

TARTU ÜLIKOOL

Sotsiaalteaduste valdkond

Ühiskonnateaduste instituut

Ajakirjanduse ja kommunikatsiooni õppekava

Volli Soasepp

Generatiivsete tehisintellektide võimekus kontrollida väidet, võrreldes
Delfi Faktikontrolliga

Bakalaureusetöö

Juhendaja: Diana Poudel

Tartu 2025

SISUKORD

SISSEJUHATUS	4
1. TÖÖ TEOREETILISED JA EMPIIRILISED LÄHTEKOHAD	6
1.1 Tehisintellekt.....	6
1.2 Faktikontroll.....	12
1.3 Generatiivsete tehisintellektide tuntud mudelite ülevaate.....	14
2. UURIMISKÜSIMUSED	19
3. EMPIIRILISE UURINGU KAVA	20
3.1 Kasutatavad andmekogumis- ja analüüsimeetodid.....	20
3.2 Plaanitav valim.....	21
3.3 Algne hindamismaatriks.....	21
3.4 Meetodikatsetus.....	23
3.4.1 Pilootuuring.....	24
3.4.2 Uuring.....	29
4. UURIMISTÖÖ TULEMUSED	31
4.1 Faktidele toetumine.....	33
4.2 Usaldusväärsed allikad.....	34
4.3 Viidete kasutamine.....	36
4.4 Läbipaistvus.....	38
4.5 Erapooletus.....	39
4.6 Põhjalikkus.....	40
4.7 Aktuaalsus.....	42
4.8 Faktikontrolli tulemused.....	43
4.9 Järeldused.....	45
5. DISKUSSIOON	47
KOKKUVÕTE	49
KASUTATUD KIRJANDUS	52
Lisa 1. DeepSeek V3 faktikontroll	62

Lisa 2: Gemini 2.0 Flash faktikontroll.....	65
Lisa 3. Grok 3 faktikontroll.....	67
Lisa 4. Claude 3.7 Sonnet faktikontroll.....	71
Lisa 5. ChatGPT 4o faktikontroll.....	73
Lisa 6. Lihtlitsents lõputöö reprodutseerimiseks.....	77

SISSEJUHATUS

Uurimistöö eesmärgiks on välja selgitada, millisel määral suudab tehisintellekti keelemudel läbi viia täiemahulist faktikontrolli autori poolt esitatud sisendi (ingl. *prompt*) põhjal. Siinse töö autorile teadaolevalt on Eesti faktikontrollijate seas ühine arusaam, et tehisintellekt on tööriist, mis aitab faktikontrolli ajal infot töödelda ja analüüsida, kuid ei suuda iseseisvalt kvaliteetselt ja järjepidevalt väiteid kontrollida. Varasemates uurimistöodes on välja toodud, et manuaalne faktikontroll ei suuda sammu pidada laialdase valeinfo levikuga (Nikopensius, 2022). Uurimistöös uuritakse, kas tehisintellekt suudab olla enam kui tööriist ja oskab iseseisvalt informatsioon leiduvat informatsiooni kinnitada või ümber lükata.

Keelemudelid ja nende võimekused arenevad iga mööduva päevaga (Kaplan ja Haenlein, 2019). Kui kümme aastat tagasi meenus tehisintellekti mainides Apple'i Siri või Amazoni Alexa, siis tänapäeval on tehisintellekti mõiste palju laiem. Tehisintellektist on abi tervishoiust ja transpordist kuni hariduse, meelelahutuse ja äriini välja. Kiire arengu tõttu kuulutas Eesti president Alar Karis, et 2025. aasta sügiseks on plaan tagada ka ligipääs juhtivatele tehisintellektil (OpenAI, Anthropic, Google ja Microsoft) põhinevatele haridusrakendustele 20 000 õpilase jaoks (TI-Hüpe, 2025). Nende põhjuste tõttu on kiirelt arenevas informatsioonil oluline mõista tehisintellekti võimekusi ja piiranguid.

