



TILMÆLI um notkun gervigreindar hjá Reykjavíkurborg

Nýting gagna með aðstoð gervigreindar gerir borginni kleift að veita þjónustu með skilvirkari hætti og bæta gæði í þjónustu. Ýmis siðferðileg álitamál snúa að nýtingu gervigreindar og því hafa eftirfarandi tilmæli verið gefin út.

Tilmælin byggja á eftirfarandi viðmiðum sem byggð eru á þjónustustefnu borgarinnar og skiptast í þrjá meginflokka:

Fagmennska

1. Skiljanlegt
Notendur fá skýrar og skiljanlegar upplýsingar um vinnslu þar sem gervigreind er nýtt.
2. Gagnsætt
Notendum er ávallt ljóst hvar gervigreind er nýtt í starfi borgarinnar og með hvaða hætti það er gert.
3. Örugg
Tryggt er að kerfi sem nýta gervigreind eru vel varin og undir eftirliti.
4. Ábyrgðaraðili
Öll þjónusta sem nýtir gervigreind þarf að hafa tilgreindan ábyrgðaraðila sem auðvelt er að hafa samband við.

Notendamiðuð þjónusta

5. Gott fyrir fólk
Nýting gervigreindar þarf að leiða til bættrar þjónustu við notendur.
6. Fyrir öll
Tryggja þarf að gögn sem gervigreind nýtir endurspegli fjölbreytileika notenda.
7. Réttlát
Tryggt er að þjónusta sem nýtir gervigreind sé unnin af heilindum og taki mið af réttindum notenda.
8. Stjórnað af fólki
Ábyrgðaraðilar þjónustu sem nýtir gervigreind skal fylgjast með, grípa inní og bregðast við ef þess þarf.

Skilvirkni

9. Skilvirk og mælanleg
Gervigreind er einungis nýtt þegar ljóst er að þjónusta verði betri með notkuninni. Gæði þjónustu sem nýtir gervigreind skal vera mælanleg.

Tilmæli vegna notkunar gervigreindar

Vandamál má leysa á ólíka vegu og því þarf að taka upplýstar ákvarðanir um hvaða verkfæri henta best til að leysa þau vandamál sem borgin stendur frammi fyrir. Gervigreind er ein af þeim verkfærum og skal hún einungis nýtt þegar ljóst er að þjónusta verði betri með notkuninni. Markmið með tilmælum þessum er að tryggja að notkun gervigreindar sé í takt við verkefnaáherslur, öryggiskröfur og siðferðileg gildi Reykjavíkurborgar.

Tilmæli þessi leggja áherslu á þau atriði sem þarf að hafa í huga þegar tekin er ákvörðun um notkun gervigreindartóla. Sérstaklega skiptir máli að meta hvaða gögn og töl á að vinna með og hvaða verkefni eða úrlausn vonast er til að leysa. Við þessa greiningu er stuðst við gervigreindarreglugerð Evrópusambandsins (reglugerð ESB 2024/1689, e. AI Act), sem skilgreinir áhættuflokka gervigreindartóla, og öryggisflokkun gagna íslenska ríkisins.



Þjónustu- og nýsköpunarsvið (PON) framkvæmir gangsæismat gervigreindarlausna, aðstoðar við innleiðingu og veitir ráðgjöf í tengslum við tilmæli þessi.

Hvert svið innan borgarinnar ber ábyrgð á að taka afstöðu til notkunar gervigreindar í samræmi við faglegar áherslur og verkefni sviðsins. Þessi afstaða ætti að byggja á stefnu borgarinnar í stafrænum málum og tilmælum þessum.

