

EQUITECH



Co-funded by
the European Union

REPORT OF THE TESTING THE PUBLIC ADM SYSTEMS IN ESTONIA

(Estonian version)

Project 101144709 — EquiTech - Improving response to risks of discrimination, bias and intolerance in automated decision-making systems to promote equality

The Gender Equality and Equal Treatment Commissioner's Office of Estonia

The Office of the Equal Opportunities Ombudsperson of Lithuania

Ministry of Justice and Digital Affairs of Estonia

Tallinn University of Technology

20.12.2025



Co-funded by
the European Union

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



REPUBLIC OF ESTONIA
MINISTRY OF JUSTICE
AND DIGITAL AFFAIRS



GENDER EQUALITY AND
EQUAL TREATMENT COMMISSIONER



Lygių galimybių
kontrolieriaus
tarnyba

**TAL
TECH**

Sisukord

1. Sissejuhatus	4
2. AI ja ADM kasutamise hetkeseis Eesti avalikus sektoris	6
3. Valideerimisintervjuude metoodika	8
3.1 Valideerimise eesmärk	8
3.2 Kasutusjuhtumite valik.....	8
3.3 Intervjuude läbiviimine ja testitud materjalid	9
3.4 Metoodika piirangud	9
4. Valideeritud kasutusjuhtumid	10
4.1 Otsustustoetussüsteem toetustaotluste hindamiseks (PRIA).....	10
4.2 Vestlusrobot “Bürokratt”	10
4.3 Masinõppel põhinev analüütiline tööriist ettevõtete elujõulisuse hindamiseks (MKM–STAT).....	11
4.4 Kasutusjuhtumite mitmekesisuse põhjendus.....	12
5. Kasutusjuhtumipõhised hindamistulemused.....	13
5.1 Otsustustoetussüsteem toetustaotluste hindamiseks (PRIA).....	13
5.2 Vestlusrobot Bürokratt.....	15
5.3 Masinõppel põhinev analüütiline tööriist ettevõtete elujõulisuse hindamiseks (MKM–STAT).....	17
6. Valideerimise peamised järeldused	21
6.1 Valideerimistulemuste kasutamine tööriistade lõplikul kujundamisel.....	24
7. Kokkuvõte ja järeldused	25

1. Sissejuhatus

Tehisintellekti (AI) ja algoritmiliste süsteemide kasvav kasutamine avalikus sektoris on loonud uusi võimalusi tõhustada teenuste osutamist, parandada otsuste järjepidevust ning suurendada haldusprotsesside kvaliteeti. Samal ajal võivad sellised süsteemid kaasa tuua või võimendada riske, mis on seotud kallutatuse, diskrimineerimise ja ebavõrdse kohtlemisega, eriti olukordades, kus algoritmilised väljundid mõjutavad inimesi või rühmi puudutavaid otsuseid.

Nende riskide teadvustamise tulemusena vajavad avaliku sektori asutused üha enam praktilisi tööriistu, mis toetaksid AI- ja automatiseeritud otsustussüsteemide (ADM) vastutustundlikku kavandamist, kasutuselevõttu ja rakendamist. Sellised tööriistad on eriti olulised kontekstides, kus täielikult automatiseeritud otsuseid kasutatakse veel piiratud määral, kuid AI-põhised lahendused toetavad inimotsustamist või pakuvad teavet otse avalikkusele.

Käesolev raport esitab valideerimistegevuse tulemused, mis viidi läbi osana laiemast projektist, mille eesmärk on vähendada diskrimineerimise ja kallutatuse riske avaliku sektori AI- ja ADM-süsteemides. Valideerimise eesmärk oli hinnata kolme projekti raames välja töötatud omavahel seotud tööriista praktilist rakendatavust ja kasulikkust:

1. algoritmilise kallutatuse riskihalduse meetodikat,
2. sellega seotud juhendmaterjale ning
3. struktureeritud riskihindamise töölehte.

Raporti eesmärk ei ole hinnata konkreetsete AI-süsteemide tehnilist toimivust ega auditeerida õigusaktidele vastavust. Selle asemel keskendub raport sellele, kas loodud materjalid toetavad avaliku sektori asutusi potentsiaalsete diskrimineerimis- ja kallutatusriskide tuvastamisel, analüüsimisel ja leevendamisel, kui neid rakendatakse reaalses AI-lahendustes.

Valideerimine põhines kvalitatiivsetel intervjuudel avaliku sektori organisatsioonidega, kes rakendasid meetodikat, juhendeid ja töölehte arenduses või kasutuselevõtu faasis olevate AI-süsteemide puhul. Need süsteemid valiti seetõttu, et need on kas funktsioonilt lähedased automatiseeritud otsustamisele või omavad otsest või kaudset mõju inimestele. Intervjuud andsid ülevaate sellest, kuidas materjalid praktikas toimivad, milliseid väljakutseid nende kasutamisel esines ning millistes valdkondades on vajalik edasine täiendamine.

Raporti ülesehitus on järgmine: peatükk 2 annab lühikese ülevaate AI ja ADM kasutamise hetkeseisust Eesti avalikus sektoris, tuginedes varasemale projektitööle. Peatükk 3 kirjeldab valideerimisintervjuude metoodikat. Peatükk 4 tutvustab valideerimise käigus hinnatud AI kasutusjuhtumeid. Peatükk 5 esitab hinnangud kasutusjuhtumite lõikes. Peatükk 6 koondab peamised valideerimistulemused. Peatükk 7 selgitab, kuidas valideerimise tulemusi kasutati tööriistade lõplikul kujundamisel. Raport lõpeb kokkuvõtte ja järeldustega.

2. AI ja ADM kasutamise hetkeseis Eesti avalikus sektoris

Eesti tunnustatakse laialdaselt arenenud digitaalse valitsemise ning andmepõhiste lahenduste kasutamise eest avalikus sektoris. Viimase kümnendi jooksul on avaliku sektori asutused üha enam rakendanud algoritmilisi süsteeme haldusprotsesside toetamiseks, teenuste parandamiseks ja analüütilise võimekuse suurendamiseks. Viimastel aastatel on neid süsteeme täiendamas ka tehisintellektil põhinevad lahendused, eeskätt andmeanalüüsi, otsustustoe ja kodanikega suhtlemise valdkonnas.

