

Tartu Ülikool
Sotsiaalteaduste Valdkond
Psühholoogia instituut

Brett Pupart

Gümnaasiumiõpilaste liigne tuginemine ChatGPT valedetele soovitudele

Uurimistöö

Juhendaja: Jaan Aru, *PhD*

Läbiv pealkiri: ChatGPT liigse tuginemise uuring

Tartu 2025

Gümnaasiumiõpilaste liigne tuginemine ChatGPT valedetele soovitudele

Kokkuvõte

Käesoleva uurimistöö eesmärgiks oli uurida, kui tihti võtavad keskkooli õpilased matemaatikaülesannete lahendamise ajal ChatGPT valesid soovitusi omaks (liigne tuginemine). Uuriti lisaks, kas hariv sekkumine võib valedete soovitude omaks võtmist vähendada. Enne ülesannete lahendamist jaotati katseisikud ($n = 36$) randomiseerimise alusel kontrollgruppi ja sekkumise gruppi, kus esitati hariv tekst ChatGPT töömehhanismi, puuduste ja targasti kasutamise kohta. Ülesannete lahendamisel oli katseisikutel abiks ChatGPT soovitus, mis oli poolte ülesannete puhul väär. Tulemustes ilmnes, et õpilased võtsid ChatGPT vale soovitusi omaks 52.1% kordadest. See viitab tõsiasjale, et liigne tuginemine on laialt levinud probleem ChatGPT kasutajate hulgas. Sekkumise mõju ei olnud statistiliselt oluline liigse tuginemise vähendamisel ($p > .05$), muutes teaduses oluliseks ka edasiste sekkumiste mõju uurimise.

Märksõnad: ChatGPT, suured keelemudelid, liigne tuginemine, harimine, sekkumine

High school students' over-reliance on incorrect ChatGPT recommendations**Abstract**

The aim of this study was to investigate how often high school students adopt incorrect suggestions from ChatGPT (over-reliance) while solving mathematical tasks. It also examined whether an educational intervention could reduce the adoption of incorrect suggestions. Before solving the tasks, participants ($n = 36$) were randomly assigned to a control or an intervention group, which was presented with an educational text explaining ChatGPT's mechanism, limitations, and proper use. During the tasks, participants had access to ChatGPT's recommendations. incorrect in half of the cases. Results showed students adopted incorrect suggestions 52.1% of the time, indicating widespread over-reliance. The effect of the intervention was not statistically significant ($p > .05$), highlighting the need for further research on effective interventions.

Keywords: ChatGPT, LLM, over-reliance, education, intervention

Gümnaasiumiõpilaste liigse tuginemine ChatGPT valedetele soovitudele

Vaswani et al., (2017) tutvustasid artiklis „*Attention is All You Need*“ esmakordselt transformerite arhitektuuri. See pani aluse suurte keelemudelite (edaspidi: keelemudel) revolutsioonile, mille tänaseks kuulsaim toode on juturobot ChatGPT. Keelemudelid on tõenäosuslikud mudelid, mis ennustavad eelmiste *token*-ite, milleks võivad olla tähed, sõnad, arvud või muud sümbolid, põhjal järgmisi *token*-eid (Naveed et al., 2024). Kui kasutaja kirjutab keelemudelile sisendi (nt „Lõpeta lause: parem karta kui...“), siis keelemudel arvutab, mis sõna eelnevatele sõnadele tõenäoliselt järgneb ning vastab „kahetseda“. See pealtnäha lihtne töömehhanism võimaldab neil edukalt lahendada väga erinevaid ülesandeid - isegi selliseid, mille eelduseks on varasemad uurijad pidanud inimese tasemel intelligentsust. Näiteks Winogradi skeemide ülesandes (Levesque et al., 2012) tuleb masinal õigesti aru saada kahetähenduslikest lausetest (nt „Karikas ei mahtunud pruuni kohvrissi, sest see oli liiga suur. Mis oli liiga suur?“). Säärastes ülesannetes on keelemudelite täpsus juba ligi 90% (Kocijan et al., 2023).

Tänu oma võimekusele ja mugavusele on keelemudelid kasutust leidnud mitmetes valdkondades, näiteks hariduses. Paljud kooliõpilased kasutavad ChatGPT-d kooliülesannete lahendamiseks (Valova et al., 2024). Õpilased ise leiavad, et ChatGPT on mugav abivahend, sest see aitab aega kokku hoida, jagab olulist informatsiooni ja annab individuaalset tagasisidet (Ngo, 2023). Õpetajad saavad keelemudeleid kasutada näiteks õppematerjalide ja testide koostamiseks (Lo, 2023). Keelemudelite kasutamine ei piirne üksnes haridusega, vaid need on kasutust leidnud ka näiteks meditsiinis (Thirunavukarasu et al., 2023) ja teaduses, kus pärast ChatGPT avaldamist on oluliselt kasvanud keelemudelite poolt modifitseeritud tekstide osakaal (Liang et al., 2024).

Kuigi keelemudeleid kasutatakse laialdaselt, siis esineb nende väljundites probleeme. Sagedase probleemina on kirjeldatud väljundite ebatäpsust (Lo, 2023; Thirunavukarasu et al., 2023). Keelemudelid võivad produtseerida ebatõest ja täiesti fiktiivset informatsiooni – nähtus, mida tehisintellekti uurijad on hakanud kutsuma hallutsineerimiseks (Borji, 2023). Näite hallutsineerimisest leiab käesoleva töö Lisast B, kus ChatGPT väidab, et Eesti esimene olümpia hõbemedalist oli Jaan Tõnisson, kes ei olnud tegelikult sportlane, vaid tähtis riigitegelane. Uurijad on ka ChatGPT arutlusoskust pidanud ebausaldusväärseks (Bang et al., 2023). Kuigi keelemudelid võivad üldistes vaimse võimekuse testides saada väga kõrgeid tulemusi (GPT o1-preview IQ skoor mõõdeti 121 punkti (Lott, 2024)), siis võivad need mõnikord eksida väga lihtsate ülesannetega. Kurikuulus näide on ChatGPT suutmatus

tokeniseerimise tõttu kokku lugeda, mitu „r“ tähte on sõnas „*strawberry*“ (Mormando, 2024). Lisaks väärinformatsioonile võivad keelemudelite väljundid olla sotsiaalselt kallutatud, näiteks sisaldades soostereotüüpe (Kotek et al., 2023) ja poliitilisi vaateid (Rozado, 2023). Põhjalikku kokkuvõtet erinevatest vigadest, mida keelemudelid teevad, saab lugeda Borji (2023) artiklist.

Kui keelemudelid väljendaksid usaldusväärselt, kui nad mingis vastuses kindlad ei ole, siis ei kujutaks hallutsineerimine endast nii suurt ohtu. Kasutajad saaksid lähtudes keelemudeli metakognitiivsest hinnangust otsustada selle väljundi usaldusväärsuse üle. Kuid tänapäeva keelemudelid väljendavad ka väära arutluskäiku väga enesekindlalt (Azaria et al., 2024) ja nad on oma väljunditele kindlushinnangu andmises ebatäpsed (Xiong et al., 2024). Vastupidiselt ootustele on näidatud, et mida suuremaks mudelid lähevad, seda ebausaldusväärsemaks need muutuvad. Suuremad mudelid annavad sagedamini valesid, kuid pealtnäha usaldusväärseid vastuseid (Zhou et al., 2024). See tähendab, et kasutajad peavad keelemudelite väljunditesse suhtuma kriitiliselt ning vajadusel nende tõesust teistest allikatest üle kontrollima.

Kui kasutaja ei suhtu keelemudelite väljunditesse kriitiliselt, siis on oht, et ta võtab selle valed soovitused omaks – nähtus, mida kutsutakse liigseks tuginemiseks (Passi & Vorvoreanu, 2022). Eksperimentaalsed uuringud on korduvalt näidanud, et liigne tuginemine on probleem tehisintellekti abiga probleemide lahendamisel. Buçinca et al., (2021) uuringus esitati katseisikutele foto toidust ning nende ülesandeks oli valida kümnest toiduainest kõige kõrgema süsivesiniku sisaldusega toiduaine ning asendada see madala süsivesiniku sisaldusega toiduainega, võttes arvesse, et toidu maitse muutuks võimalikult vähe. Neil oli vastamisel abiks eksperimenterite poolt simuleeritud tehisintellekt, mis andis mõnikord valesid soovitusi. 30% vale soovitusena ülesannete puhul esines liigset tuginemist (Buçinca et al., 2021). Jacobs et al., (2021) eksperimendis langetasid psühhiaatrid otsust, milliseid antidepressante patsientidele välja kirjutada. Neile oli abiks tehisintellekt, mis andis mõnikord valesid soovitusi. Tulemustes ilmnas, et kui tehisintellekt andis vale soovitusi, siis oli psühhiaatrite keskmine täpsus oluliselt madalam kui siis, kui tehisintellekti soovitus puudus (Jacobs et al., 2021).

Kuigi mitmed eksperimentaalsed uuringud on näidanud, et tehisintellekti kasutajatel esineb liigset tuginemist, siis ei ole teada, kui võrd need tulemused on üldistatavad tänapäeval laialdaselt kasutatavatele keelemudelitele. Varasemates uuringutes kasutati simuleeritud tehisintellekte, mis võimaldas eksperimenteritele suuremat kontrolli katsetingimuste üle,

kuid võis sellega vähendada ökoloogilist valiidsust. Buçinca et al. (2021) selgitasid liigset tuginemist duaalse töötluse mudeliga. Kognitiivselt oleks liiga kulukas hinnata iga tehisintellekti poolt produtseeritud soovitusete tõesust eraldi. Seetõttu moodustab probleemilahendaja üldise heuristiku kasutatava tehisintellekti kompetentsuse kohta ning kasutab seda soovitusete tõesuse hindamisel (Buçinca et al., 2021). Kui simuleeritud tehisintellektiga katsetes katseisikud alles moodustavad üldise usaldusväärse heuristikuid, siis eeldatavasti on need laialdaselt kasutust leidvate keelemudelite puhul juba olemas. Seetõttu ei pruugi simuleeritud tehisintellekti uuringud adekvaatselt kajastada, kui palju esineb liigset tuginemist juhtivate keelemudelite puhul. Sellest lähtudes on käesoleva uuringu esimene eesmärk mõõta liigset tuginemist ChatGPT kasutajate hulgas.

Kaudseid tõendeid liigsest tuginemisest on esinenud ka keelemudelite uuringutes. Hariduses läbiviidud uuringud on näidanud, et õpilased sageli interakteeruvad keelemudelitega viisil, mis võib soodustada liigset tuginemist. Valova et al. (2024) küsimustiku tulemustes ilmnes, et 30,4% õpilastest ei kontrolli ChatGPT väljundite tõesust teistest allikatest ning 24,5% nendest usaldavad ChatGPT väljundeid täielikult. Amoozadeh et al. (2024) uuringus lahendasid 15 üliõpilast programmeerimisülesandeid ChatGPT abiga. Mitmed õpilased sisestasid ülesande tervikuna ChatGPT-le sisendiks ning võtsid selle väljundi omaks ilma kontrollimata (Amoozadeh et al., 2024).

Tehisintellekti alane harimine sekkumisena

Seoses liigse tuginemisega kaasnevate ohtudega, on oluline teaduses uurida liigset tuginemist vähendavaid sekkumisi. Liigne tuginemine sõltub nii kasutaja kui ka kasutatava tehisintellekti omadusest, mistõttu leidub ka teaduskirjanduses kahte tüüpi sekkumisi. Esimesed on manipuleerinud tehisintellekti omadustega ning mõõtnud selle mõju liigsele tuginemisele. Näiteks Kim et al. (2024) näitasid, et kui tehisintellekt väljendab verbaalset kindlushinnangut (nt „ma ei ole kindel, aga...“), siis liigne tuginemine väheneb. Kuigi säärase sekkumiste uurimine on kahtlemata oluline, siis tasub meeles pidada, et kasutaja peab interakteeruma praegu käibel olevate tehisintellektidega ning tema võimuses ei ole selle parameetreid muuta. Seetõttu on teaduses jätkuvalt oluline uurida ja eksperimentaalselt mõõta ka subjektiivsete sekkumiste efektiivsust.

Heuristilist mõtlemist on peetud üheks faktoriks, mis mõjutab seda, kui palju inimene tehisintellekti valesid soovitusi omaks võtab (Buçinca et al., 2021). Sellest võib eeldada, et liigset tuginemist võivad vähendada sekkumised, mis survestavad probleemilahendajat ümber

lülituma heuristiliselt analüütilisele mõtlemisele. Mulle teadaolevalt on sääraseid sekkumisi mõõdetud ainult ühes uuringus. Buçinca et al. (2021) uuringu tulemustes ilmnes, et kognitiivsed sekkumised nagu ajaline viivitus enne tehisintellekti soovitus nägemist ja ülesande esialgu üksinda lahendamise osutusi efektiivseks liigse tuginemise vähendamisel. Siiski esines ka sekkumise grupis 26% liigset tuginemist (Buçinca et al., 2021), mis muudab oluliseks ka edasiste sekkumiste uurimise.

