

20. október 2026

Kjarnaverkefni máltækniáætlunar 2026

Áform um ráðstöfun fjármagns og kall eftir sjónarmiðum hagaðila

1 Um áformaskjalið

Menningar-, nýsköpunar- og háskólaráðuneytið semur árlega um framkvæmd kjarnaverkefna fyrir **100 m.kr.** skv. máltækniáætlun 2. Verkefnastjórn máltækniáætlunar gerir tillögur til ráðherra ár hvert um hvaða verkefni eigi að ráðast í fyrir það fjármagn. Af fjármagni ársins 2026 hefur þegar verið samið um 10 m.kr. til framhalds verkefnis sem hófst árið 2025. Verkefnastjórnin vinnur nú að því að gera tillögur til ráðuneytisins að ráðstöfun þeirra 90 m.kr. sem eftir standa.

Í áformaskjali þessu koma fram sjónarmið verkefnastjórnarinnar um hvernig fjármagni til kjarnaverkefna ársins 2026 skuli vera varið í grófum dráttum. Tilgangurinn með útgáfu skjalsins er tvíþættur; annars vegar að upplýsa helstu hagaðila um áformin og vinnulagið framundan við val verkefna og hins vegar að gefa þeim færi á að koma sjónarmiðum sínum og athugasemdum við fyrirætlanirnar á framfæri áður en gengið er frá þeim. Opið er fyrir umsagnir til og með 28. október.

Það er mat verkefnastjórnarinnar að það fjármagn sem eftir stendur, **90 m.kr.**, skili mestu virði sé því varið í tvennt: Aukna gagnaöflun og mælipróf fyrir íslensku í tækni.

2 Tímalína

20. október

Áformaskjal birt

**Til og með 28.
október**

Opið fyrir umsagnir

**Að umsagnarfresti
loknum**Verkefnastjórn fer
yfir umsagnir og velur
próf í kallið**2. til 13. nóvember**Opið kall eftir tillögum
um smíði mæliprófa

3 Gagnaöflun: 60 m.kr.

Áformað er að verja **60 m.kr.** af fjármagni ársins í gagnaöflun. Verkefnastjórnin metur nú, í samvinnu við menningar-, nýsköpunar- og háskólaráðuneytið, í hvaða gagnaverkefni sé fýsilegast að ráðast fyrir það fjármagn. Þar er sérstaklega horft til þess hvaða verkefni geti skilað mestu magni nýrra, opinna og nothæfra texta- og/eða talgagna fyrir íslensku.

Greining er þegar hafin á því hvar stærstu og gagnlegustu ósóttu gagnalindirnar er að finna í samráði við ráðuneytið. Á meðal þess sem er til skoðunar þar er að ráðast í eitt stærra verkefni fyrir megnið af fénu. Þar mætti sem dæmi sjá fyrir sér stærra verkefni sem miðaði að því að vinna gögn frá RÚV, sem hafa ekki verið aðgengileg undir opnum leyfum fram að þessu, inn í Risamálheildina, eða verkefni sem miðaði að því að vinna stóran hluta safnkosts Landsbókasafns Íslands inn í Risamálheildina. Önnur verkefni af minni skala eru einnig til skoðunar.

Verkefnastjórnin telur mikilvægt að þessi vinna sé unnin í nánu samráði við ráðuneytið. Þar eru í undirbúningi stærri áform um gagnaöflun og stafvæðingu menningararfsins, og mikilvægt er að verkefni máltækniáætlunar 2, á þessu síðasta úthlutunarári hennar, og framtíðaráform ráðuneytisins í gagnamálum fari saman.

4 Mælipróf fyrir íslensku: 30 m.kr.

Mælipróf eru söfn verkefna og spurninga sem lögð eru fyrir mállíkön og aðrar máltæknilausnir til að mæla hversu vel tæknin ræður við íslensku eða hversu vel hún stendur sig í ákveðnum verkefnum. Prófin gefa almenna mynd af því hvar íslenskan stendur í heimi tækninnar og hvar helst vantar upp á. Þau nýtast líka þeim sem velja og kaupa lausnir og síðar við þjálfun íslenskra líkana. Ýmis íslensk mælipróf eru þegar til og niðurstöður flestra eru birtar á opnum stigatöflum. Á mörgum sviðum vantar þó enn próf sem mæla ýmislegt sem skiptir máli fyrir íslenska notendur.

Áformað er að verja **30 m.kr.** af fjármagni ársins í smíði nýrra mæliprófa og að opið kall verði haldið þar sem leitað verður eftir aðilum til að þróa þau próf. Áætlað er að kallið verði opið frá 2. til 13. nóvember.

