiriapõhiseks teoorialoomeks. Käesoleva magistritöö tarbeks kasutasin induktiivset ehk empiiriapõhist teoorialoomet. Induktiivne lähenemine nimelt lubab selekteerida analüüsiks vajalikku materjali (Mayring, 2000). Induktiivse meetodi korral ei ole analüüsi lähtekohaks juba eelnevalt paika pandud teooria, vaid uurimismaterjal, mida tuleb süstematiseerida (Lagerspetz, 2017). See oli oluline ka minu magistritöö kontekstis, kuna sisuanalüüsi läbi viimiseks olid mul olemas intervjuude transkriptsioonid, mida analüüsima hakata. Lagerspetz'i (2017) järgi seostatakse induktiivse meetodi korral hiljem tulemusi mõne taustteooriaga. Süstematiseerimise tarbeks luuakse kategooriad, mida tekstist otsitakse. Analüüsi kulg jaguneb induktiivses analüüsis kodeerimise, temaatilise rühmitamise ja kokkuvõtmise vahel. Kodeerimisel leitakse materjalist

olulised ütlused, need jagatakse temaatiliselt suurematesse gruppidesse ning lõpuks püütakse loodud struktuur teoreetilisel tasandil kokku võtta (Lagerspetz, 2017).

Lagerspetzi (2017) järgi võimaldab kvalitatiivne sisuanalüüs analüüsitavat teksti käsitleda tervikuna, mitte mingite osade nagu näiteks sõnade ja väidete summat. Teksti sõnum ei seisne esitatud väidetes vaid teemade ja väidete vahel peituvates hierarhiates, loogilistes seostes ja mustrites. Kvalitatiivset meetodit kasutades saab keskenduda teksti latentsele sisule. Lisaks võivad tulemused erineda sellest, mida analüüsi alustades eeldati ning niisugune avatus on kvalitatiivsete meetodite kasutamise põhjuseid (Lagerspetz, 2017). Kalmuse (2015) järgi ei ole erinevalt standardiseeritud kontentanalüüsist kvalitatiivse sisuanalüüsi eesmärgiks uuritavat teksti analüüsiühikute kaupa kodeerida ega koodide esinemissagedust määrata. Leian, et kvalitatiivne sisuanalüüs oli sobilik, kuna ekspertidega läbi viidud intervjuude transkriptsioonide analüüsimisel oli oluline keskenduda arusaamade mõistmisele tehisintellektist ja selle läbipaistvusest, mitte polnud oluline esinevaid teemasid ja mõtteid loendada. Ka Kalmuse (2015) järgi on kvalitatiivne sisuanalüüs tundlik ja täpne. Tähelepanu on võimalik pöörata ka harva esinevatele või unikaalsetele nähtustele tekstis. Analüüs on tihedalt tekstipõhine, selle käigus ei taandata tekstide sisurikkust ega nüansse numbrilistele koodidele, ei lihtsustata ega moonutata uuritavat nähtust ega liiguta sellest liiga kaugele. Küll aga on Kalmus (2015) välja toodud kriitikat, et kvalitatiivne sisuanalüüs analüüs ei võimalda erinevaid tekste täpsetel alustel võrrelda. Lisaks on toodud kriitikana välja seda, et kvalitatiivne sisuanalüüs loob uurijale võimaluse valikulise tõendusmaterjali kogumiseks, mis toimub sageli mitteteadlikult (Kalmus, 2015). Sellegi poolest leian, et just kvalitatiivse sisuanalüüsi avatus on see, mis minu magistritöö läbi viimiseks oli poolstruktureeritud intervjuude analüüsimeetodi valikul määravaks.

2.2 Valim

Kontentanalüüsis kuulusid valimisse Eesti, Soome, Rootsi, Norra, Leedu, Läti ja Islandi riiklikke tehisintellekti strateegia aruandeid. Lõplikusse valimisse kuulusid aga Eesti, Soome, Rootsi, Norra ja Leedu tehisintellekti strateegia aruandeid (vt Tabel 1). Analüüsist on välja jäetud Island, kuna magistritöö kirjutamise ajaks ei olnud nende tehisintellekti aruanne valmis. Analüüsi ei olnud võimalik kaasata ka Lätit. Euroopa Komisjoni veebilehel „AI Watch“ (AI Watch veebilehekülg, ia), mis tutvustab iga Euroopa Liidu riigi tehisintellektialaseid dokumente, on küll viidatud, et Lätil valmis tehisintellekti raport 2020. aastal, kuid siiski ei leidnud nende aruannet *Googl*e otsingu

kaudu. Dokumendi saamiseks saatsin ka Läti valitsuse kodulehel märgitud üldmeilile kirja, mis suunati edasi, aga millele kahjuks vastust ikkagi ei saanud.

Tabel 1. Valimisse kuulunud tehisintellekti strateegiad.

Jrk nr	**Riik**	**Tehisintellekti strateegia**
1.	Eesti	Eesti tehisintellekti kasutuselevõtu eksperdirühma aruanne (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020)
2.	Soome	Finland's Age of Artificial Intelligence (Ministry of Economic Affairs and Employment, 2017)
3.	Rootsi	National Approach to Artificial Intelligence (Ministry of Enterprise and Innovation, 2018)
4.	Norra	National Strategy for Artificial Intelligence (Norwegian Ministry & of Local Government and Modernisation, 2020)
5.	Leedu	Lithuanian Artificial Intelligence Strategy (Ministry of the Economy and Innovation, 2019)

Poolstruktureeritud intervjuude läbi viimiseks intervjueerisin nelja Eesti avaliku sektori eksperti, kes on omavad kogemusi ja teadmisi tehisintellekti arendustest ja projektidest Eestis (vt Tabel 2). Intervjueeritavate täpsed ametipositsioonid ning avaliku sektori asutused, kus nad töötavad, on mulle teada. Siiski võttes arvesse, et tehisintellektide arenduste ja projektidega tegelevate inimeste ring Eestis on kitsas, siis olen magistritöös kirjeldanud nende üldist rolli, mitte nende täpset ametipositsiooni, et kaitsta intervjueeritavate anonüümsust. Intervjueeritavateni jõudsin läbi tuttavate, kes mulle vastavate inimeste kontaktid andsid. Samuti otsisin ka intervjueeritavaid läbi avalike sektori asutuste kodulehtede läbi nende märgitud kontaktandmete läbi ühendust võttes.

Tabel 2. Valimisse kuulunud intervjueeritavad.

Jrk nr	**Tähis**	**Seotus tehisintellektide arendustega**
1.	INT_1	Tarkvarainsener
2.	INT_2	Andmehalduse arhitekt
3.	INT_3	Andmehalduse ekspert
4.	INT_4	Tarkvara arhitekt

2.3 Andmeanalüüs

Andmete kogumiseks kontentanalüüsi jaoks otsisin riiklikud tehisintellekti aruanded Eesti, Soome, Rootsi, Norra, Leedu kohta. Eesti tehisintellekti eksperdirühma aruande sain veebilehelt, mis hõlmab krattide projekti (Krattide veebileht, i.a.). Soome ja Leedu riiklikud tehisintellekti strateegia aruanded leidsin Euroopa Komisjoni veebilehelt „AI Watch“ (AI Watch veebilehekülg, ia). Need aruanded, mis Euroopa Komisjoni veebilehel puudusid, leidsin kasutades *Google* otsingumootorit. Riiklike strateegiate tekstide analüüsis kasutasin kontentanalüüsi. Esmalt moodustasin esialgse koodijuhise programmis MS Word, mille alusel sai tekste lugema hakata. Pärast tekstide esmast läbi vaatamist täiendasin ning valmis lõplik kodeerimisjuhis (vt LISA 1), mille alusel hakkasin tekste põhjalikult lugema ning koodijuhise abil kodeerima programmis MS Excel, mis võimaldas hiljem ka erinevaid jooniseid teha.

Magistritöö tarbeks viisin andmete kogumiseks läbi 4 poolstruktureeritud intervjuud. Enne intervjuude läbi viimist panin paika poolstruktureeritud intervjuukava (vt LISA 2). Arvestades 2021. aasta kevade COVID-19 viiruse levikut, ei olnud võimalik intervjueeritavatega näost-näkku intervjuu tarbeks kohtuda, seega viisin intervjuud läbi Skype veebikeskkonna ning lindistasin. Enne intervjuude algust selgitasin intervjueeritavatele, et intervjuu lindistusi kasutan edasises andmeanalüüsis anonümiseeritud kujul. Intervjuude pikkuseks oli 40 – 70 minutit. Infotehnoloogia vallas töötavate ekspertide näol on tegemist tiheda töögraafikuga inimestega. Seetõttu jäi ühe intervjuu pikkuseks 40 minutit, kuna intervjueeritaval ei olnud võimalik leida rohkem aega. Siiski sai ka nende 40 minutiga intervjuu kavale tuginedes magistritöö jaoks põhilised olulised teemad kaetud. Intervjuust on kergem terviklikku ülevaadet saada siis, kui seda mitte ainult ei kuulata, vaid see pannakse paberile või tekstfaili kirja (Lagerspetz, 2017). Käesoleva magistritöö jaoks said intervjuud lindistatud ning edasise analüüsi tarbeks transkribeeritud sõna-sõnalt kasutades programmi MS Word. Tihti piisab lihtsamast transkriptsioonist, kus intervjueerija ja intervjueeritava kõne pannakse kirja sõna-sõnalt, aga ilma erimärkideta, tähistades üksnes pikemaid pause (Lagerspetz, 2017). Käesoleva magistritöö tarbeks oli eelkõige eesmärk intervjueeritavate arvamuste ja hoiakute kindlaks tegemine, seega piisas transkribeerimisel sõna-sõnalisest kirja panekust. Erimärgid nagu kõne rütm ja hääletooni muutus ei olnud käesolevat magistritöö teemat arvesse võttes olulised. Selleks, et säilitada intervjueeritavate anonüümsus, on transkriptsioonidest peidetud nende täpsed ametipositsioonid, ettevõtted ja/või projektid, millega nad on seotud, kuna Eesti avaliku sektori infotehnoloogia ekspertide kitsas ringkonnas võiks arvata isiku identiteedi tuginedes tema tööalastele tegemistele. Käesolevas magistritöös said intervjuude

transkriptsioonid analüüsitud kvaliteetse sisuanalüüsi läbi. Lugemisel said märgitud sarnased olulised ütlused ning väited programmis MS Word. Otsisin välja korduvad mõtted, mida ma hiljem kodeerisin teemade alla programmis MS Excel, seejärel kasutasin neid uurimuse analüüsis ning hilisemalt järelduste tegemisteks.

2.4 Uurija reflektsioon

Potentsiaalsetele ekspertidele kirjutades vastati mulle, et ollakse meelsasti nõus minu intervjuudes osalema, kuna teemal vestlemine tundus huvitav. Mõnevõrra raske oli aga kokku leppida sobivat aega. Tegemist on tiheda töögraafikuga ekspertidega, mille tõttu intervjuu jaoks leidus vaba aeg tihti nädalate pärast. Intervjuu ajal raskendas asjaolu, et pidin oma küsimused mahutama lühemasse ajaraami, kui olin planeerinud. Sellepärast oli ühe intervjuu pikkuseks 40 minutit. Intervjuu läbiviijana taotlesin neutraalset rolli. Ma ei suunanud vastajaid soovitud vastuste poole ning lasin neil luua minu välja toodud teemadele nende endi tähendused. Tunnetan, et suutsin säilitada intervjuude vältel neutraalset ja professionaalset rolli. Dokumendianalüüsis lähenesin dokumentide lugemisele ilma, et mul oleksid kindlad hoiakud uuritava teema suhtes. Teadvustasin endale, et lähtudes uurimishuvist, olin endale eelnevalt selgeks teinud erinevad läbipaistvust puudutavad kontseptsioonid. Tunnetan, et suutsin siiski dokumentide üle vaatamisel olla neutraalne ning seeläbi suutsin välja selekteerida selle, mida iga riik oma tehisintellekti strateegia aruandes läbipaistvuse kohta on kirjutanud.

3. TULEMUSED

Tulemuste esimese peatükina kajastan tulemusi tehisintellekti riiklike strateegiate kontentanalüüsist. Järgmisest peatükist leiab intervjuude tulemused Eesti avaliku sektori ekspertidega.

