



Íslenskan okkar, alls staðar

Áætlun um íslenska máltækni 2.0



Stjórnarráð Íslands
Menningar- og viðskiptaráðuneytið

Unnið af stýrihóp um gerð næstu máltækniáætlunar, skipuðum af menningar- og viðskiptaráðherra í maí 2023. Hópin skipuðu Björgvin Ingi Ólafsson formaður, Lilja Dögg Jónsdóttir og Páll Ásgeir Guðmundsson. Með hópnum starfaði Óttar Kolbeinsson Proppé f.h. menningar- og viðskiptaráðuneytisins.

Ábyrgð: Menningar- og viðskiptaráðuneytið

Sölvhólgata 7 – 101 Reykjavík
545 9800 | mvf@mvf.is

Útgefandi: Stjórnarráð Íslands

Textavinnsla: Stjórnarráð Íslands

Umbrot: Einar Guðmundsson

Prófarkalestur: Örn Úlfar Sævarsson, Óttar Kolbeinsson Proppé

Myndir, ef annað er ekki tekið fram:

Unsplash og Freepik

Mars 2024

ISBN: 978-9935-9743-5-8

© 2024 – Menningar- og viðskiptaráðuneytið

Efnisyfirlit

Samantekt	5
1. Inngangur	9
1.1 Afmörkun efnistaka	9
1.2 Aðdragandi	11
1.3 Ábyrgð og stefna Íslands gagnvart íslenskri tungu	12
1.4 Framsækin máltækni til almennrar hagnýtingar	16
1.5 Viðtæk hagnýting	16
1.6 Áframhald á alþjóðlegri sókn	17
2. Rekstur máltækniáætlunar	21
2.1 Gervigreind og máltækni	22
2.2 Máltækni- og gervigreindarmiðstöð Íslands	23
2.3 Umsýsla næstu ára	26
2.4 Fyrirkomulag fjárveitinga til máltæknimála	27
3. Hagnýting máltækni	31
3.1 Fjármagn til hagnýtingarverkefna	31
3.2 Ráðgjöf til íslensks atvinnulífs og almenn kynning máltæknilausna	32
3.3 Samskipti við erlend tæknifyrirtæki	33
4. Kjarnaverkefni	37
4.1 Sveigjanleiki áætlunar og nýtt fyrirkomulag	37
4.2 Kjarnaverkefni og samkeppnissjónarmið	40
4.3 Kjarnaverkefni í forgrunni	40
4.4 Viðhald innviða og kjarnalausna	48

Samantekt

Fyrri máltækniáætlun stjórnvalda, sem lauk 2023, fól í sér metnaðarfulla framtíðarsýn um verndun íslenskrar tungu og gerði það að verkum að Ísland og íslenskan voru betur í stakk búin til að takast á við hraðar tæknibreytingar í gervigreind og máltækni en flest minni málsvæði.

Í stjórnarsáttmála núverandi ríkisstjórnar kemur fram að áfram verði unnið að því að styrkja stöðu íslenskunnar í stafrænum heimi með áherslu á máltækni. Í því skyni skipaði menningar- og viðskiptaáðherra stýrihóp í maí 2023 til að meta og greina með hvaða hætti máltækni sé best fyrir komið innan stjórnsýslunnar til framtíðar og leggja drög að næstu máltækniáætlun stjórnvalda.

Stýrihópurinn vann í upphafi með starfshópi um máltækni, sem menningar- og viðskiptaráðherra skipaði samhliða stýrihópnum. Í honum sátu 12 fulltrúar máltækniyrirtækja, háskóla, stofnana og samtaka þeirra sem máltækni getur nýst í daglegu lífi auk meðlima stýrihópsins. Hópurinn átti fjóra fundi á skipunartíma sínum þar sem tókst að draga fram mikilvæg sjónarmið hagsmunaaðila og áhersluatriði þeirra fyrir næstu máltækniáætlun. Auk þess var verktími fyrir máltækniáætlunar krufinn með það fyrir augum að læra af reynslunni og greina tækifæri til úrbóta fyrir næstu skref máltækni á Íslandi. Reyndist það afar gagnleg undirbúningsvinna fyrir gerð þeirra tillagna sem hér verða lagðar fram.

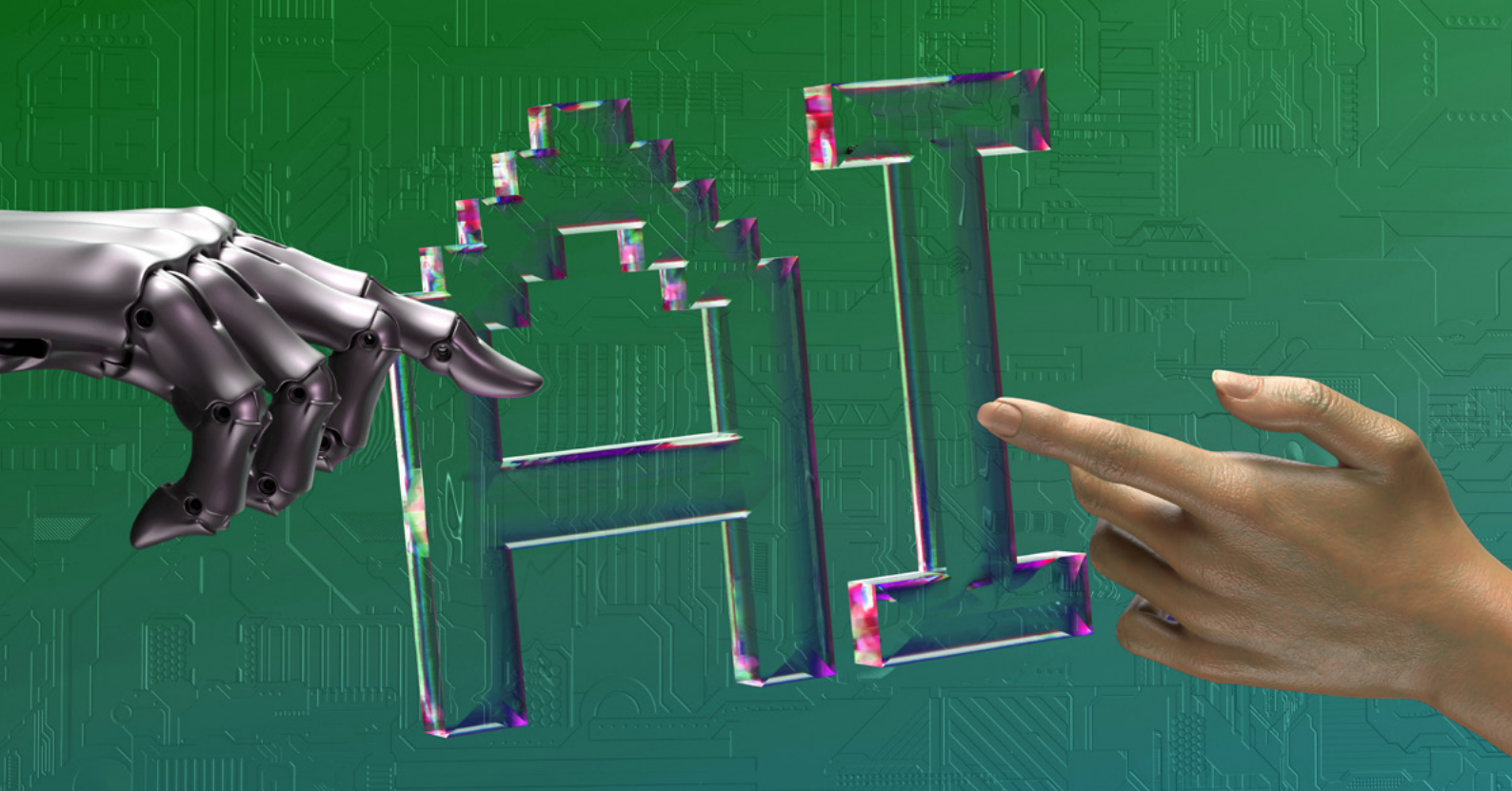
Í skýrslunni er fjallað um þrjú meginviðfangsefni:

Viðhald og áframhaldandi þróun innviða sem þróaðir voru á verktíma fyrir máltækniáætlunar stjórnvalda.

Hagnýting og innleiðing máltækni í íslensku atvinnulífi, stjórnsýslu og samfélagi. Sérstök áhersla er lögð á **kynningu á íslenskri máltækni** með það að markmiði að koma íslenskum máltæknilausnum að í hugbúnaðarþróun, bæði hérlendis og hjá erlendum tæknifyrirtækjum.

Sambætting málefna gervigreindar og máltækni til að styrkja samkeppnisstöðu Íslands og auka velsæld í samfélaginu með tungumálið í forgrunni.

Í skýrslunni eru tillögur að ráðstöfun fjármuna til máltækni á árunum 2024-2026 til að ná þeim markmiðum og áherslum sem stýrihópurinn leggur til. Einnig er lagt til að hafin verði vinna við að koma á framtíðarfyrirkomulagi sameiginlegrar umgjarðar í íslenskri stjórnsýslu fyrir máltækni og gervigreind.



Mikilvægt er að heildarstefnumörkun í málefnum máltækni og gervigreindar verði gerð fyrir lok árs 2024 og æskilegast er að slík stefna verði samþykkt á Alþingi.

Taka þarf afstöðu til þess hvort koma eigi á fót stofnun um máltækni og gervigreind að norrænni fyrirmynd og hvernig fjármögnun slíkrar stofnunar yrði. Að auki er nauðsynlegt að meta hvernig málefnasvið gervigreindar og máltækni verði betur samþætt innan stjórnsýslu.

Stýrihópur leggur til að menningar- og viðskiptaráðuneyti leiti samstarfs við háskóla-, iðnaðar- og nýsköpunarráðuneyti um að hefja vinnu við slíka stefnumörkun með fýsileikakönnun um nýja máltækni- og gervigreindarstofnun. Mikilvægt er að finna þessum málefnum sem hentugasta heimilisfesti og er lagt til að á tímabilinu 2024-2026 ráðist menningar- og viðskiptaráðuneytið í eftirfarandi:

1

Máltækni- og gervigreindarsetur:

Menningar- og viðskiptaráðuneytið leiti samstarfs við háskóla-, iðnaðar- og nýsköpunaráðuneyti um að kanna fýsileika þess að stofna nýja máltækni- og gervigreindarstofnun, með hliðsjón af norrænum fyrirmyndum. Fýsileikakönnunin verði gerð á árinu 2024 og í kjölfarið vinni ráðuneytin saman að því að leggja fyrir Alþingi stefnumörkun í málefnum gervigreindar og máltækni til næstu tíu ára.

2

Hagnýtingarstyrkir:

60 m.kr. verði varið árlega í hagnýtingarverkefni í máltækni. Styrkirnir miði að því að stuðla að þróun endalausna sem nýta máltækniinnviði sem þróaðir hafa verið undir máltækniáætlunum stjórnvalda en einnig að innleiðingu þeirra lausna í kerfi og þjónustu fyrirtækja.

3

Kynningarstarf og ráðgjöf:

Stóraukin áhersla verði lögð á kynningar- og ráðgjafastarfsemi fyrir máltækni á Íslandi. Kynningarstarf verður aukið til muna á þeim innviðum sem byggðir hafa verið upp fyrir íslenska máltækni og tilheyrandi tækifærum fyrir íslenskt atvinnulíf, félagasamtök og almenning. Einnig verði fylgst á markvissan hátt með tækniþróun erlendra tækniyrirtækja og unnið að því að koma íslenskri máltækni að innan þeirra.

4

Kjarnaverkefni:

A.m.k. 100 m.kr. verði varið árlega í áframhaldandi þróun kjarnaverkefna í máltækni á árunum 2024-2026. Þau verkefni sem lögð eru til hér í skýrslunni verði endurmetin árlega, undir nýju fyrirkomulagi verkefnastjórnar áætlunarinnar, í samræmi við hraðar tæknibreytingar.

5

Nýtt viðhaldsfyrirkomulag:

Nýtt viðhaldsfyrirkomulag fyrir máltækniinnviði verði sett á laggirnar og 40 m.kr. verði varið í viðhald árlega út árið 2026. Verkefnastjóri haldi utan um fyrirkomulagið og úthluti fjármunum til verkefna eftir þörfum í samstarfi við verkefnastjórn.

6

Stjórnvöld leiðandi í innleiðingu máltækni:

Stjórnvöld verði leiðandi í innleiðingu máltæknilausna í þjónustu sinni, bæði innan ráðuneyta og undirstofnana þeirra.

7

CLARIN-samstarf:

Þátttaka Íslands í CLARIN-samstarfinu, rannsóknarinnviðaverkefni á vegum Evrópusambandsins, verði endurmetin og úttekt gerð á umfangi þess og kostnaði við samstarfið, sem nú er hýst hjá Árnastofnun samkvæmt samningi til ársins 2026. Úttektin verði gerð samhliða vinnu ráðuneytisins við stefnumörkun í máltækni- og gervigreind sem tæki við eftir að verktíma þessarar áætlunar lyki árið 2026.

Kaflí 1



1. Inngangur

1.1 Afmörkun efnistaka

Með skipunarbréfi, dags. 15. maí 2023, voru Björgvin Ingi Ólafsson, sem formaður, Lilja Dögg Jónsdóttir og Páll Ásgeir Guðmundsson skipuð í stýrihóp um gerð næstu máltækniáætlunar. Var hópnum gert að skila tillögum til menningar- og viðskiptaráðherra að fyrirkomulagi framkvæmdar máltækniverkefna, með hliðsjón af hlutverki stjórnvalda og að forgangsraða máltækniverkefnum fyrir tímabilið 2023-2026. Helstu markmiðin eru að tryggja að áframhaldandi framþróun eigi sér stað í þróun íslenskra máltæknilausna eftir að verktíma fyrri máltækniáætlunar, *Máltækni fyrir íslensku 2018-2022*, lauk auk þess að tryggja að þeir innviðir sem smíðaðir hafa verið undir þeirri áætlun komist í aukna og almenna notkun meðal almennings og atvinnulífs. Með því er aukin áhersla sett á hagnýtingu máltækni.

