

TARTU ÜLIKOOL  
Arvutiteaduse instituut  
Informaatika õppekava

**Joosep Lember**

**Kursuse “Andmebaasid” kodutööde  
lahendamine keelemudelitega  
võrdluses üliõpilaste tulemustega**

**Bakalaureusetöö (9 EAP)**

Juhendaja: Piret Luik, PhD

Tartu 2025

# **Kursuse “Andmebaasid” kodutööde lahendamine keelemudelitega võrdluses üliõpilaste tulemustega**

## **Lühikokkuvõte:**

Viimastel aastatel on tehisintellektil põhinevate keelemudelite, nagu ChatGPT ja Microsoft Copilot, kasutamine hariduses muutunud tavapäraseks. Samas pole selge, kui hästi suudavad need tööriistad keerukamaid ülesandeid iseseisvalt lahendada. Bakalaureusetöö eesmärk oli hinnata keelemudelite võimekust lahendada Tartu Ülikooli kursuse „Andmebaasid“ kodutöid ning võrrelda tulemusi 2025. aasta kevadsemestri üliõpilaste omadega. Kodutööd lahendati keelemudelitega kahel viisil: esmalt andmebaasi struktuuri eelinfota ja seejärel koos vastava teabega. Tulemusi hinnati automaatkontrolliga ning analüüsiti veatüüpe ja täpsust. Mudelid saavutasid paremaid tulemusi, kui neil oli juurdepääs andmebaasi struktuurile, kuid jäid üldiselt üliõpilastele alla, eriti keerukate ja mitmeti mõistetavate ülesannete puhul.

**Võtmesõnad:** tehisintellekt (TI), andmebaasid, SQL-keel, ChatGPT, Copilot

**CERCS:** P175 Informaatika, süsteemiteooria, S281 Arvuti õpiprogrammide kasutamise meetoodika ja pedagoogika

## **Solving the “Databases” SQL Assignments with Language Models in Comparison with Students’ Results**

### **Abstract:**

In recent years, the use of AI-based language models such as ChatGPT and Microsoft Copilot has become common in education. However, it remains unclear how well these tools can independently solve more complex tasks. The aim of this bachelor’s thesis was to evaluate the ability of these language models to solve homework assignments from the University of Tartu’s “Databases” course and compare the results with those of students from the 2025 spring semester. The assignments were completed using the language models in two ways: first without any prior information about the database structure, and then with the relevant structural information provided. The results were evaluated using automatic grading, and the types and accuracy of errors were analyzed. The models performed better when they had access to the database structure, but overall still fell short of student performance, especially in complex and ambiguous tasks.

**Keywords:** artificial intelligence (AI), SQL language, ChatGPT, Copilot

**CERCS:** P175 Informatics, systems theory, S281 Computer-assisted education

# Sisukord

<b>Sisukord.....</b>	<b>3</b>
<b>Sissejuhatus.....</b>	<b>4</b>
<b>1. Keelemudelid ja nende mõju haridusele.....</b>	<b>6</b>
1.1 Keelemudelid.....	6
1.2 Keelemudelite kasutamine hariduses.....	8
1.3 Keelemudelite mõju arvutiteaduse õppimisele.....	10
<b>2. Tartu Ülikooli kursus “Andmebaasid”.....</b>	<b>12</b>
2.1 Kursuse sisu.....	12
2.2 Kursusel kasutatavad programmid ja andmebaasi kirjeldus.....	12
2.3 Hinnatav osa.....	13
2.4 Kodutööde kontrollimine.....	14
<b>3. Metoodika.....</b>	<b>15</b>
3.1 Valitud keelemudelid.....	15
3.2 Mudelite kasutamine.....	15
3.3 Valimi kirjeldus ja tulemuste arvutamise metoodika.....	16
3.4 Kodutööd.....	17
<b>4. Tulemused.....</b>	<b>18</b>
4.1 Keelemudelite lahendused.....	18
4.1.1 Kodutöö 1.....	18
4.1.2 Kodutöö 2.....	19
4.1.3 Kodutöö 3.....	20
4.1.4 Kodutöö 4.....	21
4.1.5 Kodutöö 5.....	22
4.1.6 Kodutöö 6.....	24
4.1.7 Kodutööde tulemuste koondvaade.....	25
4.2 Keelemudelite tulemuste võrdlus üliõpilaste tulemustega.....	26
<b>5. Arutelu.....</b>	<b>28</b>
<b>Kokkuvõte.....</b>	<b>30</b>
<b>Viidatud kirjandus.....</b>	<b>31</b>
I. Ülesande näidisviip ChatGPT-s.....	35
II. Ülesande näidisviip Copilot-is.....	36
III. Eelinfo keelemudelile.....	37
IV. Keelemudeli antud lahendus.....	38
V. Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks.....	39

## Sissejuhatus

Keelemudelite areng viimastel aastatel on toonud kaasa uusi võimalusi nii teadusuuringutes kui ka hariduses. Tehisintellektil põhinevad süsteemid suudavad genereerida inimkeelset teksti ja lahendada mitmesuguseid ülesandeid, sealhulgas arvutiteaduse ja programmeerimise alaseid harjutusi õpingutel [1]. Samas on nende mudelite kasutamine hariduses tekitanud küsimusi seoses nende täpsuse, usaldusväärsuse ning mõjuga õppimisprotsessile [2]. Kuna keelemudelid võivad pakkuda kiireid ja struktureeritud vastuseid, on oluline hinnata, kas nende abil saadud tulemused on võrreldavad üliõpilaste omadega. Kuigi üliõpilased kasutavad keelemudeleid järjest enam õppimise toetamiseks, ei ole teada, kui hästi need mudelid suudavad iseseisvalt lahendada Tartu Ülikooli kursuse „Andmebaasid“ (LTAT.03.004) kodutöid. Selle kursuse eesmärk on anda ülevaade relatsiooniliste andmebaaside ja sellega kaasneva teooria kohta [3]. Keelemudeleid kasutatakse tihti abivahendina olukorras, kus õppija jääb hätta ja soovib saada vihjeid või soovitusi edasiliikumiseks, kuid kui mudel pakub ebatäpseid või eksitavaid lahendusi, võib see õppijat hoopis valele teele suunata.

Bakalaureusetöö eesmärk on selgitada keelemudelite suutlikkust andmebaaside aine koduülesannete lahendamisel ja võrrelda keelemudelite sooritust üliõpilaste tulemustega. Uuringu käigus lahendatakse andmebaaside aine ülesandeid, kasutades keelemudeleid ChatGPT ja Microsoft Copilot, ning võrreldakse nende tulemusi 2025 aasta üliõpilaste lahenduste ja tulemustega. Selles uurimuses teisi selle kursusega seotud hindelisi ülesandeid ei käsitleta. Vastavalt eesmärgile on püstitatud uurimisküsimused, millele töös keskendutakse:

1. Millised on keelemudelite poolt lahendatud "Andmebaasid" kursuse koduülesannete tulemused, kui kasutada hindamiseks automaatkontrolli skoori ning kuidas erinevad tulemused sõltuvalt sellest, kas andmebaas on ette antud SQL kujul või mitte?
2. Kui hästi suudavad keelemudelid lahendada andmebaasi koduülesandeid võrreldes üliõpilaste tulemustega?

Töö on jaotatud viieks peatükiks. Esimeses peatükis antakse ülevaade keelemudelite arengust ja nende kasutamisest hariduses, sealhulgas programmeerimise ja andmebaaside õpetamisel. Teises peatükis kirjeldatakse kursuse sisu, hindamiseks kasutatud tarkvara ning hinnatavat kursuse osa. Kolmandas peatükis kirjeldatakse töö metoodikat, sealhulgas viipade (*prompt*)

ülesehitust, keelemudelite kasutamist ning tulemuste hindamise kriteeriumeid. Neljandas peatükis esitatakse analüüsi tulemused ning võrreldakse keelemudelite lahendusi üliõpilaste omadega. Viiendas peatükis arutletakse saadud tulemuste üle vastates uurimisküsimustele.

# 1. Keelemudelid ja nende mõju haridusele

Peatükis antakse ülevaade keelemudelite arengust, nende tööpõhimõtetest ja kasutusvõimalustest hariduses. Keskendatakse sellele, kuidas keelemudelid võivad toetada õppimist arvutiteaduses. Samuti kirjeldatakse, milliseid võimalusi ja väljakutseid keelemudelid pakuvad.

## 1.1 Keelemudelid

Keelemudelid (*Large Language Models*, LLM) tähistavad olulist edasiminekut loomuliku keele töötlemises (*Natural language processing*, NLP), milles keskendatakse inimese ja arvuti vahelisele suhtlusele [2]. Need mudelid kasutavad süvaõppe meetodeid, et analüüsida ja mõista keeleandmetes peituvaid mustreid ja struktuure. Oluliseks omaduseks on võime töödelda suuri andmehulkasid, sealhulgas struktureerimata teksti, ning tabada semantilisi seoseid sõnade ja fraaside vahel [4]. Sellest tulenevalt suudavad keelemudelid genereerida vastuseid esitatud küsimustele, ülesannetele või lihtsalt kaasa arutleda esitatud teemadel. Ka hariduses on hakatud keelemudeleid rohkem kasutama ning keelemudelite ja tehisintellekti areng on muutnud tehisaru kasutamise õppetöös igapäevaseks tööriistaks [5]. Keelemudelid võivad olla tõhusad tööriistad teadmiste omandamisel, aidates õppijatel teooriat mõista ja praktilisi oskusi arendada - keelemudelite kasutamine hariduses on kiiresti arenev valdkond [3, 6, 8].

Keelemudelite areng sai alguse statistilistest keelemudelitest, mis kasutasid tõenäosuslikke lähenemisi, näiteks n-gramme, peidetud Markovi mudeleid (HMM) ja maksimaalse entroopia mudeleid, et ennustada järgmise sõna tõenäosust eelnevate sõnade põhjal [4]. Kuigi mudelid võimaldasid teksti loomist ja sõnade ennustamist, olid nende võimalused keeruliste kontekstipõhiste seoste tabamisel piiratud. Siiski väidavad Hadi jt [4], et tänapäevased keeletehnoloogiad, nagu LLM-id saavad keerulisemate ülesannetega, nagu programmeerimine, paremini hakkama, pakkudes oluliselt suuremat võimet mõista ja genereerida inimesele sarnast keelt. Keelemudelid on masinõppe ja loomuliku keele töötlemise kiire arengu tulemus, pakkudes murrangulisi lahendusi, mis suudavad luua inimesele sarnaseid vastuseid keerulistele sisenditele [7]. Nende üks enimtuntuid näiteid on ChatGPT, mis on saavutanud laialdase tuntuse tänu oma võimele vastata väga erinevat tüüpi küsimustele ja selle kasutuselevõtt on kaasa toonud muutusi nii ühiskondlikes kui ka

tööstuslikes tööstiilides, sealhulgas hariduses, kus selle mõju võib olla nii positiivne kui ka negatiivne [8]. Mitmed suured keelemudelid on viimastel aastatel loodud, kiirelt arenenud ja laialdaselt kasutusse tulnud, näiteks: ChatGPT, DeepSeek, Microsofti Bing Chat (nüüd tuntud kui Copilot) ja Google'i Bard (nüüd nimega Gemini) [9].

