Lisa 1.

Valimislubaduste AI-tööriista eksperimentaal-arendustöö lähteülesanne

1. Taust ja vajadus

Rahandusministeerium on koostanud valimisplatvormide eelarvemõju esmahinnanguid alates 2007. aastast ja plaanib seda teha ka 2027. aasta riigikogu valimiste eel. Analüüs annab avalikkusele poliitiliselt erapooletu ülevaate valimislubaduste rahalisest mõjust ning toetab teadlikke valikuid ja sisukamat avalikku arutelu. Pärast valimisi aitab see erakondadel ja ametkondadel hinnata koalitsioonileppe tegevuste maksumust ning toetab valitsuse tööplaani ja riigieelarve koostamist.

Seni on sellist analüüsi tehtud käsitsi, sest varasemad tehnilised lahendused ei suutnud piisavalt hästi omavahel seostada ja analüüsida eri vormingutes valimisplatvorme, Exceli dokumentides olevaid eelarvetabeleid ning Wordi- ja Pdf-dokumentides olevat infot. Tehisintellekti kiire areng loob uusi võimalusi eri vormingutes olevate dokumentide seostamiseks ja analüüsi toetamiseks, kuid nende võimaluste rakendatavus avaliku sektori analüütilistes tööprotsessides vajab täiendavat katsetamist ja hindamist. Seetõttu on vajadus uurida ja katsetada, kas ja kuidas oleks võimalik luua ühtne ning AI-sõbralik andme- ja päringulahendus, mis muudab valimislubadused, avalikud arengudokumendid ja osa ministeeriumi varasemast analüütilisest teadmusest paremini leitavaks, võrreldavaks ja taaskasutatavaks.

Praegu paikneb vastav teadmus eri failides, tabelites ja kaustastruktuurides, mistõttu selle kasutamine on ajamahukas, ebaühtlane ning sõltub suurel määral üksikute inimeste teadmistest ja käsitööst. Samas on varasemate valimislubaduste analüüsi kaudu süstematiseeritud kujul talletatud hulk metainfot selle kohta, kuidas valimisprogrammides kirjas olevat tekstilist infot seostada analüütiliste järelduste, eelarveinfo ning arengukavadega.

2. Projekti eesmärk ja tellitava töö sisu

Projekti eesmärk on luua korduvkasutatav lahendus, mis koondab valimislubadused, eelarvedokumendid, arengukavad ja fiskaalpoliitilised analüüsid ühtsesse masinloetavasse süsteemi ning võimaldab nii Rahandusministeeriumi analüütikutel kui ka avalikkusel teha selles süsteemis päringuid ning saada allikaviidetega vastuseid.

Projekt peab lahendama kolm peamist probleemi:

Esiteks tuleb vähendada analüütikute ajakulu varasemate hinnangute, dokumentide ja eelarveandmete leidmisel. Praegune tööprotsess eeldab mitmete andmeallikate käsitsi läbivaatamist ning puudub ühtne otsingumehhanism, mille kaudu seostada valimislubadus või küsimus infoga, mis juba varasemates dokumentides eksisteerib.

Teiseks tuleb suurendada valimislubaduste ja nende eelarvemõjude analüüside läbipaistvust. Avalikkus, meedia ja teadlased peavad saama kasutada avalikku päringusüsteemi, et saada

samasugune nii veebipõhine kui masinloetav ülevaade valimislubadustest ning lubaduste seostest süsteemi sisestatud avalikuks kasutamiseks mõeldud dokumentidega.

Kolmandaks tuleb luua tehniline alus AI-põhiseks teadmiste kasutamiseks viisil, kus mudeli vastused põhinevad alati tuvastatavatel allikatel, mitte üksnes keelemudeli üldteadmistel.

Tegemist on eksperimentaal-arendustööga, mille olemus eeldab teadmiste loomist, erinevate lähenemiste katsetamist ning tulemuste põhjal tehtavaid valikuid. Seetõttu ei kirjelda tellija ette üksikasjalikult eksperimentaal-arendustegevuse metoodikat ega tegevuskava, vaid jätab taotlejale piisava autonoomia valida eesmärgi saavutamiseks sobivaimad uurimis- ja arendusmeetodid. Taotluses esitatav tervikkontseptsioon ja tegevuskava on lähtepunkt töö elluviimiseks, kuid neid võib projekti käigus põhjendatult täiendada või kohandada vastavalt katsete tulemustele, ilmnevatele piirangutele ning uutele teadmistele. Selline lähenemine võimaldab rakendada teaduslikku loogikat ning suurendab tõenäosust leida probleemi lahendamiseks kõige toimivam ja tõenduspõhisem lahendus.