Í framhaldinu var einnig óskað eftir því að stýrihópurinn tæki út fyrirkomulag umsýslu fyrri máltækniáætlunar og gerði tillögu að fyrirkomulagi til lengri tíma auk þess sem sérstaklega yrði horft til þess hvernig tryggja mætti nýtingu þess fjármagns sem gert var ráð fyrir í fjárlögum ársins 2023.

Samhliða skipun stýrihóps skipaði ráðherra starfshóp um íslenska máltækni sem átti að vera stýrihópnum til ráðgjafar og samráðs eftir þörfum í upphafi ferlisins og sem samráðsvettvangur fyrir skýrslugerð stýrihópsins. Í starfshópi sátu fulltrúar 12 hagsmunaaðila; máltækniyfirtækja, háskóla, ráðuneyta, opinberra stofnana og félagasamtaka auk þriggja meðlima stýrihópsins. Björgvin Ingi Ólafsson var einnig skipaður formaður starfshópsins og leiddi fundi hans.

Starfshópur átti fjóra fundi á tímabilinu maí-ágúst 2023 auk þess sem sendir voru út spurningalistar til allra þeirra sem sæti áttu í hópnum þar sem kallað var eftir sjónarmiðum um ýmsa þætti sem taka þyrfti afstöðu til í framhaldinu. Auk þess átti stýrihópurinn fjölda funda með hagsmunaaðilum, erlendum sérfræðingum og fulltrúum stjórnsýslunnar. Þar tókst að draga fram mikilvæg sjónarmið hagsmunaaðila varðandi næstu máltækniáætlun. Auk þess var verktími fyrri máltækniáætlunar krufinn með það fyrir augum að greina hvað hefði betur mátt fara og hvar helstu tækifæri lægju til úrbóta fyrir næsta áfanga máltækni vinnu á Íslandi.

Við vinnu starfshópsins kom í ljós að sjónarmiðin voru margþætt og ýmis álitafni til staðar. Hagsmunaaðila greindi nokkuð á um ýmsa áhersluþætti og hvaða leiðir væru skynsamlegastar við mótun framtíðarfyrirkomulags íslenskrar máltækni. Þá má einnig nefna að þeir hagsmunaaðilar sem kallaðir voru að vinnunni með skipun starfshópsins höfðu haft mismikla aðkomu að máltækni málum áður. Það átti einna helst við fulltrúa ráðuneytanna í hópnum.

Helstu áskoranir:

1

Heimilisfesti máltækni og gervigreindar

Samtvinna þarf heimilisfesti gervigreindar og undirþáttar hennar, máltækni.

2

Skýr verkskipting

Samþætta þarf tæknilega þróun, hagnýtingu, kynningarstarf og samskipti við erlend fyrirtæki og stjórnvöld. Skilgreina þarf betur hlutverk mismunandi aðila en gert var í síðustu áætlun. Misjafnir þættir kalla á ólíka hæfni og er mikilvægt að tryggja yfirsýn yfir fjölbætt verkefni undir einni verkstjórn sem ber ábyrgð á framkvæmd máltækniáætlunar.

3

Máltækni innan stjórnsýslu

Efla þarf þekkingu innan stjórnsýslunnar á máltækni. Vilji og umboð fagráðuneyta til að styrkja eða vinna markvisst að innleiðingu máltæknilausna í opinberri þjónustu er ekki nógu ljós.

4

Ný tækni

Misjöfn sjónarmið um möguleika gervigreindar og stórra mállíkana og hverju sú tækni breytir fyrir íslenska máltækni. Einnig eru skiptar skoðanir um hver þörfin er fyrir áframhaldandi þróun innlendra lausna sem gerðar eru frá grunni þegar aðlögun erlendrar tækni gæti skilað betri árangri og kostað minna.

5

Samkeppni tryggð

Mikilvægi samkeppnissjónarmiða og þess að hið opinbera keppi ekki við innlend tæknifyrirtæki með þróun kjarnaverkefna, sem eru alfarið fjármöggnuð af hinu opinbera. Samhliða því þarf að skilgreina hversu langt hugtakið kjarnaverkefni nær og hvenær hagnýting atvinnulífs tekur við.

6

Víðhald innviða

Tryggja þarf víðhald á máltækniinnviðum sem hafa verið þróaðir og ákveða hvaða lausnum hið opinbera þarf nauðsynlega að viðhalda.

1.2 Aðdragandi

Máltækniáætlun stjórnvalda sem kom út í júní 2017, *Verkáætlun máltækni fyrir íslensku 2018-2022*, markaði tímamót í þróun máltækni á Íslandi. Með henni var fjárfest í smíði innviða máltækni með myndarlegu fjárframlagi hins opinbera og markvissu átaki sérfræðinga undir merkjum Samstarfs um íslenska máltækni (SÍM). SÍM var hópur íslenskra háskóla, stofnana, fyrirtækja og félagasamtaka sem vann að rannsóknum og þróun innan áætlunarinnar samkvæmt samningum við sjálfseignarstofnunina Almennaróm. SÍM vann að þróun kjarnaverkefna í máltækni en Almennarómur fór með verkefnastjórn fyrir hönd ráðuneytis. Innan máltækniáætlunarinnar voru skilgreind 6 kjarnaverkefni, sem hafa lagt grunninn að þróun máltæknilausna fyrir íslensku. Um var að ræða kjarna- verkefni í eftirtöldum flokkum, sem voru gefin út í opnum hugbúnaðarhirsalum:

Málföng

Öll máltækni byggist á málföngum: textum og/eða hljóðupptökum. Þau eru nauðsynleg við greiningu máls, söfnun orðaforða og til að finna reglur og mynstur í tungumálinu. Út frá málgögnum er þannig hægt að „kenna“ tölvum það sem máli skiptir fyrir þann hugbúnað sem verið er að þróa, eða láta hugbúnað finna reglur og mynstur í miklu magni gagna. Innan máltækniáætlunar hefur verið unnið að stórum textasöfnum og þau undirbúin fyrir notkun í máltækni, bæði einmála íslensk textasöfn og tvímála samhliða málheildir sem innihalda íslenska og enska texta. Einnig fara fram upptökur á tali í miklum mæli, bæði í gegnum lýðvirkjunarverkefnið Samróð og hágæða upptökur í hljóðveri fyrir þróun talgervla. Jafnframt hefur verið unnið að gagnasöfnum sem geyma upplýsingar um einstaka þætti tungumálsins eins og orðaforða, framburð og merkingu.

Málrýni

Málrýni aðstoðar við að leiðrétta texta og skrifa rétt, jafnvel með viðeigandi mál-sniði. Málrýni hefur einnig brýnu hlutverki að gegna við þróun annars máltækni-hugbúnaðar þar sem villur í texta geta haft áhrif á frekari sjálfvirka vinnslu hans. Markmið máltækniáætlunar hefur verið að þróa almennan málrýni sem ræður við að finna og leiðrétta algengustu villurnar sem gerðar eru. Markmiðið er að til verði þekking á eðli ritvillna hjá mismunandi hópum og að aðferðir verði þróaðar til þess að laga kerfið að mismunandi þörfum, m.a. með tilliti til þjálfunar og kennslu.

Talgerving

Talgerving breytir rituðum texta í talað mál. Tvö meginvið talgervingarhugbúnaðar eru upplestur og (radd)samskipti. Talgervlar eru notaðir til þess að lesa texta, til dæmis af vefsíðum eða jafnvel heilu bækurnar. Fólk sem af einhverjum ástæðum getur ekki lesið sjálft eða á í erfiðleikum með það treystir á talgervlatækni í daglegu lífi. Samskipta-kerfi, þar sem talgreinir nemur það sem notandi segir, þurfa á talgervlum að halda til

Þess að hægt sé að svara með rödd. Innan máltækniáætlunar hefur verið lögð áhersla á að þróa nýjar talgervilsraddir fyrir íslensku, m.a. til þess að notendur geti valið rödd sem þeim þykir þægileg áheyrnar.

Talgreining

Talgreining snýst um að breyta töluðu máli í ritmál. Hún er forsenda þess að við getum átt samskipti við tölvur og tæki með þeim hætti sem flestum er eðlilegast: Með því að tala. Markmið máltækniáætlunar hefur verið að til verði almennur íslenskur talgreinir sem megi nota með vefþjónustu. Allar aðferðir og gögn verði einnig aðgengileg sem grunnur fyrir þróun sérhæfðra talgreina.

Vélpýðingar

Vélpýðingar eru sjálfvirkar pýðingar milli tungumála. Þær eru nú þegar orðnar gagnlegar fyrir ýmis tungumálapör, bæði til þess að hjálpa fólki við að átta sig á innihaldi texta á tungumáli sem það er ekki læst á og til þess að flýta fyrir vinnu þýðenda við tungumál sem þeir eru sérfræðingar í. Enn sem komið er ræður þó enginn pýðingarhugbúnaður við að skila pýðingum sem eru nálægt fullnægjandi gæðum, alltaf þarf að yfirfara texta og laga ef pýðing á að vera nákvæm. Markmið máltækniáætlunar hefur verið að til verði opin pýðingarvél sem þýðir milli íslensku og ensku. Hún á að gagnast við pýðingar á ákveðnum sviðum svo að þýðendur geti fullunnið texta hraðar.

1.3 Ábyrgð og stefna Íslands gagnvart íslenskri tungu

Markmið í málefnum íslenskrar tungu

Í stjórnarsáttmála núverandi ríkisstjórnar er sérstaklega fjallað um öra tækniþróun og stafræna umbreytingu sem á sér stað á heimsvísu. Þar er kveðið á um að á yfirstandandi kjörtímabili verði lögð áhersla á að styrkja enn frekar tæknilega innviði auk þess sem sérstök áhersla er lögð á nýsköpun og hagnýtingu hugvits.

Þar segir meðal annars:



Þá verður áfram unnið að því að styrkja stöðu íslenskunnar í stafrænum heimi með áherslu á máltækni.

Auk þess er í sáttmálanum að finna fyrirheit um að hagnýta nýja tækni og gervigreind og styðja sérstaklega við umbreytingar í menntakerfinu en einnig í heilbrigðis- og lífvísindum, hugbúnaðarþróun og grænni tækni og skapandi greinum.

Einnig segir í lögum um stöðu íslenskrar tungu og íslensks táknmáls nr. 61/2011 að:



Ríki og sveitarfélög bera ábyrgð á að varðveita og efla íslenska tungu og skulu sjá til þess að hún sé notuð.

Þegar þessi skýrsla er rituð hefur tillaga til þingsályktunar um aðgerðaáætlun í málefnum íslenskrar tungu fyrir árin 2023-2026 verið lögð fyrir þingið. Sú tillaga er afrakstur samstarfs fimm ráðuneyta þeirra ráðherra sem sæti eiga í ráðherranefnd um íslenska tungu. Aðgerð 16. *Framtíð máltækni* kveður á um að



...unnið verði að mótun nýrrar máltækniáætlunar sem miðli forgangsröðun stjórnvalda um áframhaldandi þróun, viðhald og innleiðingu máltæknilausna. Áhersla í máltækni verkefnum til ársins 2026 verði m.a. á lausnir og verkefni sem nýtast almenningi.



Íslensk málstefna

Ný íslensk málstefna 2021-2030 var gefin út af íslenskri málnefnd í september 2021. Í henni er sérstaklega fjallað um máltækni og mikilvægi hennar fyrir framtíð íslenskrar tungu ítrekuð:



Eigi íslenska áfram að vera opinbert mál og þjóðtunga landsmanna þarf nauðsynlega að fylgja eftir þróun í máltækni. Slíkt verkefni má ekki vera áttak til skamms tíma heldur verður að líta á það sem verkefni sem aldrei lýkur heldur þróast í takt við hraða tæknipróun.

Í málstefnunni er einnig kveðið á um að stjórnvöld haldi úti máltækniáætlun, styðji þróunarverkefni og sprotafyrirtæki í máltækni, kynni íslenska máltækni og hvetji landsmenn til að nýta sér hana.

Fjórða iðnbyltingin og gervigreindarstefna

Um mitt ár 2018 skipaði forsætisráðherra nefnd um fjórðu iðnbyltinguna sem gert var að „greina frá umræðu um fjórðu iðnbyltinguna á alþjóðavettvangi, afleiðingum hennar fyrir íslenskt samfélag og tækifæri Íslands í þessum breytingum“. Rauður þráður í niðurstöðu nefndarinnar er mikilvægi þess að íslenskt samfélag taki tæknibreytingum tveim höndum og vinni út frá þeirri forsendu að áhrif tækninnar fari eftir því hvernig mannfólkið notar hana. Einnig lagði nefndin mikla áherslu á að tryggja þyrfti að framleiðni- aukning af aukinni sjálfvirknivæðingu dreifðist með sanngjörnum hætti í samfélaginu.