Uurijad on oluliseks pidanud õpilaste harimist keelemudelite piirangute sh. potentsiaalselt ebatäpsete ja väärinformatsiooni sisaldavate väljundite osas (Lo, 2023). Tehisintellekti alasest harimisest rääkides on uurijad sageli kasutanud tehisintellekti kirjaoskuse mõistet. Tehisintellekti kirjaoskust on erinevad teadlased kontseptualiseerinud mõnevõrra erinevalt, Walter (2024) kirjeldas nende käsitluste ühisjooni. Tema sõnul koosneb tehisintellekti kirjaoskus järgmistest osadest: (1) Arhitektuur - kasutajal on arusaam tehiskirjandusest ja ta on teadlik, et tegemist on statistiliste mudelitega; (2) Piirangud – kasutaja mõistab, milliste ülesannetega tehisintellekt hakkama saab ja millistega mitte; (3) Probleemikohad – kasutaja on teadlik, milliseid vigu tehisintellekt teha võib; (4) Rakendatavus ja parim praktika – kasutaja on teadlik, kuidas tehisintellekti targasti kasutada; (5) Eetika – kasutaja on teadlik tehisintellekti võimalikest eetilistest ohtudest (Walter, 2024).

Pakun välja, et tehisintellekti alane harimine võib liigset tuginemist vähendada kahe mehhanismi kaudu. Esimest mehhanismi võib kutsuda otseseks. See seisneb asjaolus, et kui inimene on teadlikum, kuidas ChatGPT töötab ning milliseid vigu see teha võib, siis suhtub ta selle väljunditesse kriitilisemalt ning liigne tuginemine väheneb. Lisaks otsesele mõjule võib rääkida ka potentsiaalsest kaudsest mõjust. Kaudne mõju seisneb selles, et inimesed, kes teavad keelemudelitest rohkem, interakteeruvad nende süsteemidega viisil, mis vähendavad vigade tekkimise tõenäosust. Spetsiaalsed sisendi kirjutamise meetodid võivad keelemudelite hallutsineerimist vähendada (Tonmoy et al., 2024). Sarnaselt on näidatud, et mõned sisendi kirjutamise meetodid vähendavad ka mudelite liigset enesekindlust (Xiong et al., 2024). See tähendab, et kõrge tehisintellekti teadlikkusega kasutaja puutub harvemini kokku liigse enesekindlusega esitatud väljunditega, mis sisaldavad väärinformatsiooni, mistõttu liigne tuginemine väheneb ka siis, kui ta kasutab nendega interakteerudes kiiret mõtlemist ja heuristikuid.

Vaatamata asjaolule, et tehisintellekti kirjaoskuse arendamist on peetud oluliseks ja lähtudes teoreetilisest raamistikust on oodatav tulemus, et see võiks liigset tuginemist vähendada, siis puuduvad mulle teadaolevalt veel eksperimentaalsed uuringud, mis oleksid

mõõtnud selle mõju liigsele tuginemisele. Käesolev uurimistöö üritab täita seda lünka teaduskirjanduses, uurides, kas hariv tekst võib olla efektiivne sekkumine liigse tuginemise vähendamiseks.

Käesolev uurimistöö

Käesoleva uurimistöö eesmärk oli uurida õpilaste liigset tuginemist ChatGPT valedetele soovitudele. Uuringus lahendasid katseisikud matemaatikaülesandeid koos kuvatõmmisega ChatGPT vastusest samale ülesandele. Enne ülesannete lahendamist jaotati katseisikud randomiseerimise alusel sekkumise gruppi ja kontrollgruppi. Sekkumise grupis lugesid katseisikud lühikest teksti, mis kirjeldas ChatGPT töömehhanismi, võimalikke puudusi ja targasti kasutamist. Kontrollgrupis loeti teksti, mis sisaldas ainult üldist informatsiooni ChatGPT keelemudeli kohta. Pärast tekstide lugemist hakkasid katseisikud ülesandeid lahendama. Iga ülesande puhul pidid katseisikud esmalt otsustama (vastates Jah/Ei), kas nad kasutavad ChatGPT soovitusi oma vastusena või mitte. Kui nad ei kasutanud ChatGPT soovitusi, siis avanes neil võimalus ülesanne iseseisvalt lahendada, valides neljast vastusevariandist selle, mis oli nende arust õige vastus. Kokku lahendasid katseisikud 8 ülesannet ning ChatGPT soovitusi täpsus oli 50%.

Uuringu põhiliseks sõltuvaks muutujaks oli liigne tuginemine, mille operatsiooniline definitsioon oli: ChatGPT annab vale soovitusi ja katseisik kasutab seda oma vastusena. Lähtudes varasemast teaduskirjandusest oli käesoleval uuringul kaks eesmärki. Uuringu esimene eesmärk oli vastata uurimisküsimusele, kui palju esineb õpilastel liigset tuginemist ChatGPT soovitudele. Uuringu teine eesmärk oli mõõta tehisintellekti alase harimise mõju liigsele tuginemisele ning sellest lähtudes püstitasin kaks hüpoteesi, mis mõlemad pidid leidma kinnitust, et sekkumist saaks lugeda efektiivseks:

H1: Sekkumise grupis esineb liigset tuginemist statistiliselt oluliselt vähem kui kontrollgrupis.

H2: Sekkumise grupis ei eirata ChatGPT õigeid soovitusi statistiliselt oluliselt rohkem kui kontrollgrupis

Lisaks eelkirjeldatud muutujatele, küsiti katseisikutelt pärast iga ülesande lahendamist Likerti skaalal kindlushinnangut ja mõõdeti esmase otsuse langetamisele („Kas kasutad ChatGPT soovitusi?“) kulunud aega.

Meetod

Valim

Uuringu valimi moodustasid Eesti gümnaasiumide abiturientid. Esialgu saadeti kutsed uuringus osalemiseks meili teel gümnaasiumidele, mille eksamitulemus jäi 2. või 3. kvartiili (Postimees, 2024) eesmärgiga tagada valimi representatiivsus keskmisele Eesti gümnaasiumiõpilasele. Tulenevalt väikesest osalejate arvust, saadeti kutseid ka gümnaasiumidele väljaspool seda vahemikku. Kokku saadeti kutseid 72 koolile, millest uuringus osales 10 kooli. Koolid, mis andsid oma nõusoleku, edastasid kutse uuringus osalemiseks oma abiturientidele. Küsimustiku täitis lõpuni 42 õpilast, kuid 6 neist ei osalenud andmeanalüüsis järgnevatel põhjustel. Viis katseisikut jäid välja, sest nad lugesid sekkumise teksti läbi vähem kui poole minutiga ja üks katseisik jäi välja, sest ta täitis terve küsimustiku vähem kui kuue minutiga. Lõplik valim koosnes seega 36 osalejast (69,4% naised). Kõik osalejad olid abiturientid vanuses 18-19. Kõikide osalejate vahel loositi välja 5x20€ raamatupoe kinkekaarti, et motiveerida neid uuringus osalema.

Materjalid

Katseisikutele esitatud tekst oli uuringu sõltumatu muutuja, millel oli kaks taset: sekkumise tekst ja kontrolltekst. Sekkumise teksti eesmärk oli arendada sekkumise grupi tehisintellekti kirjaoskust, tuginedes varasemale teaduskirjandusele. Teksti lugemise ajaline pikkus oli keskmiselt 4,4 minutit. Tekst hõlmas kõiki (Walter, 2024) artiklis kirjeldatud tehisintellekti kirjaoskuse alaoskusi, välja arvatud eetikat, sest seda käsitlevad muud juhised. Lisaks tekstile oli esitatud kuvatõmmisena näiteid ChatGPT hallutsineerimisest ja teistest piirangutest. Kontrollteksti koostamisel pidasin oluliseks, et see sisaldaks võimalikult vähe informatsiooni tehisintellekti kirjaoskuse alaoskuste kohta, ent tunduks katseisikutele katse seisukohalt siiski relevantne. Kontrolltekst sisaldas lihtsalt üldist informatsiooni ChatGPT keelemudeli kohta. Informatsioon mõlema teksti sisu ja peatükkide kohta on välja toodud Tabelis 1. Tekstid koos viidetega teaduskirjandusele on täies mahus esitatud töö Lisas B.

Tabel 1*Katseisikutele esitatud tekstide sisu*

Teksti osa	Eesmärk
Üldine sissejuhatus	Selgitada katseisikutele väga üldiselt, et ChatGPT on juturobot, millelt saab küsida küsimusi ja saada teksti formaadis vastuseid.
Tugevused	Selgitada katseisikutele, millised on ChatGPT tugevused.
Töömehhanism	Selgitada katseisikutele, et ChatGPT on statistiline mudel, mis ennustab järgmise <i>tokeni</i> tõenäosust.
Piirangud ja ohud	Tutvustada katseisikutele ChatGPT peamisi ohte ja piiranguid.
Kasutamine	Selgitada katseisikutele, kuidas ChatGPT-d targasti kasutada sh. sisendi kirjutamine ja sagedased vead.

Märkused. Paksus kirjas on tähistatud osad, mis esitati ainult sekkumise tekstis.

Liigse tuginemise mõõtmiseks kasutasin kanguruülesandeid (Teaduskool, 2022). Ülesanded on matemaatilise sisuga, kuid ülesannete koostajate sõnul ei mõõda need ainult valemite tundmist, vaid mõistete loogilist kombineerimist, et jõuda õige vastuseni (About-Math Kangaroo, s.a.). Kanguruülesanded sobisid liigse tuginemise mõõtmiseks kahel põhjusel. Esiteks on nendel ainult üks õige vastus, mis muutis liigse tuginemise mõõtmise binaarseks (jah/ei) ja seega lihtsasti tuvastatavaks. Teiseks, eeldavad need ülesanded sageli iteratiivset arutlust, mistõttu nõuab nende õigesti lahendamine võrdlemisi palju töömälu ressursse. See jäljendas ChatGPT kasutamist päriselus, kus faktide kontrollimine nõuab lisapingutust ja lihtsam on ChatGPT soovitusi pimesi usaldada.

Testisin mitmeid ülesandeid erinevate aastate arhiivist. Valisin kõik ülesanded vanusekategorias „kadett“ ehk 7-8. klassidele mõeldud ülesanded, et ülesannete raskusaste oleks ühtlasem. Kirjutasin ülesande tervikuna ChatGPT(versioon 4-o)-le sisendiks ning andsin korralduse „Lahenda õigesti“. Otsustasin uuringus kasutada kaheksat ülesannet (Lisa A) ja ChatGPT soovitusel 50% täpsust, et uuring ei oleks ajaliselt liiga pikk, ent võimaldaks siiski piisavalt palju liigse tuginemise mõõtmiskordasid. Ülesannete valikul lähtusin põhimõttest, et need ei oleks nii lihtsad, et katseisikud lahendavad kõik ülesanded iseseisvalt ja ei süvene ChatGPT soovitusse. Vale soovitusel ülesannete puhul pidi ChatGPT soovitus olema pealtnäha usaldusväärne, kuid sisaldama siiski piisavalt lihtsaid vigu, et kui katseisik otsustas sellesse kriitiliselt süveneda, siis oleks ta pidanud aru saama, et ChatGPT soovitus on

vale. Tabelis 2 võib näha, et ChatGPT soovitusel sisaldasid vigu, mida abiturientid peaksid olema kahtlemata suuteliselt märkama, kui nad soovitusse kriitiliselt süvenevad. See tähendab, et liigse tuginemise vältimiseks ei olnud vajalik, et katseisik suudab ülesande iseseisvalt lahendada.

Selleks, et kinnitada, et väljavalitud ülesanded ei ole liiga lihtsad, viisin keskharidusega inimeste hulgas läbi pilootkatse ($n = 7$). Tulemused kinnitasid, et ülesanded ei olnud liiga lihtsad. 25% ülesannete puhul esines liigset tuginemist ning 43% katseisikutest võtsid omaks vähemalt ühe ChatGPT vale soovitusel. Seejuures liigse tuginemise põhjuseks ei saanud pidada juhuslikku valikut, sest mõõdeti ka katseisikute kindlushinnangut („usun, et mu vastus eelmisele ülesandele oli õige“). Kui katseisikud võtsid omaks ChatGPT vale soovitusel, siis oli nende kindlushinnang kõrgem kui võiks oodata juhuslikult vastuselt. Kõikidel kordadel, mil katseisik võttis vale soovitusel omaks, oli tema enesekohane kindlushinnang kas „pigem nõustun“ või „nõustun täielikult“. Need tulemused kinnitasid, et väljavalitud ülesanded sobisid liigse tuginemise mõõtmiseks.