Opið kall eftir tillögum

Í kallinu verður leitað eftir tillögum frá aðilum sem lýsa yfir vilja til að taka að sér smíði tiltekins prófs eða prófa. Tillögurnar skulu fylgja verklagi sem lýst verður í kallinu og innihalda verkefnaáætlun og kostnaðaráætlun. Umsækjendur lýsa sjálfir aðferðafræði og mælikvörðum, taka fram hvernig matið fer fram, hvort sem það er mannlegt, sjálfvirk, með dómara líkani eða blanda af þessu, og hvort þeir staðfæra próf sem er til erlendis eða leggja til nýja aðferð.

Áður en kallið er opnað leggur verkefnastjórnin hér fram lista yfir próf sem vantar fyrir íslensku og íslenskt samhengi í tækni. Listinn er ekki tæmandi og ekki settur fram eftir neinni sérstakri forgangsörðun. Verkefnastjórnin kallar hér með eftir sjónarmiðum hagaðila um það hvaða próf liggi mest á að smíða ásamt rökstuðningi fyrir þörfinni. Einnig gefst hér tækifæri til að benda á önnur próf sem verkefnastjórninni kann að hafa yfirsést.

Ljóst er að aðeins verður hægt að smíða fáein próf fyrir þær 30 m.kr. sem áformaðar eru í þennan hluta kjarnaverkefna ársins 2026. Því verða aðeins þau próf sem metin eru í hæsta forgangi sett í opna kallið í ár.

Almannarómur er þó mjög áfram um að öll prófin á listanum verði til á endanum og mun vinna að því að finna þeim prófum sem ekki komast að í þessari umferð farveg og fjármagn í náginni framtíð.

Kröfur til prófa í kallinu

- Að mati verkefnastjórnarinnar skal meginreglan vera sú að prófin verði gefin út með opnu leyfi. Umsækjendur geta þó lagt til að próf eða hluti þess verði lokaður ef það er nauðsynlegt til að koma í veg fyrir að prófið rati í þjálfunargögn líkana og verði þar með fljótt úrelt. Slíkri tillögu þarf að fylgja hvar lokaði hlutinn verður geymdur, hverjir fá aðgang að honum, hvort aðrir geti nýtt hann til að meta líkön og hvort umsækjandi keyri prófið sjálfur til að skila niðurstöðu.
- Efni prófanna er samið eða yfirfarið af fólki með íslensku að móðurmáli. Vélþýddur texti fer ekki inn án nákvæmrar mannlegrar yfirferðar.
- Æskilegt er að óháðir aðilar fari yfir úrtak úr prófinu áður en það er birt. Almannarómur getur aðstoðað við að finna slíka aðila.
- Í kostnaðaráætlun er gert ráð fyrir að prófið sé keyrt einu sinni á helstu mállíkönunum, opnum sem lokuðum, og niðurstöður birtar.

Tækifæri til athugasemda

Verkefnastjórnin veitir hér hagaðilum tækifæri á að koma sjónarmiðum sínum um áform þessi á framfæri og óskar jafnframt eftir sjónarmiðum um eftirfarandi, fyrir fimmtudaginn **29. október**:

- Athugasemdir við áformin í heild sinni; fyrirkomulag og skiptingu fjármagnsins í gagnaöflunarverkefni annars vegar og ný mælipróf hins vegar.
- Hvaða próf af listanum hér á eftir ætti að smíða fyrst, og hvers vegna.
- Ábendingar um próf sem kann að vanta á listann.
- Athugasemdir við lýsingar á einstökum prófum, ef einhverjar eru.

Umsagnir eru sendar í gegnum eyðublað á tally.so/r/4a4405 eða á almannaromur@almannaromur.is.

5 Próf sem koma til greina

Lýsingarnar hér á eftir eru settar fram til að lýsa því um hvers kyns próf sé að ræða, tilgangi þess og hvaða þætti það ætti helst að mæla. Reynt er að halda orðalagi opnu svo þeir sem taka þátt í kallinu geti sjálfir lagt fram aðferðafræði um hvernig þeir hyggist mæla viðkomandi þætti. Dæmi um sambærileg próf og erlendar fyrirmyndir hér á eftir eru aðeins til viðmiðunar. Prófunum er raðað eftir sviðum og röðin segir ekkert um forgangsmat verkefnastjórnarinnar.