Tellija ootused lahendusele on järgmised:

2.1. Töövõtja aitab Rahandusministeeriumil **luua automaatse töövoo**, mille abil saavad ametnikud indekseerida valimislubaduste analüüsimiseks vajalikud eestikeelsed dokumendid *retrieval-augmented generation* (RAG) töövoo sõbralikule kujule. Meetodi (skriptipõhine, AI-agendid või midagi muud) selliseks töövooks pakub välja töövõtja, kuid oluline on, et dokumentide sisestamine oleks ametnikele lihtne ja toimuks näiteks veebipõhise kasutajaliidese või muu sarnase lahenduse kaudu. Koos dokumendi üleslaadimisega sisestab ametnik vajadusel ka muu vajaliku metainfo dokumendi kohta.

2.2. Töövõtja **loob: 1) andmebaasi**, kus see indekseeritud info asub; 2) **otsinguserveri andmete otsimiseks;** 3) **liidesed sellest andmebaasist info otsimiseks**. Kasutajaliideseid peaks olema otsinguserveril kaks – nii veebipõhine märksõnaotsing dokumentidest kui masinloetav API põhine otsingusüsteem, mis on tehisintellekti päringuteks ühendatav kasutaja poolt valitava AI pakkujaga. Märksõnaline otsing peab leidma sisestatud dokumentidest üles tekstiosad, kus vastavast teemast on juttu ning andma nii dokumendi sisulõigud kui ka konkreetsed viited ja vajaliku metainfo (peatüki, lõigu või lehekülje täpsusega), millises dokumendis ja kus täpselt otsitud teemat käsitleti.

2.3. Dokumentide indekseerimine ja nende otsing andmebaasist peab olema **optimeeritud eestikeelsete tekstide leidmiseks ja tooma otsingu tulemusena kasutajale mitte üksnes märksõna vasted, vaid sisulised temaatilised vasted**. Seetõttu on tellitava töö oluline osa eesti keele eripäradele suunatud eksperimentaalarendus, kus töövõtja peab standardlahenduse rakendamise asemel katsetama eesti keelele kohandatud võtteid, näiteks lemmatiseerimist ja sünonüümide käsitlemist, ning valima parima lähenemise dokumentide indekseerimiseks ja otsingu realiseerimiseks empiiriliste tulemuste põhjal. Töövõtja peab ka lõppraportis kirjeldama, kuidas lahendus arvestab eesti keele terminoloogilisi erinevusi, sünonüüme, liitsõnu ja sõnavorme, ning valideerima seotud dokumentide leidmise ka ühiste märksõnade puudumisel.

2.4. Otsinguserveril peab olema kaks vaadet: **avalik vaade** ja **ametkonnale suunatud piiratud vaade**. Avaliku vaadet võivad valimislubaduste, eelarvedokumentide ja arengukavade omavaheliseks seostamiseks nii veebiotsingus kui AI-assistentides kasutada kõik soovijad. Ametkondlik vaade on mõeldud ametnikele ning see otsing tagastab lisaks avaliku vaate tulemustele ka süsteemi sisestatud kuid mitteavalikuks kasutamiseks märgitud sisemised dokumendid. Tellija esialgsel hinnangul isikuandmeid ning avaliku teabe seaduse mõttes ametkondlikuks kasutamiseks mõeldud dokumente selles süsteemis ei käidelda, kuid enne töö tellijale üleandmist peab töövõtja teostama andmekaitse analüüsi. Autoriseerimiseks tuleb kasutada riigi TARA autentimislahendust. Töövõtja peab ette nägema avalikule otsinguserverile *throttle* võimaluse, et vajaduse tekkimisel hajutada serveri koormust (prioritiseerides *throttle* rakendamisel autoriseeritud kasutajate ehk ametnike päringuid).

