refrAIme

POLIITIKADOKUMENT

INIMKESKNE TEHISINTELLEKTI INNOVATSIOON JA KASUTUS: EUROOPA LIIDU PÕHIÕIGUSTE HARTA ROLL

Tatyana Novossiolova, PhD
Center for the Study of Democracy

Co-funded by the European Union

CSD CENTER FOR THE STUDY OF DEMOCRACY

CENTRE FOR EUROPEAN CONSTITUTIONAL LAW
THEMISTOKLES AND DIMITRIS TSATSOS FOUNDATION

European Center for Not-for-Profit Law

ESTONIAN HUMAN RIGHTS CENTRE

L-Università ta' Malta Faculty of Laws | Department of European & Comparative Law

Tehisintellekt (TI) on murranguline tehnoloogia, mis toob kaasa olulisi muutusi erinevates sektorites ja tööstusharudes [1]. Digitaliseerimise transformatiivne potentsiaal, mis on märgatav mitmetes valdkondades, on sillutanud teed andmeanalüütika, prognoosimise, loomise ja mustrite avastamise süsteemide ja tööriistade kasutuselevõtuks mitmesuguste rakendustega. TI toega süsteemid kujundavad üha enam ümber töökohti, kunsti, teadusuuringuid ja juhtimist.

Avaliku ja erasektori investeeringud tehisintellekti on viimase kümnendi jooksul hüppeliselt kasvanud. TI innovatsioon on tulus nišš, mis võib luua taskukohaseid ja juurdepääsetavaid tehnoloogilisi lahendusi. TI toodete ülemaailmne turg laieneb ja praegused prognoosid näitavad, et järgmise kolme aasta jooksul võib selle väärtus ulatuda umbes triljoni dollarini; ainuüksi generatiivse TI turg võib 2032. aastaks ületada triljoni dollari piiri [2].

Kuid TI innovatsioon on rohkem kui majandusliku ja tehnoloogilise võimekuse element. See õhutab uut rahvusluse vormi, kuna riigid võistlevad suveräänse TI väljatöötamise nimel, mis tugineb andmete kättesaadavusele ja andmetöötluse taristule siseriiklikul tasandil. See võib pidurdada rahvusvahelist koostööd TI arendamisel, koormata oluliste tehniliste komponentide (nt pooljuhid, kiibid) tarneahelaid ja süvendada digilõhesid nii riikide vahel kui ka riikide sees.

Nagu iga teinegi suur tehnoloogia viimaste sajandite jooksul, pakub TI kasu, kuid tekitab ka riske, mille kontrollimata jätmisel võivad olla tõsised tagajärjed. TI innovatsioonist ja juurutamisest saadava kasu maksimeerimiseks on vaja võtta asjakohased sammud ja meetmed võimalike negatiivsete mõjude maandamiseks, eelkõige inimõiguste valdkonnas.

ELi tasandil on Euroopa Liidu põhiõiguste harta (edaspidi "harta") ainulaadne riigiülene õiguslikult siduv raamistik, mis tagab ELi kodanike poliitilised, majanduslikud, sotsiaalsed ja kultuurilised õigused. Harta sätete järgimine on hädavajalik usaldusväärse TI arendamiseks ja inimkeskse TI innovatsiooni tagamiseks.

Käesolev poliitikadokument kaardistab lähenemisviise ja mehhanisme, mille abil maandada TI-süsteemide negatiivset mõju põhiõigustele. Suur osa käesolevast analüüsist on seotud masinõppe edusammudega, sealhulgas süvaõppega, ja dokument algab nende mõistete ülevaatega. Seejärel kirjeldatakse harta põhielemente, sellega hõlmatud õigusi ja tagatisi. Järgmisena uuritakse, kuidas TI toega süsteemide arendamine ja kasutamine võib põhiõigusi ohustada. Dokument lõpeb praktiliste soovitustega, et tagada tasakaalustatud lähenemisviis riskide maandamiseks ilma TI innovatsiooni lämmatamata. Dokument põhineb uuringu “Taustauuring TI kasutamise mõju kohta põhiõigustele” peamistele järeldustele [3].

1. TI, masinõpe ja süvaõpe

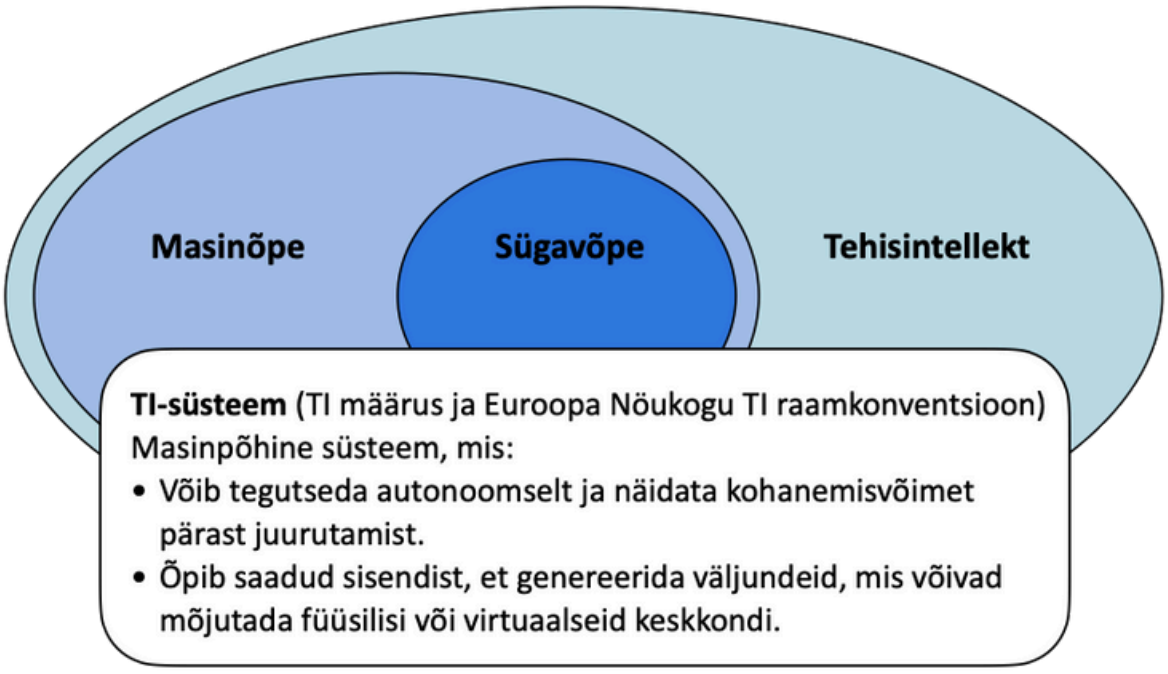
Tehisintellekt on üldmõiste, mis hõlmab mitmeid tehnoloogiaid ning uurimis-, innovatsiooni- ja tootearenduse valdkondi. Lihtsalt öeldes tähendab TI arvutite või arvutipõhiste süsteemide võimet simuleerida kognitiivseid funktsioone ja täita tegevusi, mis on omased või iseloomulikud inimestele. TI toega süsteemide võimekus on lai ning näitlik loetelu hõlmab loomuliku keele töötlust, sisuloomet (nt tekst, heli, video, pildid), probleemilahendust, andmeanalüüsi, aga ka korduvate lihtsate ülesannete optimeerimist (nt protsesside automatiseerimine), sõidukite juhtimist (nt isesõitvad autod) ja hooldustöid (nt abirobotid).