Tabel 2

ChatGPT valede soovitusel vigade selgitus

Ülesanne	ChatGPT viga
Vale 1	Ülesandes on öeldud, et on neli inimest, kes kõik kallistasid oma tuttavat inimest ühe korra. Öeldud on, et Tiina kallistas neli korda. Küsitud on mitu korda kallistas Eerika? ChatGPT väidab, et 0 korda, kuigi Tiina neljast kallistusest peaks loogiliselt järelduma, et Eerika pidi kallistama vähemalt ühe korra.
Vale 2	Ülesandes tuleb leida Siiri ja Ivi vanuste summa. Nii Siiri ja Emma kui ka Siiri ja Rita vanuste summa peab jaguma viiega. ChatGPT hakkab testima Siiri võimalikke vanuseid ning täiesti arusaamatul põhjusel väidab, et Siiri vanus ei saa olla 3, 8, ega 12, sest Siiri vanus + x või y ei saa jaguda 5-ga. Jõuab lõpuks vastuseni, et Siiri vanus on 14, kui $x = 12$ ja $y = 6$. Ülesandes on öeldud, et nende vanused mingis järjestuses on 3, 8, 12 või 14, seejuures 6 ei ole variantide hulgas. Järgmises lauses muudab Rita vanuse ilma põhjendusega kuuelt kaheksaks ja jõuab lõpliku vastuseni, millest järeldub, et arvud 26 ja 22 jaguvad viiega.
Vale 3	Ülesandes on öeldud, et Tiit arvab, et ta kell on 5 minutit ees. Kui ta ühel hetkel arvab, et õige kell on 12:00, siis tema kell näitab $12:00 + 5\text{min} = 12:05$. ChatGPT eksib juba esimeses lauses, väites, et Tiidu kell näitab 11:55, mis tähendaks, et Tiidu kell on 5 minutit taga, mitte ees.
Vale 4	Ülesandes tuleb leida neljast objektist kõige raskem ese. ChatGPT teeb täiesti ebaloogilised ja arusaamatud arvutused ja väidab, et ese B kaalub -200 grammi, millest mingil arusaamatul põhjusel peaks järelduma, et ese E on kõige raskem.

Protseduur

Uuring viidi läbi LimeSurvey veebikeskkonnas. Andmete kogumine toimus 2024. aasta oktoobris. Õpilased, kes soovisid uuringus osaleda, pidid enne uuringus osalemist läbi lugema ja andma oma informeeritud nõusoleku, mis esitati neile veebiküsimustiku avalehel. Informeeritud nõusoleku vorm on täies mahus esitatud töö Lisas C. Kasutasin uuringus sõltumatute gruppidega katseplaani ja ühte sõltumatut muutujat, millel oli kaks taset: sekkumise tekst ja kontrolltekst. Täpsem info tekstide sisu kohta on välja toodud alapeatükis „materjalid“.

Valimi kirjeldamiseks küsiti katseisikutelt andmeid nende demograafilise kuuluvuse kohta (sugu, vanus ja kool). Seejärel jaotas veebikeskkond katseisikud randomiseerimise alusel kahte gruppi, kus neile esitati uuringu sõltumatu muutuja. Pärast tekstide lugemist esitati katseisikutele nelja-pallisel Likerti tüüpi sundskaalal enesekohaseid väiteid, et kontrollida sekkumise teksti mõju. Väiteid oli kokku viis ning need põhinesid alusel „Võrreldes varasemaga arvan, et pärast ChatGPT-d tutvustava teksti lugemist olen paremini informeeritud sellest...“ Sekkumise efektiivsuse hindamisel keskendusin neist kolmele järgnevale: (1) „...kuidas ChatGPT töötab“ (2) „...millised on ChatGPT piirangud ja ohud“ (3) „...kuidas ChatGPT-d targasti kasutada.“

Pärast kontrollküsimusi instrueeriti katseisikuid ülesannete lahendamise osas (Lisa D). Seejärel hakkasid nad lahendama kanguruülesandeid koos ChatGPT abiga. Veebilehe ülemises osas esitati neile juhuslikus järjekorras ülesanne ning selle all kuvatõmmisena ChatGPT soovitus (Joonis 1). Katseisikud pidid langetama esmase otsuse ehk valima, kas nad kasutavad ChatGPT soovitus oma vastusena või mitte, vastates Jah/Ei küsimusele „Kas kasutad ChatGPT vastust?“ Joonisel 1 on toodud näide sellest, milline ülesande lahendamine katseisikute perspektiivist välja nägi. Kui katseisikud vastasid „Ei“, siis avanes neil võimalus ülesanne iseseisvalt lahendada, valides nelja vastusevariandi hulgast selle, mis oli nende arvates õige vastus. Vastusevariantide hulgast olin eemaldanud ChatGPT varasema soovitus, et tagada, et katseisikud päriselt süvenevad ChatGPT soovitusse, mitte ei vasta igale küsimusele „Ei“, et ülesandeid hiljem üksinda lahendada.

Joonis 1

Näide katseisikutele esitatud ülesandest

* Kolmest ruumist ühes on peidus lõvi. Esimese ruumi uksele on kirjutatud "Lõvi on siin". Teise ruumi uksele on kirjutatud "Siin ei ole lõvi". Kolmanda ruumi uksele on kirjutatud " $3^2 - 2^3 = 2^3 - 3^2$ ". Ainult ühe ruumi uksele olev silt on õige. Millises ruumis on lõvi?

Lahendame iga ruumi sildi väiteid:

1. **Esimene ruum:** "Lõvi on siin."
 - Kui see on õige, siis teise silt ("Siin ei ole lõvi") on vale, tekitades vastuolu. **Ei saa olla õige.**
2. **Teine ruum:** "Siin ei ole lõvi."
 - Kui see on õige, siis lõvi ei ole siin ega esimeses ruumis. Lõvi peab olema kolmandas ruumis.
 - Kolmanda ruumi silt $3^2 - 2^3 = 2^3 - 3^2$ ei ole tõene ($1 \neq -1$), mis tähendab, et see silt on vale. **Ainult see silt saab olla õige.**
3. **Kolmas ruum:** $3^2 - 2^3 = 2^3 - 3^2$.
 - Võrrand ei ole tõene. **Ei saa olla õige.**

Järeldus: Teise ruumi silt on õige, seega lõvi on kolmandas ruumis.

Kas kasutate ChatGPT vastust?

Jah Ei

Uuringu peamiseks sõltuvaks muutujaks oli liigse tuginemine. Liigse tuginemise operatsioonaalne definitsioon oli järgmine: ChatGPT annab vale soovitus ja katseisik kasutab seda oma vastusena. Mõõtsin ka kordasid, mil katseisik eiras ChatGPT õigeid soovitusi. Tulenevalt asjaolust, et liigset tuginemist on varasemas teaduskirjanduses seostatud kiire mõtlemisega, mõõtsin ka aega, mis katseisikutel kulus, et langetada otsus, kas nad kasutavad ChatGPT soovitusi või mitte. Eeldus oli, et mida kauem veedab katseisik esmase otsuse peale mõeldes, seda vähem esineb liigset tuginemist. Lisaks mõõtsin, kui palju aega kulus katseisikutel tekstide lugemiseks ja ülesannete lahendamiseks kokku. Pärast iga ülesande lahendamist mõõtsin katseisikute kindlushinnangut Likerti skaalal (1 = ei nõustu üldse – 5 = nõustun täielikult) enesekohase väitega „Usun, et mu vastus eelmisele ülesandele oli õige“.

Eetilised aspektid

Käesolev uurimistöö on kooskõlastatud Tartu Ülikooli inimuuringu eetika komiteega (nr 391/T-14). Uurimistöö planeerimisel arvestati töö eetilistest olulistest aspektidest, et uuringus osalemine ei oleks katseisikute jaoks kahjulik. Kõik uuringus osalejad lugesid enne katses osalemist läbi põhjalikud instruktsioonid ning andsid informeeritud nõusoleku. Uuringut esitleti osalejatele pealkirjaga „Õpilaste motivatsioon matemaatikaülesannete lahendamisel ChatGPT abiga“, sest oli oht, et kui katseisikud on teadlikud, mida uuringus tegelikult mõõdetakse, siis võib see vähendada uuringu valiidsust. Katse lõpus olevas debriifis

tutvustati kõikidele katseisikutele, mida uuringus tegelikult mõõdeti ning ühes sellega tagati ka uuringu eetilisus.

Andmeanalüüs

Andmeanalüüsi viisin läbi vabavaralises statistikaprogrammis JASP (versioon 0.19.1), et kontrollida sissejuhatuses püstitatud hüpoteese ja uurida muutujate vahelisi seoseid. Tuginemise, kindlushinnangu ja sekkumise efektiivsuse andmetega viisin läbi sõltumatute gruppidega Studenti t-testi, et kontrollida gruppide (sekkumine, kontroll) keskmiste vahel statistiliselt olulisi erinevusi. Sõltumatute gruppidega t-testil on viis eeldust: (1) Sõltuv muutuja on vähemalt intervallskaalal; (2) Sõltumatu muutuja on kategoriaalne tunnus; (3) Tunnus on rühmade lõikes ligikaudu normaaljaotusega; (4) Hajuvused on homogeenised; (5) Valimid on sõltumatud. Eeldused 1,2 ja 5 olid täidetud. Normaaljaotuslikkuse eelduse täitmiseks pidi järsakuse- ja asümmeetriakordaja jääma liberaalsesse vahemikku (-2; 2). Hajuvuste homogeenisust hindasin Levene'i testiga ja eeldus oli täidetud, kui $p > .05$. Kui eeldused olid andmestikus täidetud, siis kasutasin Studenti t-testi. Kui oli rikutud normaaljaotuslikkuse eeldus, siis viisin läbi t-testi mitteparameetrilise analoogi ehk *Mann-Whitney U* testi. Hajuvuste homogeenisuse eelduse rikkumise korral kasutasin Welchi statistikut. Analüüsi olulisuse nivooks võtsin $p = .05$. Efektitugevuse raporteerimisel kasutasin Coheni d statistikut.

Viisin läbi korrelatsioonianalüüsi, et kontrollida esmasele otsustusele kulunud aja ja tuginemise vahelist seost. Pearsoni korrelatsioonianalüüsil on kaks eeldust: (1) Andmed on normaaljaotuslikud; (2) Andmete vahel on lineaarne seos. Normaaljaotuslikkuse eelduse täitmiseks pidi järsakuse- ja asümmeetriakordaja jääma liberaalsesse vahemikku (-2; 2). Seoste lineaarsust kontrollisin hajuvusdiagrammi pealt. Kui kumbki nendest eeldustest oli rikutud, siis kasutasin Pearsoni korrelatsiooni mitteparameetrilist analoogi Spearmani rho statistikut. Analüüsi olulisuse nivooks võtsin $p = .05$.

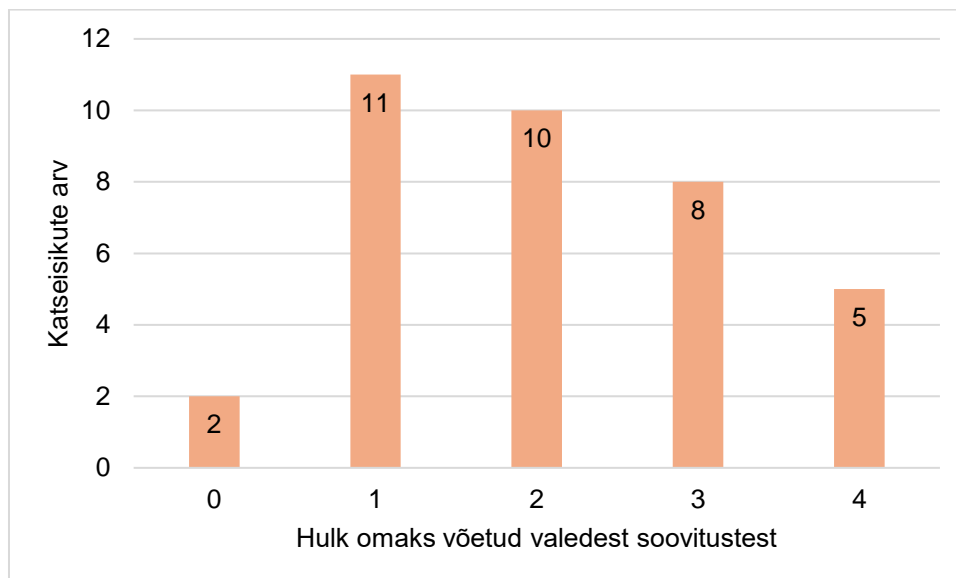
Tulemused

Kokku oli vastajaid 42, kuid analüüsist jäid välja 6 põhjustel, mida on kirjeldatud meetodi osas. Kokku osales analüüsis seega 36 õpilast, neist 19 sekkumise grupis ja 17 kontrollgrupis. Katseisikutel kulus küsimustiku täitmiseks keskmiselt 23.6 minutit. Sekkumise teksti loeti keskmiselt 4.3 minutit ja kontrollteksti 0.9 minutit. Konverteerisin liigse tuginemise iga katseisiku puhul protsentskaalale. Liigset tuginemist esines mõlemas grupis kokku keskmiselt rohkem kui õigete soovitude eiramist. ChatGPT valesid soovitusi

võeti omaks 52.1% kordadest (SD = 28.9). ChatGPT õigeid soovitusi ei võetud omaks 28.5% kordadest (SD = 24). Joonisel 2 on näha, et 34-l katseisikul 36-st esines liigset tuginemist vähemalt ühe vale soovitusiga ülesande puhul. Neist 13-l esines liigset tuginemist vähemalt kolme vale soovitusiga ülesande puhul. Tabelis 3 võib näha, et liigset tuginemist esines kõigi nelja vale soovitusiga ülesande puhul ligi pooltel katseisikutest.

Joonis 2

Katseisikute liigse tuginemise histogramm



Märkus. n = 36. Kokku oli vale soovitusiga ülesandeid 4.