3.1 Läbipaistvus tehisintellekti strateegiates

Läbipaistvus on tehisintellekti strateegiates sõnastatud erinevalt, osa riike on läbipaistvuse sätestanud läbi strateegiates paika pandud printsiipide. Leedu tehisintellekti strateegias (2019) on nimelt sätestatud eetilised ja õiguslikud printsiibid ning nende rakendamise mehhanismid tehisintellekti kasutamiseks ja arendamiseks, puudutades sealjuures ka läbipaistvust. Leedu strateegias (2019) on ühe printsiibina välja toodud luua usaldust tehisintellekti üle valitsevate normide, eeskirjade ja seaduste vastu. Printsiibi rakendamise mehhanismina nähakse lisa jõupingutuste ja investeeringute tegemist eesmärgiga luua selgitatavust, läbipaistvust, usaldust, kontrollimist, valideerimist, rünnakute vastast turvalisust ning pikaajalist tehisintellekti ohutust ja väärtuste ühtlustamist. Teise printsiibina on Leedu tehisintellekti strateegias (2019) välja toodud rakenduste läbipaistvuse ja aususe põhimõtetest lähtumise julgustamine. Printsiibi rakendamise mehhanismidena nähakse uuringute tegemise toetamist, et minimiseerida tehisintellekti kallutatust. Mehhanismina nähakse ka antud strateegias riikliku interdistsiplinaarse tehisintellekti keskuse loomise hõlbustamist, et edendada tehisintellekti eetikaga seotud arutelusid. Välja on toodud ka asjaolu, et paljud algoritmid, sealhulgas need, mis põhinevad süvaõppel, on kasutajatele läbipaistmatud; seega peab Leedu looma kaitsemehhanismi, mille abil teadlased töötaksid välja süsteemid, mis oleksid läbipaistvad ja võimelised ka kasutajatele tulemuste põhjuseid selgitama (Ministry of the Economy and Innovation, 2019).

Soome tehisintellekti strateegias on sätestatud printsiibid võrreldes Leeduga pigem üldised. Nimelt on Soome sätestanud põhimõtted, mille läbi on võimalik luua „hea tehisintellekti ühiskond“ ning mille hulka kuulub läbipaistvus (Ministry of Economic Affairs and Employment, 2017). Põhimõtete all peetakse antud strateegias läbipaistvust, aruandekohustust ja ulatuslikku ühiskondlikku kasu. Soome tehisintellekti strateegias (2017) on tõdetud, et tuleb veel täpsustada,

mida need põhimõtted erinevate osalejate ja reguleerivate süsteemide seisukohast praktikas tähendavad.

/../*Mida täpsemalt tähendab hea tehisintellekti ühiskond? Selle üldpõhimõtetena peetakse läbipaistvust, aruandekohustust ja ulatuslikku ühiskondlikku kasu. Küll aga on veel täpsustamata, mida need põhimõtted praktikas erinevate osalejate ja reguleerivate süsteemide seisukohast tähendavad. Hea tehisintellekti ühiskonna määratluse kindlaksmääramiseks on töös vaja ettevõtete, valdkonna ekspertide, teadlaste, poliitiliste otsustajate ja kodanike panust. Seda tööd on juba Soomes alustatud ja see nõuab meie kõigi panust*/../Soome tehisintellekti strateegia „Finland's Age of Artificial Intelligence" (Ministry of Economic Affairs and Employment, 2017, eesti keelde tõlgitud minu poolt)

Osad riigid käsitlevad läbipaistvust läbi koostöögruppide. Läbipaistvus on Eesti ja Norra tehisintellekti strateegiates sätestatud erinevate koostöögruppide ülestes juhistes ja suunistes (vt Tabel 3). Tehisintellekti valdkonnas on tegusamad rahvusvahelised organisatsioonid Euroopa Liit, OECD ning Põhjala- Balti koostöögrupp.

Tabel 3. Läbipaistvus riiklikes strateegiates koostöögruppide üleselt.

	Eesti	Soome	Rootsi	Norra	Leedu
EL tasandil ekspertrühma suunistes	x			x	
OECD tasandil tehisintellekti soovitustes	x				
Põhjala- Balti koostöögrupi koostöösuunistes	x				

Nii Eesti kui Norra on tehisintellekti strateegias välja toonud **läbipaistvuse eetilise suunisena**, mida töötas välja Euroopa Komisjoni poolt kokku kutsutud 52- liikmelise sidusrühma esindajatest koosnev eksperdirühm, kes sai ülesandeks välja töötada tehisintellekti arendamise ja kasutamise eetikasuuniseid. Eetikasuuniste kohaselt peab krattide usaldusväärsuse saavutamiseks olema täidetud kolm tingimust: kratt peab vastama seadusele, olema kooskõlas eetika põhimõtetega ning olema töökindel. Nendele tingimustele vastamiseks on omakorda välja toodud seitse põhinõuet, mille hulgas on läbipaistvus. Eesti tehisintellekti strateegias on suunised mõeldud kõigile sidusrühmadele kasutamiseks, kes kratte arendavad, juurutavad või kasutavad. Suunised ei ole õiguslikult siduvad, kuid on hea tavana soovituslikud (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020). Norra tehisintellekti strateegias (2020) on eksperdirühma sätestatud läbipaistvuse põhimõttest kirjutatud detailsemalt ning ühtlasi on

kirjutatud, et valitsus võtab Norras tehisintellekti vastutustundliku arendamise ja kasutamise aluseks just need samad põhimõtted. Seega on Norra suhtunud eksperdirühma paika pandud põhimõtetesse tõsisemalt. Sellele viitab ka see, et Norra riiklik tehisintellekti aruanne (2020) on eksperdirühma läbipaistvuse eetilise põhimõtte seletanud lahti viidates, et see tähendab, et üksikisikutel või juriidilistel isikutel peab olema võimalus saada ülevaade sellest, kuidas neid mõjutav otsus tehti. Seega lähtudes Norra strateegiast (2020) saavutatakse läbipaistvus muu hulgas andmesubjekti teavitamisega andmete töötlemisest. Antud strateegias on läbipaistvus seotud ka sellega, et arvutisüsteemid ei pretendeeri inimeseks, nimelt on Norra strateegias kirjas, et inimesel peab olema õigus teada, kas nad suhtlevad tehisintellekti süsteemiga või mitte (Norwegian Ministry & of Local Government and Modernisation, 2020).

Eesti tehisintellekti strateegias (2020) on kirjutatud läbipaistvusest ka OECD tasandil tehisintellekti soovitustes. OECD tasandil on Eesti tehisintellekti eksperdigrupp ja liikmesriikide esindajad koostanud tehisintellekti soovitused, mis kinnitati ministrite kohtumisel mais 2019. aastal. Soovitused käsitlevad põhimõtteid usaldusväärse tehisintellekti vastutustundlikuks haldamiseks. Paika on pandud viis soovituslikku põhimõtet, mida ka Eesti tehisintellekti strateegias loetletakse. Üheks põhimõtteks on sealjuures läbipaistvus ja selgitatavus ehk krattidest arusaamiseks on vaja mõistlikku info jagamist, sealhulgas juhul, kui krattide tegevusel on teatud osapooltele olnud negatiivne mõju (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020). Eesti strateegias (2020) on märgitud, et põhimõtete rakendamist oodatakse kõigilt sidusrühmadelt, kuid need ei ole õiguslikult siduvad. Seega on antud OECD tasandil koostatud soovitused pigem soovituslikud.

Eesti tehisintellekti strateegias (2020) on kirjutatud ka läbipaistvuse mõiste käsitlemisest Põhjala-Balti tasandi suunistes. 2018. aastal kirjutasid Põhjala ja Balti digiteemade eest vastutavad ministrid alla tehisintellekti deklaratsioonile, millega kaardistati Põhjala ja Balti riikide vahelise koostöö teemad tehisintellekti alal. Muu hulgas rõhutati ühiseid sihte parema ligipääsu tagamiseks andmetele piiriüleselt, tehisintellekti rakendamisega seotud oskuste edendamiseks, üle reguleerimise vältimiseks ja eetilise ning läbipaistva tehisintellekti loomiseks (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020).

Võrreldes Eestiga ei ole teistes analüüsitud Põhja- ja Baltimaade strateegiates detailselt toodud välja erinevate koostöögruppide üleseid juhiseid ja suunised, mis hõlmaksid endas ka läbipaistvust. Küll aga on Rootsi strateegias (2018) kirjutatud, et paljud regulatiivsed raamistikud

ja suunised, millega Rootsi peab arvestama, pärinevad Euroopa Liidust. Antud strateegias on sätestatud, et kui Rootsi soovib saada kasu Euroopa Liidu tehtud algatustest, on oluline, et oleksid riigis valmis struktuurid ja kompetents. Seetõttu peab Rootsi osalema Euroopa tehisintellekti üle peetavates arutluses ja osalema ELi jõupingutustes, et saada kasu, mida tehisintellekti kasutamine võib tuua (Ministry of Enterprise and Innovation, 2018). Norra strateegias (2020) on toodud välja veel ka tehisintellekti koostöö Põhjamaades, mis viitab sellele, et tehisintellekti osas kattuvad huvid ja väärtused. Norra tehisintellekti strateegias (2020) on nimelt ka detailsemalt kirjutatud sellest, et Põhjamaad teevad juba ministrite nõukogu kaudu koostööd mitmetes tehisintellektiga seotud valdkondades ning üks neist valdkondadest puudutab ka andmeid. Moodustatud on töörühm, et teha kindlaks andmekogumid, mida saab Põhjamaade vahel vahetada ja luua Põhjamaade ettevõtetele - nii avalikele kui ka eraõiguslikele ettevõtetele - lisaväärtust, austades samas Põhjamaade eetilisi aspekte ning väärtusi (Norwegian Ministry & of Local Government and Modernisation, 2020). Leedu strateegias (2019) on välja toodud, et Euroopa Liidul on juba olemas mõni regulatsioon, mis kehtib tehisintellekti kohta, kuid Leedus see puudub. Täpsemalt Euroopas Liidus kehtivast regulatsioonist räägitud ei ole. Leedu strateegias on aga sätestatud, et Leedu peab ise välja töötama reeglid, standardid, suunised, normid ja eetika põhimõtted, et suunata tehisintellekti eetilist ja jätkusuutlikku arengut ning tehisintellekti kasutamist (Ministry of the Economy and Innovation, 2019). Leedu strateegias (2020) on juba ka pandud eetilised ja õiguslikud suunised, mis viitab, et Leedu on arvamusel, et tuleb rohkem riigi siseselt suunata tehisintellekti eetilist ja jätkusuutlikku arengut ning kasutamist, kui rakendada soovituslikke eetilisi suunised mõne koostöögrupi läbi.

Kokkuvõtvalt saab öelda, et Leedu tehisintellekti strateegias on paika pandud eetilised printsiibid, mille hulgas on ka kirjeldatud läbipaistvuse olulisust ning välja on toodud ka mehhanismid, kuidas läbipaistvust saavutada. Soome tehisintellekti strateegias küll väärtustatakse läbipaistvust printsiibina, mis on aluseks „hea tehisintellekti ühiskonnale“, kuid täpsemalt pole kirjutatud, milles läbipaistvus peaks seisnema või kuidas seda saavutada. Eestis kirjeldatakse läbipaistvust läbi erinevate koostöö organisatsioonide, siiski on läbipaistvus seal märgitud ühe soovitusliku eetilise suunisena ning detailsemalt mõistet lahti pole seletatud. Sarnaselt Eestile on Norra oma strateegias välja toonud Euroopa Komisjoni poolt kokku kutsutud eksperdirühma suunised, mille hulka kuulub läbipaistvus. Norra on seletanud sealset suunist detailsemalt ning märkinud, et valitsus võtab Norras läbipaistvuse põhimõtte tehisintellekti vastutustundliku arendamise ja kasutamise aluseks.

3.1.1 Tehisintellektiga kaasnevad riskid

Osad riikide strateegiad sätestasid läbipaistvust riskina, mis kaasneb tehisintellekti kasutuselevõtuga. Mitmete riikide strateegiad tõid välja teisigi tehisintellektiga kaasnevaid riske (vt Tabel 4).

Tabel 4. Strateegiates välja toodud tehisintellektiga kaasnevad riskid.

