Kaimar Karu

Tehisintellekti keerukad küsimused[*1]

Euroopa Komisjoni tehisintellekti arendamise koordinatsiooni ja potentsiaalse regulatsiooni[*2], samuti Eesti „kratiseaduse“[*3] koostamise kontekstis on tõusnud mitu olulist küsimust ning keerukat ülesannet, mille ilmingud ja praktiline lahendamise vajadus on tihti juriidika, kuid juured ja lahendusvõimalused pigem ühiskonna, filosoofia ja tehnoloogia valdkonnas.

Alljärgnevas tekstis käsitlen sedasorti keerukaid põhimõttelisi küsimusi, millega tegelemisest ei pääse ei tehisintellekti reeglite väljatöötamisel ega praktiliste rakendus- ja kontrollmehhanismide juurutamisel. Enamasti on tegemist „enne kui“, mitte „alles siis, kui“ küsimustega, millele vastamata jätmisel jääb tehisintellekti vundament nõrgaks ning seatud eesmärkide saavutamine ebatõenäoliseks.

1. Alustame mõistetest

Tehisintellekti võib jagada kolmeks tüübiks vastavalt selle võimekuse tasemele.[*4] Esimene tase ehk **ahas tehisintellekt** (ATI) eeldab algoritmidelt inimintellektitaolist hakkamasaamist üht tüüpi ülesannetega, sisaldab piiratud õppimisvõimet ning on mõeldud kitsa funktsiooni täitmiseks. Teine tase ehk **inimesetaoline tehisintellekt** (ITI) oskab kombineerida ning lahendada eri tüüpi ülesandeid inimintellektiga võrdväärsel tasemel ja toimib sarnaselt sellele, kuidas me kujutame ette teadvusega indiviidi oskuslikku toimetamist. Kolmas tase ehk **ülim tehisintellekt** (ÜTI) ületab inimkonna kombineeritud intellekti taset olulisel määral.

Lisaks nimetatakse tehisintellektiks kohati tarkvaralahenduse automatiseerimisvõimekust, mille reeglistik põhineb etteantud tingimuslausetel ning millel õppimisvõime puudub. Selline funktsionaalsus võib olla suureks abiks rügamisvajaduse (ingl *toil*) vähendamisel, kuid tegemist pole tehisintellektiga.

Käesolevas artiklis kasutan sõna „tehisintellekt“ edaspidi ahta tehisintellekti tähenduses, kui pole märgitud teisiti.

Võimaluste mitmekesisuse ja võimekuse piiride ebaselguse tõttu erineb tehisintellekt enamikust teistest viimaste aastakümnete jooksul tavakasutusse jõudnud tehnoloogiatest. Samamoodi nagu varasemaid uuenduslikke ideid on ka tehisintellekti mõistet aeg-ajalt rakendatud turundusvankri ette ilma sisu lisamata, kuid tunnetuslikult on see tehnoloogia varasemast oluliselt erinev, tihti ka ohtlikuna näiv.[*5]

Tehisintellekt kõigutab vähemalt potentsiaalselt nii harjumusi kui ka tavapäraseid protsesse ja protseduure ka nendes valdkondades, kus uute tehnoloogiate rolli on varem nähtud pigem piiratud võimekusega abistaja, mitte suunajana. Sellest tuleneva ebakindluse ja eelmainitud ohutunde üheks väljundiks võib olla soov tundmatut (üle)reguleerima asuda.

Selleks et midagi reguleerida, tuleb see selgelt defineerida ja piiritleda. Püüdluses tehisintellekti defineerida liigutakse juba aastaid konkreetsuse ja üldsõnalisuse äärmuse vahel ning ka värske ja põhjalik erinevate definitsioonide analüüs[*6] pakub vaid esimesed sammud olukorra parandamiseks.

---

1 Artikkel põhineb 08.–09.10.2020 toimunud 36. Eesti õigusteadlaste päevadel peetud ettekandel.

2 Artificial Intelligence. European Commission. Arvutivõrgus: https://ec.europa.eu/digital-single-market/en/artificial-intelligence.

3 Algoritmiliste süsteemide mõjude reguleerimise väljatöötamise kavatsus („krati VTK“). Arvutivõrgus: https://adr.rik.ee/jm/dokument/7458502.

4 Ingl *Artificial Narrow Intelligence* (ANI), *Artificial General Intelligence* (AGI), *Artificial Superintelligence* (ASI).

5 Benefits and risks of Artificial Intelligence. Future of Life Institute. Arvutivõrgus: https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/.

6 AI WATCH. Defining Artificial Intelligence. European Commission. Arvutivõrgus: https://ec.europa.eu/jrc/en/publication/ai-watch-defining-artificial-intelligence.

Täpne definitsioon on muuhulgas oluline näiteks selleks, et tehisintellekti puudutava ning praegu paraku osaliselt ka tehnoloogilisest võõrahirmust ajendatud regulatsiooni alla ei satuks lihtsalt kõik need tarkvaralahendused, mis toimivad millisel iganes viisil, millest tavakasutaja aru ei saa.

Üheks näiteks saab tuua eespool mainitud süsteemid, kus rakendatakse kasutaja jaoks nähtamatut protsesside automatiseerimist – kirjeldus, mille alla mahub suur osa tänapäevastest infosüsteemidest. Selline eriregulatsiooni liialt laiahaardeline rakendus halvaks riikide ja tervikuna Euroopa Liidu arenguvõimalused selles valdkonnas ning kaotaks meie (potentsiaalsed) konkurentsieelised.

Kuigi ühest ja laialdaselt aktsepteeritud definitsiooni praegu veel pole, saame me selleteemalistes vestlustes üksteisest justkui aru. Tegelikkuses tekib aga definitsioonide mitmekesisusest ning laialivalguvusest palju segadust, sest erinevad on ka eeldused ja ootused. Lisaks eespool toodud kolmele võimekuse tasemele (ATI, ITI ja ÜTI), mis aitavad piiritleda seda, kas räägime tänasest või tuleviku tehisintellektist, on minu hinnangul kasu ka eelduste kaardistamisest. Seetõttu pakun siinkohal välja kolm eelduste ja konteksti piiritlemise postulaati, mis on ka artikli alusprintsiipideks.

Esiteks: tehisintellektil puudub teadvus. Tänapäevased ning lähitulevikus loodavad tehisintellekti rakendused ei tunne, arva, eelista ega unista selliselt, nagu seda teevad inimesed. Meil tuleb vältida tehisintellekti antropomorfiseeritud käsitlust seni, kuni tehisintellekti võimekus pole teadvuse tasemele tõusnud. Tehisintellekti puhul ei saa me rääkida vabast tahtest, tahtlikkusest ja uskumustest.

Teiseks: tehisintellekti rakendused tõlgendavad tõenäosuslikult, mitte tunnetuslikult. Mustrite avastamine ja analüüs on keerukates süsteemides statistilise iseloomuga ning põhjuslikkust võib, kuid ei pruugi esineda.

Kolmandaks: tehisintellekti rakendused ei anna sisulist (õiguslikku, moraalset vms) hinnangut. Meil tuleb eristada tehisintellektipoolset andmetöötlust ja esitletud tulemusi ning nende põhjalt tehtavaid järeldusi, otsuseid ja tagajärgi.