ChatGPT<sup>1</sup> on OpenAI poolt välja töötatud generatiivne keelemudel, mis sai alguse neurovõrkude ja masinõppe algoritmide kiirest arengust [10]. Keelemudel põhineb *Generative Pre-trained Transformer* (GPT) süsteemil, mis on tehisintellekti mudel, mille eesmärk on luua võimalikult täpseid tekstilisi vastuseid etteantud kirjalike sisendite põhjal [6]. Tegemist on keelemudeliga, mis on treenitud suurte tekstikorpuste peal ja pakub tuge keelepõhiste ülesannete lahendamisel [6]. Keelemudelid nagu ChatGPT on võimelised vastama ainult sisenditele, mille kohta neil on eelnevalt informatsioon olemas, sest vastasel juhul võib vastus olla kokku pandud sarnase konteksti põhjal, kuid mitte korrektselt [6]. Keelemudelid nagu ChatGPT on läbinud mitmeid arenguetappe, mille käigus on mudelid muutunud järjest suuremaks ja keerukamaks [4]. GPT-seeria, alates GPT-1-st kuni GPT-4-ni, on näidanud suurt kasvu parameetrite arvu osas, millega neid treeniti: GPT-1 sisaldas sadu miljoneid parameetreid, samas kui GPT-4 on jõudnud 1,7 triljoni parameetrini [4]. See areng on võimaldanud täiustada keele mõistmise ja genereerimise võimekust.

Teine tuntumatest keelemudelitest on Microsoft Copilot<sup>2</sup>. See on tehisintellektil põhinev tööriist, mis on olemuselt sarnane ChatGPT-le, kuid rohkem levinud hariduses ja koolides ning omab potentsiaali muuta õpetamis- ja õppimismeetodeid [3]. Copilot toob haridusse uuenduslikke lahendusi ja võib põhjalikult muuta õpetamis- ja õppemeetodeid [3]. Selle kasutamisel öeldakse olevat väga palju positiivseid külgi, näiteks kiire algne tagasiside või esmane vastuste kontrollimine, mis muudab õppetöö palju efektiivsemaks, aga selleks oleks vaja üliõpilastele õpetada selle õigest kasutamist ning erinevaid võimalusi [3].

Veel üks tuntud ja laialdaselt levinud keelemudel on Google Gemini<sup>3</sup>. Nagu teised arenenud AI mudelid, on ka Gemini keelemudel, mis on treenitud suurel hulgal tekstil. See võimaldab tal genereerida inimlikku teksti, tõlkida keeli, luua erinevat tüüpi loomingulist sisu ja vastata küsimustele informatiivselt [11]. Erinevalt teistest eristub Gemini oma multimodaalse disainiga, mis ühendab teksti, pilte ja heli, võimaldades tal töödelda ja luua väljundeid

---

<sup>1</sup> <https://chatgpt.com/>

<sup>2</sup> <https://copilot.microsoft.com/>

<sup>3</sup> <https://gemini.google.com/app>

erinevatest andmetüüpidest. Lisaks sellele on see mudel ka integreeritud teistesse Google'i teenustesse ja ka nende Google Pixel telefonidesse [11].

DeepSeek<sup>4</sup> keelemudel kirjeldab ise ennast järgnevalt: “DeepSeek on täiustatud keelemudel, mis on loodud tehisintellekti ja masinõppe tehnoloogiate abil, et mõista ja genereerida inimkeelt. Kuna tegemist on uue keelemudeliga, siis järgnevat informatsiooni tuleks kohelda kriitiliselt, sest seda ei ole jõutud veel liigselt uurida. See mudel põhineb sügavõppevõrgustikel, mis on treenitud suurel hulgal tekstil, et pakkuda täpset ja kontekstipõhist vastust. DeepSeek võimaldab kasutajatel suhelda loomulikus keeles, vastates küsimustele, pakkudes seletusi ja genereerides loovat sisu.” [12]. Mudeli peamised omadused on keelest arusaamine, kontekstipõhised vastused, mitmekülgsus, kohandatavus ja eetiline kasutamine [12]. Keelemudel DeepSeek väidab, et selle rakendused on mitmekülgsed, näiteks saab seda kasutada hariduses, klienditoe automatiseerimises ja ka tõlkimisel, see suudab kohaneda erinevate ülesannete ja kontekstidega, mis muudab selle väärtuslikuks vahendiks nii teadlastele kui ka igapäeva kasutajatele [12]. Seega tundub DeepSeek nagu iga teine keelemudel, kuid DeepSeeki teadlaste poolt tehtud uuringus väidetakse, et tulemuste järgi on nende keelemudel konkureerivatest keelemudelitest nagu OpenAI-o1 mudel, mis on hetkel kõige levinum vabalt kasutatav mudel, teatud valdkondades tugevam ja teistes jällegi nõrgem [13]. Näiteks matemaatiliste ja algoritmiliste ülesannete lahendamisel saadi samaväärsed tulemused, kuid inseneerimisele keskendunud kodeerimisülesannetes oli OpenAI mudel paremate tulemustega [13].

Kokkuvõtvalt võib öelda, et keelemudelid on viimastel aastatel palju arenenud ning muutunud igapäevaselt kasutatavaks tööriistaks hariduses. Erinevad mudelid, nagu ChatGPT, Copilot, Gemini ja DeepSeek, pakuvad mitmekülgsed võimalusi teksti genereerimiseks ning õppetöö edendamiseks. Igal mudelil on oma eripärad ja tugevused, kuid nende kasutamine nõuab teadlikkust nii võimalustest kui ka piirangutest.

## 1.2 Keelemudelite kasutamine hariduses

Viimasel ajal on ChatGPT saanud palju tähelepanu läbi selle eeliste (näiteks koodijuppide kirjutamine) ja puuduste (näiteks plagiati) erinevates valdkondades ning selle, millist mõju see kõrgharidusele üldiselt avaldab [13]. Keelemudelid võivad täiendada traditsioonilist õpet

---

<sup>4</sup> <https://www.deepseek.com/>

pakkudes personaliseeritud tuge ja kiiret tagasisidet [14]. See on eriti oluline keerukates programmeerimise ainetes, kus õpilased sageli vajavad nii teoreetilist tuge kui ka praktilisi näiteid [14]. Õppimise efektiivsus sõltub nii teooria mõistmisest kui ka selle praktilisest rakendamisest ülesannetes. Matemaatika on üks valdkond, kus keelemudelitel on suur mõju olnud, sest need aitavad selle raske aine palju lihtsamalt mõistetavaks teha [7]. Neil Andersoni jt [6] on näidanud, et keelemudelid, nagu ChatGPT, suudavad õpilaste küsimustele vastata, tuua näiteid ja pakkuda parandusi nende kirjutatud viipadele. Samuti leiavad Anderson jt, et keelemudelid võivad aidata vähendada õpetajate koormust, pakkudes õpilastele iseseisvaid õppimisvõimalusi ja hõlbustades tööde parandamist ning tundideks ettevalmistumist [6].

Alati ei ole keelemudeli vastused täpsed, eriti spetsiifilistes olukordades, kus vajalikke andmeid või konteksti ei leidu mudeli treeningandmetes [15]. Üks peamine probleem on keelemudelite hallutsineerimine (*hallucination*) ehk valeinformatsiooni esitamine. Hallutsineerimine tekib, kui keelemudel genereerib vastuseid, mis ei ole faktipõhised ega tugine usaldusväärsetele allikatele [16]. Need on tavaliselt veenvalt sõnastatud vale sisuga vastused, näiteks esitades arvutustehte, kus on viga sees väites, et see on korrektne [17]. Hiljutine rahvusvaheline uuring leidis, et väiksemad mudelid hallutsineerivad sagedamini kui suuremad (nagu ChatGPT) ning erinevates keeltes esineb hallutsinatsioon erineval määral, kuid nende esinemine ei sõltu otseselt keele levikust [17].

Liigne tuginemine sellistele tööriistadele võib vähendada õpilaste iseseisva mõtlemise arengut. Gökhan jt. [18] uuring viitab, et paljud õpilased kopeerivad keelemudeli vastuseid ilma selle korrektsust kontrollimata ja kasutavad seda lootes, et vastus on õige. ChatGPT saavutab õigusteaduse alastel eksamitel umbes C+ ehk keskpärase taseme [1]. Matemaatikaülesannete puhul tuli välja, et lihtsamate ülesannete lahendamisel oli keelemudel väga heal tasemel, kuid keerukamate ülesannete puhul tekkis probleeme [1].

Farrokhnia jt [19] uurisid erinevaid allikaid, rakendas SWOT-analüüsi, et hinnata ChatGPT kasutamise eeliseid ja riske hariduses. Toodi välja, et ChatGPT tugevusteks on realistlike vastuste loomine, kohandatud vastuste pakkumine reaajas ning võime ennast täiustada ja vigadest õppida. Nõrkuste hulka kuulusid aga piiratud sügav mõistmine, kalduvused eelarvamustele ning raskused kontekstist arusaamisel ja vastuste kvaliteedi hindamisel. Ohtudeks märgiti konteksti halb mõistmine, akadeemilise aususe õonestamine, plagiaadi võimalus ja kõrgema astme kognitiivsete oskuste langus.

Kokkuvõtvalt võib väita, et keelemudelid pakuvad õppijatele tuge teadmiste omandamisel ja võimaldavad saada kiiret tagasisidet. Samas tuleb nende kasutamisel arvestada piirangutega. Mudelite vastused ei ole alati täpsed ning liigne sõltuvus võib vähendada õppija kriitilise mõtlemise ja iseseisva töö oskusi. Keelemudelid võivad olla hea lisandus kvaliteetse hariduse andmisesse, kuid nende kasutamisel tuleb olla teadlik piirangutest ja kaasnevatest ohtudest.

### **1.3 Keelemudelite mõju arvutiteaduse õppimisele**

Haridustöötajate üks peamisi muresid on see, et õpilased kasutavad ChatGPT-d rikkumaks kehtestatud akadeemilise aususe reegleid [1]. Seeläbi on ka ülikoolide õppejõududel, kes õpetavad programmeerimist, tekkinud huvi, kas ja kui palju on võimalik tuvastada keelemudeli kasutamist hindeliste tööde kontrollimisel [1]. Järgnevalt käsitletakse uurimusi erinevate programmeerimiskeelte, sealhulgas päringukeelte õppimisel ja õpetamisel.

Sun jt [20] võrdlesid ChatGPT abil programmeerimist iseseisva programmeerimisega, mille tulemused näitasid, et ChatGPT kasutamine ei mõjutanud oluliselt õpilaste sooritust, kuid parandas nende programmeerimiskäitumist ja -hoiakuid. See viitab, et programmeerimisega alustamisel võib ChatGPT kasutamine olla kasulik, tingimusel et seda tehakse õigesti ja eesmärgiga muuta õppeprotsess efektiivsemaks [20]. ChatGPT võib pakkuda reaajas personaliseeritud tuge, kuid selle vastuste täpsust ja täielikkust on vaja hoolikalt hinnata [21]. Budhiraja jt. [22] uuringus arvas enamik õpilasi, et ChatGPT on kasulik ja toetab õppimist. Samas on keelemudelite paremaks ühildamiseks õppekavadesse vaja tegeleda probleemidega nagu õpilaste liigne tuginemine tehisintellekti genereeritud lahendustele [22]. Kui õpilased esitavad vastuseid kontrollimata keelemudeli poolt antud vale vastuse kodutööna või arvestusel, siis võib see kaasa tuua halva hinde või halvemal juhul akadeemilise petturluse süüdistuse.