Masinõpet seostatakse sageli TI-ga ning kuigi need mõisted on seotud, ei tähenda need sama. Masinõpe on kõige paremini kirjeldatav kui TI alamvaldkond, mis keskendub arvutitarkvara treenimisele, et see suudaks andmete põhjal iseseisvalt ülesandeid täita. Algoritmid õpivad neile sisestatud andmetest ja kohandavad oma toimimist kogemuste põhjal. Olenevalt inimese ja masina interaktsiooni tasemest treenimise ajal võib masinõpe olla juhendatud või juhendamata. Juhendatud masinõppes kasutatakse märgendatud andmeid, näiteks sildistatud pilte, mille põhjal algoritm õpib pilte eristama ja uusi pilte õigesti vastava sildiga klassifitseerima. Juhendamata masinõppes kasutatakse märgendamata andmeid, näiteks sildistamata pilte, mille põhjal algoritm õpib avastama mustreid ja seoseid erinevate piltide vahel ning grupeerib need teatud tunnuste alusel. Juhendamata masinõpe on keerukam ja tavaliselt vähem täpne võrreldes juhendatud masinõppega. Juhendatud masinõpet kasutatakse ennustuste tegemiseks uute andmete kohta, näiteks rämpsposti filtreerimiseks e-postkastis. Juhendamata masinõpet kasutatakse mustrite tuvastamiseks, näiteks sotsiaalmeedia soovitussüsteemides, mis pakuvad kasutajate käitumise põhjal sisu.

Süvaõpe on masinõppe liik, mis kasutab inimaju matkivaid mitmekihilisi tehisnärvivõrke, et lahendada keerukaid probleeme. Süvaõppe mudelid suudavad töödelda suuri märgendamata andmekogumeid ning tuvastada ja eraldada seoseid ja mustreid. Süvaõppe mudelite näited on generatiivse TI süsteemid, mis loovad päringu alusel sisu (teksti, muusikat, videot, pilte jne), finantspettuste tuvastamise mootorid, vestlusrobotid, loomuliku keele töötlemise süsteemid (nt kõnest tekstiks programmid) ning arvutinägemine, mille puhul tarkvara analüüsib visuaalseid sisendeid ja reageerib kõrvalekalletele või muudele probleemidele, mida kasutatakse näiteks näotuvastustehnoloogias.

Euroopa Liidu TI määrus (TI määrus), mis on esimene terviklik TI regulatsioon, määratleb TI-süsteemi järgmiselt: „masinpõhine süsteem, mis on projekteeritud töötama erineval autonoomsuse tasemel ning mis võib pärast juurutamist olla kohanemisvõimeline ja mis saadud sisendist otseste või kaudsete eesmärkide saavutamiseks järeldab, kuidas genereerida väljundeid, näiteks prognoose, sisu, soovitusi või otsuseid, mis võivad mõjutada füüsilist või virtuaalset keskkonda“. See määratlus hõlmab nii juhendatud kui ka juhendamata masinõppe mudeleid, sealhulgas süvaõppe mudeleid. Sarnase määratluse on andnud ka Euroopa Nõukogu tehisintellekti ning inimõiguste, demokraatia ja õigusriigi raamkonventsioon (joonis 1).

Joonis 1. TI määratlemine



Allikas: CSD.

2. Euroopa Liidu põhiõiguste harta

ELi usaldusväärse (st seadusliku, eetilise ja kindla) TI raamistik keskendub põhiõiguste tagamisele. ELi põhiõiguste harta (edaspidi „harta“) sätestab kodaniku-, poliitilised, sotsiaalsed ja majanduslikud õigused, mis on kõigil ELi kodanikel ja elanikel. Harta on siduv ELi institutsioonidele ja liikmesriikidele juhul, kui nad tegutsevad ELi õiguse reguleerimisalas. See tähendab, et liikmesriigid peavad hartat arvesse võtma ELi direktiivide ülevõtmisel ja ELi õigusaktide ning määruste rakendamisel riiklikul tasandil.

Harta koosneb 50 õigusest, mis on jagatud kuue valdkonna alla: väärikus, vabadused, võrdsus, solidaarsus, kodanike õigused ja õigusemõistmine. Paljud neist õigustest on käsitletud ka erinevates rahvusvahelistes lepingutes, näiteks kodaniku- ja poliitiliste õiguste rahvusvaheline pakt, majanduslike, sotsiaalsete ja kultuurialaste õiguste rahvusvaheline pakt, Euroopa inimõiguste konventsioon ning valdkondlikud erikonventsioonid (nt naiste diskrimineerimise kõigi vormide likvideerimise konventsioon, rassilise diskrimineerimise kõigi vormide kõrvaldamise rahvusvaheline konventsioon, piinamise vastane konventsioon, lapse õiguste konventsioon, puuetega inimeste õiguste konventsioon). Harta eristab õigusi ja põhimõtteid. Põhimõtete näidetena võib tuua artikli 25, mis käsitleb eakate õigusi, artikli 26, mis puudutab puuetega inimeste integreerimist, ja artikli 37, mis käsitleb keskkonnakaitset. Kuigi nii õigused kui põhimõtted on võrdselt siduvad, saavad kodanikud õigustele tugineda otse riiklikes kohtutes, samas kui põhimõtteid saab kasutada vaid koos rakendava seadusandliku akti või haldusaktiga [4].