2. Tänased ootused tehisintellektile

Tehnoloogiasektorile loomuomaselt vaevleb ka tehisintellekt (mõnede eranditega[*7]) suurte ootuste ning nendega võrreldes aeglaselt saabuva tulemuslikkuse konfliktis. Kolm „talve“[*8] üle elanud valdkond on tänu pilvetehnoloogia poolt toetatud arvutusvõimsuse kättesaadavuse märkimisväärsele paranemisele muutunud küll oluliseks, sisuliseks, äriliseks ja administratiivseks lisavõimekuseks, kuid ka tänasel päeval tasub alati kontrollida, mida tähendab tarkvarasüsteemi, otsustusprotsessi või „uue revolutsioonilise lahenduse“ müügimaterjalides kasutatav sõna „tehisintellekt“.

Riiklikke[*9] ja rahvusvahelisi[*10] visioonidokumente lugedes näeme, et tehisintellekti kasu nähakse laiemalt tehnoloogilise uuendusena ning sellele seatakse ootusi mitmest ühiskondlikust aspektist. Euroopa Komisjon näiteks jaotab tehisintellekti ökosüsteemi arendamisest loodetu kolmeks: kodanikku, äriettevõtteid ja avalikku huvi teenivaks kasuks.[*11]

Riigiaparaadi toimimise kontekstis on oodatud kasutegurina tihti arutlusel kulude kokkuhoid, kuid siinkohal kutsuks ma üles vaoshoitusele. Tehnoloogiliste uuenduste uksest sisse saamiseks kasutatakse seda argumenti sageli, reaalsuses seab see aga vale fookuse ning viib tihti ka pettumuseni.

Pigem võib oodata seda, et uue tehnoloogia kasutuselevõtu järel kasvavad mingil määral kulud ning olulisel määral (loodetavasti) ka tulemuslikkus. Targad investeeringud tehnoloogiasse aitavad meil tuttavaid ülesandeid lahendada efektiivsemalt ning asuda lahendama uusi ülesandeid, mis valdkonna või terviku võimekust suurendavad. Selliste investeeringute puhul on kulude kasv võrreldes produktiivsuse kasvust tulenevate kasudega vastuvõetaval tasemel.

---

7 AlphaFold: a solution to a 50-year-old grand challenge in biology. – DeepMind, 30.11.2020. Arvutivõrgus: https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology.

8 *AI winter* ehk eriloomuliste kõrgendatud ootuste mitterahuldamisele järgnenud populaarsuse (ja rahastuse) langusperioodid 1960-ndatel, 1970-ndatel ning 1980-ndatel.

9 Kratid Eesti heaks. Arvutivõrgus: https://www.kratid.ee/.

10 United Nations Activities on Artificial Intelligence (AI). International Telecommunication Union. Arvutivõrgus: http://handle.itu.int/11.1002/pub/8120d5d5-en.

11 White Paper on Artificial Intelligence – A European approach to excellence and trust. COM(2020) 65 final. Brussels, 19.02.2020. Arvutivõrgus: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

Kirjeldatud eesmärgi saavutamiseks tuleb vaadata tervikpilti, mitte pelgalt selle tehnoloogilist komponenti. Tehnoloogia peab aitama kaasa eesmärkide saavutamisele ning toetama selleks loodud protsesse ja protseduure. See kõik aga nõuab omakorda nii eesmärkide selgeks rääkimist, strateegias kokkuleppimist, protsesside kogupaketi ülevaatamist-uuendamist kui ka sisukate mõõdikute seadmist.

Eesti kontekstis on sagedamini mainimist leidnud kaks konkreetset eesmärki: menetlustoimingute automatiseerimine ning nähtamatute (sündmus)teenuste disain. Mõlemad vajavad konkreetsete eelduste täitmist.

3. Avaandmed kui vundament ja võimaldaja

Mitme tehisintellektile seatud eesmärgi realiseerumise üheks oluliseks eelduseks on kvaliteetsete andmete kättesaadavus, sest tehisintellekti rakenduste aluseks oleva masinõppe tulemuslikkus sõltub muuhulgas algoritmide treenimise edukusest. See põrkub harilikult aga mitme takistuse vastu.

Kui 2019. aasta Euroopa Andmeportaali avaandmete küpsuse uuringu[*12] tulemusena paiknesime me teiste riikidega võrreldes küll esimeses pooles, kuid kaugeltki mitte liidrite hulgas, siis värske 2020. aasta uuringu[*13] põhjal oleme tõusnud juba viiendale kohale. Me oleme lühikese ajaga suutnud oluliselt parandada enda tulemusi peaaegu kõigis 16 kriteeriumis, mille järgi riike hinnatakse. Need kriteeriumid on omakorda jaotatud neljaks temaatiliseks aspektiks: riiklik andmepoliitika, andmete ligipääsukeskkond, andmete kvaliteet ning andmete kasutamise mõju. Just viimases aspektis on meie tulemused tänu majandusliku mõju selgele tuvastamisele märkimisväärselt paranenud. Vaatame kõiki nelja nüüd lähemalt.

Esiteks andmete kättesaadavus. 2015. aastal avatud Eesti avaandmete portaalis[*14] on hetkeseisuga ligi 800 andmestikku rohkem kui 100 teabevaldajalt. Nendes andmestikes olevad andmed toetavad nii statistilist analüüsi, prognooside tegemist kui ka uute tehnoloogiate arendamist. Portaalis on vabalt kättesaadavad ka rohkem kui 60 andmepõhist rakendust põllumajandusest demograafiliste muudatusteni, palgaandmetest COVID-19 levikuni Eestis.

Teiseks andmete kvaliteet. Kuna andmekasutus otsuste tegemisel on traditsiooniliselt pigem tagasihoidlik, on raske andmete kvaliteedi parandamiseks vajalikke tegevusi ja investeeringuid põhjendada. Isegi siis, kui andmete korrastamise vajadust on mõistetud, toimub see tihti hoogtöö korras ühekordse või juhusliku sagedusega tegevusena. Samuti on meie (ning muidugi ka teiste riikide) andmekogusid juba esimestest päevadest alates kummitanud andmekogu struktuuri ja sisu selgitavate metaandmete halb kvaliteet või suisa puudumine.

Tänuväärselt on näiteks Statistikaamet otsustanud kogutud andmed ka andmete esitajate vahetusse teenistusse rakendada, avades selleks juhtimislaudade portaali.[*15] Lisaks sellele, et info kahesuunaline liiklus on hea koostöö märk, mõjutab see pikema aja jooksul loodetavasti ka andmete esitajate käitumist, sest tekib selge seos esitatud andmete ning ametkondlikult koostatud ja eraettevõtluse kasutusse antud analüüsi kvaliteedi vahel.

See toob meid kolmanda punkti, andmete kasutamise mõju juurde. Ei riigid ega ettevõtted pole harjunud sellega, et ühtäkki on meil ligipääs suurtele andmehulkadele, mida vähemasti potentsiaalselt on võimalik kasutada oluliste ja kohati kiirelt nõutud otsuste tegemisel, mis omakorda vähemalt potentsiaalselt annab tänases maailmas olulise konkurentsieelise.

Kogunenud andmeid on varem tihti käsitatud arhiivimaterjalina, mille puhul andmete säilitamine võib olla ka (seaduslik) kohustus, kuid andmete hilisem kasutamine on ebaselge kasuteguriga või „kellegi teise mure“. On tõenäoline, et tehisintellekti rakenduste pakutav potentsiaal kiirendab andmete kasutamise kasuteguri paremat mõistmist nii paremate otsuste tegemise kui ka nende põhjendamise tarbeks, sest nii kodanikud kui ka kliendid nõuavad otsustelt üha suuremat läbipaistvust ning endale suuremat jälgitavat rolli otsuste mõjutamisel.