Keelemudelid võivad olla kasulikud mitmesuguste ülesannete jaoks, näiteks suudavad need pakkuda erinevaid varjante mingi kindla programmeerimisülesande lahenduseks, toota kvaliteetset õppematerjali, genereerida alustuskoodi ning aidata koodi siluda [22]. Samas tuleb nende tööriistade laiemaks kasutuselevõtuks tegeleda mitmete eetika ja turvalisusega seotud küsimustega [22]. ChatGPT tõhusus on piiratud kasutaja programmeerimisoskustega, ilma vastavate erialaste teadmisteta kasutaja ei pruugi ChatGPT genereeritud koodi korrektsust osata kontrollida ega korrektselt viipasid sõnastada [23]. Keelemudelid ei mõista

koodi nagu inimesed ja teevad otsuseid statistiliste mustrite pealt, mis on saadud suurte andmemassiivide läbi töötamisest [24]. Seega keelemudelite toimimismehhanism ei ole tavakasutajale lõpuni selge ja ei tea kunagi täpselt, kuidas need väljundini jõuavad. Lisaks on leitud, et edukaks kasutamiseks peab inimene oskama esitada õigeid küsimusi ja lisainformatsiooni, kuna tehisintellektil endal puuduvad vajalikud kriitilise mõtlemise ja probleemide lahendamise oskused [23]. Keelemudelid automatiseerivad ja kiirendavad koodi genereerimist, silumist ja optimeerimist, võimaldades arendajatel keskenduda keerukamatele ja loovamatele ülesannetele [25].

Seega saab järeldada, et keelemudelite kasutamine arvutiteaduse õppimisel pakub kiiret personaalset tagasisidet ja aitab arendada programmeerimisoskuseid. Siiski eeldab keelemudelite kasutamine, et kasutaja on kursis valdkonna mõistetega ja käesoleva ülesandega. Samuti on oluline vastuste kriitiliselt hindamine ja ebaõigete väljundite ära tundimine, et vältida valeinformatsioonile tuginemist. Viimaks peab keelemudeleid kasutades silmas pidama akadeemilist ausust, et vältida ülikooli reeglitega vastuollu sattumist.

## **2. Tartu Ülikooli kursus “Andmebaasid”**

Selles peatükis kirjeldatakse kursuse „Andmebaasid“ (LTAT.03.004) ülesehitust ning täpsemalt kodutöid, millele see töö keskendub. Tegemist on esimese aasta informaatika ja matemaatilise statistika bakalaureuseõppe üliõpilastele kohustusliku kursusega, mis annab üliõpilastele ülevaate relatsiooniliste andmebaaside olemusest, struktuurist ja kasutusest. Aine mahuks on 6 EAP ja tegemist on hindelise kursusega.

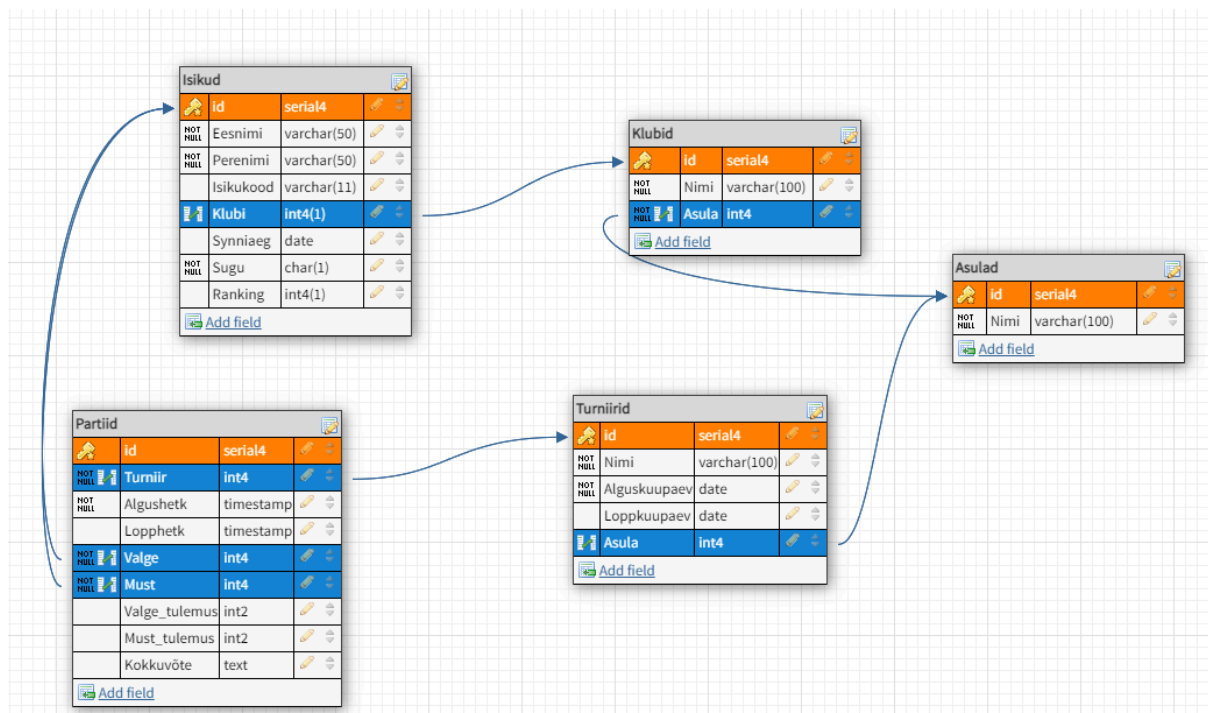
### **2.1 Kursuse sisu**

Järgnev lõik tugineb Tartu Ülikooli Õppeinfosüsteemist [26] saadud infole. Kursuse LTAT.03.004 "Andmebaasid" eesmärgiks on anda üliõpilastele arusaam relatsiooniliste andmebaaside loomise, kasutamise ja teooria kohta. Kursuse läbinu tunneb valdkonna mõisteid, oskab disainida andmebaase kasutades ER-diagramme ja teisendada neid relatsioonilisteks mudeliteks. Samuti annab kursus oskuse SQL keeles andmebaase kirjeldada ning koostada efektiivseid päringuid, sealhulgas filtreerida, grupeerida ja agregeerida andmeid. Lisaks praktilistele oskustele pööratakse tähelepanu ka andmebaaside teooriale, nagu funktsionaalsetele sõltuvustele, normaalkujudele ja relatsioonialgebrale. Kursuse jooksul õpitakse looma vaateid, kirjutama ja muutma funktsioone, protseduure ning trigereid. Tutvutakse ka andmekaevega ning vaadeldakse andmebaasidega seotud teemasid, nagu andmelaod, andmekaitse ja süsteemi töökindlus.

### **2.2 Kursusel kasutatavad programmid ja andmebaasi kirjeldus**

Tartu Ülikooli kursusel “Andmebaasid” kasutatakse PostgreSQL-i koos DBeaveri rakendusega, mis võimaldab üliõpilastel andmebaasidega töötada ja päringuid luua [30]. PostgreSQL on avatud lähtekoodiga objekt-relatsiooniline andmebaasihaldussüsteem, mis pakub tugevat arhitektuuri, kõrget töökindlust ja laialdast arendatavust [27]. See on avatud lähtekoodiga objekt-relatsiooniline andmebaasihaldussüsteem, mida iseloomustab päringute efektiivsus ja sobivus ka keerukamateks äri lahendusteks [28]. DB-Engines edetabeli [29] andmetel oli PostgreSQL 2025. aasta aprillis populaarsuselt neljas andmebaasihaldussüsteem maailmas. PostgreSQL töötab kõigil suurematel operatsioonisüsteemidel ning on tasuta

kasutatav [27]. Kursusel kasutatav välja mõeldud andmebaas “Male” koosneb viiest tabelist ehk relatsioonist. Andmebaasi mudel on kujutatud joonisel 1.



Joonis 1. Male andmebaasi mudel. Allikas: TÜ kursuse “Andmebaasid” Moodle [36]

Andmebaas koosneb viiest tabelist: isikud, partiid, turniirid, klubid ja asulad. Tabelitevahelised ühendused on näidatud nooltega.

## 2.3 Hinnatav osa

Kursuse “Andmebaasid” raames toimuvad praktikumid ning tuleb esitada kuus kodutööd, mis annavad kokku ühe lävendi “Kodutööd ja praktikumid”. Selle aineplokki eest on võimalik kokku teenida 21 punkti ja lävendi ületamiseks peab teenima vähemalt 11 punkti. Nendest 9 punkti on võimalik teenida praktikumidest osavõtmise eest ja tehtud tunnitöö esitamise eest ning ülejäänud 12 punkti saab kodutööde lahendamise eest. Kursuselt on valitud siia uuringusse hinnatavaks osaks selle kodutööd. Need ülesanded valiti käesoleva töö praktiliseks osaks, kuna need on struktureeritud viisil, mis võimaldab keelemudelil neid tõhusalt lahendada ning samal ajal katavad nad andmebaaside aine õppimise põhietape. Samuti on need kõige tõenäolisemad ülesanded, mida üliõpilased keelemudelitega lahendaksid, sest need sooritatakse väljaspool praktikumi aega.

## 2.4 Kodutööde kontrollimine

Üliõpilaste poolt esitatud kodutöid kontrollitakse automaatse hindamissüsteemi abil, milleks on kursusel kasutusel olev Lahenduse<sup>5</sup> keskkond. Tegemist on veebipõhise platvormiga, mis võimaldab üliõpilastel oma lahendusi jooksvalt testida, saada tagasisidet ning esitada ülesannete lõplikud versioonid hindamiseks. Lahenduse keskkond on tihedalt seotud kursuse struktuuriga ja toetab iseseisvat õppimist, pakkudes igale üliõpilastele võimalust oma lahendusi enne esitamise tähtaega mitmel korral proovida ja tuues välja nende veakohad. See vähendab tehniliste vigade hulka, võimaldab õppijatel paremini mõista oma eksimusi. Automaatkontrollid on loodud nii, et need hindavad lahenduse vastavust oodatud tulemusele, kontrollides näiteks SQL-päringute õigsust, tabelistruktuuride sobivust või vajalike andmete olemasolu. Samuti tagab selline süsteem hindamise objektiivsuse ja järjepidevuse, võimaldades sama hindamisloogikat rakendada kõigile lahendustele [31].

---

<sup>5</sup> <https://lahendus.ut.ee/courses>

### **3. Metoodika**

Selles peatükis selgitatakse, mis keelemudelid valiti käesoleva uurimuse raames andmebaaside õppeaine kodutööde lahendamiseks. Samuti selgitatakse, kuidas neid keelemudeleid uurimuse läbiviimiseks kasutatakse ning millega saadud tulemusi hindama ja võrldema hakatakse.

#### **3.1 Valitud keelemudelid**

Töö praktilise osa tegemiseks otsustati kasutada ChatGPT-3.5 mudelit ja Microsoft Copiloti keelemudelit, mis põhineb GPT-4 arhitektuuril. ChatGPT-3.5 keelemudel valiti, sest erinevalt ChatGPT-4 mudelist on tegemist tasuta mudeliga. Kuna ChatGPT-3.5 on tasuta versioon, siis eeldati, et üliõpilased kasutaksid seda versiooni kursuse ülesannete lahendamisel tõenäolisemalt kui tasulist versiooni. Samuti küsimustele vastanud üliõpilastest üle 95%<sup>6</sup> väitsid, et nad kasutasid just ChatGPT mudelit sellel kursusel. Microsoft Copilot sai uuringusse valitud, sest Tartu Ülikool pakub tasulist versiooni sellest üliõpilastele tasuta, kui kasutada sisse logimiseks kooli emaili [32]. See muudab selle keelemudeli kasutamise üliõpilaste seas tõenäolisemaks.

