

Kas tehisaru räägib oskuskeelt? Tehistaip kui tugi, partner ja segaja terminitöös¹

Karit Jäärats

Tartu Ülikooli Pärnu kolledži inglise keele nooremlektor

Artiklis käsitlen tehisaruna (TA) suuri keelemudeleid ning nende baasil arendatud kommertsvestlusplatvorme, näiteks ChatGPT, Google Gemini, Claude, Microsoft Copilot jt. TA arendamine ja kasutamine on olnud kiire ja laialdane ning ligipääs TA-le on pea igapähe, kelle käsutuses veebiühendusega arvuti või nutiseade. TA kasutamiseks ei ole vaja sügavamaid teadmisi keelemudelite toimimisest, nende infotehnoloogilisest poolest ega andmehaldusest.²

Kuna TA on loodud genereerima eelkõige teksti, puutub iga TA kasutaja tahes-tahtmata kokku keele, sageli mingil määral ka oskuskeelega. Kuna TA-le on järjepidevalt õpetatud eesti keelt ja selle tase on üha parem, siis eeldab lõppkasutaja suhteliselt häid tulemusi.³ Selge on see, et treenimine jätkub ja arendajatele tuleb anda aega. Väikekeelte puhul on iseäranis oluline sisestada mudelitele võimalikult palju autentset ja korrektset materjali, et TA saaks piisavalt ja terviklikke andmeid nii kõne- kui kirjakeele ja erinevate registrite, sh erialakeele kohta. Samas peab arvestama, et väikekeelte vähene keeleressurs ei võimalda mudelitel toimida suurkeeltega samal tasemel.⁴ Väikekeelte jaoks puuduvad terviklikud hindamisstandardid ja piisavad treenimisandmed, mis raskendab mudeli toimivuse süstemaatilist hindamist ja parandamist.⁵

Meeles tuleb pidada sedagi, et TA koolitamiseks kasutatavad andmed võivad olla subjektiivsed ja eelarvamuslikud ning põhinevad peamiselt veebis kajastatavatel ja vabalt kättesaadavatel tekstidel. Kuigi TA-le saab anda konteksti, ei ole veel teada, kuidas täpselt see konteksti kasutab või sellest aru saab. See mõjutab ka TA võimet mõista ja genereerida täpseid termineid.⁶ Ikka ja jälle on täheldatud, et TA eksib, eriti kui küsida fakte või kasutada TA-d kui otsimootorit. See tuleneb asjaolust, et TA pakub tõenäolist teavet ning võib seega anda tekstides hästi kõlavaid ja professionaalsena näivaid, kuid väljamõeldud sõnu.

TA-l puudub sisseehitatud kontrollisüsteem ja teadlikkus eksimisest. Kui TA-le esitatud viip on ähmane või liiga üldine, paneb see TA täitma lünki tõenäoliste vastuste ja oletustega. Bluffida võib TA ka aegunud või piiratud treenimisandmete tõttu.⁵

¹ Artikkel põhineb Tartu Ülikooli, Eesti Maaülikooli ja Kaitseväge Akadeemia ühiselt korraldatava terminipäevaku 10. toimumiskorral peetud ettekandel.

² **T. T. Procko, T. Elvira, O. Ochoa.** Dawn of the dialogue: AI's leap from lab to living room. 2024.

³ **K. Taro.** Tehisaru mudelid vajavad kvaliteetset eesti keelt. 2025.

⁴ **M. Žordania.** Targem tehnik, targem elu: tehisaru võimalused ja ohud. – Õiguskeel 2025, nr 3.

⁵ **R. Dargis jt.** Evaluating open-source LLMs in low-resource languages: insights from Latvian high school exams. Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities. 2024.

⁶ **K. M. Adavala, O. Adavala.** Domain-specific knowledge and context in large language models: challenges, concerns, and solutions. 2025.

Siin tuleb TA platvormidele appi avatud standard Model Context Protocol⁷, mille abil saavad mudelid turvaliselt ja standarditud viisil kasutada väliseid andmeallikaid ja tööriistu. See toimib justkui vahelüli TA ja lisakonteksti vahel, et kasutaja saaks TA-lt küsida ka treenimisandmetest puuduvat ajakohast teavet, näiteks praeguse nädala ilma või euro ja dollari hetkekurssi.