Neljandaks ja äärmiselt oluliseks punktiks on riiklik andmepoliitika ning selle praktiline rakendamine. Siin on Eesti teinud palju sisulist tööd ja saavutanud häid tulemusi nii kohalikus kontekstis kui ka

12 Open Data in Europe 2019, https://www.europeandataportal.eu/en/dashboard/2019.

13 Open Data in Europe 2020, https://www.europeandataportal.eu/en/dashboard/2020.

14 Eesti avaandmete portaal. Arvutivõrgus: https://opendata.riik.ee/.

15 Eesti statistika juhtimislauad. Statistikaamet. Arvutivõrgus: https://juhtimislauad.stat.ee/.

rahvusvahelises võrdluses. Andmete ühekordse küsimise põhimõtte[*16] strateegiline üleriigiline rakendamine on eeskujuks ka Euroopa teistele riikidele. Eestil on juba mitu aastat ametis riigi *Chief Data Officer*.[*17] Koostamisel on avaandmete tegevuskava[*18], mis hõlmab nii seadusandlikke aspekte kui ka näiteks tarkvaraarenduslikke nõudeid, infomaterjalide koostamist ning avaliku sektori spetsialistide koolitusi.

Selleks et avalikus sektoris säiliks praegune hoog ning et ka erasektor tuleks avaandmete ja tehisintellekti teemaliste arutelude ja tegudega rohkem kaasa, on tarvis, et need teemad muutuksid riiklikul tasemel strateegilisteks rohkem kui paari ministeeriumi haldusalas ning et arendustegevusteks tagataks ka vajaminev püsiv rahastus. Ka siinkohal võib eeldada, et tehisintellekti rakenduste hoomatav kasu hoiab teema hästi nähtaval, kuid see ei taga, et lubadustele järgneb võimekuse kasv.

4. Kehtivad reeglid ja soovituslikud printsiibid

Kõige olulisem sisuline küsimus, millele rahuldaval tasemel vastamiseni suures kasutuselevõtu tuhinas tihti ei jõutagi, on selgus tehisintellekti rakendamise eesmärkides. Mis on need otsused, mida meil on tarvis teha, kuid milleks meil praegu puudub piisav informatsioon ning mille puhul me usume, et tehisintellekti rakendused meile vajaliku informatsiooni annavad? Millised on need (praegu veel piiratud) intellekti nõudvad tegevused, millest me soovime inimesed vabastada, asendades nende rolli masina omaga? Miks me arvame, et just tehisintellekt on see, mida me probleemi lahendamiseks vajame?

Olles olukorda analüüsides jõudnud arusaamale, et just tehisintellekti rakendus on tõepoolest see, mida me tõenäoliselt vajame, peame enda jaoks tegema selgeks kehtivad rahvusvahelised ja kohalikud reeglid, seatud või (tehnoloogiliselt) paratamatud piirangud, mõju hindamiseks vajalikud huvirühmad, olemasolevate lahenduste võimalused ning tehnoloogilised ja ühiskondlikud riskid.

2018. aastal kehtima hakanud isikuandmete kaitse üldmäärus[*19] otsesõnu tehisintellekti ei maini, kuid mitu määruses kirjeldatud reeglit mõjutavad (tänu oma üldisele sõnastusele) ka tehisintellekti rakenduste kasutuselevõttu. Euroopa Parlament on koostanud ka põhjalikuma ülevaate[*20] kahe teema kokkupuutepunktidest, mistõttu keskendun siinkohal vaid ühele olulise mõjuga aspektile.

Määruse artikli 22 lõige 1 sätestab järgmist:

> „Andmesubjektil on õigus, et tema kohta ei võetaks otsust, mis põhineb üksnes automatiseeritud töötlusel, sealhulgas profiilianalüüsil, mis toob kaasa teda puudutavaid õiguslikke tagajärgi või avaldab talle märkimisväärset mõju."

Seega peab tehisintellekti rakendust disainides analüüsima algoritmide rakendamise potentsiaalset mõju ning tõmbama selge piiri algoritmi tulemuse esitlemise ja selle tulemuse põhjal otsuste tegemise vahele. Lisaks tuleb otsustada, kas ja milliseid otsuseid tehakse automatiseeritult ja inimsilmale nähtamatult ning millised vajavad inimeste sekkumist. Lühidalt kokku võttes: selgelt tuleb välja tuua see, millised volitused me algoritmile anname.

Mitmed määruses toodud põhimõtted – õigus väljendada tehtud otsuste suhtes oma seisukohta, õigus otsuseid vaidlustada ning kohustus andmete töötlemisel ja otsuste tegemisel minimeerida potentsiaalseid diskrimineerimisriske – on kaetud ka erinevate soovituslike rahvusvaheliste raamistikega. Sõltuvalt raamistikust keskendutakse selles kas masinõppele või tehisintellektile ning käesoleva artikli kontekstis saame neid käsitleda ühtsetel alustel, kus masinõpe on tehisintellekti rakenduse üheks komponendiks.

---

16 Ingl *once-only principle*, nt teenuste korraldamise ja teabehalduse alused (RT I, 31.05.2017, 7), § 9 lg 1. Vt ka T. Puusaar. Key Factors Influencing the Implementation of the Once-only Principle: Case study of Estonia. Master's thesis. Tallinn University of Technology 2019. Arvutivõrgus: https://digikogu.taltech.ee/et/item/8e9c2389-d28d-4f01-8aeb-eb596c40f281.

17 Eesti riik sai endale CTO ja CDO. Majandus- ja Kommunikatsiooniministeerium. Arvutivõrgus: https://medium.com/digiriik/eesti-riik-sai-endale-cto-ja-cdo-1ba37af31407.

18 Avaandmete tegevuskava (kavand). Majandus- ja Kommunikatsiooniministeerium. Arvutivõrgus: https://mkm.ee/sites/default/files/avaandmete_tegevuskava_2021-22.pdf.

19 Euroopa Parlamendi ja nõukogu määrus (EL) 2016/679, 27. aprill 2016, füüsiliste isikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise ning direktiivi 95/46/EÜ kehtetuks tunnistamise kohta (isikuandmete kaitse üldmäärus). Arvutivõrgus: https://eur-lex.europa.eu/legal-content/ET/TXT/PDF/?uri=CELEX:32016R0679&&from=ET.

20 The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. European Parliament Think Tank. Arvutivõrgus: https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2020)641530.

2018. aasta märtsis avaldas Maailma Majandusfoorum (WEF) valge raamatu pealkirjaga „How to Prevent Discriminatory Outcomes in Machine Learning“*[21], milles muuhulgas tuuakse ära neli printsiipi, mida peaks tehisintellekti rakenduste puhul järgima, selleks et minimeerida rakenduste kasutamisest potentsiaalselt tulenevat diskrimineerivat mõju.

a. **Aktiivne kaasamine**, mille kohaselt tuleb rakenduste disaini ja arenduse käigus tagada sisendite selline mitmekesisus, mis vastab tehisintellekti süsteemi hilisemast kasutamisest mõjutatud ühiskonnagruppide mitmekesisusele.
b. **Õiglus**, mille kohaselt tuleb rakenduse arhitektuuri, loomeprotsessi ning mõõdikute seadmise aspektides lahti mõtestada ja rakendada olukorras kõige sobilikumat õigluse kontseptsiooni.
c. **Arusaamisõigus**, mille kohaselt peab avalikustama otsuste tegemisel rakendatud masinõppefunktsioonid ning nende funktsioonide tulemuste loogiline jada peab olema otsusest mõjutatud inimestele arusaadav ning spetsialistide poolt hinnatav.
d. **Edasikaebamise võimalus**, mille kohaselt on rakenduste loojatel vastutus tuvastada nende rakenduste potentsiaalsed mõjud inimõigustele ning tagada mõjutatutele ligipääs kahju toovate ebaõigete otsuste edasikaebamise protsessile.