#### **3.2 Mudelite kasutamine**

Kursuse kodutööd lahendati kahel viisil. Esimesel korral lahendati kodutööd mõlema keelemudelig viisil, kus eelinfo andmebaasi kohta ei olnud mudelitele teada. Enamus kodutöödest sisaldasid mitut alamülesannet, mistõttu said keelemudelit iga ülesande eraldi viibana. Viibad ei sisaldanud muud informatsiooni, kui ülesande teksti ja tulemuse kirjeldust (Lisa 2-3). Ülesanded olid eesti keeles, seega pidi mudel ise tegema tõlke ja vastas samuti eesti keeles. Vastused sisaldasid lahendust SQL koodi formaadis ning ka kirjeldust, mida mingi koodi osa teeb. Tulemus saadi ilma keelemudelile selgitamata, mis veeru- ja tabelinimed on andmebaasis olemas. Lihtsamad vead nagu valed veeru- või tabelinimed parandati jälgides, mida üliõpilased ise suudaksid parandada ilma mudeli antud loogikat muutmata. Saadud tulemused sisestati Lahendus keskkonda, kus lasti tulemust hinnata

---

<sup>6</sup> Andmed on saadud kursuse “Andmebaasid” vastutavalt õppejõult

automaatkontrollidel. Juhul kui tegemist oli vale tabeli nimega, siis see parandati ja samuti ka veerunimede puhul, et automaatkontroll saaks loogikat hinnata. Seda oli tehti juhul, kui keelemudelil puudus eelinformatsioon ning seda tehti igas kodutöös, sest veerunimede vahetamine on asi mida ka üliõpilane teeks küsides keelemudelilt abi.

Teisel korral lahendati kursuse kodutöid andes keelemudelitele ette andmebaasi struktuur koos olemasolevate seostega. Alustuseks küsiti keelemudelilt, mida oleks vaja, et nad oskaksid aidata kursuse kodutöö ülesannete lahendamiseks aidata (Lisa 4). Hindamaks mudelite suutlikkust lahendada kursuse kodutöid, said mõlemad keelemudelid enne ülesannete kirjeldusi andmebaaside struktuuri kohta informatsiooni nagu millised tabelid eksisteerivad, mis on nende ülesehitus ja andmetüübid. See oli kogu informatsioon, mida keelemudelid enne ülesande saamist teadsid. Kuna osade kodutööde eelduseks oli sellele eelneva praktikumi läbi lahendamine, sest muidu oli osa andmebaasist puudu, siis sellest tulenevad struktuurimuudatused anti keelemudelitele ette. See tagas, et mudelitel oleks ülesandeid lahendades olemas korrektne andmebaasi struktuur. Koodi kontrollimiseks kasutati eelnevalt mainitud Lahendus keskkonda. Saadud tulemused salvestati (Lisa 5) ja ka weakohad märgiti ära, et leida mustreid ja korduvaid vigu, mida keelemudelid tegid.

### **3.3 Valimi kirjeldus ja tulemuste arvutamise meetodika**

Teisele uurimisküsimusele vastamiseks võeti aluseks Tartu Ülikooli kursuse „Andmebaasid“ raames esitatud kodutööde hindetabel, mis sisaldab protsentuaalseid tulemusi alamülesannete kaupa ning kodutööde koondhindeid punktisüsteemis (0–2 punkti). Iga üliõpilase kohta on märgitud, kas konkreetne ülesanne on esitatud või mitte; esitamata tööde korral esineb andmestikus sümbol „-“, mis arvutuste käigus asendati puuduva väärtuse tähisega NULL. Valimi hulgas olid 337 aastal 2025 ainele registreerunud üliõpilast, kes võtsid vastu kutse Lahenduse keskkonda. Autor on selle aine praktikumijuhendaja juba teist õppeaastat ning sai seetõttu ligipääsu hindetabelile Moodle'i keskkonnast. Üliõpilaste hinnete kasutamine on kooskõlastatud Tartu Ülikooli eetikakeskuse loaga nr 3851T-22, mis on välja antud 01.02.2024.

Õppijate keskmiste tulemuste arvutamisel ei ole arvesse võetud esitamata töid. See tähendab, et iga keskmine on arvutatud nende õppurite põhjal, kes vastava ülesande ka tegelikult esitasid. Näiteks kui konkreetse kodutöö esitas 280 üliõpilast 347-st, siis keskmine arvutati

nende 280 õppija tulemuste põhjal. Lisaks keskmisele arvutati ka mediaan, standardhälve, miinimum ja maksimum, et kirjeldada tulemuste jaotust ning varieeruvust.

### **3.4 Kodutööd**

Uuringus lahendatakse kuus koduülesannet, millest igatüki keskendub kindlale andmebaaside valdkonnale. Iga kodutöö eest on võimalik teenida 2 punkti, mis olenevalt kodutööst võib jaguneda alamülesanneteks.

Esimeses kodutöös 'Lihtsad päringud' käsitletakse lihtsaid päringuid, mille abil õpitakse välja tooma, filtreerima ja sorteerima andmeid andmebaasist, kasutades selliseid käskke nagu SELECT, WHERE ja ORDER BY. See loob aluse kõigi järgmiste teemade mõistmiseks.

Teine kodutöö 'Päringud mitme tabeli pealt' keskendub päringutele mitme tabeli pealt, mille puhul kasutatakse erinevaid liitumisviise (JOIN, UNION jne), et siduda eri tabelites paiknevat informatsiooni. Samuti olid selles kodutöös ka teemaks alampäringud.

Kolmas kodutöö 'Tabelite loomine, andmete sisestamine ja välisvõti' hõlmab tabelite loomist, andmete sisestamist ja välisvõtmete kasutamist. Selle käigus kujundatakse andmebaasi struktuuri ning määratletakse tabelite vahelised seosed, mis on oluline andmete sidususe ja loogilise ülesehituse tagamiseks.

Neljanda kodutöö eesmärk oli harjutada vaadete ja materialiseeritud vaadete loomist andmebaasis. Ülesannetes pidi koostama mitmesuguseid vaateid, mis koondasid ja töötlesid infot turniiride, klubide ja mängijate kohta. Näiteks loodi vaade, mis näitab iga partii kohta, kes võitis ja kus mäng toimus, ning vaade, mis arvutab, kui palju partiisid on iga klubi mängijad mänginud.

Viiendas kodutöös pidid üliõpilased looma erinevaid funktsioone ja ühe protseduuri, mille eesmärk oli rakendada andmebaasis loogikat, mis võimaldab arvutada vanust, leida klubiliikmete keskmist ränkingut, kuvada turniiri parimaid mängijaid ning sisestada uusi turniire koos vajaliku lisainfoga. Ülesannete lahendamisel tuli arvestada nii sisendparameetrite kasutamise kui ka tingimusliku väljundi koostamisega.

Kuuenda kodutöö ülesannetes oli vaja luua indekseid ja trigereid. Üliõpilased pidid looma erineva ülesehitusega indekseid andmete sortimise ja päringukiiruse optimeerimiseks ning koostama trigereid, mis kontrollivad sisestatavate andmete loogikat või lisavad puuduva info automaatselt.

## 4. Tulemused

Selles peatükis tuuakse välja keelemudelite poolt lahendatud kodutööde tulemused ning nende võrdlus üliõpilaste saadud tulemustega. Bakalaureusetöö tulemused keskenduvad sellele, kui võrd suudavad ChatGPT ja Copilot lahendada kursuse „Andmebaasid“ ülesandeid ning millised olid nende peamised vead lähtuvalt püstitatud uurimisküsimustest.

### 4.1 Keelemudelite lahendused

Esimene uurimisküsimus oli, et millised on keelemudelite poolt lahendatud "Andmebaasi" kursuse koduülesannete tulemused, kui kasutada hindamiseks automaatkontrolli skoori ning kuidas erinevad tulemused sõltuvalt sellest, kas andmebaas oli ette antud SQL kujul või mitte? Järgnevalt on toodud ülevaade keelemudelite tulemustest kodutööde kaupa.

#### 4.1.1 Kodutöö 1

Kodutöö tulemused olid keelemudelitel kõrged (vt tabel 1). Selles kodutöös oli vaja koostada neli päringut, mis kõik põhinesid ainult Isikud tabelil. Keelemudelid saavutasid alamülesannete 1.1 ja 1.2 puhul maksimaalse tulemuse ehk 100% sõltumata sellest, kas neil oli eelinfo andmebaasi kohta või mitte. Erinevused tekkisid aga järgnevates alamülesannetes, kus ülesandes 1.3 saavutasid andmebaasi eelinfoga mudelid maksimaalse tulemuse, kuid eelinfota saavutasid tulemuseks 60%. Viimases alamülesandes 1.4 said eelinfota mudelid tulemuseks vaid 40%. Kokkuvõttes said mudelid andmebaasi eelinfota 1,5 punkti. Teades andmebaasi ülesehitust said mõlemad mudelid maksimaalse punktisumma ehk 2 punkti.

Tabel 1. Keelemudelite tulemused esimeses kodutöös

Ülesanded	1.1	1.2	1.3	1.4	Kokku
ChatGPT punktisumma andmebaasi infota	100%	100%	60%	40%	1,5 p
Copilot-i punktisumma andmebaasi infota	100%	100%	60%	40%	1,5 p
ChatGPT punktisumma andmebaasi infoga	100%	100%	100%	100%	2 p
Copilot-i punktisumma andmebaasi infoga	100%	100%	100%	100%	2 p

Peamised vead ilmnesisid olukordades, kus mudelitel puudus informatsioon andmebaasi struktuuri kohta. Sageli kasutati valesid veeru- või tabelinimesid. Alamülesandes 1.4 tehti viga sootunnuse kontrollimisel. Keelemudelid üritasid võrrelda väärtust kujul *sugu = 'mees'*, kuigi õige vaste andmebaasis oli *sugu = 'm'*. Tegemist oli lihtsa veaga, mida üliõpilane suudaks automaatkontrolli abil hõlpsasti märgata ja parandada. Seega oli see viga pigem tehniline ja mitte loogikaviga. Kuid alamülesandes 1.3 eksisid mudelid tulemuste arvutamisel. Jagamistehtes ei teisendatud jagajat kümnendarvuks ning seetõttu toimus täisarvuline jagamine. Selle tulemusel ümardusid kõik jagatised täisarvudeks, mistõttu automaatkontroll ei tuvastanud neid õigete vastustena. See oli suurem loogikaviga, mida üliõpilane ei pruugi kohe avastada.

#### 4.1.2 Kodutöö 2

Selle kodutöö keskmes olid päringud mitme tabeli pealt, mis eeldasid JOIN-käsu oskuslikku kasutamist. Samuti oli selles kodutöös vaja kasutada esimeses kahes ülesandes alampäringut ning viimases käsku UNION. Keelemudelite tulemused näitasid suuremat varieeruvust võrreldes esimese kodutööga (vt tabel 2). Andmebaasi eelinfota suutis ChatGPT lahendada enamik ülesandeid heal tasemel, samas kui Copilot ei saanud ühegi ülesande eest punkte.