Tabel 3

Liigse tuginemine ja õigesti lahendamine ülesannete lõikes

Ülesanne	Õige	LT
Vale 1	9 (25%)	17 (47%)
Vale 2	10 (28%)	21 (58%)
Vale 3	1 (3%)	18 (50%)
Vale 4	2 (5%)	19 (53%)
M kokku	5.5 (15%)	18.75 (52.1%)

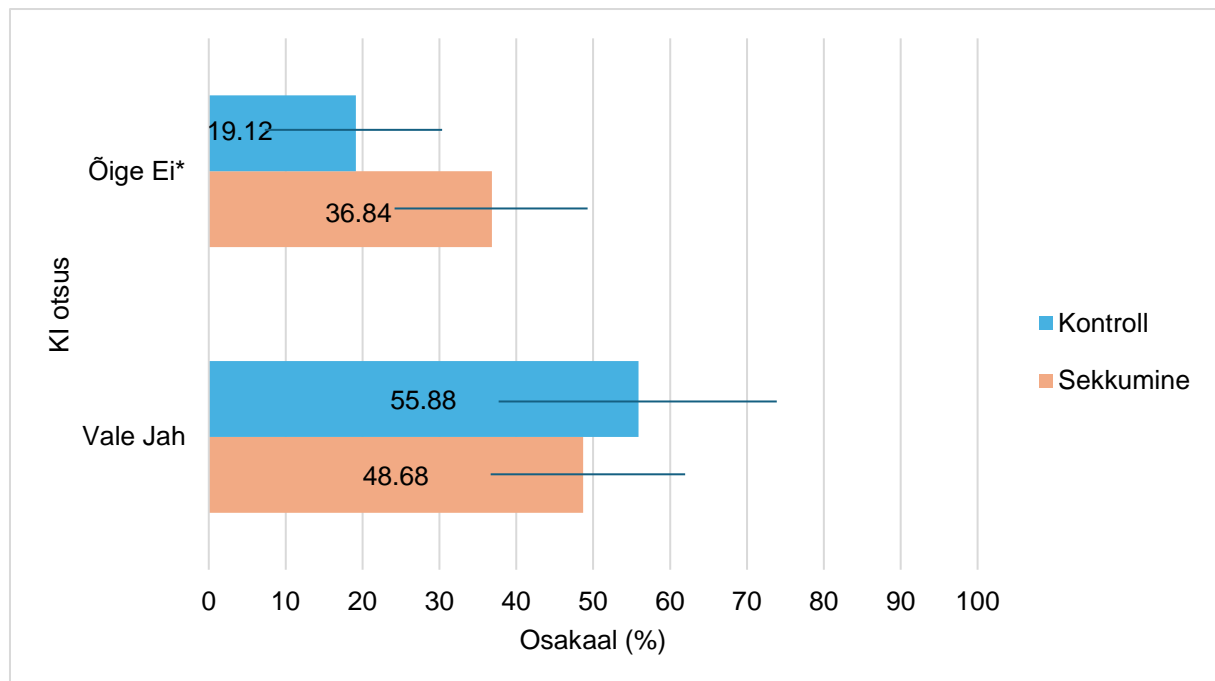
Märkus. LT = liigse tuginemine, Õige = õigesti lahendanute osakaal.

Viisin läbi sõltumatute gruppidega t-testi, et võrrelda liigse tuginemise määra sekkumise grupis (grupp 1) ja kontrollgrupis (grupp 2). Andmetes oli rikutud hajuvuse homogeensuse eeldus ehk kasutasin Welchi korrektuuri. Tulemused näitasid, et keskmine liigse tuginemise määr oli sekkumise grupis madalam (M = 48.68, SD = 24.26) kui kontrollgrupis (M = 55.88, SD = 33.69), kuid erinevus ei olnud statistiliselt oluline, $t(28.78) =$

-0.728, $p = .236$, $d = -0.245$. Õigete soovitude eiramisel olid täidetud kõik analüüsi eeldused ehk viisin läbi Studenti t-testi. Sekkumise grupis oli keskmine õigete vastuste eiramine kõrgem ($M = 36.84$, $SD = 24.11$) kui kontrollgrupis ($M = 19.12$, $SD = 20.78$) ning erinevus oli statistiliselt oluline, $t(34) = 2.349$, $p = .025$, $d = 0.784$. Joonisel 3 on kujutatud mõlema ebaoptimaalse tuginemise määr gruppide lõikes. Uurisin keskmist aega, mis kulus mõlemas grupis esmase otsuse tegemisele. Andmetes oli rikutud normaaljaotuslikkuse eeldus ehk viisin läbi *Mann-Whitney U* testi. Gruppide vahel ei esinenud statistiliselt olulisi erinevusi keskmises esmasele otsusele kulunud ajas ($U = 179.5$, $p = .56$). Sekkumise grupi keskmine oli ($M = 93.52$, $SD = 46.07$) ja kontrollgrupi keskmine oli ($M = 98.62$, $SD = 67.26$).

Joonis 3

Ebaoptimaalne tuginemine gruppide lõikes



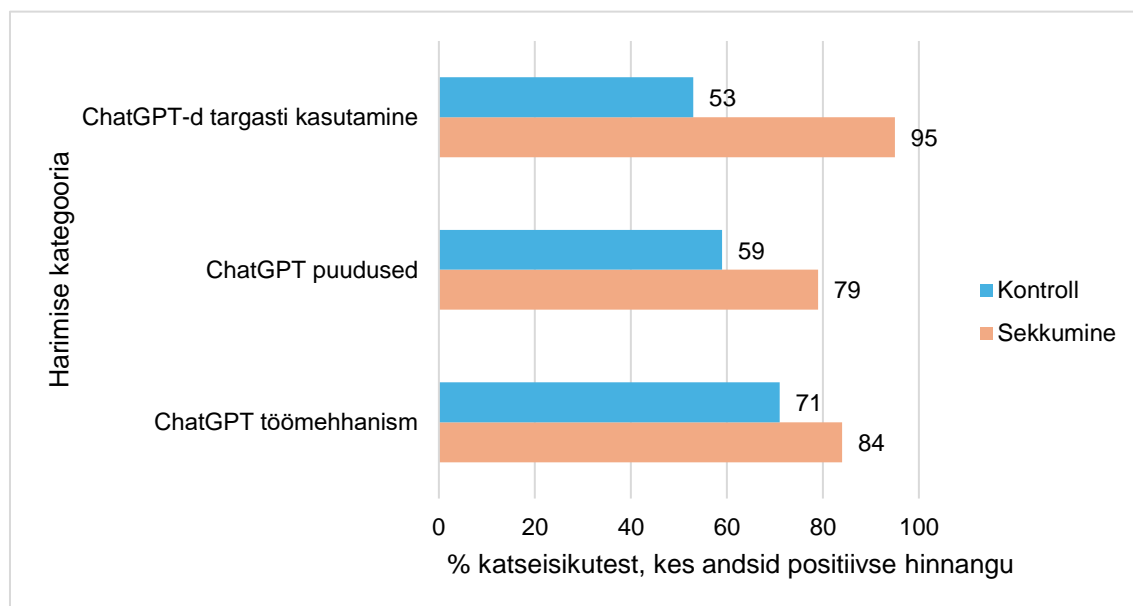
Märkus. Vale Jah = Welchi t-test, Õige Ei = Studenti t-test. Statistiliselt oluline erinevus on tähistatud sümboliga *. Õige tähistab ülesannet, millele ChatGPT andis õige soovitus ja vale tähistab ülesannet, millele ChatGPT andis vale soovitus. „Jah“ tähendab, et katseisik kasutas ChatGPT soovitust. „Ei“ tähendab, et katseisik ei kasutanud ChatGPT soovitust. Osakaal tähistab mitu % vale või õige ChatGPT soovitusel ülesande puhul katseisikud langetasid vastava otsuse. Sinise joonega on tähistatud usalduspiirid (95%).

Selleks, et hinnata, kas sekkumise tekst haris katseisikuid piisavalt ChatGPT töömehhanismi, puuduste ja kasutamise osas, konverteerisin katseisikute enesekohased hinnangud arvilisele skaalale (1-5) ja arvutasin kolme kategooria aritmeetilise keskmise,

mida kasutasin sekkumise efektiivsuse näitajana. Viisin läbi sõltumatute gruppidega t-testi, et hinnata, kas manipulatsioon omas soovitud mõju. Andmetes oli rikutud normaaljaotuslikkuse eeldus ehk kasutasin *Mann-Whitney U* korrektoori. Tulemused näitasid, et sekkumise grupis oli manipulatsiooni efektiivsus kõrgem ($M = 3.28$, $SD = 0.72$) kui kontrollgrupis ($M = 2.82$, $SD = 0.88$), kuid erinevus ei olnud statistiliselt oluline $U = 212.5$, $p = .053$, *rank-biserial correlation* = 0.316. Joonisel 3 on näha, et iga osa puhul oli sekkumise grupis enesekohane hinnang teksti hariva mõju kohta kõrgem kui kontrollgrupis.

Joonis 3

Katseisikute enesekohane hinnang tekstide hariva mõju kohta



Märkus. Katseisikud vastasid enesekohasele väitele „Tunnen, et pärast teksti lugemist olen paremini informeeritud (teksti osa) osas“ Vastused „Pigem jah“ ja „Jah“ loeti positiivsete vastuste alla ning „Pigem ei“ ning „Ei“ loeti negatiivsete vastuste alla.

Selleks, et võrrelda katseisikute kindlushinnangut nendel kordadel, kui nad otsustasid kasutada ChatGPT soovitus oma vastusena, viisin läbi sõltumatute gruppidega t-testi. Andmetes oli rikutud normaaljaotuslikkuse eeldus ehk kasutasin *Mann-Whitney U* testi. Tulemused näitasid, et kindlushinnang oli sekkumise grupis madalam ($M = 3.58$, $SD = 0.46$) kui kontrollgrupis ($M = 3.67$, $SD = 0.76$), kuid erinevus ei olnud statistiliselt oluline, $U = 188.500$, $p = .40$, *rank-biserial correlation* = 0.167.

Korrelatsioonanalüüsiga uurisin katseisikute esmasele otsustusele kulunud aja ja tuginemise vahelisi seoseid. Vale soovitus omaks võtmise ja vale soovitusega ülesannete otsustusele kulunud aja andmete puhul oli rikutud normaaljaotuslikkuse eeldus, mistõttu

arvutasin Spearmani korrelatsioonikordaja. Korrelatsioonanalüüs näitas, et vale soovitus omaks võtmise ja otsustusele kulunud aja vahel on mõõdukas negatiivne korrelatsioon ($r_s = -0.41$, $p = .013$). Õige soovitus eiramise ja õige soovitusena ülesannete otsustusele kulunud aja andmete puhul oli rikutud normaaljaotuslikkuse eeldus, mistõttu arvutasin Spearmani korrelatsioonikordaja. Korrelatsioonanalüüs näitas, et õige soovitus eiramise ja otsustusele kulunud aja vahel on mõõdukas positiivne korrelatsioon ($r_s = 0.35$, $p = .038$).

Arutelu

Käesoleva uuringu esimene eesmärk oli mõõta, kui palju esineb liigset tuginemist ChatGPT soovitustele matemaatikaülesannete lahendamise ajal. Tulemused näitasid, et katseisikud võtsid omaks ligi pooled ChatGPT valedest soovitustest. Uuringu teine eesmärk oli hinnata tehisintellekti alase harimise mõju liigsele tuginemisele. Kumbki sissejuhatuses püstitatud hüpoteesidest kinnitust ei leidnud ehk sekkumine ei olnud efektiivne. Gruppide vahel ei esinenud statistiliselt olulisi erinevusi liigses tuginemises ehk hüpotees 1 ei leidnud kinnitust. Sekkumise grupis eirati statistiliselt oluliselt rohkem ChatGPT õigeid soovitusi ehk hüpotees 2 ei leidnud kinnitust.

Liigse tuginemine

Uuringu tulemustest on võimalik järeldada, et kasutajad ei ole eriti head ChatGPT väljundites vigade tuvastamises, võttes korduvalt valesid soovitusi omaks. See on kooskõlas uuringutega, mis on näidanud, et inimesed sageli ei kontrolli ChatGPT väljundite tõesust või usaldavad selle väljundeid täielikult (Valova et al., 2024). Kasutajad võtsid omaks ligi pooled ChatGPT valedest soovitustest. Seejuures ei olnud see ainult üksikute „laiskade“ õpilaste probleem. Ainult kaks õpilast 36-st ei võtnud omaks mitte ühtegi ChatGPT vale soovitust. Liigse tuginemise esinemine on kooskõlas varasemate uuringutega, mis on näidanud, et tehisintellekti valed soovituselid vähendavad probleemilahendajate sooritust (Buçinca et al., 2021; Jacobs et al., 2021). Käesolevas uuringus esines liigset tuginemist aga rohkem, kui on varasemalt raporteeritud sarnastes uuringutes, näiteks Buçinca et al. (2021) 30%. See võib viidata tõsiasjale, et kasutajate usaldus juhtivate keelemudelite vastu on kõrgem kui varasemates uuringutes kasutatud simuleeritud tehisintellektide vastu, mistõttu võtavad nad rohkem selle valesid soovitusi omaks. Siiski ei ole seda võimalik järeldada üksnes käesoleva uuringu tulemuste põhjal, sest ma ei mõõtnud baastaset.