2019. aasta aprillis avaldas Euroopa Komisjon usaldusväärse tehisintellekti arendamise eetikasuunised*[22], milles käsitletakse usaldusväärsuse kolme aspekti: seaduslikkust, eetilisust ning töökindlust.

a. **Seaduslik tehisintellekt**: süsteemid peavad järgima kehtivaid rahvusvahelisi ja kohalikke seadusi ning oluline on silmas pidada nii seadustes kehtestatud keelde kui ka kohustusi ja võimalusi.
b. **Eetiline tehisintellekt**: tagatud peab olema kooskõla eetikanormidega.
c. **Töökindel tehisintellekt**: süsteemid peavad toimima ohutult, turvaliselt ning usaldusväärselt; tahtmatute kahjulike mõjude vältimiseks tuleks ette näha kaitsemeetmed. Silmas tuleb pidada nii tehnilist külge kui ka sotsiaalseid seisukohti.

Eetika alampunktis käsitletakse nelja põhiprintsiipi: inimeste sõltumatuse austamine, kahju tegemisest hoidumine, õiglus ning selgitatavus. Need on paljuski sarnased WEF-i dokumendis kirjeldatud printsiipidega.

Dokument käsitleb lisaks sisulistele suunistele ka usaldusväärse tehisintellekti hindamise protseduure ning pakutakse välja kontrollnimekiri – küsimused, mida hindamise käigus käsitleda.

Mõnevõrra laiapõhjalisemat ning mitut teaduslikku, ühiskondlikku ja tehnilist aspekti lühisõnaliselt katva printsiipide loetelu*[23] on koostanud Future of Life Institute, mis koondab hulgaliselt teadlasi, ettevõtjaid ning ühiskonnategelasi tervest maailmast, sh kaasasutaja Jaan Tallinn ning teadusliku nõuandva koja liikmed Nick Bostrom ja Elon Musk. Nendele printsiipidele on oma toetuse praeguseks andnud üle 5300 inimese, neist üle 1600 on tehisintellekti ja robootika valdkonna teadlased.

Hoolimata headest kavatsustest on sellised printsiipide kogumikud omakorda problemaatilised.

5. Tänased proovikivid

Riikide huvi tehisintellekti parima võimaliku, kuid samas ka võimalikult kiire rakendamise vastu on lisaks eeldatud otsestele kasudele seotud ka rahvusvahelise võidujooksuga tehisintellekti vallas. Tehisintellekti aruka rakendamise abil on võimalik parandada nii otsuste tegemise kiirust ja kvaliteeti kui ka optimeerida eri majandusharude ning riigiaparaadi tegevusi.

Enamjaolt vajavad tehisintellekti rakendused treeningandmeid, mille abil algoritmide reeglistikku ja täpsust kalibreerida. Mida rohkem andmeid (omavaheliste seoste ja avastatavate mustritega) on treenimiseks kasutada, seda paremini treenimine õnnestub. Mida kvaliteetsemad on need andmed, seda paremini treenimine õnnestub.

[21] White paper: How to Prevent Discriminatory Outcomes in Machine Learning. World Economic Forum 2018. Arvutivõrgus: https://www.weforum.org/whitepapers/how-to-prevent-discriminatory-outcomes-in-machine-learning.

[22] Ethics guidelines for trustworthy AI. European Commission. Arvutivõrgus: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[23] Asilomar AI principles. Future of Life Institute. Arvutivõrgus: https://futureoflife.org/ai-principles/.

5.1. Rahvusvaheline konkurents

Eri riigid on ennast aja jooksul ise tehisintellekti liidriteks krooninud ning ükskõik kui väga me sooviksime esirinnas näha Eestit või mõnd teist Euroopa Liidu riiki, on reaalsus teine. 2019. aastal avaldatud analüüsi[*24] kohaselt on kõige suurem tehisintellektialaste patenditaotluste päritoluriik USA, millele järgnevad Hiina, Jaapan, Korea ning alles siis esimese Euroopa riigina Saksamaa. Eri hinnangutel on eksponentsiaalse kasvuga Hiina USA-st möödumas või tänaseks päevaks juba möödunud.

Suurtel riikidel on selles valdkonnas kindel eelis: neil on kasutada suuremad andmehulgad. Olulise ühiskondliku mõjuga tehisintellekti rakenduste arendust ning laialdast kasutuselevõttu soodustab – ilma siinkohal väärtushinnanguid andmata! – ka demokraatia ning inimõiguste käsitlus, mis kohati erineb (oluliselt) Euroopa Liidu riikide tavadest ja tõekspidamistest.

Kuna üksikult ei suuda ükski Euroopa riik suurriikidega võistelda, ongi Euroopa Liit võtnud tehisintellekti teema lauale keskse strateegilise küsimusena. Selles mudelis on aga mitu proovikivi, mille halvasti lahendamine või suisa lahendamata jätmine tähendab meie regiooni jaoks olulist mahajäämust.

Esiteks **andmete kättesaadavuse küsimus**. Selleks et luua Euroopa Liidu üleseid tehisintellekti rakendusi, mis suudaksid algoritme treenida mitmest riigist pärit suurte andmekogude peal, on tarvis lahendada andmete piiriülese liikumise probleem. Sellega tegeleb nii Euroopa Komisjon kui laiemalt ka OECD[*25] ning Eesti on asjakohastes aruteludes häälekalt sõna võtnud ja aktiivselt kaasa löönud.

Riikide koostöös tuleb lahendada nii juriidilised kui ka tehnilised raskused, kuid iga riik omaette peab tegelema ka andmete kvaliteedi probleemiga, sest andmeanalüüsis kehtib „rämps sisse, rämps välja" reegel. Metaandmete ja andmevahetusprotokollide kirjeldamiseks vajalik koostöö peab käima paralleelselt läbirääkimiste ning andmekorrastustöödega.

Teiseks **regulatsiooni tasakaalu küsimus**. Meie eesmärgiks on maksimeerida tehisintellekti rakendustest saadavat kasu ning minimeerida potentsiaalset kahju. Liiga suur ettevaatlikkus võib viia innovatsiooni piiramiseni – reeglid muutuvad nii kitsaks, et nende raames pole võimalik efektiivselt toimetada. Liiga vähene ettevaatlikkus võib aga äärmuslikel juhtudel viia ohuni indiviidide elule.

Veebruaris 2020 Euroopa Komisjoni avaldatud valge raamatu „On Artificial Intelligence – A European approach to excellence and trust"[*26] avaliku konsultatsiooni[*27] käigus koguti tagasisidet rohkem kui 1200 panustajalt, kelle hulka kuulusid nii üksikisikud, ettevõtted, avaliku sektori organisatsioonid kui ka teadusasutused. Lisaregulatsiooni pooldas üle 70% vastajatest, kuid selle kehtivuse mõjuvälja suhtes – kas eristada tehisintellekti ohutasemeid ning neile kehtivaid reegleid – lahknesid arvamused olulisel määral.

Eesti koos 13 teise riigiga allkirjastas ka ebapaberi[*28], milles kutsuti üles tehisintellekti regulatsioonile lähenema innovatsiooni toetavalt, tõendipõhiselt ning tervemõistuslikult, arvestades erinevaid riskiprofiile ja kasutades vabatahtlikke, mitte kohustuslikke kategoriseerimissüsteeme. Kõrge riskitasemega tehisintellekti rakendusi nähti seal pigem erandi kui reeglina.