2.5. Iga andmebaasis oleva dokumendi juurde kuulub täiendav metainfo (selle sisestab ametnik dokumendi sisestamisel) ning kasutajatel peab olema võimalik otsingu või päringute tegemisel **selgelt sisse ja välja lülitada, millistest dokumentidest otsinguvasteid antakse**.

2.6. Indekseerimine **ei tohi algdokumente muuta**. Samuti on soovitatav ehitada süsteemi sisse võimalus, et indekseerimisalgoritmi hilisemal täiendamisel või parandamisel on kasutajal võimalik ka **hiljem ümber indekseerida juba andmebaasi sisestatud dokumente**.

2.7. Töövõtja võib lahenduse realiseerimiseks vajaliku tehnilise lahenduse ja tarkvara valida oma äranägemisel, kuid eelistades võimalusel **vabavaralist ja üldlevinud tarkvara**.

2.8. Lahendus antakse üle Kubernetesel käivitatava konteinerlahendusena. Tellijale antakse üle lahenduse lähtekood, konfiguratsioonifailid, juurutusjuhised ja tehniline dokumentatsioon, sealhulgas konteineris sisalduvate töövoogude, otsinguserveri, eesti keelele suunatud eksperimentaalarenduse komponentide ning lahenduses kasutatavate AI-agentide lähtekood, konfiguratsioonid, töövood, süsteemijuhised, promptid ja muu tehniline dokumentatsioon. Samuti antakse üle kõik lahenduse käitamiseks, taasesitamiseks, valideerimiseks, hooldamiseks ja edasiarendamiseks vajalikud artefaktid. Dokumentatsioon peab võimaldama tellijal lahendust iseseisvalt käitada, hallata ja edasi arendada ilma töövõtja täiendava osaluseta. **Tehnilise lahenduse valik ja kavandatav arhitektuur tuleb enne rakendamist kooskõlastada ning valideerida koostöös Rahandusministeeriumi Infotehnoloogiakeskusega (RmIT)**.

2.9. Töövõtja koostab **AI agendi või AI agentide süsteemi**, mis koostab lahendusse sisestatud valimisprogrammi põhjal automaatselt eeltäidetud valimislubaduste analüüsi tabeli vastavalt tellija poolt ette antud kujule.

2.10. Töövõtja aitab tellijal seadistada vajalikud **MS365 Copilot pluginad** ja **agendid** liidestamaks ametnike tööarvutid lahenduse masinotsinguga, et ametnikud saaksid kasutada süsteemi sisestatud infot otsinguserveri kaudu Copiloti AI lahendustes.

2.11. Enne lahenduse lõplikku üleandmist koostab töövõtja tellijale täidetud „Algoritmi kasutuse ülevaate“ vormi või selle vormi kõikidele väljadele vastava samaväärse dokumentatsiooni[1].

2.12. Töövõtja peab lahenduse kavandamisel, arendamisel, juurutamisel ja üleandmisel lähtuma RmITi üldistest ristfunktsionaalsetest, mittefunktsionaalsetest ja tehnilistest nõuetest (lisa 7 RFN-2.9.0 ning RmIT IT-profiilist (lisa 8 IT-profiil-2.9.0).

Täpsem kasutatavate andmeallikate ja sisestatavate dokumentide ülevaade on toodud tehnilises lisas (vt lisa 2).

3. Eksperimentaal-arendustöö peab vastama teadus-arendustegevuse tunnustele:

Uudne. Tellitav teenus on uudne. Teadaolevalt oleks tegemist Eesti avalikus sektoris uudse ja rahvusvaheliselt eristuva lahendusega, mis teeb süstematiseeritud valimislubaduste andmestiku kättesaadavaks nii avaliku päringusüsteemi kui ka AI-agentide poolt kasutatava API kaudu. Uudsus seisneb aga ka selles, et lahenduse väljatöötamine eeldab eesti keele eripäradele suunatud eksperimentaalarendust. Standardsete otsingu- ja indekseerimislahenduste rakendamisest ei piisa, kuna eesti keel on rikkaliku morfoloogia, liitsõnade ja terminoloogiliste variatsioonidega keel. Seetõttu peab eksperimentaal-arendustöö käigus katsetama ja võrdlema erinevaid eesti keelele kohandatud lähenemisi, sealhulgas lemmatiseerimist, sünonüümide käsitlemist ja muid semantilise otsingu võtteid, ning valima dokumentide indekseerimiseks ja otsingu realiseerimiseks sobivaima lahenduse empiiriliste tulemuste põhjal.