Samas on täielikult automatiseeritud otsustussüsteemide (ADM) kasutamine Eesti avalikus sektoris endiselt suhteliselt piiratud. Enamasti kasutatakse AI- ja algoritmilisi süsteeme inimotsustamise toetamiseks, mitte selle täielikuks asendamiseks. Süsteemide väljundid võivad toetada hindamist, prioriseerimist, prognoosimist või teabe edastamist, kuid lõplikud otsused jäävad üldjuhul ametnike teha.

Hoolimata täielikult automatiseeritud otsuste vähesest levikust on varasem projektitöö (D3 audit diskrimineerimise ja kallutatuse kohta Eesti avaliku sektori ADM-süsteemides) näidanud, et ka otsustustoetussüsteemid ja AI-põhised teenused võivad kätkeada kallutatuse ja diskrimineerimise riske. Need riskid võivad tuleneda näiteks kallutatud treeningandmetest, algoritmidesse sisse kirjutatud disainivalikutest või sellest, kuidas inimesed süsteemi väljundeid tõlgendavad ja kasutavad. Otsese avalikkusega suhtlevates süsteemides võivad lisanduda riskid, mis on seotud valeinfo, ebavõrdse ligipääsu või kasutajate erineva kohtlemisega.

Varasem audit tõi samuti esile, et paljud avaliku sektori AI-lahendused toimivad keerukas regulatiivses ja organisatsioonilises keskkonnas, kus vastutus on jaotunud mitme osapoole vahel ning läbipaistvuse ja vastutuse mehhanismid on ebaühtlased. Samuti rõhutati, et olemasolevad juhtimisraamistikud ei paku alati piisavalt praktilisi juhiseid kallutatuse riskide tuvastamiseks ja leevendamiseks süsteemide kavandamise ja rakendamise käigus.

Sellest tulenevalt on selge vajadus praktiliste ja kontekstitundlike tööriistade järele, mida saab rakendada erinevat tüüpi AI- ja ADM-süsteemide puhul. Need tööriistad peavad arvestama nii täielikult automatiseeritud otsustega kui ka otsustustoetussüsteemide ja AI-lahendustega, mis mõjutavad tulemusi kaudselt või kujundavad avaliku sektori ja inimeste vahelist suhtlust.

Käesolevas raportis kirjeldatud valideerimistegevus tugineb nendele tähelepanekutele, keskendudes uute riskihaldusvahendite praktilisele rakendamisele reaalsetes avaliku sektori kontekstides. Metoodika, juhendite ja töölehe testimine erinevates AI kasutusjuhtumites aitab

ületada lõhe kõrgetasemeliste põhimõtete ja igapäevase praktika vahel ning toetab ühtlasemat ja teadlikumat lähenemist diskrimineerimise ja kallutatuse riskide juhtimisele avaliku sektori AI-süsteemides.

3. Valideerimisintervjuude metoodika

3.1 Valideerimise eesmärk

Valideerimisintervjuude eesmärk oli hinnata projekti raames välja töötatud algoritmilise kallutatuse riskihalduse metoodika, juhendmaterjalide ja riskihindamise töölehe praktilist rakendatavust, selgust ja kasulikkust. Valideerimine keskendus sellele, kuidas need materjalid toimivad reaalses avaliku sektori praktikas ning kas need toetavad organisatsioone diskrimineerimise ja kallutatuse riskide tuvastamisel ja leevendamisel.

Arvestades, et Eesti avalikus sektoris ei kasutata täielikult automatiseeritud otsustussüsteeme, ei olnud valideerimise eesmärk testida täielikult toimivaid ADM-süsteeme. Selle asemel keskenduti AI-põhiste süsteemidele, mis on kas funktsioonilt lähedased automatiseeritud otsustamisele või millel võib olla otsene või kaudne mõju inimestele otsustustoe või teabe edastamise kaudu.

3.2 Kasutusjuhtumite valik

Valideerimine põhines kolmel kvalitatiivsel intervjuul avaliku sektori organisatsioonidega, kes olid valmis rakendama loodud materjale arenduses või kasutuselevõtu faasis olevate AI-lahenduste puhul. Organisatsioonidega võeti ühendust otse ning valideerimises osalesid ainult need asutused, kes ka intervjuudes osalesid.

Kasutusjuhtumite valikul kasutati sihipärast valimit, mille eesmärk oli katta AI-lahendusi, mis erinevad oma eesmärgi, tehnilise iseloomu ja kasutajatega suhtlemise viisi poolest. Valikut mõjutasid järgmised kaalutlused:

PRIA kaasati pärast seda, kui asutus avaldas huvi riskihaldusvahendite vastu, olles teadlik nende väljatöötamisest. Valideerimise ajal oli PRIA arendamas otsustustoetussüsteemi toetuste taotluste hindamiseks. Tegemist on süsteemiga, mis on oma mõjult lähedane automatiseeritud otsustamisele ning võib mõjutada isikute jaoks olulisi tulemusi.

Välisministeerium osales valideerimises seoses Bürokrati vestlusroboti rakendamisega. Bürokratt suhtleb otseselt inimestega ning pakub teavet avalike teenuste kohta. Seetõttu peeti oluliseks hinnata kallutatuse, valeinfo ja ebavõrdse kohtlemise riske avalikkusega suhtlevas süsteemis.

Majandus- ja Kommunikatsiooniministeeriumi ning Statistikaameti (MKM-STAT) kasutusjuhtum hõlmas masinõppel põhinevat analüütilist lahendust, mis pakub ettevõtjatele

äriühenduse hinnanguid ja prognoose. Kuigi süsteem ei tee siduvaid otsuseid, võivad selle väljundid mõjutada kasutajate otsuseid. See kasutusjuhtum valiti, et testida tööriistade rakendatavust analüütiliselt keerukates ja kaudse mõjuga süsteemides.

Lisaks võimaldas nende kasutusjuhtumite valik täiendada projekti varasemat auditit (D3), mis keskendus peamiselt teistsugustele AI- ja ADM-lahendustele.

3.3 Intervjuude läbiviimine ja testitud materjalid

Kokku viidi läbi kolm poolstruktureeritud intervjuud veebipõhiselt, millest igaüks kestis ligikaudu ühe tunni. Osalevatele organisatsioonidele edastati eelnevalt: algoritmilise kallutatuse riskihalduse meetoodika, juhendmaterjalid ja riskihindamise tööleht.

Organisatsioonidele anti ligikaudu kaks nädalat aega, et materjalidega tutvuda ja proovida töölehte oma AI-lahenduse kontekstis. Intervjuude käigus kirjeldasid osalejad oma kogemusi materjalide kasutamisel, tõid esile arusaadavuse ja kasutusmugavusega seotud küsimusi ning arutlesid riskide tuvastamise ja dokumenteerimise üle.