Andmebaasi andmete lisamisel paranes Copiloti tulemus märgatavalt, ületades isegi ChatGPT oma.

Tabel 2. Keelemudelite tulemused teises kodutöös

Ülesanded	2.1	2.2	2.3	Kokku
<b>ChatGPT punktisumma andmebaasi infota</b>	100%	100%	60%	1,6 p
<b>Copilot-i punktisumma andmebaasi infota</b>	0%	0%	0%	0 p
<b>ChatGPT punktisumma andmebaasi infoga</b>	100%	100%	70%	1,7 p
<b>Copilot-i punktisumma andmebaasi infoga</b>	100%	60%	100%	1,8 p

ChatGPT eksis ainult alamülesandes 2.3 ja tegi seal mõlemal juhul vea UNION käsu kasutamisel, kasutama oleks pidanud UNION ALL käsku ja seeläbi jäi palju väärtuseid kajastamata. Eelinfota keelemudel tegi ka vea tulemuste järjestamisel, millest tekkis 10% erinevus tulemuses. Copilot seevastu ebaõnnestus, kui andmebaasi eelinfo puudus. Eelinfota arvas Copilot, et kõik andmed olid kättesaadavad ühest tabelist. Päringus oli vaja kasutada mitut tabelit ja selles oli ka veakoht. Eelinfoga eksis Copilot ainult alamülesandes 2.2, kus jäi kasutamata DISTINCT käsklus ja seega oli saadud kirjeid liiga palju.

#### 4.1.3 Kodutöö 3

Selles kodutöös tuli koostada SQL-päringuid, mis hõlmasid tabelite loomist, andmete sisestamist ja seoste defineerimist välisvõtmete kaudu. Kolmandas kodutöös olid keelemudelite tulemused üsna stabiilsed, kuigi ka üllatavad (vt tabel 3). Alamülesandes 3.1-6, kus oli vaja luua uus tabel ja muuta kirjeid olemasolevates tabelites ning alamülesandes 3.7 suutsid keelemudelid genereerida korrektseid lahendusi ja seeläbi saadi tulemuseks 100%. Samas ilmnes ülesandes 3.8, kus oli vaja koostada päring, mis väljastaks millistes asulates mängides on mängija nimega Kalamees valgetega võitnud, tekkis erisusi. Andmebaasi

eelinfota saavutas Copilot 0% tulemuse. Huvitava tulemuse andis ChatGPT mudel, millel ei olnud ligipääsu andmebaaside eelinfole, kuna see saavutas parema tulemuse kui ChatGPT mudel, millel vastav ligipääs oli. Eelinformatsiooniga mudelid saavutasid vaid 30% tulemuse.

Tabel 3. Keelemudelite tulemused kolmandas kodutöös

Ülesanded	3.1-6	3.7	3.8	Kokku
<b>ChatGPT punktisumma andmebaasi infota</b>	100%	100%	100%	2 p
<b>Copilot-i punktisumma andmebaasi infota</b>	100%	100%	0%	1,7 p
<b>ChatGPT punktisumma andmebaasi infoga</b>	100%	100%	30%	1,79 p
<b>Copilot-i punktisumma andmebaasi infoga</b>	100%	100%	30%	1,79 p

Andmebaasi eelinfota Copilot saavutas alamülesandes 3.8 tulemuse 0%, sest sellel puudus arusaam, et kasutama peab kahe tabeli asemel kolme. Mõlemad eelinfoga keelemudelid said viimases alamülesandes sama tulemuse, kasutades küll vajalikku kolme tabelit, kuid kontrollisid perenime asemel eesnime vastavust “Kalamehe” tingimusele ja seeläbi ei saanud ühtegi tulemust. Kui teha kitsendus, kus kontrollida valet veergu sellise tingimuse vastu, kus ei saagi sellist nime eksisteerida olid tulemused valed. Keelemudelid ei saanud aru, et tegemist oli perekonnanimega.

#### 4.1.4 Kodutöö 4

Kodutöös oli vaja koostada vaateid, mis ühendasid andmeid mitmest tabelist ning esitasid tulemused õiges formaadis. Ülesannete hulgas oli ka materialiseeritud vaate loomine. Neljanda kodutöö tulemused näitasid erinevust keelemudelite sooritusel sõltuvalt sellest, kas neile anti eelnevalt andmebaasi eelinfo või mitte (vt tabel 4). Nii ChatGPT kui ka Copilot ebaõnnestusid täielikult andmebaasi eelinformatsioonita. Kõigi alamülesannete tulemused jäid 0% peale. Kui mudelitele anti ette andmebaasi ülesehitus, paranesid keelemudelite

tulemused oluliselt ja maksimaalseid punkte ei teenitud vaid kahes alamülesandes. ChatGPT koos andmebaasi eelinfoga suutis saada parema tulemuse kui eelinformatsiooniga Copilot edestades seda 0,12 punktiga.

Tabel 4. Keelemudelite tulemused neljandas kodutöös

Ülesanded	4.1	4.2	4.3	4.4	Kokku
<b>ChatGPT punktisumma andmebaasi infota</b>	0%	0%	0%	0%	0 p
<b>Copilot-i punktisumma andmebaasi infota</b>	0%	0%	0%	0%	0 p
<b>ChatGPT punktisumma andmebaasi infoga</b>	100%	100%	100%	100%	2 p
<b>Copilot-i punktisumma andmebaasi infoga</b>	75%	100%	100%	100%	1,88 p

Kuna selles kodutöös oli vaja luua vaateid kasutades mitmeid tabeleid, siis andmebaasi infota keelemudelite vead olid arusaadavad. Neil puudus arusaam veergudest, mis oli mingis tabelis olemas ja nad ei suutnud neid tabeleid korrektselt ühendada ja seeläbi vaateid koostada. Mõlemad mudelid suutsid küll teha õige struktuuriga vaated, ehk vajalikud veerud olid olemas, kuid loogika info saamiseks nendesse oli vale ja vaated ei kompüleerunud. Eelinfoga keelemudelid erinesid 4.1 alamülesandes 25% võrra, kus tuli luua vaade turniiridel mängitud partiide tulemuste kuvamiseks. Erinevus tekkis seetõttu, et Copilot lahendas valesti toimumiskoha leidmise, kasutades selleks valesid tabelite seoseid. Mudel otsis toimumiskohta läbi klubide tabeli, mis ei olnud korrektne, sest küsitud oli turniiri toimumiskoha nime. Seega saadi tulemuseks klubide asukohad ja seeläbi ka liiga palju kirjeid, ehk automaatkontrollis kuvati see veakohana.

#### 4.1.5 Kodutöö 5

Selle kodutöö ülesannetes tuli luua erinevaid funktsioone ja protseduure, mis töötlevad sisendandmeid ja teevad muudatusi andmebaasis vastavalt kehtestatud loogikale. Kodutöö

tulemused näitasid, et esimesed kaks alamülesannet olid kõigi keelemudelite jaoks lihtsad, need lahendati edukalt sõltumata sellest, kas mudelile oli andmebaasi eelinfo ette antud või mitte (vt tabel 5). Suuremad erinevused ilmneseid aga ülesannetes 5.3 ja 5.4. Eelinfota ei suutnud Copilot neid lahendada, samas kui ChatGPT teenis osalise tulemuse. Kui mudelitele lisati andmebaasi eelinformatsioon, paranesid mõlema tulemused ning mõlemad saavutasid viimases ülesandes maksimaalse tulemuse.

Tabel 5. Keelemudelite tulemused viiendas kodutöös

Ülesanded	5.1	5.2	5.3	5.4	Kokku
<b>ChatGPT punktisumma andmebaasi infota</b>	100%	100%	40%	0%	1,2 p
<b>Copilot-i punktisumma andmebaasi infota</b>	100%	100%	0%	0%	1 p
<b>ChatGPT punktisumma andmebaasi infoga</b>	100%	100%	40%	100%	1,7 p
<b>Copilot-i punktisumma andmebaasi infoga</b>	100%	100%	40%	100%	1,7 p

Alamülesandes 5.3 oli vaja luua funktsioon, mis peaks turniiri id põhjal väljastama selle turniiri kümne parima mängija nime ja saadud punktisumma. Ainuke mudel, mis siin punkte ei teeninud oli eelinfota Copilot, sest eeldas, et vaates, mida ülesandes kasutama pidi olid andmed teisel kujul kui nad tegelikult olid. Nimelt kasutas mudel õiget vaadet selle funktsiooni loomisel, kuid arvas, et seal vaates olid eesnimi ja perenimi eraldi veerud. Seeläbi kasutas keelemudel käsku CONCAT nende kokkupanemiseks, kuid vaates oli nimi juba õigel kujul olemas veerunimega “mängija”, mida oleks tulnud kasutada. Teistel mudelitel tekkis viga punktide arvutamisel. Mudelid eeldasid, et saavad õiged tulemused vaatest v\_edetabelid, kuid sealsed punktid oleks pidanud kokku summeerima, mida mudelid ei teinud. Ülesannet oleks saanud ka lahendada ilma olemasolevat vaadet kasutamata, kuid mudelid seda ei teinud. Viimases ülesandes 5.4, kus oli vaja luua protseduur, mis pidi sisestama uue turniiri ja kui asulat pole asulate tabelis, siis ka selle asula, ei saanud eelinfota mudelid ühtegi punkti. Mudelid lahendasid esimese poole alamülesandest korrektset, kuid mõlemad jätsid tegemata

osa, kus pidi sisestama turniiri turniiride tabelisse ja kuvama turniiriinfo koos selle kestusega. See oli selle ülesande peamine raskuskoht ja ilma selleta polnud võimalik punkte teenida. Siin tuleb hästi välja eelis, mille annab ligipääs andmebaasi informatsioonile, kuna mõlemad mudelid lahendasid selle ülesande vastava eelinfo põhjal õigesti.

#### 4.1.6 Kodutöö 6

Viimases kodutöös keskenduti indeksite ja trigerite koostamisele. Eesmärk oli arendada andmebaasi toimimise kiirust ja lisada kontrollmehhanisme andmete muutmisel. Kodutöö tulemused olid ühtlased olenemata andmebaasi eelinfo olemasolust (vt tabel 6). Alamülesandes 6.1-2 suutsid nii ChatGPT kui Copilot saavutada 100% tulemuse, mis näitab, et indeksi loomise loogika, mis oli vastava alamülesande eesmärk, oli hästi arusaadav. Alamülesannetes 6.3 ja 6.4 saavutasid samuti kõik keelemudelid võrdse tulemuse 75%.