Það að íslenska sé notuð í nýrri tækni er einmitt lykilforsenda þess að svo megi verða eins og kemur svo skýrt fram í aðgerðaráætlun sem unnin var á grundvelli skýrslu nefndarinnar af verkefnastjórn sem forsætisráðherra skipaði í júlí 2019. Tvær aðgerðir af 27 í aðgerðaáætluninni fjalla sérstaklega um íslenska máltækni:

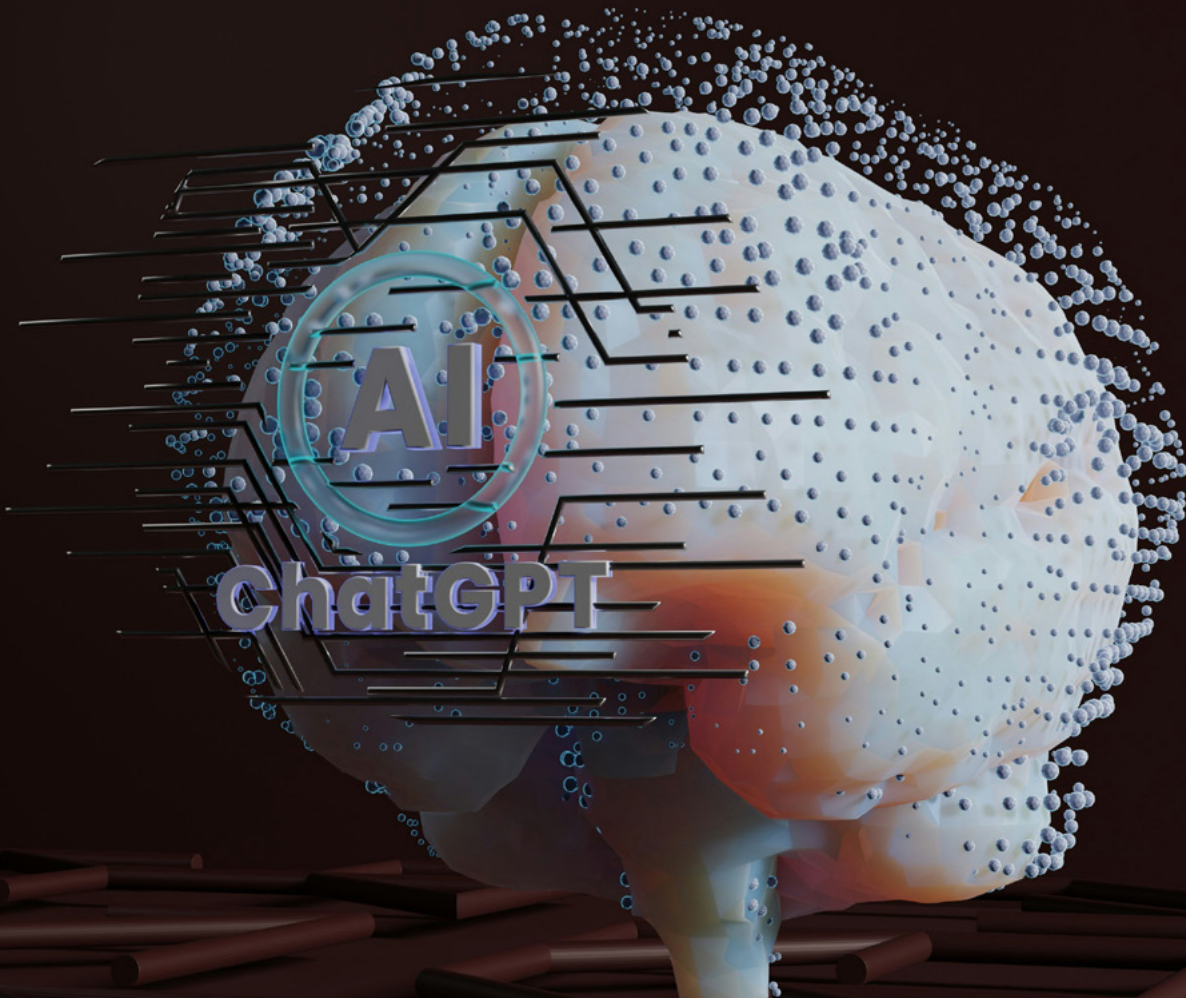


[Aðgerð 26] Tryggt verði að íslenska verði gjaldgeng í stafrænum heimi. Verkefnisstjórnin leggur til að gerð sé áætlun um hvað taki við eftir að 5 ára máltækniáætlun lýkur. [...]

[Aðgerð 27] Unnið verði að því að búa til sjálfvirkt þýðingarkerfi á milli ensku og íslensku, í því skyni að hægt verði að texta efni á íslensku hraðar og með minni tilkostnaði. Sjálfvirkt þýðingarkerfi væri einnig nothæft á stafrænum veitum.

Önnur aðgerð í áætluninni varðaði mikilvægi þess að Ísland markaði sér stefnu í málum gervigreindar en slíkt hafði ekki áður verið gert hér á landi. Í október 2020 skipaði forsætisráðherra nefnd um ritun gervigreindarstefnu sem skilaði tillögum til ráðherra í apríl 2021. Þar er fjallað sérstaklega um íslenska tungu og mikilvægi rannsóknar í máltækni til að tryggja að hægt verði að vinna með gögn á íslensku í gervigreind.

Við gerð þessarar skýrslu voru ofangreindar stefnur og þingsályktanir hafðar til hliðsjónar og samræmast þær tillögur sem hér eru settar fram markmiðum ríkisstjórnarinnar og samþykktum Alþingis, bæði í málefnum íslenskrar tungu og gervigreindar.



1.4 Framsækin máltækni til almennrar hagnýtingar

Máltækniáætlun stjórnvalda, sem lauk 2023, var metnaðarfull framtíðarsýn um verndun íslenskrar tungu og gerði það að verkum að Ísland og íslenskan voru í betur stakk búin til að takast á við þessa umbreytingu en flest minni málsvæði.

Með máltækniáætluninni var tryggð smíði innviða, sem eru nauðsynlegir til að hlúa að íslensku máli í stafrænum heimi og um leið treystar stoðir þess að íslenskan verði okkar mál til allrar framtíðar.

Á verk tíma fyrri máltækniáætlunar, *Máltækni fyrir íslensku 2018-2022*, lögðust tugir sérfræðinga á eitt við að smíða opnar kjarnalausnir undir forystu SÍM, Samstarfs um íslenska máltækni. Þegar litið er til baka var áætlunin farsæl og markaði Íslandi eftirtektarverð spor í þróun máltækni og sýndi hvernig minni tungumál geta gert sig gildandi í stafrænum heimi.

Í þeim áfanga máltækniáætlunar sem hér er lagður fram er lögð rík áhersla á hvernig tryggja megi sem best hagnýtingu þeirra innviða sem smíðaðir hafa verið hjá fyrirtækjum, stofnunum og sveitarfélögum hér á landi en jafnframt meðal alþjóðlegra fyrirtækja þar sem er mikilvægt að íslenskan sé á meðal tungumála í tæknilausnum þeirra. Að auki er lögð áhersla á viðhald og áframhaldandi þróun þeirra innviða sem hafa verið smíðaðir.

Stefnumótun málefna sem snerta tækni er um margt snúin enda er þróunin oftari en ekki hröð og oft ófyrirséð. Á síðustu árum hefur vægi stórra mállíkana og möguleikar þeirra orðið langt umfram það sem búist var við í fyrri máltækniáætlun. Búast má við að umbylting tækninnar haldi áfram en hvernig hún mun þróast er óljóst sem fyrr. Þó má spá því að leysa megi margþætt vandamál máltækni á grunni stórra mállíkana og með víðtækari hætti en áður var mögulegt. Um þetta er þó mikil óvissa. Þróunin undirstrikar mikilvægi þess að stefnumótun sem þessi sé bæði framsýn og skýr en þó ekki greytt í stein og geti tekið breytingum í takt við breyttar aðstæður.

1.5 Víðtæk hagnýting

Mikilvægt er að næsta máltækniáætlun leggi megináherslu á hagnýtingu þeirra innviða máltækni sem þegar hafa verið þróaðir. Ný máltækniáætlun felur því í sér að hagnýting máltækni í víðum skilningi verði nú í meiri forgrunni samhliða viðhaldi og þróun máltæknilausna. Lögð verður meiri áhersla en áður á kynningu á þeim innviðum og lausnum sem til eru og verða þróaðar til þess að hagnýting verði eins mikil og kostur er. Sérstök áhersla verður lögð á hagnýtingu máltækni innan menntakerfisins og heilbrigðiskerfisins því ljóst er að máltækni og önnur gervigreindardrífni tækniþróun mun geta haft margvísleg jákvæð áhrif á bæði kerfin.

Þá verður að auki lögð mikil áhersla á að viðskiptadrifin nýsköpun eigi sér stað á grunni þeirra máltækniinnviða sem hafa verið smíðaðir. Nú þegar hafa sprotafyrirtæki kynnt til sögunnar máltæknilausnir sem byggja að hluta á þeim innviðum sem hafa verið smíðaðir auk þess sem stærri fyrirtæki hafa markað sér stefnu um hagnýtingu máltækni og ráðið til sín sérfræðinga til að drífa áfram frekari þróun máltækni innan fyrirtækjanna. Þó er ljóst að enn eru mikil ónýtt tækifæri hjá fyrirtækjum og stofnunum við innleiðingu og nýtingu máltæknilausna. Mikilvægt er að stjórnvöld stuðli að frekari framgangi hagnýtingar máltækni hjá fyrirtækjum bæði með skilvirkari upplýsingagjöf um þá innviði og lausnir sem til staðar eru auk beins stuðnings við innleiðingar og hagnýtingarverkefni í atvinnulífinu.

1.6 Áframhald á alþjóðlegri sókn

Á síðustu árum hefur máltækniþróun Íslendinga meðal annars falist í því að kynna þá innviði sem hafa verið smíðaðir fyrir stórum erlendum tæknifyrirtækjum. Lögð hefur verið áhersla á að einfaldara sé að gera íslensku hluta af tungumálamengi alþjóðlegra lausna vegna þeirra innviða sem hafa verið smíðaðir á Íslandi. Haldnar hafa verið kynningar fyrir leiðandi tæknifyrirtæki á borð við Microsoft, Amazon, Apple, Google og Meta. Þrátt fyrir að vel hafi gengið að ná eyrum margra þessara fyrirtækja þarf að vinna með markvissari hætti að því að tryggja að íslenskan verði hluti af framboði tungumála. Bæði eru dæmi um nýlegar lausnir þar sem íslenska skipar veglegan sess og eins dæmi um lausnir þar sem íslenskan er fær ekki að vera með.

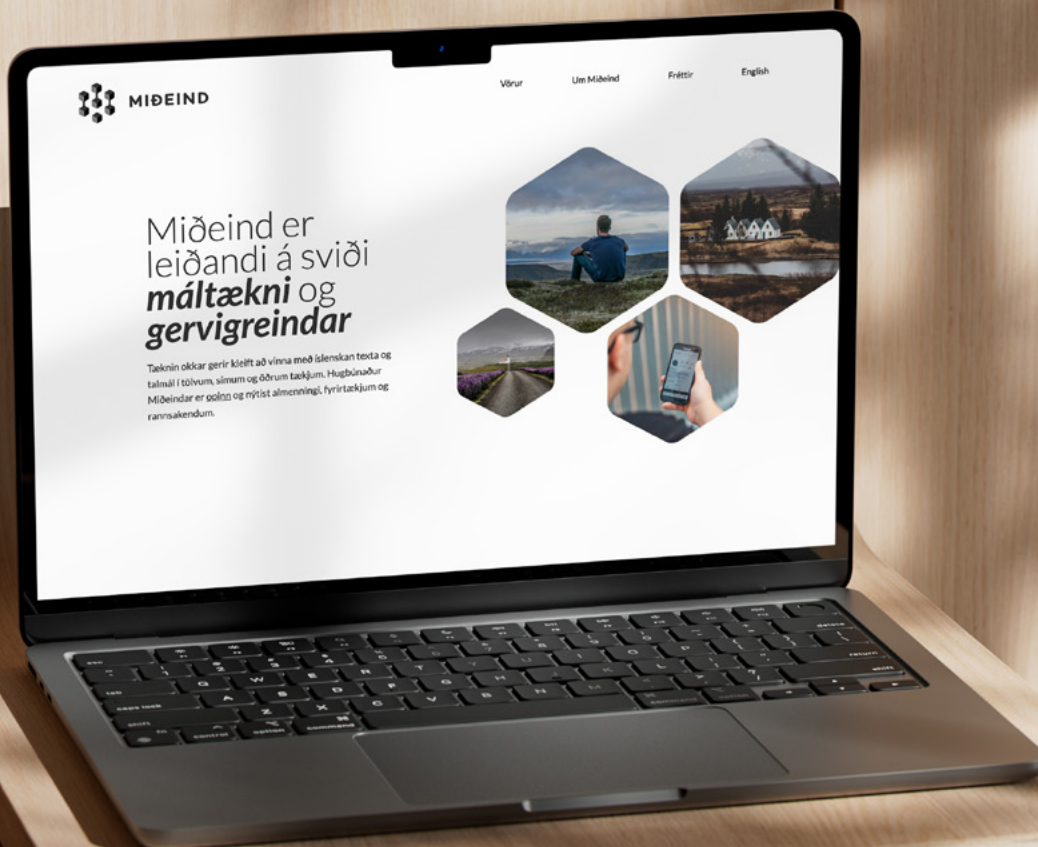
Með kynningu á afurðum fyrri máltækniáætlunar myndaðist sterkt samband við Open AI, höfund mállíkansins Chat GPT. Þegar Chat GPT-4 kom út í nóvember 2023 var sérvalið íslenskt gagnasafn felld inn í mállíkanið. Þannig urðu miklar framfarir á öllum íslenskum texta, bæði hvað varðar málfræði en þó helst í skilningi líkansins á íslensku máli og getu til að þýða íslenskan texta yfir á ensku.

Frá því að Chat GPT kom fyrst fram á sjónarsviðið hefur skilningur líkansins á íslensku aukist stórkostlega með virku samstarfi við íslenska máltæknifyrirtækið Miðeind. Hæfni Chat GPT til að skrifa gott íslenskt mál hefur stöðugt batnað og að sama skapi hafa möguleikar til þess að smíða frekari lausnir á grunni þess á íslensku aukist. Úrbætur og framfarir í íslenskuhæfni líkansins sýna vel hve mikilvægt er að fjárfesta í þjálfun slíkra mállíkana með þróun á þjálfunar- og prófunargögnum. Að sama skapi sýnir samstarfið við Open AI að það er eftir miklu að slægjast ef viðlíka árangur næðist í samstarfi við aðra, til dæmis Google, sem nýlega sendi frá sér mállíkanið Gemini.