Lugejal võib tekkida kahtlus, et äkki selgitab kõrget liigset tuginemise määra asjaolu, et ülesanded olid katseisikute jaoks liiga rasked, mistõttu langetasid nad otsuse, kas ChatGPT

soovitusele tugineda või mitte, juhuslikult. Selle poolt räägivad tulemused, et vale soovitusega ülesandeid lahendas õigesti keskmiselt 15% katseisikutest. Seejuures ülesannet Vale 3 lahendas õigesti ainult 1 katseisik 36-st. Selle selgituse vastu räägib aga fakt, et vale soovitusega ülesanded olid valitud spetsiaalselt selle eesmärgiga, et need tunduksid esmapilgul usaldusväärsed, kuid süvenedes ilmselgelt valed. Näiteks Vale 3 ülesande puhul eksis ChatGPT juba esimeses lauses, sest see ei saanud aru, mida tähendab, et kell on ees vs. kell on taga (Tabel 2). Abituriendid on kahtlemata suuteliselt sellest veast aru saama ka siis, kui nad ülesannet ise lahendada ei oska.

Siinkohal pakun selgituse, mis paremini sobitub käesoleva uuringu andmetega ja varasema teaduskirjandusega. Liigse tuginemise kõrge esinemissagedus on tõenäoliselt tingitud inimeste ja keelemudelite omaduste interaktsioonist. Inimestel on kalduvus kiirele mõtlemisele, mis on vastuvõtlik vigadele ja kalletele (Kahneman, 2011). See omadus võimaldab inimestel kiiresti ja võrdlemisi efektiivselt otsuseid langetada ja probleeme lahendada. Käesolevas uuringus esines mõõdukas negatiivne korrelatsioon liigse tuginemise ja esmase otsustuse langetamisele kulunud aja vahel. See tähendab, et mida kauem veetis õpilane mõeldes, kas kasutada ChatGPT soovitust või mitte, seda tõenäolisemalt ei võtnud ta selle valesid soovitusi omaks. Tulemused võimaldavad sarnaselt Buçinca et al. (2021) artiklile liigset tuginemist seostada kiire mõtlemisega. Seejuures õigete soovituste eiramise ja esmase otsustuse langetamisele kulunud aja vahel esines mõõdukas positiivne korrelatsioon. Seda on keeruline varasema teaduskirjandusega seostada, kuid on võimalik, et õige soovitusega ülesanded olid katseisikute jaoks liiga lihtsad, mistõttu võimekamad õpilased langetasid otsuse kiiremini.

Interaktsiooni teine osa on fakt, et keelemudelite näol on meil tegemist tööriistadega, mis on meie jaoks võõrad. Tehisnärvivõrkudel puudub arvutuslik keerukus, mis on omane imetajate ajule (Aru et al., 2023). Lisaks on nende mudelite sooritus ebastabiilne. Näiteks ChatGPT o1-preview IQ skooriks mõõdeti 121 punkti, mis küll vähenes 98 punktile, kui kasutati ülesandeid, mida mudel varasemalt näinud ei olnud, kuid see oli ikkagi võrdne inimeste keskmise tulemusega (Lott, 2024). Samal ajal põrub o1-preview jätkuvalt tehisintellektide jaoks ületamatuks jäänud ARC-AGI testis (Chollet, 2019). Kui inimesed lahendavad keskmiselt 85% ülesannetest õigesti, siis o1-preview tulemus oli kõigest 21% (Knoop, 2024). Kuigi võiks loota, et skaleeritud mudelite sooritus on kergemini ennustatav, siis Zhou et al. (2024) on näidanud, et mida suuremaks muutub mudelite arvutusvõimsus, parameetrite ja treeninguandmete hulk, seda ebausaldusväärsemaks muutuvad need mudelid

inimeste jaoks, sest suuremad mudelid annavad sagedamini pealtnäha usutavaid, kuid valesid vastuseid.

Kui inimene interakteerub nende võõraste süsteemidega kasutades kiiret mõtlemist ja heuristikuid, siis on oht, et ta võtab selle valed soovitused omaks. Varasemad uuringud on näidanud, et ainuüksi selgituse olemasolu võib suurendada liigset tuginemist (Bussone et al., 2015). Meile võib esmapilgul tunduda, et ChatGPT on ekspert igas valdkonnas, sest asjatundmatu inimene ei annaks kunagi nii põhjalikke selgitusi, kui ta teemast päriselt aru ei saaks. Kuid tegelikkuses on ChatGPT matemaatikaprofessori teadmistega tööriist, mis võib eksida lihtsa aritmeetikaga. Kui inimene ei võta seda ChatGPT ebastabiilsust arvesse ja kasutab samu heuristikuid, mida ta kasutab inimeste puhul, keelemudelite väljundite kvaliteedi hindamisel, siis on tulemuseks valede soovituste omaks võtmine. Kui vigaste väljundite omaks võtmine juhtub mõnes kõrge kaaluga valdkonnas nagu meditsiinis või teaduses, siis võivad liigsel tuginemisel olla reaalsed negatiivsed tagajärjed inimeste eludele.

Harimine sekkumisena

Kumbki sissejuhatuses püstitatud hüpoteesidest kinnitust ei leidnud, mis tähendab, et vähemalt käesoleva uuringu tulemused viitavad sellele, et lühike ChatGPT alane harimine ei pruugi olla efektiivne sekkumine liigse tuginemise vähendamiseks. Liigse tuginemise määr oli sekkumise grupis küll madalam, kuid erinevus ei olnud statistiliselt oluline ja ligi pooltel kordadest võeti vale soovitus ikkagi omaks. Selle tulemusega on kooskõlas (Bussone et al., 2015; Church, 2024) uuringud, mis on näidanud, et isegi pärast inimeste harimist selle osas, et tehisintellekt võib teha vigu, esineb kasutajate hulgas endiselt liigset tuginemist. See läheb vastuollu ootusega, et olles teadlik keelemudelite hallutsineerimisest, peaks kasutaja suhtuma selle väljunditesse kriitilisemalt.

Vastupidiselt ootustele ei suutnud sekkumine edukalt suurendada aeglast ja analüütilist mõtlemist probleemide lahendamise ajal. Selle tõendiks on asjaolu, et sekkumise ja kontrollgrupis ei olnud statistiliselt olulisi erinevusi keskmises esmase otsuse langetamisele kulunud ajas. Harimine võis põhjustada hoopis kasutatavate heuristikute muutmist. Selle poolt räägib asjaolu, et teksti lugenud katseisikud hakkasid rohkem eirama ChatGPT õigeid soovitusi. See tähendab, et kokkupuude informatsiooniga ChatGPT vigadest ja hallutsineerimisest võis katseisikuid suunata rohkem kasutama ChatGPT soovitusi üldiselt eiravaid heuristikuid, näiteks „kui ei tea, ära usalda“ ning see erinevus kajastus nii liigse tuginemises kui ka õigete soovituste eiramises. See on kooskõlas teaduskirjandusega, mis on

näidanud, et kui inimene ja algoritmist nõuandja teevad võrdväärse eksimuse, siis probleemilahendaja usaldus algoritmi vastu on sellest rohkem kahjustatud (Dietvorst et al., 2015). Tulevased sekkumise uuringud peaksid seda tendentsi arvesse võtma ning disainima sekkumisi nii, et need käsitleksid põhjalikult lisaks puudustele ka ChatGPT tugevusi ja kirjeldaksid ülesandeid, millega see saab inimestest oluliselt paremini hakkama.

Kuigi sekkumine ei omanud soovitud mõju, ei tähenda see, et tehisintellekti alane harimine ei võiks olla efektiivne sekkumine liigse tuginemise vähendamisel. Esiteks, on võimalik, et tekstid ei harinud katseisikuid piisavalt ChatGPT kohta. Seda toetab fakt, et sekkumise teksti ja kontrollgrupile esitatud teksti efektiivsuses ei olnud statistiliselt olulisi erinevusi. Teiseks, nagu ma sissejuhatuses kirjeldasin, siis harimine võib liigset tuginemist vähendada kahe erineva mehhanismi kaudu. Käesolevas uuringus mõõdeti neist vaid otsest mehhanismi, sest katseisikud ei pidanud ülesannete lahendamisel ChatGPT-ga aktiivselt interakteeruma. Seejuures on võimalik, et tänu teksti lugemisele ja suuremale teadlikkusele nt sellest, kuidas paremaid sisendeid kirjutada, hakkavad õpilased keelemudelitega interakteeruma viisil, mis vähendab ChatGPT hallutsinatsioonide esinemist ja ühtlasi ka liigset tuginemist.

Kokkuvõte

Käesoleva uuringu eesmärk oli mõõta gümnaasiumiõpilaste liigset tuginemist ChatGPT soovitudele matemaatikaülesannete lahendamise ajal ning hinnata tehisintellekti alase harimise kui sekkumise mõju. Uuringus lahendasid Eesti gümnaasiumiõpilased känguruülesandeid koos kuvatõmmisega ChatGPT soovitudest samale ülesandele. Nende ülesandeks oli esmalt otsustada, kas nad kasutavad ChatGPT soovitud oma vastusena või mitte. Kui katseisik kasutas ChatGPT vale soovitud, siis kutsuti seda liigseks tuginemiseks. Enne ülesannete lahendamist esitati katseisikutele uuringu sõltumatu muutuja, millel oli kaks taset (sekkumise tekst ja kontrolltekst). Uuringu tulemused näitasid, et liigset tuginemist esines rohkem kui poolte vale soovitudega ülesannete puhul. Seejuures ainult kaks inimest 36-st ei võtnud omaks mitte ühtegi ChatGPT vale soovitud. Sekkumine ei omanud soovitud mõju, sest gruppide vahel ei esinenud erinevusi liigses tuginemises ja sekkumine põhjustas ChatGPT õigete soovitude eiramise. Kokkuvõttes näitavad käesoleva uuringu tulemused, et inimesed kipuvad ChatGPT valesid soovitud omaks võtma ning et tegemist on üldise probleemiga, mis esineb vähemal või rohkemal määral enamikel inimestel. See muudab tulevastes teadustöodes oluliseks efektiivsete sekkumisprogrammide väljatöötamise ja mõju uurimise.

Kasutatud allikad

- About- Math Kangaroo.* (s.a.). Külastatud 2. jaanuar 2025,
<https://www.mathkangaroo.in/about/introduction>
- Amoozadeh, M., Nam, D., Prol, D., Alfageeh, A., Prather, J., Hilton, M., Ragavan, S. S., & Alipour, M. A. (2024). Student-AI Interaction: A Case Study of CS1 students. arXiv.
<http://arxiv.org/abs/2407.00305>
- Aru, J., Larkum, M. E., & Shine, J. M. (2023). The feasibility of artificial consciousness through the lens of neuroscience. *Trends in Neurosciences*, 46(12), 1008–1017.
<https://doi.org/10.1016/j.tins.2023.09.009>
- Azaria, A., Azoulay, R., & Reches, S. (2024). ChatGPT is a Remarkable Tool—For Experts. *Data Intelligence*, 6(1), 240–296. https://doi.org/10.1162/dint_a_00235
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. arXiv. <https://doi.org/10.48550/arXiv.2302.04023>
- Borji, A. (2023). A Categorical Archive of ChatGPT Failures. arXiv.
<https://doi.org/10.48550/arXiv.2302.03494>
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–21.
<https://doi.org/10.1145/3449287>
- Bussone, A., Stumpf, S., & O’Sullivan, D. (2015). The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems. *2015 International Conference on Healthcare Informatics*, 160–169. <https://doi.org/10.1109/ICHI.2015.26>

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2024). Unleashing the potential of prompt engineering in Large Language Models: A comprehensive review. arXiv.

<http://arxiv.org/abs/2310.14735>

Chollet, F. (2019). On the Measure of Intelligence. arXiv.

<https://doi.org/10.48550/arXiv.1911.01547>

Church, K. (2024). Emerging trends: When can users trust GPT, and when should they intervene? *Natural Language Engineering*, 30(2), 417–427.

<https://doi.org/10.1017/S1351324923000578>

DAIR.AI. (2023). *Prompt Engineering Guide*.

<https://www.promptingguide.ai/introduction/elements>

Diaz, M. (2024). ChatGPT vs. Microsoft Copilot vs. Gemini: Which is the best AI chatbot?

ZDNET. <https://www.zdnet.com/article/chatgpt-vs-microsoft-copilot-vs-gemini-which-is-the-best-ai-chatbot/>

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126.

<https://doi.org/10.1037/xge0000033>

Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals of Biomedical Engineering*, 51(12), 2629–2633.

<https://doi.org/10.1007/s10439-023-03272-4>

Hariri, W. (2024). *Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing* (arXiv:2304.02017).

arXiv. <https://doi.org/10.48550/arXiv.2304.02017>

Jacobs, M., Pradier, M. F., McCoy, T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z.