14 próf í 6 flokkum

Tal 2	Málkunnátta og ritfærni 3	Þekking 3	Verkefni og rökhugsun 2
Öryggi 1	Starfsgreinar 3		

Talgreiningarpróf

Að mæla almenn gæði íslenskrar talgreiningar á venjulegu tali, bæði á vönduðum upptökum eins og útvarpsefni og við algengar aðstæður í daglegu lífi. Sérstakir hlutar fyrir tal barna og eldra fólks koma til greina.

Dæmi: Orðavillutíðni, villur í nöfnum og tölum, sjálfsprottið tal og símtöl.

Erlendar fyrirmyndir: Open ASR Leaderboard, FLEURS (Google), CHiME-keppnirnar.

Talgervingarpróf

Að mæla hve eðlilegur og skýr almennur framburður helstu talgervilslausna er á íslensku, og hvernig þær ráða við texta sem eru erfiðir í lestri.

Dæmi: Skilgreindur verkefnalisti sem fólk með íslensku að móðurmáli metur með MUSHRA-samanburði frekar en einkunnakvarða. Mælt er meira en eðlileiki: hvort raddeinkenni haldist, hrynjandi, hæfi við samhengi og skiljanleiki. Sjálfvirk mat til viðbótar, til dæmis talgreinir á úttakið og MOS-líkön til viðmiðunar. Erfiðir textar eins og tölur sem beygjast eftir falli og kyni, dagsetningar, skammstafanir og erlend nöfn.

Erlendar fyrirmyndir: Blizzard Challenge, TTS Arena, VoiceMOS Challenge.

Málkunnátta og ritfærni

Íðorðapróf fyrir lykilsvið

Að mæla hvort líkön kunni og noti íslensk íðorð á sviðum þar sem íslenskan á í samkeppni við ensku.

Dæmi: Heilbrigðismál, lögfræði og stjórnsýsla, upplýsingatækni, fjármál, raunvísindi, sjávarútvegur og orkumál. Að þekkja íslenska heitið fyrir erlent hugtak, skýra merkingu þess og nota orðið rétt beygt í samhengi í stað enskrar slettu, til dæmis með stuðningi Íðorðabankans. Hluti prófsins þarf að vera nýr og lokaður, því íðorðasöfnin eru opin.

Fyrirmyndir: Íðorðapróf Miðeindar (innlent), íðorðakeppni WMT (2021, 2023 og 2025).

Ritfærniþróf í verkefnum

Að mæla hvort líkön skili íslenskum texta sem þjónar tilgangi sínum, umfram það að málið sé rétt.

Dæmi: Samantektir, umorðun á auðlesið mál, formleg bréf og svör við erindum, metin af fólki. Enskuleg áhrif á stíl, til dæmis nafnorð og lýsingarorð þar sem íslenska notar sagnorð.

Erlendar fyrirmyndir: WritingBench, SummEval.

Fyrirmælapróf

Að mæla hvort líkön fylgi fyrirmælum á íslensku nákvæmlega.

Dæmi: Fyrirmæli sem forrit getur sannreynt, svo sem um lengd, snið og orðaval, ásamt íslenskum skilyrðum eins og þéringum eða tilteknu málsniði.

Erlendar fyrirmyndir: IFEval, M-IFEval.

Þekking

3 próf

Fagþekkingarpróf eftir skólastigum

Að mæla hvort líkön búi yfir þeirri þekkingu sem er kennd og prófuð á íslensku, á ólíkum skólastigum og í starfsréttindanámi.

Dæmi: Spurningar úr íslensku námsefni og prófum á grunnskóla-, framhaldsskóla- og háskólastigi og úr starfsréttindaprófum, meðal annars í lögfræði og raungreinum. Spurningarnar þurfa að vera nýjar eða óbirtar.

Erlendar fyrirmyndir: INCLUDE, C-Eval (kínverska), KMMLU (kóreska).

Hversdagsþekking á Íslandi

Að mæla hvort líkön þekki daglegt líf og venjur á Íslandi.

Dæmi: Matur, siðir, frídagar og veður. Stækkun á íslenska hluta Global PIQA, sem nú er 100 dæmi.

Erlendar fyrirmyndir: Global PIQA, BLEnD.

Íslenska í alþjóðleg fjöltyngd próf

Að íslenska sé með í þeim alþjóðlegu prófum sem gjarnan eru notuð við þjálfun málíkana.

Dæmi: Fagþýðing og yfirferð á Global MMLU Lite (400 spurningar). MGSM (250 stærðfræðiþrautir) er þegar til vélþýtt og þarf aðeins yfirferð. Stækka má í allt MMLU síðar ef vel gengur.

