ESUKA – JEFUL 2025, 16–1: 145–172

TEKSTIROBOTID JA LAIENDATUD VAIM. KUIDAS MÕISTA TEHISINTELLEKTI ROLLI AKADEEMILISE TEKSTI LOOMES?

Vivian Puusepp
Tartu Ülikool, EE
vivian.puusepp@ut.ee

Kokkuvõte. Siinses artiklis arutlen kahe küsimuse üle: kui õppija kasutab akadeemilise teksti kirjutamisel tekstiroboti abi, siis (1) kuidas võiks mõista tekstiroboti rolli selles protsessis ning (2) kui akadeemilise teksti kirjutamist käsitleda laiendatud kognitiivse protsessina, siis kas tekstirobot laiendab inimvaimu? Artiklis tekstiroboti võimalikke rolle kaaludes selgub, et tekstirobotit ei saa paigutada ühtegi kategooriasse, millega oleme akadeemilise kirjutamise kontekstis seni harjunud: see pole allikas, (kaas)autor, vestluspartner ega pelgalt tööriist. Teise küsimuse üle arutledes võtan aluseks laiendatud vaimu (ingl *extended mind*) lähenemise, mille järgi laienevad inimese vaimsed protsessid mõnikord väljapoole tema aju ja keha, hõlmates väliseid protsesse või vahendeid kognitiivse süsteemi osana. Artiklis selgitan, et sellest, kuivõrd vastab tekstiroboti kasutusviis usaldusväärse sidususe (ingl *reliable coupling*) kriteeriumidele, sõltub ka tekstiroboti kasutamise mõju inimvaimule. Arutluskäigust ilmneb, et inimvaimu laiendamiseks on määravaim usalduse kriteerium, mis eeldab kasutaja oskust tekstiroboti väljundit kriitiliselt hinnata. Kui kasutaja usaldab tekstirobotit pimesi ja kasutab väljundit seda kriitiliselt hindamata, siis tekstiroboti kasutamine pigem ahendab inimvaimu, pärssides kasutaja kirjutamis- ja mõtlemisoskuse arengut.

Märksõnad: suured keelemudelid, tekstirobot, tehisintellekt, TI, laiendatud vaim, akadeemiline tekst, kirjutamine, õppimine

DOI: https://doi.org/10.12697/jeful.2025.16.1.06

1. Sissejuhatus

Digiuuendused on toonud kaasa suuri muudatusi selles, kuidas inimesed tekste kirjutavad. Tekstirobotite laialdane kasutuselevõtt on muutmas õpetamise põhialuseid ja arusaamu nii haridusest kui

ka akadeemilise teksti kirjutamisest. Pärast ChatGPT-3[1] väljatulekut 2022. aasta novembris selgus, et see tekstirobot suudab genereerida teksti, mis sarnaneb inimese loodud tekstiga. Seetõttu leidsid paljud haridustöötajad end olukorrast, kus tuli teksti kirjutamise roll ümber mõtestada ning kaaluda, milliseid kirjalikke ülesandeid on uutes tingimustes mõttekas õppijatele anda. Praegu seisavad õpetajad ja õppejõud silmitsi küsimusega, kuidas juhendada õppijaid kaasama tekstirobotit õppe- ja kirjutamisprotsessi nii, et see õppimist toetaks, mitte pärsiks.

Artiklis uurin, kuidas tuleks käsitleda tekstirobotite rolli sellises õppeprotsessis, mille käigus kirjutatakse akadeemilisi tekste. Akadeemilise teksti all pean silmas argumenteerivat ja allikatel põhinevat teksti, mida kirjutatakse ülikoolis jt haridusasutustes. Nende tekstide hulka kuuluvad nt argumenteeriv essee, seminari- või lõputöö, teadusartikkel, retsensioon vms. (Hint, Jürine & Leijen 2022: 327) Lähtun uurimisel küsimusest, et kui õppija kirjutab akadeemilist teksti ja kasutab selleks tekstiroboti abi, siis kuidas mõista tekstiroboti rolli selles protsessis. Kusjuures tekstirobotit võidakse kasutada eri viisidel ja ulatuses. Esmalt arutlen, kas tekstirobotit võiks käsitleda ühe kirjaliku allikana teiste seas, teksti (kaas)autori, tööriista või vestluspartnerina. Seejärel laiendan küsimuse fookust ja vaatlen tervikuna kirjutamis- ehk tekstiloomeprotsessi, milles osalevad nii inimene kui ka tekstirobot. Nendevahelise koostöö mõistmiseks tutvustan vaimufilosoofias ja kognitiivteaduses levinud laiendatud vaimu (ingl *extended mind*) lähenemist (Clark & Chalmers 1998, 2009).

Kuna tekstirobotite kasutamine akadeemilise teksti kirjutamise abivahendina muutub üha tavalisemaks ning ühtlasi on Eestis algatatud TI-Hüpe[2], on oluline mõista, kuidas tekstirobotite kasutamine mõjutab lisaks õppija ja tekstiroboti ühiselt valminud tekstidele ka õppija vaimu.

---

1 ChatGPT (*Chat Generative Pre-Trained Transformer*) on OpenAI loodud generatiivne tekstirobot, mis on disainitud jäljendama inimestevahelisi kirjalikke vestlusi. ChatGPT rakendus tehti tavakasutajatele tasuta kättesaadavaks 2022. aasta novembris ning väidetavalt kogus see kahe kuuga üle saja miljoni kasutaja, mistõttu mitmed meediakanalid kirjeldasid seda kui ajaloo kiireima tarbijate kasvuga rakendust (Greene 2024). Pärast ChatGPT rakenduse avalikustamist on hoogustunud võidujooks eri ettevõtete vahel, mis arendavad generatiivseid tekstiroboteid, nt Google'i Gemini ja Microsofti Copilot.

2 TI-Hüpe 2025 on Tiigrihüppest inspireeritud ning avaliku ja erasektori koostöös sündiv haridusprogramm, mille raames antakse Eesti koolinoortele ja õpetajatele juurdepääs maailma juhtivatele tehisintellekti õpirakendustele. Vt lähemalt: https://tihupe.ee/ (20.03.2025).

Lähtudes üldisemast ideest, et tehisintellekti (TI) võib käsitleda inimvaimu laiendusena (Hernández-Orallo 2025; Pellegrino & Garasic 2020), keskendun artiklis suurtel keelemudelitel põhinevate tekstirobotite rolli uurimisele akadeemilise teksti kirjutamisel. Artikli eesmärk on välja selgitada, millised tingimused peavad olema täidetud selleks, et tekstiroboti tegevust saaks pidada inimvaimu laiendavaks. Sealjuures on minu eesmärk näidata, et tekstirobotid võivad inimvaimu ka hoopis ahendada. See oht on olemas siis, kui õppija laseb tekstirobotil töö enda eest ära kirjutada, sest õppija näiteks (1) ei oma piisavalt akadeemiliste tekstide kirjutamise vilumust või teemakohaseid teadmisi, (2) ei oska koostada asjakohaseid viipasid (ka prompt, ingl *prompt*) või (3) õppijal puudub võime tekstiroboti väljundit kriitiliselt hinnata. Loodan, et artikkel pakub teoreetiliselt huvitavat ja praktiliselt kasulikku raamistikku teadlastele, õpetajatele, õppejõududele ja õppijatele. Raamistik võiks neil aidata otsustada, millal ja kuidas tasub tekstirobotit akadeemilise teksti kirjutamisel kasutada ning millal oleks õppijal kasulikum oma tekstiga iseseisvalt tegelda.

2. Tekstiroboti võimalikud rollid akadeemilise teksti kirjutamisel

2.1. Tekstiroboti väljund ei ole allikas

Akadeemilises maailmas ollakse valdavalt seisukohal, et tekstiroboti genereeritud teksti esitamine enda nime alt on nagu teise inimese loodud teksti esitamine enda nimega ning seda tuleb käsitleda akadeemilise petturlusena. Ülikoolid üle maailma, sealjuures Eestis näiteks Tartu Ülikool (TÜ), Tallinna Ülikool (TLÜ) ja Tallinna Tehnikaülikool (TalTech), on töötanud välja suunised tekstirobotite kasutamiseks õppetöös. Neis nõutakse, et kui tekstiroboti genereeritud teksti oma kirjalikus töös kasutatakse, tuleb sellele korrektselt viidata. (Puusepp 2025: lisa 2, 3, 4, 6)

Viitamiskohustuse tõttu võib esmapilgul tunduda, et tekstiroboti genereeritud teksti saab käsitleda kirjaliku allikana. Siiski kirjutatakse ülikoolide suunistes selgelt, et allikana tekstirobotit käsitleda ei tohi (näited 1 ja 2).

(1) „(t)ekstirobot **ei ole avaldatud allikas** [autori rõhutus], vaid tekstiloomise mudel, mis võib olenevalt suhtlusolukorrast anda erinevaid vastuseid“[3] (Puusepp 2025: lisa 2).

(2) „Tehisintellekt **ei sobi iseseisvaks allikaks** [autori rõhutus], seega õppijal tuleb tõenduseks ikkagi viidata tegelikele inimeste loodud algallikatele. Iseseisva algallikana on tehisintellekti poolt loodud sisu võimalik viidata juhul kui antud töö eesmärgiks on uurida või luua tehisintellekti kasutust.“ (Puusepp 2025: lisa 6)

Põhjus võib olla ChatGPT jt tekstirobotite töötamispõhimõttes. Tekstirobot genereerib iga kord uue tekstiväljundi, lähtudes tohutust tekstimassist, mille peal on seda treenitud, ning sisestatud lähtetekstist ehk viibast, võttes vahel arvesse ka pooleliolevat vestlust. See tähendab, et tagantjärele ei ole võimalik sama viipa kasutades kontrollida, millise väljundi on tekstirobot varem genereerinud. Seega pole tekstiroboti genereeritud väljund avalik allikas, mida saaks kontrollida nii, nagu on võimalik kontrollida mistahes muu internetis või trükis avaldatud allika viidet.