Intervjuud keskendusid praktilistele kogemustele ning sellele, kuidas materjalid toetasid sisemisi arutelusid ja otsustusprotsesse.

3.4 Meetoodika piirangud

Valideerimine oli kvalitatiivne ning hõlmas piiratud arvu kasutusjuhtumeid. Seetõttu ei ole tulemused statistiliselt üldistatavad kõigile avaliku sektori AI- ja ADM-süsteemidele. Valideerimise eesmärk ei olnud üldistamine, vaid tööriistade põhjalik testimine erinevates realistlikes kasutusolukordades.

Kuna täielikult automatiseeritud otsustussüsteemid hetkel ei ole Eestis kasutusel, keskendus valideerimine otsustustoetussüsteemidele ja AI-lahendustele, mille mõju inimestele on kaudne. Tulevikus võiks valideerimist laiendada, hõlmates rohkem ADM-süsteeme nende leviku kasvades.

4. Valideeritud kasutusjuhtumid

Käesolev peatükk annab ülevaate valideerimise käigus hinnatud AI kasutusjuhtumitest. Kirjeldused keskenduvad iga süsteemi eesmärgile, funktsionaalsusele ja võimalikule mõjule, mitte tehnilistele rakendusdetailidele. Peatüki eesmärk on luua piisav kontekst edasiste hindamistulemuste mõistmiseks, vältides samas tundliku või süsteemispetsiifilise teabe avalikustamist.

Valideerimisel hinnatud kasutusjuhtumid esindavad erinevat tüüpi AI-rakendusi avalikus sektoris, ulatudes otsustustoetussüsteemidest kuni avalikkusega otseselt suhtlevate lahendusteni. Koos annavad need ülevaate kontekstidest, kus diskrimineerimise ja kallutatuse riskid võivad tekkida erineval viisil.

4.1 Otsustustoetussüsteem toetustaotluste hindamiseks (PRIA)

Esimene kasutusjuhtum puudutab Põllumajanduse Registrate ja Informatsiooni Ameti (PRIA) poolt arendatavat otsustustoetussüsteemi. Süsteemi eesmärk on toetada toetustaotluste hindamise protsessi, aidates hindamiskomisjonidel analüüsida ja võrrelda taotlejate esitatud äriplaanide.

Süsteem on kavandatud koondama ja struktureerima teavet, mis on seotud eelnevalt määratletud hindamiskriteeriumidega, tuginedes taotlustes esitatud andmetele ning täiendavale kontekstuaalsele teabele. Süsteemi peamine eesmärk on vähendada hindajate halduskoormust ning toetada järjepidevamat ja informeeritumat hindamisprotsessi. Valideerimise ajal oli süsteem arendusfaasis ning ei olnud kasutusel täielikult automatiseeritud otsustussüsteemina.

Kuigi süsteem ei tee lõplikke otsuseid, võivad selle väljundid mõjutada hindamisprotsessi, kujundades viisi, kuidas teavet otsustajatele esitatakse. Seetõttu on tegemist AI-rakendusega, mis on oma mõjult lähedane automatiseeritud otsustamisele, hoolimata inimjärelvalve säilimisest.

4.2 Vestlusrobot “Bürokratt”

Teine kasutusjuhtum on seotud Bürokrati rakendamisega – tegemist on riigi loodud tehisintellektil põhineva digitaalse assistendiga, mille eesmärk on toetada avaliku sektori asutusi kaasaegse ja tõhusa klienditeeninduse pakkumisel. Bürokratt võimaldab asutustel

kasutada suuri keelemudeleid (LLM-e), et mõista loomulikus keeles esitatud päringuid ning vastata kasutajate küsimustele kiiresti, viitepõhiselt ja ööpäevaringselt.

Bürokratt on mõeldud inimestele, kes soovivad suhelda riigiasutustega ja saada teavet avalike teenuste kohta. Samal ajal on see kavandatud ühishahendusena avaliku sektori asutustele, sealhulgas nii riigiasutustele kui ka kohalikele omavalitsustele, kes soovivad parandada klienditeeninduse kättesaadavust, tõhusust ja järjepidevust. Tüüpilised kasutusjuhtumid hõlmavad korduma kippuvatele küsimustele vastamist, teabenõuete haldamist ning rutiinsete suhtluste automatiseerimist, mis muidu eeldaksid ametnike käsitsi sekkumist.

Välisministeeriumi kontekstis rakendatakse Bürokratti kasutajatega suhtlemise toetamiseks seoses välisesindustega. Süsteem hangib ja genereerib vastuseid avalikult kättesaadava teabe ning asutusesisese sisu põhjal. Valideerimise ajal oli Bürokratt testimise ja rakendamise etapis.

Seda kasutusjuhtumit iseloomustab otsene suhtlus inimestega, mis toob kaasa spetsiifilised kaalutlused seoses diskrimineerimise ja kallutatuse riskidega. Kuigi Bürokratt ei tee haldusotsuseid, võivad selle vastused mõjutada kasutajate arusaama menetlustest, õigustest ja kohustustest. Sellest tulenevalt võivad potentsiaalsed riskid seisneda ebatäpses, aegunud või kallutatud teabes, kasutajate ebavõrdses kohtlemises või süsteemi väljundite väärkasutuses. Need omadused muudavad Bürokrati eriti asjakohaseks kasutusjuhtumiks riskihaldusvahendite testimisel, mille eesmärk on käsitleda kallutatuse ja diskrimineerimise riske avaliku sektori AI-süsteemides, mis suhtlevad otseselt inimestega.

4.3 Masinõppel põhinev analüütiline tööriist ettevõtete elujõulisuse hindamiseks (MKM-STAT)

Kolmas kasutusjuhtum puudutab Statistikaameti ja Majandus- ja Kommunikatsiooniministeeriumi koostöös arendatud masinõppel põhinevat analüütilist tööriista. Süsteem pakub ettevõtjatele, eelkõige väikestele ja keskmise suurusega ettevõtetele, analüüsi ja prognoose ettevõtte elujõulisuse kohta.

Tööriist töötleb andmeid, mis on juba avalikule sektorile esitatud, ning rakendab masinõppemudeleid analüütiliste väljundite ja prognooside loomiseks. Teenus on vabatahtlik ja informatiivne ega loo siduvaid otsuseid ega kohustusi.