	Eesti	Soome	Rootsi	Norra	Leedu
Vastutuse puudumine				x	
Diskrimineerimine			x		
Eetilised riskid			x		
Kallutatud või manipuleeritud andmed			x	x	
Küberrünnakud			x		
Läbipaistvuse puudumine			x	x	
Ohud demokraatia toimimisele			x		
Rahaline kahju			x		
Tehisintellekti väär- või vaenulik kasutamine			x		
Töökohtade kaotus		x	x	x	x
Usalduse kaotus			x		

Enamikes strateegiates peale Eesti mainiti tehisintellekti kui ohtu **töökohtade kaotamisele**. Nimelt Soome, Rootsi, Norra ja Leedu on oma strateegiates välja toonud riski, et tehisintellekti kasutuselevõtt kaotab tulevikus töökohad, mida tehisintellekt suudab ise ära teha (Ministry of the Economy and Innovation, 2019; Ministry of Economic Affairs and Employment, 2017; Norwegian Ministry & of Local Government and Modernisation, 2020; Ministry of Enterprise and Innovation, 2018). Kõige rohkem on tehisintellektiga kaasnevaid riske välja toonud Rootsi. Nimelt on Rootsi tehisintellekti strateegias (2018) välja toodud, et tehisintellekti kasutusele võtmisega kaasnevad riskid nagu uut tüüpi intelligentsed **küberrünnakud**. Rootsi riiklik strateegia (2018) tõdeb, et tehisintellekti kasutuselevõtuga kaasnevad erinevad väljakutsed ning on oluline, et Rootsi saaks nendega hakkama. Rootsi strateegias on kirjutatud, et tehisintellekti kasutamisel võivad olla ettenägematud tagajärjed, mis võivad tekkida **kallutatud või manipuleeritud andmete**, **läbipaistvuse puudumise**, **väärkasutuse või vaenuliku kasutamise tagajärjel**. See võib põhjustada omakorda **diskrimineerimist**, **usalduse kaotust, rahalist kahju ja tagajärgi demokraatia toimimisele**. Nendel põhjustel on Rootsi jaoks oluline aktiivselt töötada teemadel

ja probleemidel, mida tehisintellekt endaga kaasa toob (Ministry of Enterprise and Innovation, 2018). Rootsi riikliku tehisintellekti strateegia (2018) järgi ei saa eetilisi, ohutuse ja turvalisuse kaalutlusi jätta tagantjärele lahendamiseks vaid need peavad olema varase projekteerimise etapi lahutamatu osa. See viitab sellele, et Rootsi strateegias pannakse rõhku tehisintellektiga kaasnevate riskide maandamiseks juba varajastes arenguetappides. Rootsi strateegias (2018) on märgitud, et tehisintellektiga seotud riskid pole mitte ainult tehnilised, vaid ka **eetilised**, eriti mis puudutavad tehisintellektil baseeruvate rakenduste arendamist ja kasutamist avalikus sektoris. Antud strateegias ollakse arvamusel, et tehisintellekti kasutamine nõuab moraalsete ja õiguslike probleemidega arvestamist ning esitab väljakutseid, mis on seotud õigusriigi menetluste ja otsuste automatiseerimisega. Väljakutsete näiteks on Rootsi strateegias (2018) toodud välja laialdaselt arutatud eetiline küsimus, et kuidas peaks autonoomne sõiduk mõtlema ja tegutsema, kui ta on hädaolukorras sunnitud valima kahe tulemuse vahel, mis mõlemad tähendavad, et inimene võib saada viga. Rootsi võib võtta tehisintellekti eetilise, ohutu, turvalise ja jätkusuutliku kasutamise juhtpositsiooni, töötades aktiivselt selle teemaga riiklikul tasandil ning propageerides seda rahvusvaheliselt (Ministry of Enterprise and Innovation, 2018). Rootsi tehisintellekti strateegia (2018) on sätestanud, et tehisintellekti algoritmide kasutamine peab olema läbipaistev ja arusaadav ning tehisintellekti väljatöötamisel ja kasutamisel tuleb juhinduda normidest ja eetika põhimõtetest vähendades nii ühiskonnale kui ka üksikisikutele tekkivaid riske. See pole ainult teadlaste ja inseneride küsimus, vaid puudutab ka kõiksugu ametkondi (Ministry of Enterprise and Innovation, 2018).

Riiklikes tehisintellekti strateegiates on ka Norra toonud välja tehisintellekti kasutuselevõtuga kaasnevaid riske. Norra tehisintellekti strateegia (2020) on arvamusel, et tehisintellekti kasutamine avalikus sektoris toob kaasa kasu, näiteks automatiseerimine võib edendada võrdset kohtlemist ja määruste järjepidevat rakendamist. Sellegipoolest on Norra strateegias tõdetud, et algoritmi hinnang olema vähemalt sama usaldusväärne, kui inimese kaalutlused, mida see asendab. Selle tagamiseks vajame läbipaistvust ja süsteeme, mille toimimist on võimalik selgitada (Norwegian Ministry & of Local Government and Modernisation, 2020). Sarnaselt Rootsile on Norra tehisintellekti strateegias (2020) tehisintellekti kasutuselevõtuga seonduvatest riskidest välja toodud **läbipaistvuse puudumist.** Täpsemalt on Norra strateegia (2020) toonud välja, et tehisintellekti väljakutse on läbipaistvuse puudumine mõnes süvaõppel põhinevates lahenduses, nimelt mõnda süvaõppe algoritmi saab võrrelda nn „musta kastiga“, kus inimesel pole juurdepääsu mudelile, mis seletaks, miks antud sisend annab kindla tulemuse. Siiski on Norra strateegias (2020) kirjutatud, et enamik tehisintellektil põhinevaid süsteeme ei ole siiski „mustad kastid“ ning

võimaldavad mõista ja dokumenteerida, kuidas otsuseid tehakse. Norra strateegia toob välja, et nendes kohtades, kus seletatavus on oluline, võib sobivam olla alternatiivne lähenemine süvaõppele, näiteks keskenduda „seletatava tehisintellekti" valdkonnale, mille eesmärk on muuta „musta kasti" algoritmid selgitatavaks. Norra tehisintellekti strateegias (2020) on viidatud sellele, et tehisintellekti puhul ei ole läbipaistvuse tagamine sama, mis koodi avaldamine algoritmi taga või juurdepääsu andmine täielikule andmekogumile, sest selline lähenemine võib rikkuda intellektuaalomandi õigusi ja andmekaitseseadusi. Selle asemel saab seletatav tehisintellekt analüüsida, millistel andmetel oli tulemuse jaoks oluline tähendus ja millist tähtsust omasid erinevad elemendid, ning seletada seeläbi tulemuse loogikat (Norwegian Ministry & of Local Government and Modernisation, 2020). Ühe võimalusena näeb Norra strateegia (2020) läbipaistvuse tagamiseks vastavate suuniste ja regulatsioonide loomist, nimelt riigihalduse süsteemides tuleb tehisintellekti kasutamisel kehtestada läbipaistvuse ja vastutuse nõuded. Norra strateegia (2020) kohaselt on sellised algatused juba seotud näiteks autonoomse transpordiga, mis viitab sellele, et Norra on juba enne strateegia avaldamist pööranud tähelepanu läbipaistvuse ja vastutuse nõuetele. Läbipaistvuse puudumist tehisintellekti juures nähakse Norra strateegia (2020) poolt ka väljakutseks tarbijatele. Ühelt poolt nähakse, et tehisintellekti kasutamine pakub tarbijatele palju eeliseid, näiteks üha suurema hulga uute igapäevaelu lihtsustavate teenuste väljatöötamist (Norwegian Ministry & of Local Government and Modernisation, 2020). Teiselt poolt nähakse Norra strateegias (2020) ka probleemkohti eraelu puutumatuse, läbipaistvuse ja tarbijaõiguste osas. Lisaks on antud strateegias toodud välja, et tarbijad on eriti haavatavad, kui tehisintellekti kasutatakse personaalsete teenuste ja sihipärase turunduse arendamiseks, mis põhineb tarbijate isikuandmete kogumisel ja töötlemisel. Norra strateegias mõistetakse ka probleemkohta eraelu puutumatuse, läbipaistvuse ja tarbijaõiguste osas laiemal tasandil kui riiklikul, nimelt on strateegias kirjutatud, et rahvusvaheline murekoht on see, et ettevõtted ei võta tarbijate privaatsust piisavalt tõsiselt (Norwegian Ministry & of Local Government and Modernisation, 2020).

Sarnaselt Rootsile on Norra tehisintellekti strateegias (2020) välja toodud ka ühe tehisintellekti kasutuselevõtu riskina **andmete manipuleerimist ja kallutatust**. Norra (2020) tehisintellekti strateegias on aga seletatud täpsemalt, kuidas võib andmete kallutatus mõjutada tehisintellekti. Antud strateegia järgi ilmneb kallutatus siis, kui andmekogumid sisaldavad teavet ainult osa asjakohaste lähteandmete kohta. Norra strateegias (2020) on kallutatuse ilmestamiseks välja toodud näide, et kui koerte kujutiste äratundmiseks mõeldud algoritmi koolitatakse ainult pallidega mängivate koerte piltide abil, võib algoritm koera pilti mitte ära tunda, kui pildil palli pole. Näitena

treeningandmete kallutatusest on Norra strateegias (2020) välja toodud ka, kui näotuvastuseks mõeldud algoritmi koolitatakse ühe etnilise rühma näopiltidele, siis teisi etnilisi rühmasid algoritm pildi pealt ei tuvasta. Strateegia kohaselt võib kallutatus tekkida muudel põhjustel, näiteks võib andmekogum sisaldada kallutatust, mis tuleneb inimeste väärarvamustest või ajaloolistest eelarvamustest lähteandmetes (näiteks mehi nähakse teatud tüüpi ametikohtadel töötavatena rohkem kui naisi või kui andmed sisaldavad rohkem naiste kujutisi köögivalamu juures kui meeste omi) (Norwegian Ministry & of Local Government and Modernisation, 2020). Norra strateegias välja toodud näited viitavad sellele, et treeningandmestiku kallutatust tehisintellekti juures teadvustatakse ning nähakse seda probleemkohana.

Samuti toob ka Leedu tehisintellekti strateegia (2019) välja tehisintellektiga kaasnevad riskid, kuid üldiselt lähtuvalt sätestatud õiguslike ja eetilise printsiipidega ning nende rakendusmehhanismidega. Mis viitab sellele, et Leedu pole küll konkreetseid riske oma strateegias välja toonud, kuid printsiibid on koostatud lähtuvalt tehisintellektiga kaasnevatest riskidest. Sellele viitab ka see, et Leedu strateegias on tõdetud, et tehisintellekt, mis suudab indiviididele ja ühiskonnale tohutut kasu tuua, tekitab ka teatud riske, mida tuleks korralikult kontrollida (Ministry of the Economy and Innovation, 2019). Seega on strateegia kohaselt vaja tehisintellekti inimkeskset lähenemist. Leedu strateegia on arvamusel, et usaldusväärne tehisintellekt on eetiline ja tehniliselt kindel, kuna isegi head kavatsused võivad tehnoloogilise meisterlikkuse puudumise tõttu põhjustada kahju. Küll aga tõdetakse, et arvestades, et tehisintellekti kasu kaalub tervikuna üles selle riskid, tuleks tagada kesktee, mis maksimeerib tehisintellekti eelised, minimeerides samal ajal selle riskid (Ministry of the Economy and Innovation, 2019).

Kokkuvõtvalt saab öelda, et Soome ja Leedu on konkreetse riskina tehisintellekti kasutuselevõtu juures välja toonud strateegiates töökohtade kaotust. Rootsi ja Norra strateegiates on lisaks töökohtade kaotusele pööratud rõhku muudele tehisintellekti kasutusele võtmisega seotud riskidele. Läbipaistvuse puudumist on sealjuures samuti nähtud tehisintellekti kasutusele võtmise riskina just Rootsi ja Norra strateegiates. Leedu on viidanud tehisintellektiga kaasnevate riskidele strateegias üldiselt ning sätestanud probleemkohtade vältimiseks printsiibid. Eesti tehisintellekti strateegias riskidest kirjutatud ei olnud.