2.2. Tekstirobot kui teksti (kaas)autor?

Kui tekstiroboti kasutaja on olnud piisavalt osav, on keeruline tuvastada, kas ja mil määral on ta teksti loomisel kasutanud tekstiroboti abi (vt nt Makiev et al. 2023). Sealjuures pole tuvastamine ka tervitatav. Eesti haridussüsteem sai oma esimesed vitsad siis, kui Haridus- ja Noorteamet süüdistas Saaremaa gümnaasiumi abiturienti tehisintellekti abiga eksami tegemises, kuid kaotas kohtus vaidluse (Sepp 2024). Samuti on näiteks TÜ avaldanud tuvastamisvastase seisukoha: „TI tuvastuse tarkvara kasutamine õppetöös üliõpilaste kirjalike tööde kontrollimiseks ei ole põhjendatud ega soovituslik.“ (Hint 2024) Kuna suur osa õppijatest kasutab kirjutamisel tehisintellekti abi (vt nt Digital

3 TÜ suunist on pärast esimese versiooni ilmumist muudetud ning töörühm tegeleb selle uuendamisega ka edaspidi. Artiklis viitan esimesele, 2023. aasta aprillis avaldatud versioonile ning 2024. septembris uuendatud versioonile. Kuna ülikoolide suuniseid muudetakse, on artikli valmimise ajal kehtivad suunised, millele siin artiklis viidatakse, avaldatud artikli lisadena teadusandmete publitseerimise repositooriumis DataDOI (vt Puusepp 2025).

Education Council Global AI Student Survey 2024), on üha olulisem arutleda selle üle, kuidas mõtestada tehisintellekti rolli kirjutamisprotsessis. Kui tekst on loodud tekstiroboti abiga, siis kas tekstirobotit võiks käsitleda teksti autori või kaasautorina?

Pärast ChatGPT-3 rakenduse väljatulekut hakkas ilmuma teadusartikleid ja muud kirjandust, mille autorite loetelus oli ChatGPT. ChatGPT kaasautoriks nimetamine tõi akadeemilises maailmas peagi kaasa jõulise vastuseisu. Mitu juhtivat teadusajakirja ja -kirjastust keelas tekstirobotite nimetamise autoriks või kaasautoriks (Sample 2023; Stokel-Walker 2023). Näiteks ajakirja Nature toimetus avaldas seisukoha ChatGPT jt keelemudeli rakendustel põhinevate tööriistade kasutamise kohta kõigis Nature Springeri ajakirjades:

> „Esiteks, ühtegi suurtel keelemudelitel põhinevat tööriista ei aktsepteerita teadusartikli autorina. See tuleneb asjaolust, et igasugune autorluse omistamine toob kaasa vastutuskohustuse tehtud töö eest ja tehisintellekti tööriistad ei saa sellist vastutust võtta. Teiseks, teadlased, kes kasutavad suurtel keelemudelitel põhinevaid tööriistu, peaksid nende kasutamist dokumenteerima meetodite või tänuavalduste rubriigis. Kui artikkel ei sisalda selliseid rubriike, saab suurte keelemudelite kasutamise dokumenteerimiseks kasutada artikli sissejuhatust või mõnd muud sobivat osa.“ (Nature Editorial 2023) [Autori tõlge]

Sarnase seisukoha on võtnud ka ülikoolid jt haridusasutused (näited 3–5).

(3) „(…) tehisintellekti ei tohiks kunagi nimetada teksti või muu kirjaliku väljundi autoriks. Tehisintellekt ei saa võtta vastutust teksti sisu eest – sina vastutad ülesannete sisu eest, mille esitad.“ (Puusepp 2025: lisa 5) [Autori tõlge]

(4) „(t)ekstiroboti väljundi kasutamise eest vastutab selle kasutaja, kellel peavad selle väljundi hindamiseks olema sobivad teadmised.“ (Puusepp 2025: lisa 3)

(5) „(t)ekstirobot ei ole teksti (kaas)autor, vaid vahend, mida kasutada teksti koostamisel.“ (Puusepp 2025: lisa 2)

Ehkki autoriks võiks pidada teose mistahes loojat, inimest või masinat, on mõiste „autor“ spetsiifiline sotsiaalne konstruktsioon oma ajaloolise kujunemislooga ja kindlate funktsioonidega ühiskonnas (vt nt Foucault 1979). Johannes Fritz (2024: 552) väidab rahvusvahelise autoriõiguste seaduse ja vastava Euroopa Liidu seadusandluse analüüsi põhjal, et autoriks saab pidada vaid isikut, kellele on võimalik omistada teose loomise kavatsust ja kes kasutab oma loomingulist vabadust teose kogu loomisprotsessi vältel. Tekstirobot neid tingimusi ei täida, mistõttu ei saa teda pidada teose autoriks – üldjuhul saab autor olla vaid inimene. (Fritz 2024: 552)

Autori mõiste üks olulisi funktsioone on reguleerida loodud teose õigusi. Autoriõigusi saab omistada inimestele, mitte aga masinatele või arvutiprogrammidele. Näiteks seisab Eesti Autorite Ühingu kodulehel, et „Autor on üldises tähenduses teose looja. Autoriks saab olla vaid inimene ehk autoriks ei ole masinad, loomad ega ka organisatsioonid“ (Eesti Autorite Ühing 2024). Eestis jagunevad autoriõigused seaduse järgi isiklikeks ja varalisteks õigusteks, millest esimesed „on autori isikust lahutamatud ning ei ole üleantavad“ ja viimased „on üleantavad kas üksikute õigustena või õiguste kogumina, kas tasu eest või tasuta“ (Autoriõiguse seadus 2024, § 11). Tekstirobotile ei saa anda autoriõigusi või omistada autorsust teoste eest, mis kuuluvad autoriõiguse seaduse alla (Lee 2023).

Autorsusega kaasneb ka vastutus. Kui tekst tekitab kellelegi kahju, solvab või laimab kedagi või desinformeerib lugejat, peab olema võimalik teksti autor vastutusele võtta, näiteks kaevata ta kohtusse ning nõuda teose levitamise peatamist ja/või kahju hüvitamist. Eestis on olnud mitu autori vastutusega seotud vaidlust, näiteks Kadri Kõusaare filmi „Magnus“ kaasus (Jõgeda 2008) või juhtum, kus Gea Kangilaski kaebas blogijad Inno Tähismaa ja Irja Tähismaa laimu eest kohtusse (Kuul 2018). Tekstirobotit ei ole aga vähemalt praegu kuidagi võimalik vastutusele võtta. Kuigi tekstirobot võib genereerida lugejale semantiliselt tähenduslikku teksti, puuduvad sel kavatsused ja see ei saa anda nõusolekut väljundi avaldamiseks. Kui tekstiroboti kaasabil loodud tekst põhjustab kellelegi kahju, saab vastutusele võtta ainult inimese, kes on selle teksti loomise ja levitamisega seotud. Kokkuvõtvalt võib öelda, et tekstirobotit ei tohiks kunagi, eriti akadeemilises kontekstis käsitleda teksti autori ega kaasautorina.

2.3. Lihtsalt järjekordne tööriist või hoopis vestluspartner?

Tekstiroboteid nimetatakse sageli tehisintellekti tööriistadeks (ingl *AI tools*) või vahenditeks. Tartu Ülikooli suuniseski (2024) nimetatakse tekstirobotit kas vahendiks või abivahendiks (näide 6).

(6) „(n)agu kalkulaatorite, õigekirjakorrektorite ja keeleredaktorite, otsingumootorite ja muude taoliste **vahendite** [autori rõhutus] puhul, pole üldjuhul mõtet keelata ka tekstirobotite kasutamist“ (Puusepp 2025: lisa 3).

Selle näite järgi võib tekstirobotit pidada samasuguseks vahendiks, nagu on keelekorrektor või tekstiredaktor. Tuleb siiski tähele panna, et kui nende jt taoliste vahendite kasutamisele viitamist akadeemilistes tekstides reeglina ei nõuta, siis tekstiroboti genereeritud sisu kasutamist sellele viitamata peetakse akadeemiliseks petturluseks. Näiteks TÜ suunise hilisemas versioonis (2024) tuuakse välja, et tehisintellekti kasutamisele lõputöö kirjutamisel tuleb viidata (näide 7), v.a juhul, kui seda kasutatakse üliõpilase loodud teksti toimetamisel ja vormistamisel töö viimistlemise etapis (näide 8). Ka Haridus- ja Teadusministeeriumi koostatud juhendis eristatakse kahte olukorda, millest ühes pole vaja viidata, aga teises on (näide 9).

(7) „Kirjaliku töö puhul tuleb kirjeldada TI kasutamise viisi ja viidata tekstirobotile korrektselt. Tekstiroboti loodud sisu esitamine enda nime all on akadeemiline petturlus.“ (Puusepp 2025: lisa 3)

(8) „Samuti võivad TI rakendused olla abiks üliõpilase loodud teksti toimetamisel ja vormistamisel töö viimistlemise etapis. **Sellisel juhul ei pea rakendustele viitama** [autori rõhutus].“ (Puusepp 2025: lisa 3)

(9) „1. Kui rakendust kasutatakse tööriistana (teksti redigeerimine või tõlkimine, töölehe loomine, teksti koostamine, ideede kogumine), pole viitamine vajalik. 2. Kui rakendusest saadud väljundit kasutatakse sisulises mõttes (tekstiroboti poolt pakutud tekstilõik, pildigeneraatori poolt loodud pilt), tuleb kasutatud rakendusele viidata, järgides koolis kehtivat viitamise süsteemi.“ (Puusepp 2025: lisa 1)

Näiteid vaadates tekib küsimus, millal on tekstirobot lihtsalt üks tööriist ja millal on tegemist tööriistaga, mis mingite omaduste poolest erineb oluliselt senistest teksti toimetamise abivahenditest.