Hoolimata mittesiduva iseloomuga väljunditest võivad süsteemi pakutavad hinnangud mõjutada kasutajate ärilisi otsuseid. Süsteemi tõenäosuslik ja analüütiline olemus tõstatab ka

spetsiifilisi väljakutseid seoses ebakindluse mõistmise, mudelite piirangute ning treeningandmetes või modelleerimisvalikutes sisalduva võimaliku kallutatusega.

4.4 Kasutusjuhtumite mitmekesisuse põhjendus

Kolm hinnatud kasutusjuhtumit valiti teadlikult selleks, et esindada erinevaid AI kasutamise viise avalikus sektoris:

- otsustustoetussüsteemid, mis mõjutavad haldusprotsesse,
- vestlusrobotid, mis suhtlevad otseselt inimestega,
- analüütilised süsteemid, mis pakuvad tõenäosuslikku teavet kasutajatele.

Selline mitmekesisus võimaldas valideerimise käigus hinnata, kas loodud meetoodika, juhendmaterjalid ja tööleht on piisavalt paindlikud, et käsitleda diskrimineerimise ja kallutatuse riske erinevates tehnilistes ja organisatsioonilistes kontekstides. Samuti võimaldas see võrrelda, kuidas riskihindamise väljakutsed erinevad sõltuvalt süsteemi eesmärgist, kasutajatega suhtlemise viisist ja automatiseeritud otsustamisele lähedusest.

5. Kasutusjuhtumipõhised hindamistulemused

5.1 Otsustustoetussüsteem toetustaotluste hindamiseks (PRIA)

PRIA kasutusjuhtum pakkus põhjaliku ja praktikale orienteeritud vaate algoritmilise kallutatuse riskihalduse meetodika, juhendmaterjalide ja riskihindamise töölehe rakendamisele arendamisel oleva otsustustoetussüsteemi kontekstis. Valideerimine toimus etapis, kus süsteemi eesmärgid ja peamised kasutusjuhtumid olid määratletud, kuid tehnilised teostusvalikud alles kujunemas.

Materjalide rakendamine praktikas

PRIA-s algatas valideerimisprotsessi projektijuht, kes tutvus esmalt iseseisvalt juhendmaterjalide ja meetodikaga ning alustas seejärel töölehe täitmist sissejuhataval kohtumisel arutatud teabe põhjal. Materjale vaadati üle korduvalt ning töölehte ajakohastati iteratiivselt, kui arusaam meetodikast süvenes.

Materjalide läbivaatamine ja töölehe täitmine toimus etappide kaupa, mitte üheainsa tööseansina. Kokku võttis valideerimisprotsess, sealhulgas ettevalmistavad arutelud, ligikaudu 17–18 tundi. Valideerimise ajal ei olnud tehnilised eksperdid ega süsteemi kasutajad kättesaadavad, mistõttu täitis töölehe suuresti üks roll. Samas märgiti, et edasises etapis kaasatakse täiendavaid meeskonnaliikmeid, et ühiselt üle vaadata tuvastatud riskid ning kokku leppida riskide käsitlemise otsustes.

Tagasiside juhendmaterjalidele ja meetodikale

Juhendmaterjale hinnati väga positiivselt, eeskätt neis sisalduvate konkreetsete näidete ja algoritmilise kallutatuse ning diskrimineerimise taustaselgituste tõttu. Neid peeti abistavaks teema mõistmisel ning kättesaadavaks ka isikutele, kellel on AI-süsteemidega piiratud varasem kokkupuude.

Samas tuvastati valideerimise käigus mitmeid valdkondi, kus materjale saaks tugevdada:

- Osalejad märkisid, et meetodika viitab erinevatele standarditele, kuid ei tee alati piisavalt selgeks, kuidas neid praktikas rakendada. Soovitati selgemalt näidata, kas konkreetne standard on antud projekti puhul asjakohane ning milliseid samme selle rakendamine eeldab.
- Terminoloogia osutus kohati segadust tekitavaks. Eelkõige ei olnud alati selge, kas mõisted „AI-süsteem“ ja „algoritmiline süsteem“ viitavad konkreetsele mudelile või laiemale,

mitmest komponendist koosnevale süsteemile. Lisaks raskendas materjalide järjepidevat kasutamist rollide ja vastutuste erinev kirjeldus metoodikas ja juhendmaterjalides.

- Osalejad tõid välja, et kallutuse leevendusmeetmete näited keskendusid peamiselt mudeli arendusfaasile. Soovitati lisada rohkem näiteid olukordade jaoks, kus mudelit ennast ei ole võimalik muuta ning kus riskide maandamine peab toimuma juhtimis-, protsessi- või inimjärelvalve meetmete kaudu.

Tagasiside töölehele ja eeltäidetud näitele

Riskihindamise töölehte peeti kasulikuks riskide struktureerimiseks ja eelduste dokumenteerimiseks, kuid selle praktiline rakendamine tõi esile mitmeid piiranguid.

Korduv probleem ilmnis süsteemide puhul, mis tuginevad mitmele paralleelselt toimivale mudelile. Tööleht ei pakkunud selget võimalust dokumenteerida rohkem kui ühte mudelit sama süsteemi raames. Selle tulemusena tuli tabeleid käsitsi dubleerida ning eri mudelite nimetamine ja eristamine ei olnud alati üheselt selge, mis suurendas segaduse riski mitme kasutusjuhtumi teabe koondamisel.

Mõned töölehe osad olid osalejate hinnangul ebapiisavalt selgitatud. Eelkõige jäi ebaselgeks viidete ja allikatega seotud väljade eesmärk: millist tüüpi teavet oodatakse ning miks on viited riskihindamise osana vajalikud.

Samuti seati kahtluse alla teatud rollide, nagu „importija“ ja „levitaja“, asjakohasus. Riskihindamise seisukohast soovitati keskenduda pigem püsivatele elementidele, nagu mudel ise ja operatiivne keskkond (nt pilveteenuse asukoht), mitte rollidele, mis võivad muutuda sõltuvalt hankest või rakenduslahendusest.

Eeltäidetud näidis-töölehte peeti kasulikuks üldise illustratsioonina, kuid liiga pealiskaudseks, et toimida tugeva võrdlusmaterjalina. Eriti riskide käsitlemise otsuseid puudutavad osad sisaldasid üksnes üldtasemel märksõnu ilma selgitava tekstita, mistõttu oli keeruline mõista, milline näeb välja praktikas hästi täidetud hindamine. Osalejad soovitasid, et abivahendina mõeldud eeltäidetud tööleht oleks detailsem ja sisuliselt lähedasem oodatavale lõpptulemusele.