Õppe- ja teadustöös, mis paratamatult kätkeb ka tööd terminitega, võib TA täita kolme rolli: tugi, partner ja segaja. Toeks võib TA-d pidada, kui kasutaja teeb tekstiga suurema töö ära ise, TA vaid kinnitab ja kontrollib. Näiteks võib masintõlke järelkontrollis kasutada mudelit nn peegeldajana, kui kasutajal endal on õiged terminid teada, kuid ta soovib teksti ülevaatamisel veenduda terminite ühtlases kasutamises, definitsioonide kooskõlas või mõne fraasi sobivuses. Toeks on TA kindlasti siis, kui ollakse ise valdkonna asjatundja ning osatakse anda TA-le viibas piisavalt palju infot. See võib hõlmata näiteks seda, mis valdkonnast termin on, mis on selle tõlkevaste, milline on eelitermin ning millele peaks vastus tuginema, kas terminibaasile, standardile, erialatekstidele või teaduskirjandusele. Assisteerivas rollis on TA suuteline kiiresti ja ootustele vastavalt muutma etteantud teksti näiteks stiililt ametlikumaks või vähem ametlikuks, tuvastama sõnakordusi ja vigu jms. Tallinna Prantsuse Lütseumi õpetaja Merike Saar sõnas Õpetajate Lehes: „Kui tehisintellekt hakkab tulevikus ka õpilastele soovitusi andma nii, et ta ei paku valmis vastuseid, vaid suunab ja nügib mõtlema, on sellel väga suur potentsiaal. Ootan huviga ka õpilastele mõeldud litsentsi.“⁸

Enamaks kui assistendiks saab TA, kui kirjutaja enda panus on tekstitöös väiksem ja TA roll suurem. Sel juhul kasutatakse TA-d partnerina ideede kogumiseks, piisava konteksti saamiseks termini, teema, valdkonna, nende seoste jms kohta. Kui TA-lt küsida materjali, taustateavet ja allikaid, annab see üldjuhul kiiresti ja struktureeritult termini kohta ülevaate ning aitab läbi mõelda variandid, leida tüüpilised segiajamised ja sõnastada lauseid vastavalt soovitule. Küllap on nii õppijad ja õpetav personal kui uurimuste koostajad juba saanud TA-lt hulganisti abi akadeemilises kirjutamises, sest TA on väga võimekas tekste analüüsima.⁹ Ka teksti kompaktsemaks muutmise ning kokkuvõtte tegemisega saavad keelemudelid suurepäraselt hakkama. Näiteks võib projektitaotluste täitmisel olla TA samuti tõeline partner, kui on vaja püsida etteantud tähe märkide piires ning hoida TA abil kokku aega ja vaeva. Tekstide analüüsimise kõrval ei saa eirata ka TA suutlikkust tekste tõlkida. Siin tuleb aga eristada suuri keelemudeleid ja masintõlget, sest esimesed on võimelised teksti looma ning seega ka termineid pakkuma ja teksti ümber sõnastama, kuid masintõlke puhul on fookus eelkõige terminite vastendamisel ja tõlkevariandi valikul.

Paraku võib TA töös terminite ja tekstidega olla nõu ja abi andmise kõrval ka tõeline segaduse külvaja. Kuigi TA-d kasutatakse tõhusalt näiteks võõrkeelsest materjalist kokkuvõtte tegemiseks, rõhutas Zürichi Ülikooli teadur dr Sara Petchey oma 2025. aasta septembris Tartu Ülikoolis peetud koolitusel „Õppesisu türanniast: mida, kui palju ja kuidas õppeaines õpetada“, et kui kasutaja laseb terminirohkest teadusartiklist olulisima leida TA-l, annab ta sellega

⁷ Model Context Protocol. <https://modelcontextprotocol.io>. Vaadatud 07.02.2026.

⁸ S. Pärismaa. Tehisaru kasutades ei tohiks oma loovus kängu jääda. – Õpetajate Leht 2026, 20. jaanuar.

⁹ B. R. Cibu jt. Exploring the educational applications of large language models: a systematic review and topic analysis. 2025.

otsustuspädevuse TA-le, loobub oma mõttekäigust ja individuaalsusest ega saa olla seejuures kindel, et TA rõhutab tekstis tähtsaimana sama, mida kasutaja ise oleks esile toonud.