Siinses artiklis ei arutle ma täpsemalt selle üle, kas tekstirobotil võib olla mingit sorti teadvus, mistõttu oleks „teda" ebaeetiline kohelda pelga tööriistana. Küll aga on vaimufilosoofe tehisintellekti võimalik teadvus huvitanud alates arvutite leiutamisest. Pärast ChatGPT rakenduse väljatulekut on filosoofid hakanud arutlema ka selle üle, kas suurtel keelemudelitel saab olla mingisugune teadvus. Kuigi valdava arusaama järgi ei ole tekstirobotid (veel) teadvusel (vt nt Chalmers 2023), leidub ka autoreid, kelle hinnangul ei saa seda kindlalt väita (nt Lloyd 2024).

Ehkki tekstirobotit saab ka kasutada tekstiredaktori või tõlkeprogrammina, on sellel võimalusi, mis tekstirobotit teistest tekstitöötlusvahenditest eristab. Peamise erinevusena suudab tekstirobot lisaks kasutaja loodud tekstile genereerida *uut* teksti. Võrreldes aga otsingumootoritega ei vahenda tekstirobot lihtsalt avalikult kättesaadavat infot, vaid genereerib *uut* infot, sh väärinfot.

Tekstiroboteid nimetatakse eesti keeles ka suhtlusrobotiteks ja vestlusrobotiteks (ingl *chatbots*), kuna nendega saab suhelda ja vestelda. Nii võib tekstiroboti üheks rolliks pidada teksti kirjutajale vestluspartneriks olemist (näide 10).

(10) „(k)ohati soovitatakse tekstiroboti kasutamisele viidata kui suhtlusele sellega ..."[4] (Puusepp 2025: lisa 2).

Näites 10 nimetatud soovituse alus tundub olevat praktika viidata kirjalikes töödes eravestlustele ja erakirjavahetustele. Andmeanalüütik Indrek Seppo (2024) on samuti arvanud, et ChatGPT võib tavakasutajale olla ideaalne vestluspartner ja kaaslane:

> „Talle tuleb ette kirjeldada, mis sinu ootused tema suhtes on ja siis suhtled juturobotiga nagu teise inimesega. Kellegagi, kes on tulnud sulle külla ja sinust mitte midagi ei tea. Algul tuleb talle üsna palju ette öelda, konteksti lisada, mis sind huvitab, millest sa tahad temaga rääkida."

4 TÜ juhendi hilisemas versioonis (2024) sellist soovitust enam pole.

Tekstirobotid suudavad üsna hästi täita vestluspartneri rolli ja haarata kasutaja interaktiivsesse dialoogi, näiteks keeleõppe, meelelahutuse või muul eesmärgil. Mõni tekstirobot on spetsiaalselt disainitud vestlema nii, et rahuldada kasutaja sotsiaalseid vajadusi või toetada kasutaja vaimset tervist, näiteks imiteerides vestlust psühhoterapeudiga (Kim et al. 2024). Suhtlusrobotite (ingl *social chatbots*) jt tehisintellektil põhinevate robot-terapeutide kasutuselevõtt võib osaliselt aidata asendada mõningaid vaimset tervist toetavaid teenuseid, kuid sellega kaasnevad mitmesugused eetilised probleemid ja küsitavused (Khawaja & Bélisle-Pipon 2023).

Nii nagu mõiste *autor* puhul, saab ka mõistes *vestluspartner* eristada selle üldisemat ja spetsiifilisemat tähendust. Üldises tähenduses on vestluspartner ehk vestluskaaslane „see, kellega vesteldakse“ (vt sõna „vestluskaaslane“ Sõnaveebist 2025), kusjuures asesõna „see“ võiks osutada nii inimesele kui ka tehisintellektile/tekstirobotile. Spetsiifilisemas tähenduses viitab „vestluspartner“ aga „isikule, kellega kõneleja suhtlussituatsioonis vestleb“ (Sõnaveeb 2025). Mõiste *isik* tähendusvälja on analüüsinud nii keeleteadlased kui ka filosoofid ja üldjuhul on leitud, et kuigi mõiste *isik* ei kattu inimolendi mõistega, peetakse isiku all enamasti silmas olendit, kellele saab omistada teatavaid vaimseid omadusi, nagu mõtlemisvõime, moraalsus ja eneseteadvus (vt nt Frankfurt 1971; Taylor 1985). Tekstirobotit praegu (veel) isikuks (pigem) ei peeta. Küll aga kaasnevad tekstiroboti isikuna tajumisega mitmesugused ohud, näiteks see, et tekstirobotiga suhtlemine hakkab asendama tõelisi inimsuhteid ja tekitama sõltuvust. Näiteks uurisid Tianling Xie ja Irina Pentina (2022: 2046) kiindumusteooria perspektiivist vestlusrobot Replika kasutamise psühholoogilist mõju kasutajatele ja tõid välja, et

> „[---] stressirikastes tingimustes ja inimkaaslaste puudumise korral võivad inimesed suhtlusrobotitesse kiinduda. [---] suhtlusroboteid saab kasutada vaimse tervisega seotud ja terapeutilistel eesmärkidel, kuid need võivad põhjustada sõltuvust ja kahjustada lähisuhteid päriselus [---]“.

Eelneva põhjal saab öelda, et tekstirobotit võib küll käsitleda tööriistana, kuid seejuures pole tegu pelga tööriistaga, nagu on tekstiredaktor või otsingumootor. Erinevalt programmidest, mis üksnes töötlevad olemasolevat teksti, suudab tekstirobot luua täiesti uut teksti. Teisest küljest võib tekstirobotit käsitleda vestluspartnerina, kuid

tuleb meeles pidada, et robotiga suhtlemise liiga tõsiselt võtmisel ja selle isikustamisel võivad olla kasutajale kahjulikud psühholoogilised tagajärjed. Seetõttu oleks parem tekstirobotit pidada vestluspartneriks vaid ülekantud tähenduses. Tekstiroboti tajumist tööriista või vestluspartnerina võib käsitleda ühe kontiinuumi kahe erineva otsana – mida enam suhtub kasutaja tekstirobotisse kui vestluspartnerisse, seda vähem tajub ta seda pelga masina või tööriistana. Kuigi tekstirobotid võivad aidata vähendada kirjutamisega seotud üksildust, ei ammenda kumbki neist variantidest tekstiroboti rolli akadeemilise teksti loomes.

3. Kas tekstirobotid laiendavad meie vaimu?

Eelmise jaotisega sai ilmseks, et tekstiroboti rolli mõistmine akadeemilise teksti kirjutamisel on paras pähkel – seda ei saa paigutada kategooriatesse, millega oleme akadeemilise kirjutamise kontekstis harjunud. Akadeemilise teksti kirjutamisel ei saa tekstiroboti väljundit käsitleda allikana ning tekstirobotit ei saa (ega tohi) nimetada teksti (kaas)autoriks. Tekstirobotit võib küll nimetada tööriistaks või vestluspartneriks, kuid need variandid ei hõlma täielikult tekstiroboti rolli tekstiloomes. Tekstiroboti rolli mõtestamise muudab keeruliseks asjaolu, et tekstiloomes võib muutuda hägusaks piir selle vahel, mida teeb tekstirobot ja mida teeb inimene.

Rääkides tekstiroboti kasutamisest akadeemilise teksti kirjutamisel, pean silmas selle väga erinevaid kasutusviise – seetõttu ei keskendu ma siin üksikute näidete analüüsimisele. Küll aga võtan appi Andres Karjuselt (2023) laenatud metafoori, mille kohaselt ei ole tehisaru kasutamine sisse-välja-lüliti, vaid sarnaneb pigem keeratava helivaljuse nupuga. Kui kirjutan kogu teksti ise, kuid küsin tekstirobotilt mõnda sünonüümi, on heli vaevu kuulda. Kui aga palun tekstirobotil kogu teksti valmis kirjutada ja esitan selle muutmata kujul, on heli põhja keeratud. Nende kahe äärmuse vahele jääb lai spekter erinevaid kombinatsioone.

Järgnevalt vaatlen, kas ja kuidas saaks tekstirobotite kasutamist akadeemilisel kirjutamisel käsitleda vaimufilosoofias ja kognitiivteadustes tuntud laiendatud vaimu (ingl *extended mind*) filosoofiast lähtuvalt. Eelkõige huvitavad mind juhtumid, mis paiknevad Karjuse (2023) kirjeldatud spektri keskosas ehk olukorrad, kus kirjutamisprotsessi käigus toimub pidev interaktsioon kasutaja ja tekstiroboti vahel. Samas ma ei

välista, et inimese vaim võib teatud tingimustel laieneda tekstiroboti abiga ka spektri kumbagi otsa jäävate kasutusviiside korral.

3.1. Laiendatud vaim

Laiendatud vaimu idee sõnastasid filosoofid Andy Clark ja David Chalmers artiklis „Laiendatud vaim“[5] aastal 1998. Viimasel veerandsajandil on see lähenemine leidnud nii vaimufilosoofias kui ka kognitiivteadustes edasiarendamist ja poolehoidjaid. Lihtsalt seletatuna viitab laiendatud vaimu tees ideele, et inimese vaimsed (ingl *mental*) – eeskätt kognitiivsed ehk tunnetuslikud (ingl *cognitive*) – protsessid ja seisundid võivad laieneda väljapoole tema aju ja keha. Teisisõnu: mõtlemine ja muud kognitiivsed protsessid ei piirdu ainult sellega, mis toimub inimese peas, vaid on midagi sellist, mida inimene teeb oma aju, keha ja vahel ka keskkonna abil.

Vaimufilosoofias on tavaks eeldada, et vaimuseisunditel, nagu uskumused ja soovid, ning vaimsetel protsessidel, nagu uskumine ja soovimine, on *sisu* ja sellel sisul on alati mingisugune kandja – nt aju teatud protsesside puhul. Laiendatud vaimu teesi järgi kannavad inimese vaimuseisundeid või vaimseid protsesse osaliselt tegurid, mis asuvad väljaspool tema bioloogilist organismi (Rowlands, Lau & Deutsch 2020: 19).

