

## Kuidas tagada eestikeelne ja -meelne tehisaru?

*Marek Tamm ja Krister Kruusmaa*

Ei nõua kuigi elavat kujutlusvõimet, et näha juba lähitulevikus maailma, mis on lahutamatult põimunud kõikvõimalike tehisaru rakendustega. Nii nagu oleme praeguseks harjunud elama digitaalselt vahendatud eluilmas, nii tuleb meil õppida elama tehisaru võimestatud keskkonnas.

Kui selle visiooniga nõustuda, siis on ilmselge, et eesti keele ja kultuuri jaoks pole tänapäeval palju tähtsamaid küsimusi kui see, kuidas tagada, et see tehisaruga tembitud uus maailm oleks eestikeelne ja eestlasele kodune.

Esmane eeldus eestikeelse tehisaru arendamiseks on selliste masinõppe mudelite olemasolu, mis on kohandatud eesti keele eripäradega. Eestil puudub võime iseseisvalt suuri keelemudeleid algusest lõpuni arendada, seega jääb üle kaks võimalust. Esimene on pakkuda suurtele arendajatele oma keeleandmeid ja eksperditeadmisi vastutasuks mudelite kohandamise eest. Oluline on sealjuures soosida avatud lähtekoodiga lahendusi, mida saame iseseisvalt mugandada kohalikuks kasutuseks.

Teine võimalus on olemasolevaid avatud mudeleid peenhäälestada eestikeelsete andmetega, nii et need mõistaks meie keelt paremini. Mõlemal juhul on vältimatu koguda ja kasutada kvaliteetseid keeleandmeid, see tähendab kaasata ilukirjandust, ajakirjandust, teadustekste, rahvaluulet ja muid heatasemelise keelekasutuse näiteid.

Kuid lisaks eesti keele valdamisele on oluline, et tehisaru mõistaks konteksti, viiteid ja väärtusi, mis on eestlastele olulised. Selle saavutamiseks on vaja kahte asja: esiteks sisendandmeid, mis hõlmavad Eesti kultuuriajalugu ja tänapäevast kultuurikonteksti, ning teiseks kvaliteetseid ja kättesaadavaid mudeleid, mida on nende andmetega treenitud.

Seega peame tehisarule andma selliseid andmeid, mis peegeldavad võimalikult hästi eestlaste maailmapilti ja kogemust. Kõik see ilus plaan eeldab aga autoriõiguslike probleemide adekvaatset lahendust. Meil ei ole vaja neid lahendusi otsida tühjalt kohalt, vaid saame õppida teiste riikide kogemusest. Ühed võimalikud eeskujud on Island ja Norra, kes teevad oma keelte kaitsmise nimel koostööd suurfirmadega, aga on ka edukalt kaasanud erialaorganisatsioone ja loomeinimesi, et nendega ühistes põhimõtetes kokku leppida. Konkreetset tuge võiks pakkuda Islandil koostatud [teejuht](#) selle kohta, kuidas väikekultuurid saavad keele- ja kultuuritundlikke keelemudeleid arendada.

Seadusloome tasandil on Eestil vaja leida uued lahendused, kuidas tehisaru treenimiseks sisu kasutamine õiglaselt ja paindlikult reguleerida. Vajalik on ka Euroopa Liidu ühisesse keeleressursside hankesse panustamine, et tulevased avatud mudelid sisaldaksid rohkem eesti keelt. Eesti huvides on tagada, et meie kultuuripärandi digitaalne kättesaadavus teeniks avalikku huvi ega kahjustaks loojaid.

# TEADLASE PILGUGA

Tehisarule eesti keele ja eestluse õpetamise kõrval pole vähem tähtis see, kuidas me selle oskusi ja teadmisi hindame. Keelemudelite kvaliteeti hinnatakse peamiselt võrdlusandmestike abil, mis kujutavad endast justkui suuri kontrolltõid tehisarule. Tehisaru kasutuselevõtt sellistes valdkondades nagu näiteks haridus eeldab, et me oskame mõõta eri mudelite oskusi, sisuteadmisi ja väärtushinnanguid. Seega tuleks kiiresti luua juurde valdkondlikke eestikeelseid võrdlusandmestikke. See ei vaja niivõrd tehnilisi, kuivõrd sisulisi eksperditeadmisi – eriti oluline on, et just humanitaar- ja sotsiaalteaduste asjatundjad nende loomisse panustaksid.

Kui tahame, et tehisaru arvestaks eestlastega, on meil vaja see enda vajadustega joondada. Kõige parem viis selleks on õiguslike takistuste ületamine andmete jagamisel ja kohapealse hindamisvõime loomine. Hästi joondatud tehisaru võiks aga ühtlasi käsitleda kui uut võimalust Eesti kultuuripärandi säilitamiseks. Suur keelemudel on teataval määral nagu andmebaas, kuhu saab talletada keelt ja kultuuri puudutavat infot. Kuid erinevalt tavapärasest andmebaasist suudab mudel nende andmete põhjal teha üldistusi ja tuletusi. Nii võib tehisarust kujuneda tulevikus omalaadi virtuaalne kultuurikiht, millest uued põlvned saavad eesti keeles ammutada oma kultuurialaseid teadmisi.

Kokkuvõtlikult on tehisaru keelelise ja kultuurilise joondamise puhul võtmetähtsusega leida mõistlik tasakaal kahe eesmärgi vahel: tagada heatasemeline eestikeelne ja -meelne tehisaru, samal ajal tagades, et kellegi autoriõiguseid ei saaks mõtlematult rikutud. Kui see õnnestub, võib tehisarust saada eesti kultuuri uus oluline väljendusvorm ja suure tähtsusega keelehoidja.

[Ilmunud ajalehes Postimees 15. märtsil 2025](#)