Olenevalt keelesuunast, andmetest ja lünkade täitmise vajadusest võib TA tõlkimisel piirduda sõnasõnalise väljendusviisiga ning anda seega kasutajale usaldusväärsema näivaid, kuid ometi kohmakaid, vähe levinud või lausa vääri vasteid. Näiteks on TA tõlkinud üliõpilase kodutöös ingliskeelse termini *composite service* eesti keelde „komposiitteenuseks“. ÕS 2025 järgi on „komposiit“ „komposiitmaterjali“ sünonüüm ning tähendab liitmaterjali, mille üks koostisosa on kõva ja tugev, teine plastne ja elastne. Inglise keeles võib aga omadussõna *composite* tähendada muu hulgas lihtsalt midagi, mis koosneb eri osadest, seega termini *composite service* all on mõeldud enam kui üht teenust hõlmavat liitteenust, mitmikteenust või kompleksteenust. Kui eriala alles omandav õppija kopeerib TA genereeritud mehaaniliselt, võivad jääda tõlgitud tekstidesse ka sellised vead nagu inglise majandustermin *triple bottom line* eesti keeles „kolmekordne kasumiaruanne“, kuigi õige oleks „kolmekordne tulem“. TA muutub segajaks ka siis, kui sel puudub piisav kontekst või kui kasutaja viip on liiga üldine, sõnastatud mitmeti mõistetavalt või ähmaselt. Ühtlasi võib TA loodud tekst olla oma ala eksperdile, teadlasele või õppejõule liiga algeline ja pealiskaudne nii sisus kui keelekasutuses.

Vaikimisi on TA nõustaja, nn meele järgi olija. TA-d on arendatud nii, et hoida kasutajat dialoogis, teda suunata ja pakkuda muudkui uusi võimalusi. See on kasutaja jaoks hea ja mugav, kui TA annab suuna kätte ja juhib teda vajalikuni. Nii saab kasutaja kirjutada viipa järjest konkreetsema sõnastusega ja vahest jõuda tänu TA-le ka üksikasjade või vaatenurkadeni, mille peale ta iseseisvalt ei oleks tulnud. Kui aga TA tekstis ei ole silmanähtavaid loogika- ega sisuvigu, ei pruugi kasutaja tabada ka terminieksimusi.

Vaatleme näiteks õigusterminite „preambul“. Enam kui kümme aastat tagasi avaldatud „Keelesäutsus“ sõnab keeleteoimetaja Maris Jõks, et ÕS-i järgi on õige vaid „preambul“, mitte „preambula“, nagu esineb vahel üld- ja kõnekeeles. Jõks selgitab termini ladinakeelset päritolu ja lisab, et kuigi 2005. a ladina-eesti õigussõnastikus on *preambula* kirjas kui ladinakeelne õigustermin ja me võiksime seda eesti keeles võõrsõnana *praeambulum*’i lühenenud vormina kasutada, on meil olemas „preambul“ ja uuesti laenamiseks puudub põhjus. Ta toonitab, et eelistada võiks kuju „preambul“ ja tähele panna, et see sõna käändub *preambuli* : *preambulit* ilma *a*-lõputa üheski vormis.¹⁰ Katsetasin MS Copilotit küsimusega, kumb termin on õige, ja sain vastuse, et eesti keeles peetakse õigemaks ja normikohaseks sõnaks „preambul“. Copilot tõi esile ka venekeelse *преамбула* mõju eesti keeles „preambula“ kasutamisele ning rõhutas, et keelekorraldus on järjekindlalt hoidnud ladinapärasest kuju „preambul“. Kui küsisin, kust on see info pärit, loetles Copilot allikatena „Ametniku soovitusõnastikku“, ERRi „Keelesäutsu“, „Eesti keele seletavat sõnaraamatut“ ja Vikipeediat. Ühtegi viidet ÕS-ile ei esinenud ega ka soovitus sellest järele vaadata. Uusima õigekeelsussõnaraamatu veebiversioonist leiab aga, et „preambul“ ja „preambula“ on nüüd mõlemad sobivad kirjakujud. See näitab, kuidas TA võib ajada kasutaja segadusse, toonitades normikohasust ja

¹⁰ M. Jõks. Kas preambul või preambula? – Keelesäutsud I. Tartu: Keelehooldakeskus, 2014, lk 31–32.

keelekorraldust, kuid jättes viitamata kirjakeele normi allikale – uusimale ÕS-ile. Oskuskeeles varieerumist ei soosita, seega on õigusvaldkonna eelistermin ikka „preambul“.