Aitamaks otsustada, millal inimese vaim laieneb väljapoole tema bioloogilise keha piire, sõnastasid Clark ja Chalmers (2009: 2078) nn võrdväärsuse põhimõtte (ingl *parity principle*):

> „Kui olukorras, kus me oleme vastamisi mõne ülesandega, funktsioneerib mõni osa maailmast protsessina, mida, kui seda tehtaks peas, me peaksime kõhklematult tunnetusprotsessi osaks, siis see osa maailmast ongi osa tunnetusprotsessist.“

Võrdväärsuse põhimõtte iva on selles, et vaimseid protsesse tuleb määratleda nende funktsiooni, mitte asukoha järgi. Et laiendatud vaimu ideed paremini mõista, vaatame järgnevalt kahte mõttelist eksperimenti Clarki ja Chalmersi (2009) artiklist.

---

5 Artikkel on eesti keeles ilmunud Silver Rattasepa tõlkes ajakirjas Akadeemia (vt Clark & Chalmers 2009).

Esimesel juhul kutsuvad autorid kujutlema end Tetrist mängimas: olukorda, kus istutakse arvutiekraani ees, mis kuvab erinevaid kahemõõtmelisi geomeetrilisi kujundeid, ning mängijal palutakse vastata küsimustele kujundite võimaliku sobivuse kohta ekraanil kujutatud „pesadesse“. Sobivuse hindamiseks on kolm võimalust.

(a) Pöörata kujundeid vaimusilmas, joondades neid pesadega.
(b) Pöörata kujundit füüsiliselt, vajutades pööramisnupule, mis võimaldab kujundit ekraanil pöörata ligi kolm korda kiiremini kui vaimusilmas.
(c) Pöörata kujutist närviimplantaadi abil, mis asub ajus ja võimaldab vaimusilmas sooritada pööramisoperatsiooni sama kiiresti kui füüsilise nupulevajutusega variandis b.

Sealjuures on oluline mainida, et kuigi tänapäeval juba katsetatakse inimajju sisestatavate implantaatidega (Levy 2024), pole punktis c kirjeldatud närviimplantaati veel leiutatud. Nüüd võiks mõelda, millist varianti kolmest oldaks valmis nimetama vaimseks protsessiks ja millal tahaks vaimsete protsesside kõrval rääkida vaimuvälistest protsessidest.

Autorid eeldavad, et lugeja on valmis pidama kujundite vaimusilmas pööramist vaimseks protsessiks nii variantide a kui ka c puhul. Sealt edasi püüavad nad lugejat veenda, et kõik kolm varianti kirjeldavad protsesse, millel on tunnetuses sama funktsionaalne roll. Selle põhjal väidavad Clark ja Chalmers (2009), et kõiki kolme varianti peaks käsitlema kognitiivsete protsessidena, ainult et variandi b puhul laieneb vaim kehast keskkonda.

Teine mõtteline eksperiment, mille varal Clark ja Chalmers laiendatud vaimu ideed selgitavad, ei nõua sedavõrd küberpungiliku maailma kujutlemist. Lugejal tuleb mõelda kahest tavalisest inimesest: Ingast ja Ottost. Tähtsaim erinevus nende vahel seisneb selles, et Inga on täiesti terve, aga Otto põeb kerget Alzheimeri tõbe, mistõttu on tal mäluprobleemid. Mäluprobleemide kompenseerimiseks kannab Otto kõikjal kaasas märkmikku, kuhu ta kirjutab üles olulise uue info, millega ta päeva jooksul kokku puutub, et seda vajalikul hetkel järele vaadata. Kui Inga soovib minna kunstimuuseumi, siis ta mõtleb hetke, talle meenub, et muuseum asub 53. tänaval, ja ta läheb muuseumisse. Otto aga vaatab märkmikust järele, et muuseum asub 53. tänaval, ning seab samuti sammud samasse muuseumisse. Loo mõte on selles, et märkmikuga konsulteerimine mängib Ottole samasugust funktsionaalset rolli nagu

bioloogilise mäluga konsulteerimine Ingale. Clark ja Chalmers väidavad, et Otto mäluprotsessid ja nende tulemusel teadvusse kerkinud uskumused laienevad märkmiku kasutamise võrra ajust välja, keskkonda.

3.2. Tekstirobot inimvaimu laiendajana

Laiendatud vaimu ideed on eri autorid rakendanud eri valdkondades. Filosoof Richard Menary (2007: 261) väidab, et ka kirjutamine on laiendatud vaimne protsess ehk laiendatud mõtlemine[6]: „Kirjalike kandjate loomine ja manipuleerimine on osa meie kognitiivsest töötlusest ja seega muundab kirjutamine meie kognitiivseid võimeid.“ Teisisõnu: ma võin mõlgutada mõtteid oma peas, aga ma võin mõtelda ka paberi ja pliiatsi või ekraani ja klaviatuuri abil, kusjuures viimased võivad mul aidata mõelda täiesti uutel viisidel. Menary väitel võimaldab kirjapanek osa inimese sisemistest kognitiivsetest ressurssidest (nt osa protsesse, mis hõlmavad lühi- ja pikaajalist mälu) vabastada, mis omakorda võimaldab esile kerkida uutel kognitiivsetel ressurssidel.[7]

Ka psühholoogid Keith Oatley ja Maja Djikic (2008: 9) väidavad, et „pastakas on masin, mille abil mõelda“. Idee pole iseenesest uus – Kenneth Craik (1943) leidis juba 1940ndatel, et mõtlemine hõlmab mõne maailma aspekti tõlkimist vaimseks mudeliks, millega manipuleerimine loob uue vaimse seisundi, kusjuures mudeliga manipuleerimine *ongi* mõtlemine, olenemata sellest, kas seda tehakse peas või paberil.

Mõeldes tekstiroboti abil teksti loomisele, kerkib üles küsimus: kui kirjutamine on laiendatud mõtlemine, siis kas tekstirobotit tuleks käsitleda inimvaimu laiendusena? Kuigi mitu autorit on hiljuti rakendanud laiendatud vaimu ideed tehisintellektile üldiselt (nt Hernández-Orallo 2025; Pellegrino & Garasic 2020) ja mitmesugustele digitaalse

6 Vihje sellele ideele leiab juba Clarki ja Chalmersi artiklist, kus nad kirjutavad: „Mõelgem (...) filosoofile, kes mõtleb kõige paremini kirjutades ja arendab oma mõtteid sedamööda, kuidas ta edeneb. Vahest on keel arenenud osalt selleks, et meil oleks võimalik laiendada kognitiivseid ressursse niiviisi aktiivselt seotud süsteemide piires.“ (Clark & Chalmers 2009: 2083).

7 Kirjaoskus tundub aitavat kaasa abstraktse mõtlemise kujunemisele. Nt Luria (1976) uuris 1930-ndatel Usbekistanis kirjaoskajate ja kirjaoskamatute mõtlemist ja leidis, et valdav osa kirjaoskamatutest ei suutnud lahendada abstraktse mõtlemise ülesandeid, millega kirjaoskajad kerge vaevaga hakkama said.

kirjutamise abivahenditele, nagu otsingumootorid ja tekstiredaktorid (Overstreet 2022, Overstreet, Akhmedjanova & Vaccino-Salvadore 2023), ei käsitle nad tekstirobotite rolli kirjutamisel.

Üks laiendatud vaimu teesi levinud vastuväiteid on nn lahtiseotavuse argument: erinevalt ajust on välised vahendid ja protsessid meist liiga kergesti lahtiseotavad, seega pole tegu vaimsete protsessidega (Clark & Chalmers 2009: 2081). Aju, mille abil mõelda, on kogu aeg inimesega – sama ei saa aga öelda tekstirobotite kohta. Teiseks on laiendatud vaimu teooriale ette heidetud vaimu laialivalgumise ohtu ehk seda, et laiendatud vaimu tees muudab vaimu mõiste absurdseks, kuna selle järgi ei saagi enam vaimu selgelt piiritleda. Näiteks võib tekkida olukord, kus inimese vaim on internetti laiali valgunud.

> „Kui laused Otto märkmikus võivad minna arvesse tema mõtetena, siis miks mitte laused, mis leiduvad entsüklopeedias, või laused, mis leiduvad internetis – ka neile mõlemale on Ottol suhteliselt lihtne ligi pääseda?“ (Rowlands, Lau & Deutsch 2020: 21)

Sellele vastuväitele on avatud Overstreeti (2022) käsitlus, kus vaadeldakse digitaalsete abivahendite kasutamist kirjutamisel inimese vaimu laiendajatena. Nii kerkivad esile järgmised küsimused: kui tekstirobot laiendab inimese vaimu, siis kas see vaim valgub tekstirobotisse laiali? Kui ma kasutan mingit kindlat tekstirobotit samal ajal paljude teiste kasutajatega, siis kas minu ja teiste kasutajate vaim sulab ühte?

Nn lahtiseotavuse ja laialivalgumise vastuargumentide tõrjumiseks argumenteerivad Clark ja Chalmers (2009: 2082), et millegi vaimseks liigitamisel pole oluline selle kaasaskantavus, vaid miski, mida nad nimetavad *usaldusväärseks sidususeks* (ingl *reliable coupling*):[8]

> „Juhtumisi toimub enamik usaldusväärsest sidumisest ajus, aga usaldatav sidusus võib kergesti tekkida ka keskkonnaga. Kui taskuarvuti või märkmik on mul **alati käepärast, kui teda tarvis läheb** [autori rõhutus], on ta minuga seotud just nii usaldusväärselt kui tarvis. Sisuliselt on nad siis osa kognitiivsete ressursside põhipaketist, mida ma igapäevaselt maailma kallal rakendada saan.“ (Clark & Chalmers 2009: 2082)

8 Lähtun Silver Rattasepa tõlkest (Clark ja Chalmers 2009), kus ingliskeelse mõiste *reliable coupling* eestikeelne vaste on *usaldusväärne sidusus*.