Populaarsed keelemudelid on tuntud ka enda hallutsineerimise poolest ehk need lisavad vastusesse infot, mida algallikas tegelikult ei sisalda, mis võivad takistada nende usaldusväärsust (Hughes, 2024). Selle jaoks tuleb välja selgitada, milline keelemudel on andmekogumises ja -analüüsimises kõige usaldusväärsem ja mis mudeli usaldusväärsust mõjutab. Samal ajal on kiiresti arenenud ka faktikontrolli valdkond, kuhu on viimase kümne aasta jooksul lisandunud sadu faktikontrolli organisatsioone (Stencel jt, 2024). Eestis on faktikontrolli eesmärgiks jälgida avalikus informatsioonil avaldatud väidete täpsust ja võidelda desinformatsiooni vastu (Eesti Päevaleht, 2018; International Fact-Checking Network, i.a). Euroopa komisjon määratleb desinformatsiooni kui vale, ebatäpset või eksitavat teavet, mida levitatakse teadlikult (EEAS, 2018). Samuti arvatakse, et desinformatsiooni üks eesmärgid on ühiskonnas eksisteerivaid lõhesid suurendada (Wong, 2024).

Bakalaureusetöö koosneb viiest osast: teoreetilistest ja empiirilistest lähtekohtadest, uurimisküsimustest, empiirilise uuringu kavast, tulemustest ja diskussioonist. Esimene peatükk tutvustab ajakirjanduslikku faktikontrolli, tehisintellekti ja selle erinevaid keelemudeleid. Peatükis kirjeldab autor, kuidas tehisintellekt sündis ja kuidas keelemudelid üksteisest erinevad. Järgmises peatükis esitleb autor uurimisküsimusi, millele vastuseid otsitakse. Kolmandas peatükis esitletakse uurimistöö meetodit kasutatavat andmekogumis- ja analüüsimeetodit. Neljandas peatükis näidatakse uurimistöö tulemusi ja järeldusi. Viimases peatükis arutleb autor uuringutulemuste, meetodi ja uurimisküsimuste üle.

Töö eripära seisneb autori poolt koostatud hindamismaatriksis (Tabel 5), mille eesmärk on struktureeritult hinnata tehisintellekti faktikontrolli kvaliteeti. Autorile teadaolevalt ei ole varasemates teadusuuringutes või -artiklites kasutatud hindamismaatriksit, mis võimaldaks võrrelda tehisintellekti keelemudelite vastuseid eksperthinnanguga. Tegemist on originaalse lähenemisega, mille eesmärk on hinnata keelemudeleid kaheksa kategooria põhjal: faktidele toetumine, allikate usaldusväärsus, viidete kasutamine, läbipaistvus, erapooletus, käsitluse põhjalikkus, aktuaalsus ja faktikontrolli tulemus.

1. TÖÖ TEOREETILISED JA EMPIIRILISED LÄHTEKOHAD

Tehisintellekti teoreetilised alused pandi paika eelmise sajandi keskpaigas ning esimeste praktiliste katsetusteni jõuti 1960. aastatel (Toosi jt, 2021; Koit, 2015). Kui algselt keskenduti sellele, kuidas masinad saaksid jäljendada inimese mõtlemist, siis tänapäeval on tehisintellekt laienenud erinevatesse eluvaldkondadesse ning muutunud paljuski nähtamatuks, kuid samas igapäevaseks abiliseks. Alates tervishoiust, transpordist ja haridusest kuni meelelahutuse, ärimaailma ja eetiliste küsimusteni. Tehisintellekti kasutamine on laiendanud tehnoloogia rolli valdkondades nagu tööprotsesside automatiseerimine, otsuste vastuvõtmine ning teatud loovülesannete täitmine, mida varem peeti tööks inimesele. Tehisintellekti arenguga on viimastel aastatel kasvanud ka faktikontrolli valdkond, kuhu on lisandunud sadu faktikontrolli organisatsioone (Stencel, 2024). Käesolevas peatükis annan ülevaate, kuidas tehisintellekt on oma koha leidnud erinevates eluvaldkondades ja kuidas näeb välja faktikontrollime Eestis.