Harta annab ühtse õiguslikult siduva raamistiku, mis tugineb ELi õiguse tugevusele, millel on ülimuslikkus riigisisese õiguse ees. Harta suurendab põhiõiguste nähtavust ja täpsustab õigusi ja põhimõtteid, mis ei pruugi olla teistes inimõiguste dokumentides sõnaselgelt kajastatud. Selle demonstreerimiseks on kasulik võrrelda harta ja Euroopa inimõiguste konventsiooni (EIÕK) poolt tagatud õigusi ja vabadusi (vt tabel 1).

Tabel 1: Harta ja Euroopa inimõiguste konventsiooni sätete võrdlus

Harta osa	Harta artiklid	Võrdlus Euroopa inimõiguste konventsiooniga (EIÕK)
Väärikus	**Art. 1 Inimväärikus**	**EIÕKis vastet ei ole**
	Art. 2 Elu	EIÕKiga samaväärne kaitse
	Art. 3 Isikupuutumatus	**EIÕKis vastet ei ole**
	Art. 4 Piinamine ning ebainimlik või alandav kohtlemine	EIÕKiga samaväärne kaitse
	Art. 5 Orjapidamine ja sunniviisiline töö	
Vabadused	Art. 6 Vabadus ja turvalisus	
	Art. 7 Era- ja perekonnaelu	
	Art. 8 Isikuandmed	**Ulatuslikum kui EIÕK**
	Art. 9 Abielu ja pere loomine	
	Art. 10 Mõtte-, südametunnistuse- ja usuvabadus	EIÕKiga samaväärne kaitse
	Art. 11 Sõna- ja teabevabadus	
	Art. 12 Kogunemine ja ühinemine	**Ulatuslikum kui EIÕK**
	Art. 13 Kunst ja teadus	EIÕKiga samaväärne kaitse
	Art. 14 Haridus	**Ulatuslikum kui EIÕK**
	Art. 15 Kutsevabadus ja töö tegemine	**EIÕKis vastet ei ole**
	Art. 16 Ettevõtlus	
	Art. 17 Omand	EIÕKiga samaväärne kaitse
	Art. 18 Varjupaik	**EIÕKis vastet ei ole**
	Art. 19 Tagasisaatmine, väljasaatmine, väljaandmine	EIÕKiga samaväärne kaitse
Võrdsus	Art. 20 Võrdsus	
	Art. 21 Diskrimineerimise keeld	
	Art. 22 Kultuuriline, usuline ja keeleline mitmekesisus	**EIÕKis vastet ei ole**
	Art. 23 Naiste ja meeste võrdõiguslikkus	EIÕKiga samaväärne kaitse
	Art. 24 Lapse õigused	**EIÕKis vastet ei ole**
	Art. 25 Eakate õigused	
	Art.26 Puuetega inimeste integreerimine	

Solidaarsus	**Art. 27 Töötajate õigus olla informeeritud ja ära kuulatud** **Art. 28 Kollektiivläbirääkimised ja kollektiivne tegutsemine** **Art. 29 Tööhõiveteenuste kasutamine** **Art. 30 Põhjendamatu vallandamine** **Art. 31 Head ja õiglased töötingimused** **Art. 32 Keeld kasutada laste tööjõudu; noorte töötajate kaitse** **Art. 33 Perekonna- ja tööelu** **Art. 34 Sotsiaalkindlustus ja sotsiaalabi** **Art. 35 Tervishoid** **Art. 36 Võimalus kasutada üldist majandushuvi pakkuvaid teenuseid** **Art. 37 Keskkonnakaitse** **Art. 38 Tarbijakaitse**	**EIÕKis vastet ei ole**
Kodanike õigused	Art. 39 Õigus hääletada ja kandideerida EP valimistel Art. 40 Õigus hääletada ja kandideerida kohalikel valimistel Art. 41 Hea haldus Art. 42 Õigus tutvuda dokumentidega Art. 43 Euroopa ombudsman Art. 44 Petitsioon (EP) Art. 45 Liikumis- ja elukohavabadus Art. 46 Diplomaatiline ja konsulaarkaitse	ELi kontekstipõhised õigused (pole EIÕKis kaetud)
Õiguse-mõistmine	**Art. 47 Tõhus õiguskaitsevahend ja õiglane kohtulik arutamine**	**Ulatuslikum kui EIÕK**
	Art. 48 Süütuse presumptsioon ja kaitseõigus Art. 49 Kuritegude ja karistuste seaduslikkuse ja proportsionaalsuse põhimõte	EIÕKiga samaväärne kaitse
	Art. 50 Mitmekordse kohtumõistmise ja karistamise keeld	**Ulatuslikum kui EIÕK**

Allikas: Euroopa Liidu Põhiõiguste Amet, *Applying the Charter of Fundamental Rights of the European Union in Law and Policymaking at National Level – Guidance*, 16.10.2018.

3. TI toega süsteemid ja põhiõigused

Võimalike riskide tõhusa haldamise ja maandamise jaoks on oluline mõista, kuidas TI toega süsteemide disain, juurutamine ja kasutamine (ja väärkasutamine) võib mõjutada põhiõigusi. Selles jaotises vaadeldakse mõningaid levinumaid viise, kuidas TI võib põhiõigusi riivata. [5]

TI kallutatus

TI-süsteemid võivad anda tulemusi, mis peegeldavad ja võimendavad sotsiaalseid, kultuurilisi või ajaloolisi eelarvamusi, mis on seotud vanuse, etnilise päritolu, soo või rahvusega. Näideteks võib tuua TI toega värbamistööriistad, mis eelistavad mehi naissoost taotlejatele, pidades neid teatud tööstusharude jaoks sobivamaks ja paremini kvalifitseerituks; ennustavat politseitööd toetavad TI-süsteemid, mis suunavad ebaproportsionaalselt palju tähelepanu majanduslikult ebasoodsas olukorras olevatele või etniliselt mitmekesistele piirkondadele, pidades neid kõrgema kuritegevuse riskiga aladeks; ning sotsiaaltoetuste väärkasutuse tuvastamise tööriistad, mis peavad teatud majandusliku tausta või rahvuse esindajaid suurema tõenäosusega petturiteks.