Autorid sõnastavad välise vahendi või protsessi usaldusväärse sidususe hindamiseks neli kriteeriumi (Clark & Chalmers 2009, vt ka Clark 2010: 46).

1. Vahendi pidev käepärasus. Vahend peab olema kättesaadav parasjagu siis, kui seda on tarvis. Nt kui Ottol on märkmik alati käepärast siis, kui ta soovib midagi meenutada, ning ta ei vasta küsimusele enne „ma ei tea", kui on oma märkmikuga konsulteerinud, siis on märkmik temaga samavõrd usaldusväärselt seotud, nagu on Inga bioloogilise mälu aluseks olevad ajuprotsessid seotud Ingaga.[9] Siinjuures ei pea vahend olema 100% alati käepärast – ka bioloogilises ajus tekkivad mälestused ei ole inimesele alati kättesaadavad, kuna aju võib aeg-ajalt nt une, alkoholijoobe või kõrgendatud emotsioonide tõttu kaotada osa oma töövõimest.
2. Info lihtne kättesaadavus vahendi kaudu. Nii nagu mõtted ja mälestused tulevad pähe ilma, et peaks selleks väga palju vaeva nägema, peaks ka välise vahendiga opereerimine infot hõlpsasti vahendama. Inga pidi muuseumi aadressi teadasaamiseks hetkeks seda meenutama, nii nagu Otto leidis oskuslikult märkmikku käsitsedes muuseumi aadressi raskusteta üles.
3. Info enam-vähem automaatne usaldamine ja omaksvõtt. Inimene peab vahendi kaudu saadavat infot usaldama ja selle enam-vähem automaatselt omaks võtma sarnaselt sellega, kuidas ta usaldab ja võtab omaks bioloogilise aju genereeritud mälestusi ja uskumusi.
4. Info varasem teadlik heakskiit. Inimene peab olema vahendi kaudu saadava info heaks kiitnud. Nt Ottol on põhjust usaldada oma märkmiku sissekandeid, sest ta on märkmikku kantud info mingil varasemal ajahetkel heaks kiitnud.

Clark ja Chalmers (2009: 2090) möönavad, et neljanda kriteeriumi üle võib vaielda. Seda põhjendatult, kuna ka meie ei pruugi olla enda uskumuste aluseks olevat infot eelnevalt teadlikult heaks kiitnud. Kolmel esimesel kriteeriumil on nende arvates aga otsustav roll. Kas ja kuivõrd on aga tekstiroboti kasutamisel usaldusväärse sidususe kriteeriumid täidetud?

9 Clarki ja Chalmersi (2009) artikkel on kirjutatud ajal, mil nutitelefonid ei olnud veel laialt levinud. Praegu loodab suur osa inimesi tõenäoliselt näiteks aadresside või sünnipäevade meenutamisel pigem nutiseadme kui oma bioloogilise mälu peale.

Esimene ja teine kriteerium on täidetud juhul, kui inimesel on pidev ligipääs arvutile, internetile ja ChatGPT-le vm tekstirobotile ning kui ta suudab vastava tehnoloogiaga sujuvalt ümber käia. See-eest neljas kriteerium ei ole kindlasti täidetud, sest tekstirobot genereerib iga kord kasutajale uue teksti, mistõttu ei saa kasutaja olla tekstiroboti genereeritud infot varem heaks kiitnud. Kuid nagu Clark ja Chalmers ise möönavad, ei ole neljas kriteerium usaldusväärse sidususe hindamisel otseselt tarvilik, nii et jätan selle siinkohal kõrvale.

Otsustavaks saab kolmas kriteerium ehk tingimus, et vahendi kaudu saadav info oleks inimesele sama usaldusväärne, nagu on oma mõtete vahendatud info, ja ta võtaks selle enam-vähem automaatselt omaks, samamoodi nagu oma uskumused ja mälestused. Tekstirobotite kasutamise ohuna rõhutatakse sageli, et kasutaja ei peaks kunagi usaldama tekstiroboti väljundit, vaid suhtuma sellesse kriitiliselt. Tekstirobotid võivad luua kallutatud sisu ja n-ö hallutsineerida ehk genereerida enesekindlalt väärfakte. Sellest võiks järeldada, et tekstirobotite kasutamisel ei ole kolmas kriteerium kunagi täidetud. Selline järeldus oleks aga ennatlik.

Siinkohal vajab äramärkimist, et ma määratlen mõisteid „usaldus“ ja „usaldusväärne“ laiemalt, kui kasutatakse ingliskeelses eetikakirjanduses mõisteid *trust* ja *trustworthy*.[10] Nimelt väidab osa moraalifilosoofe, et masinaid pole võimalik usaldada (Jones 1996: 14); neile saab üksnes kindel olla või toetuda, kuna usaldust on võimalik reeta, kuid millegi töökindluses saab ainult pettuda (Hatherley 2020; Hawley 2014; Holton 1994). Siiski leidub ka moraalifilosoofe, kes kaitsevad seisukohta, et usaldus on hoiak, mis saab olla ka masinatele suunatud (vt nt Nickel 2013).

Kui uurida, kuidas inimesed tekstirobotitega suheldes tekstiroboteid tajuvad, selgub, et tavakasutajad ei pruugi neid tajuda lihtsalt masinatena, vaid kipuvad neid eri viisil antropomorfiseerima. Näiteks leidsid Oliver Jacobs, Farid Pazhoohi & Alan Kingstone (2024), et pärast tekstirobotite kasutamist väheneb kasutajatel tunne, et teatud vaimsed omadused (agentsus ja võime kogeda) on inimestele ainuomased. Inju Lee & Sowon Hahn (2024) aga avastasid, et tekstirobotiga suhtlemisel kogevad inimesed emotsionaalset tuge suurema tõenäosusega siis, kui

10 Tänan retsensenti, kes juhtis mu tähelepanu usalduseteemalisele diskussioonile eetikaalases kirjanduses.

nad implitsiitselt omistavad tekstirobotile vaimseid omadusi ehk tajuvad „teda“ vaimu omavana (ingl *mind perception*). Jätan siinkohal lahtiseks küsimuse, kas inimesed saavad tekstiroboti käitumises üksnes pettuda või tunda end ka tekstiroboti poolt reedetuna – see küsimus vajaks empiirilist uurimist. Igatahes tundub üksnes tekstirobotite töökindlusest rääkimine eesti keeles liiga kitsendav, eriti arvestades, et eesti keeles on tavaks rääkida ka näiteks ettevõtete ja teatmeteoste usaldusväärsusest (vt sõna „usaldusväärsus“ 2009. a EKSS-ist). Lisaks on viimastel aastatel ilmunud teadusuuringuid erinevate tegurite mõju kohta sellele, kuivõrd inimesed tekstiroboteid usaldavad ja milles see usaldus seisneb (De Cicco, Silva & Alparone 2020; McLean et al. 2020; Xie, Pentina & Hancock 2023).

Siinses artiklis pole rõhk niivõrd küsimusel, kas inimesed usaldavad ja saavad usaldada tekstirobotit kui sellist, kuivõrd küsimusel, mil määral inimesed usaldavad ja on õigustatud usaldama igale konkreetsele viibale vastuseks genereeritud teksti. Näiteks võib kasutaja usaldada rohkem ChatGPT-4 genereeritud väljundit inglise keeles kui eesti keeles. See tundub õigustatud, kuna tekstiroboti treenimiseks on kasutatud märkimisväärselt rohkem ingliskeelseid tekste kui eestikeelseid ning seetõttu on see inglise keeles ka võimekam.

Oluline on silmas pidada, et tekstiroboti genereeritud väljundi kvaliteet sõltub viiba loomise oskusest, mistõttu eeldab tekstiroboti väljundi usaldamine enamasti adekvaatset viipade koostamise oskust (ingl *prompt engineering skills*). Viimast võib määratleda kui „oskust formuleerida täpseid ja hästi struktureeritud juhiseid, et saada suurtelt keelemudelitelt soovitud vastuseid või teavet, optimeerides interaktsiooni efektiivsust“ (Knoth et al. 2024, autori tõlge). Kuigi veebilehtedel, sh ülikoolide kodulehtedel, jagatakse praktilisi soovitusi viipade koostamiseks akadeemilises kontekstis (vt nt Viiba koostamine; Prompts 2023), on viipade koostamise oskusi praeguseks veel vähe uuritud. Nils Knoth et al. (2024) seostavad viipade koostamise oskusi TI-kirjaoskusega (ingl *AI literacy*). TI-kirjaoskuse uurimine on samuti alles tärkav uurimisvaldkond ning selle mõiste all peetakse silmas erinevaid asju (vt Ng et al. 2021), kuid TI-kirjaoskuse ühe olulise komponendina käsitletakse enamasti teadmisi selle kohta, kuidas tehisintellekt töötab.

Milline alus on väitel, et inimesed võtavad omaenese uskumuste ja mälestuste kaudu esitatud info automaatselt omaks, pidades seda usaldusväärseks, kuid tekstirobotite genereeritud infot inimesed ei

usalda ega tohikski kunagi usaldada? Tuleb tähele panna, et kuigi inimesed kipuvad küll uskuma ja usaldama oma uskumusi ja mälestusi, ei ole see alati ratsionaalselt põhjendatud. Psühholoogia uuringutest on teada, et inimpsüühikale on omased mitmesugused kognitiivsed kallutatused (Korteling & Toet 2022), mis peegelduvad ka inimeste kirjutatud tekstides. Kuna suurtel keelemudelitel põhinevad tekstirobotid on treenitud inimautorite kirjutatud tekstide põhjal, peegelduvad inimestele omased mõtlemistendentsid ka tekstirobotite genereeritud tekstides (Acerbi & Stubbersfield 2023). Seega võiks väita, et inimese pähe n-ö kerkivate mõtete sisu pole põhjust eriti rohkem usaldada kui tekstiroboti genereeritud teksti sisu. Samuti pole võimalik väita, et ükski kasutaja ei usalda kunagi ühtegi tekstiroboti väljundit. Siinjuures eristaksin kahte sorti usaldust: n-ö pimedat ehk algaja usaldust ja oskuslikku ehk eksperdi usaldust.