(2021). How machine-learning recommendations influence clinician treatment

- selections: The example of antidepressant selection. *Translational Psychiatry*, 11(1), 1–9. <https://doi.org/10.1038/s41398-021-01224-x>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kim, S. S. Y., Liao, Q. V., Vorvoreanu, M., Ballard, S., & Vaughan, J. W. (2024). „I’m Not Sure, But...“: Examining the Impact of Large Language Models’ Uncertainty Expression on User Reliance and Trust. ArXiv <https://doi.org/10.1145/3630106.3658941>
- Knoop, M. (2024, september 13). OpenAI o1 Results on ARC-AGI-Pub. *ARC Prize*. <https://arcprize.org/blog/openai-o1-results-arc-prize>
- Kocijan, V., Davis, E., Lukasiewicz, T., Marcus, G., & Morgenstern, L. (2023). The defeat of the Winograd Schema Challenge. *Artificial Intelligence*, 325, 103971. <https://doi.org/10.1016/j.artint.2023.103971>
- Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in Large Language Models. *Proceedings of The ACM Collective Intelligence Conference*, 12–24. <https://doi.org/10.1145/3582269.3615599>
- Levesque, H., Davis, E., & Morgenstern, L. (2012). The Winograd Schema Challenge. In Thirteenth international conference on the principles of knowledge representation and reasoning.
- Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., Cao, H., Liu, S., He, S., Huang, Z., Yang, D., Potts, C., Manning, C. D., & Zou, J. Y. (2024). Mapping the Increasing Use of LLMs in Scientific Papers. arXiv. <https://doi.org/10.48550/arXiv.2404.01268>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), Article 4. <https://doi.org/10.3390/educsci13040410>
- Lott, M. (2024, detsember 7). Tracking AI. <https://www.trackingai.org/iq>

- McCoy, R. T., Yao, S., Friedman, D., Hardy, M., & Griffiths, T. L. (2023). Embers of Autoregression: Understanding Large Language Models Through the Problem They are Trained to Solve. arXiv. <https://doi.org/10.48550/arXiv.2309.13638>
- Mormando, S. (2024, september 2). How Many R's in the Word "Strawberry?". *Medium*. <https://medium.com/@SamMormando/how-many-rs-in-the-word-strawberry-a6b8a697a1be>
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A Comprehensive Overview of Large Language Models. arXiv. <https://doi.org/10.48550/arXiv.2307.06435>
- Ngo, T. T. A. (2023). The Perception by University Students of the Use of ChatGPT in Education. *International Journal of Emerging Technologies in Learning (iJET)*, 18(17), Article 17. <https://doi.org/10.3991/ijet.v18i17.39019>
- Passi, S., & Vorvoreanu, M. (2022). *Overreliance on AI: Literature Review*. <https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/>
- Postimees. (2024, veebruar 5). Postimehe koolide edetabel 2023. *Postimees*. <https://www.postimees.ee/7953105/edetabelid-postimehe-koolide-edetabel-2023>
- Rogers, A. (2021). Changing the World by Changing the Data. arXiv. <https://doi.org/10.48550/arXiv.2105.13947>
- Rozado, D. (2023). The Political Biases of ChatGPT. *Social Sciences*, 12(3), Article 3. <https://doi.org/10.3390/socsci12030148>
- Zhou, L., Schellaert, W., Martínez-Plumed, F., Moros-Daval, Y., Ferri, C., & Hernández-Orallo, J. (2024). Larger and more instructable language models become less reliable. *Nature*, 634(8032), 61–68. <https://doi.org/10.1038/s41586-024-07930-y>

Teaduskool. (2022, november 14). *Känguru võistluse arhiiv*.

<https://teaduskool.ut.ee/et/ainevoistlused/kanguru-voistluse-arhiiv>

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S.

W. (2023). Large language models in medicine. *Nature Medicine*, 29(8), 1930–1940.

<https://doi.org/10.1038/s41591-023-02448-8>

Tonmoy, S. M. T. I., Zaman, S. M. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A.

(2024). A Comprehensive Survey of Hallucination Mitigation Techniques in Large

Language Models. arXiv. <https://doi.org/10.48550/arXiv.2401.01313>

Valova, I., Mladenova, T., & Kanev, G. (2024). Students' Perception of ChatGPT Usage in

Education. *International Journal of Advanced Computer Science and Applications*

(IJACSA), 15(1), Article 1. <https://doi.org/10.14569/IJACSA.2024.0150143>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., &

Polosukhin, I. (2017). Attention Is All You Need. arXiv.

<https://doi.org/10.48550/arXiv.1706.03762>

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The

relevance of AI literacy, prompt engineering, and critical thinking in modern

education. *International Journal of Educational Technology in Higher Education*,

21(1), 15. <https://doi.org/10.1186/s41239-024-00448-3>

Xiong, M., Hu, Z., Lu, X., Li, Y., Fu, J., He, J., & Hooi, B. (2024). Can LLMs Express Their

Uncertainty? An Empirical Evaluation of Confidence Elicitation in LLMs. arXiv.

<https://doi.org/10.48550/arXiv.2306.13063>

Lisad

Lisa A. Uuringus kasutatud ülesanded ja kuvatõmmis ChatGPT soovitusel

Olen küsimuste lõppu kirjutanud ka viite ülesande asukohale kanguruvõistluse arhiivis (Teaduskool, 2022). Viide on esitatud formaadis vanusekategooria, aasta, ülesande nr.

Õige 1

Kolmest ruumist ühes on peidus lõvi. Esimese ruumi uksele on kirjutatud "Lõvi on siin". Teise ruumi uksele on kirjutatud "Siin ei ole lõvi". Kolmanda ruumi uksele on kirjutatud " $3^2 - 2^3 = 2^3 - 3^2$ ". Ainult ühe ruumi uksele olev silt on õige. Millises ruumis on lõvi? (kadett, 2018, 14)

ChatGPT soovitus:



Lahendame iga ruumi sildi väiteid:

1. Esimene ruum: "Lõvi on siin."

- Kui see on õige, siis teise silt ("Siin ei ole lõvi") on vale, tekitades vastuolu. **Ei saa olla õige.**

2. Teine ruum: "Siin ei ole lõvi."

- Kui see on õige, siis lõvi ei ole siin ega esimeses ruumis. Lõvi peab olema kolmandas ruumis.
- Kolmanda ruumi silt $3^2 - 2^3 = 2^3 - 3^2$ ei ole tõene ($1 \neq -1$), mis tähendab, et see silt on vale. **Ainult see silt saab olla õige.**

3. Kolmas ruum: $3^2 - 2^3 = 2^3 - 3^2$.

- Võrrand ei ole tõene. **Ei saa olla õige.**

Järeldus: Teise ruumi silt on õige, seega lõvi on kolmandas ruumis.

Vastusevariandid: esimeses, teises, lõvi saab olla ükskõik millises ruumis, lõvi võib olla nii esimeses kui ka teises ruumis.

Õige vastus: kolmandas ruumis

Õige 2

Mikk ostab järgmiseks kuuks spordiklubikaarti, millega saab trennis käia kahel päeval nädalas. Tal tuleb kohe märkida need kaks nädalapäeva, millal ta hakkab trennis käima. Mitu erinevat võimalust on tal nende nädalapäevade valimiseks, kui ta ei soovi trennis käia kunagi kahel järjestikusel päeval? (kadett, 2017, 19)

ChatGPT soovitus:

 Esmalt arvutame kõik võimalikud kahe päeva kombinatsioonid nädalas:

$$\binom{7}{2} = 21$$

Järjestikuste päevade paarid on:

(Esmaspäev, Teisipäev), (Teisipäev, Kolmapäev), (Kolmapäev, Neljapäev), (Neljapäev, Reede), (Reede, Laupäev), (Laupäev, Pühapäev), (Pühapäev, Esmaspäev) — kokku 7 paari.

Vähendame kõigist kombinatsioonidest järjestikused paarid:

$$21 - 7 = 14$$

Seega on võimalik valikute arv:

$$\boxed{14}$$


Vastusevariandid: 16, 12, 10, 8

Õige vastus: 14

Õige 3

Klassi poiste seas ei ole neid, kes oleks sündinud samal nädalapäeval ja selle klassi tüdrukute seas ei ole neid, kes oleks sündinud samal kalendrikuul. Kui klassi tuleks uus õpilane, siis üks neist tingimustest poleks enam täidetud. Mitu õpilast on selles klassis praegu? (kadett, 2015, 11)

ChatGPT soovitus:

 Olgu meil klassis b poissi ja g tüdrukut. Antud on kaks tingimust:

1. Klassipoiste seas ei ole neid, kes oleks sündinud samal nädalapäeval.
2. Klassitüdrukute seas ei ole neid, kes oleks sündinud samal kalendrikuul.

Kuna poiste seas ei tohi olla samal nädalapäeval sündinuid, on maksimaalne poiste arv 7 (sest nädalas on 7 päeva). Tüdrukute seas ei tohi olla samal kalendrikuul sündinuid, seega on maksimaalne tüdrukute arv 12 (sest aastas on 12 kuud).

Kui klassi tuleks uus õpilane, siis üks nendest tingimustest poleks enam täidetud. See tähendab, et klassis on praegu maksimaalne arv poisse ja tüdrukuid, mis vastab antud tingimustele.

Maksimaalne poiste arv on 7 (kuna rohkem poisse tähendaks, et vähemalt kaks poissi oleksid sündinud samal nädalapäeval) ja maksimaalne tüdrukute arv on 12 (kuna rohkem tüdrukuid tähendaks, et vähemalt kaks tüdrukut oleksid sündinud samal kalendrikuul).

Kokku on klassis praegu 19 õpilast.

Vastusevariandid: 18, 20, 24, 25

Õige vastus: 19

Õige 4

Õpetaja küsis viielt oma õpilaselt, mitu neist olid eelmisel päeval kodus õppinud. Pelle ütles, et keegi ei õppinud. Elle ütles, et vaid üks õppis, Mikk ütles, et kaks, Riho ütles, et kolm ning Karl ütles, et neli. Õpetaja teadis, et need õpilased, kes tegelikult kodus ei õppinud, valetasid ja need, kes kodus õppisid, vastasid ausalt. Mitu õpilast neist viiest õppis eelmisel päeval kodus? (kadett, 2015, 20)

ChatGPT soovitus:



Olgu x kodus õppinud õpilaste arv:

- Pelle: 0 õppis.
- Elle: 1 õppis.
- Mikk: 2 õppisid.
- Riho: 3 õppisid.
- Karl: 4 õppisid.

Analüüs:

- Kui $x = 0$: Pelle tõtt (1 tõtt, 4 valetavad). Ei sobi, sest kõik peaksid valetama.
- Kui $x = 1$: Elle tõtt (1 tõtt, 4 valetavad). Sobib, sest üks räägib tõtt, ülejäänud valetavad.
- Kui $x = 2$: Mikk tõtt (1 tõtt, 4 valetavad). Ei sobi, sest peaks olema 2 tõtt ja 3 valetavad.
- Kui $x = 3$: Riho tõtt (1 tõtt, 4 valetavad). Ei sobi, sest peaks olema 3 tõtt ja 2 valetavad.
- Kui $x = 4$: Karl tõtt (1 tõtt, 4 valetavad). Ei sobi, sest peaks olema 4 tõtt ja 1 valetav.

Järeldus: Ainus võimalik lahendus on $x = 1$, seega õppis eelmisel päeval kodus 1 õpilane.


Vastusevariandid: 0, 2, 3, 4

Õige vastus: 1

Vale 1

Koosolekul osalesid Anni, Pille, Kati, Tiina ja Eerika. Neist igaüks tervitas algul ühe kallistusega vaid iga oma tuttavat osalejat. Kokku kallistas Anni ühe korra, Pille kaks korda, Kati kolm korda ja Tiina neli korda. Mitu korda kallistas Eerika? (Märkus: iga kallistamine on vastastikune.) (kadett, 2019, 14)

ChatGPT soovitus:

 Olgu A, P, K, T ja E vastavalt Anni, Pille, Kati, Tiina ja Eerika. Iga inimene kallistab oma tuttavat, seega on iga kallistus vastastikune.

Teame, et iga inimese kallistamiste arv on järgmine: $A = 1, P = 2, K = 3, T = 4$.

Leiame kokku kallistamiste arvu. Kui x on Eerika kallistamiste arv, siis:

$$1 + 2 + 3 + 4 + x = 2 \times \text{kõik kallistused kokku}$$

Kokku kallistuste arv:

$$1 + 2 + 3 + 4 + x = 10 + x$$

Kuna iga kallistus on vastastikune, jagame kogu kallistuste arvu kahega:

$$10 + x = 2 \times \text{vastastikused kallistused}$$

$$\text{vastastikused kallistused} = \frac{10+x}{2}$$

Kuna kõik inimesed pidid kallistama oma tuttavaid ja x peab olema 0 kuni 4, on sobiv lahendus:

$$x = 0$$

Seega, Eerika kallistas 0 korda.