Fyrirvarar um notkun gervigreindar

Til þess að tryggja ábyrga og örugga notkun gervigreindar í opinberri þjónustu skal hafa í huga að:

- Persónuvernd og öryggi gagna verða að vera í forgrunni.
- Gögn og reiknirit þurfa að vera rýnd til að tryggja réttmæti niðurstaðna.
- Gagnsæismat er nauðsynlegt við innleiðingu gervigreindar, þ.e. mat á því hvort og hvernig hægt er að skilja virkni gervigreindar, með áherslu á gagnavinnslu, öryggi, rekjanleika og ákvörðunartöku.
- Notendur þurfa að vita hvenær þeir eiga í samskiptum við gervigreind.
- Gervigreind tekur aldrei ákvarðanir er snerta réttindi og skyldur borgaranna.

Áhættuflokkun notkunar gervigreindar

Notandi þarf að þekkja þá áhættu sem fylgir notkun gervigreindar sem og áhættuflokkun þeirra gagna sem hann vinnur með.

Reykjavíkurborg skilgreinir fjóra áhættuflokka í notkun gervigreindartóla og notar til samræmis öryggisflokkun gagna íslenska ríkisins.

Mikilvægt er að notendur kynni sér öryggisflokkun gagna, sem finna má í töflunni hér að neðan:

Skilgreining öryggisflokunar gagna íslenska ríkisins

Flokkur 1 Opin gögn	Ópersónugreinanleg gögn eða gögn sem eru opin og aðgengileg til notkunar og endurnotkunar. Svo gögn teljist opin þurfa þau að vera tiltæk án umsókna / beiðna og vera aðgengileg óháð tíma.
Flokkur 2 Varin gögn	Öll gögn önnur en opin gögn sem eru hluti af daglegum rekstri ríkisaðila. Varin gögn geta þó verið misviðkvæm og krafist sérsníðinna öryggisúrræða í samræmi við niðurstöður áhættumats.
Flokkur 3 Sérvarin gögn	Gögn sem vegna viðkvæmrar stöðu m.t.t. tímasetninga eða innihalds geta valdið víðtæku og langvarandi tjóni fyrir hópa einstaklinga, lögaðila eða ríkisaðila.
Flokkur 4 Afmörkuð gögn	Gögn sem eru viðkvæm fyrir samfélagið í heild eða stöðu þjóðarinnar á alþjóðavettvangi.

Áhættuflokkun gervigreindar hjá Reykjavíkurborg byggist á samspili á milli tilgangs notkunar tólsins og viðkvæmni þeirra gagna (sjá öryggisflokkun) sem nauðsynleg eru til að ná þeim tilgangi og mótar þannig hvort og hvernig gervigreind er viðeigandi í tiltekinni starfsemi borgarinnar.



Óheimilt er að stofna aðgang með netfangi Reykjavíkurborgar í gervigreindartólum sem ekki hafa verið samþykkt af þjónustu- og nýsköpunarsviði, sbr. reglur um tölvunotkun hjá Reykjavíkurborg.

Áhættuflokkur 1: Heimil notkun

Opin gervigreindartól sem ekki hafa fengið sérstakt samþykki

Öryggisflokkun gagna: 1

Áhætta: Lítil/Hverfandi áhætta

Reykjavíkurborg getur ekki tekið afstöðu til allra þeirra opinna gervigreindartóla sem eru í boði á markaðnum í dag, einkum þar sem mörg þeirra eru opin og aðgengileg almenningi, ýmist ókeypis eða gegn vægu gjaldi. Við notkun slíkra tóla ber að sýna aðgát, fylgja leiðbeiningum um persónuvernd og forðast að setja inn upplýsingar sem ekki mega vera opinberar eða gætu talist viðkvæmar.

Tól

Opin gervigreindartól sem eru aðgengileg almenningi, eins og t.d. ChatGPT, Claude.ai, Google Gemini.

Gögn

Þegar opin gervigreindartól eru notuð skal einungis vinna með efni sem Reykjavíkurborg hefur þegar birt opinberlega. Þetta nær til útgefins efnis sem er aðgengilegt almenningi, svo sem greina, opinna fundargerða, skýrslna og annarra opinberra gagna borgarinnar.

Verkefni

Opin gervigreindartól má nýta í fjölbreytt verkefni eins og þýðingar, samantektir, gerð kynningarefnis, textasmíð, svörun tölvupósta og myndgerð, svo fátt eitt sé nefnt.