Pimeda usaldusega on tegu näiteks siis, kui õppija, kes kirjutab esimest korda elus argumenteerivat esseed, palub tekstirobotil selle valmis kirjutada ning esitab õppejõule muutmata kujul tekstiroboti genereeritud teksti, kuna usub, et ta ise ei oskaks niikuinii paremat teksti kirjutada. Ta ei oska tekstiroboti väljundit piisavalt kriitiliselt hinnata ning ta pole ka oskuslik viipade koostaja. Kui kasutaja usub, et tekstirobot suudab esseed temast paremini kirjutada, olemata seejuures võimeline tekstiroboti loodud teksti süvitsi mõistma ja kriitiliselt hindama, on tegu pimeda ehk ratsionaalselt põhjendamatu usaldusega.

Oskusliku usaldusega on tegu aga näiteks siis, kui õppija, kel on piisavalt oskusi argumenteeriva essee kirjutamiseks ning piisavalt teadmisi teemast, mille kohta ta esseed kirjutab, koostab parajalt detailse viiba ja palub tekstirobotil oma töövaeva kergendamiseks genereerida mingi osa tekstist. Ta loeb selle läbi ja, hinnanud väljundi kvaliteedi sobilikuks, lõimib selle oma muudatustega pooleliolevasse teksti. Sarnaselt võib toimida näiteks professionaalne programmeerija, kes annab tekstirobotile oskuslikud juhtnöörid mõne koodiosa kirjutamiseks ja pärast tuvastamist, et tekstirobot on genereerinud korrektse koodijupi, integreerib tulemuse oma töösse ning hoiab seejuures aega kokku. Näiteks ütleb Indrek Seppo (2024): „Mina kasutan seda [ChatGPT-d] päris palju. Minu põhitegevus on andmeanalüüs, mis hõlmab koodi kirjutamist. Ma kirjeldan talle ära, missuguseid tulemusi ma soovin saada, mida kood peaks tegema ja ta kirjutab mulle koodi ette.“ Oluline erinevus seisneb aga selles, et koodi puhul saab tulemust enamasti objektiivselt testida –

kood kas töötab või ei tööta –, samas kui teksti sobivuse hindamine põhineb inimese subjektiivsel tõlgendusel.

Kui kasutajal on poolelioleva teksti teemal piisavalt teadmisi ja viipade kirjutamise oskusi ning ta on kogenud, et tekstirobot saab mingi töölõiguga piisavalt hästi (ja enamasti kasutajast mitu korda kiiremini) hakkama, ei pruugi ta tekstiroboti genereeritud väljundisse suhtuda suurema umbusuga kui enda loodud teksti või koodi. Kirjutamisprotsessi käigus kohendab professionaal tavaliselt niikuinii jooksvalt juba kirjapandud teksti: õnnestunud koostöö puhul tekstirobotiga ei pruugi olla vahet, kas kirjutaja võtab aluseks enda loodud või tekstiroboti genereeritud teksti. Samamoodi võib kasutaja põhjendatult usaldada tekstirobotit oma loodud teksti toimetama, näiteks kui ta on veendunud, et tekstiroboti toimetatud teksti kvaliteet on võrdne kasutaja toimetamiskvaliteediga.[11] Kui lõppväljund on kasutaja ja tekstiroboti vahelise interaktsiooni tulemus, mille kasutaja on põhjalikult läbi töötanud ja heaks kiitnud, on tegemist tekstiroboti oskusliku usaldamisega.

Eelnevast mõttekäigust näeb, et inimesed vahel usaldavad tekstiroboti väljundit ja et teatud tingimustel on seda ka ratsionaalne teha. Iseäranis oluline on siinjuures, et kasutaja ei usaldaks tekstiroboti väljundit pimesi, vaid teeks seda võimalikult oskuslikult. Selle hindamiseks, kui oskuslik või pime on kasutaja usaldus, ei ole lihtsat viisi, kuid kokkuvõtvalt saab öelda, et see sõltub kontekstist ning sellest, (1) millised on kasutaja iseseisvad akadeemilise teksti kirjutamise oskused; (2) kui hästi tunneb kasutaja teemat, millest ta kirjutab; (3) milline on kasutaja TI-kirjaoskus ja kui asjakohased on tema koostatud viibad ning (4) kui hästi suudab kasutaja eelnimetatud teadmiste valguses tekstiroboti väljundit kriitiliselt hinnata. Seega võib järeldada, et teatud tingimustel suudab tekstirobot täita kolme usaldusväärse sidususe kriteeriumi ja sellistel puhkudel saab rääkida ka tekstiroboti abil laiendatud inimvaimust.

11 TÜ uurimisrühm on loonud akadeemilise teksti kirjutamise protsessi toetava veebimaterjali, kus kirjeldatakse ka tehisintellektil põhinevate tööriistade katsetamise tulemusi. Vt https://www.teadustekst.ut.ee/tehisintellektil-pohinevate-tooriistade-kaasamine-kirjutamisprotsessi/ (21.04.2025).

3.3. Tekstirobot inimvaimu ahendajana

Kui kasutaja usaldab tekstirobotit pimesi, nt laseb oma akadeemilise teksti tekstirobotil täienisti valmis genereerida, seda kriitiliselt hindamata ja oskamata ise teksti kirjutada, on kohane rääkida pigem tekstiroboti tõttu ahendatud inimvaimust, kuna selline teguviis võtab kasutajalt võimaluse arendada kirjutamis- ja mõtlemisoskust. Õppimist ja õppijate arengut pärssiva mõju pärast on mures paljud haridustöötajad. Mure üks aluseid on mõttekäik, et kui õppija delegeerib olulised mõtlemise aluseks olevad ülesanded tekstirobotile, siis tema võime iseseisvalt mõelda kärbub või jääb välja arenemata. Iseäranis tõsine on see mure laste ja noorte puhul, kel pole iseseisva mõtlemise aluseks olevad kognitiivsed baasoskused veel välja kujunenud.

Praegu ei ole veel empiirilisi andmeid tekstirobotite ülemäärase kasutamisega[12] seotud pikaajaliste mõjude kohta laste, noorte ja täiskasvanute kognitiivsetele võimetele. Lühiajalisi mõjusid on juba mõnevõrra uuritud ning tulemused viitavad sellele, et tekstirobotite kiirustav ja läbi mõtlemata kasutuselevõtt ei pruugi õppimist toetada (Tragel & Hint 2025). Tekstirobotite ülemäärane kasutamine võib soovimata suunas mõjutada kasutaja võimet iseseisvalt mõelda – sarnaselt sellega, kuidas pidev GPSi kasutamine halvendab kasutajate ruumilist mälu (Dahmani & Bohbot 2020). Näiteks Muhammad Abbas et al. (2024) on leidnud, et ChatGPT kasutamine suurendab üliõpilaste kalduvust õppimist edasi lükata ning on seotud mälu halvenemise ja kehvemate akadeemiliste tulemustega. Hamsa Bastani et al. (2024) leidsid, et keskkooliõpilased, kes kasutasid katses matemaatikaülesannete lahendamisel ChatGPT abi, olid ülesannete lahendamises küll edukamad, kuid kui neil hiljem ChatGPT-d kasutada ei lubatud, said nad ülesannete lahendamisel kehvemaid tulemusi võrreldes õpilastega, kes polnud katse varasemas etapis ChatGPT abi kasutanud.

Tekstiroboti kasutamine akadeemiliste tekstide loomisel võib mõjuda kriitilisele mõtlemisele sarnaselt sellega, kuidas otsingumootoritele toetumine mõjutab mäluprotsesse. Nimelt avastasid Betsy Sparrow, Jenny Liu ja Daniel M. Wegner (2011) nn Google'i efekti: kui inimesed

12 Mõiste „tekstirobotite ülemäärane kasutamine" ei ole kirjanduses veel üheselt defineeritud, kuid laias laastus peetakse selle all silmas sellist tekstirobotite kasutamist, mis ühel või teisel moel kahjustab kasutajat.

usuvad, et mingi teave on neile otsingumootorite kaudu kättesaadav, siis nad meenutavad vähem seda teavet ennast ja keskenduvad hoopis meenutamisele, kust vastavat teavet leida. Hiljutises Microsofti uurimisrühma tehtud uuringus võeti luubi alla generatiivse tehisintellekti kasutamise mõju teadlaste ja spetsialistide (ingl *knowledge workers*) kriitilisele mõtlemisele (Lee et al. 2025). Uuringus lähtuti Bloomi (1956) kriitilise mõtlemise definitsioonist ja taksonoomiast, mis jagab kriitilise mõtlemise kuueks komponendiks:

1) teadmine (ideede meenutamine);
2) mõistmine (ideedest arusaamise demonstreerimine);
3) rakendamine (ideede praktikasse viimine);
4) analüüs (ideede võrdlemine ja seostamine);
5) süntees (ideede kombineerimine);
6) hindamine (kriteeriumide alusel ideede üle otsustamine).