Kolmandaks **väärtuste küsimus**. Artiklis eespool kirjeldatud printsiibid lähtuvad Euroopa väärtus- ja õigusruumist. Meie rahvusvahelised konkurendid lähtuvad aga teistsugustest väärtustest. Kuna riigid üldjuhul ei eksisteeri enda hinnangul väärate väärtuste ruumis ning ajalooliselt on õigeks peetavad väärtused olnud oluliseks ekspordiartikliks, põrkume juba praegu väärtuskonfliktide vastu.

Meil tuleb järjepidevalt analüüsida ning ümber mõtestada seda, millised praktilised käitumisviisid on meie väärtuste parimaks kandjaks kõnealuses kontekstis ning kas me oleme nõus oma väärtustes tegema järeleandmisi (ja milliseid). Mõnel puhul võivad potentsiaalsed järeleandmised olla vabatahtlikud või ennetavad, mõnel juhul tuleb meil aga valida kahest halvast variandist vähem halb.

---

24 Artificial Intelligence – A Worldwide Overview of AI Patents. Intellectual Property Office 2019. Arvutivõrgus: https://www.gov.uk/government/publications/artificial-intelligence-a-worldwide-overview-of-ai-patents.

25 V. Spiezia, J. Tscheke. International Agreements on Cross-border Data Flows and International Trade. OECD 2020. Arvutivõrgus: http://www.oecd.org/digital/international-agreements-on-cross-border-data-flows-and-international-trade-b9be6cbf-en.htm.

26 White Paper on Artificial Intelligence: a European approach to excellence and trust (viide 11).

27 White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust. European Commission. Arvutivõrgus: https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence.

28 Innovative and trustworthy AI: two sides of the same coin. Arvutivõrgus: https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf.

5.2. Tehisintellekti erapoolikus

Üheks sagedamini välja toodud väärtustega seotud hoiatuseks ning etteheiteks tehisintellekti rakendustele on nendes kasutatavate algoritmide potentsiaalne erapoolikus (ingl *bias*), mille tulemuseks on ebaõiglased tulemused ja otsused. See teema on mitmekihiline.

Esiteks erapoolikus kui nähtus. Mugavusest omistame me seda aeg-ajalt ka inimeste poolt loodud olemitele (nt „erapoolik kirjutis"), kuid elututel, teadvuseta olemitel puudub toimijalisus. Need võivad kirjeldada, kuid mitte omada väärtushinnanguid ning eelistusi, olles selle juures pelgalt teadvusega indiviidide sellesuunalised väljendusvahendid.

Erapoolikus eeldab eelistust või mitte-eelistust, mis omakorda nõuab teadvust ja toimijalisuse olemasolu. Algoritmidel need puuduvad, mistõttu algoritmid ei saa olla erapoolikud; küll aga saavad erapoolikud olla (teadvuse ja toimijalisusega) inimesed ning nende tegevused, sealhulgas algoritmide parameetrite seadmine võib põhineda – või põhineb vältimatult – erapoolikutel uskumustel. Siin on probleemi lahendamise seisukohalt oluline vahe, sest erapoolikuse eemaldamine või korrigeerimine muutub indiviidi ja vahendi eristamisel abstraktsest konkreetseks ning aitab liikuda sõnadest tegudeni.

Kirjeldan järgnevalt masinõppe ja tehisintellekti toimimise lihtsustatud tehnilisi põhimõtteid, mis on erapoolikuse kontekstis protseduuride ja tegevusahelate mõistmiseks oluline aspekt.

Masinõppele toetuva tehisintellekti rakenduse toimimismehhanismi saab kirjeldada kahe etapi kaudu. Esimese etapi eesmärgiks on algoritmi treenimine vastavalt seatud eesmärkidele, millele eelneb andmeallika valik, andmete kogumine ning kogutud andmetest lõpliku valiku tegemine.

Algoritmi kohandatakse ning kohendatakse kas automaatselt või inimestepoolse sekkumisega andmete ja tulemuste analüüsi kaudu seni, kuni algoritm ja selle parameetrid näivad eesmärgi saavutamiseks sobilikud. Algoritmide treenimiseks on kolm peamist võimalust:

a) **juhendatud** (ingl *supervised*) meetod on sobilik eelkõige sildistatud andmete analüüsiks (nt töötades läbi foto-taimenimetuse andmepaarid selleks, et hiljem tuvastada tõenäolisim taim kasutaja poolt edastatud fotol). Juhendatud õppemeetodi puhul eelneb kategoriseerimismudel andmetele;

b) **iseseisvat** (ingl *unsupervised*) meetodit kasutatakse enamasti andmetes mustrite leidmiseks ning siin eelnevad andmed kategoriseerimismudelile. Algoritmile ei anta ette tähendusi ning vastavusi, mistõttu suudab see andmete analüüsi käigus tuvastada ka selliseid seaduspärasusi, mida inimesed pole osanud seal (ette) näha. Muuhulgas on selline meetod kasutusel meditsiiniliste piltide baasil diagnoosimiseks;

c) **tugevdav** (ingl *reinforcement*) meetod püüab leida optimaalseid lahendusi läbi preemiate ja karistuste struktuuri vastavalt sellele, millised tulemused klassifitseeritakse edukate ning millised mitteedukatena. Sellist meetodit kasutatakse harilikult näiteks automatiseeritud soovituste andmiseks toodete või teenuste valikul.

Teise etapi eesmärgiks on uute andmete puhul algoritmi edukas rakendamine nende analüüsimiseks ning soovitud vormis tulemuse väljastamiseks. Näiteks peaks lihtne kõnesünteesi rakendus suutma tuvastada etteantud helifailist treeningust tuttavaid üksikuid sõnu, keerukama rakenduse puhul aga analüüsima lauseehitust ning lausutu konteksti. Lihtne fotoanalüüsi rakendus peaks etteantud pildifaililt tuvastama treeningust tuttavaid objekte, keerulisem rakendus aga suutma vaatajat ära petta süvavõltsingu (ingl *deep fake*) abil loodud kujundite abil.

Algoritmi treenimise etapis tehtud valikutest sõltub olulisel määral algoritmi efektiivsus ja eesmärgipärasus, aga ka tulemuste tajutav korrektsus. Nii näiteks on tehisintellekti rakendusi süüdistatud tumedanahalistevastases erapoolikuses nii näotuvastuses[*29] kui ka retsidiivsuse analüüsis[*30], naistevastases erapoolikuses palkamisprotseduurides[*31], iseloomustustes[*32] ja emotikonide eelistamisel[*33] ning

[29] N. Singer. Amazon Is Pushing Facial Technology That a Study Says Could Be Biased. – The New York Times, 24.01.2019. Arvutivõrgus: https://www.nytimes.com/2019/01/24/technology/amazon-facial-technology-study.html.

[30] J. Angwin, J. Larson, S. Mattu, L. Kirchner. Machine Bias. – Pro Publica, 23.05.2016. Arvutivõrgus: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[31] J. Dastin. Amazon scraps secret AI recruiting tool that showed bias against women. – Reuters, 11.10.2018. Arvutivõrgus: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

[32] N. Sonnad. Google Translate's gender bias pairs "he" with "hardworking" and "she" with lazy, and other examples. – Quartz, 29.11.2017. Arvutivõrgus: https://qz.com/1141122/google-translates-gender-bias-pairs-he-with-hardworking-and-she-with-lazy-and-other-examples/.

[33] D. Beres. Type 'ceo' into your iPhone keyboard for a sexist surprise. – Mashable, 04.03.2017. Arvutivõrgus: https://mashable.com/2017/03/03/apple-emoji-predictions-sexist/.