3.2 Ekspertide arusaamad läbipaistvusest

Tulemuste teise peatükina esitlen läbi poolstruktureeritud intervjuude kogutud ja kvalitatiivse sisuanalüüsi läbi analüüsitud tulemusi Eesti avaliku sektori ekspertidega, kes omavad teadmisi ja kogemusi tehisintellekti vallas.

3.2.1 Arusaam tehisintellektist

Tehisintellekt tähendab enamus intervjueeritud ekspertide jaoks mudelit, mida on võimalik treenida. Osa eksperte kirjeldasid mudeleid masina nimega, millele mingisuguse sisendi alusel õpetatakse tegema tegevusi. Intervjueeritavaid tõid sisse ka paralleele inimese ja tehisintellekti vahel. Andmehalduse arhitekt tõi välja selle, et tehisintellekt on oma algsesse tähendusesse tagasi läinud, kus on mingisugune masin, mis käitub sarnaselt inimesele. Tarkvara arhitekt tõi välja, et tehisintellekt ei ole rohkemat, kui hästi õpetatud süsteem ning tõdes sealjuures, et tehisintellekti termin on üles kiidetud.

/../Kohati tundub tehisintellekt selline üle haibitud asi ja selles mõttes mulle natukene meeldib võib-olla see Eesti lähenemine nagu krati nime all, siis ta ei kõla suurelt ja vägevalt. Kõlab muidu täpselt niisugune müügijutt nagu Big Data /suurandmed/. Aga ta ei ole nagu mitte midagi rohkemat, kui lihtsalt hästi-hästi õpetatud süsteem. Mis oskab lihtsalt hästi nagu oma ressursse hästi ära kasutada, oma teadmisi ära kasutada, et haip on alati suurem nagu asi ise. Selle tõttu ei maksa vist tõesti karta, et mõni tolmuimeja üks kord su kodu üle võtab, et seda niipea ei juhtu/../ (INT_4)

Mõistmaks arusaamu tehisintellektist, uurisin intervjueeritavatelt, et mida on vaja tehisintellekti arendamiseks. Enamik intervjueeritavaid tõid välja asjaolu, et lisaks teadmistele ja oskustele on vaja aru saada, millist probleemi hakatakse lahendama. Kõik intervjueeritavad olid ühel arvamusel, et tehisintellekti arendamisel on vaja laiemaid matemaatilisi ja statistilisi teadmisi, mis viitab sellele, et ainult tarkvara alased teadmised jäävad tehisintellekti arendamise koha pealt väheseks. Toodi ka välja seda, et tarkvara arendajad tehisintellekti algoritme ise ei realiseeri.

/../Ma tarkvara arendajate pärast väga ei muretse selles mõttes, et ega nemad ka tehisintellekti algoritme üldiselt ise realiseeri, et kasutavad olemasolevaid teenuseid, et nende jaoks on see rohkem nagu must kast/../ (INT_2)

Enamus eksperte olid ühel arvamusel, et andmeteadlane mõjutab tehisintellekti läbi treeningandmestiku valiku, mis viitab sellele, et võib tekkida treeningandmestiku kallutatus. Sellele omakorda viitab ka tarkvara arhitekti välja toodud näide, kus tehisintellekt ei oska tunda eelmise sajandi pildilt mobiiltelefoni, kuna teda ei ole piisavalt andmetega treenitud.

/../*tehisintellekti on treenitud andes ette tuhandeid ja miljoneid fotosid, et vot õpi selgeks, selline asi on mobiiltelefon./../ Sellele samale tehisintellektile anti ette pilt aastast 1920, kus inimestel oli käes sigaretikarp, nüüdseks tilluke asi ja no otse loomulikult tuvastas ta pildi pealt mobiiltelefoni, sellepärast, et ta oli õppinud nagu sellise materjali pealt, et noh, et sellised asjad peavad olema mobiiltelefonid*/../ (INT_4)

Treeningandmestiku kallutatust nähti ka kõigi ekspertide seas tehisintellekti arendamise ja kasutuselevõtu juures probleemkohana. Treeningandmestiku kalde kohta toodi välja ühe eksperdi poolt lisaks seda, et tegemist on murekohaga, küll aga pigem on tegemist teisejärgulise probleemiga. Nimelt andmehalduse arhitekti poolt lisati ka seda, et kui teha masinaga masinõppe teel algoritme ja see töötab paremini kui inimene, siis kalle on võib-olla isegi teisejärguline. Mis viitab sellele, et kui tehisintellekt täidab oma eesmärki, siis ei pea ekspert kallet peamiseks probleemiks. Tarkvarainsener tõi tehisintellekti puhul probleemina välja selle, et mõned algoritmid ja mudeleid on raskesti tõlgendatavad. Ekspert teadvustab, et tehisintellekti puhul on oma olemuselt tegu keerukate ning raskesti tõlgendatavate süsteemidega. Tarkvarainsener lisas ka seda, et tehisintellekt keskendub tulemustele ning tihtipeale ei ole aru saada, kuidas ta selle tulemuseni jõuab. Mis viitab sellele, et tehisintellekti puhul ei ole alati läbipaistvust, et mõista kas treeningandmestik on korrektselt klassifitseeritud.

/../*Nojah, ma ütlen, et see on nagu, see täiesti oleneb algoritmist ja mis nurga alt seda lahendatud on, kindlasti on nagu juhte, kus nagu sul ongi raske aru saada nagu miks. Miks see selline on, aga noh, andmed on lihtsalt no, andmete põhjal on ta treenitud, sul on mingisugused kaalud tekkinud. Ja lihtsalt sul on nagu raske tõlgendada*/../ (INT_1)

Probleemkohtadena nähti ka tehisintellekti laiemat ühiskondlikku mõju nagu töökohtade kaotus. Andmehalduse ekspert nimelt tõi välja asjaolu, et süvaõppe algoritm GPT-3 on võimeline kirjutama ise koodi nii, et ühel hetkel võib-olla pole enam vaja koodi kirjutamise jaoks programmeerijat. Tarkvara arhitekt tõi välja probleemkoha, et masintöötluse läbi võib-olla

treeningandmestikust võimalik tuvastada inimese identiteet. See viitab, et nähakse probleemi tehisintellekti kasutuselevõtuga inimesele privaatsusele ja turvalisusele.

/../Võta nüüd see avaandmete filosoofia, et noh, ma võtan ühe avaandmete komplekti, kus ma ei suuda kedagi tuvastada ja teise avaandmete komplekti ja võib-olla võtta kolmanda ka. Ja kõik on nagu täpselt nii nagu peab, et kõik on täitsa normaalsed andmed ja inimestega kokku ei pane, aga kui ma panen kõik kolm andmekomplekti kokku, siis võib-olla üks inimene ei suuda nagu sellest midagi välja lugeda, aga kui väike masintöötlus sinna peale teha, siis selgub küll, et ahaa, et kuule, aga näed, et see mingi Juhan seal Kõrvekülast. Et sellised asjad tegelikult tekivad, kui lihtsalt saad õpetada masina tegema midagi, mida inimene ei saaks teha/../ (INT_4)

Üks ekspert tõi probleemkohtade juures välja huvitava asjaolu, et inimesed hakkavad tehisintellekti reguleerima ilma, et nad tegelikult teaksid, et mida ta võimaldab. Diskussioonides pannakse vähe rõhku sellele, et võrrelda inimese tekitatud vigu masina tehtud vigadega. Nimelt oli ekspert arvamusel, et inimene teeb samuti vigu ja diskussioonides peaks sellele rohkem rõhku panema ning masina vigadesse kergemini suhtuma.

/../Samal ajal väga vähe rõhku pandi selle peale, et võrrelda masina poolt välja pakutud ütleme, kas siis juhtimisotsuseid, muid otsuseid, võrreldes inimese poolt välja pakutuga. Ehk siis nagu diskussioon ei ole mitte sisul vaid on nagu vormil. Et kui inimene teeb vigasid, siis sellest palju ei räägita, kui masin ühe vea teinud kunagi, siis see on väga-väga ohtlik asi/../ (INT_2)

Küsisin arvamust Euroopa Komisjoni 2018. aastal kokku kutsutud eksperdirühma poolt kokku pandud tehisintellekti arendamise ja kasutamise eetikasuuniste kohta. Need põhimõtted on: inimese toimevõime ja järelevalve; tehniline töökindlus ja ohutus; privaatsus ja andmehaldus; läbipaistvus; mitmekesisus, mittediskrimineerimine ja õiglus; ühiskondlik ja keskkonnaalane heaolu; vastutuse võtmine. Eksperdid jäid suuniste osas eriarvamusele. Ühe eksperdi poolt toodi välja seda, et kõikides töölõikudes ei peaks neid põhimõtteid taga ajama, näiteks tõi välja ettevõtte müügi tiimi, kelle eesmärk on müüki kasvatada. Mis viitab sellele, et ekspert ei pea läbipaistvust igat tüüpi tehisintellektide puhul oluliseks olenevalt tehisintellekti kasutusalast. Sellele viitab ka see, et eksperdi arvates peaks olema süsteem läbipaistvam siis, kui hakatakse tegema otsuseid, mis mõjutavad 100 miljonit või 500 miljonit inimest.

/../Noh, ongi näiteks firma müügi tiim, nemad nüüd tahavad müüki kasvatada, on vaja sellist head segmenteerimist ja nende sihtgruppide seast, eks ju. Ja noh masin seal nii moodi midagi teeb ja müük tõuseb, super. See läbipaistvus ja muud sellised asjad, et nende jaoks ei ole üldsegi olulised/../ (INT_2)

Osa eksperte tõi eetikajuhiste kohta välja asjaolu, et samad põhimõtted peaksid kehtima ka inimestele. Andmehalduse arhitekt ei näe nimelt erinevust inimese ja masina vahel, kuna ka inimene teeb otsuseid, siis need samad põhimõtted peaksid olema olulised ka siis. Tarkvara arhitekt tõi samuti välja, et põhimõtted peaksid rakenduma absoluutselt igas eluvaldkonnas igas ametipostpositsioonis, mitte ainult tehisintellektidele. Lisati ka seda, et kohati on veider, kuidas inimene võiks tehisintellektile põhimõtteid ette kirjutada.

*/../Kohati mulle tundub, et kuidas see tehisintellekt nagu puutub nagu nendesse punktidesse ausalt öeldes/../muidugi on üldse väga veider, kuidas inimene võiks tehisintellektile mingit põhimõtteid ette kirjutada/../ (*INT_4).

Teised eksperdid aga pidasid põhimõtteid mõistlikuks, küll aga tarkvarainseneri arvates Euroopa Komisjoni eksperdirühma suuniste reaalne rakendamine ei ole tema arvamusel lihtne. Nii tarkvarainseneri kui ka andmehalduse eksperdi arvates aitaks põhimõtete täitmist hinnata sertifitseerimine, tuues paralleeli majandusaruannete auditeerimisega. Mis viitab sellele, et hetkel on Euroopa Komisjoni eksperdirühma poolt paika pandud põhimõtteid keeruline tehisintellektide puhul rakendada, kuna puuduvad vastavad kontrollmehhanismid, millega nende täitmist hinnata.

/../meil ei ole ka täna ju veel välja kujunenud sellist nii-öelda, nagu majandusaruandeid auditeeritakse, et keegi, kes siis nii-öelda enne, kui see label /tähis/ peale pannakse, sertifitseerib selle nii-öelda protsessi ja tõenäoliselt noh, tekivad seal mingeid astmed, viie palli skaalal, et jah, võime kasutada, aga nende agadega, et jah, siin on see excellent /korras/, siin on kõik korras/../ (INT_3)

Kokkuvõtteks saab öelda, et enamik eksperte mõistab tehisintellekti all mudelit, mida on võimalik treenida. Seega on intervjueeritud eksperdid ühel arusaamal tehisintellekti olemusest. Enamus avaliku sektori eksperte on ühel nõul, et andmeteadlane mõjutab tehisintellekti läbi treeningandmestiku valiku, millest võib tekkida treeningandmete kallutatus, mida omakorda nähti tehisintellektiga kaasneva probleemkohana. Lisaks toodi murekohtadena ka välja mudelite raskesti

tõlgendatavust ja laiemat ühiskondlikku mõju töökohtade kaotuse näol. Seega teadvustavad eksperdid tehisintellekti arendamise ja kasutuselevõtuga kaasnevaid probleeme. Toodi aga ka välja seda, et tehisintellekti probleemkohtadest rääkides tuleks arvesse võtta asjaolu, et ka inimesed teevad vigu. Euroopa Komisjoni eksperdirühma suuniste kohta jäid intervjueeritud eksperdid eriarvamusele. Osa eksperte ei näinud vajadust, miks peaks põhimõtteid eraldi välja töötama tehisintellektide jaoks, kuna nendest peaks lähtuma ka inimesed. Lisaks oldi arvamusel, et antud eetiliste suuniste järgmine ei ole alati vajalik lähtuvalt tehisintellekti eesmärgist. Näiteks pole põhimõtted vajalikud tehisintellektide juures, mis on orienteeritud ettevõtte müügile. Teised eksperdid pidasid suuniseid mõistlikeks, kuid toodi välja keerukust nende rakendamisel. Siiski pakuti lahendusi näiteks sertifitseerimise ja auditeerimise näol.