Í nýrri máltækniáætlun þarf að tryggja verulegar fjárfestingar í að halda áfram að byggja upp samband við erlend tæknifyrirtæki og tryggja að íslenskar máltæknilausnir séu þekktar innan fyrirtækjanna og þau skilji hversu auðvelt er að gera íslensku að

hluta tungumálaframboðs í lausnum þeirra. Einnig verður að tryggja að íslenskan verði hluti af evrópskum samstarfsverkefnum í máltækni, sem fylgt geta stórir styrkir frá Evrópusambandinu. Styrkur íslensks máltækniíðnaðar og gildi á alþjóðlegum vettvangi hafa þegar sannað sig vel. Í nóvember 2023 var tilkynnt að Háskóli Íslands og Miðeind hefðu hlotið stóran styrk úr Horizon-rannsóknaráætlun Evrópusambandsins vegna þátttöku í verkefni sem lýtur að gerð stórs gervigreindarmállíkans fyrir germönsk tungumál, þar á meðal íslensku. Með þátttöku í slíkum verkefnum getur falist mikill ávinningur fyrir íslenska máltækni og framtíð tungumálsins.

Í nýrri máltækniáætlun þarf að tryggja verulegar fjárfestingar í að halda áfram að byggja upp samband við erlend tæknifyrirtæki [...]



Kaflí 2



2. Rekstur máltækniáætlunar

Ávöxtur þessa starfs sést skýrt á þeim árangri sem náðist á verktíma fyrri máltækniáætlunar. Nú tekur við næsti fasi með breyttum áherslum og mikilvægt er að öll umgjörð og rekstur máltækni endurspegli þær breyttu áherslur.

Mikilvægt er að tryggja samfellu í íslensku máltækni starfi. Samstarf stofnana, háskóla og atvinnulífs og vel skilgreint hlutverk þeirra við vinnu að framgangi íslenskrar máltækni sannaði sig á verktíma fyrri máltækniáætlunar stjórnvalda. Samstarfið var tryggt með aðkomu sjálfseignarstofnunarinnar Almennaróms að verkefninu sem hélt utan um gerð samninga um þróun máltækniinnviða og lausna fyrir hönd ráðuneytis. Sérstakur vettvangur var myndaður um þróun kjarnaverkefna, Samstarf um íslenska máltækni (SÍM), sem leiddi vinnu við þróun skilgreindra kjarnaverkefna innan verk-áætlunarinnar Máltækni fyrir íslensku 2018-2022. Þá var sett á stofn fagråd þriggja erlendra sérfræðinga sem voru Almennarómi til ráðgjafar hvað varðar tæknilega útfærslu kjarnaverkefna. Fagráðið rýndi þannig bæði tæknilýsingar samstarfssamnings Almennaróms og SÍM í samhengi við máltækniáætlun og áfangaskýrslur SÍM áður en greiðslur vegna verkefna fóru fram. Almennarómur annaðist einnig regluleg skýrsluskil til ráðuneytisins um stöðu og framgang máltækniáætlunar.

Þetta fyrirkomulag reyndist í meginatriðum henta vel til að tryggja gæði og framgang auk þess að tryggja að allt fjármagn væri nýtt sem best.

Ávöxtur þessa starfs sést skýrt á þeim árangri sem náðist á verktíma fyrri máltækniáætlunar. Nú tekur við næsti fasi með breyttum áherslum og mikilvægt er að öll umgjörð og rekstur máltækni endurspegli þær breyttu áherslur.

2.1 Gervigreind og máltækni

Svið gervigreindar og máltækni skarast mikið og væri hvort um sig býsna takmarkað án hins. Til einföldunar má segja að gervigreind sé einskona regnhlífahugtak en máltækni svið sem að mestu leyti heyrir þar undir. Stefnumótun í þessum málaflokum hefur hingað til verið aðskilin þó ávallt sé tæpt bæði á máltækni og gervigreind í afurðum slíkrar vinnu. Í gildandi stefnu Íslenskra stjórnvalda um gervigreind frá árinu 2021 er til að mynda sérstaklega fjallað um máltækni og ítrekað mikilvægi áframhaldandi framþróunar á sviðinu.

Þar er einnig bent á að á Íslandi skorti samfellu og tengingu frá rannsóknum og námi annars vegar og að innleiðingu og notkun tækni gervigreindar hins vegar. Þetta er áhugavert og í samræmi við þær niðurstöður sem fjallað er um í þessari skýrslu og varða hagnýtingu máltækni.

Einnig er ljóst að umsýsla og umgjörð fyrir bæði máltækni og gervigreind á Íslandi er skemmra á veg komin en í mörgum þeirra landa sem við berum okkur saman við. Flest nágrannalönd okkar hafa komið sérstökum gervigreindarstofnunum á laggirnar sem gegna lykilhlutverki í þessu ferli. Athyglisvert er að stofnanirnar hafa allar einnig með máltækni að gera. Lykilatriði er að þessar stofnanir eru studdar sameiginlega af fyrirtækjum, háskólum og hinu opinbera.

Hugmyndin um að sambærileg stofnun verði sett á fót á Íslandi er ekki ný og er til dæmis meðal þess sem lagt er til í gervigreindarstefnu Íslands. Þar segir meðal annars:



Til þess nýtast svokallaðar hagnýtar grunnrannsóknir sem hafa það sérkenni að þær stjórnast af þörfum atvinnulífsins. Þær sinna hlutverki sem hvorki háskólar né atvinnuvegir einir geta sinnt: að yfirfæra tækni og hugmyndir á hagnýtanlegt form og yfirfæra lausnir úr einni iðngrein eða markaði yfir á annan. Yfirfærsla á tækifærum og tækni gervigreindar á milli atvinnugreina, markaða og starfssviða gerist ekki af sjálfu sér, og það sérstaklega ekki hjá smærri og meðalstórum fyrirtækjum sem hafa takmarkaðri sérhæfingu og fjármagn. Slík fyrirtæki eru meiri hluti fyrirtækja á Íslandi.

Þetta er í samræmi við þær lykiláherslur sem lagðar eru að grundvelli næstu máltækniáætlunar. Gagnlegt er að glöggva sig frekar á gervigreindarstofnunum nágrannalandanna til að sjá tækifærin sem geta falist í sterkri gervigreindar- og máltæknistofnun hér á Íslandi.

2.2 Máltækni- og gervigreindar- miðstöð Íslands

Við gerð þessarar skýrslu var sérstaklega horft til fyrirkomulags máltækni- og gervigreindar á Norðurlöndunum. Hópurinn fundaði með Alexandra Institute, danskri sjálfseignarstofnun sem sérhæfir sig í rannsóknum og hagnýtingu máltækni- og gervigreindarlausna, AI Sweden, sem fer með svipað hlutverk í Svíþjóð og norsku einingunni NorwAI. Sambærileg eining er til í Finnlandi.¹ Þessar stofnanir eru á mismunandi laga-grundvelli og fjármagnaðar hver á sinn hátt.

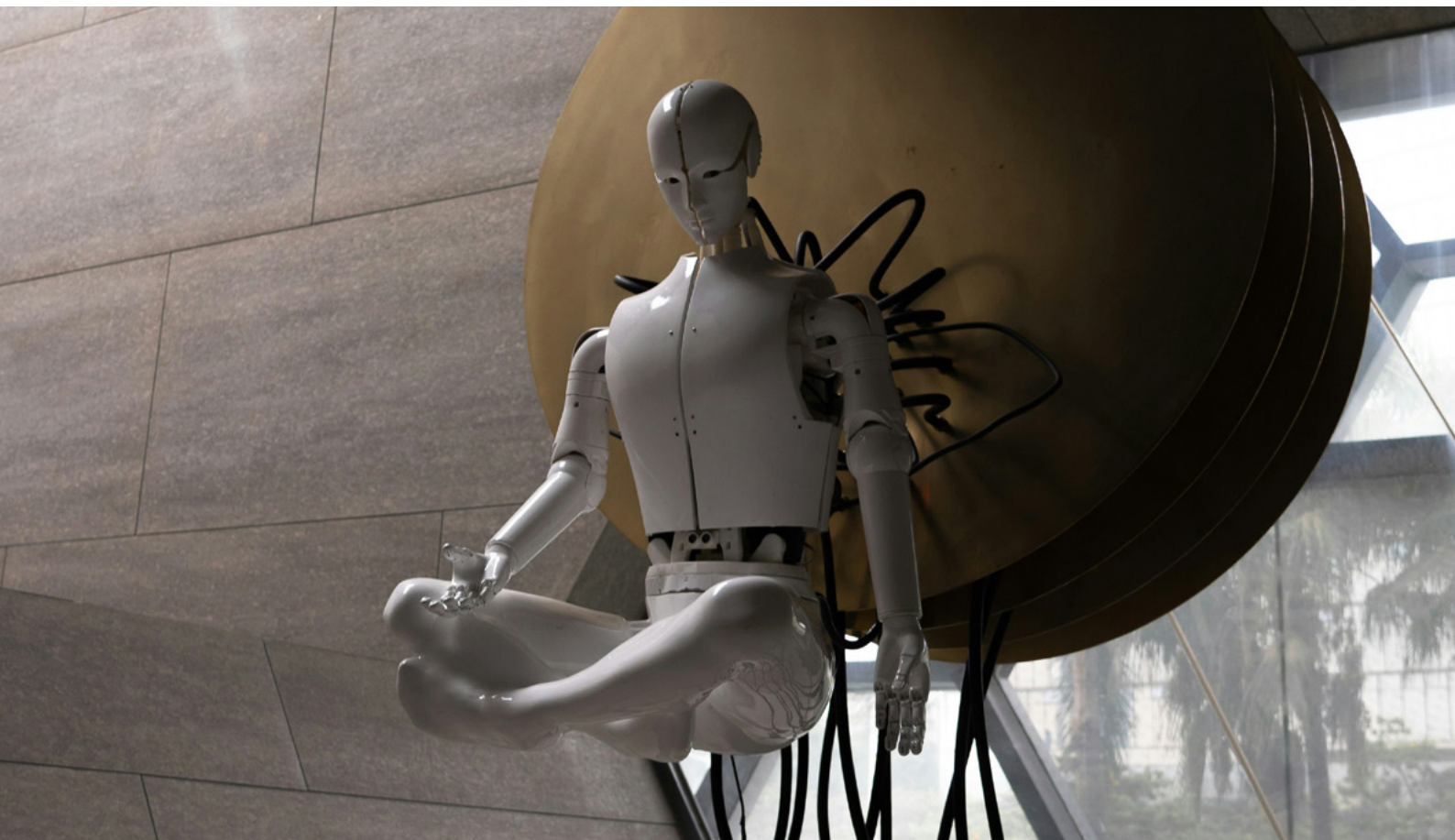
Alexandra Institute er ein sjö stofnana í Danmörku sem falla undir viðurkennda skilgreiningu mennta- og vísindamálaráðherra (d. uddannelses- og forskningsminister) sem svokallaðar RTO-stofnanir (e. Research and Technology Organisations). Þær eru allar óhagnaðardrífna sjálfseignarstofnanir sem gegna því hlutverki að sinna rannsóknum á nýjustu tækni og þróa tæknilausnir sem gagnast dönsku atvinnulífi. Alexandra er sú RTO-stofnun í Danmörku sem sérhæfir sig í máltækni- og gervigreindarþróun og -ráðgjöf til fyrirtækja.

Stofnunin er fjármögnuð að um þriðjungi til frá danska mennta- og vísindamálaráðuneytinu, að þriðjungi með styrkjum fyrir sérstök verkefni frá Evrópusambandinu og að þriðjungi frá dönsku atvinnulífi, í gegnum sameiginlega þróun á lausnum, rannsóknum og ráðgjöf við innleiðingu lausna. Stofnuninni er ætlað að auka samkeppnishæfni dansks atvinnulífs með því að bjóða því upp á nýjustu máltækni- og gervigreindarlausnir á dönsku, sem flestum fyrirtækjum væri ógerlegt að ráðast sjálf í að þróa vegna flókinna rannsókna og mikils kostnaðar. Með þessu móti býðst dönskum fyrirtækjum að innleiða lausnirnar með minni tilkostnaði þar sem þær eru þróaðar að stórum hluta til með aðkomu danska ríkisins og Evrópusambandsins. Lítil og meðalstór dönsk fyrirtæki nýta sér þjónustu RTO-stofnana mest. Á fundi hópsins með Jens Kaas Benner, sem fer fyrir Kaupmannahafnarskrifstofu Alexandra, greindi hann frá því að meginviðfangsefni stofnunarinnar væri að tryggja hagnýtingu máltækni- og gervigreindarlausna hjá dönskum almenningi og atvinnulífi, en þetta er helsta áskorun Íslands í dag.

AI Sweden, í Svíþjóð, er rannsóknarmiðaðri stofnun sem hefur aðra nálgun á innleiðingarhlutverk sitt en Alexandra. Stofnunin er einnig óhagnaðardrífna sjálfseignarstofnun en fellur ekki undir sambærilegt regluverk og RTO-stofnanir Danmerkur. Stofnunin er rekin með fjármagni sem kemur að helmingi til frá ríkisreknu nýsköpunarstofnuninni Vinnova en hinn helmingurinn kemur frá um 120 sænskum fyrirtækjum og háskólum, bæði stórum og smáum, sem eru aðilar að sjálfseignarstofnuninni AI Sweden. Stofnunin skilgreinir hlutverk sitt frekar að framleiða hagnýtar máltækni- og gervigreindarlausnir, sem síðan muni komast í notkun, en að einblína á að vinna sjálf að almennri innleiðingu þeirra eins og Alexandra Institute gerir.