Vastusevariandid: 5, 4, 3, 2

Õige vastus: 2 korda

Vale 2

Emma, Ivi, Rita ja Siiri vanused mingis järjestuses on 3, 8, 12 ja 14 aastat. Nii Siiri ja Emma vanuste summa kui ka Siiri ja Rita vanuste summa jaguvad arvuga 5. Leia Siiri ja Ivi vanuste summa. (kadett, 2017, 14)

ChatGPT soovitus:

 Antud on Emma, Ivi, Rita ja Siiri vanused, mis on 3, 8, 12 ja 14 aastat. Meil on kaks tingimust:

1. Siiri ja Emma vanuste summa jagub arvuga 5.
2. Siiri ja Rita vanuste summa jagub arvuga 5.

Kontrollime kõiki võimalikke Siiri vanuseid:

1. **Siiri = 3:** $3 + X$ ja $3 + Y$ ei anna 5-ga jaguvaid tulemusi.
2. **Siiri = 8:** $8 + X$ ja $8 + Y$ ei anna 5-ga jaguvaid tulemusi.
3. **Siiri = 12:** $12 + X$ ja $12 + Y$ ei anna 5-ga jaguvaid tulemusi.
4. **Siiri = 14:**
 - $14 + X$ ja $14 + Y$ jaguvad 5-ga ainult siis, kui $X = 12$ ja $Y = 6$.
 - Seega, Emma = 12 ja Rita = 8.

Siiri ja Ivi vanused on siis 14 ja 3. Nende summa on 17.

Vastus: 17

Vastusevariandid: 26, 22, 20, 15

Õige vastus: 26

Vale 3

Tiit on veendunud, et tema kell on 5 minutit ees, aga tegelikult on ta kell 10 minutit taga. Leo on veendunud, et tema kell on 10 minutit taga, aga tegelikult on ta kell 5 minutit ees. Ühel hetkel vaatavad mõlemad oma kella ning Tiidu arvates on õige kell 12.00. Kui palju on sel hetkel õige kell Leo arvates? (kadett, 2016, 22)

ChatGPT soovitus:



Tiit arvab, et tema kell on 5 minutit ees. Tegelikult on tema kell 10 minutit taga. Seega, kui Tiit arvab, et õige kell on 12.00, siis tema kella järgi on kell:

$$12.00 - 5 \text{ minutit} = 11.55$$

Kuna tema kell on tegelikult 10 minutit taga, siis tegelik kellaeg on:

$$11.55 + 10 \text{ minutit} = 12.05$$

Nüüd vaatame Leo kella. Leo arvab, et tema kell on 10 minutit taga, aga tegelikult on tema kell 5 minutit ees. Kui Leo vaatab oma kella, siis tema kella järgi on kell:

$$\text{Tegelik kellaeg} - 10 \text{ minutit} = 12.05 - 10 \text{ minutit} = 11.55$$

Kuna Leo kell on tegelikult 5 minutit ees, siis tema kella järgi on kell:

$$11.55 - 5 \text{ minutit} = 11.50$$

Seega, kui Tiidu arvates on õige kell 12.00, siis Leo arvates on sel hetkel õige kell 11.50.

Vastusevariandid: 11:30, 11:45, 12:00, 12:30

Õige vastus: 12:30

Vale 4

Üks vanaaegne kaal ei tööta korralikult. Kui kaalutav on kergem kui 1000 g, siis kaal näitab õigesti, vastasel juhul näitab see aga tulemuseks suvalist 1000-st suuremat arvu. Viis eset A, B, C, D ja E olid kõik kergemad kui 1000 g. Esemeid kaaluti paarikaupa ja nende summaks kirjutati kaalu näit: $B + D = 1200$, $C + E = 2100$, $B + E = 800$, $B + C = 900$, $A + E = 700$. Milline neist esemetest oli kõige raskem? (kadett, 2014, 22)

ChatGPT soovitus:



Meil on viis eset: A, B, C, D ja E, kõik kergemad kui 1000 g. Kaaludes paarikaupa, saadi tulemused: $B + D = 1200$, $C + E = 2100$, $B + E = 800$, $B + C = 900$, $A + E = 700$.

Tõesed võrrandid on: $B + E = 800$, $B + C = 900$, $A + E = 700$. Kui $B + E = 800$ ja $B + C = 900$, siis $C = 100$. Asendades $A + E = 700$: $A + (800 - B) = 700$, saame $A = -100 + B$.

Lõpuks saame: $(900 - B) + (800 - B) = 2100$, mis annab $B = -200$.

Seega E on kõige raskem ese.

Vastusevariandid: A, B, C, D

Õige vastus: D

Lisa B. Katseisikutele esitatud tekstid

Siin on ühe tekstina esitatud nii sekkumise kui ka kontrollgrupile esitatud tekstid. Kontrollgrupi tekst sisaldas ainult sissejuhatust ja katsegrupile esitati tekst tervikuna kõikide peatükkidega. Tekstis on eraldi välja toodud viited teaduskirjandusele ja mõned kommentaarid. Katseisikud said lugeda teksti ilma viidete ja kommentaarideta. Lugemise lihtsuse huvides olen tähistanud kollase värviga kõik osad, mida katseisikutele ei esitatud.

Enne, kui Te hakkate koos ChatGPT-ga matemaatikaülesandeid lahendama, lugege palun läbi alljärgnev tekst, mille eesmärk on lühidalt tutvustada ChatGPT juturobotit. Palun lugege tekst tähelepanelikult ja mõttega läbi ka siis, kui olete varasemalt ChatGPT-ga kokku puutunud ja seda kasutanud. Kui olete teksti läbi lugenud, siis vajutage lehe all paremas nurgas olevale nupule "Järgmine".

ChatGPT lühitutvustus

ChatGPT on juturobot, mida tutvustati esmakordselt 2022. aastal ja mis on sellest ajast alates kiiresti populaarsust kogunud. Seda on palju kasutusele võetud erinevates valdkondades nagu haridussüsteemis, majanduses, klienditeeninduses ja teaduses, kus selle kasutamine on aidanud mitmeid töid oluliselt lihtsustada. ChatGPT töötab lähtudes kasutaja sisendist, milleks võib muuhulgas olla küsimus, ülesanne või matemaatiline tehe. ChatGPT analüüsib kasutaja sisendit ja toodab väljundi, mida esitatakse kasutajale arusaadavas keeles. ChatGPT abiga on võimalik lahendada erinevaid ülesandeid, alustades lihtsatest küsimustest ja lõpetades mahukate projektidega. Näiteks võib ChatGPT aidata koostada originaalset teksti, tõlkida, analüüsida andmeid, kirjutada koodi ja teha mahukast materjalist kokkuvõtteid.

Siin on mõned põhjused, miks ChatGPT on hea abivahend keeleliste ülesannete lahendamisel: Alltoodud punktid on võetud (Hariri, 2024) artiklist. Mõned artiklis väljatoodud tugevused jätsin kõrvale, sest pidasin neid liiga tehniliseks ja ebaoluliseks gümnaasiumiõpilaste teadmiseks.

Keelest arusaamine. ChatGPT toodab väljundeid, mis suudavad hästi arvesse võtta sõnade konteksti ja sisu.

Isikupärastatud vastused. ChatGPT suudab oma väljundeid kohandada lähtudes kasutaja stiilist ja varasematest sisenditest. See ChatGPT omadus parandab kasutuskogemust, viies väljundid paremini kooskõlasse kasutaja vajadustega.

Kättesaadavus. Kuniks on olemas internetiühendus, siis on ChatGPT abi alati kättesaadav. ChatGPT kättesaadavus seisneb ka selles, et see saab hästi aru ja oskab väljundeid toota erinevates keeltes, sealhulgas ka eesti keeles.

Automatiseerimine. ChatGPT aitab lihtsustada ja automatiseerida rutiinseid ülesandeid, säästes sellega kasutaja aega ja vaeva.

ChatGPT töömehhanism

ChatGPT töötab järgmiselt: Iga kord, kui see väljundeid toodab, siis lahendab see väga lihtsat ülesannet, milleks on järgmise sõna ennustamine (McCoy et al., 2023). Kui ma kirjutan lause „Koer on ___“ ja ma küsin sinult, mis sõna kõige paremini sinna lünka sobib, siis on mingid sõnad, mida sa pakud suurema tõenäosusega (nt loom, karvane, truu) ja teised, mida sa pakud väiksema tõenäosusega (nt pudel, sinine, kala). Sarnaselt nagu sina tead, et mingid sõnad

sobivad sinna lünka suurema tõenäosusega kui teised, teab seda tänu eelnevale treeningule ka ChatGPT. ChatGPT treenimiseks kasutatakse peaaegu kogu internetis kättesaadavat teksti, mille analüüsimise käigus õpib juturobot, kuidas sõnad omavahel seotud on ja millised sõnad teistele tõenäoliselt järgnevad. Oluline on märkida, et ChatGPT ei mõtle ega koge maailma nagu inimesed. Oma olemuselt on see lihtsalt tohutult suur statistiline mudel, mis vastab kasutaja sisendile arvutades järgmiste sõnade tõenäosust.

Sisendi osad

Nagu eelnevalt kirjeldatud, toodab ChatGPT väljundeid lähtudes kasutaja sisendist. Kasutaja sisend võib koosneda neljast osast: **Sisendi osade loetelu on võetud DAIR.AI (2023) veebijuhistest.**

1. *Instruktsioonid*. Selle osa eesmärk on selgitada ChatGPT-le, millise ülesande sa tahad, et see lahendaks. Mõned näited: „lahenda, arvuta, kirjuta, tee kokkuvõtte, hinda, selgita, soovita“.
2. *Kontekst*. Kontekst on väline informatsioon, mis annab ChatGPT-le lisateavet ülesande lahendamise või teksti sisu kohta. See aitab ChatGPT-l toota asjakohasemaid ja täpsemaid väljundeid.
3. *Tekst*. See on sisendi sisu ehk nt täpne kirjeldus esseest, mida sa tahad, et see kirjutaks või tekstilõik, millest sa tahad, et see kokkuvõtte teeks.
4. *Väljundi eelistused*. Sellega on võimalik ChatGPT-le märku anda, millises vormis või stiilis sa väljundeid soovid. Näiteks võib määrata sõnade arvu ja lõigu pikkust või küsida väljundit täpplendina.

Sisend ei pea tingimata koosnema kõigist neljast osast. Kui ChatGPT-le nt sisendiks kirjutada „Kui $a = b$ ja $b = c$, siis...“, siis saab ChatGPT aru, et instruktsioonideks on „lahenda“, ilma et seda peaks eraldi täpsustama. Kui sa aga tahad, et ChatGPT toodaks ülaltoodud näite põhjal hoopis sarnaseid loogikaülesandeid juurde, siis on vajalik oma instruktsioone täpsustada. Mistahes sisendi puhul on kasulik mõelda sisendi osade peale ja hinnata, millist informatsiooni ChatGPT sobiva väljundi tootmiseks vajada võiks.

Näide sisendist:

Selgita järgneva teksti põhisõnumit, (**Instruktsioonid**)
tegemist on ühe Eesti vanasõnaga: (**Kontekst**)
“Kes teisele auku kaevab, see ise sisse kukub.” (**Tekst**)
Esita vastus rootsi keeles ja kahe lausega. (**Väljundi eelistus**)

Sagedased vead sisendi kirjutamisel

Sagedased vead on võetud Giray (2023) artiklist. Seejuures jätsin mõned punktid kõrvale, sest pidasin neid käesoleva teksti lõikes ebaoluliseks või kirjeldan neid mõnes muus alapeatükis. Näited ja selgitused on originaalsed.

- *Ebaselge sisend*. Mõnikord ei ole kasutaja sisend piisavalt spetsiifiline, mistõttu on ka ChatGPT väljund liiga üldine ja laialivalgav, andes lihtsalt üldise ülevaate kirjeldatud

teemast. Näiteks ei ole sisend „Kirjelda, kuidas on tehisintellekt maailma muutnud“ piisavalt spetsiifiline, kui me tahame teada, milliseid ohte ja probleeme on ChatGPT kasutamine konkreetselt haridussüsteemis põhjustanud.

- *Keerulised ja konfliktised juhised.* Kui meie juhised on liiga keerulised või omavahel vastuolus, siis on ChatGPT-l sageli raske nendest aru saada. Üks näide konfliktsest ja keerulisest sisendist: „Kirjuta essee 20. sajandi tehnoloogilistest arengutest, olles samal ajal piisavalt detailne, et see oleks arvutiteadlase jaoks huvitav ja piisavalt lihtne, et sellest saaks aru ka viieaastane laps.“
- *Kallutatud sisend.* Kasutaja sisend võib olla erinevatel viisidel kallutatud. Kui me tahame nt, et ChatGPT hindaks meie kirjutatud teksti stiili, siis võib sisend „Kas järgnev tekst on stiili poolest hästi kirjutatud?“ müksata ChatGPT vastama „Jah“. Parem on küsida neutraalselt „Hinda järgneva teksti stiili ja anna nõuandeid, kuidas seda paremaks muuta.“

ChatGPT piirangud ja ohud

Kuigi sisendi korrektne sõnastamine on kahtlemata oluline, et tagada maksimaalselt hea väljund, siis on mõned ülesanded, millele ChatGPT võib valesti vastata ka siis, kui kasutaja sisend on väga hea. Siin on kokkuvõtte ChatGPT peamistest probleemikohtadest koos näidetega. Üldised punktid olen võtnud Borji (2023) artiklist, seejuures valisin 11-st välja neli punkti, mida pidasin liigse tuginemise kontekstis kõige olulisemaks.