Erlendar fyrirmyndir: Global MMLU, MGSM.

Verkefni og rökhugsun

2 próf

Svör við fyrirspurnum til stofnana

Að mæla hvort líkön geti svarað almennum fyrirspurnum til stofnana rétt út frá opinberum íslenskum upplýsingum, án þess að skálda.

Dæmi: Spurningar um skatta, bætur, leyfi og skráningar. Hvort svarið styðjist við heimildir, hvort líkanið viðurkenni þegar svarið finnst ekki og hvort það vísi á rétta stofnun. Spurningarnar eru orðaðar eins og fólk skrifar þær, líka fólk sem talar íslensku sem annað mál.

Erlendar fyrirmyndir: RAGTruth, FACTS Grounding (Google DeepMind), CRAG (Meta).

Forritunarþrautir á íslensku og ensku

Að mæla hvort líkön leysi forritunarþrautir jafn vel þegar þær eru settar fram á íslensku og á ensku, og hvernig lausnir þeirra standast samanburð við lausnir sem fólk hefur skrifað.

Dæmi: Þrautir úr forritunarkeppnum framhaldsskóla, þar sem hluta er haldið eftir til prófunar og nýjar þrautir bætast við á hverju ári. Lausnir líkana má bera saman við lausnir keppenda. Mælt er hvort lausnin virki fyrir öll inntök og hve góð hún er, til dæmis lengd og tíma- og minnisflækja.

Fyrirmyndir: Völundarhúsapróf HÍ (MazeEval), LiveCodeBench.

Bjaga- og öryggispróf

Að mæla hvort líkön svari eftir staðalmyndum um hópa í íslensku samfélagi, þar á meðal eftir kyni, hvort þau séu örugg í almenntri notkun á íslensku og hvort þau endurtaki algengar ranghugmyndir.

Dæmi: BBQ-snið lagað að íslenskum aðstæðum. Starfsheiti og samræmi, kynjaval í þýðingum, ráðgjöf eftir kyni og lýsingar á fólki, til dæmis meðmælabréf fyrir Jón og Guðrúnu með sömu ferilskrá. Áhættuflokkar á borð við Alluminate og ranghugmyndir á borð við TruthfulQA. Öryggispróf sem unnið er að hjá Háskóla Íslands ná yfir hluta af þessu. Erlendar fyrirmyndir: BBQ og KoBBQ, WinoMT, Alluminate, TruthfulQA.

Starfsgreinar

3 próf

Opinber þjónusta og stjórnsýsla

Að mæla hvort líkön geti unnið dæmigerð verkefni starfsfólks stofnana á íslensku, af sambærilegum gæðum og fagfólk.

Dæmi: Að svara erindi borgara, skrifa ákvarðanir eða formleg bréf á skýru máli, skrifa minnisblöð, draga saman umsagnir um mál.

Erlendar fyrirmyndir: GDPval, τ^2 -bench.

Lögfræði

Að mæla hvort líkön geti unnið lögfræðileg verkefni á íslensku og út frá íslenskum rétti.

Dæmi: Að draga saman dóm, finna viðeigandi lagaákvæði, útskýra réttarstöðu á mannamáli.

Erlendar fyrirmyndir: LegalBench, LexGLUE.

Heilbrigðisþjónusta

Að mæla hvort líkön geti stutt íslenska heilbrigðisþjónustu og samskipti við sjúklinga á réttan og öruggan hátt.

Dæmi: Að útskýra upplýsingar fyrir sjúklingi á skiljanlegu máli eða svara spurningum um heilbrigðiskerfið. Aðeins eru notuð tilbúin dæmi, ekki raunveruleg sjúklingagögn.

Erlendar fyrirmyndir: HealthBench (OpenAI), MedQA.

6 Næstu skref

Að loknum umsagnarfresti fer verkefnastjórnin yfir þau sjónarmið sem berast og ákveður hvaða próf fara í opið kall. Kallið verður auglýst opinberlega og er áætlað að það verði opið frá 2. til 13. nóvember. Þar kemur fram verklag, tímalína og hvernig tillögur verða metnar.

Í verkefnastjórn máltækniáætlunar sitja Lilja Dögg Jónsdóttir, Magga Dóra Ragnarsdóttir og Hallgrímur J. Ámundason. Varamenn eru Ragnheiður H. Magnúsdóttir og Óttar Kolbeinsson Proppé.

Almannarómur, almannaromur@almannaromur.is