¹ Hópnum tókst ekki að funda með finnsku stofnuninni Finnish Centre for Artificial Intelligence við gerð þessarar skýrslu en hefur þó hafið samtal við hana sem starfsmaður menningar- og viðskiptaráðuneytisins og tengiliður hópsins gæti tekið lengra.

Báðar stofnanirnar leiða saman rannsóknaraðila og þá sem á endanum eiga að nýta þær lausnir sem rannsóknirnar ganga út á að styðja við. Þær eru brúin milli innviða og endanotenda; vísindamanna og atvinnulífs. Þær styðja báðar við það markmið að máltæknilausnir lítilla tungumála á borð við dönsku og sænsku verði þróaðar og nýta til þess aðkomu ríkis, styrki vegna Evrópuverkefna og framlög frá atvinnulífnum. Þetta fyrirkomulag fjármögnunar er nauðsynlegt á tungumálasvæðum sem þjóna eins litlum markaðssvæðum og raun ber vitni á Norðurlöndunum öllum. Stofnanirnar eiga þá í nánú samtali við atvinnulífið og ráðast ekki í verkefni eða þróun lausna án þess að eiga víðtækt samtali við mögulega notendur og greiningu á þörf fyrir þær lausnir. Allt er þetta hýst á einum og sama stað; stofnanirnar ráða vísindamenn og rannsóknaraðila sem vinna við hagnýtar rannsóknir og tækniþróun auk þess sem þær selja lausnirnar áfram til endanotenda og eiga í góðu samstarfi við aðra rannsóknaraðila, á borð við háskóla, og atvinnulíf. Jafnframt eru allar lausnir sem þróaðar eru með þessu fyrirkomulagi opnar (e. open source) og geta þá nýst víðar í samfélaginu. Þessar stofnanir leiða einnig saman rannsóknir og þróun í máltækni annars vegar og gervigreind hins vegar, sem er æskileg leið til framtíðar vegna þess hve gervigreind og máltækni eru orðin samofin.



Þetta fyrirkomulag er ólíkt því sem var hérlendis á verktíma fyrri máltækniáætlunar þar sem ríkið skilgreindi fjögur ár fram í tímann hvaða verkefni máltækniáætlunin ætti að ráðast í og fól svo sjálfseignarstofnuninni Almennarómi að hafa milligöngu um þau verkefni og annast eftirlit með því að þeim væri sinnt. Tæknileg þróun þeirra verkefna fór síðan fram utan sjálfseignarstofnunarinnar, ýmist hjá háskólum eða máltækniáætlunum. Þetta þýddi að sú þekking sem varð til við þróunina býr meðal þessara háskóla og fyrirtækja sem og hinar eiginlegu lausnir sem þróaðar voru, þrátt fyrir að þær séu sannarlega gefnar út undir opnum leyfum. Þá hefur Almennarómi ekki haft það skilgreinda hlutverk að tryggja aðkomu atvinnulífsins að verkefnum eða að tryggja þróun lausna sem henta atvinnulífinu, í nánú samráði við það.

Jafnvel þó þetta fyrirkomulag hafi sem fyrr segir reynst ágætlega á verktíma fyrri máltækniáætlunar þá er það álit stýrihópsins að markmiðum næstu áætlunar verði betur þjónað með öðru móti.

Það er álit hópsins að æskilegt sé að stjórnvöld stefni að því að koma á fót sambærilegri gervigreindar- og máltækniáætlun á Íslandi og þekkist í öðrum Norðurlöndum og víðar í Evrópu. Helstu kostir þess væru meðal annars:

-
- Skýr heimilisfesti gervigreindar og máltækni á Íslandi. Þannig sé tryggð ábyrgð, samfella og stöðugleiki málaflökkanna beggja.
-
- Fyrirkomulag fjármögnunar sem kæmi að hluta frá hinu opinbera, að hluta frá fyrirtækjum og að hluta frá alþjóðlegum styrkjum væri áhrifaríkur hvati í átt að hagnýtingu tækninnar.
-
- Samnýting á mikilvægum aðföngum á borð við reiknigetu væru auðsóttari.
-
- Stöðug uppbygging þekkingar á sviði gervigreindar og máltækni og miðlun þeirrar þekkingar innan stofnunarinnar. Þetta er ekki síst mikilvægt í ljósi þess hve fáir starfa á sviðinu hér á landi og með þessu móti yrði til ákveðið „stofnanaminni“ í málaflöknum.
-

Ljóst er að margt þarf að skoða betur; hvaða fyrirkomulag henti best hér á landi, hvort ríkið eigi að hafa beina aðkomu að slíkri stofnun eða styrkja starfsemi sjálfseignarstofnunar áþekkri þeirri sem þekkist í Danmörku, kanna hug fyrirtækja á að koma með fjármagn inn í slíka stofnun og gera nákvæma athugun á lagaumhverfi fyrir slíkar stofnanir hér miðað við það sem þekkist í Norðurlöndunum. Þá er einnig ljóst að það er vandasamt verk sem mun taka sinn tíma að koma slíkri stofnun á laggirnar verði það niðurstaða eftir að allri greiningarvinnu lýkur. Með þetta í huga leggur stýrihópurinn til að menningar- og viðskiptaráðuneytið ráðist strax í að leita samstarfs við háskóla-, iðnaðar- og nýsköpunaráðuneytið, sem heldur utan um málefni gervigreindar, um að gera fýsileikakönnun á stofnun máltækni- og gervigreindarstofnunar á Íslandi, með hliðsjón af norrænum fyrirmyndum. Við könnunina verði gerð skilmerkileg grein fyrir umsýslu slíkra stofnana á Norðurlöndum, mögulegri hagræðingu og kostum þeim sem fengjust við að koma gervigreind og máltækni undir sama hatt, hvort sú stofnun geti einfaldað og tryggt betri sveigjanleika í þróun máltæknilausna fyrir endanotendur og mögulega tryggt fjármagn víðar að en úr ríkissjóði í verkefnum. Fýsileikakönnunin verði framkvæmd strax á næsta ári, árið 2024, og skilað til ráðuneytanna með tillögu að aðgerðum. Ráðuneytin tækju í kjölfarið ákvörðun um hvort ráðast ætti í að koma slíkri stofnun á fót og sameina þá vinnu við undirbúning að gerð næstu máltækniáætlunar.

2.3 Umsýsla næstu ára

Á meðan slík máltækni- og gervigreindarmiðstöð er ekki starfandi þarf þó að tryggj rekstur þeirra tillagna sem lagðar eru fram í skýrslunni til ársins 2026 og leggur stýrihópurinn til að þeim verkefnum verði útvístað til eins aðila eða stofnunar sem sinni eftirtöldu:

- 1 Verkefnastjórn yfir máltækniáætlun. Viðkomandi verði falið að sinna því hlutverki sem Almennarómur fór með á verk tíma síðustu áætlunar og halda þannig utan um gerð samninga fyrir hönd ráðuneytisins um framkvæmd kjarnaverkefna.
- 2 Umsjón með úthlutunum hagnýtingarstyrkja.
- 3 Kynningar- og ráðgjafahlutverki fyrir hönd stjórnvalda og vinni að því markmiði að koma íslenski máltækni að hjá íslensku atvinnulífi og í lausnum erlendra tæknifyrirtækja.

Umfang þessara verkefna er talsvert og leggur hópurinn til að menningar- og viðskiptaráðuneytið leggi a.m.k. 85 m.kr. árlega í þennan rekstur, sem felur einnig í sér gríðarlega stórt hlutverk í hagnýtingarþætti áætlunarinnar. Nánar er fjallað um þessa þætti starfseminnar í síðari köflum skýrslunnar.

Hjá Almennarómi hefur skapast mikilvæg reynsla og þekking af rekstri fyrri máltækniáætlunar og hefur það fyrirkomulag sannað gildi sitt. Því er lagt til að Almennarómur verði miðstöð máltækni á Íslandi út verktíma næstu áætlunar og að sjálfsseignarstofnunin sjái til þess að verkefni áætlunarinnar verði framkvæmd af þeim sérfræðingum, stofnunum og fyrirtækjum sem best þykja til þess fallin. Auk þess vinni stofnunin stöðugt að því meginmarkmiði áætlunarinnar að styðja við og tryggja hagnýtingu íslenskrar máltækni fyrir almenning.

Verkefnastjóri hjá stofnuninni sjái um eftirlit með framkvæmd kjarnaverkefna og viðhaldi innviða. Stofnunin myndi einnig hafa umsjón með hagnýtingarstyrkjum og koma að úthlutun þeirra í samstarfi við menningar- og viðskiptaráðuneytið auk þess sem kynningarstarf og útrás íslenskrar máltækni væri í hennar höndum. Með því að hafa öll þessi málefni undir sama hatti skapast góð yfirsýn yfir heildarmarkmið áætlunarinnar auk þess sem þjónusta við fyrirtæki og ráðgjöf við innleiðingu máltæknilausna ætti að stórukast og bjóða upp á góðar tengingar beint inn í rannsóknarsamfélag máltæknilausna.

2.4 Fyrirkomulag fjárveitinga til máltæknimála

Þrátt fyrir að vel hafi tekist til við uppbyggingu innviða í gegnum kjarnaverkefni fyrri máltækniáætlunar er nú nauðsynlegt að koma á fót heildstæðri stjórnun fjármagns og verkefna. Í grunninn má segja að fjárveitingar vegna fyrri máltækniáætlunar hafi farið í gegnum tvo farvegi, annars vegar bein framlög frá menningar- og viðskiptaráðuneytinu (áður mennta- og menningarmálaráðuneytið) og hins vegar með fjárveitingum í gegnum samkeppnissjóði, þ.e. Markáætlun um tungu og tækni sem úthlutað hefur verið í gegnum Rannís.

Markáætlun í tungu og tækni er opinn samkeppnissjóður sem starfar samkvæmt lögum um opinberan stuðning við vísindarannsóknir. Markáætlun veitir styrki samkvæmt þeim áherslum sem stjórn markáætlunar skilgreinir fyrir sjóðinn, og á grundvelli mats fagráðs á gæðum verkefna, hversu víðtækt verkefnið er, þörf á afurðum verkefnisins, færni þeirra einstaklinga sem stunda rannsóknirnar og aðstöðu þeirra til að sinna verkefninu. Hlutverk markáætlunarinnar er að auðvelda nýtingu nýrrar samskiptatækni í íslensku samfélagi, til hagsbóta fyrir almenning, stofnanir og fyrirtæki.

Á verk tíma fyrri máltækniáætlunar, fimm ára tímabili frá 2018-2022, komu alls um 1.713 m.kr. beint frá menningar- og viðskiptaráðuneyti (áður mennta- og menningar- málaráðuneyti) og um 542 m.kr. í gegn um Markáætlun í tungu og tækni. Gert var ráð fyrir að kjarnaverkefni máltækniáætlunar yrðu fjármöggnuð af ráðuneytinu en að samhliða því starfi yrði komið á laggirnar sérstökum hagnýtingarsjóði sem væri ætlað að styðja við verkefni í máltækni sem hefðu það að markmiði að hagnýta þá innviði sem þróaðir væru undir kjarnaverkefnahluta áætlunarinnar.

Ljóst er að sá hagnýtingarsjóður sem lagt var upp með í fyrri máltækniáætlun, sem átti að styðja við þróun endalausna og hagnýtingu máltækniinnviða með mótframlagi frá atvinnulífnum, varð aldrei að veruleika. Þeir styrkir sem veittir hafa verið til máltækniverkefna í gegn um markáætlun hafa ekki verið til þess fallnir að markmið um þátttöku atvinnulífsins næðu fram að ganga. Nær öll verkefni sem hlutu styrk úr markáætlun voru að einhverju eða öllu leyti unnin af aðilum sem eru hluti af rannsóknarsamfélaginu, sem jafnframt vann að uppbyggingu innviða í máltækni. Aðkoma fyrirtækja á neytendamarkaði hefur því verið hverfandi.

Mikilvægt er að tryggja með betri hætti aðkomu atvinnulífsins að máltækni, bæði með tilliti til markmiða máltækniáætlunar en ekki síður til að hvetja fyrirtæki til að fjárfesta í hagnýtingu máltækni og gervigreindar.

Ein stærsta áskorun íslenskrar máltækni felst í því að koma máltækniáætlun til notenda með skilvirkum hætti. Núverandi fyrirkomulag fjárveitinga til verkefna mætir ekki að fullu þeim áskorunum sem áframhald íslenskrar máltækni felur í sér. Mikilvægt er að endurskoða fyrirkomulag fjárveitinga, bæði til að veita stjórnvöldum betri yfirsýn yfir þau verkefni sem verið er að vinna en ekki síður til að tryggja að fjármagni sé veitt í samræmi við stefnu stjórnvalda um forgangsverkefni næstu máltækni áætlunar. Í því ljósi er skynsamlegt að fjármunum til hagnýtingarverkefna verði í meira mæli veitt með miðlægum hætti í stað þess að ætla Markáætlun í tungu og tækni það hlutverk. Með því mætti forgangsraða betur verkefnum með tilliti til þeirrar stefnu sem mörkuð verður og tryggja að nægjanlegu fjármagni verði varið til að auka notkun þeirra máltækniáætlunar sem þróaðar eru í gegnum kjarnaverkefnahluta áætlunarinnar. Leggja þarf mun meiri áherslu á hagnýtingu þeirra innviða sem þegar hafa verið gerðir og því þarf umgjörð fjármögnunar að taka mið af því.

