

Efni: Umsögn Skákgreindar ehf. um stefnu Íslands um gervigreind, 15 mars 2021.

Skákgreind er Sprotafyrirtæki, sem þróar kennsluhugbúnað sem nýtir gervigreinda tækni fyrir einstaklingsmiðaða þjálfun ásamt námsefni á sviði fullorðinsfræðslu, sem undirbýr almenning og starfsmenn fyrir áskoranir tengdar innleiðingu gervigreindar.

Fjallað er um mikilvægi þess að Stefna um gervigreind sé fyrsta skref í átt að hornsteini þar sem fram koma viðmið, sem breið samstaða er um, standist tímans tönn, leggi áherslu á sjálfbærni og ferli sem hægt er að viðhalda í sibreytilegu tækniumhverfi sem í eðli sínu kallar á ítrunarnálgun.

Sérstaklega er fjallað um sýnidæmi (myndræn tækni í löggæslu, sjá umsögn Persónuverndar), til að greina hvað getur farið úrskeiðis í gervigreindri tækni og hvernig er hægt að gera betur.

Þá er lögð áhersla á mikilvægi fræðslu um gervigreind bæði til að virkja almenning í upplýstri umræðu, að almenningur hafi betri forsendur til þess að veita eða veita ekki upplýst samþykki, gæta að hagsmunum þeirra sem starfa með og fá þjónustu frá gervigreindarstuddri tækni.

Lögð er áhersla á gildi og viðmið sem halda gildi sínu í dýnamísku umhverfi.

Fyrir umfjöllun um hvaða þjóðfélagsþætti tækniþróunin getur haft áhrif á, er vísað í umfjöllun Vitvélastofnunar, Landssamtaka um Þroskahjálp, VR, ASÍ og SA.

1. Hvaða gildi á íslenskt samfélag að hafa að leiðarljósi við innleiðingu nýrrar tækni um gervigreind?

- Markmið:

Mikilvægasta gildið er að leggja áherslu á markmið beitingu tækninnar. Leyfa ætti notkun sem fer fram í þeim tilgangi að bæta lífsgæði samborgara, efla starfshæfni og verðmætasköpun, styðja við þá sem minna mega sín og leiða velmegunar þjóðfélagsins. Mikilvægt er að styðja við þá þjóðfélagsþegna, sem eiga undir högg að sækja og bera hagsmuni þeirra fyrir brjósti. Það getur verið vegna félagslegs- og eða menntunarlegs bakgrunnar, fötlunar og einnig vegna sjúkdóma sem gervigreind tækni getur stutt við að meðhöndla. Einnig er mikilvægt að hafa í fyrirrúmi hagsmuni þeirra sem upplýsinga er aflað um til úrvinnslu í gervigreindu kerfi. Síðast en ekki síst er mikilvægt að innleiðing tækninnar sé sjálfbær. Að allir njóti góðs af þróuninni og að enginn sé skilinn útundan. Að tæknin þannig lyfti upp þeim sem eiga undir högg að sækja á sama tíma og hún eflir og skapar sóknarfæri í leik og starfi.

Í umsögn Landssamtakanna Þroskahjálpar eru margvísleg dæmi nefnd tengd mannréttindum, þar sem færa má rök fyrir því að markmiðið sé umdeilanlegt og gangi gegn hagsmunum þeirra sem aflað er upplýsinga um. Leggja ber áherslu á gagnsæi.

Aðgerð: leggja áherslu á að markmiðin, sem tæknin á að ná fram séu jákvæð fyrir þjóðfélagið, þar með talið þá einstaklinga sem eiga undir högg að sækja t.d. vegna sjúkdóma og eða fatlanir sem gervigreind tækni getur gert bærilegri. Leggja áherslu á að markmiðin séu til hagsbóta fyrir þann sem veitir aðgengi að upplýsingum, sem gervigreind vinnur úr.

- Þjálfun:

Það er nauðsynlegt að leggja áherslu á rétta þjálfun þeirra sem vinna með tæknina. Þjálfunin snýr bæði að rétti starfsmanna til að þekkja sitt vinnuumhverfi og einnig rétti þeirra sem njóta þjónustu sem studd er með gervigreind til áreiðanlegrar, hágæða þjónustu.

Fyrir starfsmenn er nauðsynlegt að fá tímanlega upplýsingar um þá tækni sem til stendur að innleiða og viðeigandi þjálfun til að vinna með gervigreinda tækni þ.a. þau geti undirbúið sig og mætt auknum starfstengdum áskorunum. Starfsmenn þurfa að hafa forsendur til þess að smíða vandaða verkferla, þar sem áratuga þekking og reynsla á viðfangsefninu er studd með tækninni.

Einnig þarf að huga að rétti samborgara og atvinnulífs til þess að þau sem vinna með gervigreindar upplýsingar hafi rétta þjálfun þ.a. gæði þeirrar þjónustu sem er veitt sé áreiðanleg og traust.