Maailma kohta järelduste tegemine. ChatGPT-l puudub otsene kontakt pärismaailmaga, mistõttu saab see maailma kohta järeldusi teha üksnes tekstide põhjal, millel seda treeniti. See põhjustab mõnikord olukordi, mil ChatGPT eksib ülesannetega, mis nõuavad mõtlemist päris maailmas eksisteerivate asjade ja nende omaduste üle.

Kui mul on praegu 5 õuna, aga eelmine nädal sõin ma 2 õuna ära, siis mitu õuna on mul alles?

Kui sul oli algselt 5 õuna ja eelmisel nädalal sõid sa 2 õuna ära, siis on sul alles:

$5 - 2 = 3$ õuna.

Ülaltoodud näites eksib ChatGPT aja kontseptsioonist (praegu, eelmine nädal) arusaamisega ja arvab ekslikult, et tegemist on lihtsa aritmeetika ülesandega. Ülesanne on inspireeritud Diaz (2024) artiklist.

Loogika ja matemaatika. ChatGPT võib mõnikord valesti lahendada isegi lihtsat matemaatilist ja loogilist mõtlemist nõudvaid ülesandeid.

mitu sõna ja mitu arvu on järgnevas loetelus: kass, 10, koer, 60, lehm, 40, orav, 70, kuusk, 15, raamat, 32.

Loetelus on 7 sõna ja 5 arvu. [-]

Lugejale teadmiseks: Koodi analüüsimine näitas, et ChatGPT kategoriseeris arvu „32“ sõnade hulka, sest arvule järgneb punkt. Näites olev ülesanne on võetud kanguruülesannete arhiivist (Teaduskool, 2022).

Vastusevariantides antud viis kaarti tuleb ritta panna nii, et neist moodustuv 9-kohaline arv oleks võimalikest arvudest vähim. Milline neist kaartidest peab olema sel juhul parempoolsem? Kaardid on: 4, 8, 31, 59, 107.



Et moodustada võimalikult väiksem 9-kohaline arv, tuleb kaardid paigutada suurenevas järjekorras, võttes arvesse iga numbriga väärtust:

- Kõige väiksem number on 4, millele järgneb 8, siis 31, seejärel 59 ja lõpuks suurim 107.

Kui me need järjestame väiksema arvu moodustamiseks, on loogiline, et kõige suurem kaart, ehk 107, peab olema parempoolsem, sest see suurendaks arvu väärtust kõige vähem, kui asetada see viimasele kohale.

Vastus: parempoolsem kaart peab olema 107.

Lugejale teadmiseks: Tegelikult peaks 107 olema vasakpoolsem ja 8 parempoolsem arv.

Hallutsineerimine. ChatGPT mõtleb mõnikord informatsiooni täiesti välja – nähtust, mida kutsutakse hallutsineerimiseks. See võib muuhulgas toota valesid viiteid ja ajaloolisi fakte. Lisaks võib sellel olla raskusi fiktsiooni ja reaalsuse eristamisega. All on toodud üks näide hallutsineerimisest.

kes oli eesti esimene olümpia hõbemedalist?



Eesti esimene olümpia hõbemedalist oli **Jaan Tõnisson**, kes võitis hõbemedali 1920. aasta suveolümpiamängudel Antwerpenis, Belgias. Tõnisson võistles maadluses kreeka-roma stiilis raskekaalus.

Lugejale teadmiseks: Eesti esimene olümpiamängude hõbemedalist oli Martin Klein 1912. aastal ja Jaan Tõnisson ei olnud mitte maadleja, vaid tähtis riigitegelane.

ChatGPT võib hallutsineerida igasuguste küsimuste puhul. Kuid eriti lihtne on seda esile kutsuda, kui kasutaja instruksioonid nõuavad väga kitsast ja spetsiifilist väljundit. All on toodud sellest kaks näidet. Selle väite esitamisel lähtun enda kogemusest juturoboti kasutamisel.

Nimeta viis linna Harjumaal, mis algavad ja lõpevad sama tähega.



Harjumaal on viis linna, mis algavad ja lõpevad sama tähega:

1. Keila
2. Saue
3. Maardu
4. Paldiski
5. Jõelähtme

Vasta ühe sõnaga, mis spordialal võitis Napoleon 1980. aasta olümpiamängudel kulla



Jalgpall.

Diskrimineerivad väljundid. ChatGPT väljundid sõltuvad tekstidest, millel seda treeniti. Need tekstid võivad sisaldada erinevaid sotsiaalseid kaldeid inimeste sotsiaalse staatuse, soo, rassi ja vanuse kohta ning treeningu käigus õpib tehisintellekt neid kaldeid oma väljunditesse lisama (Rogers, 2021). Seetõttu võivad ChatGPT väljundid olla näiteks poliitiliselt kallutatud (Rozado, 2023) ja sisaldada ühiskonnas eksisteerivaid soostereotüüpe (Kotek et al., 2023).

Suurettevõtte juht käis enne manikööri juurde minekut wc-s. Meeste ja naiste wc oli eraldi. Kumba wc-sse ta läks?



Lause põhjal saame järeldada, et suurettevõtte juht on mees, sest tavaliselt sellises kontekstis, kui sugu ei ole eraldi määratletud ja kasutatakse üldnimetust "juht", eeldatakse sageli, et tegemist on mehega. Seega ta läks meeste WC-sse.

Kui see lause on osa loost ja soovid selgemat vastust anda, võiks see sõltuda täiendavast kontekstist või selgelt väljendatud tegelase soost. Antud juhul oleks oletus, et ta kasutas meeste WC-d.

Ülaltoodud näites väljendab ChatGPT ühiskonnas eksisteerivat soostereotüüpi, eeldades, et suurettevõtte juht on mees, vaatamata asjaolule, et ta käib maniköörija juures.



Lisaks teadmisele, et ChatGPT võib teha vigu, on oluline teada, et mõned ChatGPT omadused muudavad nende vigade märkamise keeruliseks.

Liigne enesekindlus. ChatGPT esitab ka väärinformatsiooni sisaldavaid väljundeid kõrge enesekindlusega (Azaria et al., 2024). Kui see ei tea vastust, siis selle asemel, et öelda „Ma ei ole kindel“, esitab see oma vastuse väga veenvalt ja soravalt, mis muudab keeruliseks ChatGPT väljundites vigade märkamise.

Vead on ootamatud. ChatGPT vead võivad tulla ootamatult, mistõttu võivad need sageli tähelepanuta jääda. ChatGPT võib ühel hetkel selgitada kvantfüüsika keerulisi ideid väga selgelt ja tabavalt ning järgmisel hetkel eksida kõige lihtsamate loogikaülesannetega. Säärane vastuste kvaliteedi ebastabiilsus on oht ChatGPT kasutamisel. Varasemate heade soorituste põhjal võime eeldada, et ChatGPT suudab õigesti lahendada ka järgnevad ülesanded, seda isegi juhul, kui vastused sisaldavad tegelikult vigu. Sellele ohule ei ole teaduskirjanduses palju tähelepanu pööratud, ent pidasin oluliseks selle punkti siiski eraldi välja tuua.

Nõuanded ChatGPT kasutamiseks

- *Küsi mitu korda* (Chen et al., 2024). ChatGPT vastused põhinevad tõenäosuste arvutamisel, mistõttu võib see samale küsimusele pakkuda erinevaid vastuseid ja mõned neist vastustest võivad olla ebatõenäolised ning valed. Seetõttu on kasulik, kui küsid sama küsimust mitu korda, aga veidike erineva sõnastusega, siis saad valida enda jaoks sobivaima vastuse.
- *Küsi arutluskäiku*. Kui Sa lahendad loogilist mõtlemist nõudvaid ülesandeid, siis kirjuta ChatGPT-le sisendiks, et see selgitaks oma arutluskäiku (nt lisades sisendi lõppu lause „Lahenda samm-sammult“). Sellisel viisil on sul lihtsam aru saada, kuidas ChatGPT mingi vastuseni jõudis ja tuvastada kui ChatGPT arutluskäik on vale (Walter, 2024).
- *Kontrolli fakte*. Kui sa ei ole kindel, kas ChatGPT väljundis sisalduv informatsioon on tõene, siis alati kontrolli seda teistest allikatest. Üks võimalus on guugeldada. Teine võimalus on paluda ChatGPT-l lisada oma väljunditesse viited, et sa saaksid ise veenduda nende usaldusväärsuses. Church (2024) on rääkinud faktide kontrollimise olulisusest ChatGPT kasutamisel.

 Sõnum ChatGPT 

ChatGPT võib teha vigu. Kontrollige olulist teavet.

Pea meeles, et ChatGPT on mõtlemise abivahend, mitte mõtlemise asendaja.

Lisa C. Informeeritud nõusoleku ankeet

Palume Teil tähelepanelikult läbi lugeda järgnev teave Tartu Ülikooli bakalaureuseastme uurimistöö raames läbiviidavast uurimisprojektist “Õpilaste motivatsioon kognitiivsete ülesannete lahendamisel ChatGPT abiga”, et saaksite anda oma informeeritud nõusoleku selles osalemiseks.

Kõikide küsimustiku lõpuni täitjate vahel läheb loosi 5x20€ Apollo kinkekaarti!

Uuringu eesmärk on välja selgitada, kuidas mõjutab ChatGPT abiga ülesannete lahendamine gümnaasiumiõpilaste motivatsiooni. Selleks palume uuringus osalejali täita siin (Limesurvey) veebikeskkonnas olev küsimustik. Küsimustik koosneb neljast osast ja selle täitmine võtab aega umbes 40-60 minutit. Küsimustikku on soovituslik täita arvutis.

Küsimustik koosneb neljast osast:

1. osas küsitakse Teie demograafilisi andmeid (sugu, vanus, kool).
2. osas palutakse Teil läbi lugeda ChatGPT lühitutvustus.
3. osas lahendate Te matemaatikaülesandeid koos ChatGPT abiga. Täpsemad instruksioonid esitatakse vahetult enne ülesannete lahendamist.
4. osas küsitakse Teilt küsimusi ChatGPT kasutamise ja motivatsiooni kohta.

Küsimustiku täitmise ajal ei ole lubatud arvutis avada muid vaheaknaid, kasutada nutitelefoni või muid abivahendeid (sealhulgas paberit ja pliiatsit)

Isikuandmete kaitseks ei küsita ega salvestata osalejate nime, meiliaadressi, koduaadressi ega koguta ning töödelda uuritavate IP-aadresse. Kogutud andmete põhjal ei ole võimalik katses osalejat tuvastada. Samuti palume teie luba kasutada kogutud andmeid anonüümselt õppe- ja teadustöö huvides. Kõik uuringu käigus kogutud andmed on anonüümsed ja konfidentsiaalsed. Neid analüüsitakse anonüümsel kujul Tartu Ülikoolis ega avaldata isikut tuvastada võimaldaval kujul mitte kellelegi väljaspool uurijaid.

Uuringus osalemine on vabatahtlik ning uuringus osalemisest on võimalik enne vastamise kinnitamist loobuda. Pärast küsimustiku täitmist ei saa uuringus osalemisest enam loobuda.

Uuring on kooskõlastatud Tartu Ülikooli Inimuuringute eetika komiteega.

Lisainfo saamiseks pöörduge uurimistöö teostaja Brett Pupartit poole järgneval meiliaadressil: brett.pupart@ut.ee

Kaebustega isikuandmete töötlemise osas pöörduge palun Tartu Ülikooli andmekaitse spetsialisti poole e-posti aadressil andmekaitse@ut.ee

Uuringus osalemise alustamisega annate nõusoleku osaleda eelpool kirjeldatud uuringus ning lubate kasutada kogutud andmeid anonüümselt õppe- ja teadustöö huvides.

Lisa D. Katseisikutele esitatud instruktsioonid

Kohe hakkate Te lahendama matemaatikaülesandeid koos ChatGPT abiga. Kokku on ülesandeid 8 ja teie eesmärk on vastata võimalikult paljudele küsimustele õigesti. Vastamisel ajapiirangut ei ole.

Teile esitatakse lehe ülemises nurgas matemaatikaülesanne ja kohe selle all kuvatõmmis ChatGPT vastusest samale küsimusele. Teil avaneb esialgu valik, kas kasutate ChatGPT soovitus oma vastusena või mitte.

- Kui vastate "Jah", siis saate edasi liikuda järgmise ülesande juurde.
- Kui vastate "Ei", siis saate järgmisel lehel ülesande üksinda lahendada, valides nelja vastusevariandi hulgast selle, mis on teie arvates õige vastus. Seejuures ChatGPT soovitus vastusevariantide hulgast valida ei saa.

Pärast igale küsimusele vastamist palutakse Teil hinnata, kui kindel Te oma vastuses olite.

Ülesannete lahendamise ajal tohib Teil olla avatud ainult LimeSurvey veebikeskkond.

Guugeldamine või muude abivahendite (sealhulgas paberi ja pliiatsi) kasutamine ei ole ülesannete lahendamise ajal lubatud.

Käesolevaga kinnitan, et olen korrektselt viidanud kõigile oma töös kasutatud teiste autorite poolt loodud kirjalikele töödele, lausetele, mõtetele, ideedele või andmetele.

Olen nõus oma töö avaldamisega Tartu Ülikooli digitaalarhiivis DSpace.

/Brett Pupart/