Uuringust selgus, et generatiivse tehisintellekti kasutamine muudab kriitilist mõtlemist kolmel viisil: teadmiste ja mõistmise komponendi puhul nihkub inimeste pingutus teabe kogumiselt teabe kontrollimisele; rakendamise puhul tegeletakse ülesande lahendamise asemel pigem tehisintellekti vastuste integreerimisega ning analüüsi, sünteesi ja hindamise puhul nihkub põhirõhk ülesannete lahendamiselt pigem „järelevaataja“ rollile. Ilmnes ka seos kriitilise mõtlemise ning omaenda kirjutamisvõimetesse uskumise ja generatiivse tehisintellekti võimetesse uskumise vahel. Nimelt korreleerus kasutajate suurem usk sellesse, et nad suudavad ise vastavat kirjutamisülesannet lahendada, parema kriitilise mõtlemise ning suurema kasutajapoolse pingutusega, samal ajal kui suurem usk generatiivse tehisintellekti võimekusse korreleerus vähem kriitilise mõtlemisega ning väiksema kasutajapoolse pingutusega. (Lee et al. 2025)

4. Arutelu

Mida siis teha akadeemilises hariduses selleks, et tekstirobotid laiendaksid õppijate vaimu, mitte ei ahendaks seda? See on küsimus, millele pole lihtsat vastust ja millele vastamiseks pole veel ka piisavalt empiirilisi uuringuid. Õpetajad ja õppejõud saavad aga sellegipoolest juba praegu üht-teist ära teha. Näiteks oleks kasulik õppijatele selgitada,

milliseid kognitiivseid baasoskusi akadeemilise teksti iseseisev kirjutamine arendab, miks on need oskused olulised ja kuidas ülemäärane tekstirobotile toetumine võib nende oskuste arengut takistada. Valmis teksti ehk lõpptulemuse kõrval tuleks rohkem väärtustada ka kirjutamisprotsessi – hinnata mitte ainult lõpptulemust, vaid ka kirjutamise erinevaid etappe ning seda, kui palju vaeva õppija mustanditega näeb. Koolid ja ülikoolid peaksid arendama nii õppijate iseseisvaid kognitiivseid oskusi kui ka TI-kirjaoskust ning looma keskkonna, kus ühiselt arutletakse, milliseid ülesandeid on akadeemiliste tekstide kirjutamisel mõistlik tekstirobotile anda ja milliseid mitte.

Üks lihtne retsept otsustamaks, kas tekstiroboti kasutamine mingi konkreetse kirjutamisülesande sooritamisel õppuri kirjutamis- ja mõtlemisoskuste arengut toetab, oleks paluda õppuril iga tekstiroboti väljundi puhul endalt küsida: „Kas ma oleksin võinud põhimõtteliselt ise selle teksti kirjutada, olgugi et see oleks mul kauem aega võtnud?“ Kuni vastus sellele küsimusele on „ei“ või „pigem mitte“, võiks õppurile soovitada pigem ise oma tekstiga töötamist. Väljaspool õppimist ei pruugi eeltoodud kontrollküsimus nii oluline olla – näiteks olukorras, kus kellelgi on vaja kirjutada mõni tüütu bürokraatlik tekst, kuid ta ei soovi bürokraatlike tekstide loomise oskust arendada. Võib juhtuda, et too inimene vastab küsimusele „Kas ma oleksin ise suutnud sellise teksti kirjutada?“ eitavalt, kuid kasutab edukalt tekstiroboti väljundit oma vaeva kergendamiseks ja aja kokkuhoiuks. Sellisel juhul võib tekstirobotist küll kirjutamisel abi olla, kuid kasutaja vaimu see ei laienda ning tema kirjutamisoskusi ei arenda.

Haridustöötajatel lasub igal juhul keeruline ülesanne disainida õppijatele selliseid kirjutamisülesandeid, mida ei oleks lihtne sooritada tekstirobotite pimeda usalduse abil. Tarvis on leida viise, kuidas toetada nii õppijate iseseisvaid kirjutamis- ja mõtlemisoskusi kui ka oskust teatud kirjutamisülesandeid tekstirobotile targalt delegeerida, et laiendada tekstiroboti abil oma vaimu. Alles siis kui õppija on omandanud teatavad iseseisva mõtlemise ja teksti kirjutamise baasoskused, saab tekstirobot hakata toimima tema vaimu laiendusena.

5. Kokkuvõte

Kui inimene kasutab (akadeemilise) teksti kirjutamisel tekstiroboti abi, siis kuidas tuleks mõista tekstiroboti rolli selles protsessis? Tekstirobot ei sobitu kategooriate alla, millega oleme harjunud akadeemilise kirjutamise kontekstis – tegu pole allika, (kaas)autori, lihtsalt tööriista ega vestluspartneriga. Vaimufilosoofias ja kognitiivteadustes populaarsust kogunud laiendatud vaimu käsitluse järgi võivad inimese vaimsed protsessid ja seisundid laieneda väljapoole tema aju ja keha piire ning hõlmata väliskeskkonna vahendeid ja protsesse. Kirjutamisprotsessi iseenesest saab pidada laiendatud mõtlemisprotsessiks, ent see, kas kirjutamisel tekstiroboti abi kasutamisega saab inimese vaimseid protsesse laiendada, sõltub kasutaja eelnevatest mõtlemis- ja kirjutamisoskustest ning suutlikkusest tekstiroboti väljundit mõista ja kriitiliselt hinnata. Pime usaldus tekstiroboti genereeritud teksti suhtes pigem ahendab kui laiendab inimvaimu, pärssides kasutaja mõtlemis- ja kirjutamisoskuse arengut.

Kirjandus

Abbas, Muhammad, Farooq Ahmed Jam & Tariq Iqbal Khan. 2024. Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education* 21(10). https://doi.org/10.1186/s41239-024-00444-7.

Acerbi, Alberto & Joseph M. Stubbersfield. 2023. Large language models show human-like content biases in transmission chain experiments. *Proceedings of the National Academy of Sciences* 120(44). https://doi.org/10.1073/pnas.2313790120.

Autoriõiguse seadus (21.06.2024). *Riigi Teataja*. https://www.riigiteataja.ee/akt/AutÕS (22.08.2024).

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakci & Rei Mariman. 2024. Generative AI can harm learning. *The Wharton School Research Paper*, 15.07. http://doi.org/10.2139/ssrn.4895486.

Bloom, Benjamin S. (toim). 1956. *Taxonomy of educational objectives: The classification of educational goals: Handbook I: Cognitive domain. Technical Report*. New York: David Mckay Company.

Chalmers, David. 2023. Could a large language model be conscious? *Boston Review*, 09.08. www.bostonreview.net/articles/could-a-large-language-model-be-conscious/ (22.04.2025).

Clark, Andy. 2010. Memento's revenge: The extended mind, extended. Richard Menary (toim), *The extended mind*, 43–66. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0003.

Clark, Andy & David Chalmers. 1998. The extended mind. *Analysis* 58(1). 7–19. https://doi.org/10.1093/analys/58.1.7.

Clark, Andy & David Chalmers. 2009. Laiendatud vaim. Silver Ratassepp (tlk). *Akadeemia* 11. 2076–2094.

Craik, Kenneth. 1943. *The nature of explanation.* Cambridge University Press.

Dahmani, Louisa & Véronique D. Bohbot. 2020. Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports* 10(6310). https://doi.org/10.1038/s41598-020-62877-0.

De Cicco, Roberta, Susana C. Silva & Francesca Romana Alparone. 2020. Millennials' attitude toward chatbots: An experimental study in a social relationship perspective. *International Journal of Retail and Distribution Management* 48(11). 1213–1233. https://doi.org/10.1108/IJRDM-12-2019-0406.

Digital Education Council Global AI Student Survey. 2024. https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024 (17.04.2025).

Eesti Autorite Ühing. *Mis on autoriõigus?* https://eau.org/mis-on-autorioigus/ (22.08.2024).

Foucault, Michel. 1979. What is an Author? (Qu'est-ce qu'un auteur?). Josue V. Harari (toim), *Textual Strategies: Perspectives in Post-Structuralist Criticism*. 141–160. Ithaca: Cornell University Press. https://doi.org/10.7591/9781501743429.

Frankfurt, Harry G. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68(1). 5–20. https://doi.org/10.2307/2024717.

Fritz, Johannes. 2024. The notion of 'authorship' under EU law – Who can be an author and what makes one an author? An analysis of the legislative framework and case law. *Journal of Intellectual Property Law & Practice* 19(7). 552–556. https://doi.org/10.1093/jiplp/jpae022.

Greene, Jim. 2024. ChatGPT (software). *Salem Press Encylopedia of Science.* https://www.ebsco.com/research-starters/science/chatgpt-software (21.04.2025).

Hatherley, Joshua James. 2020. Limits of trust in medical AI. *Journal of Medical Ethics* 46(7). 478–481. https://doi.org/10.1136/medethics-2019-105935.

Hawley, Katherine. 2014. Trust, distrust and commitment. *Noûs* 48(1). 1–20. https://doi.org/10.1111/nous.12000.

Hernández-Orallo, José. 2025. Enhancement and assessment in the AI age: An extended mind perspective. *Journal of Pacific Rim Psychology* 19. 1–12. https://doi.org/10.1177/18344909241309376.

Hint, Helen. 2024. Tartu Ülikooli seisukoht tekstiroboti kasutamise tuvastuse kohta. *Tartu Ülikooli e-õppe ajakiri.* https://etu.ut.ee/2024/tartu-ulikooli-seisukoht-tekstiroboti-kasutamise-tuvastuse-kohta/ (20.03.2025).

Hint, Helen, Djuddah A. J. Leijen & Anni Jürine. 2022. Eestikeelse akadeemilise teksti tunnustest. *Keel ja Kirjandus* 65(4). 327–353. https://doi.org/10.54013/kk772a3.

Holton, Richard. 1994. Deciding to trust, coming to believe. *Australasian Journal of Philosophy* 72(1). 63–76. https://doi.org/10.1080/00048409412345881.

Jacobs, Oliver L., Farid Pazhoohi & Alan Kingstone. 2024. Large language models have divergent effects on self-perceptions of mind and the attributes considered uniquely human. *Consciousness and Cognition* 124. 1–6. https://doi.org/10.1016/j.concog.2024.103733.

Jones, Karen. 1996. Trust as an affective attitude. *Ethics* 107(1). 4–25. https://doi.org/10.1086/233694.

Jõgeda, Tiina. 2008. Miks kohus keelas „Magnuse" näitamise? *Eesti Ekspress*, 15.05.

Karjus, Andres. 2023. Tehisintellekt seab juba sel sügisel koolid sundvaliku ette. *Novaator*, 16.06. https://novaator.err.ee/1609007981/andres-karjus-tehisintellekt-seab-juba-sel-sugisel-koolid-sundvaliku-ette (16.04.2025).