Tabel 6. Keelemudelite tulemused kuuendas kodutöös

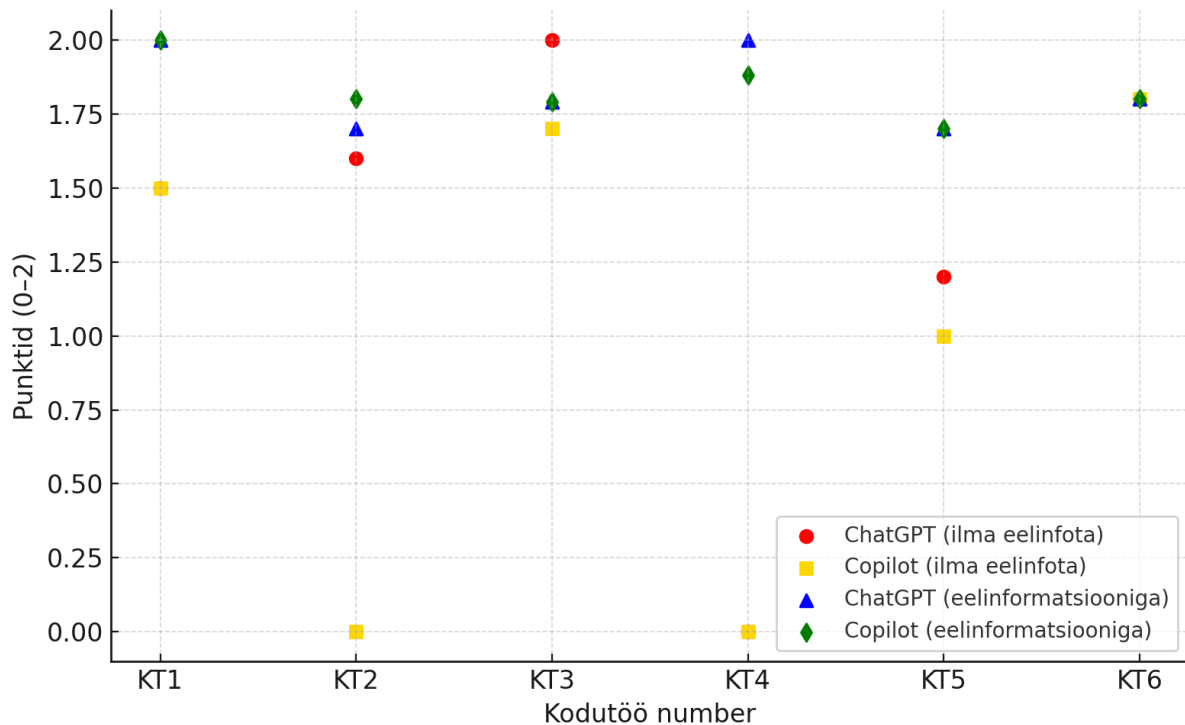
Ülesanded	6.1-2	6.3	6.4	Kokku
<b>ChatGPT punktisumma andmebaasi infota</b>	100%	100%	75%	1,80 p
<b>Copilot-i punktisumma andmebaasi infota</b>	100%	100%	75%	1,80 p
<b>ChatGPT punktisumma andmebaasi infoga</b>	100%	100%	75%	1,80 p
<b>Copilot-i punktisumma andmebaasi infoga</b>	100%	100%	75%	1,80 p

Alamülesandes 6.3 loodud triger pidi kontrollima, kas partii lõppaeg oli hilisem kui algusaeg. Selles ülesandes tegid eelinformastioonita keelemudelid vea valede veerunimede kasutamisega. Tegemist oli veaga, mille üliõpilased suudaksid koheselt parandada, seega muutes veerunimed ära said mõlemad mudelid maksimaalse tulemuse. Viimases ülesandes 6.4 oli vaja luua triger, mis pärast isiku lisamist või isiku kirje muutmist kontrollib, kas isik kuulub mingisse klubisse ja kui ei kuulu lisab isiku klubitude hulka. Selles ülesandes olid

kõikidel mudelitel samad vead. Kõigil jäi trigeri funktsiooni loogikas puudu UPDATE käsk, millega oleks isikule määratud vajalik klubi “klubitud”, kui see neil puudus. Automaatkontroll ei tuvastanud läbivaatusel ühtegi sellist isikut, seega ei saadud maksimaalset tulemust.

#### 4.1.7 Kodutööde tulemuste koondvaade

Jooniselt (vt joonis 2) tuleb välja, et keelemudelite tulemused olid üsna stabiilselt üle poolte punktide ehk 1 punkti piiri. See visualiseerib ka eelinformatsiooni abiks olemist, sest oli näha, et mudelid koos selle infoga ei saavutanud üheski kodutöös tulemust alla poolte punktide.



Joonis 2. Keelemudelite tulemused kodutööde kaupa

Jooniselt on ka näha, et kõige madalama skoori saavutas keelemudelistest eelinformatsioonita Copilot, mis saavutas kahes kodutöös 0 punkti ning oli veel ühes alla poolte punktide piiri.

## 4.2 Keelemudelite tulemuste võrdlus üliõpilaste tulemustega

Teine uurimisküsimus “Kui hästi suudavad keelemudelid lahendada andmebaasi koduülesandeid võrreldes üliõpilaste tegelike tulemustega?” käsitles keelemudelite ChatGPT ja Microsoft Copilot võimekust lahendada andmebaasi koduülesandeid võrreldes üliõpilaste saavutatud tegelike tulemustega. Selleks võrreldi mõlema keelemudeli nii eelinfota kui ka eelinfoga saadud tulemusi üliõpilaste 2025. aasta kevadsemestri keskmiste sooritustega kuues kodutöös (vt tabel 7). Kõikides kodutöödes üliõpilaste tulemuste minimaalne väärtus oli 0 ja maksimaalne 2.

Tabel 7. Keelemudelite ja üliõpilaste tulemused kodutöödes

Kodutööd	ChatGPT eelinfota	Copilot eelinfota	ChatGPT eelinfoga	Copilot eelinfoga	Üliõpilaste aritmeetiline keskmine tulemus (standardhälve)
<b>Kodutöö 1</b>	1,5 p	1,5 p	2 p	2 p	1,99 p (0,1)
<b>Kodutöö 2</b>	1,6 p	0 p	1,7 p	1,8 p	1,95 p (0,26)
<b>Kodutöö 3</b>	2 p	1,7 p	1,79 p	1,79 p	1,96 p (0,25)
<b>Kodutöö 4</b>	0 p	0 p	2 p	1,88 p	1,92 p (0,27)
<b>Kodutöö 5</b>	1,2 p	1 p	1,7 p	1,7 p	1,84 p (0,33)
<b>Kodutöö 6</b>	1,80 p	1,80 p	1,80 p	1,80 p	1,85 p (0,33)
<b>Kokku</b>	8,1 p	6 p	10,99 p	10,97 p	11,51 p (0,24)

Kõikide kodutööde puhul oli üliõpilaste mediaaniks 2 punkti, mis tähendab, et vähemalt pooled õppuritest saavutasid igas ülesandes maksimaalse tulemuse. Standardhälbed jäid vahemikku 0,1–0,33, viidates tulemuste tugevale koondumisele keskmise väärtuse ümber. Üliõpilaste seas oli ka alati esitlus, mille eest teeniti 0 punkti, seega iga kodutöö minimaalne väärtus oli sama. Seega oli otstarbekas võrrelda keelemudelite tulemusi üliõpilaste mediaani, minimaalsete tulemustega ning standardhälvega aritmeetilisest keskmisest tulemustest.

Andmebaasi eelinfota jäid keelemudelid mitmel juhul alla üliõpilaste minimaalsetele tulemustele. Näiteks kodutöödes kaks ja neli said mõlemad mudelid tulemuseks 0 punkti,

kuigi üliõpilaste madalaimad arvestatud tulemused olid sama madalad siis nende mediaantulmus oli endiselt 2 punkti. See annab teadmise, et eelinformatsioonita keelemudelid saavutasid madalamad tulemused kui üliõpilased.

Kui mudelile anti ette andmebaasi struktuur, paranesid tulemused märgatavalt. Kuigi keelemudelite lahendused, olenemata eelinformatsiooni olemasolust, jäid kõikides kodutöodes üliõpilaste tulemusvahemiku sisse. Mitmes ülesandes (nt KT1 ja KT4) küündisid need ka mediaanini ehk maksimaalse tulemuseni või edestasid üliõpilaste aritmeetilist keskmist. Selline sooritus tähendab, et õige kasutamise korral võiksid keelemudelid toetada õppijaid, mis muidu jääksid alla mediaani ning koostöös oleks võimalik saavutada ilmselt parem tulemus.

Kõige suurem vahe üliõpilaste ja keelemudelite vahel tuli välja neljanda kodutöö puhul, kus eelinfota saavutasid mõlemad mudelid null punkti, kuid üliõpilased saavutasid keskmiseks tulemuseks 1,92 punkti ning standardhälve oli 0,27. Kuuendas kodutöös saavutasid kõik mudelid tulemuseks 1,8 p, jäädes nii üsna lähedale üliõpilaste tulemuste mediaanile ning mudelite tulemused olid tugevasti üle ühe punkti paremad üliõpilaste miinimumtulemusest.

Kokkuvõtlikult võib öelda, et keelemudelid suudavad saavutada tulemusi, mis mõnes kodutöös ületavad või lähenevad üliõpilaste keskmisele, aga üldiselt olid nende tulemused ebaühtlased. Kuid parima tulemuse saavutamiseks vajavad need tööriistad selget sisendit ja tulemuste kontrollimist.

## 5. Arutelu

Käesoleva bakalaureusetöö eesmärk oli hinnata keelemudelite (ChatGPT ja Microsoft Copilot) võimekust lahendada Tartu Ülikooli „Andmebaasid“ kursuse kodutöid ning võrrelda nende tulemusi üliõpilaste omadega.

Bakalaureusetöö esimene uurimisküsimus oli, “Millised on keelemudelite poolt lahendatud "Andmebaasid" kursuse koduülesannete tulemused, kui kasutada hindamiseks automaatkontrolli skoori ning kuidas erinevad tulemused sõltuvalt sellest, kas andmebaas oli ette antud SQL kujul või mitte?”. Uuringu tulemused näitasid, et kui keelemudeid saavad enne päringu genereerimist ülevaate andmebaasi skeemist, tõuseb nende lahenduste keskmine skoor märgatavalt, kuigi tüüpilised vead, nagu vale arvutusloogika, või kitsenduste rakendamine esinesid endiselt. ChatGPT suutis skeemi olemasolul oma tulemust ligikaudu veerandi võrra parandada, ning Copilot parandas enda tulemust üle kahe korra. See tulemus oli loogiline ja kooskõlas varasemate töödega, kus keelemudelite edukus oli otseselt seotud neile antava kontekstiga [20, 21, 33]. Sellised vead eristavad keelemudelite lahendusi üliõpilaste omadest, kuna nemad suudavad sageli läbi vigade analüüsimise neid kohti vältida. Õppejõud võiks seda arvesse võtta ülesannete koostamisel ja hindamisel. Näiteks jättes osa kontekstist välja või sõnastades tingimusi mitmeti mõistetavalt, saab vigade korral tuvastada, kas tegemist oli keelemudelilt saadud koodiga. Sarnasele tulemusele jõudis ka Liu jt [37]. Samuti aitavad tuvastada keelemudelite loodud koodi valede veerunimedega kasutamine, juhul kui üliõpilased jätavad selle parandamata. See tuli välja eelinformatsioonita keelemudelite lahendustest, kus tegid mudelid palju vigu veeru- ja tabelinimedega kasutamisel. Need olid vead, mis loeti piisavalt kergeteks, et tudengid märkaksid neid ja automaatkontrollis selliseid vigu ei hinnatud.