Lisaks märgiti, et töölehe täitmine aktiivse arendusfaasi ajal võib olla keeruline, kuna süsteemi funktsionaalsus ja eeldused võivad muutuda. See tõstas küsimuse, millise detailsusega teavet tuleks dokumenteerida arenduse eri etappides, ning viitas vajadusele parema toe järele iteratiivsete ajakohastuste tegemisel.

Üldhinnang

Kokkuvõttes näitas PRIA kasutusjuhtumi valideerimine, et metoodika, juhendmaterjalid ja riskihindamise tööleht on kasulikud ja asjakohased tööriistad otsustustoetussüsteemide puhul, mis võivad kaudselt mõjutada haldusotsuste tulemusi. Materjalid toetasid süstemaatilist riskide läbimõtlemist ning aitasid esile tuua küsimusi, mida projektis ei olnud varem selgesõnaliselt käsitletud.

Samas tuvastati valideerimise käigus konkreetsed ja praktikast lähtuvad parendusvõimalused, eelkõige seoses:

1. selgema ja järjepidevama terminoloogia kasutamisega dokumentide lõikes;
2. otsese toe pakkumisega mitme mudeli dokumenteerimiseks ühe süsteemi raames;
3. detailsemate ja kontekstitundlikumate leevendusmeetmete näidete lisamisega;
4. konkreetsete töölehe väljade täitmise selgema juhendamisega;
5. ning parema toe pakkumisega tööriistade kasutamiseks iteratiivse süsteemiarenduse käigus.

5.2 Vestlusrobot Bürokratt

Bürokraati kasutusjuhtum andis ülevaate algoritmilise kallutatuse riskihalduse metoodika, juhendmaterjalide ja riskihindamise töölehe rakendamisest vestlusrobotis, mis suhtleb otseselt inimestega. Erinevalt PRIA kasutusjuhtumist ei toeta see süsteem haldusotsuste tegemist, kuid selle väljundid võivad mõjutada kasutajate arusaama avalikest teenustest, menetlustest ja kohustustest. Otsene suhtlus avalikkusega muutis eriti oluliseks kallutatuse, valeinfo ja mainekahju riskide hindamise.

Materjalide rakendamine praktikas

Bürokraati puhul algas valideerimisprotsess kõigi materjalide lühikese ülevaatamisega, millele järgnes riskihindamise töölehe otsene rakendamine. Riskihindamine viidi läbi koostöös süsteemi arendaja esindajatega, kuna asjakohane tehniline ja operatiivne teave oli arendusmeeskonna sees kättesaadav.

Tööleht täideti suhteliselt tõhusalt; esmane riskihindamine võttis ligikaudu pool tööpäeva. Osalejad märkisid, et hindamise lõplikuks viimistlemiseks ja formaalseks dokumenteerimiseks

oleks vaja täiendavat aega. Võrreldes teiste kasutusjuhtumitega kasutati töölehte pigem praktilise töövahendina, mitte algusest peale detailse aruandlusdokumendina.

Juhendmaterjale kasutati peamiselt pärast töölehe esmast täitmist, et kontrollida, kas olulised riskid olid tuvastatud. See viitas sellele, et praktikas toimus tööleht esmase sisenemispunktina riskihindamisse, samas kui juhendid täitsid toetava rolli ega olnud kasutusel samm-sammulise käsiraamatuna.

Tagasiside juhendmaterjalidele ja metoodikale

Osalejad leidsid, et juhendmaterjalid ja metoodika on sisuliselt põhjalikud, kuid mõnel juhul tajuti neid mahukate ja keeleliselt keerukatena, eriti ajapiirangute tingimustes. Seetõttu ei kasutatud materjale alati esialgse riskihindamise käigus, vaid pigem hiljem tuvastatud riskide kinnitamiseks või täiendamiseks.

Esitati mitmeid parendusettepanekuid:

- Tehti ettepanek koostada juhendmaterjalidest lühem ja kokkuvõtlikum versioon, mis toimiks kiireks sissejuhatuseks olukordades, kus hindamine tuleb läbi viia tõhusalt ja piiratud ajaga.
- Mõned juhendite osad olid sõnastuselt liialt tehnilised. Näiteks AI-süsteemi elutsükli ja andmete kogumist käsitlevaid selgitusi peeti raskesti kiiresti mõistetavaks ning need võiksid olla esitatud kasutajasõbralikumas keeles.
- Osalejad rõhutasid vajadust selgema juhise järele olukordades, kus ühtegi riski taset ei peeta vastuvõetavaks. Bürokrati kontekstis peeti näiteks valeinfo edastamise riski põhimõtteliselt lubamatuks, mistõttu oli keeruline rakendada riskide skoorimist või jääkriski aktsepteerimise loogikat.

Lisaks soovitati, et metoodika käsitleks selgemalt kommunikatsiooniga seotud riske, eriti avalikkusega suhtlevate süsteemide puhul. Märgitakse, et organisatsiooniline ja mainekahju võib ulatuda kaugemale ühest asutusest ning mõjutada usaldust avaliku halduse vastu laiemalt.

Tagasiside riskihindamise töölehele

Riskihindamise töölehte hinnati üldiselt kasulikuks struktuuriks riskide tuvastamisel ja arutelude algatamisel, eriti sisemise töövahendina. Osalejad rõhutasid, et tööleht toimib hästi esmase kaardistamise tööriistana, mitte lõpliku aruandlusdokumendina.

Töölehe kasutamisel tuvastati mitmeid konkreetseid kitsaskohti:

- Kasutusjuhtumitega seotud väljade arv tundus suur. Soovitati piirata algses etapis sisestatavate kasutusjuhtumite hulka, et muuta protsess varajases hindamisfaasis hoomatavamaks.
- Mõned riskiosaga seotud kvantitatiivsed küsimused, näiteks hinnang mõjutatud isikute arvule või riski realiseerumise kestusele, ei olnud avaliku sektori kontekstis alati sisukad. Soovitati need küsimused ümber sõnastada, paremini selgitada või asendada kvalitatiivsete hindamisvõimalustega.
- Jääkriski aktsepteerimise kontseptsioon osutus selles kontekstis problemaatiliseks. Osalejad leidsid, et teatud riskide, eriti valeinfo ja ebavõrdse kohtlemisega seotud riskide puhul, ei ole realistlik riskitaluvus. Soovitati selgemat juhust olukordade käsitlemiseks, kus kehtib nulltolerants.