TI kallutatuse põhjused võivad olla erinevad [6]. Üks põhjus on algoritmi koolitamiseks kasutatud andmekogumite kvaliteet. Andmestikud, milles teatud rühmad on alaesindatud või üleesindatud, mõjutavad masinõppe protsessi ning kordavad andmetes sisalduvaid eelarvamusi. Kui koolitamiseks kasutatud andmed keskenduvad peamiselt migrantide või vähemusrahvuste poolt toime pandud kuritegudele, õpib algoritm seostama kuritegelikku käitumist nende gruppidega ja võib seetõttu alahinnata riski, et valge inimene kuriteo toime paneb.

Kallutatus võib tuleneda ka programmeerimisvigadest või -puudustest, mida nimetatakse algoritmiliseks kallutatuseks. Algoritmiline kallutatus võib olla seotud vigase andmestikuga, kuid ka algoritmi loomise viisiga. Algoritmid teevad otsuseid eelnevalt määratud näitajate alusel ning nende näitajate (nt sõnavara, haridustase, kvalifikatsioonid) valik ja neile antav kaal võivad viia kallutatud tulemusteni. Näiteks värbamisalgoritm, mis lükkab automaatselt tagasi kandidaadid, kellel puudub kindel ülikoolikraad, võib diskrimineerida neid, kellel on samaväärne töökogemus. Algoritmiline kallutatus võib areneda ka aja jooksul tagasisideahelate kaudu. Tagasisideahel tekib siis, kui süsteemi tehtud prognoosid mõjutavad andmeid, mida kasutatakse sama süsteemi uuendamiseks [7]. Selline tagasisideahel võib tekkida näiteks ennustava politseitöö puhul, mis kasutab kuritegevuse määra andmeid. Algoritm suunab rohkem patrulle piirkondadesse, kus kuritegevuse määr on kõrgem. Selle tulemusena registreeritakse seal rohkem kuritegusid ning need andmed sisestatakse uuesti algoritmi, mis omakorda suurendab politsei kohalolu samades piirkondades ja võib kaasa tuua liigse järelevalve mõnes piirkonnas teiste piirkondade arvelt.

Kolmas TI kallutatuse allikas on kognitiivne kallutatus. TI-süsteemid on loodud inimeste poolt ja inimesed võivad tahtmatult sisestada algoritmidesse oma vaated, eelistused või eelarvamused. Kognitiivne kallutatus võib mõjutada koolitusandmestike valikut, algoritmide kirjutamise protsessi, arendatavate rakenduste valikut ning TI-süsteemide lõppkasutust. Õpetlik näide kognitiivsest kallutatusest on rühmavälise homogeensuse eelarvamus – kalduvus pidada enda gruppi mitmekesiseks, samal ajal kui teisi gruppe nähakse ühtse ja sarnase massina [8]. Selline kallutatus vähendab algoritmi võimet eristada vähemusgruppe ja toob kaasa ebatäpsusi ning valeandmeid. Kognitiivse kallutatusega on seotud ka süsteemne või institutsionaalne kallutatus [9]. Kui tehnoloogiaettevõtte organisatsioonikultuuris on juurdunud teatud stereotüübid, mõjutab see tõenäoliselt arendatavate TI-süsteemide valikut kui ka nende disaini.

TI kallutatusel on olulised tagajärjed põhiõigustele. See võib mõjutada diskrimineerimise keeldu, õigust võrdsusele (nt sooline võrdõiguslikkus), süütuse presumptsiooni ja kaitseõigust ning takistada juurdepääsu olulistele avalikele teenustele, sealhulgas sotsiaalabile ja tervishoiule.

TI hallutsinatsioonid

TI-süsteemide hallutsinatsioonid tähendavad ekslike või vigaste tulemuste loomist. Hallutsinatsioonid on generatiivse TI puhul loomupärased ning kuuluvad nende ülesehituse juurde, mistõttu ei ole võimalik neid täielikult kõrvaldada [10]. TI hallutsinatsioonid ilmnevad eriti selgelt suurte keelemudelite (LLM) puhul, kui need teevad faktilisi vigu, loovad eksitavat teavet või pakuvad ebatõenäolisi seletusi selle kohta, kuidas nad vastava vastuseni jõudsid. Keelemudelid võivad segamini ajada akadeemilisi viiteid ning valesti tõlgendada arve ja fakte.

Näiteks võib tervishoiuvaldkonnas kasutatav mudel anda vale diagnoosi ja soovitada tarbetuid meditsiinilisi sekkumisi või hallutsineeriv uudiste robot võib vastata küsimustele kujuneva olukorra kohta kontrollimata teabega ning levitada valeinfot [11]. Veelgi problemaatilisem on olukord, kus LLM-id toodavad teavet, mis näib õige, kuid tegelikult seda ei ole. Ühel juhul viis generatiivse TI kasutamine õigusdokumendi koostamisel selleni, et tekst sisaldas olematuid kohtulahendeid [12]. Kuigi sellised juhtumid paljastavad tehnoloogia puudused, toovad need ühtlasi esile professionaalse järelevalve vajalikkuse ja vajaduse kontrollida, kas genereeritud teave on õige ja usaldusväärne. Kuna LLM-e kasutatakse väga erinevates valdkondades, võib ebapiisav järelevalve nende tulemuste üle kahjustada erinevaid põhiõigusi. Hea halduse õigus on siinkohal asjakohane näide, kuna ebatäpne teave võib takistada kvaliteetsete avalike teenuste osutamist ja neile juurdepääsu ning lõppkokkuvõttes õõnestada avalikkuse usaldust vastavate riigiasutuste vastu.

TI, privaatsus ja andmekaitse

Reaalses maailmas kasutamiseks sobivad TI-süsteemid nõuavad reaalseid andmeid. See tõstatab paratamatult küsimusi isikuandmete kaitse kohta, kuna üha raskem on jälgida, milliseid andmeid, millal ja mis eesmärgil kogutakse [13]. Digitaalne jälgimine on juba reaalsus, kuid TI muudab olukorda oluliselt mastaabi ja kiiruse poolest. Internetis saadaolevaid andmeid, sealhulgas isikuandmeid, saab kraapida ja kasutada koolitusalgoritmide jaoks, sageli ilma mõjutatud isikute teadmata või nõusolekuta. Samuti on võimalik kraapida pilte ja videosalvestisi, mis on jäädvustatud avalike kaameratega, näiteks CCTV süsteemidega, et kasutada neid näotuvastuse TI-süsteemide koolitamisel.