Sem dæmi nefnir Persónuvernd að notkun andlitsgreiningartækni í löggæslu hafi verið ábótavant. Færa má rök fyrir því að af því gefnu að markmiðin séu göfug, þá sé það að sinna ekki þjálfun starfsmanna ein megin ástæða þess að slík tækni hefur ekki í öllum tilvikum nýst sem skyldi. Starfsmenn s.s. lögreglumenn þurfa að skilja styrkleika og takmarkanir tækninnar, undir hvaða forsendum hún virkar og hvenær þarf að fylgjast sérstaklega vel með greiningunni. Þjálfun þarf í að beita tækninni og við smíði verkferla á þann veg að mannleg skynsemi og hugsun sé studd með tækninni. Líkt og t.d. hjá flugmönnum, sem vinna ábyrgðarstarf sem stutt er með tækni, þarf að viðhalda hæfni til að vinna með ákveðnar tegundir gervigreinda kerfa.

Í ljósi áherslna um að efla stafræna þjónustu við borgara er nauðsynlegt að huga sérstaklega að opinberum starfsmönnum. Æskilegt er að starfsmenn fái fyrirvara í formi upplýsinga um hvaða verkfæri, sem byggja á hvaða gervigreindartækni, til stendur að innleiða á þeirra vinnustað og þeim sé gert kleyft að búa sig undir breytingarnar, sem verða á þeirra störfum með fyrirvara ásamt því að þeim sé veitt aðgengi að þjálfun, sem undirbýr þau undir breytt starfsumhverfi.

Þjálfun styrkir stoðir þess sem VR leggur til, þ.e. lýðræði á vinnustað. Það er æskilegt að allir starfsmenn taki þátt í upplýstri umræðu og hafi augun opin fyrir tæknitengdum tækifærum.

Aðgerðir:

Við mat á árangri innleiðingar stafrænnar tækni ætti að hafa í heiðri sjálfbær gildi og skoða sérstaklega hvernig til tókst með þjálfun starfsmanna. Var reynslumiklum starfsmönnum sagt upp vegna þess að þau náðu ekki að halda í við breytingar á sínu starfi, sem urðu vegna innleiðingar tækninnar?

Fengu starfsmenn fyrirvara og upplýsingar um tæknina sem var innleidd?

Var boðið uppá viðeigandi fræðslu sem hafði það að markmiði að starfsmennirnir gætu tekist á við breyttar starfstengdar áskoranir sem innleiðing tækninnar hefur í för með sér?

Náðu starfsmenn með öflugri sí- og endurmenntun að færast yfir í annað hlutverk innan stofnunarinnar eða fyrirtækisins, þar sem þau nýta gervigreinda tækni til að efla verðmætasköpun og auka framleiðni?

Fókusinn er á sjálfbærni og heildræn áhrif. Það er dýrt fyrir þjóðfélagið ef starfsfólk sem byggt hefur upp áratuga sérfræðireynslu á ákveðnu sviði heltist úr starfi vegna þess að það nær ekki að halda í við tækniframfarirnar. Það mun ekki vera auðveldara að fá starf á öðru sviði þar sem viðkomandi hefur ekki áratuga sérfræðiþekkingu.

- Þjálfunargögn - íslensk menning og sérkenni ekki „utan dreyfingar“

Gervigreind tækni er þróuð fyrir ákveðin verkefni, sem eru framkvæmd undir ákveðnum kringumstæðum af ákveðnum aðilum. Hver aðferðafræði hefur styrkleika og veikleika. Aðferðirnar sem í dag leiða þróunina eru m.a. byggðar á tauganetum, sem í eðli sínu byggja á tölfræði. Við hagnýtingu tækninnar ber að hafa í huga að ekkert gervigreindarlíkan er betra en þau gögn sem það er þjálfað á. Ef þjálfunargögnin hafa skekkju t.d. mismuna blökkumönnum, þá mun gervigreindarlíkan sem byggt er á þeim gögnum t.d. andlitsgreiningartækni fyrir löggæslu bera þá skekkju og mismunun áfram, sjá nánar dæmi m.a. í umsögn Landssamtaka Proskahjálpar. Þar er ekki við tæknina, þ.e. líkanið eða algorithmann, sem slíka að sakast, heldur þau gögn sem er unnið með.

Það er mikilvægt að taka tillit til ólíkra og viðkvæmra þjóðfélagshópa við þróun gervigreindra kerfa. Það hefur verið talað um svokölluð WEIRD áhrif (White, Educated, Industrialized, Rich, Democratic). Í öllum heiminum eru ca. 8 billjón manns, þar af eru minna en 1 billjón WEIRD. Skv.

Daniel Haun, 2017 voru 92% af vísindarannsóknnum í þróunarsálfræði framkvæmdar með WEIRD þátttakendum, sjá einnig Kupferschmidt, 2019.