Lisaks tehti ettepanek, et tööleht võiks selgemalt julgustada AI-põhiste abivahendite kasutamist riskide ja stsenaariumite kaardistamisel, eeldusel et tundlikku teavet ei jagata. Võimaliku parendusena nähti ka näidispäringute või vastutustundliku kasutamise juhiste lisamist.

Üldhinnang

Kokkuvõttes näitas Bürokrati kasutusjuhtumi valideerimine, et meetodika, juhendmaterjalid ja riskihindamise tööleht on kasulikud ja asjakohased tööriistad diskrimineerimise ja kallutatuse riskide hindamiseks avalikkusega suhtlevates vestlusrobotites. Tööriistad toetasid riskide süstemaatilist tuvastamist ning löid ühise raamistiku aruteludeks projekti- ja arendustiimide vahel.

Samas rõhutas valideerimine vajadust kohandada materjale kontekstidesse, kus:

1. teatud riske peetakse põhimõtteliselt vastuvõetamatuks;
2. süsteemid suhtlevad otseselt inimestega;
3. maine- ja usaldusmõjud võivad ulatuda kaugemale ühest organisatsioonist;
4. ning praktiline rakendatavus, selgus ja ajasäästlikkus on määrava tähtsusega.

5.3 Masinõppel põhinev analüütiline tööriist ettevõtete elujõulisuse hindamiseks (MKM–STAT)

MKM–STAT kasutusjuhtum andis ülevaate algoritmilise kallutatuse riskihalduse meetodika, juhendmaterjalide ja riskihindamise töölehe rakendamisesest masinõppel põhinevas analüütilises

süsteemis, mis pakub lõppkasutajatele tõenäosuslikke väljundeid. Kuigi süsteem ei tee automatiseeritud haldusotsuseid, võivad selle tulemused mõjutada ettevõtjate ärilisi otsuseid. See kaudne mõju muutis kasutusjuhtumi asjakohaseks tööriistade rakendatavuse valideerimisel ka väljaspool traditsioonilisi otsustuskontekste.

Materjalide rakendamine praktikas

Selle kasutusjuhtumi puhul algas valideerimisprotsess metoodikat tutvustavate sissejuhatavate materjalide ülevaatamisega, millele järgnes juhendmaterjalide sihipärane läbitöötamine. Arvestades, et osalejad olid süsteemi ja selle tehnilise kontekstiga hästi kursis, ei vaadatud juhendmaterjale enne töölehe täitmist detailsemalt läbi.

Riskihindamise töölehe täitis peamiselt süsteemi arendaja (Statistikaamet), kellel oli ligipääs vajalikele tehnilistele ja andmetega seotud üksikasjadele. Telliva asutusega (Majandus- ja Kommunikatsiooniministeerium) konsulteeriti eelkõige rollide ja vastutuste täpsustamiseks. Kuigi töölehe täitmist kirjeldati põhimõtteliselt koostööpõhise protsessina, tehti suurem osa tööst ühe isiku poolt. Kogu protsessile kulunud aeg oli mõõdukas.

See kogemus näitas, et töölehte on võimalik tõhusalt rakendada olukordades, kus tehniline kompetents on kiiresti kättesaadav, kuid tõi samas esile rollisulguse olulisuse, eriti mitme organisatsiooni koostöös arendatud süsteemide puhul.

Tagasiside juhendmaterjalidele ja metoodikale

Osalejad hindasid metoodika üldist lähenemist positiivselt, märkides, et see suunas süsteemsele ja struktureeritud riskide käsitlemisele ka sellistes süsteemides, kus diskrimineerimise ja kallutatuse riske ei pruugita esmapilgul teadvustada. Metoodikat peeti eriti väärtuslikuks selle poolest, et see suunas tähelepanu kallutatuse ja diskrimineerimise riskidele ka süsteemides, mis ei tooda otseseid haldusotsuseid.

Samas tuvastati mitmeid piiranguid:

Juhendmaterjale peeti liiga mahukateks kasutajatele, kes on tehnilise valdkonnaga juba tuttavad. Soovitati koostada lühem ja kokkuvõtlikum versioon, mis toimiks tõhusa sisenemispunktina, säilitades samas põhjalikuma juhendi viitena süvitsi mõistmiseks.

Osalejad märkisid, et eri andmetüüpide eristamine ei olnud piisavalt selge. Eriti oluliseks peeti selgemat vahet treeningandmete, testandmete ja operatiivses süsteemis kasutatavate sisendandmete vahel, et paremini mõista kallutatuse tekkimist ja mudelite toimivuse muutumist ajas.

Metoodikat tajuti eelkõige haldusotsuseid tegevatele või toetavatele süsteemidele suunatuna. Kuigi see ei takistanud selle rakendamist analüütilistes süsteemides, leiti, et täiendav juhendamine aitaks selgemalt mõista, kuidas hinnata kallutatuse ja diskrimineerimise riske olukordades, kus süsteemi väljundid on tõenäosuslikud ja mittesiduvad.

Tagasiside riskihindamise töölehele

Riskihindamise töölehte peeti üldiselt kasulikuks struktuuriks riskiteabe korrastamisel, kuid selle praktilisel kasutamisel ilmnis mitmeid kitsaskohti.

Osalejad märkisid, et Exceli-põhine vorming ei ole optimaalselt sobiv keerukamate süsteemide jaoks. Soovitati kaaluda veebipõhist tööriista, kus küsimused kohanduvad süsteemi tüübi järgi ning vähendavad vajadust tõlgendada, millised osad on konkreetsel juhul asjakohased.

Esitati mitu konkreetset parandusettepanekut:

- Tööleht peaks selgemalt eristama treeningandmeid ja operatiivseid sisendandmeid, võimaldades dokumenteerida nende vahelisi erinevusi ning hinnata andmetriivi ja ajas toimuvate muutustega seotud riske.
- Juhtudel, kus kvantitatiivseid andmeid ei ole võimalik või mõistlik kasutada, tuleks võimaldada kvalitatiivset riskihindamist, mitte nõuda arvulisi hinnanguid või protsente.
- Vajati selgemat juhust oodatava detailsuse taseme kohta, kuna ei olnud alati arusaadav, kui põhjalikult tuleks teavet dokumenteerida, arvestades, et tööleht on mõeldud asutusesiseseks kasutamiseks, mitte väliseks aruandluseks.
- Rollide määratlemist peeti ebaselgeks, eriti mitme organisatsiooni koostöös toimuvate projektide puhul. Selgemad juhised rollide ja vastutuse dokumenteerimiseks oleksid vajalikud järjepideva rakendamise tagamiseks.