Microsofti kõnerobotit kiirelt omandatud rassismis.*[34]

Enamikul sellistel juhtudel tuleb vaadata, milliste andmete baasil on algoritmid treenitud. Näiteks kui varasemate kohtulahendite seas on rohkem selliseid, kus kurjategijaks on mustanahaline inimene, võib iseseisvalt õppiv algoritm luua seose nahavärvi ning kuritegelikkuse tõenäosuse vahel ka siis, kui selles seoses puudub põhjuslikkus.

Kui analüüsitud juhtide elulookirjeldustest oli suurem osa meesjuhtidelt, võib algoritm luua inimese soo ning juhiomaduste vahel otsese (kuna naiste elulookirjeldusi oli andmete hulgas vähem, siis järelikult sobivad nad juhiks väiksema tõenäosusega) või kaudse (kuna juhtide elulookirjeldustes kasutati tihedamini meesspetsiifilisi omadussõnu, on edukamad juhid need, kelle elulookirjelduses selliseid sõnu kasutatakse) näiliselt põhjusliku seose.

Kui kõnerobotit on treenitud varasemate inimestevaheliste vestluste baasil, kus tihti on väljendatud rassistlikke seisukohti, võib algoritm käsitada neid tavapäraste või oodatutena ning formuleerida ka oma vastused samasuguseid konstruktsioone kasutades.

Mida sellises olukorras teha? Sisuliselt on meie ees kaks valikut.

a. Me võime **muuta treenimiseks kasutatavaid sisendandmeid**. Selleks peame me esmalt tuvastama andmete õiglase tasakaalu ning seejärel kas koguma tasakaalu parandavaid lisaandmeid, eemaldama tasakaalu häirivaid andmeid või redigeerima olemasolevaid andmeid selliselt, et tagatud oleks eelnevalt defineeritud õiglane tasakaal.

b. Samuti võime me **muuta algoritmi parameetreid**. Kui me näeme, et mõned tulemuste prognoosimiseks kasutatavad parameetrid kallutavad tulemusi eelnevalt defineeritud õiglasest tasakaalust eemale, siis saame nende parameetrite osatähtsust vähendada või lisada uusi parameetreid sobiliku osatähtsuse tasemel.

Esimese kahe valikuga kaasnevad aga olulised raskused.

Esiteks: kuidas me teame, et praegused tulemused on väärad? Millist etaloni kasutame me tulemuste õigsuse ja õigluse hindamiseks? Kui me oleme veendunud, et tulemused on väärad, siis kuidas me tuvastame selle, kas tasakaalu kadumise taga on treeningandmed või algoritm ja selle parameetrid? Kas tasakaalu puudumise tulemus on kohe nähtav analüüsi põhjal tehtavates otsustes või laieneb see ka otsuste kaudsele mõjule? Kui jah, siis kui pikka ajavahemikku me vaatlema peame?

Teiseks: kuidas me teame, mida täpselt muuta? Kui me kohendame andmeid või algoritme, siis kuidas saame me olla kindlad, et me vaid eemaldame tajutava erapoolikuse, mitte ei asenda seda uutmoodi erapoolikusega või hoopiski ei suurenda erapoolikust? Kui me valime algoritmi parameetrite muutmise tee, siis millise etaloni järgi hindame me erinevate muudetavate parameetrite õiglast kaalu?

Kolmandaks: kuidas me teame, et kohendamise järel muutunud tulemused on korrektsed? Kas me saame välistada, et meie enda arusaam sellest, mis on korrektne ning mittekorrektne, on erapoolik? Kas võib juhtuda, et me asendame ette kavatsemata ebaõigluse ettekavatsetud ebaõiglusega?

See, mis on õige ja õiglane, pole kõigi osaliste jaoks sellisena üheselt mõistetav ning tajutav. Ühiskonnas on meil sarnaste keerukate olukordade lahendamiseks kujunenud välja reeglid ning loodud institutsioonid, kuid tehisintellekti rakenduste jaoks meil praegu seesugune võimekus puudub.

Kirjeldatud probleemide tõttu on paljudes olukordades suure tõenäosusega ainuvõimalikuks lahenduseks hoopiski kolmas:

c. **muuta esitatavat küsimust**.

Me peame leppima, et tihti tuleb meil andmete puudulikkuse ja algoritmi parameetrite mõju tundmatuse tõttu esitatavat küsimust kitsendada: kui meil on kasutada vaid üht tüüpi või tendentslikud andmed, siis saame küsimusi küsida vaid selles kontekstis. Näiteks ei saa me kogu rahvastikku puudutavatele küsimustele otsida vastuseid vaid rahvastikust üht osa katvate andmete pealt.

Võimaluste piiratust aktsepteerides peame me järgnevalt koguma juurde uusi, laiemaid andmeid, et saaksime ka küsimusi järk-järgult laiemaks muuta. See on ilmselt pikem protsess, mitte ühekordne tegevus. Me peame ka algoritme nende uute andmetega üle treenima, kohati ka algusest peale uuesti treenima, et jõuda paremini toimivate tulemusteni.

[34] J. Vincent. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. – The Verge, 24.03.2016 Arvutivõrgus: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

Olukordades, kus küsimuse kitsendamine ei tule kõne alla, peame me tehisintellekti rakenduse tulemusi kohustuslikult ning teadlikult käsitama oletuslike ja tõenäosuslike, mitte põhjuslike seostena. See, et me tahame konkreetsele küsimusele ühest vastust leida, ei tähenda seda, et seda vastust ka saab leida.

5.3. Tehisintellekti heuristikud

Protsessid, mille abil inimesed mõtlevad ja otsuseid teevad, on ebatäpsed ning kallutatud. Daniel Kahneman kirjeldab oma raamatus „Thinking, Fast and Slow"[*35] kaht mõtlemissüsteemi.[*36] Süsteem 1 ehk kiiresti mõtlemine toimub väikese pingutusega või automaatselt, on väikese energiakasutusega ning toetub varem tuvastatud mustritele (ja eelarvamustele). Süsteem 2 ehk aeglaselt mõtlemine tegeleb uudsete olukordadega, nõuab pingutust ning on suurema energiakasutusega.

Kui inimesel tekib vajadus teha kiire otsus, siis rakendub süsteem 1. Kui aga hiljem palutakse tal selgitada oma otsustamisprotsessi ning otsuse tagamaid, saame me pigem süsteemi 2 toimimist kirjeldava narratiivi, mis on suure tõenäosusega küll retrospektiivselt sidus, kuid ei pruugi vastata (või pigem enamasti ei vasta) nendele mõtteprotsessidele, mis tegelikkuses toimusid – muuhulgas selle tõttu, et inimene ise pole nendest protsessidest teadlik.[*37]

Igapäevatoimingute kontekstis oleme me õppinud inimeste mõtlemise ebatäpsustega elama. On palju halle alasid, mida me aktsepteerime, sest need on sulandunud inimestevahelistesse suhetesse ning ühiskonna toimimise mehhanismidesse. Me teame, et inimestel on eelarvamused, et meie mälu on ekslik ja mõjutatav, et meie selgitusvõime on oluliselt paremini välja arenenud kui õiglaste otsuste tegemise võime ning et kuigi me tihti arvame, et teeme otsuseid ratsionaalselt, tuginevad need siiski (tagantjärele ratsionaliseeritud) emotsioonidele.