Hafa þarf í huga

Afurðir úr opnum gervigreindartólum, eins og myndir, skulu ekki notaðar í formlegt kynningarefni Reykjavíkurborgar. Í slíkum tilvikum skal ávallt styðjast við Hönnu – hönnunarkerfi borgarinnar. Höfundaréttarvarið efni, þar með talið tónlist, myndir og gagnasöfn, skal einnig umgangast með sérstakri varúð þegar unnið er með opin gervigreindartól.

Áhættuflokkur 2: Notkun háð samþykki

Notkun gervigreindar í innri starfsemi borgarinnar er háð samþykki

Öryggisflokkun gagna: 1 og 2

Áhætta: Miðlungs/Takmörkuð áhætta

Tól

Gervigreindartól sem samþykkt eru til notkunar í innri starfsemi Reykjavíkurborgar, svo sem Copilot, spjallmenni og stafrænt vinnuafli, falla undir þennan flokk. Öll slík tól skulu fara í gegnum mat á áhrifum á persónuvernd (MÁP) á vegum þess sviðs sem hyggst nýta tólið, auk gagnsæismats frá ÞON. Öll tól í þessum flokki þurfa sérstakt samþykki ÞON áður en notkun hefst.

Gögn

Við notkun gervigreindartóla sem hafa verið samþykkt í innri starfsemi borgarinnar má setja inn gögn og upplýsingar sem annaðhvort hafa þegar verið birt opinberlega eða tengjast daglegum rekstri borgarinnar, þó þær séu ekki birtar almenningi. Þó er stranglega bannað að



setja inn viðkvæmar persónuupplýsingar eða gögn sem teljast viðkvæm fyrir starfsemi borgarinnar.

Verkefni

Samþykkt gervigreindartól má nýta til margvíslegra verkefna sem styðja við innri starfsemi borgarinnar. Slík verkefni geta meðal annars falið í sér þýðingar, aðstoð við greina- og skýrsluskrif, gerð útdráttar og samantekta, vinnslu fundargerða og afritun viðtala. Einnig er hægt að nýta tólin til að búa til kynningarefni, yfirfara gögn fyrir og eftir fundi, fara yfir fjárhagsáætlanir, framkvæma einfaldar sjálfvirknivæddar aðgerðir og flytja gögn milli tölvukerfa. Þá má einnig nota tól til að meta myndir, til dæmis af kvittunum, og bera saman við upplýsingar í fjárhagskerfum.

Hafa þarf í huga

Mikilvægt er að notendur séu ávallt meðvitaðir um þegar þeir eiga í samskiptum við gervigreind, til dæmis við spjallmenni. Þá skiptir einnig máli að starfsfólk sem vinnur með slík tól hafi góða þekkingu á þeim og settar séu upp vörður þar sem réttleiki niðurstaðna er kannaður. Gervigreindartól af þessari gerð má hins vegar ekki nota til að taka ákvarðanir. Gervigreindartól í áhættuflokki 2 má t.d. ekki nýta til þess að meta nemendur, fara yfir atvinnuumsóknir, taka ákvarðanir sem varða rekstur eða sem varða réttindi og skyldur borgarbúa.

Áhættuflokkur 3: Notkun háð samþykki

Allar sértækar gervigreindarlausnir í starfsemi borgarinnar, sem vinna með viðkvæm gögn, eru háðar ströngum reglum.

Öryggisflokkur gagna: 1, 2 og 3

Áhætta: Mikil

Tól

Sértæk gervigreindartól eru sérhönnuð til að sinna afmörkuðu hlutverki og getur notkun þeirra haft áhrif á heilsu, öryggi eða grundvallarréttindi einstaklinga. Um slík tól gilda strangar reglur og kröfur og skulu þau ávallt fara í gegnum áhættumat, skráningu, ítarlegt mat á áhrifum á persónuvernd (MÁP), gagnsæismat og vera undir virku mannlegu eftirliti. Öll tól í þessum flokki þurfa sérstakt samþykki PÖN áður en notkun hefst.