3.2.2 Arusaam tehisintellekti läbipaistvusest

Läbipaistvust tõlgendati enamike ekspertide poolt üldiselt, kui võimet aru saada, kuidas tehisintellekt töötab. Osad eksperdid tõid ka välja, et kohati on see tehisintellekti puhul raske. Ühe eksperti poolt toodi välja läbipaistvuse puhul mõistmist, kuidas tehisintellekt lõpptulemini jõuab.

/../mitte seda, et siin protsessi algas ja vot nüüd on siin lõpptulem, vaid seda tahabki näha, et mis seal vahepeal ka nii-öelda tegelikult juhtus/../(INT_4).

Üks ekspert tõi välja, et termin läbipaistev pole kõige parem seletada tehisintellekti läbipaistvust. Nimelt kirjeldab tehisintellekti läbipaistvust rohkem aru saamine, kuidas mudel treenitud on ja mis on selle eesmärgid.

*/../läbipaistvus minu jaoks on see, et mitte öelda paistab läbi selles mõttes, et see on, see on minu meelest vildakas, vaid mul on võimaluse korral võimalus aru saada ehk jah, arusaamine nii-öelda, kust ta tuleb ja kuidas ta on loodud ja mis on tema eesmärgid, et see noh, ütleme, et noh, statistilises keeles lihtsalt metoodika oma parameetritega /../(*INT_3).

Andmehalduse arhitekt tõi välja, et lineaarsete mudelite puhul on võimalik mõista tehisintellekti toimimist, aga tunnuste kasvades on seda pigem raske teha, eriti lõppkasutaja vaatevinklist. Mis viitab sellele, et praegu on meil kasutuses mittelineaarsed tehisintellektid, mis on läbipaistvuse koha pealt raskemini tõlgendatavad.

/../Üldiselt mudelid on siis selliselt ehitatud, et sul on seal mingid konkreetsed tunnused, millel omad kaalud on välja toodud ja läbipaistvus siis sellest ütleme noh... Lineaarsete mudelite puhul see tähendabki seda, et sa tead siis nende võrrandi elementide kaalusid. Et tead, mis see matemaatiline mudel seal taga on. Noh, kui sul tunnuste arv kasvab näiteks, et sul ei ole neid enam sadasid näiteks, vaid on sajad tuhanded või on näiteks isegi miljonid, siis taolisel juhul inimestele.... Noh, sa pead kuidagi tulemust ära abstraheerima, et läbipaistvus tähendab siis lihtsalt võib-olla sellist nagu selgitust, et kuidas treenitud mudel üldjuhul töötab näiteks./../ (INT_2)

Eksperdid jäid eriarvamusele selles osas, kas läbipaistvus on tehisintellekti juures alati oluline. Tarkvara arhitekt tõi välja selle, et läbipaistvus ei pea olema alati oluline lõppkasutajale ning antud arvamuse illustreerimiseks lihtsustas ta tehisintellekti joogiautomaadi näitele. Nimelt, kui inimene saab mündi masinasse panekul oma joogi kätte, siis ei huvita teda see, mis protsessid masinas käisid. See viitab sellele, et ekspert on arvamusel, et kui tehisintellekt täidab lõppkasutaja jaoks eesmärki, siis ei pea olema ta tema jaoks läbipaistev. Küll aga tõi ta näite, et ka joogiautomaadi puhul peab kellelegi olema vajalik teada, kuidas see töötab, näiteks sellele, kes peab seda parandama. Mis viitab omakorda sellele, et tehisintellekt peaks olema läbipaistev näiteks tehisintellekti arendajatele ning mitte nii väga lõpptarbijale.

/../Et see on nagu see läbipaistvus, noh, kas see peab olema läbipaistev on omaette küsimus minu meelest hoopis, et jah, kuigi see oli üks põhimõtetest/../see teema, et alati ju sind ka ei huvita, et kui sa, nagu ma tea, paned mündi masinasse, valid numbri ja saad oma joogi kuskilt automaadist, et noh, kas sind huvitas, mis sellest mündist seal vahepeal sai ja kuidas see mehhanism tegelikult tööle hakkas ja kuidas sa oma joogi kätte said. Sind huvitas see, et ah nii palju tuli raha panna ja nüüd sain sellise joogi. Aga noh samas kedagi huvitab, et kedagi, kes seda masinat parandama peab, peab väga hästi teadma, kuidas ta töötab/../ (INT_4)

Kui tarkvara arhitekt oli arvamusel, et läbipaistvus ei ole oluline alati lõppkasutaja vaatenurga alt, siis andmehalduse arhitekt tõi välja, et läbipaistvus ei pea olema alati oluline üldises plaanis, kui tegemist on näiteks ettevõtte müügile orienteeritud tehisintellektiga. Mis viitab sellele, et kui tehisintellekt üleüldiselt täidab oma eesmärki, siis ei pea ta läbipaistvust oluliseks.

/../Et algoritm ei pea olema läbipaistev, kui on selline nagu ütleme, väike seltskond, kellele ta teeb töö ära näiteks./../ (INT_2)

Osad eksperdid olid aga üldiselt arvamusel, et läbipaistvus on oluline, et oleks võimalik ära hoida võimalikke probleeme, mida tehisintellektiga võib kaasneda. Andmehalduse ekspert tõi näiteks välja seda, et läbipaistvuse tagamisega saab mõista treeningandmete kallutatust. Seega treeningandmete kallutatuse probleemi lahendamiseks näeb ekspert läbipaistvust.

/../sotsiaalmeedia valimi pealt tehtud algoritmid, mis mõjutavad nii-öelda kogu valimit, see ei ole õige/../et kui see on tehtud, siis see algoritm ei tohi läbi minna nii öelda. Või noh, kui me selle pilditööstusel noh, et meil on kõik HD pildid, aga me ei ole ühtegi nii öelda madalama sagedusega või noh, madalama resolutsiooniga pilti on ju, siis noh, me ei saa rakendada algoritmi, sest me teame, et osad pildid võivad olla väga mürased ja nii edasi./../ (INT_3)

Üks ekspert tõi välja asjaolu, et läbipaistvusele hakatakse mõtlema kogemusena, kui on vajadus otsida hilisemalt süsteemi toimimise vea põhjuseid. Eksperdi arvamus viitab sellele, et ühtlasi võibki tekkida vajadus läbipaistvuse järele siis, kui tehisintellekti süsteemiga tekivadki probleemid.

/../tõenäoliselt, kui sul esialgu seda kogemust ei ole/../siis sa teed täiesti läbipaistmatu tehisintellekti, kus annad sisendi ja siis saab mingi vastuse ja siis mõtled, et kuidas ta selle vastuseni jõudis. Ja, ja siis võib-olla leiad mõne koha, kus äkki see oli see koht, kus ta komistas ja tekitad sinna mingisuguse jälje, et aru saada, et aga mis ta sellest kohast arvas/../või ressursside olemasolu näiteks, kui sul on nagu kiire või sul on ressursse lihtsalt vähe/../panustanud nagunii palju mitte kvaliteeti, vaid kvantiteeti eks, nagu lõid selle tehisintellekti, mis teeb küll asja ära, aga kui peaks tekkima mingi jama, et seda jama lahendada või uurida noh, mitte isegi keeruline võid suht võimatu ka./../ (INT_4)

Läbipaistvuse suurendamise koha pealt arvas üks ekspert, et väga oluline on metoodika kirjeldamine. Oluline on sealjuures kirjeldada, mis andmeid on kasutatud, kuidas on valitud nii-öelda treeningnäitajad ja teised näitajad, millist meetrikat kasutades hinnatakse mudelit. Samuti on oluline, milliseid katseid on tehtud ning milliste erinevate variatsioonidega nad tehtud on.

/../Et see on ka selline, nagu ütleme see dokumentatsioon mudeli kohta või mudelite gruppide kohta, et see kindlasti aitaks seda läbipaistvust tõsta, aga ta on mõeldud väga väikesele sihtgrupile

pigem nendele, kes neid mudelid teevad või siis teevad rakendusi, kus neid mudeleid kasutatakse/../ (INT_2)

Tarkvarainsener tõi läbipaistvuse suurendamise koha pealt samuti välja olulise aspekti, et vajalik on teada, mis andmeid on kasutatud tehisintellekti arendamisel. Lisas veel, et tegelikult on läbipaistvus pigem sõltuv erinevatest metoodikatest, kuidas mudeleid seletada. Vaja oleks teha mingeid täiendavaid samme või meetodeid, et paremini seletada, et miks tehisintellekt nii käitub. Mis viitab, et läbipaistvuse saavutamiseks on vaja teha lisategevusi.

Uurisin ekspertidelt ka seda, kuidas mõjutavad arendajate arusaamad läbipaistvusest tehisintellekti. Üks ekspert oli arvamusel, et arendaja eetikat tema väga kõrgelt ei hindaks, pigem ta eeldab, et nad on kursis selle teemaga ja teavad sellest. Ekspert tõi välja asjaolu, klient on ikkagi see, kes peaks teadma, mida ta tahab tellida. Seega eksperdi arvates sõltub läbipaistvus ka tehisintellekti süsteemi tellijalt. Lisati ka seda, et kui klient on nõus tehisintellektiga, milles esineb treeningandmete kallutatust, siis arendajale ei tohiks see probleemiks olla.

/../kui klient tahab tellida mingisugust asja, kus on kalle sees ja ta nõus sellega siis selle nagu ütleme arendaja poole pealt ei tohiks ka probleemiks olla/../ (INT_2)

Teised eksperdid olid arvamusel, et läbipaistvus ikkagi oleneb arendajast ja seda erinevate aspektide läbi. Tarkvarainsener arvates sõltub läbipaistvus arendaja eesmärgist. Siiski tõi ta välja asjaolu, et see sõltub arendajast, kuidas ta süsteemis paikneb. Mis viitab sellele, et kui tegemist on arendajaga, kes töötab ettevõtte või mõne tehisintellekti projekti heaks, siis pigem tuleb tal järgida ülevalt-alla nõuded, kui läbipaistev peab arendatav süsteem olema.

/../oleneb jah arendajast, et kui, kui oluliseks tema peab seda läbipaistvust. Kas tema eesmärk on saada lihtsalt hea mudel või kui tema eesmärk on ka seda kuidagi ära seletada kellegile eksju et, et see oleneb jälle, et kuidas ta seal süsteemis, kuidas arendaja seal süsteemis paikneb/../ (INT_1)

Ühe eksperdi poolt toodi välja, et olulised on arendaja kogemused ja oskused selle läbi, kuivõrd on tekkinud probleeme, et oma arendusi muuta läbipaistvamaks. Eksperdi arvamus viitab sellele, et tehisintellekti süsteeme muudetakse läbipaistvamaks probleemide vältimiseks.

/../Pigem ongi täpselt seesama kokkupuutepunkt lihtsalt, et selles mõttes kogemus, et kas sul on nagu olnud vajadus selliseid asju /läbipaistvamaks/ teha või kas sul on selliste asjadega jama/../ (INT_4)

Kokkuvõtteks võib öelda, et eksperdid olid üksmeelel üldises läbipaistvuse definitsioonis, et see on võime aru saada, kuidas tehisintellekt töötab. Osade ekspertide poolt toodi välja, et läbipaistvust tehisintellekti puhul on pigem raske defineerida ja ka saavutada. Läbipaistvuse olulisuse koha pealt jäid eksperdid erinevatele arvamustele tuues välja seda, et läbipaistvus ei pea olema alati oluline lõppkasutajale ning erinevalt tehisintellektist ei pea näiteks müügile orienteeritud tehisintellekt olema läbipaistev, kuna ta täidab oma eesmärki.