Ootused tehisintellektile on aga oluliselt suuremad. Sellelt eeldatakse läbipaistvat ja auditeeritavat otsuseni jõudmise protsessi ning ka Euroopa Komisjoni ja WEF-i kirjapandud printsiipide järgi peaksime me vältima selliste tehisintellekti rakenduste kasutamist, kus need nõuded pole täidetud. Lihtsamate algoritmide puhul pole see probleemiks, kuid sadu tuhandeid või miljoneid otsustuspuu oksi läbi käinud ning parameetrite kaalust igal sammul mõjutatud protsessi mõistmine käib inimestele üle jõu.

Kui inimeste ebatäiuslikkuse mõju vähendamiseks oleme me aja jooksul ehitanud erinevaid kontroll- ja kompensatsioonimehhanisme, siis tehisintellekti laialdane rakendamine on veel nii uus teema, et vajalikud mehhanismid on välja ehitamata. Eespool kirjeldatud erapoolikuse probleemi (praegune) lahendamatus on üks selliseid näiteid.

Tehisintellektile seatavate suurte ootuste taga on vähemalt osaliselt taas kord juba varem kirjeldatud hirm tundmatuse ees. Piltlikult öeldes me ei tea, kuidas masin toimib ja „mõtleb". Me ei tea tihti ka seda, kuidas teine inimene mõtleb, aga kuna ta on meiega sarnane, siis saame talle projitseerida enda uskumusi ning usutavana kõlavaid mõtlemisprotsesse. Algoritmid aga ei ole meie moodi. Seetõttu tahame me riskide maandamiseks muuhulgas seada reegleid, mis annavad meile teatava kontrolli (illusiooni) teistsugususe üle. Kas me aga oleme valmis hallide alade asemel eksplitsiitsete, tihti binaarsete reeglitega tegelema?

Võtame näitena isesõitvad autod. Oletame, et tekib olukord, kus avarii on vältimatu ning sellises autos rakendatud algoritmil tuleb teha otsus selle kohta, kelle vigastamine on õigem: kas auto sõidab vastu telliskiviseina ja vigastab sel juhul autojuhti või sõidab see otsa jalakäijale.

Milliste andmete põhjal võiks algoritm selle otsuse teha? Üheks variandiks on universaalsed, kõigile autodele kehtivad reeglid: näiteks et sellistes olukordades tuleb alati eelistada jalakäija elu säästmist ka siis, kui see toimub autojuhi elu hinnaga. Selline reegel tekitaks autoostjates ilmselt kõhklusi, sest mitte igas olukorras ei tunduks see lahendus õiglane. Kui aga reegel oleks vastupidine? Sellisel juhul võiks igat isesõitvat autot käsitada tuima tapamasinana ning jalakäijate valmidus sõiduteede läheduses liikuda võib oluliselt väheneda.

Kui mustvalge universaalne reegel ei näi mõistliku suunana, siis millised on alternatiivid? Kas võiks eelistus olla autojuhi poolt seadistatav? Vahemikus „Mina tahan, et minu auto valiks alati minu elu

35 D. Kahneman. Thinking, Fast and Slow. Farrar, Straus and Giroux 2011.

36 Keerukusteaduse kontekstis kutsutakse süsteemi 1 kohati ka autonoomseks (ingl *autonomic*) ning süsteemi 2 uudsusvastuvõtlikuks (ingl *novelty-receptive*).

37 Kõrvalepõikena: valede kergemaks tuvastamiseks on soovitatud paluda inimesel varem toimunut kirjeldada vastupidises ajalises järjestuses, alustades kõige viimasest sündmusest. Koherentse valenarratiivi tagurpidi ehitamine on keerulisem kui voolavas ajalises järjestuses sündmuste narratiivi sidumine.

päästmise, isegi kui see tähendab teise inimese surma“ kuni „Mina tahan, et minu auto püüaks alati päästa teise inimese elu isegi siis, kui see tähendab minu surma“? Kas sellisel juhul tekib autojuhile ühiskondlik või seaduslik vastutuskoorem? Kas autodele peaks lisama välised eelistuste indikaatorid?

Loomulikult on võimalik otsuseid teha ka juhuslikult, kuid kas see tundub õiglane? Me ju ei arva, et avariisse sattunud autojuht teeb otsuseid juhuslikult – pigem püüab ta kasvõi alateadlikult leida parima võimaliku lahenduse, arvestades konkreetset olukorda. Kas algoritm ei saaks teha sama?

Siinkohal tuleb meil aga vastata küsimusele, et mis on „parim võimalik lahendus“ ning milliseid kriteeriumeid peaks algoritm arvestama. Kas ehk oleks mõistlik tuvastada ohtu sattunud jalakäija isik, teha kiire, aga põhjalik indiviid *vs.* indiviid väärtusanalüüs ning ohverdada vähem väärtuslik osapool?

Normatiivne eetika annab meile siinkohal kätte mõned mõttesuunad. Konsekventsialism hindab tegude moraalsuse puhul nende tagajärgi – näiteks utilitarismi (suurim hüve suurimale võimalikule hulgale inimestele, minimeerides negatiivseid mõjusid) või egoismi (suurim hüve minule) paradigmas. Deontoloogiline eetika aga hindab tegude moraalsust sõltumata nende tulemustest, kusjuures olukorrale võib läheneda nii kohusepõhiselt kui ka õigustepõhiselt. Otsustusmudelid on võimalik välja töötada, kuid on tõenäoline, et kõik nendega ei nõustu.

Tavaliselt ei puutu enamik inimesi teadlikult kokku ei moraalifilosoofia ega olukordadega, kus tuleb otsustada enda või teiste elu üle. Seetõttu kõlab ehk ka klassikaks saanud trammiprobleem ebarelevantse mõtteeksperimendina – kuni jõuame olukorda, kus me peame hakkama ühiskonna sõnastamata väärtusi reeglitena kirja panema ning nende reeglite järgmist ka jälgima, hindama ning vajadusel kiitma või karistama.

5.4. Kas tehisintellekti võimu all?

Reeglite optimaalse hulga ning kehtivusulatuse suhtes aitab meil otsusteni jõuda konkreetse eesmärgi seadmine: mida me tehisintellekti rakendustelt täpsemalt siiski ootame? Kui me seame uhkeid, kõlavaid eesmärke ning samas lubame teadlastel ning praktikutel toimetada vaid väga kitsastes piirides, siis on pettumus vältimatu. Samuti viib piirangute ebamõistlik rakendamine selleni, et regioonina jääme me konkurentsis kaotajateks ning nii teadusrahastus kui ka talendid liiguvad mujale.

Kui palju peaksime me aga laskma tehisintellektiga seotud ambitsioonidel ning selle rakendamisest tulenevatel nõudmistel kujundada meie laiemat strateegiat – eeldusel, et meil see ühel hetkel tekib – ning igapäevaelu? Kas tehnoloogiaga peaks tegema kompromisse?

Avaliku sektori kontekstis on ühiskonna jaoks oluline mõista seda, millistel juhtudel kavatseme me kasutada tehisintellekti rakendusi selleks, et aidata inimestel paremini (efektiivsemalt, kiiremini, tervislikumalt, odavamalt) oma ülesandeid täita ja otsuseid teha, ning millal kavatseme me ohjad käest anda. Ehk siis: millistes olukordades asendaks juhirollis inimest tehisintellekt?

Rõhutan siinkohal üle, et tehisintellekti puhul pole tegemist peamiselt infotehnoloogia probleemiga. Sellega seotud väljakutseid ei saa delegeerida IT osakonnale ega ole mõistlik eeldada, et just tehnoloogiaspetsialistid lahendused lauale tooks.

