

Umsögn við viðmið um notkun gervigreindar hjá Reykjavíkurborg

Umsögn við viðmið um notkun gervigreindar hjá Reykjavíkurborg

Inngangur	1
Helstu áskoranir og umbótatillögur	2
a. Skortur á verklagsreglum og ábyrgð	2
b. Lýðræðisleg þátttaka og samráð við borgarbúa	2
c. Gagnsæi og aðkeyptar lausnir	3
d. Umhverfisáhrif og sjálfbærni	3
e. Hæfni og menntun starfsfólks	3
f. Samvinna og samræming	3
Tilfni til frekari stefnumótunar	4
Lokaorð	4

Inngangur

Við teljum það jákvætt og tímabært skref að Reykjavíkurborg sé að setja sér viðmið um notkun gervigreindar í starfsemi sinni. Slík viðmið marka mikilvægt upphaf í átt að ábyrgri og siðferðilegri nálgun á nýtingu þessarar öflugu tækni. Hins vegar verður að tryggja að viðmiðin séu ekki aðeins viljayfirlýsing heldur raunverulegur rammi sem styður við örugga, gagnsæja og réttláta framkvæmd.

Í því samhengi er mikilvægt að skýra betur grundvallarhugtök og aðgreina þau. Gervigreind er regnhlífarhugtak sem nær yfir fjölbreytta tækni, til dæmis sjálfvirka ákvarðanatöku, myndgreiningu, röklega hugsun og málvinnslu. Stór tungumálalíkön, á borð við þau sem knýja spjallvélar eins og ChatGPT, eru aðeins einn undirflokkur innan þessarar heildar. Ef notkunarviðmið eiga að styðja við skilning og ábyrga notkun þarf að útskýra þessi hugtök skýrt, bæði fyrir starfsfólki og almenningi, og forðast að orða tæknina á óljósan eða almennt yfirborðskenndan hátt.

Einnig má horfa til alþjóðlegrar þróunar þar sem gagnsæi og stjórn á gögnum hafa orðið kveikja að endurskoðun á notkun ákveðinna lausna. Í Danmörku hefur ríkið til dæmis ákveðið að hætta notkun á Microsoft í opinberum stofnunum vegna óvissu um hvernig persónugögn

eru meðhöndluð, hvert þau fara og hver hefur aðgang að þeim. Slíkar ákvarðanir minna okkur á að siðferðileg og traust notkun gervigreindar krefst skýrra reglna, rekjanleika og skilnings á innviðum tækninnar.

Við teljum því mikilvægt að Reykjavíkurborg haldi áfram að þróa viðmiðin með það að leiðarljósi að styrkja framkvæmd þeirra, skýra hugtök og efla fræðslu og þátttöku.

Helstu áskoranir og umbótatillögur

a. Skortur á verklagsreglum og ábyrgð

Við teljum brýnt að Reykjavíkurborg styrki framkvæmd viðmiðanna með skýrum verkferlum. Sérstaklega þarf að tryggja reglulega greiningu og viðbrögð við mögulegri hlutdrægni gervigreindarkerfa. Það krefst ferla sem meta sanngirni gagnvart ólíkum hópum og taka mið af þeim fjölbreytileika sem borgin þjónustar.

Við leggjum áherslu á reglulegt endurmat á sjálfum viðmiðunum í samræmi við hraða þróun tækni og siðferðilegra álitaefna. Einnig þarf að skilgreina með skýrum hætti hvernig eftirlit með gervigreindartólum fer fram, hver ber ábyrgð, og hverjar hæfniskröfur eru gerðar til þeirra sem koma að notkun og innleiðingu þeirra. Þar með talið að gera kröfu um símenntun í gervigreind fyrir ábyrgðaraðila.

b. Lýðræðisleg þátttaka og samráð við borgarbúa

Viðmiðin nefna ekki hvernig borgarbúar fá aðkomu að ákvörðunum sem hafa áhrif á líf þeirra. Við teljum nauðsynlegt að koma á reglubundnum samráðsferlum við borgarbúa og hagsmunaaðila, bæði áður en gervigreind er tekin í notkun og á meðan hún er í virkri þjónustu.

Jafnframt þarf að vera aðgengilegur rammi utan um endurgjöf og kvartanir frá notendum, svo hægt sé að bregðast við hugsanlegum skaða eða villum. Mikilvægt er að borgarbúar eigi einnig rétt á að vita hvort og hvernig ákvörðun sem þá snertir byggði á notkun gervigreindar. Þennan rétt þarf að skýra betur í viðmiði 2 um gagnsæi.

c. Gagnsæi og aðkeyptar lausnir

Við leggjum áherslu á að gagnsæi þurfi að ná til allrar innleiðingar, ekki aðeins til ytra viðmóts, heldur einnig gagna og gervigreindarlíkana sem notuð eru í aðkeyptum lausnum, jafnvel þótt borgin hafi ekki aðkomu að þjálfun þeirra. Einnig ætti gegnsæismatið að ná til þeirra kvaðninga (e. prompt) sem notast er við til að fá þær niðurstöður sem sóst er eftir.