Gildi fyrir íslenskt samfélag ættu að leggja áherslu á að gervigreind tækni sé þróuð með alla þjóðfélagshópa í huga, ekki bara WEIRD fólk. Þá ætti að leggja áherslu á að á Íslandi séu virkar rannsóknir, sem tryggja að gervigreind kerfi takið mið af menningarlegum sérkennum Íslendinga. Tungumálið er mikilvægur þáttur okkar menningar, en íslensk menning afmarkast ekki við tungumálið. Í t.d. mennta- og heilbrigðistengdu samhengi kann að vera að séríslensk menningaráhrif sé að finna á fjölbreyttum sviðum, t.d. í tækni sem greinir upplifun og líðan fólks út frá atriðum á borð við blæbrigði raddarinnar, svipbrigði, áslátt á lyklaborð og líkamsstöðu.

Svo framanlega sem markmið tækninnar eru göfug (e. altuist), starfsfólk sem vinnur með tæknina vel þjálfað, unnið eftir vel skilgreindum verkferlum og gætt vel að persónuverndarsjónarmiðum s.s. upplýst samþykki, er það hagur íslensks samfélags að tæknin nái yfir íslensk sérkenni og menningu í víðasta skilningi. Með öflugum innlendum rannsóknnum og þátttöku í tilraunum er tryggt að tæknin nýtist íslenskum þjóðfélagsþegnum til fulls og að íslensk sérkenni og menning sé ekki utan tölfræðilegu dreyfingarinnar (e. out of distribution) sem unnið er með þegar tæknin er þróuð, sem annars myndi þýða að tæknin nýtist síður.

Tekið er undir áherslu VR um að þjálfunargögn séu vönduð og laus við skekkju eða mismunun eftir aldri, kyni, kynhneigð, þjóðerni, litarafti, trú eða stjórnmálaskoðunum.

Aðgerðir: leggja áherslu á og stuðla að innlendar rannsóknir og hagnýtingu gervigreindartækni, þar sem tekið er mið að íslenskum sérkennum og menningu með það að markmiði að tækniframfarirnar nýtist öllum íslenskum þjóðfélagsþegnum til fulls, ekki síst þeim sem minna mega sín. Að tæknin sé þannig fyrir alla (e. inclusive).

- Aðgengi að gögnum – aðferðir sem gera gögn ekki persónugreinanleg

Gögn eru verðmæt auðlind sem hægt er að nýta í jákvæðum tilgangi fyrir þjóðfélagið. Það er mikilvægt að gera gögn og upplýsingar aðgengilegar með því fyrirkomulagi að nýjustu rannsóknir á sviði þess að gera gögn ópersónugreinanleg séu nýttar.

Aðgerð: stuðla að aðgengi að gögnum þar sem þess er gætt að háþrúðum aðferðum sé beitt við að gera gögnin ekki persónugreinanleg.

- Samhygð (empathy) og heilbrigði

Það er mikilvægt að leggja áherslu á að gervigreind kerfi hafi mannvirðingu og samhygð (e. empathy) í fyrirrúmi. Að tæknin stuðli að þjóðfélagi þar sem uppbyggileg samskipti, jákvæð orka og hvatning er leiðarljós. Til þess að gera þá þróun mögulega þurfa gervigreind kerfi að geta skynjað líðan og upplifun. Einnig gegna slík kerfi mikilvægum heilsufarslegum tilgangi, sjá hér að neðan. Sem dæmi þá getur gervigreind tækni sem greinir líðan og upplifun bætt lífsgæði þeirra sem glíma við sjúkdóma sem hafa áhrif á getu til að greina líðan samborgara. Einnig getur hún nýst við að greina fyrstu skref taugahrönnunarsjúkdóma á borð við Alzheimer, sjá:

Screening of Alzheimer's disease by facial complexion using artificial intelligence eftir Yumi Umeda-Kameyama et. al., 2021, aðferðin byggir á rannsóknnum sem sýndu að Alzheimer sjúklingar eiga erfitt með að greina svipbrigði, sjá: Impaired Recognition of Facial Expressions of Emotion in Alzheimer's Disease eftir Rita Hargrave et. al, 2002.

Sýnst hefur verið framá svipaðar niðurstöður fyrir Parkinson, sjá: Specific impairments in the recognition of emotional facial expressions in Parkinson's disease eftir Uraina S.Clark et. al, 2008 og Facial expression recognition in people with medicated and unmedicated Parkinson's disease eftir R.Sprengelmeyer et. al., 2003.

Einnig fyrir Parkinson hefur tæknin nýst vel, sjá: Review of automated emotion-based quantification of facial expression in Parkinson's patients eftir Sonawane et. al, 2020.

Aðgerð: leggja áherslu á markmið, þjálfun og að tæknin sé nýtt í réttu samhengi. Huga við lagasetningu að því að takmarka ekki möguleika tækninnar til góðra verka.