Me peame oma plaanides ja ambitsioonides olema ausad selle suhtes, kas me soovime tehisintellekti rakendustelt ametnike toetamist või nende asendamist. Kui mõnes konkreetses valdkonnas on realistlik protsesside ulatuslik automatiseerimine, siis lasub meil vastutus tõsta selle valdkonna spetsialistide võimekust sellisele tasemele, et nad kas suudaksid täiendatud süsteemis tekkinud uutes rollides hästi hakkama saada või suudaksid leida sisulist, eneseteostusvõimalusi pakkuvat rakendust muudes valdkondades.

Kui tehisintellekti rakenduste rolliks saab muuhulgas parimana käsitatava otsustusvariandi pakkumine, tuleb meil luua analüüsitulemuste konfliktiolukordade lahendamise protseduur. Kuidas tuleb käituda sellises olukorras, kus tehisintellekti rakendus pakub andmete analüüsimise järel välja ühe lahenduse, valdkonnaspetsialistist ametnik jõuab enda tehtud analüüsi tulemusena aga teistsuguse lahendusvariandini?

Väga lihtsustatult kokkuvõttes saab seda küsimust vaadata kui „Kes jääb süüdi?“ dilemmat. Ametniku järeldus põhineb reeglitel, faktidel ning eelneval kogemusel, tehisintellekti rakenduse järeldus aga reeglitel, faktidel ning etteantud piiride raames suurte andmemahtude, sh eelnenud otsuste analüüsil.

Kui tegemist on olukorraga, kus otsuse korrektsust saab faktide ja reeglite põhjal hinnata ühemõtteliselt, on üks lahendusvariant objektiivselt korrektne ning teine lahendusvariant objektiivselt väär. Kui aga otsus nõuab hinnangute andmist, kaob objektiivsus ning binaarne tõene-väär võimalus.

Võtame näitena olukorra, kus meil on kaks otsuse varianti X ja Y ning hiljem tehtud analüüsi tulemusena selgub, et otsuse variant X oli korrektne ning Y väär. Meil on siinkohal kaheksa kombinatsiooni:

Tehisintellekt	Ametnik	Ametniku otsus	Tulemus	Initsiatiiv
X	Y	Y	Väär	Ametnik
X	Y	X	Õige	Tehisintellekt
X	X	X	Õige	Ühine
X	X	Y	Väär	Ametnik
Y	X	X	Õige	Ametnik
Y	X	Y	Väär	Tehisintellekt
Y	Y	Y	Väär	Ühine
Y	Y	X	Õige	Ametnik

Kuidas lahendame me siinkohal vastutuse küsimuse ning kui tõenäoline on see, et ametnik valib variandi, mis läheb tehisintellekti poolt pakutavaga vastuollu? See sõltub muuhulgas otsusele järgneva hindamise tulemustest ning sellest, kas ametniku otsusele võib järgneda kiitus, laitus või karistus. Eksperimentaalne filosoofia*[38] on varasematele küsimustele ja ideedele ehitades teinud viimastel aastatel selles valdkonnas mitmeid uuringuid ning tuvastanud väljateenitud kiituse-laituse tunnetuses ja karistuste nõudmises olulise asümmeetria.

Lisaks: kas vastutust saab kanda vaid teadvusega spetsialist või saab ka tehisintellekti rakendusele vastutust rakendada? Kui viimane on võimalik, siis kes jääb tegelikult vastutavaks olukorras, kus otsuse tegijal puudub toimijalisus? Variante on siin palju: näiteks tehisintellekti rakendust kasutav ametnik, rakendust ja selle parameetreid seadistav spetsialist, rakenduse kasutuselevõtu otsustanud juht, rakenduse loonud ettevõtte juht, rakenduse algoritmide programmeerija, algoritmi treenimiseks kasutatud andmete valija, algoritmi treenimiseks kasutatud andmete koguja, algoritmi treenimiseks kasutatud andmete algne looja …

Kui harjumuspärase töövahendi defektist tuleneva kahju puhul on meil vastutuse rakendamise kohta varasem kogemus olemas, siis tehisintellekti puhul on olukord erinev vähemalt kahest aspektist. Enamikke teisi töövahendeid ei käsita me otsuste tegijana ning defektsust on vähemalt teoreetiliselt võimalik määrata binaarselt, sõltuvalt valitud abstraktsioonitasemest. Samuti räägime me tehisintellekti rakenduste puhul tõsikindla teadmise ning etteprogrammeeritud reeglistiku asemel tõenäosuslikkusest ning paljude parameetrite koosmõjust tulenevatest mõjutajatest. Nagu inimeste puhul, kuid (seni veel) ilma teadvust omamata.

6. Kokkuvõtteks

Usun, et peaksime Eestis püüdma jääda tehisintellekti pragmaatilise rakendamise globaalsele eesliinile ning toetama paindlikku regulatsiooni koos turvaliste „liivakastidega“ erinevate eksperimentide läbiviimiseks. Samuti usun, et peaksime jätkama Euroopa Liidu sisest võitlust korrastatud andmete turvaliseks liikumiseks üle riigipiiride, sest ei suured probleemid ega nutikad lahendused riigipiire ei tunnista.

Krattide arendamisel ja rakendamisel on tegijate ning vajajate vahelises koostöös lepitud kokku mitmes printsiibis, mis aitavad valdkonda loodetavasti kestlikult ning turvaliselt arendada. Meie eelistus on olnud väikeste korduskasutatavate rakenduste, mitte suurte monoliitsete süsteemide ehitamine. Me usume jätkuvalt andmete detsentraliseerituse tugevustesse ning eelistame superandmebaaside asemel arendada turvalisi andmevahetuskeskkondi ja koostöövõimelisi kratte. Me oleme mõistnud keeletehnoloogia olulisust väikese keele (ja kultuuri!) säilimisel ning mitmed enimkasutatud kratid (nt Salme kohtutes ja Hans Riigikogus) on just keeletehnoloogiapõhised.

Loomulikult on meil veel palju ära teha. Mitte kõik asutused pole enda käsutuses olevate andmete korrastamisega veel väga kaugele jõudnud ning tihti pole ka korrastatud andmetest tulenevad potentsiaalsed

[38] J. Knobe, S. Nichols. Experimental Philosophy. – The Stanford Encyclopedia of Philosophy. E. N. Zalta (ed.). Winter 2017 Edition. Arvutivõrgus: https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/.

otsesed ja kaudsed kasud piisavalt selged – muuhulgas seetõttu, et erinevate ebaefektiivsete protseduuridega ollakse aja jooksul lihtsalt ära harjunud ja alternatiivid olid mõeldamatud.

Kindlasti peab jätkuma juhendmaterjalide koostamine, koolituste korraldamine ning teadlikkuse suurendamine. Üheks efektiivseks viisiks kasude demonstreerimisel on seda teha näidete, mitte pelgalt lubaduste abil. Selleks puhuks on abiks kiiresti loodavad piiratud funktsionaalsusega kratiprototüübid ja pilootrakendused ning loomulikult ka kasusaajate tulemuslik kaardistus. Sündmusteenused on heaks näiteks kasutajapõhisest lähenemisest ja praegune fookus suurema hulga (nähtamatute) sündmusteenuste arendamisel toetab nii kodanike heaolu, ametnike ja riigiaparaadi efektiivsust kui ka tugeva tehnoloogilise vundamendi ehitamist.

Nii nagu tehisintellekti temaatika üldiselt pole eelkõige tehnoloogiline, pole ka avaliku sektori tehisintellekti rakenduste temaatika mitte pelgalt ametnike teema, vaid ametkondade, erasektori ning teadusasutuste koostöö küsimus. Siin on meil, ma usun, ees suur tööpõld.

Autorist: MA Kaimar Karu on ettevõtja.