Tervise- ja haiguslood, sealhulgas retseptid, testid ja geneetiline teave, moodustavad veel ühe tundlike andmete kogumi, mida saab digitaalsel kujul edastada algoritmidele, et optimeerida tervishoiuteenuste osutamist, varude hankimist ja ravimite jaotamist. Kuid nõusoleku küsimus on vaid üks osa probleemist. TI mudelitesse sisestatud andmeid on äärmiselt raske kustutada, mis omakorda tekitab küsimusi andmete edasise kasutamise kohta, kui isik oma nõusoleku tagasi võtab või kui süsteemi hakatakse kasutama uuel eesmärgil [14]. Samuti ei ole TI mudelite koolitamiseks kasutatud andmete ebatäpsuste parandamine lihtne protsess.

Teatud TI-süsteemid võivad olla eriti intrusiivsed. Näidetena võib tuua TI tööriistad, mis võimaldavad reaalajas biomeetrilist kaugtuvastamist, käitumise ja/või emotsioonide tuvastamist ja jälgimist ning isikute hindamist ja järjestamist nende isikuomaduste või käitumise põhjal. Sellised tööriistad põhjustavad tõsiseid murekohti privaatsuse kaitse osas ja mõjutavad mitmeid põhiõigusi, sealhulgas inimväärikust, vabaduse ja isikupuutumatuse õigust, era- ja perekonnaelu puutumatuse õigust, isikuandmete kaitse õigust, sõna- ja teabevabadust, koosoleku- ja ühinemisvabadust ning diskrimineerimise keeldu.

TI sotsiaalsed mõjud

Samamoodi nagu on oluline tagada, et TI-süsteemid on kujundatud viisil, mis ei kahjustaks põhiõigusi, on ka oluline tagada, et nende süsteemide kasutamine ja rakendamine edendaks põhiõiguste täitmist. TI tehnoloogiate üha laialdasem juurutamine eri valdkondades toob ühiskonnas kaasa ulatuslikke muutusi. Automatiseerimine ja TI tööriistad ähvardavad asendada arvukalt töökohti, eriti sellistes valdkondades nagu tootmine, kontoritöö ja teenindus [15]. Sõltuvalt töökohtade kadumise kiirusest ja ulatusest võib see trend põhjustada tööpuuduse kiire kasvu, süvendada majanduslikku ebavõrdsust ja suurendada survet sotsiaalkaitsesüsteemidele. Nende ja teiste muutustega kohanemine nõuab aega ja hoolikat planeerimist. Samuti on vaja edendada oskuste arendamist ja ümberõppeprogramme, et tööjõud ja haavatavad rühmad saaksid TI tehnoloogiate kasutuselevõtust kasu.

TI arendamine on ressursimahukas valdkond, mis nõuab märkimisväärseid investeeringuid, ulatuslikku taristut ja hästi koolitatud tööjõudu. Kuna riigid soovivad püsida globaalses TI konkurentsis, võivad nad olla sunnitud suunama ressursse ümber teistest sektoritest, et tehnoloogia arendamisel mitte maha jääda. TI andmekeskustel on märkimisväärne keskkonnamõju [16]. Selline taristu vajab ka taskukohast ja stabiilset energiavarustust, mistõttu on energiajulgeolek TI innovatsiooni lahutamatu osa. Samuti on oodata, et TI sõjalised rakendused saavad veelgi suuremat tähelepanu, kuna TI võimekused muutuvad püsivalt riigikaitse planeerimise osaks.

TI väärkasutus

Üks täiendav TI juhtimise valdkond puudutab TI-süsteemide väärkasutuse ennetamist ja tõkestamist. TI areng pakub võimsaid tööriistu, mis võivad pahatahtlikul kasutamisel põhjustada märkimisväärset kahju nii üksikisikutele kui ka kogukondadele. See kehtib isegi näiliselt kahjututе tööriistade, näiteks generatiivsete TI-süsteemide kohta. TI-põhised pildi- ja videoloomise tööriistad võimaldavad võltsitud sisu ja süvavõltsingute tootmist, et diskrediteerida konkreetseid isikuid või toetada valeinfo levitamist eesmärgiga mõjutada demokraatlikke protsesse [17]. Nakatunud robotvõrgustikud võivad sellise sisu hetkega laialdaselt levitada. Pildi loomise ja töötlemise tööriistade väärkasutus võib raskendada ka õiguskaitseorganite tööd, näiteks muudetud lapsporno korduva kasutamise tuvastamisel [18]. Kurjategijad ja vägivaldsed äärmuslased võivad kasutada suuri keelemudeleid (LLM-e), et hankida tehnilisi teadmisi, andmeid konkreetsete asukohtade kohta või infot relvade arendamise kohta, et kavandada ja läbi viia rünnakuid kogukondade vastu.

Nagu teisedki digitehnoloogiad, on TI-süsteemid haavatavad turvarikkumiste ja manipuleerimise suhtes. Riskid hõlmavad näiteks koolitusandmestike mürgitamist, mille tulemusel algoritmid omandavad kahjulikke mustreid, algoritmide pahatahtlikku ümberkasutamist, tehniliste rikete ja muude tõrgete põhjustamist ning süsteemidesse sisenemist tundlike andmete saamiseks. TI-süsteeme saab väärkasutada ka poliitilistel eesmärkidel, näiteks poliitiliste vastaste jälgimiseks, meeleavaldajate tuvastamiseks ja kinnipidamiseks või veebiplatvormidel tsensuuri jõustamiseks, eemaldades sisu, mis seab kahtluse alla ametlikud seisukohad. Tõendid viitavad sellele, et ebademokraatlikud riigid kasutavad selliseid tööriistu üha enam teisitimõtlemise mahasurumiseks ja avaliku arutelu piiramiseks [19]. Nendes riikides puuduvad sageli piisavad kontrolli- ja tasakaalumehhanismid ning süstemaatiline lugupidamatus inimõiguste ja nende kaitsemehhanismide vastu jätab vähesed praktilised võimalused TI väärkasutuse eest vastutusele võtmiseks.