Khawaja, Zoha, & Jean-Christophe Bélisle-Pipon. 2023. Your robot therapist is not your therapist: Understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health* 5. 1–13. https://doi.org/10.3389/fdgth.2023.1278186.

Kim, Myunsung, Seonmi Lee, Sieun Kim, Jeong-in Heo, Sangil Lee, Yu-Bin Shin, Chul-Hyun Cho & Dooyoung Jung. 2025. Therapeutic potential of social chatbots in alleviating loneliness and social anxiety: Quasi-experimental mixed methods study. *Journal of Medical Internet Research* 27. 1–14. https://doi.org/10.2196/65589.

Knoth, Nils, Antonia Tolzin, Andreas Janson & Jan Marco Leimeister. 2024. AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence* 6. 1–14. https://doi.org/10.1016/j.caeai.2024.100225.

Korteling, J. E. Hans & Alexander Toet. 2022. Cognitive biases. Sergio Della Sala (toim), *Encyclopedia of Behavioral Neuroscience* [2. trükk], 610–619. Elsevier Science. https://doi.org/10.1016/B978-0-12-809324-5.24105-9.

Kuul, Marek (toim). 2018. Tartu volikogulane kaebas Tähismaad laimu eest kohtusse. *ERR*, 17.12. https://www.err.ee/885372/tartu-volikogulane-kaebas-tahismaad-laimu-eest-kohtusse (15.04.2025).

Lee, Ju Yoen. 2023. Can an artificial intelligence chatbot be the author of a scholarly article? *Journal of Educational Evaluation for Health Professions* 20(6). 1–6. https://doi.org/10.3352/jeehp.2023.20.6.

Lee, Inju & Sowon Hahn. 2024. On the relationship between mind perception and social support of chatbots. *Frontiers in Psychology* 15. 1–10. https://doi.org/10.3389/fpsyg.2024.1282036.

Lee, Hao-Ping (Hank), Ian Drosos, Advait Sarkar, Sean Rintel, Nicholas Wilson, Lev Tankelevitch & Richard Banks. 2025. *The impact of Generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers*. https://www.microsoft.com/en-us/research/wp-content/uploads/2025/01/lee_2025_ai_critical_thinking_survey.pdf (15.04.2025).

Levy, Rachel. 2024. Neuralink implanted second trial patient with brain chip, Musk says. *REUTERS*, 05.08. https://www.reuters.com/technology/neuralink-implanted-second-trial-patient-with-brain-chip-musk-says-2024-08-04/ (07.08.2024).

Lloyd, Dan. 2024. What is it like to be a bot? The world according to GPT-4. *Frontiers in Psychology* 15. 1–9. https://doi.org/10.3389/fpsyg.2024.1292675.

Luria, Alexander R. 1976. *Cognitive development: Its cultural and social foundations*. Cambridge: Harvard University Press.

Makiev, Konstantinos G., Maria Asimakidou, Ioannis S. Vasios, Anthimos Keskinis, Georgios Petkidis, Konstantinos Tilkeridis, Athanasios Ververidis & Efthymios Iliopoulos. 2023. A study on distinguishing ChatGPT-generated and human-written orthopaedic abstracts by reviewers: Decoding the discrepancies. *Cureus* 15(11). https://doi.org/10.7759/cureus.49166.

McLean, Graeme, Kofi Osei-Frimpong, Alan Wilson & Valentina Pitardi. 2020. How live chat assistants drive travel consumers' attitudes, trust and purchase intentions: The role of human touch. *International Journal of Contemporary Hospitality Management* 32(5). 1795–1812. https://doi.org/10.1108/IJCHM-07-2019-0605.

Menary, Richard. 2007. Writing as thinking. *Language Sciences* 29(5). 621–632. https://doi.org/10.1016/j.langsci.2007.01.005.

Nature Editorial. 2023. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 613, 612. https://doi.org/10.1038/d41586-023-00191-1.

Ng, Davy Tsz Kit, Jac Ka Lok Leung, Samuel Kai Wah Chu & Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2. 1–11. https://doi.org/10.1016/j.caeai.2021.100041.

Nickel, Philip J. 2013. Trust in technological systems. Marc J. Vries, Sven Ove Hansson, Anthonie W. M. Meijers (toim), *Philosophy of Engineering and Technology: Norms in Technology* 9, 223–237. https://doi.org/10.1007/978-94-007-5243-6_14.

Oatley, Keith & Maja Djikic. 2008. Writing as thinking. *Review of General Psychology* 12(1). 9–27. https://doi.org/10.1037/1089-2680.12.1.9.

Overstreet, Matthew. 2022. Writing as extended mind: Recentering cognition, rethinking tool use. *Computers and Composition* 63. 1–12. https://doi.org/10.1016/j.compcom.2022.102700.

Overstreet, Matthew, Diana Akhmedjanova & Silvia Vaccino-Salvadore. 2023. Brainbound *vs*. extended: Contrasting approaches to second-language research writing in digital environments. *Journal of Second Language Writing* 61. 1–16. https://doi.org/10.1016/j.jslw.2023.101019.

Pellegrino, Ganfranco & Mirko Daniel Garasic. 2020. Artificial intelligences as extended minds. Why not? *Rivista Internazionale di Filosofia e Psicologia* 11(2). 150–168. https://doi.org/10.4453/rifp.2020.0010.

Prompts = Getting started with prompts for text-based Generative AI tools. 2023. *Harvard University Information Technology veebileht*, 30.08. https://www.huit.harvard.edu/news/ai-prompts (18.03.2025).

Puusepp, Vivian. 2025. Lisad artiklile „Tekstirobotid ja laiendatud vaim. Kuidas mõista tehisintellekti rolli akadeemilise teksti loomes?". https://doi.org/10.23673/re-510.

Rowlands, Mark, Joe Lau & Max Deutsch. 2020. Externalism about the mind. Edward N. Zalta (toim), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2020/entries/content-externalism/ (22.08.2024).

Sample, Ian. 2023. Science journals ban listing of ChatGPT as co-author on papers. *The Guardian*, 26.01. https://www.theguardian.com/science/2023/jan/26/science-journals-ban-listing-of-chatgpt-as-co-author-on-papers (15.04.2025).

Sepp, Andres. 2024. Harno süüdistas Saaremaa koolilõpetajat AI abil eksami tegemises, kuid kaotas vaidluse. *Saarte Hääl*, 25.06. https://saartehaal.postimees.ee/8047401/harno-suudistas-saaremaa-koolilopetajat-ai-abil-eksami-tegemises-kuid-kaotas-vaidluse (15.04.2025).

Seppo, Indrek. 2024. Chat GPT on tavainimesele vestluspartner, kaaslane, teabeallikas ja õpetaja. *ERR*, 18.01. https://menu.err.ee/1609225710/chat-gpt-on-tavainimesele-vestluspartner-kaaslane-teabeallikas-ja-opetaja (15.04.2025).

Sparrow, Betsy, Jenny Liu & Daniel M. Wegner. 2011. Google effects on memory: Cognitive consequences of having information at our fingertips. *Science* 333(6043). 776–778. https://doi.org/10.1126/science.1207745.

Stokel-Walker, Chris. 2023 ChatGPT listed as author on research papers: Many scientists disapprove. *Nature* 613. 620–621. https://doi.org/10.1038/d41586-023-00107-z.

Taylor, Charles. 1985. The concept of a person. *Philosophical Papers 1: Human Agency and Language*. 97–114. Cambridge University Press. https://doi.org/10.1017/CBO9781139173483.005.

Tragel, Ilona & Helen Hint. 2025. TI-hüpe üle õpetajate. *ERR*, 05.03. www.err.ee/1609622519/ilona-tragel-ja-helen-hint-ti-hupe-ule-opetajate (15.03.2025).

usaldusväärsus. Eesti keele seletav sõnaraamat 2009. Eesti Keele Instituut. https://arhiiv.eki.ee/dict/ekss/index.cgi?Q=usaldusväärsus (17.04.2025).

vestluspartner. Sõnaveeb 2025. Eesti Keele Instituut. https://sonaveeb.ee/search/unif/dlall/dsall/vestluspartner/1/est (17.04.2025).

Viiba koostamine. *Tehisintellekt Tartu Ülikoolis veebileht,* https://sisu.ut.ee/ti/viiba-koostamine/ (08.03.2025).

Xie, Tianling & Iryna Pentina. 2022. Attachment theory as a framework to understand relationships with social chatbots: A case study of Replika. *Proceedings of the 55th Hawaii International Conference on System Sciences.* 2046–2055. https://doi.org/10.24251/HICSS.2022.258.

Xie, Tianling, Iryna Pentina & Tyler Hancock. 2023. Friend, mentor, lover: Does chatbot engagement lead to psychological dependence? *Journal of Service Management* 34(4). 806–828. https://doi.org/10.1108/JOSM-02-2022-0072.

Abstract. Vivian Puusepp: Chatbots and the extended mind. How to understand the role of artificial intelligence in academic writing? In this article, I discuss two issues: when a learner uses the assistance of a chatbot in academic writing, (1) how could the role of the chatbot be understood in this process? and (2) assuming academic writing is considered an extended cognitive process, does the chatbot extend the human mind? Studying the possible roles of the chatbot shows that the latter cannot be placed into any category that we are accustomed to in the context of academic writing: it is neither a source,

a (co-)author, a conversation partner, nor a mere tool. Discussing the second question, I draw on the extended mind approach, which holds that human mental processes sometimes extend beyond the brain and body, incorporating external processes or tools as parts of the cognitive system. I argue that the extent to which the use of a chatbot meets the criteria of reliable coupling also determines its impact on the human mind. In the article, I clarify that the most crucial criterion for extending the human mind is the criterion of trust, which presupposes the user's ability to critically evaluate the chatbot's output. If the user blindly trusts the chatbot and relies on its output without any critical reflection, then the use of the chatbot may actually narrow the human mind, hindering the development of the user's writing and thinking skills.

Keywords: large language model, LLM, chatbot, AI, extended mind, academic writing, academic texts, learning