Teine uurimisküsimus “Kui hästi suudavad keelemudelid lahendada andmebaasi koduülesandeid võrreldes üliõpilaste tegelike tulemustega?”, mis tõi välja keelemudelite taseme võrdluses üliõpilastega. Kui keelemudelite saadud tulemusi võrreldi üliõpilaste keskmistega, ilmnes, et ka parima tulemusega andmebaasi eelinformatsiooniga ChatGPT koondtulemus jäi siiski ligikaudu ühe punkti võrra alla üliõpilaste keskmisele ja langes mediaanist kaugemale keerukamate ülesannete korral. See näitab, et keelemudelid küündivad üldjoontes keskmise üliõpilase tasemeni, kuid nende lahenduste kvaliteet sõltub ülesande raskusastmest ja konteksti olemasolust. Uuringu põhjal saab öelda, et keelemudelite tulemused olid pigem nõrgemad või võrdväärseid keskmise üliõpilase tulemusega. Samale

järeldusele on jõutud ka varasemates uuringutes [34, 35]. Ka teistes valdkondades, näiteks matemaatika õppimisel, oli leitud, et keelemudelid saavad hästi hakkama lihtsamate ülesannetega, kuid keerulisemates olukordades tekivad sagedamini vead või ebatäpsed lahendused [1, 6, 7].

## Kokkuvõte

Käesoleva bakalaureusetöö eesmärk oli hinnata tehisintellektil põhinevate keelemudelite, ChatGPT ja Microsoft Copiloti, võimekust lahendada Tartu Ülikooli „Andmebaasid“ kursuse kodutöid ning võrrelda nende tulemusi üliõpilaste omadega. Selleks lahendati kursuse kodutööd keelemudelitega läbi kahel viisil: esmalt andmebaasi eelinformatsioonita ning seejärel andes ette ka olemasolev informatsioon andmebaasi kohta. Saadud tulemusi võrreldi üliõpilaste omadega.

Eelinfo puudumisel esines rohkem sisulisi ja loogilisi vigu, nagu valede veerunimede kasutamine või vale tabeliloogika. Samas esines ka üksikuid erandeid, kus keelemudel suutis ilma struktuurita ülesande edukalt lahendada. Tulemused näitasid, et keelemudelite tulemused paranesid, kui neile anti ette andmebaasi struktuur koos tabelleid siduva loogikaga. Võrdluses üliõpilastega jäi keelemudelite keskmine tulemus üldjuhul alla üliõpilaste tulemustele, seda eriti keerulisemates ülesannetes. Lihtsamates ülesannetes saavutati aga mõnikord maksimaalne punktisumma.

Et õppejõud saaksid vähendada keelemudelite efektiivsust enda kursustel, oleks vaja ülesandeid formuleerida viisil, mis sisaldab mitmetähenduslikke või kontekstist sõltuvaid sõnu. Näide sellest oli kolmandas kodutöös esinenud tingimuse kontroll, kus perekonnanimi pidi olema “Kalamees”, mis eksitas mudeleid. Uuringu tulemusena oli näha, et üliõpilased suudavad saavutada kõrgemaid ja stabiilsemaid tulemusi, kui keelemudelid. Samas keelemudelite efektiivne kasutamine juhul kui seda võib teha eeldab, et õppija oleks ise eelnevalt läbi töötanud ülesande eelduseks oleva andmebaasi struktuuri ning suudab vajadusel mudeli antud lahendusi täiendada või parandada.

Töö üheks piiranguks oli asjaolu, et tulemused põhinevad 2025. aasta kevadsemestri üliõpilaste andmetel ning keelemudelite versioonid arenevad kiiresti. Keelemudelite kiire areng võib tulevikus kaasa tuua paremaid tulemusi Seetõttu tuleks antud tulemusi käsitleda kui hetkeseisu kaardistust, mis vajab tulevikus värskendamist kasutades uuendatud mudeleid. Edasistes uuringutes soovitaksin kindlasti jälgida viipade ülesehitust, mida keelemudelitele anda. Näiteks kui oleks tarvis, et keelemudel unustaks varasema konteksti, siis tuleb seda eraldi viibana öelda.

## Viidatud kirjandus

- [1] Ellis, M., Casey, K., Hill, G. ChatGPT and Python programming homework. *Decision Sciences Journal of Innovative Education*. 2024. no. 22. [10.1111/dsji.12306](https://doi.org/10.1111/dsji.12306).
- [2] Luik, P., Keivabu, C. and Orav-Puurand, K. (2025). Using ChatGPT 3.5 to Reformulate Word Problems for State Exam in Mathematics. In Proceedings of the 17th International Conference on Computer Supported Education - Volume 1: CSEDU; ISBN 978-989-758-746-7; ISSN 2184-5026, SciTePress, pages 180-190. DOI: 10.5220/0013152200003932
- [3] Ramos, R. F., De Angel, R. M., Ruetas, A. P., Lagman, A. C., Morano, J., Payongayong, R. T. Effective Lesson Planning and Assessment Design Using Leveraging Microsoft Copilot Implementation. 2024 IEEE 15th Control and System Graduate Research Colloquium (ICSGRC), SHAH ALAM, Malaysia, 2024, pp. 331-336, doi: [10.1109/ICSGRC62081.2024.10691289](https://doi.org/10.1109/ICSGRC62081.2024.10691289).
- [4] Hadi, M. U., Al Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Hassan, S. Z., Shoman, M., Wu, J., Mirjalili, S., & Shah, M. Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects. 2023. <https://doi.org/10.36227/techrxiv.23589741.v7>.
- [5] Baidoo-Anu, David & Ansah, Leticia. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. 2023. pp. 52-62. [10.61969/jai.1337500](https://doi.org/10.61969/jai.1337500).
- [6] Anderson. N., McGowan. A., Hanna. P., Cutting. D., Galway. L., Collins. M. Using ChatGPT in Software Development Education. Belfast, Suurbritannia: *School of Electronics, Electrical Engineering and Computer Science Queen's University Belfast*. 2024. pp. 1-5, doi: <http://dx.doi.org/10.1109/EDUCON60312.2024.10578808>.
- [7] Lucas, J., Madyatmadja, E. D., Syahchari, D. H. Analyzing the Impact of ChatGPT on Enhancing Academic Skills and Learning Concepts Among University Students in Jakarta. Balikpapan, Indonesia. 2024. pp. 254-258. [10.1109/ICSINTESA62455.2024.10748220](https://doi.org/10.1109/ICSINTESA62455.2024.10748220).

- [8] Xue, Y., Chen, H., Bai, G. R., Tairas, R., Huang, Y. Does ChatGPT Help With Introductory Programming? An Experiment of Students Using ChatGPT in CS1. 2024. pp. 331 - 341. <https://doi.org/10.1145/3639474.3640076>.
- [9] Giannakopoulos, K., Kaklamanos, E. G., & Makrygiannakis, M. A. Evidence-based potential of generative artificial intelligence large language models in orthodontics: a comparative study of ChatGPT, Google Bard, and Microsoft Bing. *European Journal of Orthodontics*. 2024. 46. <https://doi.org/10.1093/ejo/cjae017>.
- [10] OpenAI. (2024). ChatGPT 4o. <https://chat.openai.com>. (12.05.2025)
- [11] Google Gemini. (2024). Gemini 2.0 Flash. <https://gemini.google.com/app>. (12.05.2025)
- [12] DeepSeek. (2024). DeepSeek AI keelemudel. <https://www.deepseek.com/>. (12.05.2025)
- [13] Kosar, T., Ostojić, D., Liu, Y.D., Mernik, M. Computer Science Education in ChatGPT Era: Experiences from an Experiment in a Programming Course for Novice Programmers. 2024. no. 12. p. 629. <https://doi.org/10.3390/math12050629>.
- [14] Banerjee, P, Srivastava, A, Adjeroh, D, Reddy, Y., Karimian, N. Understanding ChatGPT: Impact Analysis and Path Forward for Teaching Computer Science and Engineering. 2023. [10.36227/techrxiv.22639705](https://arxiv.org/abs/10.36227/techrxiv.22639705).
- [15] Zirar, A. Exploring the impact of language models, such as ChatGPT, on student learning and assessment. Huddersfield, Suurbritannia: *Department of Management, Huddersfield Business School, University of Huddersfield*. 2023. <https://doi-org.ezproxy.utlib.ut.ee/10.1002/rev3.3433>.
- [16] Kait Krull. Suunis tehisintellekti kasutamiseks õppetöös. <https://ut.ee/et/sisu/suunis-tehisintellekti-kasutamiseks-oppetoos>. (01.05.2025)
- [17] Islam, S. O. ul, Lauscher, A., Glavaš, G. How Much Do LLMs Hallucinate across Languages? On Multilingual Estimation of LLM Hallucination in the Wild. 2025. <https://doi.org/10.48550/arXiv.2502.12769>.
- [18] Gökhan, A., Elif, S. AI chatbots in programming education: guiding success or encouraging plagiarism. *Discov Artif Intell* 4. 2024. Article 87. <http://dx.doi.org/10.1007/s44163-024-00203-7>.

- [19] Farrokhnia, M., Banihashem, S. K., Noroozi, O., Wals, A. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*. 2023. 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>.
- [20] Sun, D., Boudouaia, A., Zhu, C. et al. Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study. *International Journal of Educational Technology in Higher Education* 21, no. 14. 2024. <https://doi.org/10.1186/s41239-024-00446-5>.
- [21] Sandu, R., Gide, E., Elkhodr, M. The role and impact of ChatGPT in educational practices: insights from an Australian higher education case study. *Discov Educ* 3. no. 71. 2024. <https://doi.org/10.1007/s44217-024-00126-6>.
- [22] Budhiraja, R., Joshi, I., Challa, J.S., Akolekar H.D., Kumar, D. 2024. "It's not like Jarvis, but it's pretty close!" - Examining ChatGPT's Usage among Undergraduate Students in Computer Science. *ACE '24: Proceedings of the 26th Australasian Computing Education Conference*. 2024. p. 124–133. <https://doi-org.ezproxy.utlib.ut.ee/10.1145/3636243.3636257>.
- [23] Qadir, J. Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. *IEEE Global Engineering Education Conference (EDUCON)* IEEE. 2023. pp. 1–9. [10.36227/techrxiv.21789434](https://doi.org/10.36227/techrxiv.21789434).
- [24] Li, Y., Zhang, T., Luo, X., Cai, H., Fang, S., Yuan, D. Do Pretrained Language Models Indeed Understand Software Engineering Tasks?. *IEEE Transactions on Software Engineering*. 2024. vol. 49. no. 10. pp. 4639-4655. <https://doi.org/10.48550/arXiv.2211.10623>.
- [25] Haque, Md. A. LLMs: A Game-Changer for Software Engineers? 2024. <https://doi.org/10.48550/arxiv.2411.00932>.
- [26] Tartu Ülikooli Õppeinfosüsteem. Andmebaasid (6 EAP). <https://ois2.ut.ee/#/courses/LTAT.03.004/details>. (22.03.2025)
- [27] PostgreSQL: About. <https://www.postgresql.org/about/>. (05.04.2025)
- [28] Ünal, H. T., Mete, S., Vurgun, Ö. U., Mendi, A. F., Özkan Ö., Nacar, M. A. PostgreSQL Database Management System: ODAK. *Innovations in Intelligent Systems and Applications*

Conference (ASYU). Sivas, Turkiye. 2023. pp. 1-5. doi: 10.1109/ASYU58738.2023.10296600.