Loominguline. Eksperimentaalarenduse läbiviijale jäetakse vabadus täpsustada lahendatavaid küsimusi, meetodeid ja allikaid lähtuvalt nende nägemusest selle kohta, kuidas on kõige mõistlikum püstitatud ülesannet täita. Tellija näeb ette, et nimetatud valikuid ja püsitatud küsimusi saab täpsustada ka eksperimentaalarenduse käigus, lähtudes lisanduvatest teadmisest või piirangutest valdkonna kohta.

Ettemääramatu tulemusega. Tellija ei määra täielikult ette, millistele teoreetilistele lahendustele või millistele praktikatele töö teostaja peab keskenduma. Tellija ei määra ette, milline on töö lõplik väljund, parima lahenduseni jõutakse uuringuprotsessi jooksul valideerides vahetulemeid. Kuna selliseid lahendusi pole varasemalt kusagil kasutatud, valitseb ettemääramatus, kas lahenduse väljatöötamine õnnestub.

Süstemaatiline. Teenus on süstemaatiline, tegevus dokumenteeritakse ning arendustööde kood ja tulemused (sh funktsionaalne rakendus) antakse tellijale üle. Teekond valimislubaduste analüüsi lahenduste ja töövahendite kujundamiseni peab koosnema põhjendatud valikutest, mis algavad teoreetiliste raamistike ja praktikate valikust, mis eksperimentaal-arendustöösse kaasatakse. Muudatused lahendatavates küsimustes ja metodoloogias arutatakse läbi töö teostaja ja Rahandusministeeriumi esindajatega.

[1] AI ülevaade | Kratid

Eksperimentaal-arendus lõpeb reflektsiooniga selle kohta, millised on välja töötatud lahenduse piirangud ning edasi arendamise võimalused.

Samuti peab taotleja kirjeldama, kuidas lahendus arvestab terminoloogilisi erinevusi, sünonüüme, liitsõnu ja sõnavorme ning valideerima, et süsteem suudab leida sisuliselt seotud dokumente ka olukorras, kus dokumentides puuduvad ühised märksõnad või identsed sõnavormid. Selline eesti keele eripäradest lähtuv eksperimentaalarendus on tellitava töö keskne osa ning loob eeldused lahenduse usaldusväärseks kasutamiseks Eesti avalikus sektoris.

Ülekantav ja/või korratav. Kõik metodoloogilised sammud peavad olema põhjendatud ja kirjeldatud lõppraportis ning välja peavad olema toodud eksperimentaal-arendustöö ja lahenduse piirangud. See muudab uuringu korratavaks. Tellitava töö põhisisu on korduvkasutatav tehniline lahendus: andmemudel, API, otsingu- ja seostusloogika ning haldus- või toimetusvaade, mille abil saab tellija hiljem ise uusi dokumente ja kirjeid jooksvalt lisada, parandada ja uuesti otsingukõlblikuks teha. Ülekantavust toetab ka eksperimentaalarenduse käigus välja töötatav lähenemine eesti keele eripärade käsitlemiseks. Taotleja peab kirjeldama, kuidas lahendus arvestab terminoloogilisi erinevusi, sünonüüme, liitsõnu ja sõnavorme ning valideerima, et süsteem suudab leida sisuliselt seotud dokumente ka olukorras, kus dokumentides puuduvad ühised märksõnad või identsed sõnavormid. Selliste lahenduste dokumenteerimine ja valideerimine võimaldab sama metoodikat ja tehnilisi põhimõtteid rakendada ka teistes eestikeelsetes dokumendikogudes ning avaliku sektori andmestikes. Seega on tellitav teenus ülekantav ja korratav ning tagatud on teadmiste ja tehniliste lahenduste siirdamise võimalus teistesse valdkondadesse.