- Leggja áherslu á að gervigreindar aðferðir innihaldi upplýsingar um hversu viss þau eru í sinni sök

Það er veikleiki hjá þeim aðferðum sem leiða tækniþróunina og byggð eru á svokölluðum djúpum tauganetum að þau eiga erfitt með að útskýra niðurstöður sínar og veita mat á skekkju eða hversu viss þau eru í sinni sök. Mikill þungi er í rannsóknnum sem miða að því að bæta hér úr. Í dag er engu að síður hægt að veita visst mat með því að greina gæði inntak líkansins með forvinnslu algoritnum. Sem dæmi ef inntakið er ljósmynd af andliti, þá er hægt að greina hversu vel helstu kennileiti andlits sjást.

Það er mikilvægt að huga að því að gervigreind tækni er nýtt í ákveðnu samhengi til að leysa ákveðið verkefni. Tvenn mistök geta átt sér stað t.d. í læknisfræðilegri greiningu þá getur verið að sjúklingur sé ranglega greindur með sjúkdóm eða að sjúklingur sem er með sjúkdóm sé ranglega greindur að hann sé full frískur. Gervigreind kerfi þurfa að vera hluti af ferli, þar sem meðvituð ákvörðun er tekin m.a. um hvers konar mistök eru bærilegri. Er betra að þrátt fyrir t.d. léleg myndgæði skili greiningin því að sjúklingurinn gæti verið með sjúkdóm frekar en að algorithminn segi að hann geti ekki svarað? Í sumum tilvikum t.d. löggæslu gæti verið betra að algorithminn segi að myndgæði séu ekki næginlega góð og því vilji hann ekki skila niðurstöðu, eða skili niðurstöðu með sterkum fyrirvara. Skilgreining á því er hluti af vönduðum verkferlum.

Það er mikilvægt að í kaupum á gervigreindum kerfum sé hugað að þessu. Ef t.d. ætlunin er að nýta tæknina í löggæslu, að þá sé ekki ódýrarasta kerfið keypt, sem spítir útúr sér niðurstöðum án mats á því hvort inntaksgögn séu næginlega góð.

Aðgerð: leggja áherlsu á að gervigreindri tækni er ekki beitt í tómarúmi, heldur er hún hluti af ferli eða keðju, þar sem skilgreina þarf vel markmið, s.s. hvers konar mistök er betra að lifa við, og hvernig farið er með niðurstöður gervigreindrar greiningar. Tæknin sjálf er bara einn hlekkur í þeirri keðju. Mikilvægt er að mat á vissu fylgi gervigreindri greiningu.

- Réttur starfsmanna til að þekkja þá tækni sem unnið er með

Í umfjöllun New York Times um mistök sem voru gerð í notkun gervigreindar (sjá þema sem Persónuvernd velur), kemur fram að ákveðin deild hjá Detroit Police Department gerði mistökin. Starfsmenn eru yfirleitt stoltir af sínu starfi og vinnustað og bera virðingu stofnunarinnar eða fyrirtækisins fyrir brjósti. Það má halda því fram að það að setja starfsmenn í þá stöðu, sem virðist hafa gerst, að þeir þekki ekki þá tækni sem unnið er með og hafi ekki forsendur til að beita henni, vegi að þeirra mannvirðingu og heiðri og sé skerðing á mannréttindum. Starfsmenn ættu að hafa réttindi um aðgengi að upplýsingum um þá tækni sem þeim er falið að vinna með.

Fyrir dæmið sem hér er unnið með, notkun andlitsgreiningarhugbúnaðar í löggæslu, þá hefði verið æskilegt að lögreglumennirnir og sá aðili sem sér um að fræða þá hefðu fengið tímanlega að vita hvers konar tækni greiningarhugbúnaðurinn byggir á og aðgang að fræðslu sem undirbýr þau fyrir að vinna með tækninni. Í þessu tilfelli gæti t.d. verið að ræða fléttutauganet (e. convolutional neural network), sem hefur enga forvinnslu og leggur því ekki mat á gæði inntaksgagna. Netið skilaði mögulega niðurstöðu fyrir öll inntaksgögn (það hefði mögulega verið hægt að gefa sem inntak mynd af fjölskyldu hundinum eða kettinum, það myndi samt reyna að finna fólk í gagnagrunninum sem myndin líktist). Netið var þjálfað á gagnagrunni sem innihélt x myndir af einstaklingum af uppruna y. Gögnin voru fengið með hætti z (mögulega um WEIRD þjálfunargögn að ræða með fáum blökkumönnum).

Með upplýsingar á borð við þessar í höndunum geta fræðsluaðilar og starfsmenn áttað sig á því um hvers konar tól er að ræða, hvaða takmarkanir það hefur og hverju þarf að huga að bæði við gerð verkferla og einnig þjálfun starfsmanna. Það er mögulegt að slík greining myndi leiða til þeirrar niðurstöðu að akkúrat sú útfærsla sem unnið er með henti ekki vel fyrir það verkefni (mistök í innkaupum) og það ábyrgðarhlutverk sem úrvinnslan er hluti af sbr. löggæsla, eða að

það verði að taka niðurstöður gervigreindu greiningarinnar með miklum fyrirvara. Það er þá almennt ekki við tækniáðferðina slíka (það er hægt að hafa forvinnslu o.s.frv.), heldur útfærsluna að sakast.