1.1 Tehisintellekt

Tehisintellekti (TI) tähendus ja selle võimed on viimaste aastatega palju muutunud. Esimesed märgid tehisintellektist on juba aastast 1950, kui inglasest matemaatik Alan M. Turing kirjutas artikli nimega „Arvutusmasinad ja intellekt” (Turing, 1950). Turing uskus, et inimese intellekti on võimalik arvuti abil modelleerida (Turing, 1950). Ta pakkus välja testi, mille eesmärgiks oli otsustada, kas arvutil on intellekt või mitte: inimesest kohtunik esitab telegraafiaaparaadi kaudu inglise keeles küsimusi kahele katsealusele, et määrata, kumb on arvuti ja kumb inimene (Koit, 2015).

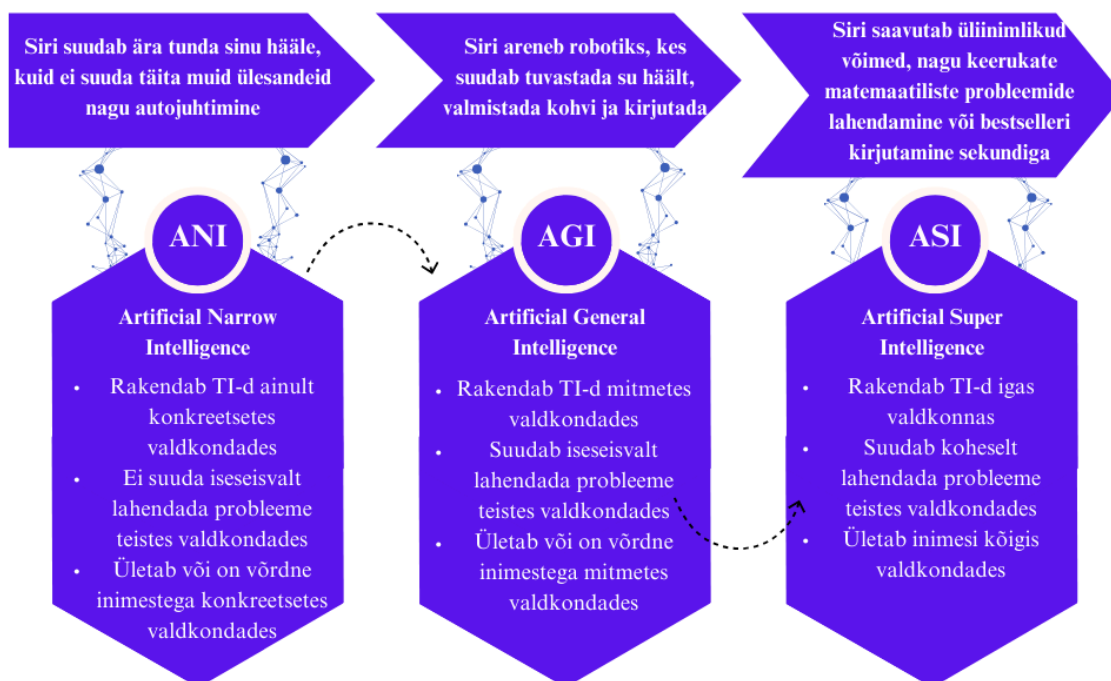
Arvuti intelligentsuses veenduti siis, kui inimesest kohtunik ei suuda viie minuti jooksul otsustada, kumb on kumb. Üksteist aastat hiljem 1961 avaldas ameeriklasest psühholoog Marvin Minsky teedrajava artikli nimega „Sammud tehisintellekti poole” ja aastal 1963 Massachusettsi tehnoloogiainstituudis (MIT) koostas James Slagle arvutiprogrammi SAINT, mis loodi integraalarvutuste ülesannete lahendamiseks, mis näitas esimesena, et arvuti suudab teha midagi, mida enne suutis teha vaid inimene (Koit, 2015; Toosi jt, 2021). Esimene vestlusrobot tehisintellekti ajaloos loodi Joseph Weizenbaumi poolt MIT-is 1966. aastal (Toosi jt, 2021). Vestlusrobot ELIZA töötati välja virtuaalse terapeudina, et esitada küsimusi ja pakkuda vastuseks järelküsimusi patsiendile (Salecha, 2016).