4. Riskide maandamise võimaluste kaardistamine

TI määrus keelustab teatud TI-süsteemid, mis kujutavad endast vastuvõetamatul tasemel ohtu põhiõigustele. See on pretsedenditu samm, mis juhib tähelepanu asjaolule, et teatud TI-süsteemide kättesaadavus ja kasutamine võivad põhjustada kahjulikke tagajärgi ELi kodanikele. Kuigi need keelud ei ole absoluutsed, on neil olulised tagajärjed põhiõiguste kaitse tagamisele kui TI-süsteemide ELis kasutuselevõtu olulisele eeltingimusele. Näiteks käsitlevad need kahesuguse kasutusviisiga TI-süsteemide võimalikku ümberkasutamist – st TI-süsteemide, mis on mõeldud riikliku julgeoleku, riigikaitse ja sõjaliste eesmärkide jaoks, kasutamist tsiviilotstarbel – et ennetada nende süsteemide ebaseaduslikku juurutamist ja väärkasutust [20]. Määrus keelab üheksa TI-süsteemi:

- **Alalävisele tajule suunatud võtteid või manipuleerivaid või petlikke võtteid kasutavad TI-süsteemid, mis võivad põhjustada olulist kahju**: need süsteemid võivad õõnestada inimeste autonoomiat otsuste tegemisel ja valikuvabadust ning meelitada inimesi ennasthävitavale käitumisele. Näideteks on voogedastusteenused, mis peidavad märkamatuid sõnumeid videotesse või filmidesse; sotsiaalmeediaplatvormid, mis edendavad algoritmiliselt emotsionaalselt laetud sisu kasutajate tunnete mõjutamiseks ja platvormil veedetud aja pikendamiseks; ning reklaamimeetodid, mis kasutavad petlikke või manipuleerivaid võtteid.
- **TI-süsteemid, mis kasutavad ära isikute haavatavust viisil, mis võib põhjustada olulist kahju**: need süsteemid võivad mõjutada inimeste käitumist, kasutades ära selliseid haavatavusi nagu vanus, tervislik seisund, majanduslik olukord või puue. Näideteks on TI-süsteemid, mis rakendavad andmeanalüüsi, et luua väga isikustatud veebireklaame, mis kasutavad tundlikke isikuandmeid inimeste valikute või ostusageduse mõjutamiseks.
- **Sotsiaalpunktide määramiseks kasutatavad TI-süsteemid**: TI-süsteemide kasutamine inimeste liigitamiseks nende sotsiaalse käitumise või isikuomaduste alusel võib viia diskrimineerimise ja marginaliseerimiseni. Näiteks TI-süsteemid, mis analüüsivad tööle kandideerijate sotsiaalmeedia tegevust ja teevad värbamisotsuseid tegurite põhjal, mis ei ole seotud töösooritusega, näiteks poliitilised vaated, usulised veendumused või kuulumine kindlatesse rühmadesse.
- **TI poolt isiksuseomaduste hindamisel põhinev ennustav politseitöö**: keeld hõlmab TI-süsteeme, mis hindavad inimese võimalikku kuritegelikku käitumist ainult profiilianalüüsi põhjal (nt rahvus, demograafilised näitajad, iseloom, majanduslik olukord) ilma mõistliku kahtluseta.
- **Näokujutiste kindla suunitluseta kraapimine näotuvastuse andmebaaside loomiseks**: need süsteemid on mõeldud näotuvastuse andmebaaside loomiseks või laiendamiseks, kogudes sihitult näopilte internetist või CCTV videosalvestistest. See keeld tugevdab õigust privaatsusele ja on oluline massilise jälgimise kultuuri ennetamisel.

- **TI-süsteemid emotsioonide tuvastamiseks töökohtades ja haridusasutustes**: keelatud on TI-süsteemid, mis püüavad analüüsida emotsionaalset seisundit töökohtades ja hariduskeskkondades. On mure, et need süsteemid ei ole usaldusväärsed ning võivad viia diskrimineerimiseni ja privaatsuse vähenemiseni. Keeld ei laiene TI-rakendustele, mis on mõeldud tervise või ohutuse eesmärkidel (nt meditsiinilistes või terapeutilistes keskkondades).
- **TI-süsteemid biomeetriliseks liigitamiseks tundlike isikuomaduste järeldamiseks**: keelatud on TI-süsteemide kasutamine inimeste liigitamiseks biomeetriliste andmete, näiteks näojoonte või sõrmejälgede analüüsi abil, et teha järeldusi nende rassi, poliitiliste vaadete, ametiühingusse kuulumise, usuliste või filosoofiliste veendumuste, seksuaalse orientatsiooni või seksuaalelu kohta. Õiguskaitse eesmärgil on lubatud erandid, näiteks piltide kategoriseerimine tunnuste alusel, nagu näiteks juuste või silmade värv, seaduses sätestatud eesmärkidel.
- **TI-süsteemid reaalajas biomeetriliseks kaugtuvastamiseks avalikus ruumis õiguskaitse eesmärkidel**: reaalajas biomeetriline kaugtuvastamine hõlmab biomeetriliste andmete jäädvustamist, võrdlemist ja tuvastamist peaaegu koheselt, ilma märkimisväärse viivituseta. Nende süsteemide kasutamine on keelatud, kuid teatud erandid on lubatud, näiteks tõsiste kuritegude ohvrite sihitud otsing, avalikkuse vastu suunatud vahetu rünnaku ennetamine ning raske kuriteo toimepanemises süüdistatava isiku tuvastamine ja tema asukoha määramine [21].