4. JÄRELDUSED JA DISKUSSIOON

Käesolevas peatükis on lähtuvalt uurimisküsimustest ning poolstruktureeritud intervjuude ja kontentanalüüsi tulemustest esitletud järeldused ning diskussioon. Uurimisküsimused olid minu magistritöös järgnevad:

- Millised on arusaamad tehisintellektist infotehnoloogia vallas töötavate ekspertide seas?
- Millised on arusaamad tehisintellekti läbipaistvusest infotehnoloogia vallas töötavate ekspertide seas?
- Kuidas on tehisintellekti läbipaistvus sõnastatud Põhja- ja Baltimaade riiklikes strateegiates?

Järgmisena toon oma magistritöös välja järeldused ja diskussiooni.

4.1 Järeldused ja arutelu

Arusaamad tehisintellektist ja tema toimimisest olid intervjueeritud avaliku sektori ekspertide seas sarnased. Sarnaselt käsitletud teooriale (Kaplan & Haenlein, 2019) toodi välja tehisintellekti olemust, kui miskit, mida treenitakse kindla probleemi lahendamiseks. Üks intervjueeritud ekspertidest lisas ka, et tehisintellekti termin on üles kiidetud nagu suurandmete puhul, mida on kirjeldanud ka käesoleva magistritöö teoorias (Strauß, 2015; Gandomi & Haider, 2015). Lisati ka, et tänasel päeval ei ole meil veel tehisintellekti süsteeme, mis meid üle võtaksid, mis viitab käsitletud teooriale, kus hetkel on kasutuses kitsad tehisintellekti süsteemid ning inimese intelligentsusega võrreldavad tehnoloogiad on veel kauges tulevikus (Kaplan & Haenlein, 2019; Kurzweil, 2005).

Uurides arusaamu tehisintellektist toodi intervjuudes välja ka tehisintellekti probleemkohti. Tõdeti ühe murekohana treeningandmestike kallutatust sarnaselt Haenlein & Kaplani (2019) teooriale, kus iga kallutatus, mis on olemas tehisintellekti arendusperioodil algandmetes, on hilisemalt olemas ka tehisintellektis endas (Haenlein & Kaplan, 2019). Treeningandmete kallutatuse

ilmestamiseks toodi näide tehisintellektist, mis ei oska mobiiltelefoni varasema sajandi piltidelt ära tunda, kuna sellele ei ole varasemaid pilte õpetatud. Analüüsist selgus ka see, et enamus eksperte olid ühel arvamusel, et andmeteadlane mõjutab tehisintellekti läbi treeningandmestiku valiku. Seega treeningandmestiku kallutatus võib tekkida lähtuvalt andmeteadlase tõekspidamistest, mis võib saada tulevase tehisintellekti osaks ning millele viitavad ka magistritöös käsitletud autorid (Ananny, 2016; O'Neil, 2016; Eubanks, 2018; Noble, 2018).

Tulemuste analüüsist selgus, et kui tehisintellekt täidab oma eesmärki, siis ei pea intervjueeritud ekspert treeningandmestiku kallutatust peamiseks probleemiks ning tuleks mõelda hoopis sellele, kas inimese kalle on parem kui masina kalle. Saab järeldada, et eksperdid mõistavad, et treeningandmestike valikust sõltub tehisintellekti enda kallutatus, siiski ei näe kõik seda ilmtingimata peamise probleemina. Peab aga tõdema, et inimese kallet ei saa võrrelda tehisintellekti kaldega. Inimese puhul saab enamasti kindlaks teha tema tegevusi ja ütlusi analüüsides, millest on viga tekkinud. Küll aga on seda raske teha läbipaistmatu süsteemi puhul. Nagu on kirjeldatud ka teoorias, siis tehisintellekti puhul on oma olemuselt tegemist süsteemidega, mille toimimist on raske seletada (Kemper & Kolkman, 2019; Larsson & Heintz, 2020). Tulemuste analüüsist selgus, et osad eksperdid mõistsid seda ning tõid ka probleemkohana välja asjaolu, et tehisintellekti puhul on raskusi mudeli tõlgendamisel.

Tehisintellekti puhul toodi probleemkohtadena intervjueeritud ekspertide poolt välja ka töökohtade kaotust. Strateegiate analüüsi tulemusena selgus, et riiklikud tehisintellekti strateegiad näevad samuti tehisintellektide kasutuselevõtuga erinevaid probleeme. Töökohtade kaotamist nägid tehisintellekti puhul probleemidena enamus analüüsitud strateegiad nagu Soome, Rootsi, Norra ja Leedu. Rohkem keskenduvad probleemidele Rootsi ja Norra, kes on välja toonud sarnaselt teooriale näiteks probleeme treeningandmete kallutatusest (Haenlein & Kaplan, 2019). Üheks Rootsi ja Norra toodud tehisintellekti probleemiks on ka läbipaistvuse puudumine. Eesti kontekstis ei ole tehisintellekti strateegias välja toodud konkreetseid tehisintellektiga seonduvaid probleeme, millest saab järeldada, et enamus avaliku sektori eksperdid mõtestavad erinevaid probleemkohti, kuid Eesti strateegias nendele tähelepanu ei pöörata. Siit tekib küsimus, et kas Eesti on oma strateegias keskendunud liialt tehisintellektist saadavale majanduslikule ja ühiskondlikule kasule ning tähelepanuta jätnud võimalikud probleemid, mida on toonud välja teised tehisintellekti aruanded. Asjaolu, et Eesti strateegias ei ole keskendutud tehisintellekti kasutuselevõtmisega kaasnevate probleemidele võib olla selles, et hetkel on kasutuses vaid kitsad tehisintellektid, mis ainult automatiseerivad mõne töölõigu inimese tööst ning oma toimimiseks

vajavad ka inimese assisteerimist (Kaplan & Haenlein, 2019; Kurzweil, 2005). Seega võib järeldada, et Eesti hetkel ei näe, et tehisintellektiga võib kaasneda probleeme, kuna need pole täielikult autonoomsed. Sellegi poolest kaasneb ka kitsaste tehisintellektidega riske, nagu treeningandmete kallutatus ning tekib küsimus, miks sellest lähtuvalt pole Eesti strateegias murekohti välja toodud.

Arusaamad läbipaistvusest tehisintellekti kontekstis üldise definitsioonina avaliku sektori ekspertide seas sarnanesid ning oma olemuselt tähendas neile läbipaistvus, kui võimet aru saada, kuidas tehisintellekt töötab. Sarnaselt teooriale (Kemper & Kolkman, 2019; Larsson & Heintz, 2020; Ananny & Crawford, 2018) eksperdid tõdesid, et läbipaistvust on raske tõlgendada. Intervjueeritud eksperdid oskasid läbipaistvust defineerida detailsemalt, näiteks statistilises keeles tähendab läbipaistvus tehisintellekti kontekstis metoodikat oma parameetritega. Oluline on siin kohal märkida, et ühe eksperdi arvates ei olegi läbipaistvus parim definitsioon tehisintellektile, kuna see tähendab läbi paistmist, tehisintellekti puhul tuleb hoopis aru saada, kuidas see toimib ja seevastu on arusaamine parem definitsioon tehisintellekti läbipaistvusele. Tehisintellektist aru saamine on kahtlemata olulisem kui läbi paistmine, seda kinnitab teooria, et läbipaistvus ei tähenda tehisintellekti kontekstis näiteks ainult koodi nägemist (Kemper & Kolkman, 2019; Ananny & Crawford, 2018) vaid hõlmab mõistmist, kuidas on saadud treeningandmestikud ning milliseid metoodikaid on kasutatud mudelite õpetamiseks. Saab järeldada, et eksperdid mõistavad raskusi läbipaistvuse käsitlemisel ja selle saavutamisel, kuna oma olemuselt on tehisintellekt keerukas süsteem.

Intervjuude tulemuste analüüsist selgus tõsiasi, et eksperdid on tehisintellekti läbipaistvuse olulisuse suhtes eriarvamusel. Intervjuudes toodi näiteks välja, et läbipaistvus ei pea olema alati oluline lõppkasutajale. Lähtudes magistritöös käsitletud teooriast peaks läbipaistvus kasutaja jaoks tähendama aru saamist, mida tehisintellekti süsteem teeb ja miks (Cheng et al., 2021). Läbipaistvuse läbi on võimalik ka tehisintellekti süsteem vastutusele võtta (Spielkamp & Loi, ia). Isegi, kui süsteem ei ole tehtud lõppkasutajale avalikult arusaadavaks, siis peaks olema tagatud vastav auditeerimissüsteem, mille tulemustega saab lõppkasutaja tutvuda (Spielkamp & Loi, ia). Siit tekib küsimus, et kuidas saab süsteemi vastutusele võtta, kui lõppkasutajale pole tehisintellekt läbipaistev? Intervjuudest selgus ka, et läbipaistvus ei ole oluline, kui tehisintellekt täidab oma eesmärgi, näiteks müügile orienteeritud ettevõttes. Olenemata sellest, kas tegemist on erasektori müügile orienteeritud või avaliku sektori tehisintellektiga peaks olema mõista, kuidas tehisintellekt toimib ning eesmärgini jõuab. Läbipaistvus on nimelt sätestatud Eesti tehisintellekti

arenduste juures ühe väärtusena. Eesti e-riigi (*E-estonia*) kontseptsioon on luua suurema läbipaistvuse, usalduse ja tõhususega ühiskond (e-Estonia veebilehekülg, ia). Strateegiate analüüsi tulemustest selgus ka, et Läbipaistvus on sätestatud ka Eesti tehisintellekti aruandes erinevate koostöögruppide üleselt printsiipidena juhiste ja suuniste läbi. Euroopa Komisjoni eksperdirühma suunised, millele viitab ka Eesti tehisintellekti aruanne, on kahtlemata olulised, millele viitab ka teooria (Agudo & Matute, 2021). Eetikajuhiste eesmärk on suurendada tehisintellekti usaldusväärsust. Siiski usuvad Agudo ja Matute (2021), et usaldusväärne tehisintellekt on ainult osa lahendusest, sest suurenenud usaldus tehisintellekti vastu võib suurendada ka potentsiaalseid ohte. Siit tekib küsimus, et kui läbipaistvus on läbivaks väärtuseks ja eetiliseks juhiseks tehisintellekti arenduste juures Eestis, siis miks ei pööra läbipaistvuse olulisusele kõik intervjueeritud Eesti avaliku sektori eksperdid piisavalt tähelepanu?

Läbipaistvust nähakse sarnaselt teooriale (Jobin et al., 2019) riiklikes Põhja- ja Baltimaade strateegiates eetilise printsiibina. Eesti tehisintellekti strateegiates ei ole rohkem läbipaistvuse kontseptsiooni selgitatud, kui läbi erinevate koostöögruppide üleselt paika pandud printsiipidena, millest saab järeldada, et Eesti peab küll läbipaistvust oluliseks, aga siit tekib omakorda küsimus, mida läbipaistvus Eesti tehisintellekti kontekstis tähendab. Intervjuude tulemuste analüüsist selgus, et Eesti avaliku sektori eksperdid olid erinevatel arusaamadel Euroopa Komisjoni eksperdirühma eetilistest juhistest. Ühelt poolt peeti suuniseid mõistlikeks, aga teiselt poolt toodi välja seda, et kõikides tehisintellekti töölõikudes ei peaks neid põhimõtteid taga ajama. Lisati ka seda, et põhimõtted peaksid rakenduma absoluutselt igas eluvaldkonnas igas ametipostpositsioonis, mitte ainult tehisintellektidele. Ühelt poolt mõistetakse eetilisi printsiipe väärtustena, mis kehtivad ka teistes elu valdkondades, aga teiselt poolt ei nähta nende vajalikkust piisavalt, et igas töölõigus neid rakendada. Siit tekib küsimus, kas Eesti strateegias jääb koostöögruppide üleselt sätestatud printsiibid, mille hulgas ka läbipaistvus, liiga väheseks, et kõik intervjueeritud avaliku sektori eksperdid peaksid neid tehisintellekti kasutuselevõtu puhul tähtsaks?