Er því lagt til að Markáætlun í tungu og tækni standi óbreytt út árið 2026 þar til heildar- mynd er komin á máltækni- og gervigreindarmálefni landsins. Í millitíðinni verði komið á fót sérstökum hagnýtingarstærkjum fyrir íslenska máltækni þar sem bæði verði horft til styrkja fyrir endalausnaverkefni sem nýta máltækniinnviði en einnig stuðnings við almenna innleiðingu þeirra innviða inn í þegar tilbúin kerfi.

Kafli 3



3. Hagnýting máltækni

Þrátt fyrir að vel hafi gengið með þróun innviða fyrir íslenska máltækni og samstarf háskóla, stofnana, fyrirtækja og félagasamtaka, undir stjórn Almennaróms, verið til fyrirmyndar er framtíð íslenskunnar í stafrænum heimi enn fjarri því að vera tryggð. Án áframhaldandi markvissra aðgerða er framtíð hennar enn ógnað. Því er afar mikilvægt að byggt verði ofan á þá innviði máltækni sem þegar hefur verið fjárfest í.

Megináhersla fyrri máltækniáætlunar var að byggja upp innlenda máltækniinnviði til að tryggja að íslenskan yrði gjaldgeng í stafrænum heimi. Mikið hefur áunnist á þessum tíma: Íslenskur talgreinir, talgervill, grunnvélþýðing milli íslensku og ensku, yfirlestrartól og margt fleira sem þróað var í síðustu áætlun. Nú er svo komið að tryggja verður að þeir innviðir komist í almenna notkun, bæði meðal almennings í daglegu lífi en einnig í þjónustu fyrirtækja og stjórnsýslu.

brennt skiptir þar meginmáli:

- 1 Að fyrirtæki og stofnanir samfélagsins sem vilja innleiða eða þróa nýjar endalausnir, sem byggja á opnum máltækniinnviðum geti sótt um styrki til þeirra gegn eigin mótframlagi.
- 2 Að fyrirtæki og stofnanir sem vilja innleiða eða hagnýta opna máltækniinnviði geti fengið ráðgjöf og að stuðlað sé að markvissri innleiðingu þeirra.
- 3 Að samskiptum sé haldið áfram við erlend tæknifyrirtæki sem þróa lausnir sem hafa náð, eða geta náð, mikilli útbreiðslu meðal íslensks almennings og unnið sé að því með markvissum hætti að koma íslenskri máltækni að í lausnum þeirra fyrirtækja.

3.1 Fjármagn til hagnýtingarverkefna

Eins og nefnt var í kafla 2.4 hér að ofan varð sá hagnýtingarsjóður sem lagt var upp með að yrði komið á fót í fyrri máltækniáætlun aldrei að veruleika og hefur því skort nauðsynlegan stuðning við hagnýtingarverkefni. Í slíka styrki gætu íslensk fyrirtæki sótt fjármagn til að innleiða máltæknilausnir í þjónustu sína eða nýta máltækniinnviði til að þróa áfram hagnýtar endalausnir fyrir notendur. Slíka styrki hefur sárlega vantað í það styrkjaumhverfi sem er til staðar á Íslandi í dag og hafa máltækniyrirtæki, menntatæknifyrirtæki og fleiri sem hafa lýst yfir áhuga á að hagnýta íslenska máltækni kvartað yfir því að hafa ekki getað sótt nauðsynlega styrki til verkefna sinna.

Lagt er til að menningar- og viðskiptaráðuneytið leggi hagnýtingarverkefnum til 60 m.kr. árlega. Almennarómur fari með umsýslu þess fjár, auglýsi eftir umsóknum og sjái um greiðslu styrkjanna sem veittir verða. Skipuð verði úthlutunarnefnd sem metur umsóknir og hvaða verkefni hljóti styrki. Í henni sitji verkefnastjóri máltækniáætlunar, fulltrúi frá Almennarómi og fulltrúi skipaður af menningar- og viðskiptaráðherra.

Tryggt verði að bæði fullmótuð þróunarverkefni sem skila endalausnum fyrir notendur verði styrkt en einnig að hægt verði að sækja þar styrki til innleiðingar máltæknilausna sem þegar hafa verið þróaðar inn í kerfi og þjónustu fyrirtækja.

Skilyrði fyrir styrkveitingum yrðu bundin við hagnýtingu máltækniinnviða og afurða í verkefnum sem hefðu að markmiði að gera notendum stafrænna lausna kleift að nota íslensku. Farið yrði fram á 50% mótframlag í umsóknum og atvinnulífið tæki þannig virkan þátt í að styðja máltækniðnaðinn í landinu.

Lagt er til að horft verði til þess sérstaklega til að byrja með að styrkja lausnir sem hagnýtast innan heilbrigðis- eða menntakerfisins. Það liggur nokkuð ljóst fyrir hvernig nýta megi máltækni innan þessara geira og eru þegar komnar fram lausnir sem nýta sér máltækni sem er sérstaklega þróuð fyrir þessi svið. Þá verði leitast við að styrkja lausnir sem geti nýst við kennslu á íslensku sem öðru máli, bæði innan skólakerfisins og fyrir fólk á vinnumarkaði, og nýta sér íslenska máltækni til þess. Við úthlutun styrkja verði það meginmarkmið máltækniáætlunar, að tryggja framtíð íslenskunnar í stafrænum heimi, ávallt haft að leiðarljósi og að verkefni séu til þess fallin að styrkja við stöðu íslenskrar tungu í samfélaginu.

3.2 Ráðgjöf til íslensks atvinnulífs og almenn kynning máltæknilausna

Við samráð og greiningarvinnu stýrihópsins á upphafsvikum skipunartíma varð fljótt ljóst að gera hefði mátt betur ráð fyrir kynningu á því starfi sem unnið var á verktíma síðustu áætlunar. Almennarómur sem hélt utan um rekstur áætlunarinnar hafði aðeins einn starfsmann í vinnu á verktímanum og ekki skýrt hver hefði það hlutverk að kynna máltæknilausnir sem þróaðar voru undir áætluninni; fyrirtæki sem þróuðu lausnina, Almennarómur eða menningar- og viðskiptaráðuneytið.

Fyrir vikið hefði mátt upplýsa íslenskt atvinnulíf, félagasamtök og almenning betur um þá möguleika sem felast í innviðum íslenskrar máltækni. Auk þess var ekki til miðlæg ráðgjöf fyrir þau fyrirtæki sem höfðu áhuga á að þróa áfram endalausnir. Upplýsinga- og ráðgjöf var því af skorum skammti um hvaða innviðaútgáfur hentuðu þeim best, hvaða uppfærslur innviða voru nýjastar og hvaða íslensku máltækni fyrirtæki væru best til þess fallin að aðstoða þau við þróun endalausnarinnar.

Lagt er til að Almannarómur taki þetta hlutverk að sér og gert sé ráð fyrir starfsmanni innan stofnunarinnar sem hafi það hlutverk að vera íslenskum fyrirtækjum og stofnunum til ráðgjafar og aðstoða við innleiðingu máltæknilausna. Almannarómur sjái um kynningarstarf til almennings og atvinnulífs á íslenskri máltækni, haldi utan um ráðstefnur o.fl.. Einnig er lagt til að stofnunin fái fjármagn í að uppfæra heimasíðu sína og kynni þar betur það sem er í boði innan íslenskrar máltækni í dag og hver stefnan sé næstu árin.

3.3 Samskipti við erlend tæknifyrirtæki

Á síðustu árum hafa tilraunir Almannaróms, íslenskra stjórnvalda og innlendra máltækniyrirtækja til að koma íslenskri máltækni að sem víðast í lausnum erlendra tæknifyrirtækja skilað nokkrum árangri.

Ljóst er að áframhaldandi vinna við það verkefni er einn mikilvægasti þáttur þess að tryggja notkun almennings á íslenskri máltækni og þar með tryggja stafræna framtíð tungumálsins í heimi gríðarlega hraðrar tækniþróunar.

Ljóst er að áframhaldandi vinna við það verkefni er einn mikilvægasti þáttur þess að tryggja notkun almennings á íslenskri máltækni og þar með tryggja stafræna framtíð tungumálsins í heimi gríðarlega hraðrar tækniþróunar. Engar leiðir eru til að tryggja að fyrirtækin taki upp íslenska máltækni en reynsla síðustu ára hefur þó sýnt að fyrirtækin eru flest opin fyrir því að taka upp íslenska máltækni fylgi því ekki of mikil fyrirhöfn eða kostnaður. Því er mikilvægt að reynt verði með markvissum hætti að funda með erlendum tæknifyrirtækjum um íslenskar máltæknilausnir og tryggja sess íslenskunnar í lausnum þeirra.

OpenAI

Samstarf íslenska máltækniyrirtækisins Miðeindar við bandaríska tækniyrirtækið OpenAI er til marks um hve góðir máltækniinnviðir hafa verið smíðaðir hér á landi á síðustu árum undir fyrri máltækniáætlun og hvernig hægt er að hagnýta þá með markvissum hætti. Samstarfið er tilkomið vegna fundar sendinefndar forseta Íslands, menningar- og viðskiptaráðherra og Almennaróms með Sam Altman, forstjóra og stofnanda OpenAI, í San Francisco í maí 2022. Í ferðinni var íslensk máltækni kynnt fyrir erlendum tækniyrirtækjum með það að markmiði að koma íslensku að í sem flestum tæknilausnum sem almenningur notar í daglegu lífi. Í kjölfar fundar sendinefndarinnar með OpenAI ákvað fyrirtækið að velja íslensku sem fyrsta málið utan ensku sem gervigreindarmállíkanið Chat GPT-4 væri þjálfað sérstaklega í og var það mögulegt með hjálp íslenskra máltæknilausna. Líkanið var gefið út um miðjan mars 2023 og vakti heimsathygli vegna færni þess í samræðum og spurningasvörum.

Microsoft

Microsoft gaf út talgervilsraddirnar Guðrúnu og Gunnar á vormánuðum 2022. Raddirnar eru bæði stýrikerfisraddir og aðgengilegar í gegnum skýjaþjónustu Microsoft og því getur vafri Microsoft, Edge, lesið upp íslenskan texta líkt og vafri Google, Chrome.

Samhliða gaf Microsoft út íslenskan talgreini í fyrsta skipti. Þannig er hægt að tala við Microsoft stýrikerfið á íslensku. Einn af teymisstjórum Microsoft í verkefninu, Jacky Kang, hefur staðfest á fundi með Háskólanum í Reykjavík að gögn sem gefin voru út undir opnum leyfum úr verkáætluninni Máltækni fyrir íslensku 2018-2022 hefðu verið notuð við þróun talgreinisins. Þá hefur HR útbúið ábendingar til Microsoft um hvernig megi bæta gæðin í röddum Guðrúnar og Gunnars.

Google

Tæknirisinn Google hefur í þrígang gefið út íslenskan talgervil. Líklegt er að þrýstingur frá Háskólanum í Reykjavík og Blindrafélaginu og tengsl íslenskra máltækniþræðinga við íslenskan starfsmann hjá Google hafi skilað því að íslenskur talgervill er aðgengilegur í lausnum þess.

Google kynnti í sumarlok ársins 2012 nýjan talgreini fyrir íslensku sem hefur meðal annars verið notaður í farsímum með Android-stýrikerfi. Slík tækni breytir töluðu máli í ritaðan texta og með slíkum búnaði er hægt að stýra ýmiss konar tækjum með Android-stýrikerfi með tali.

Haustið 2015 gaf Google svo út talgervilsróddina Önnu sem byggir á raddupptökum sem Háskólinn í Reykjavík gerði í samstarfi við Google. Þróun hennar fór því fram áður en fyrri máltækniáætlun komst í gagn. Ekki hafa fengist upplýsingar frá fyrirtækinu um hvort íslenskar máltækniáætlanir sem gefnar hafa verið út undir fyrri máltækniáætlun hafi verið notaðar við uppfærslur og þróun talgervilsins.

Dæmin sanna að langt má komast með góðum samskiptum við erlend stórfyrirtæki og frumkvæði að því að kynna fyrir þeim þær íslensku máltækniáætlanir sem þau geta nálgast og innleitt í sín kerfi. Því er lykilatriði að þeir máltækniinnviðir sem þróaðir verða á næstu áætlun verði áfram gefnir út undir opnum leyfum og að ábyrgð á áframhaldandi samstarfi og samskiptum við erlend tæknifyrirtæki verði komið fyrir hjá einum aðila sem fylgist vel með þróun mála erlendis og hafi frumkvæði að því að reyna að koma íslenskri máltækni þar að. Einnig er mikilvægt að þróuð verði prófunargögn fyrir þá máltækniinnviði sem smíðaðir hafa verið því ljóst er að fyrirtækin verða að geta áttað sig á því hver gæði lausnanna sem þau taka upp hjá sér séu en nauðsynleg forsenda þess er að prófunargögn séu til staðar.

Til að tryggja framtíð íslenskunnar í stafrænum heimi skiptir öllu máli að koma íslensku að í lausnum fyrirtækja á borð við Microsoft, Google, Meta, OpenAI og fjölda annarra stórra erlendra fyrirtækja. Þá er mikill fjöldi minna þekktra fyrirtækja sem bjóða upp á vefþjónustur sem notaðar eru víða í atvinnulífi og skólasterfi sem nauðsynlegt er að kortleggja og tryggja að íslenska verði einnig hluti af þeirra lausnum.

Lagt er til að Almennarómur sinni þessu hlutverki og að gert verði ráð fyrir talsverðri vinnu við þennan þátt ef gerður verður þjónustusamningur milli stofnunarinnar og ráðuneytisins. Lagt er til að Almennarómur skipuleggi fleiri heimsóknir sendinefnda til erlendra tæknifyrirtækja til að tala fyrir innleiðingu íslenskrar máltækni í þeirra lausnum. Stofnunin mun einnig hafa það hlutverk að vakta vel verkefni sem eru í þróun þessara stóru fyrirtækja og meta hvort og hvenær þörf er á viðbrögðum til að tryggja veg íslenskunnar í þeim verkefnum.