TI määrus määratleb teatud TI-süsteemid suure riskiga süsteemideks ning kehtestab reeglid ja nõuded nende loomiseks ja juurutamiseks, tagamaks, et selliste süsteemide kasutamine ei kujutaks ohtu ELi kodanike tervisele, ohutusele ja heaolule ega takistaks põhiõiguste täielikku kasutamist ja teostamist. TI-süsteemi peetakse suure riskiga süsteemiks, kui see on järgnevalt loetletud valdkondade toode, toote ohutuskomponent või kui seda kasutatakse ühes järgmistest valdkondadest: (1) biomeetriline analüüs; (2) elutähtis taristu (nt transport); (3) haridus (nt vahendid tulemuslikkuse hindamiseks); (4) tööhõive (nt töökohtade värbamine); (5) juurdepääs olulistele avalikele ja erateenustele (nt laenud, sotsiaaltoetused); (6) õiguskaitse (nt vahendid korduvkuritegevuse ohu hindamiseks); (7) immigratsioon (nt viisataotluste automatiseeritud läbivaatamine); ja (8) õigusemõistmine ja demokraatlikud protsessid (nt TI lahendused kohtupraktika otsimiseks). Kõrge riskiga TI-süsteemide koolitamisel kasutatavad andmekogumid peavad vastama kvaliteedikriteeriumidele, et tagada usaldusväärsus ja vältida kallutatud või diskrimineerivaid tulemusi. Kõrge riskiga TI-süsteemide kasutamisel peab olema tagatud sobiv inimjärelevalve kogu nende kasutusperioodi jooksul. Samuti peavad need süsteemid vastama piisavatele vastupidavuse ja küberturvalisuse standarditele, et kaitsta neid kolmandate osapoolte sekkumise, häkkimise ja manipuleerimise eest.

Soovitused

ELi põhiõiguste kaitse raamistik on terviklik ja tugev ning loob aluse inimkesksele TI innovatsioonile ja kasutamisele. Soovitused käsitlevad TI-süsteemide juhtimise põhiaspekte, nagu inimjärelevalve olulisus, vajadus ennetada ja takistada TI väärkasutust ning eetiliste ja vastutustundlike praktikate esmatähtsus TI-süsteemide arendamisel, kavandamisel ja juurutamisel.

Inimjärelevalve tagatised peavad paigas olema

- Riiklikud poliitika- ja õigusraamistikud, mis reguleerivad TI-süsteemide arendamist, juurutamist ja kasutamist, peavad sätestama inimjärelevalve nõude kogu nende süsteemide elutsükli jooksul ning sisaldama üksikasjalikke ja konkreetseid sätteid asjakohaste mehhanismide ja meetmete rakendamise ning jõustamise kohta.
- TI-süsteemide juurutajad peavad olema kohustatud teavitama inimesi, kes võivad nende süsteemidega kokku puutuda või olla neist mõjutatud, et kasutatakse TI-tehnoloogiat. Kui TI loodud sisu kasutatakse avaliku halduse või avalike teenuste osutamise eesmärgil, tuleb see sisu selgelt märgistada ja tuvastatavaks teha.
- Inimesi tuleb alati teavitada, kui nende isikuandmeid kasutatakse TI toega otsuste tegemiseks. Samuti tuleb neid teavitada, kui nende esitatud isikuandmeid töödeldakse automaatselt või säilitatakse TI-süsteemide koolitamiseks.
- Põhiõiguste mõjuhinnangud tuleb läbi viia kehtestatud ja standardiseeritud metoodikate alusel. Heaks näiteks selles valdkonnas on Euroopa Nõukogu TI-süsteemide riskide ja mõjude hindamise metoodika lähtuvalt inimõiguste, demokraatia ja õigusriigi vaatenurgast (HUDERIA), mis kirjeldab mitme osapoole kaasamisega struktureeritud protsessi võimalike negatiivsete mõjude tuvastamiseks ja maandamiseks TI-süsteemide juurutamise ja kasutamise puhul [22]. Tõhus osapoolte kaasamine on usaldusväärse ja põhjaliku mõjuhindamise läbiviimise ning TI-süsteemide arendamise ja juurutamise üle järelevalve tagamise oluline tingimus. Osapoolte kaasamise raamistikud hõlbustavad koostööd erinevate erialaste kogukondade vahel, näiteks teadus ja akadeemia, äri ja tööstus ning avalik sektor, ning edendavad nende sotsiaalsete gruppide kaasamist, keda TI-süsteemide kasutamine võib mõjutada [23]. TI-süsteemid peavad läbima perioodilisi auditeid, et tagada nende vastavus kehtestatud põhiõiguste standarditele. Samuti peab olema olemas selge strateegia selle kohta, kuidas lõpetada teatud TI-süsteemide kasutamine, kui need ei ole enam vajalikud.

- Riigid peavad tagama, et TI toega otsustusprotsessid on läbipaistvad ning pakkuma isikutele, keda sellised otsused mõjutavad, selged menetlused nende tulemuste vaidlustamiseks ametliku protsessi kaudu. Kaebusi tuleb läbi vaadata ja hinnata viisil, mis tagab asjakohase inimjärelevalve.

TI-süsteemide väärkasutuse ennetamine

- Riigid peavad võtma ennetavaid meetmeid tagamaks, et olemasolevad karistussüsteemid suudavad piisavalt käsitleda uusi kuritegevuse vorme või muid kahjulikke tegevusi, mida TI tehnoloogiate väärkasutamine võib võimaldada. Õiguslikud lüngad tuleb nõuetekohaselt kõrvaldada. Õppetunnid teistest valdkondadest, sealhulgas küberjulgeolekust ja sotsiaalmeedia reguleerimisest, võivad aidata kaasa olemasolevate seaduste ja poliitikainitsiatiivide ülevaatamisele ja ajakohastamisele [24]. Õiguskaitseasutused peaksid tagama, et nende tegevuses on mehhanismid tehnoloogia arengusuundade hindamiseks, et tuvastada õigeaegselt tekkivaid ohte.
- TI-süsteemid peavad läbima perioodilise hindamise, et tuvastada küberjulgeoleku haavatavusi ja ennetada sekkumisi süsteemi ülesehitusse või funktsioonidesse.
- Riigid peavad kehtestama piisavad menetlused, et tagada vastutuse kindlaksmääramine TI-süsteemide kasutamisega seotud juhtumite ja põhiõiguste rikkumise korral. See hõlmab selgeid juhtumite teatamise kohustusi ja vastutust ning taassertifitseerimise nõudeid pärast toodete turuleviimist.

Eetiliste ja vastutustundlike tavade edendamine TI-süsteemide arendamisel, kavandamisel ja juurutamisel

- Põhiõiguste kaitse tuleb integreerida TI-süsteemide kujundamise, arendamise ja juurutamise ahelasse. See hõlmab eetiliste äritavade tagamist toorainete ja tehniliste komponentide hankimisel ning TI-süsteemide kasutamisest ja toimimisest tulenevate negatiivsete sotsiaalsete ja keskkonnamõjude maandamist.
- TI arendajad peavad rakendama ennetavaid meetmeid võimalike negatiivsete mõjude maandamiseks. See hõlmab kallutatuse riski käsitlemist ning asjakohase andmekaitse ja andmehalduse tagamist TI-süsteemide koolitamisel ja testimisel.