Eins og fram kemur í umsögn Landssamtakanna Þroskahjálpar, þá geta gervigreind kerfi haft áhrif á ákvarðanatöku fagaðila, sem eru líklegri að efast um dómgreind sína. Það er nauðsynlegt að fræða fagaðila þ.a. þeir hafi raunhæfar væntingar til tækninnar og geti nýtt hana til að styðja við sína sérfræðipækkingu og almenna skynsemi. Þá er mikilvægt að viðhalda gagnsæi gagnvart starfsmönnum. Í umsókn Persónuverndar er fjallað um evrópskar tilraunir þar sem borgarar fá aðgang að skrá með upplýsingum um hvaða gervigreindartækni er beitt í opinberri þjónustu. Æskilegt er að starfsmenn hafi ekki síðri upplýsingar um þá tækni sem þeir vinna með.

Aðgerð: Að upplýsingaskylda sé um þá tækni sem innleidd er á vinnustað þ.a. ekki sé vegið að heiðri og mannvirðingu starfsmanna með því að láta í hendur þeirra tól sem þeir þekja ekki og vita ekki hverjar takmarkanir tólsins eru.

- Réttur borgara til fræðslu – upplýst samþykki

Eins og vikið er að í umsögn VR, er upplýst samþykki sífellt flóknara viðfangsefni. Mikilvægt er að allur almenningur hafi grunnþekkingu á gervigreind bæði til að virkja sem flesta í upplýstri umræðu um tæknina og þann ramma sem hún vinnur innan og einnig til að almenningur hafi forsendur til að veita upplýst samþykki eða þá að velja að standa utan ákveðinnar gervigreindrar notkunar.

Um flókið viðfangsefni er að ræða, því eðli upplýsinga getur breyst s.s. geta tvö gagnasöfn sem eru sitt í hvoru lagi ekki persónugreinanleg orðið persónugreinanleg þegar þau eru sett saman. Að sama skapi geta viðbótarupplýsingar gert gagnasafn persónugreinanlegt.

Þá þarf að skilja á milli gagnasöfnunar annars vegar og úrvinnslu og aðgengis gagna hins vegar.

Aðgerð: Leggja áherslu á fræðslu almenning þ.a. allir hafi grunn þekkingu á tækninni. Efla getu til virkri þátttöku í upplýstri umræðu og ígrundaðrar ákvarðanatöku varðandi þátttöku eða standa utan söfnun ákveðinna upplýsinga eða greiningar eða aðgengis.

- Sjálfbærni

Við stafræna og gervigreinda innleiðingu þarf að huga að sjálfbærni og vinna markvisst að því að lágmarka mögulegar neikvæðar afleiðingar innleiðingarinnar. Fylgja þarf með fjármagn til að þjálfa starfsfólk í að vinna í breyttu starfsumhverfi við hlið og með tækninni. Starfsfólk þarf að fá fyrirvara í formi tímanlegra upplýsinga um hvers konar tækni til stendur að innleiða og hvaða gervigreindaraðferðir þar er unnið með ásamt aðgengi að fræðslu sem undirbýr það undir breyttar starfstengdar áskoranir.

Aðgerðir: Áhersla og fjármagn til að tryggja sjálfbærni. Við úttekt á árangri gervigreindra kerfa s.s. hjá hinu opinbera, sé skylda að huga að því hversu vel tókst til við fræðslu starfsmanna og hvort allir náðu að læra á tæknina og nýta hana. Var sett það markmið að þeir starfsmenn sem

fyrir eru fáir þjálfun og geti haldið sínu starfi þrátt fyrir breyttar áherslur á vinnustað í kjölfar innleiðingar gervigreindrar tækni? Meta starfsánægju fyrir og eftir innleiðinguna.

- Lagasetning - Læra af mistökum - ítrunarferli

Í vísindum rétt eins og annars staðar er óhjákvæmilegt að mistök séu gerð. Mikilvægt er að draga réttar ályktanir af mistökum. Sem dæmi með notkun andlitsgreiningar fyrir löggæslu, þá þarf þar eins og annars staðar að leggja áherslu á þjálfun starfsfólks, að þróa hágæða þjálfunargögn sem ekki hafa innbyggða mismunun, að forvinnsla og eða greiningartæknin meti gæði inntaksgagna og vandaða verkferla þar sem starfsfólk sem hefur hlotið þjálfun í beitingu tækninnar og túlkun greininga gegnir lykil hlutverki. Huga þarf að heildarkeðjunni og ekki líta á tæknina sem töfralausn sem þurfi ekki vandaða verkferla.