Strateegiate tulemuste analüüsist selgus, et näiteks Leedu riiklikus tehisintellekti strateegias on ise sätestanud eetilised ja õiguslikud suunised ning sealhulgas on üheks printsiibiks, et tehisintellekti rakendus peab olema läbipaistev. Sellest saab järeldada, et Leedu on panustanud mõnevõrra rohkem kui Eesti strateegia selle läbi, et läbipaistvust tähtsustada riigi siseselt ning kirjutada see strateegias eetilise printsiibina lahti. Soome tehisintellekti strateegias väärtustatakse läbipaistvust üldise printsiibina, mis on aluseks „hea tehisintellekti ühiskonnale“, mis sarnaneb Cheng et al.,

(2021) kirjeldatud sotsiaalselt vastutustundliku tehisintellekti kontseptsioonile. Tulemuste analüüsist selgus, et Soome strateegias küll ei kirjeldata täpsemalt, millest läbipaistvus seisneb või kuidas seda saavutada, siiski saab järeldada, et läbipaistvust peetakse ühiskondlikult oluliseks väärtuseks, millest tehisintellekti arendamisel ja kasutuselevõtul lähtuda.

Minu tulemuste analüüsist selgus, et Rootsi ja Norra on samuti kirjeldanud läbipaistvuse tähtsuse eetilise põhimõttena sarnaselt Leedule, küll aga on läinud ka selles mõttes teistest riikidest teemaga rohkem süvitsi, et kirjeldanud ära, mida toob kaasa endaga läbipaistvuse puudumine tehisintellekti arenduste juures. Norra on toonud välja sarnaselt teooriatega (Haenlein & Kaplan, 2019; Larsson & Heintz, 2020; Kemper & Kolkman, 201) probleemkoha, kus mõnda süvaõppe algoritmi saab võrrelda nn „musta kastiga“, kus inimesel pole juurdepääsu mudelile, mis seletaks, miks antud sisendväärtus annab antud tulemuse. Norra tehisintellekti strateegias on mõistetud ka seda, et „nn musta kasti“ selgitamine ei saa olla ainult koodi avaldamine, sest sellest ei piisaks, mida kinnitavad ka teooriad (Haenlein & Kaplan, 2019; Larsson & Heintz, 2020; Kemper & Kolkman, 201). Norra lisab veel sealjuures, et juurdepääsu andmine täielikule andmekogumile, mida algoritm loob, võib rikkuda intellektuaalomandi õigusi ja andmekaitseseadusi. Sellest saab järeldada, et nii Norra kui Rootsi on erinevalt teistest riikidest tehisintellekti läbipaistvuse puudumise riskid kaardistanud ning seletanud ka piiranguid läbipaistvuse saavutamisel.

Kokkuvõttes saab välja tuua, et arusaamad tehisintellektist ja tema toimimisest on intervjueeritud avaliku sektori ekspertide seas sarnased ning eksperdid oskasid ka välja tuua tehisintellekti arendamise ja kasutuselevõtuga kaasnevaid riske. Arusaamade kohta läbipaistvusest tõdesid eksperdid, et tehisintellekti puhul on seda raske tõlgendada. Kõige paremini tähendab läbipaistvus tehisintellektist aru saamist, kuidas see toimib. Läbipaistvust ei pidanud tehisintellekti puhul oluliseks kõik eksperdid. Tulemuste analüüsist tuli välja, et ühel juhul ei pea tehisintellekti süsteem olema alati läbipaistev näiteks lõpptarbijale, kui see täidab oma eesmärki. Teisel juhul ei ole läbipaistvus oluline siis, kui tehisintellekt täidab oma eesmärki näiteks müügiettevõttes. Strateegiate puhul saab kokkuvõttes välja tuua, et riiklikes strateegiates on läbipaistvus sõnastatud eetilise printsiibina. Eesti juures on läbipaistvus paika pandud erinevate koostöögruppide üleselt soovitusliku põhimõttena. Leedu on seevastu oma strateegias pannud paika ise eetilised printsiibid, millest tehisintellekti arendamise ja kasutamise juures lähtuda. Soome väärtustab läbipaistvust üldise ühiskondliku väärtusena sätestades selle üldise printsiibina „heast tehisintellekti ühiskonnast“. Nii Norra kui ka Rootsi on lisaks eetilise põhimõttena tõlgendanud läbipaistvuse ühe tehisintellektiga kaasneva riskina ning ühtlasi näinud ka teistest analüüsitud strateegiatest

muid võimalikke probleeme ja raskusi, mida tehisintellekti arendamine ja kasutuselevõtt kaasa toob.

Käesoleva magistritöö fookus oli Põhja- ja Baltimaade tehisintellekti strateegiate analüüsil läbipaistvusest tehisintellekti kontekstis. Tulevikus tasuks uurida ka läbipaistvuse käsitlemist strateegiates Euroopa kontekstis, mõistmaks, millised on tõlgendused erinevate Euroopa riikide lõikes ning kuidas need teineteisest erinevad. Mõistmaks arusaamu tehisintellekti läbipaistvusest, saaks tulevikus uurida Eesti olemasolevate tehisintellektide kasutajaid, mõistmaks ka nende arusaamu tehisintellekti läbipaistvusest ja ka selle olulisusest neile.

KOKKUVÕTE

Käesoleva magistritöö eesmärgiks oli välja selgitada, kuidas mõistavad tehisintellekti ja selle läbipaistvust Eesti avaliku sektori eksperdid ning kuidas on läbipaistvust tõlgendatud Põhja- ja Baltimaade tehisintellekti strateegiates. Magistritöö teema valikul sai määravaks asjaolu, et läbipaistvuse teemat ei ole Eestis tehisintellekti kontekstis varem uuritud. Samuti ei seletata seda ka tihti erinevates poliitikadokumentides lahti, mida läbipaistvus täpsemalt tähendab. Magistritöö on väärtuslik kindlasti poliitikakujundajatele, kes annavad panuse tehisintellekti hõlmavate aruannete ja arengukavade koostamisse. Samuti on tehisintellekti läbipaistvusest mõistmine oluline ka laiemas ühiskondlikus kontekstis puudutades ka tavakodanikku, keda tehisintellektil põhinevad süsteemid juba puudutavad või hakkavad tulevikus puudutama.

Magistritöö eesmärgist lähtuvalt püstitasin järgmised uurimisküsimused:

- Millised on arusaamad tehisintellektist infotehnoloogia vallas töötavate ekspertide seas?
- Millised on arusaamad tehisintellekti läbipaistvusest infotehnoloogia vallas töötavate ekspertide seas?
- Kuidas on tehisintellekti läbipaistvus sõnastatud Põhja- ja Baltimaade riiklikes strateegiates?

Magistritöö raames sai läbi viidud poolstruktureeritud intervjuud Eesti avaliku sektori ekspertidega, kes omavad teadmisi ja kogemusi tehisintellekti vallas. Intervjuud andsid väärtusliku panuse mõistmaks, millised on arusaamad tehisintellektist ja sellega seonduvast läbipaistvusest Lisaks sai viidud läbi kontentanalüüs, millesse olid kaasatud Põhja- ja Baltimaade riiklikud tehisintellekti strateegiad.

Tulemustest selgus, et Eesti avaliku sektori eksperdid on üksmeelel tehisintellekti olemusest ning mõistavad seda üldiselt kui mudelit, mida on võimalik treenida. Intervjueeritud avaliku sektori eksperdid teadvustavad tehisintellekti arendamise ja kasutuselevõtuga kaasnevaid probleeme, tuues välja näiteks treeningandmete kallutatust. Läbipaistvuse tõlgendamisel olid eksperdid samuti

üksmeelel, et see on üldiselt võime aru saada, kuidas tehisintellekt töötab. Tulemustest selgus ka see, et läbipaistvust on tehisintellekti puhul pigem raske tõlgendada. Olulise tulemusena saab välja tuua seda, et eksperdid pidasid läbipaistvust tähtsaks, siiski leidus ka vastupidist arvamust. Tehisintellekti puhul ei nähtud läbipaistvust olulisena, kui tehisintellekt täidab oma eesmärki lõpptarbija jaoks või müügile orienteeritud ettevõttes.

Enamikus Põhja- ja Baltimaade strateegiates on läbipaistvus sõnastatud eetilise printsiibina. Tulemustes selgus, et Leedu on oma tehisintellekti strateegias ise paika pandud eetilised põhimõtted, mille hulgas on ka kirjeldatud läbipaistvuse olulisust. Soome väärtustab läbipaistvust ühiskondliku väärtusena, kuna on sätestanud selle üldise printsiibina „heast tehisintellekti ühiskonnast. Oluline tulemus on see, et Eestis kirjeldatakse läbipaistvust soovitusliku eetilise suunisena läbi erinevate koostöö organisatsioonide, detailsemalt aga pole mõistet lahti seletatud. Rootsi ja Norra tehisintellekti strateegiates on seevastu nähtud läbipaistvuse puudumist üheks võimalikuks tehisintellekti arendamise ja kasutuselevõtuga kaasnevaks riskiks. Tulemustest selgus, et ka, et ka Leedu ja Soome on tehisintellekti juures välja toonud murekohti. Oluline tulemus on aga ka see, et Eesti tehisintellekti aruandes riske puudutatud pole.

SUMMARY – Transparency in artificial intelligence in Nordic and Baltic strategies and Estonian public sector experts' perceptions of it

The aim of this master's thesis was to find out how Estonian public sector experts understand artificial intelligence and its transparency, and how transparency has been interpreted in the Nordic and Baltic artificial intelligence strategies. The topic of transparency has not been studied in the context of artificial intelligence in Estonia, nor is it often explained in various policy documents, specifically, what exactly does transparency mean. The topic of the thesis was chosen for this particular reason. My master's thesis is certainly valuable for policy makers and average citizens who are already affected or will be affected in the future by artificial intelligence systems. Understanding the transparency of artificial intelligence is important in the wider societal context.

I raised the following research questions based on the aim of the master's thesis:

- What are the perceptions of artificial intelligence among IT experts?
- What are the perceptions of artificial intelligence transparency among IT experts?
- How is the transparency of artificial intelligence formulated in the Nordic and Baltic national strategies?

The framework of the master's thesis contains semi-structured interviews with Estonian public sector experts who have knowledge and experience in the field of artificial intelligence. Interviews provided valuable input to understanding the perceptions of artificial intelligence and related transparency. In addition, a content analysis, which included Nordic and Baltic national strategies, was carried out.

The results showed that Estonian public sector experts agree on the nature of artificial intelligence and generally understand it as a model that can be trained. My research showed that public sector experts are aware of the problems caused by the use of artificial intelligence (e.g bias in training data). In interpreting transparency, public sector experts agreed that it can generally be defined as

the ability to understand how artificial intelligence works. However, the results showed and some experts pointed out that transparency in artificial intelligence is rather difficult to interpret. Whilst experts generally considered transparency to be important, contrary cases did emerge during the interviews. For example, transparency was not seen as important from the final consumer perspective or in case artificial intelligence fulfills its desired purpose (e.g sales-oriented company).

Transparency is formulated as an ethical principle in most Nordic and Baltic strategies. The analysis shows that Lithuania has established ethical principles in its artificial intelligence strategy, including the importance of transparency. Finland values transparency as a universal societal value, as it has established it as a general principle of "good artificial intelligence society". The results show that in Estonia, transparency is described as a recommended ethical guideline through various co-operation organizations, but the concept has not been explained in more detail. Swedish and Norwegian artificial intelligence strategies, on the other hand, have identified the lack of transparency as one of the potential risks associated with the development and deployment of artificial intelligence. The findings indicate that Lithuania and Finland have also raised concerns about artificial intelligence. However, the analysis results importantly show that the Estonian artificial intelligence report does not address risks.

KASUTATUD KIRJANDUS

Agudo, U., & Matute, H. (2021). *The influence of algorithms on political and dating decisions.* Plos One, 16(4). https://doi.org/10.1371/journal.pone.0249454

AI Watch veebilehekülg. Kasutatud 10.04.2021, https://knowledge4policy.ec.europa.eu/ai-watch_en#country

Algorithm Watch & Bertelsmann Stiftung. (2020). *Automating Society Report.* Kasutatud 27.03.2021, https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf

Alvarez, Y., Leguizamón-Páez, M. A., & Londoño, T. J. (2021*). Risks and security solutions existing in the Internet of things (IoT) in relation to Big Data*. Ingeniería y Competitividad, 23(1), 1–13.