Gögn

Við notkun samþykktra sértækra gervigreindartóla má setja inn sérvarin og viðkvæm gögn, svo sem persónuupplýsingar eða aðrar upplýsingar sem teljast viðkvæmar fyrir starfsemi borgarinnar, í samræmi við framangreinda áhættugreiningu. Slík gögn skulu þó aldrei fara út fyrir tæknilegt umhverfi Reykjavíkurborgar. Strangar öryggiskröfur og persónuverndarsjónarmið gilda um alla meðhöndlun gagna í þessum flokki.

Verkefni

Dæmi um sértæk gervigreindartól eru:

- Öryggiskerfi í mikilvægum innviðum sem taka ákvarðanir eða framkvæma mat. T.d. samgöngum, þar sem bilun gæti ógnað lífi og heilsu borgara
- Tól í heilbrigðisstarfsemi sem kemur að mati á lyfjagjöf einstaklinga eða meðferðarúrræðum.
- Tól sem kemur að ráðningum og mannauðsstjórnun, t.d. yfirferð á atvinnuumsóknum og mat á hæfi umsækjanda eða mat á hæfi starfsfólks og stjórnenda.
- Tól sem notað er í einkunnagjöf og námsmat sem getur haft áhrif á aðgang að framhaldsmenntun og framtíð starfsferils.



Reykjavík



- Notkun gervigreindar við aðgang að nauðsynlegum þjónustum (t.d. lánshæfismat sem gæti hindrað fólk í að fá lán)
- Tól sem við löggæslu sem gæti haft áhrif á grundvallarréttindi einstaklinga (t.d. mat á áreiðanleika sönnunargagna)
- Gervigreindarlausnir í málefnum innflytjenda, hælisleitenda og landamærastjórnunar (t.d. sjálfvirk greining á umsóknum um vegabréfsáritanir)
- Gervigreindarlausnir í réttarkerfinu og lýðræðislegum ferlum (t.d. kerfi sem undirbýr dómsúrskurði)

Hafa þarf í huga

Framkvæmd ákvarðana og það hvernig sértæk gervigreindartól komast að niðurstöðum þarf að vera bæði gagnsæ og auðskiljanleg fyrir öll sem hagnýta tólið. Slík tól má aðeins nota undir ströngu eftirliti og með virku mannlegu aðhaldi. Jafnframt er mikilvægt að þau séu þjálfuð á réttum og viðeigandi gögnum sem skulu vera gögn í eigu eða umsjá Reykjavíkurborgar.

Áhættuflokkur 4: Óheimil notkun

Öll gervigreindartól sem teljast skýr ógn við öryggi, lífsviðurværi og réttindi fólks eru óheimil.

Áhætta: Óásættanleg

Óheimil notkun:

- Skaðleg gervigreindardrifin stjórnun og hverskonar blekking.
- Skaðleg gervigreindardrifin misnotkun á viðkvæmni einstaklinga.
- Félagsleg stigagjöf (e. *social scoring*).
- Mat á áhættu eða spá um refsiverða háttsemi einstaklings.
- Vefsöfnun (e. *scraping*) frá internetinu eða eftirlitsmyndavélum til að búa til eða stækka andlitsgreiningargagnagrunna.
- Tilfinningagreining á vinnustöðum og í menntastofnunum.
- Flokkun byggð á líffræðilegum einkennum til að álykta um tiltekin vernduð einkenni.
- Rauntímafjargreining á líffræðilegum einkennum í opinberu rými í löggæsluskyni.

Viðmið þessi eru endurskoðuð reglulega, en þó ekki sjaldnar en árlega.

Samþykkt í stafrænu ráði dags. 25. júní 2025

Reykjavík, 25. júní 2025,

Óskar Jörgen Sandholt,
sviðsstjóri þjónustu- og nýsköpunarsviðs Reykjavíkurborgar