[29] DB-Engines. DB Engines Ranking, 2025. Available: <https://dbengines.com/en/ranking>. (05.04.2025)

[30] DBeaver Community. <https://dbeaver.io/>. (05.04.2025)

[31] Kakk, M. Automaatkontrollide arendus õppetöö tagasisidestamiseks aines Andmebaasid. TÜ Matemaatika- ja informaatikaõpetaja õppekava magistritöö. 2023. <https://thesis.cs.ut.ee/4e327393-d301-472f-878c-17bcf8b74840>. (12.04.2025).

[32] E-õppija käsiraamat <https://eoppija.ut.ee/tekstiroboti-kasutamisest/>. (01.05.2025)

[33] Wang, C., Sui, D., Sun, H., Ding, H., Zhang, B., Tu, Z. Plug-and-Play Performance Estimation for LLM Services without Relying on Labeled Data. 2024. <https://doi.org/10.48550/arXiv.2410.07737>.

[34] Kaimre, J. ChatGPT kasutamine tarkvaraarenduse teadmiste testide lahendamisel. TÜ Tarkvaratehnika õppekava magistritöö. 2024. <https://thesis.cs.ut.ee/88905026-ead5-4922-a6a7-d7e3776c19fb>. (01.05.2025)

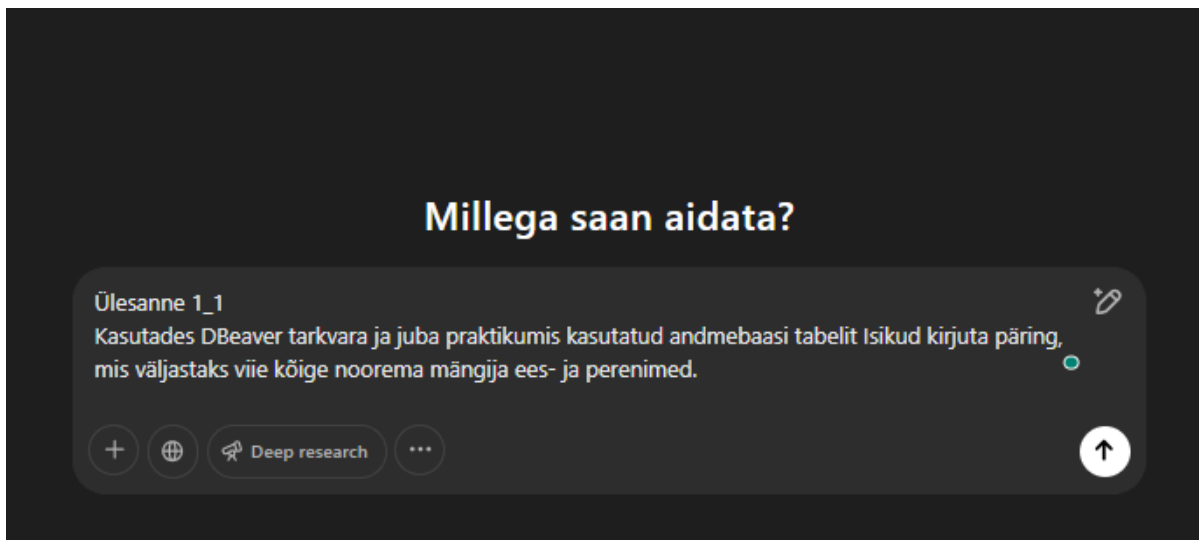
[35] Farinetti, L., Cagliero, L. A Critical Approach to ChatGPT: An Experience in SQL Learning. In Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1. Association for Computing Machinery, New York, NY, USA, 2025. pp 318–324. <https://doi.org/10.1145/3641554.3701932>.

[36] Tartu Ülikooli Moodle. <https://moodle.ut.ee>. (13.05.2025)


[37] Liu, N., Sonkar, S., Baraniuk, R. G. *Do LLMs Make Mistakes Like Students? Exploring Natural Alignment Between Language Models and Human Error Patterns*. (2025). <https://doi.org/10.48550/arXiv.2502.15140>.


## Lisad


### I. Ülesande näidisviip ChatGPT-s





## II. Ülesande näidisviip Copilot-is





 **Tulemuse kuvamine**  
Käivita see kood ja selgita seda...


 **Teadliku otsuse tegemine**  
Loo tabel, mis võrdleb enim müüdnud kontoriprintereid

 **Dekodeerimisakronüümid**  
Mida [LLM] tähendab?



 **Oma teadmiste  
võimendamine**  
Aita õppida [teema] kohta

 **Uurimistöo alustamine**  
Kuidas on maailma kirjaoskuse määrad viimase 30 aasta jooksul muutunud?

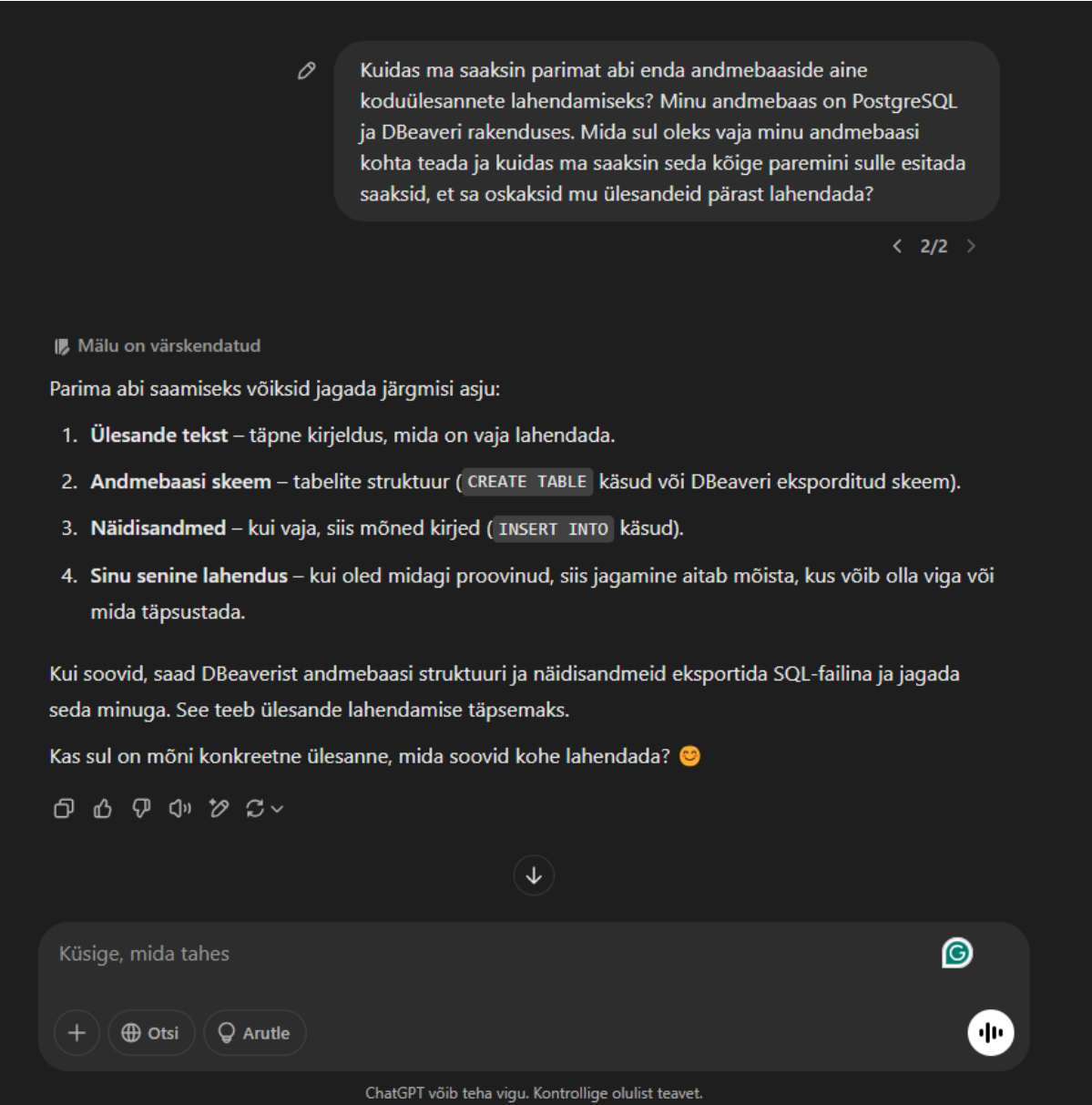
 **Trendide võrdlus**  
Saate võrrelda 10 kõige uuenduslikumat ja loominguilisemat töökohakujundust

 Kuva juhised

Ülesanne 1\_1  
Kasutades DBeaver tarkvara ja juba praktikumis kasutatud andmebaasi tabelit Isikud kirjuta päring, mis väljastaks viie kõige noorema mängija ees- ja perenimed.

### III. Eelinfo keelemudelile



Kuidas ma saaksin parimat abi enda andmebaaside aine koduülesannete lahendamiseks? Minu andmebaas on PostgreSQL ja DBeaveri rakenduses. Mida sul oleks vaja minu andmebaasi kohta teada ja kuidas ma saaksin seda kõige paremini sulle esitada saaksid, et sa oskaksid mu ülesandeid pärast lahendada?

< 2/2 >

Mälu on värskendatud

Parima abi saamiseks võiksid jagada järgmisi asju:

1. **Ülesande tekst** – täpne kirjeldus, mida on vaja lahendada.
2. **Andmebaasi skeem** – tabelite struktuur (`CREATE TABLE` käsud või DBeaveri eksporditud skeem).
3. **Näidisandmed** – kui vaja, siis mõned kirjed (`INSERT INTO` käsud).
4. **Sinu senine lahendus** – kui oled midagi proovinud, siis jagamine aitab mõista, kus võib olla viga või mida täpsustada.

Kui soovid, saad DBeaverist andmebaasi struktuuri ja näidisandmeid eksportida SQL-failina ja jagada seda minuga. See teeb ülesande lahendamise täpsemaks.

Kas sul on mõni konkreetne ülesanne, mida soovid kohe lahendada? 😊

📄 👍 🗨️ 🔊 ✍️ ↻

↓

Küsi, mida tahes

+ 🌐 Otsi 💡 Arutle 🔊

ChatGPT võib teha vigu. Kontrollige olulist teavet.

## IV. Keelemudeli antud lahendus

Ülesanne 1\_1  
Kasutades DBeaver tarkvara ja juba praktikumis kasutatud andmebaasi tabelit Isikud kirjuta päring, mis väljastaks viie kõige noorema mängija ees- ja perenimed.








Siin on SQL-päring, mis tagastab viie kõige noorema mängija ees- ja perenimed:


```
sql Kopeeri Redigeeri  
  
SELECT eesnimi, perenimi  
FROM public.isikud  
ORDER BY synniaeg DESC  
LIMIT 5;
```


**Selgitus:**





- `ORDER BY synniaeg DESC` – järjestab mängijad sünniaja järgi kahanevalt (noorimad ees).
- `LIMIT 5` – tagab, et tagastatakse ainult viis noorimat mängijat.

Kui soovid päringut testida DBeaveris, kopeeri see SQL-i vaatesse ja käivita! 😊 Anna teada, kui vajad täiendusi!



Küsi, mida tahes 

  Otsi  Arutle 

ChatGPT võib teha vigu. Kontrollige olulist teavet.

## V. Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Joosep Lember**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Kursuse “Andmebaasid” kodutööde lahendamine keelemudelitega võrdluses üliõpilaste tulemustega**, mille juhendaja on **Piret Luik**, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Joosep Lember*

**07.05.2025**