- ELi liikmesriikide elanikel peab olema hea arusaam oma põhiõigustest ja nende kaitsemehhanismidest. Riikliku tasandi pädevad asutused, sealhulgas järelevalveorganid ja sõltumatud seireasutused, samuti kodanikuühiskonna organisatsioonid, peaksid osalema avalikes aruteludes TI tehnoloogiate võimalike mõjude üle ja panustama algatustesse vastutuse tugevdamiseks.
- Valitsuse, äri-, akadeemilise ja kodanikuühiskonna sektori sidusrühmad peavad tegema koostööd sujuva digipöörde hõlbustamiseks ning tagama ühiskonna kõigile liikmetele võrdse juurdepääsu TI tehnoloogia pakutavatele võimalustele. See nõuab sihipäraseid meetmeid olemasolevate digilõhede ületamiseks, sealhulgas koolitus- ja täiendõppeprogrammide kasutuselevõtmist digitaalse kirjaoskuse suurendamiseks ning oskuste arendamise edendamiseks TI kasutamise valdkonnas.

Viited

[1] Clara Piloto, "Top 5 Disruptive Technologies That Are Changing the World".
[2] David Crawford, Anne Hoecker, and Roy Singh, "Market for AI products and services could reach up to $990 billion by 2027, finds Bain & Company's 5th annual Global Technology Report", 25.09.2024; "Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds", 01.06.2023.
[3] Link to the report.
[4] Euroopa Liidu Põhiõiguste Amet, *Applying the Charter of Fundamental Rights of the European Union in Law and Policymaking at National Level – Guidance*, 16.10.2018.
[5] TI tehnoloogiatega seotud erinevat tüüpi riskide ülevaate saamiseks vt nt *AI Risk Repository: a comprehensive database of risks from AI systems*, MIT 2024.
[6] IBM Data and AI Team, "Shedding light on AI bias with real world examples", 16.10.2023.
[7] Euroopa Liidu Põhiõiguste Amet, *Bias in algorithms – Artificial Intelligence and Discrimination*, 08.12.2022.
[8] James Holdsworth, "What Is AI Bias?", 22.12.2023.
[9] Süsteemse või institutsionaalse kallutatuse kohta vt Reva Schwartz, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, NIST Special Publication 1270, 16.03.2022.
[10] Nicola Jones, "AI Hallucinations Can't Be Stopped — but These Techniques Can Limit Their Damage", *Nature,* 21.01.2025.
[11] "*What Are AI Hallucinations?*", 01.09.2023.
[12] Ralph Artigliere, "AI Hallucinations in Court: A Wake-Up Call for the Legal Profession", *JDSUPRA*, 22.01.2025.
[13] Katharine Miller, *Privacy in an AI Era: How Do We Protect Our Personal Information?*, 18.03.2024; Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era Policy Provocations for a Data-Centric World*, White Paper, veebruar 2024.
[14] Vt nt Stephen Pastis, "A.I.'s Un-Learning Problem: Researchers Say It's Virtually Impossible to Make an A.I. Model 'Forget' the Things It Learns from Private User Data", *Fortune*, 30.08.2023.
[15] Bernard Marr, "What Jobs Will AI Replace First?", *Forbes*, 17.06.2024.
[16] Jack Marley, "Data Centre Emissions Are Soaring – It's AI or the Climate", *The Conversation*, 02.10.2024.

[17] Vt nt "Generative AI Is the Ultimate Disinformation Amplifier" *DW*, 26.03.2024.
[18] Katie McQue, "AI Is Overpowering Efforts to Catch Child Predators, Experts Warn", *Guardian*, 18.07.2024; vt ka David Thiel, "Investigation Finds AI Image Generation Models Trained on Child Abuse", *Stanford Cyber Policy Center Blog*, 20.12.2023.
[19] Vt nt Lena Masri, "Facial Recognition Is Helping Putin Curb Dissent with the Aid of U.S. Tech", *Reuters*, 28.03.2023; Mattias Carlson et al. "Kremlin Leaks: How Putin's Regime Is Building AI Surveillance Operations", *VSquare*, 27.03.2024; Abdulhakim Idris, "Application of AI in Human Rights Violations – The Uyghur Case", 27.01.2024; Steven Feldstein, "China's High-Tech Surveillance Drives Oppression of Uyghurs", *Bulletin of Atomic Scientists*, 27.10.2022.
[20] Euroopa Komisjon, Approval of the content of the draft Communication from the Commission – Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), 04.02.2025.
[21] Osman Gazi Güçlütürk ja Bahadir Vural, "Navigating Prohibited Practices under the AI Act", *Holistic AI*, 03.05.2024.
[22] Euroopa Nõukogu, *Methodology for risk and impact assessment of artificial intelligence (AI) systems from the point of view of human rights, democracy and the rule of law (HUDERIA)*, 28.11.2024.
[23] Vt nt Partnership on AI's Global Task Force for Inclusive AI, *Guidelines for Participatory and Inclusive AI*, 17.09.2024; Society Inside and European Center for Not-for-Profit Law, *Framework for Meaningful Stakeholder Engagement in AI Development*, 08.03.2023.
[24] Vt nt Miles Brundage et. al, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, veebruar 2018.

refrAIme

Reinforcing Equality and Fundamental Rights in an Artificial Intelligence-Maintained Environment - REFRAIME

See poliitikadokument täiendab uuringut "Background Study on the Fundamental Rights Implications of the Use of AI" ("Taustauuring TI kasutamise mõju kohta põhiõigustele"). Selle vaatas üle REFRAIME hindamis- ja seirekomisjon.

Väljaande kujundus: Canva

Co-funded by the European Union

Rahastatud Euroopa Liidu poolt. Avaldatud seisukohad ja arvamused on ainult autori omad ja ei pruugi kajastada Euroopa Liidu või EACEA seisukohti. Nende eest ei saa vastutusele võtta ei Euroopa Liitu ega rahastusasutust.