Lisaks tehti ettepanek, et metoodika käsitleks süsteemsemalt kaitsemeetmeid, nagu kasutajatele mõeldud selgitused ja vastutust välistavad või piiravad teated (nt lahtiütlused). Selliste meetmete käsitlemine juhuslike lahendustena ei olnud osalejate hinnangul piisav ning nende dokumenteerimine ja hindamine võiks toimuda struktureeritumalt.

Üldhinnang

Kokkuvõttes näitas MKM–STAT kasutusjuhtumi valideerimine, et meetoodika, juhendmaterjalid ja riskihindamise tööleht on rakendatavad ka väljaspool traditsioonilisi ADM-süsteeme, sealhulgas analüütilistes tööriistades, mis pakuvad tõenäosuslikke väljundeid ja otsustustoetavat teavet. Materjalid toetasid struktureeritud arutelu kallutatuse ja diskrimineerimise riskide üle ka olukordades, kus riskid on kaudsed ja vahendatud kasutajate tõlgenduse kaudu.

Samas tõi valideerimine esile vajaduse:

1. selgema eristuse järele eri andmetüüpide vahel;
2. kvalitatiivse riskihindamise parema toetamise järele;
3. täpsema juhendamise järele analüütiliselt keerukate ja tõenäosuslike süsteemide puhul;
4. ning paindlikumate töövahendite järele, mis on kohandatud erinevatele süsteemikontekstidele.

6. Valideerimise peamised järeldused

Metoodika, juhendmaterjalide ja riskihindamise töölehe valideerimine kolme erineva AI kasutusjuhtumi põhjal tõi esile mitmeid korduvaid teemasid ja üldisi järeldusi. Kuigi hinnatud süsteemid erinesid oma eesmärgi, tehnilise keerukuse ja kasutajatega suhtlemise viisi poolest, ilmnisid mitmed ühised mustrid seoses tööriistade kasutatavuse, organisatsiooniliste protsesside ning diskrimineerimise ja kallutatuse riskide olemusega avaliku sektori AI-süsteemides.

Rakendatavus erinevat tüüpi AI-süsteemides

Üheks valideerimise keskseks järelduseks oli, et loodud tööriistad on rakendatavad laias spektris AI-süsteemide tüüpides. Ka juhtudel, kus süsteemid ei teosta automatiseeritud otsustamist, toetasid metoodika ja riskihindamise tööleht sisukat arutelu diskrimineerimise ja kallutatuse riskide üle. See kinnitas, et riskihindamise tööriistad ei peaks piirduma üksnes täielikult automatiseeritud otsustega, vaid peaksid hõlmama ka süsteeme, mis mõjutavad tulemusi kaudselt teabe edastamise või otsustustoena.

Samas näitas valideerimine, et eri tüüpi süsteemid nõuavad hindamisel erinevaid rõhuasetusi. Näiteks töid avalikkusega suhtlevad süsteemid esile kommunikatsiooni ja valeinfo riskid, samas kui analüütilised süsteemid tõstatasid väljakutseid seoses andmete tõlgendamise, ebakindluse ning kaudse mõjuga.

Struktureeritud riskihindamise väärtus

Kõigi kasutusjuhtumite puhul märkisid osalejad, et metoodika ja töölehe struktureeritud lähenemine aitas esile tuua riske, mida ei olnud varem selgesõnaliselt käsitletud. Isegi olukordades, kus meeskonnad olid teatud probleemidest teadlikud, toetas hindamisprotsess nende selgemat sõnastamist, dokumenteerimist ja arutelu.

Riskihindamise tööleht toimis eriti hästi sisemiste arutelude lähtepunktina, pakkudes ühist raamistikku tehniliste, juriidiliste ja organisatsiooniliste vaatenurkade ühendamiseks. See oli eriti väärtuslik kontekstides, kus vastutus AI-süsteemide eest on jaotatud mitme rolli või organisatsiooni vahel.

Samas näitas valideerimine, et tööleht on kõige tõhusam siis, kui seda kasutatakse elava dokumendina, mitte ühekordse aruandlusvahendina. Osalejad rõhutasid vajadust ajakohastada hinnanguid süsteemide arenedes, eriti arendus- ja rakendusfaasis.

Ajastus ja lõimimine arendusprotsessidega

Kõiki kasutusjuhtumeid läbivaks teemaks oli ajastuse olulisus. Valideerimine näitas, et tööriistad on kõige kasulikamad siis, kui neid rakendatakse pärast peamiste kasutusjuhtumite ja süsteemi eesmärkide määratlemist, kuid enne süsteemide täielikku kasutuselevõttu või olukorda, kus muudatused muutuvad keeruliseks.

Liiga varajane rakendamine, kui süsteemi disain on veel väga muutuv, tekitab ebakindlust selle osas, mida tuleks dokumenteerida. Liiga hiline rakendamine piirab tööriistade mõju disainivalikutele. See rõhutab vajadust selgema juhise järel, kuidas tööriistu kasutada iteratiivselt erinevates süsteemiarenduse etappides.

Osalejad märkisid ka, et esmane täitmine toimub sageli ühe rolli poolt (nt projektijuht või arendaja), samas kui laiem kaasamine on vajalik hilisemas etapis riskide leevendamise otsuste kokkuleppimiseks.

Materjalide kasutusmugavus ja selgus

Kuigi üldist lähenemist hinnati positiivselt, tuvastas valideerimine mitmeid kasutusmugavusega seotud väljakutseid, mis olid ühised kõigi kasutusjuhtumite puhul:

- juhendmaterjalide maht ja keerukus muutsid need vähem sobivaks kiireks sissejuhatuseks, eriti ajasurve tingimustes;
- terminoloogiat ei kasutatud dokumentides alati järjepidevalt, mis tekitas aeg-ajalt segadust;
- mõned töölehe väljad olid ebapiisavalt selgitatud, mistõttu jäi ebaselgeks oodatav detailsuse tase või see, kuidas kogutud teavet hiljem kasutatakse.

Need tähelepanekud viitavad sellele, et tööriistade tõhusus sõltub lisaks kontseptuaalsele tugevusele ka kasutajakesksest disainist ning dokumentide omavahelisest kooskõlast.