Kaflí 4



4. Kjarnaverkefni

4.1 Sveigjanleiki áætlunar og nýtt fyrirkomulag

Á tímabili fyrri máltækniáætlunar var fjöldi innviða, lausna gerður aðgengilegur undir opnum leyfum. Erfitt er að mæla gæði þessara lausna þar sem góð og gild prófunargögn hafa enn ekki verið þróuð og gefin út. Ljóst er þó að þá innviði sem smíðaðir voru má enn bæta með áframhaldandi þróun undir næstu máltækniáætlun.

Með hraðri tækniþróun og nýlegri byltingu í getu stórra mállíkana hefur umhverfið gjörbreyst. Þörf fyrir kjarnalausnir innan máltækni er önnur í dag en hún var á verk tíma fyrri máltækniáætlunar. Bæði hefur skapast þörf fyrir ný kjarnaverkefni sem ekki var gert ráð fyrir í síðustu áætlun auk þess sem mismikil þörf er á úrbótum á þeim lausnum sem hafa verið þróaðar.

Því er nauðsynlegt að fyrirkomulag fjárveitinga til áframhaldandi þróunar máltækniinnviða sé sveigjanlegt og fjármagn verði ekki fest í verkefnum langt fram í tímann, heldur sé tryggt að hægt verði að bregðast við breyttum aðstæðum og veita fjármagni í ný og brýn verkefni kalli tæknibreytingar á það.

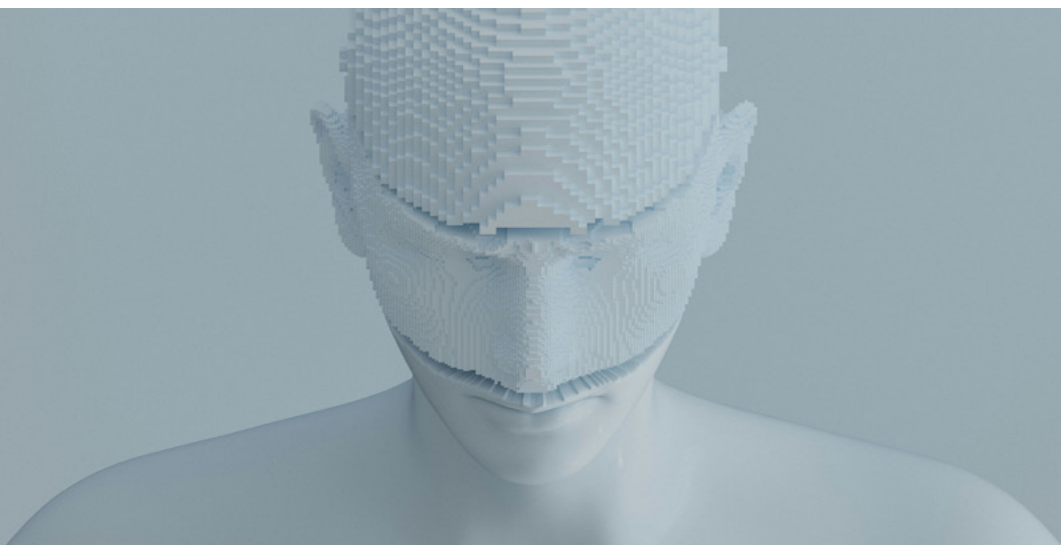
Einnig verður nú að taka tillit til óvissupátta sem fylgja hröðum tæknibreytingum. Mikilvægi og nauðsyn ákveðinna verkpátta við þróun kjarnalausna geta úrelst á svo gott sem einni nóttu með tilkomu nýrra og ófyrirséðra gervigreindarlausna að utan. Því er nauðsynlegt að fyrirkomulag fjárveitinga til áframhaldandi þróunar máltækniinnviða sé sveigjanlegt og fjármagn verði ekki fest í verkefnum langt fram í tímann, heldur sé tryggt að hægt verði að bregðast við breyttum aðstæðum og veita fjármagni í ný og brýn verkefni kalli tæknibreytingar á það.

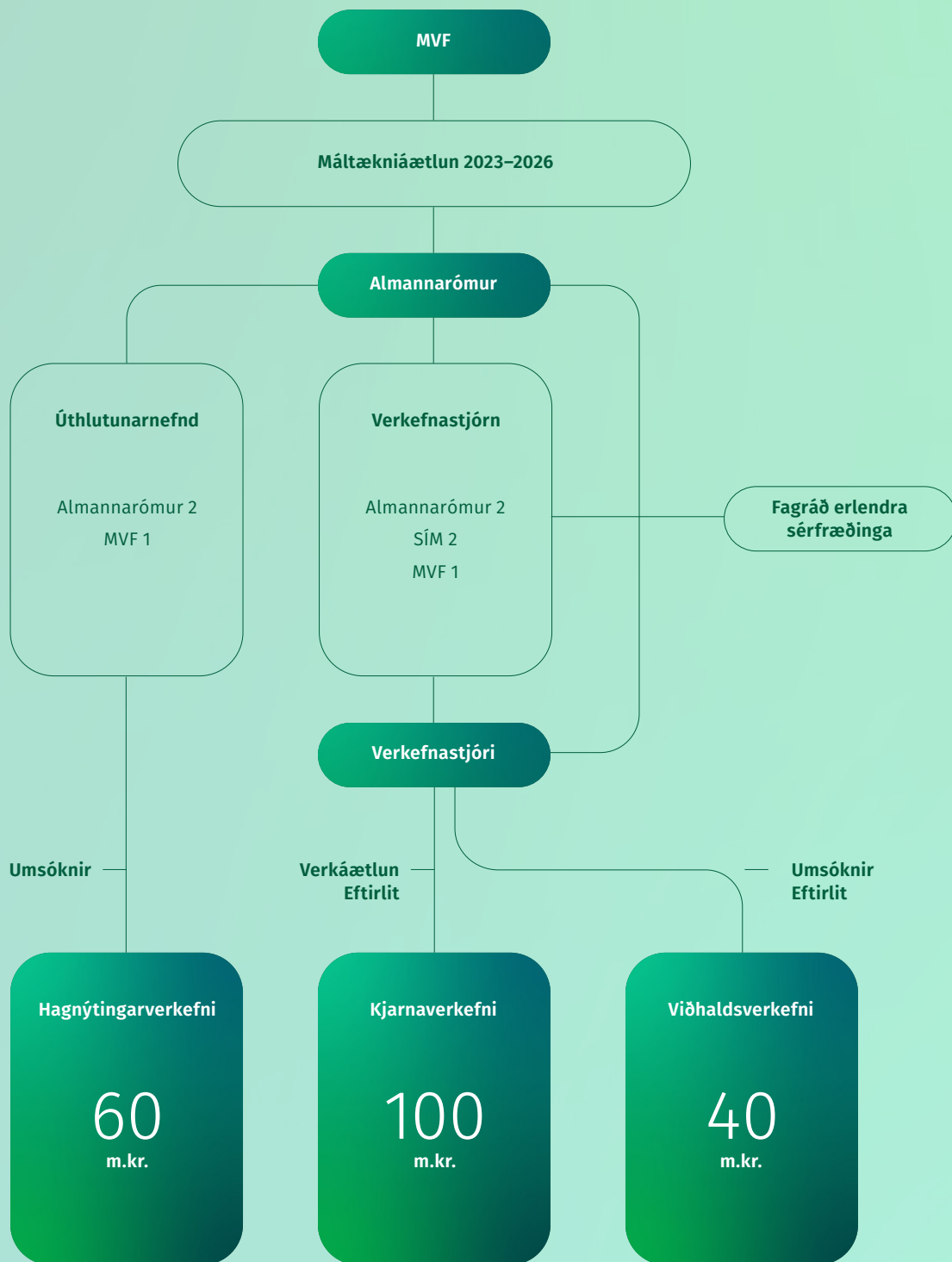
Stýrihópurinn telur að til að tryggja sveigjanleika við þróun máltækniinnviða á næstu árum þurfi að breyta fyrirkomulagi fjárveitingar til kjarnaverkefna.

Hópurinn leggur til að framlög til kjarnaverkefna nemi a.m.k. 100 m.kr. á ári en menningar- og viðskiptaráðuneytið endurmeti í samstarfi við Almennaróm fjármagnsþörf kjarnaverkefnahluta áætlunarinnar ár hvert. Samningur verði gerður við Almennaróm um að tryggja framkvæmd verkefnanna með því að útvista þeim með samningum til SÍM um framkvæmd þeirra líkt og á verktíma síðustu áætlunar. Gert er ráð fyrir að framlög til kjarnaverkefna verði sem fyrr undanþegin VSK og að allar afurðir verði áfram gefnar út með opnum hugbúnaðarleyfum og verði notendum að kostnaðarlausu. Almennarómur haldi áfram hlutverki sínu við að fylgjast með framgangi verkefnanna og upplýsa ráðuneytið um stöðu þeirra með reglulegu millibili. Fagráð erlendra sérfræðinga haldi áfram sínu hlutverki og rýni samninga og áfangaskýrslur um verkefni. Þá verði sérstakur verkefnastjóri máltækniáætlunar skipaður sem starfi hjá Almennarómi.

Lagt er til að nýtt fyrirkomulag verði tekið upp þegar kemur að vali kjarnaverkefna sem verður ráðist í ár hvert, til að stuðla enn frekar að auknum sveigjanleika áætlunarinnar.

Verkefnastjórn yfir kjarnaverkefnum og viðhaldi þeirra verði sett á laggirnar, skipuð tveimur fulltrúum Almennaróms, tveimur fulltrúum SÍM og einum frá menningar- og viðskiptaráðuneytinu. Verkefnastjórnin komi saman, þegar menningar- og viðskiptaráðuneytið hefur lokið við að skilgreina fjárhæð fjárveitingar í kjarnaverkefni komandi árs, til að meta hvort þau verkefni sem eru skilgreind í nýrri verkáætlun eigi enn rétt á sér. Þannig sé tekið tillit til nýjustu tækniþróunar og tryggt að kjarnaverkefni sem eru skilgreind í verkáætlun, eigi enn við. Þannig er tryggt að það fjármagn sem stjórnvöld hafa tilgreint í þróun kjarnaverkefna sé einvörðungu varið til verkefna sem þörf er á.





Mynd 1: Myndin sýnir umsýslu máltækniáætlunar. Almennarómur fer með umsýslu helstu verkefna. Verkefnastjóri, sem starfar hjá Almennarómi hefur umsjón með kjarnaverkefnum og viðhaldsverkefnum og leitar til verkefnastjórnar um ákvarðanir sínar. Fagráð erlendra sérfræðinga er honum til aðstoðar til að fara yfir verklýsingar og stöðuskýrslur um framgang kjarnaverkefna. Ný úthlutunarnefnd verður skipuð sem metur umsóknir um styrki Almennaróms til hagnýtingarverkefna.

4.2 Kjarnaverkefni og samkeppnissjónarmið

Tryggja verður að með þróun kjarnaverkefna sem fjármögnuð eru úr máltækniáætlun stjórnvalda verði ekki stigið inn á samkeppnissvið þeirra máltækniyrirtækja sem eru starfandi á íslenskum markaði. Skýran greinarmun verður að gera á því sem flokkast á sem kjarnaverkefni, sem gefið er út undir opnum leyfum og að öllu leyti fjármagnað úr máltækniáætlun, og þeim verkefnum sem frekar eiga heima á samkeppnismarkaði. Kjarnaverkefni eiga þannig að vera skilgreind sem þróun grunninnviða sem hægt er að nýta í þróun endalausna og gagnast jafnt öllum þeim sem vilja ráðast í slíka vinnu.

Innleiðing máltækniinnviða eða hagnýting máltæknilausna verður fjármögnuð úr hagnýtingarhluta áætlunarinnar gegn mótframlagi fyrirtækja. Máltækniinnviðir eru komnir langt í þróun eftir fyrri máltækniáætlun og eru margir þeirra þegar tilbúnir til innleiðingar í endalausnir. Mikilvægt er að framtíðarþróun kjarnaverkefna miði að því að hægt sé að nýta þá með sem bestum hætti beint inn í endalausnir sem þróaðar eru á samkeppnismarkaði.

4.3 Kjarnaverkefni í forgrunni

Stýrihópurinn leggur til að eftirfarandi kjarnaverkefni verði sett í forgrunn í nýrri verkáætlun en hópi sérfræðinga verði falið að endurmeta þau og forgangsraða á verk-tíma áætlunarinnar. Mikilvægt er að forgangsraða og horfa til þess að fjármagna hluta þeirra eftir öðrum leiðum, m.a. í gegn um Evrópusamstarf.