Kuidas siis tunda ära, kas TA räägib oskuskeelt, ja tagada, et see oleks kasutaja jaoks rohkem tugi ja partner kui segaja? TA kuvatava emakeelse teksti puhul tuleb hinnata, kas pakutud termin kõlab loomulikult, kas see tundub keeleliselt ja mõisteliselt loogiline. Kontrollida tuleb, kas oskussõnade tähendus vastab kontekstile ning kas terminid on üheselt mõistetavad.

Terminil võib olla kultuuriline või institutsiooniline eripära, näiteks võib TA kindlameelselt kasutada Eesti kohta käivas tekstis sõna „ombudsman“ või „rahvakaitaja“, kuigi meil on tegu õiguskantsleriga. Veenduda võiks selleski, kas termin on eesti keeles juba kasutusel, ning vaadata sõnastikesse, terminibaasidesse, erialakirjandusse usaldusväärse kinnituse saamiseks. Eelneva põhjal saab kasutaja teha otsuse, kas, millal ja kuidas terminit kasutada, või hoopis täpsustada dialoogi TA-ga. Julgustan panema TA-d proovile, minema dialoogis süvitsi ja jõudma seeläbi ehk suurema täpsuseni.

Samavõrd oluline on TA pedagoogiline kasutamine. Praegused mudelid ei võimalda õigesti sisestada eesti keeles näiteks jutumärke ja tsitaatsõnu. Kasutuskogemuse põhjal on TA teinud näiteks vea sõna „termin“ mitmuse osastavas ja kuvanud „termeneid“ ning eksinud kokku- ja lahkukirjutuses, nt „eesti keelne“. Õigekirja peab tundma teksti kasutaja või avaldaja ise ning sellele peab tähelepanu pöörama õpetaja või õppejõud. TA ajastul oleme silmitsi katsumusega, kuidas tagada sügav õppimine ja tähelepanu ning kuidas integreerida TA-d õppetöösse nii, et see looks peale kiirvastuste ka uusi ja harivaid võimalusi. Ka keeleteadlane, Tallinna Ülikooli emeriitteadur Peep Nemvalts on rõhutanud, et terminiõpetus peaks Eesti ülikoolide õppekavadesse kuuluma eraldi aina või erialaainetesse põimituna.

Keelemudelite treenimine valdkonnakorpusetega võib parandada nende arusaamist erialasest terminivarast ja arendada kontekstuaalset täpsust. Mõistagi tuleb seejuures anda arendajatele aega ning väikekeelte puhul võimalikult palju autentset ja korrektset materjali, et TA saaks piisavalt ja terviklikke andmeid nii kõne- kui kirjakeele ja erinevate registrite, sh oskuskeeles kohta. Samas peab arvestama, et väikekeelte vähene keeleressurss ei võimalda mudelitel toimida suurkeeltega samal tasemel.¹¹ TA treenimist teaduskirjandusega võib samuti pidada terminitöö osaks. Muidu on oht, et mõned valdkonnad, mille kohta on palju väärtuslikku paberkirjandust, jäävad keelemudelites alaesindatuks.

Kas keelt saab taandada mudeliks, andmekogumiks, on aga omaette küsimus inimliku ja tehisliku, inimese-masina kontrastis ja koosluses. Soovime ju keelemudelites näha tehisarukat keelt, mitte lihtsalt tehiskeelt. Selge on see, et oma valdkonna termineid peab tundma ja kontrollima eelkõige inimene ise. Vastavalt 10. terminipäevaku juhtteemale „Inimene-masinsüsteem terminitöös“ – las jääda inimene selles kolmikus ikka esikohale ja viimaseks otsustajaks.

¹¹ A. Harrik. Keeletehnoloog: seitsme aasta pärast mõistab tehisarvuti liivikeelset sarkasmi. 2024.