Andmete ja süsteemide keerukusega seotud väljakutsed

Andmetega seotud probleemid kujunesid oluliseks teemaks, eriti analüütiliselt keerukate süsteemide puhul. Osalejad rõhutasid vajadust:

- eristada selgemalt treeningandmeid, testandmeid ja operatiivseid sisendandmeid;
- toetada andmete ja mudelite käitumise muutuste (nt andmetriivi) dokumenteerimist ajas;
- võimaldada kvalitatiivset riskihindamist olukordades, kus kvantitatiivsed andmed ei ole kättesaadavad või asjakohased.

Mitut mudelit või komponenti hõlmavate süsteemide puhul vähendas sellise keerukuse ebapiisav käsitlemine hindamise selgust ning suurendas mittetäieliku dokumenteerimise riski.

Riskitaluvus ja jääkrisk

Valideerimine näitas, et riskitaluvuse käsitus erineb sõltuvalt süsteemi tüübist. Mõnes kontekstis, eriti avalikkusega suhtlevates süsteemides, peeti teatud riske põhimõtteliselt vastuvõetamatuks, mis muutis tavapärase jääkriski hindamise ja aktsepteerimise raamistikud raskesti rakendatavaks.

See tõi esile vajaduse selgema juhise järele kuidas käsitleda nulltolerantsiga riske; kuidas dokumenteerida riskide leevendamist olukordades, kus aktsepteerimine ei ole võimalik; ning kuidas tõhusalt kommuniqueerida realiseerunud riske ja rakendatud leevendusmeetmeid.

Organisatsioonilised ja juhtimisalased kaalutlused

Lõpetuseks rõhutas valideerimine, et diskrimineerimise ja kallutatuse riskide juhtimine ei ole üksnes tehniline ülesanne. Organisatsioonilised tegurid, nagu rollide selgus, meeskondade vaheline koostöö ja juhtimisstruktuurid, mängivad kesksel rollil riskide tõhusal tuvastamisel ja käsitlemisel.

Osalejad tõid esile vajaduse kaasata asjakohane pädevus, sealhulgas tehnilised, juriidilised ja küberturbe spetsialistid, tunnistades samal ajal praktilisi piiranguid seoses ajakulu ja töökoormusega. Tööriistu peeti kõige tõhusamaks siis, kui neid kasutatakse arutelu toetamiseks, mitte selle asendamiseks.

6.1 Valideerimistulemuste kasutamine tööriistade lõplikul kujundamisel

Valideerimistegevuse tulemusi kasutati otseselt algoritmilise kallutatuse riskihalduse meetodika, juhendmaterjalide ja riskihindamise töölehe täiendamiseks ja lõplikuks viimistlemiseks. Hinnatud kasutusjuhtumitest saadud tagasiside mõjutas nii materjalide sisulisi muudatusi kui ka struktuurseid parendusi, mille eesmärk oli parandada nende kasutusmugavust ja praktilist rakendatavust.

Valideerimine kinnitas loodud tööriistade üldist asjakohasust, tuues samal ajal esile konkreetsed valdkonnad, kus oli vajalik täpsustamine, lihtsustamine või täiendav juhendamine.

7. Kokkuvõte ja järeldused

Käesolev raport esitas valideerimistegevuse tulemused, mille eesmärk oli testida algoritmilise kallutatuse riskihalduse meetodika, juhendmaterjalide ja riskihindamise töölehe praktilist rakendatavust avaliku sektori AI- ja ADM-lähedastes süsteemides.

Valideerimine viidi läbi kvalitatiivsete intervjuude kaudu avaliku sektori organisatsioonidega, kes rakendasid loodud materjale arenduses või kasutuselevõtu faasis olevate AI-lahenduste puhul. Kuigi hinnatud süsteemid erinesid oma eesmärkide, tehnilise keerukuse ja automatiseerituse taseme poolest, pakkusid need mitmekesist ja realistlikku alust tööriistade praktilise toimimise hindamiseks.

Tulemused näitavad, et meetodika, juhendmaterjalid ja tööleht on asjakohased ja kasulikud erinevat tüüpi AI-süsteemide puhul, sealhulgas otsustustoetussüsteemides, avalikkusega suhtlevates vestlusrobotites ning analüütilistes süsteemides, mis pakuvad tõenäosuslikku teavet. Ka täielikult automatiseeritud otsuste puudumisel toetasid tööriistad süsteemset arutelu diskrimineerimise ja kallutatuse riskide üle, mis võivad tekkida kaudselt teabe edastamise, andmepõhise prioriseerimise või inimeste poolse süsteemiväljundite tõlgendamise kaudu.

Kõigi kasutusjuhtumite puhul kinnitas valideerimine struktureeritud ja selgesõnalise riskihindamise protsessi väärtust. Tööriistad aitasid esile tuua riske, mida ei olnud varem formaalselt sõnastatud, toetasid eelduste ja leevendusmeetmete paremat dokumenteerimist ning hõlbustasid arutelusid eri rollide ja distsipliinide vahel. Samal ajal tõi valideerimine esile ka praktilisi piiranguid, mis on seotud ajakulu, süsteemide arenguetapi ja organisatsioonilise suutlikkusega, rõhutades vajadust paindliku ja proportsionaalse rakendamise järele.

Valideerimine tuvastas ka konkreetseid parendusvajadusi, eelkõige seoses kasutusmugavuse, terminoloogia ühtlustamise ning juhiste täpsustamisega keerukate ja muutuvate süsteemide puhul. Need tähelepanekud kasutati otseselt meetodika, juhendmaterjalide ja töölehe täiendamiseks ja lõplikuks viimistlemiseks, et tagada nende parem sobivus tegelike avaliku sektori praktikate ja piirangutega.

Kokkuvõttes kinnitab valideerimine, et diskrimineerimise ja kallutatuse riskide tõhus juhtimine avaliku sektori AI-süsteemides eeldab lisaks tehnilistele meetmetele ka selgeid protsesse, jagatud arusaama ning organisatsioonilist pühendumust. Käesolevas projektis valideeritud tööriistad pakuvad praktilist alust nende eesmärkide saavutamiseks. AI- ja ADM-süsteemide edasise leviku korral avalikus sektoris on jätkuv testimine, täiendamine ja institutsionaalne

õppimine vältimatult vajalik, et tagada nende süsteemide õiglane, läbipaistev ja vastutustundlik arendamine ning kasutamine.