Heimurinn væri ekki betri í dag ef t.d. eðlisfræðirannsóknir hefðu verið bannaðar þegar það var uppgötvað að jörðin er ekki miðpunktur sólkerfisins. Hafa þer í huga hagsmuni komandi kynslóða, að þau geti notið góðs af tækniframförunum rétt eins og núverandi þjóðfélagsþegnar hafa gert. Einnig þarf að hafa í huga þau góðu verk sem tæknin getur unnið sbr. umsögn Landssamtaka um Þroskahjálp.

Aðgerð: Leggja áherslu á vönduð grunngildi, verkferla og fræðslu í réttri beitingu tækninnar frekar en boð og bönn.

2. Hvert á hlutverk tækni [og] gervigreindar að vera í íslensku samfélagi?

Hlutverk tækninnar er að bæta lífsgæði þjóðfélagsþegna og velmegun. Huga þer sérstaklega að því hlutverki að nýta tæknina til að styðja við þá þjóðfélagsþegna sem minna mega sín s.s. vegna sjúkdóma og eða veiks félagslegs og menntunarlegs bakgrunns. Tæknin þarf að vera fyrir alla með sjálfbærni og verðmætasköpun að leiðarljósi. Það er hagur allra að innleiða gervigreind með farsælum hætti, sem eflir lífsgæði í leik og starfi, eykur hæfni starfsfólks, verðmætasköpun og velmegun.

3. Á hvaða vettvangi ætti Ísland að ræða og leysa álitamál sem upp munu koma varðandi innleiðingu og notkun nýrrar gervigreindartækni?

Tekið er undir þá sýn sem aðrir umsagnaraðilar s.s. Landssamtök um Þroskahjálp og aðilar atvinnulífsins leggja fram.

Öflugt samstarf er nauðsynlegt til að takast á því þær áskoranir sem tækniframfarirnar hafa í för með sér.

Skapa þarf víðtæka sátt um þau gildi og það ferli sem unnið er eftir. Æskilegt er að helstu aðilar þjóðfélagsins, þar með talið atvinnulífsins taki höndum saman og vinni saman að framþróun sem styður hag allra. Stefna um gervigreind getur verið hornsteinn slíks samstarfs, þar sem áhersla er lögð á grunngildi.

Einnig er æskilegt að allir þjóðfélagsþegnar hafi tækifæri til að láta í ljós sýna skoðun og áhyggjur.

Eins og vikið er að í umsögn VR er einnig nauðsynlegt að tryggja að virkt eftirlit sé fyrir hendi til að tryggja hagsmuni almennings og virða mannréttindi ásamt því að koma í veg fyrir að svartir sauðir skemmi ekki fyrir þeim sem stunda vönduð vinnubrögð.

Það er nauðsynlegt að efla fræðslu með það að markmiði að allir þjóðfélagsþegnar geti tekið þátt í upplýstri umræðu um tæknina. Um ítrunarferli er að ræða, framþróunin er komin til að vera.

Mikilvægt er að vettvangurinn og verkferlar séu raunsæir og taki til greina takmarkanir t.d. á sviði mannafla. Þúsundir vísindagreina eru birtar á hverjum degi á sviði gervigreindar. Með áherslu á grunnild og verkferla er hægt að forðast að þurfa að fylgjast nákvæmlega með þróun gervigreindaaðferða á hverju sviði fyrir sig. Tækninni fleygir fram og áhrif hennar og geta til að kljást við fjölbreytt úrlausnarefni breytist dag frá degi.

Evrópskar reglur gilda á Íslandi. Ísland hefur tækifæri til þess að sýna frumkvæði og nýta samtakamátt og stuttar boðleiðir til að vera skrefi framur en evrópsku reglurnar í nýtingu gervigreindar þjóðfélaginu til hagsbóta.

Mikilvægt er að efla samstarf og upplýsingastreymi á milli opinberra aðila s.s. Persónuvernd, Lögreglu, Landhelgisgæslunnar, Hagstofunnar, Seðlabankans og Samkeppniseftirlitsins og fyrirtækja og rannsóknarstofnana, sem eru framanlega í nýtingu gervigreindar. Rétt eins og bent er á í áliti Persónuverndar hefur ákveðin tegund samstarfs reynst vel í Noregi.

Mikilvægt er að efla samstarf við alþjóðlegar stofnanir, sem bera farsæla innleiðingu á gervigreind fyrir brjósti og horfa þar ekki bara til Norðurlandanna og Evrópu, heldur einnig til Bandaríkjana, sem eru í dag fremst í þróun gervigreindar þó að Kína sækir á.

Einn mögulegur samstarfsaðili er: Initiative on Technology, Culture and Society við Rice Háskólann í Bandaríkjunum. Þar er um að ræða þverfaglega rannsóknarstofnun, sem hefur það að markmiði að stuðla að þjóðfélagslega og siðfræðilega ábyrgri innleiðingu gervigreindar.