Málföng

Tilgangur: Málföng eru gögnin sem máltækniþólin byggja á. Til þess að hægt sé að þróa tólin og aðlaga tæknina að íslensku þarf fjölbreytt og ríkuleg gögn. Þau eru notuð til að þjálfa tólin en einnig til að prófa þau og meta gæði þeirra og nákvæmni. Mat á gæðum er mikilvægt á öllum stigum, bæði í þróun til að tryggja sé að stefnt sé í rétta átt og þegar lokaafurð er tilbúin. Þannig er hægt að upplýsa notandann um gæði tólanna, hvort sem um er að ræða endanotanda eða þá sem notar tólin við þróun annars hugbúnaðar. Sum gagnasöfnin, t.a.m. Risamálheild, nýtast í mörg verkefni í öðrum hlutum máltækniþróunar, en sum verkefni krefjast sérhæfðra gagnasafna. Verkbættir eru eftirfarandi:

-
- Stækkun Risamálheildar
-
- Samhliða setningar á tveimur eða fleiri tungumálum
-
- Stór söguleg málheild, þáttuð skv. liðgerð og varpað yfir á UD (Universal Dependencies)
-
- Málheild fyrir viðhorfsgreiningu
-
- Málheild fyrir upplýsingaútdrátt
-
- Málheild fyrir samantekt með útdrætti og úr lengra samhengi
-
- Málheild fyrir spurningasvörun
-
- Málheild fyrir merkingareinræðingu
-
- Málheild fyrir umorðun
-
- Málheild fyrir greiningu erfiðleikastigs texta og textaeinföldun
-
- Marghátta málheild fyrir táknmál
-
- Marghátta málheild fyrir samræður augliti-til-auglitis (myndbönd af samræðum)
-
- Marghátta málheild með myndum og textalýsingu
-
- Prófunargögn fyrir samvísana- og anafóruúrlausn
-
- Prófunargögn fyrir getu stórra mállíkana (málskilningur)
-
- Viðmiðunargögn fyrir bjagagreiningu (t.d. kynjabjaga)
-

Textavinnsla og samskipti

Tilgangur: Greining texta með tilliti til ýmissa þátta, merkingarfræðilegra og málfræðilegra, er grundvöllur þess að hægt sé að laga máltæknina að þörfum tiltekinna hópa eða sviða. Það þarf að vera hægt að leysa úr margræðni til að auðvelda merkingargreiningu texta og söfnun upplýsinga úr honum. Þá er nauðsynlegt að auðvelda greiningu á íðorðum til að tryggja rétta orðanotkun í fagtextum. Þegar texta er beint að tilteknum hópi getur verið nauðsynlegt að greina hversu tyrfinn hann er, svo hægt sé að færa hann á viðeigandi málsnið og auðvelda aðgengi að innihaldinu, hvort sem það er fyrir börn, annarsmálshafa eða aðra hópa. Verkpættir eru eftirfarandi:

-
- Samvísana- og anafóruúrlausn
-
- Sjálfvirk íðorðataka
-
- Sjálfvirk orðabókagerð
-
- Erfiðleikastig texta og einföldun þeirra
-
- Máltæknilausnir fyrir íslenskukennslu og til stuðnings við annarsmálshafa
-
- Þróun Universal Dependencies-þáttara
-

Talgreining

Tilgangur: Að útvíkka notkunarmöguleika talgreina fyrir sem flesta þjóðfélagshópa, svo sem börn og annarsmálshafa, með áframhaldandi þróun. Einnig að halda áfram þróun talgreiningar fyrir ýmiss konar hugbúnaðar- og vélbúnaðarumhverfi, til dæmis stýrikerfi og snjallsíma. Viðkomandi verkþættir eru:

-
- Talgreining fyrir börn, annarsmálshafa og eldri raddir
-
- Ritarahamur fyrir íslenska talgreiningu, s.s. skipanir fyrir greinarmerki og línubil
-
- Greining tals þar sem skipt er milli tungumála (e. code switching)
-
- Gagnasöfnun á íslensku á LibriVox (opið, alþjóðlegt hljóðbókasafn, veitir aðgang að lestri langra texta í stað stakra setninga/stytttri textabúta)
-
- Mælendakennsl (e. speaker recognition) fyrir íslensku
-
- Talgreining fyrir náttúruleg samtöl, þ.m.t. bakrásarröddun (e. back channel vocalization)
-
- Mál- og hljóðlíkanagerð fyrir íslensku
-

Talgerving

Tilgangur: Að íslenskir talgervlar verði nothæfir fyrir þá sem á þurfa að halda í daglegu lífi (aðgengis-, öryggis- og réttindamál); að íslenskir talgervlar verði nothæfir í viðskipta- og þjónustuumhverfi; að hægt verði að nota íslenska talgervla til þess að lesa námsbækur og aðra nytjatexta og gefa út sem hljóðbækur, og geta þannig t.d. flýtt framleiðslu efnis frá Hljóðbókasafni Íslands og Menntamálastofnun. Verkpættir eru eftirfarandi:

-
- Áframhaldandi þróun talgervils og uppfærsla á bakenda (e. vocoder)
-
- Bæta talgervil á Android-tækjum, en blint og sjónskert fólk notar hann m.a. til aðstoðar í daglegu lífi
-
- Talgerving með barnaröddum
-
- Raddaðlögun og -blöndun
-
- Bættur lestur á texta sem inniheldur blönduð tungumál og sérfræðiorð
-
- Sérhæfð textanormun
-
- Ítónun og hljómfall
-
- Áherslu- og ítónunarstjórnun í lesnum texta
-
- Notendaviðmót: viðbætur/breytingar á orðabók, bæta við skammstöfunum o.þ.h.
-
- Talgerving á löngum textum/heilum bókum
-

Vélpýðingar

Tilgangur: Vélpýðingar eru mikilvægar til að gera íslenska texta aðgengilega þeim sem ekki tala íslensku og til að gefa íslenskumælandi fólki færi á að skilja texta á erlendum málum, t.d. í fjölmiðlum og á vefnum. Vélpýðingar eru þannig mikilvægar til að jafna og auka aðgengi almennings að upplýsingum. Þær nýtast einnig við klæðskerasaumaðar fagpýðingar, t.d. við pýðingar á reglugerðum hjá Þýðingamiðstöð Utanríkisráðuneytisins og á þjónustuvefjum hins opinbera og atvinnulífs. Vélpýðingar munu nýtast við að texta erlent sjónvarpsefni með sjálfvirkum hætti. Nauðsynlegt er að til sé innlend pýðingarvél þar sem gætt er að gæðum og blæbrigðum íslenskunnar, meðal annars m.t.t. bjaga og hatursorðræðu. Verkpættir eru eftirfarandi:

-
- Þýðingar á lengra samhengi og fjölbreyttari textategundum en í dag (fleiri svið en fréttir og reglugerðir)
-
- Samröðun lengri skjala fyrir íslensku og önnur mál en ensku
-
- Bjagar í vélpýðingum og mat á þeim
-
- Þýðing tals og samræðna í rauntíma
-
- Sjálfvirk skjátextun erlends sjónvarpsefnis á íslensku
-
- Notkun sérhæfðra orðabóka/þýðingarminnis við vélpýðingar
-
- Greining og notkun orðatiltækja við vélpýðingar
-
- Léttari þýðingarlíkön sem geta m.a. keyrt á snjallsímum án nettengingar
-
- Upplýsingaheimt þvert á tungumál
-
- Samstarf við WMT um íslensku í “shared tasks”
-

Málrýni

Tilgangur: Markmið verkefnisins er að Íslenskur almenningur hafi greiðan aðgang að tóli sem leiðréttir bæði stafsetningu og málfræði texta m.v. algengustu málstaðla og -viðmið. Viðkomandi tól þarf að vera nákvæmt og geta þjónað mismunandi notendahópum, m.a. annarsmálmálshöfum, fólki með lesblindu og börnum. Mikilvægt er að tólið sé jafnframt aðgengilegt á ólíku formi, t.d. í algengustu ritvinnslukerfum og í snjalltækjum. Málrýni getur jafnað aðstöðumun í samfélaginu þegar kemur að færni í íslensku og fordómum sem henni tengjast. Í atvinnulífinu getur málrýni gagnast við að viðhalda skilgreindu málsniði (e. tone of voice) fyrirtækis og tryggt að miðlun til viðskiptavina og annarra hagsmunaaðila sé á skýru og skiljanlegu máli sem samræmist málfarsstaðli. Verkpættir eru eftirfarandi:

-
- Áframhaldandi þróun málrýnis með útskýrðum ábendingum
-
- Áframhaldandi þróun málrýnis með hraðvirkri vörpun gallaðs texta yfir í betri texta, þar sem útskýringa á leiðréttingum er ekki þörf
-
- Viðbætur við málfarsstaðal (fleiri atriði tekin fyrir, ítarlegar leiðbeiningar fyrir hvert atriði o.s.frv.)
-
- Umbætur á almennri villumálheild
-
- Lesblinduvillumálheild
-
- Barnamálsvillumálheild
-
- Annarsmálsvillumálheild
-
- Sérniðin málrýni fyrir notendahópa, t.d. annarsmálshafa, lesblinda og börn
-
- Greining og leiðrétting flóknari atriða (kynjuð málnotkun, málsnið og stíll, orðaval)
-
- Áframhaldandi þróun viðbóta; innfelling málrýni í hugbúnað, stýrikerfi og tæki
-

Gervigreind og samræðukerfi

Tilgangur: Að nýta nýjustu gervigreindartækni, þar á meðal stór mállíkön (e. large language models), til að smíða margvísleg viðföng fyrir íslensku, sambærileg þeim sem þekkjast fyrir stærri tungumál. Þar er meðal annars átt við opna spurningasvörun, samantekt texta, einföldun texta, viðhorfsgreiningu, hjálp í þjónustuveri o.m.fl. Þessi viðföng geta jafnframt samtvinnað talgreiningu, talgervingu og samræðugreind sem þarf til þess að búa til heildstæð samræðukerfi. Markmiðið er m.a. að hægt sé að tala við tæknina á íslensku og að hún skilji og svari eins eðlilega og mögulegt er. Jafnframt getur tæknin gert atvinnulífi kleift að nota gervigreind í ýmsum innri ferlum sem byggja á texta- og talnagögnum og til að aðstoða stjórnendur.

Aðlögun gervigreindar- og samræðulausna að íslensku er lykilþáttur í að verja stöðu íslenskunnar í fjórðu iðnbyltingunni og að tryggja að tæki og hugbúnaður framtíðar geti skilið, meðhöndlað og komið upplýsingum frá sér á íslensku í stað þess að enskan taki yfir.

-
- Stór mállíkön aðlöguð fyrir íslensku og nýtt í fjölbreytt viðföng, þar á meðal:

 - » Opna spurningasvörun og upplýsingaheimt

 - » Textasamantekt, textaeinföldun og upplýsingaútdrátt

 - » Viðhorfsgreiningu

 - » Spjallmenni/þjónustuver

 - Samræðukerfi, sbr. m.a. Emblu/Siri/Alexa/Cortana, sem m.a. byggja á gervigreind

 - » Samræðustjóri (e. dialogue manager)

 - » Málskilningur (e. natural language understanding)

 - » Málmyndun (e. natural language generation)

 - » Ýmsir hlutar samræðna sem eru utan texta, svo sem hikorð og hljómfall

 - Lágörkun þjaga og skekkja í gögnum og líkönum; mælingar; greinargerðir
-

4.4 Viðhald innviða og kjarnalausna

Í máltækniáætlun er gert ráð fyrir hýsingu og viðhaldi gagna og máltækniinnviða sem hafa verið þróaðir á verk tíma fyrri máltækniáætlunar auk þeirra sem þróaðir verða á næstu. Gerður hefur verið samningur við Stofnun Árna Magnússonar í íslenskum fræðum út árið 2026 um að reka CLARIN-miðstöð á Íslandi.

CLARIN ERIC er evrópskur samstarfsvettvangur um rannsóknarinnviði í máltækni. CLARIN stendur fyrir *Common Language Resources and Technology Infrastructure* og **ERIC** fyrir *European Research Infrastructure Consortium* sem er rekstrarform viðkomandi samstarfsnets. CLARIN-samstarfsnetið gengur út á styðja við uppbyggingu, skráningu, viðhald, varðveislu, notkun og samnýtingu mállegra gagna og búnaðar til rannsókna á þess konar gögnum í hug- og félagsvísindum.

Lagt er til að fyrirkomulag við rekstur CLARIN-miðstöðvar Íslands innan Stofnunar Árna Magnússonar verði endurskoðað fyrir lok þessarar máltækniáætlunar árið 2026 en æskilegt væri að koma öllum helstu máltækni verkefnum landsins undir sama hatt, líkt og greint er nánar frá í kafla 2.2.

Huga verður vel að viðhaldi þeirra máltækniinnviða sem þróaðir hafa verið svo þeir úreldist ekki við öra tækniþróun, nýrra útgáfa forritunarmála eða stýrikerfa. Hópurinn telur æskilegast að horfið sé frá því fyrirkomulagi að úthluta einstökum verkefnum fyrir fram skilgreindu fjármagni til viðhalds. Þess í stað verði verkefni, sem talin eru upp hér að neðan, sameinuð á eitt vefsvæði, sem hýst verður af Almannarómi, og viðhald og uppfærslur fari fram eftir mati verkefnastjóra áætlunarinnar í samráði við verkefnastjórn sem greint er frá í kafla 4.1.

Verkefnastjórnin hefði 40 m.kr. árlega til að útdeila eftir þörfum og forgangs röðun lausnanna sem þyrftu á viðhaldi að halda. Verkefnastjórnin útdeilir því fjármagni til stofnana og máltækni fyrirtækja sem reka máltækniinnviði. Verkefnastjóri mun eiga í nánú samstarfi við SÍM og aðra aðila máltækniáætlunar til að meta viðhaldsþörf lausna og koma umsóknum um viðhaldsstyrki áfram til verkefnastjórnarinnar sem tekur ákvarðanir um hvaða kjarnalausnir hljóti fjármagn til viðhalds.

Viðhaldsverkefnum verður m.a. forgangsraðað eftir fjölda notenda þeirra, umfangi og hversu áriðandi uppfærslur eða lagfæringar eru hverju sinni. Skilyrði fyrir úthlutun viðhaldsfjármagns er að það sé nýtt í viðhald gagna eða innviða sem hafa verið þróuð með fjármagni úr fyrstu eða annarri máltækniáætlun stjórnvalda og gefin út undir opnum leyfum.

Þannig gæfist færi á að haga áherslum í viðhaldi þannig að þau verkefni sem verða í mestri notkun verði sett í forgang í stað þess að fjármunir yrðu bundnir fyrir fram í verkefnum sem verða mögulega í minni notkun eða verkefnum sem kynnu að úreldast með tímanum.

Einnig er talið æskilegast að ákveðnum vefþjónustum sem opnaðar hafa verið og veita aðgang að afurðum kjarnaverkefna sé haldið við og þær hýstar og reknar á einum stað.