Borgin ætti að birta opinberlega lista yfir öll gervigreindartól sem eru í notkun og hver á þau, og hvaða gervigreindarlíkön þau notast við. Einnig þarf að skilgreina skýrt hvað telst viðunandi útskýring á ákvörðun gervigreindar í því skyni að vernda sjálfræði og skilning notenda.

d. Umhverfisáhrif og sjálfbærni

Viðmiðin gera engar kröfur um mat á umhverfisáhrifum tækniinnleiðingar. Það er alvarlegur ágalli, sérstaklega í ljósi þess að borgin hefur sett sér markmið um kolefnishlutleysi. Við leggjum til að borgin framkvæmi skýrslu á tveggja ára fresti um umhverfisáhrif tækniinnleiðingar og taki tillit til sjálfbærni í vali og innleiðingu á tækni.

e. Hæfni og menntun starfsfólks

Við teljum mikilvægt að Reykjavíkurborg geri skýrar kröfur til hæfni, meðvitundar og ábyrgðar starfsfólks í notkun gervigreindartóla – bæði í formlegum verkefnum og í daglegu starfi. Þeir sem bera ábyrgð á innleiðingu þurfa að hafa djúpa innsýn í áhrif og takmarkanir gervigreindar, og slíkt krefst stöðugrar fræðslu og endurmenntunar.

Í kafla 1 um opna notkun vantar skýrar og afdráttarlausar leiðbeiningar. Starfsmenn þurfa að vita hvaða gögn má nota í opnum tólum (t.d. ChatGPT) og hvað fellur undir óheimila notkun. Orðalag á borð við „sýna aðgát“ eða „forðast að“ dugar ekki. Það þarf að skýra þessi viðmið með raunhæfum dæmum og siðferðislegum viðmiðunarramma.

f. Samvinna og samræming

Við leggjum áherslu á mikilvægi þess að Reykjavíkurborg vinni með háskólum og sérfræðingum að siðferðilegri og gagnrýnninni umræðu um gervigreind. Þá er mikilvægt að samræma stefnu og verklag við önnur sveitarfélög og ríkisstofnanir, svo ekki myndist ósamræmi í regluverki og kröfum.

Tilfni til frekari stefnumótunar

Það er mikilvægt að Reykjavíkurborg hugi að því hvernig gervigreind er skilgreind og aðgreind frá annarri tækni. Í mörgum tilfellum er gervigreind hluti af stærri tækni kerfum sem ekki eru notendum sýnileg sem slík, t.d. við stýringar á upplýsingaflæði, kortanotkun, eða þjónustustýringu. Slík blanda af gervigreindardrifiðni og annarri sjálfvirkni gerir það flókið að setja stefnu sem nær eingöngu til „gervigreindar“ en útilokar önnur tengd kerfi.

Nauðsynlegt er að líta á tæknilausnir sem heildarkerfi. Stefnumótun ætti að byggja á kortlagningu þeirra kerfa sem eru í notkun, hver tilgangur þeirra er, og hver áhrif þeirra kunna að vera, óháð því hvort þau séu formlega flokkuð sem „gervigreind“ eða ekki.

Einnig mætti skoða aðgreiningu á viðmiðum eftir því hvort um sé að ræða tækni sem þróuð er innanhúss eða aðkeypta lausn. Þau viðmið og kröfur sem eiga við um töl sem borgin hefur beina stjórn á, eins og innanhúss þróað kerfi, eru ekki endilega þau sömu og eiga við um stærri, lokuð kerfi eins og Google Workspace eða sjálfkeyrandi farartæki.

Lokaorð

Viðmiðin sem Reykjavíkurborg hefur sett fram eru mikilvægt skref í átt að ábyrgri notkun gervigreindar í opinberri þjónustu. Til að þau hafi raunverulegt gildi þarf að styðja þau með skýrum ferlum, reglubundnu mati og virkri þátttöku starfsfólks og borgarbúa. Án þess er hættu á að viðmiðin verði óljós í framkvæmd og skili ekki því gagnsæi og trausti sem að er stefnt.

Gervigreind í opinberri þjónustu snertir ekki aðeins tæknileg atriði, heldur einnig siðferðileg, samfélagsleg og jafnréttistengd sjónarmið. Því er brýnt að sérfræðipekking sé til staðar innan borgarkerfisins, sérstaklega á sviði tæknisiðfræði. Slík þekking er lykilatriði til að styðja við stefnumótun, tryggja meðvitaða ákvarðanatöku og veita markvissa fræðslu og endurmenntun til starfsfólks.

Reykjavíkurborg hefur einstakt tækifæri til að vera leiðandi í mótun faglegrar og samfélagslegra ábyrgðar notkunar gervigreindar. Slíkt hlutverk krefst stöðugs samtals, gagnrýnnar greiningar og sveigjanlegrar nálgunar í samræmi við hraða þróun tækninnar.

Við í Samtökum um mannvæna tækni erum reiðubúin til samstarfs og getum komið og gert

grein fyrir sjónarmiðum samtakanna ef þess er óskað.