Við lagasetningu er mikilvægt að hugsa til þeirra sem njóta góðs af tækninni. Sem dæmi þá nýtist andlitsgreiningartækni, þ.e. tækni sem greinir upplifun, fólki sem glímur við margvísleg heilsufarsleg vandamál (sjá tilvísanir í kafla um samhygð) og huga þarf að hagsmunum framtíðar kynslóða. Eins og vikið er að í umsögn Landssamtaka Þroskahjálpar getur tæknin stutt við fatlaða einstaklinga, aukið lífsgæði og gert þeim kleyft að taka virkan þátt í samfélaginu.

Lagasetning fyrir mælingar sem hafa það að markmiði að skilja upplifun einstaklinga s.s. blæbrigði raddarinnar, áslátt á lyklaborð, göngulag og sumar aðrar upplýsingar um lífkenni og hegðun einstaklinga ætti að taka mið af því að með réttu markmiði og nálgun er hægt að beita þessari tækni til að efla lífsgæði þjóðfélagsþegna. Tæknin samborgurum sem minna mega sín vegna ólíkra ástæðna. (sjá tilvísanir í kafla um samhygð). Mikilvægt er að takamarka ekki gerð gervigreindra kerfa, sem hafa virðingu og samhygð (e. empathy) fyrir notendum sínum að hornsteini.

Tekið er undir að yfirveguð nálgun þar sem frá upphafi er hugað að innbyggðri og sjálfgefni persónuvernd er farsæl, þó með því innleggi að eðli tækninnar og gagnanna er um ítrunarferli að ræða frekar en að hægt sé að negla niður ramma sem stendur í stað um ókomna tíð.

Huga ætti að rétttri beitingu tækninnar og forðast bönn (árinni kennir illur ræðari) því það þýðir að tæknin er ekki aðlöguð að íslenskri menningu og sérkennum ásamt því að samborgarar og komandi kynslóðir njóti ekki góðs af bættum lífsgæðum, sem tæknin getur boðið uppá.

Ísland er aðili að Evrópska efnahagssvæðinu og löggjöf þess á sviði gervigreindar. Ísland hefur engu að síður tækifæri til að taka frumkvæði og standa framanlega í ábyrgri og sjálfbærri nýtingu gervigreindar með áherslu á gagnsemi fyrir þjóðfélagið í heild og bætt lífsgæði fyrir almenning í landinu, hagsæld og velmegun atvinnugreina. Eins og vikið er að í umsögn SA gegnir skynsamleg nýting tækninnar lykilhlutverki í samkeppnishæfni atvinnulífsins og þar með þjóðfélagsins.

4. Hvað þarf svo að íslenskt atvinnulíf geti að fulli nýtt tækni gervigreindar?

Fræðsluáttak þar sem fókusinn er á að auka hæfni starfsfólks til að vinna með gervigreind og þróa vandaða verkferla auk þess að upplýsa almenning og stuðla að upplýstri umræðu. Átaksverkefni hjá Tækniþróunar- og Rannsóknarsjóði Rannís getur hér gengt mikilvægu hlutverki.

Einnig þarf íslenskt atvinnulíf og borgarar hágæða, áreiðanlega þjónustu frá hinu opinbera. Til að tryggja sjálfbærni stafrænnar innleiðingar er mikilvægt að skerpa á rétti (opinberra) starfsmanna til upplýsinga um gervigreinda tækni, sem til stendur að innleiða eða hefur verið innleidd ásamt því að taka frá fjármagn og veita aðgang að viðeigandi þjálfun, sem er viðhaldið. Æskilegt er að hafa sjálfbær sjónarmið í huga, þ.e. að starfsmenn nái að halda í við tækniframfarirnar og að vinnustaðurinn missi ekki áratuga þekkingu og reynslu vegna þess að starfsfólk fær ekki viðeigandi þjálfun. Einnig er mikilvægt að efla gæði og áreiðanleika verkferla og nýtingar tækninnar.

Til að forðast hagsmunaárekstra og auka sérhæfingu getur verið æskilegt að skilja á milli annars vegar þróunar stafrænnar tækni á opinberum vettvangi, sem hefur það að markmiði að auka þjónustu við borgara og hins vegar þjálfunar starfsmanna með sjálfbærni að leiðarljósi

Við úttækt á áhrifum stafrænnar innleiðingar, s.s. mat á sparnaði ætti að kanna heildræn, sjálfbær áhrif, þar með talið áhrifin á starfsmenn, hvernig þjálfun tókst til og hvort það tókst að endurþjálfva reynt starfsfólk þ.a. það gat fótað sig í breyttu starfsumhverfi. Einkaaðilar á borð við Skákgreind ehf. sem sérhæfa sig á þessu sviði, geta komið að því að veita þjálfunina.

Tekið er undir tillögur um samstarf m.a. Persónuverndar og fyrirtækja sem standa framanlega í nýtingu tækninnar.

Leggja ber áherslu á innlendar rannsóknir og þróun, sem tekur mið að íslenskum aðstæðum og menningu þ.a. gervigreind kerfi miðist við að þjóna íslenskum þjóðfélagsþegnum og taki fullt tillit til innlendra sérkenna og menningar.