Ananny, M. (2016). *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*. Science, Technology, & Human Values, 41(1), 93–117. https://doi.org/10.1177/0162243915606523

Ananny, M., & Crawford, K. (2018). *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.* New Media & Society, 20(3), 973–989. https://doi.org/10.1177/1461444816676645

Bowen, G. A. (2009). *Document Analysis as a Qualitative Research Method*. Qualitative Research Journal, 9(2), 27–40. https://doi.org/10.3316/QRJ0902027

boyd, danah, & Crawford, K. (2012). *Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon*. Information, Communication & Society, 15(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Cheng, L., Varshney, K. R., & Liu, H. (2021). *Socially Responsible AI Algorithms: Issues, Purposes, and Challenges*. Kasutatud 02.05.2021, https://ui.adsabs.harvard.edu/abs/2021arXiv210102032C/abstract

Couldry, N. (2020). *Recovering critique in an age of datafication*. New Media & Society, 22(7), 1135–1151. https://doi.org/10.1177/1461444820912536

E-Estonia veebilehekülg. Kasutatud 14.05.2021, https://e-estonia.com/

Elish, M. C., & boyd, danah. (2018). *Situating methods in the magic of Big Data and AI*. Communication Monographs 85(2), 1-24. https://doi.org/10.1080/03637751.2017.1375130

Euroopa Komisjon. (2020). *Tehisintellekt: Euroopa käsitus tipptasemel ja usaldusväärsest tehnoloogiast.* Kasutatud 04.04.2021, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_et.pdf

Euroopa Liidu Põhiõiguste Amet. (2020). *Põhiõiguste aruanne 2020 – FRA arvamused*. Kasutatud 04.04.2021, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-fundamental-rights-report-2020-opinions_et.pdf

Fast, E., & Horvitz, E. (2017). *Long-Term Trends in the Public Perception of Artificial Intelligence*. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1).

Gandomi, A., & Haider, M. (2015). *Beyond the hype: Big data concepts, methods, and analytics.* International Journal of Information Management, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gray, D. E. (2004). *Doing Research in the Real World*. Sage Publications Ltd.

Haenlein, M., & Kaplan, A. (2019). *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*. California Management Review, 61(4), 5–14. https://doi.org/10.1177/0008125619864925

Jobin, A., Ienca, M., & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kalmus, V. (2015). *Standardiseeritud kontentanalüüs*. Kasutatud 10.05.2021,

http://samm.ut.ee/kontentanalyys

Kalmus, V., Masso, A., & Linno, M. (2015). *Kvalitatiivne sisuanalüüs*. Kasutatud 10.05.2021, https://samm.ut.ee/kvalitatiivne-sisuanalyys

Kaplan, A., & Haenlein, M. (2019). *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*. Business Horizons, 62(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kemper, J., & Kolkman, D. (2019). *Transparent to whom? No algorithmic accountability without a critical audience*. Information, Communication & Society, 22(14), 2081–2096. https://doi.org/10.1080/1369118X.2018.1477967

Krattide veebileht. (i.a.). Kasutatud 18. aprill 2021, https://www.kratid.ee

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Penguin.

Lagerspetz, M. (2017). *Ühiskonna uurimise meetodid*. TLÜ Kirjastus.

Larsson, S. (2017). *Sustaining Legitimacy and Trust in a Data-Driven Society*. Ericsson Technology Review, 94(2), 40–49.

Larsson, S., & Heintz, F. (2020). *Transparency in artificial intelligence*. Internet Policy Review, 9(2). https://doi.org/10.14763/2020.2.1469

Loi, M., & Spielkamp, M. (ia). *Towards accountability in the use of Artificial Intelligence for Public Administrations*. Kasutatud 21.05.2021, https://algorithmwatch.org/en/wp-content/uploads/2021/05/Accountability-in-the-use-of-AI-for-Public-Administrations-AlgorithmWatch-2021.pdf

Majandus- ja Kommunikatsiooniministeerium. (2021). *Eesti digiühiskond 2030*. Kasutatud 20.05.2021, https://mkm.ee/sites/default/files/eesti_digiuhiskond_2030.pdf

Majandus- ja Kommunikatsiooniministeerium & Riigikantselei. (2020). *Eesti tehisintellekti kasutuselevõtu eksperdirühma aruanne*. Kasutatud 13.02.2021, https://f98cc689-5814-47ec-86b3-db505a7c3978.filesusr.com/ugd/0b32e3_9e397d14453b454db0b8d3615a7012ba.pdf

Marr, B. (2019). *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems.* John Wiley & Sons, Incorporated.

Mayring, P. (2000). *Qualitative Content Analysis*. Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research, 1(2). https://doi.org/10.17169/fqs-1.2.1089

Mejias, U. A., & Couldry, N. (2019). *Datafication*. Internet Policy Review, 8(4). https://doi.org/10.14763/2019.4.1428

Ministry of Economic Affairs and Employment. (2017). *Finlan d's Age of Artificial Intelligence*. Kasutatud 13.03.2021, https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

Ministry of Enterprise and Innovation. (2018). *National approach to artificial intelligence*. Kasutatud 14.03.2021, https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf

Ministry of the Economy and Innovation. (2019). *Lithuanian Artificial Intelligence Strategy*. Kasutatud 14.03.2021, http://kurklt.lt/wp-content/uploads/2019/04/DI_strategija_ENG.pdf

Muhammad, I., & Yan, Z. (2015). *Supervised Machine Learning Approaches: A Survey.* Journal on Soft Computing, *5*(3), 946–952. https://doi.org/10.21917/ijsc.2015.0133

Nordic Co-operation veebileht. *The Nordic-Baltic region: A digital frontrunner*. Kasutatud 15.03.2021, https://www.norden.org/en/declaration/nordic-Baltic-region-digital-frontrunner

Norwegian Ministry & of Local Government and Modernisation. (2020). *National Strategy for Artificial Intelligence.* Kasutatud 20.03.2021, https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf

Pepito, J. A., Vasquez, B. A., & Locsin, R. C. (2019). *Artificial Intelligence and Autonomous Machines: Influences, Consequences, and Dilemmas in Human Care*. Health, 11(07), 932. https://doi.org/10.4236/health.2019.117075

Redden, J. (2018). *Democratic governance in an age of datafication: Lessons from mapping government discourses and practices*. Big Data & Society, 5(2). https://doi.org/10.1177/2053951718809145

Robinson, S. C. (2020). *Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI)*. Technology in Society, 63. https://doi.org/10.1016/j.techsoc.2020.101421

Strauß, S. (2015). *Datafication and the Seductive Power of Uncertainty—A Critical Exploration of Big Data Enthusiasm*. Information, 6(4), 836–847. https://doi.org/10.3390/info6040836

Zeadally, S., Das, A. K., & Sklavos, N. (2019). *Cryptographic technologies and protocol standards for Internet of Things*. Internet of Things. https://doi.org/10.1016/j.iot.2019.100075

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). *Machine learning on big data: Opportunities and challenges.* Neurocomputing, 237, 350–361. https://doi.org/10.1016/j.neucom.2017.01.026

Van Dijck, J. (2014). *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology*. Surveillance & Society, 12(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Warwick, K. (2011). *Artificial Intelligence: The Basics*. Taylor & Francis Group.

LISAD

LISA 1 Kodeerimisjuhis

A. Tehisintellekti strateegia tunnused

A1 Strateegia välja andnud riik

A2 Strateegia pealkiri

A3 Ilmumise aasta

B. Läbipaistvus peamiste teemadena

B1 Läbipaistvus koostöögruppide üleselt

1. EL tasandil eksperdirühma suunistes
2. OECD tasandil tehisintellekti soovitustes
3. Põhjala- Balti koostöögrupi tasandil koostöösuundades

B2 Läbipaistvus sätestatud strateegias suunistena

1. Läbipaistvus eetilise ja õigusliku printsiibina
2. Läbipaistvus üldise printsiibina

B3 Tehisintellektiga kaasnevad riskid

1. Läbipaistvuse puudumine
2. Usalduse kaotus
3. Töökohtade kaotus
4. Tehisintellekti väärkasutamine või vaenulik kasutamine
5. Rahaline kahju
6. Ohud demokraatia toimimisele
7. Küberrünnakud
8. Kallutatud või manipuleeritud andmed
9. Eetilised riskid
10. Diskrimineerimine
11. Vastutuse puudumine

LISA 2 Intervjuukava

1. Intervjueeritavate tutvustus
    a. Mis on Teie taust infotehnoloogia vallas (haridus, valdkond)?
    b. Millisel viisil olete seotud olnud tehisintellekti hõlmavate arendustega Eestis?
2. Alustuseks
    a. Mis Te arvate, kuidas andmestumine avalikku sektorit muutnud on?
3. Tehisintellekti mõiste
    a. Palun rääkige, kuidas Teie mõistate tehisintellekti. Mis see on ja milleks see vajalik on?
    b. Millised on Teie arvates tehisintellekti kasutegurid?
    c. Oskate ehk mõne näite tuua, kus teie arvates tehisintellekti võiks kasutada või kaustatakse hästi?
    d. Millised on Teie arvates tehisintellektiga seonduvad probleemid?
    e. Oskate Te ka siin tuua mõnda näidet, kus tehisintellektide puhul võib esineda probleeme?
    f. Kas nendele probleemidele on täna lahendusi ja milliseid?
        i. Kuidas saaks neid probleeme tulevikus vältida?
4. Tehisintellekti roll Eestis
    a. Milline on olukord tehisintellekti vallas Eestis?
        i. Milline on olukord erasektoris?
        ii. Milline on olukord avalikus sektoris?
5. Tehisintellekti arendamine
    a. Mis te arvate, mida on vaja selleks, et tehisintellekti arendada?
    b. Millistel viisidel mõjutab tehisintellekti arendamist arendaja ise?
    c. Palun rääkige mida peaksid arendajad tehisintellekti arendamisel silmas pidama ja miks?
        i. Millistest põhimõtetest tuleks lähtuda?
6. Põhimõtted tehisintellekti arendamisel
    a. Euroopa Komisjon kutsus 2018. aastal kokku eksperdirühma, kes sai ülesandeks välja töötada tehisintellekti arendamise ja kasutamise eetikasuuniseid. Eksperdirühma suunised on järgmised, mille kohaselt peab krattide

usaldusväärsuse saavutamiseks olema täidetud kolm tingimust: kratt peab 1) vastama seadusele, 2) olema kooskõlas eetikapõhimõtetega ja 3) olema töökindel. Nendele tingimustele vastamiseks on omakorda välja toodud seitse põhinõuet, mis kujundavad usaldusväärse krati kontseptsiooni: inimese toimevõime (ingl human agency) ja järelevalve; tehniline töökindlus ja ohutus; privaatsus ja andmehaldus; läbipaistvus; mitmekesisus, mittediskrimineerimine ja õiglus; ühiskondlik ja keskkonnaalane heaolu; vastutuse võtmine (Majandus- ja Kommunikatsiooniministeerium & Riigikantselei, 2020). Mida Te nendest suunistest arvate?

   b. Palun rääkige oma kogemusele tuginedes, millest peaks tehisintellekti arendamisel lähtuma?

7. Läbipaistvus
   a. Mida tähendab teie jaoks läbipaistvus?
   b. Mida tähendab teie jaoks läbipaistvus tehisintellekti kontekstis?
   c. Rääkige palun, mis te arvate, kelle jaoks ja miks on läbipaistvus oluline?
   d. Mida te arvate oma kogemuste Põhjal, milliste lähenemiste kaudu on võimalik läbipaistvust saavutada?
   e. Kuidas võib mõjutada arendajate arusaam läbipaistvusest tehisintellekti?
8. Tehisintellekt tulevikus
   a. Milline on teie hinnangul tehisintellekti tulevik? Eestis?
9. Lõpetuseks
   a. Kas Te sooviksite midagi täiendavalt lisada?

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Heli Orav,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Tehisintellekte puudutava läbipaistvuse käsitlemine Põhja- ja Baltimaade strateegiates ning Eesti avaliku sektori ekspertide arusaamad sellest“, mille juhendaja on Maris Männiste, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Heli Orav
25.05.2021