Eksperimentaal-arendustööst tulenev kasu ja avalikustamine. Eksperimentaal-arendustöö tulemusel valmiv lahendus loob enamat kui valimislubaduste analüüsi toetava töövahendi. Projekt annab praktilise teadmise, kuidas kujundada avalikus sektoris tehisintellektile sobivaid teadmuskihte viisil, mis säilitab allikapõhisuse, läbipaistvuse ja kontrollitavuse. Erilist väärtust loob eestikeelse otsingu kvaliteedi parandamine, dokumentide masintöödeldavaks muutmine ning mehhanismide loomine, mis võimaldavad siduda AI-rakenduste vastused kontrollitavate allikatega. Projekti käigus loodud andmemudelid, sissevõtuprotsessid, õiguste halduse põhimõtted ning allikaviidetega kontekstipakettide loomise metoodika on rakendatavad ka teistes poliitikavaldkondades ja avaliku sektori asutustes. Seetõttu aitab projekt kaasa laiemalt teadmistepõhise poliitikakujundamise võimekuse arengule ning loob aluse usaldusväärsete ja eestikeelsete AI-lahenduste kasutamiseks avalikus sektoris.

Projekti tulemid ja info kajastab töövõtja Eesti Teadusinfosüsteemis. Rahandusministeerium kajastab töö tulemeid ka oma veebilehel. Tulemeid jagatakse kõigile huvilistele arvestades andmekaitsereeglitega.

4. Nägemus meeskonnast ja ootused koostööst.

Alljärgnevalt on esitatud esialgne nägemus soovituslikust koostööformaadist (taotleja võib esitada oma nägemuse töökorraldusest). Tellija hinnangul peaks projekti kestel moodustama fiskaalpoliitika osakonna analüüsitiimist ja töövõtja meeskonda kuuluvatest ekspertidest ühine meeskond.

Taotleja meeskonnas peab olema:

- projektijuht;
- kogenud AI ekspert, kes suudab nõustada lahenduse arhitektuuri, mudelite valiku ja juurutamise osas;
- täiendava eelise annab eestikeelsete keelemudelite või eestikeelse keeletehnoloogia rakendamise kogemusega eksperdi kaasamine meeskonda (vt hindamismetoodika).

Tegemist on koostööga, mille keskmes on regulaarselt toimuvad töökoosolekud - analüütilised seminarid/arutelud töövõtja ning ministeeriumi analüütikute vahel. Töökoosolekuid viiakse läbi regulaarselt vastavalt vajadusele, aga mitte harvem kui 2 nädala tagant. Töökoosolekutel antakse ülevaade läbi viidud ja lõpetatud tegevustest, lahendatud probleemistikest, viiakse sellele tuginedes läbi arutelu ning püstitatakse täpsem ülesanne järgmiseks töökoosolekuks. Töökoosolekute täpne sisu ja ajakava lepitakse kokku projekti jooksul.

Töö olulistes vaheetappides viiakse läbi valideerimisseminare, kuhu kaasatakse asjasse puutuvad osapooled (sh tehniliste lahenduste valikul tuleb kaasata RMITi).

Lepingu täitmise ajaks soovib Rahandusministeerium kujundada püsiva nõustamise koostööformaadi, mis toetab tulemite kasutamist poliitikakujundamises.

Nõustamise eesmärk on:

- tagada poliitikakujundajatele õigeaegne ligipääs eksperimentaal-arenduse tulemustele;
- suurendada vastastikust mõistmist töövõtja ja poliitikakujundajate vahel;
- täpsustada, millised teadmised on praktiliselt rakendatavad ning kuidas neid igapäevases töös kasutada;
- toetada eksperimentaal-arenduse tulemuste rakendamist poliitikakujundamises, sh muuta tulemused selgemaks, sõnastatumaks ja kasutatavaks;
- laiendada teadmist kasutavate sihtrühmade ringi, sh ministeeriumide ja teiste avaliku sektori asutuste üleselt.

Eksperimentaal-arendustöö käigus on soovitatav korraldada ka avalikke seminare ja arutelusid. Lisaks tuleb projekti lõpus korraldada lõpuseminar, mis keskendub valminud lahenduse tutvustamisele. Eksperimentaal-arendustöö peab olema kajastatud Eesti Teadusinfosüsteemis.