Dæmi um misheppnaða innleiðingu gervigreindar er lýst í grein Hill í New York Times 2020 <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html> (þema það sama og í umsögn Persónuverndar):

Hér fá starfsmenn ekki nægilega fræðslu, þjálfunargögnin eru af litlum gæðum, svo virðist sem ódýrasta lausnin hafi verið keypt, sem inniheldur ekki mat á gæðum greiningar líkansins og það vantar vandaða verkferla, þar sem vel þjálfað starfsfólk vinnur með tækninni með vel skilgreindum hætti. Niðurstaðan er bæði óásættanleg gæði þjónustu og einnig er vegið að heiðri og virðingu starfsfólks, þ.e. lögreglumannanna og vinnustaðarins, þ.e. stofnunarinnar.

Til að nýta að fullu tækni gervigreindar í íslensku atvinnulífi þarf að gera akkúrat öfugt: leggja áherslu á þjálfun starfsmanna þ.a. þau geti nýtt tæknina með réttum hætti, leggja áherslu á vönduð þjálfunargögn, sem taka tillit til íslenskrar menningar og sérkenna og líta ekki framhá þeim sem minna mega sín. Einnig þarf vönduð gervigreindarlíkön, sem innihalda aðgengilegar upplýsingar um hvaða forkröfur þarf að gera til inntaksgagna þ.a. þau nýtist sem skildi s.s. myndgæði. Æskilegt er að líkönin hafi forvinnslu algorithmna, t.d. mat á því hvort andlitsmynd sé næginlega skýr, ásamt úttaki sem segir til um hversu mikil gæði inntak líkansins hafði og að líkanið meti annaðhvort að gæði myndarinnar séu ekki nægilega góð til að það geti skilað niðurstöðu eða fyrirvara og mati þar sem kemur fram hversu vist líkanið er í sinni sök.

Það að gera akkúrat öfugt við sýnidæmið hefur einnig í för með sér getu til að sjá ný tækifæri fyrir tæknina, taka frumkvæði í smíði nýrra lausna og stíga skref í vöruþróun.

Þegar þessi skilyrði eru uppfyllt þá mun gervigreind styðja við og veita starfsmönnum sem vinna eftir vönduðum verkferlum upplýsingar, sem þeir geta dregið ályktanir út frá og einnig mun tæknin nýtast í tengslum við sjálfvirkni undir vökulu eftirliti starfsmanna, sem þekkja takmörk tækninnar og vita hvað þarf að fylgjast sérstaklega með.

Stefna um gervigreind getur verið hornsteinninn að sameiginlegum gildum og fyrsta skref í átt að verkferlum og samráðsvettvangi, sem stenst tímans tönn.

Fyrir hönd Skákgreindar ehf.

Héðinn Steingrímsson

ceo@skakgreind.is

Heimildir:

Clark, Uraina S et. al, Specific impairments in the recognition of emotional facial expressions in Parkinson's disease *Neuropsychologia* 46 (2008) 2300–2309,
https://www.sciencedirect.com/science/article/pii/S0028393208001115?casa_token=1PAizj-UmM8AAAAA:ijd3MvVYc0QRNqneODkT7wGNbGLJ6ZWs-aPLqg6Q1OaNVK9W-pKrK0nxg9QGA5qMGb9tNbEO

Hargrave, Rita et. al, Impaired Recognition of Facial Expressions of Emotion in Alzheimer's Disease eftir, *J Neuropsychiatry Clin Neurosci* 2002; 14: 64–71.
<https://neuro.psychiatryonline.org/doi/pdf/10.1176/jnp.14.1.64>

Haun, Daniel et. al. The persistent sampling bias in developmental psychology: A call to action, *Journal of Experimental Child Psychology* 162 (2017) 31-38,
https://www.uni-muenster.de/imperia/md/content/psyifp/aekaertner/nielsen_m_haun_d_k_rtnr_j_le_gare_c_2017_the_persistent_sampling_bias_in_developmental_psychology.pdf

Hill, Kashmir, Wrongfully Accused by an Algorithm, New York Times, 2020,
<https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>

Kupferschmidt, Kai, Is the Western mind too WEIRD to study? Science 07/2019,
<https://www.sciencemag.org/news/2019/07/western-mind-too-weird-study>

Sonawane et. al., Review of automated emotion-based quantification of facial expression in Parkinson's patients, Vis Comput (2020), <https://link.springer.com/content/pdf/10.1007/s00371-020-01859-9.pdf>

Sprengelmeyer, R et. al., Facial expression recognition in people with medicated and unmedicated Parkinson's disease, Neuropsychologia 41 (2003) 1047–1057,
https://www.sciencedirect.com/science/article/pii/S0028393202002956?casa_token=XlIXExRXpioA AAAA:iFRLydYNne-PAu012uw7tIV2Yatp3muVX9Lj_bJMlp1XJutxBnHU_OpfZh1rPv9o8opoMc3

Umeda-Kameyama, Yumi et. al., AGING 2021, Vol. 13, No. 2,
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7880359/>