Tehisintellekti mõistmiseks on vaja üle vaadata olulised mõisted. Läbi ajaloo on tehisintellekti defineeritud erinevat moodi. Tavaline ja üldkasutatud viis tehisintellekti defineerimiseks, on viitamine inimintellektile, mille peamine printsiip on lahendada probleeme või luua tooteid, millel on kultuuriline väärtus (Gardner, 1999: 33—34). Dartmouth Research Project defineeris tehisintellekti kui probleemi, mille eesmärk on: "panna masin käituma viisil, mida peetaks intelligentseks, kui inimene käituks samamoodi" (McCarthy jt, 1955). Teadlane Marvin Minsky pidas tehisintellekti "teaduseks, mis tegeleb masinate selliseks programmeerimisega, et need teeksid asju, mis nõuaksid intelligentsust, kui neid teeksid inimesed" (Minsky, 1968).

Eesti keele seletav sõnaraamat defineerib tehisintellekti kui modelleeritud ajuprotsessidest tulenevat arvuti suutlikkust jäljendada inimese vaimset tegevust või kui arvutiteaduse ja -tehnika haru, mis uurib ajuprotsesside modelleerimist elektronarvutil ja vastavate arvutisüsteemide loomise meetodeid (EKSS, 2009). Lühidalt on definitsioone ja õigeid vastuseid mitmeid, kuid selles uurimistöös keskendutakse tehisintellektile kui tehiskeelele, mis on esitatud ja loodud arvuti poolt.

Kaplani ja Haenleini (2019) järelduse järgi saab tehisintellekti liigitada analüütiliseks, inimesest inspireerituks ja inimlikuks tehisintellektiks, sõltuvalt sellest, millist tüüpi intelligentsust see näitab (kognitiivne, emotsionaalne ja sotsiaalne intelligentsus), või kitsaks, üldiseks ja superintellektiks, lähtudes selle evolutsioonilisest staadiumist. Kitsas tehisintellekt on inimesest kehvem ja nõrgem, üldine tehisintellekt on inimese tasemel ja tugev ning superintellekt on eneseteadlik, inimesest võimsam intellekt (Joonis 1). Eneseteadlik viitab tehisintellekti kontekstis arenenud tehisarule, mis suudab mõista omaenda seisundit ja rolli ümbritsevas keskkonnas (Kaplan ja Haenlein, 2019).

Tehisintellekti kolm etappi



Joonis 1. Tehisintellekti kolm etappi. *Allikas: Kaplan ja Haenlein, 2019; autori poolt tõlgitud ja kujundatud.*

Kaplani ja Haenleini loodud idee avalikustati 2019. aastal. Sellest ajast saati on tehisintellekt märkimisväärselt kiiresti arenenud, viies meid ajastusse, kus tehisintellektil põhinevad programmid on muutunud igapäevaseks osaks erinevates eluvaldkondades. Kui veel enne 2022. aastat olid kõige tuntumad tehisintellekti esindajad Apple'i Siri või Amazoni Alexa, mis suutsid mõista inimehõnet ja vastata erinevatele küsimustele, siis tänapäeval kasutatakse tehisintellekti abi tervishoiust ja transpordist kuni hariduse, meelelahutuse, äri, eetika ja privaatsuseni välja.

Tervishoius uurivad teadlased, kuidas tehisintellekti abiga saab analüüsida suurt hulka terviseandmeid, et leida tee uute meditsiiniliste avastusteni ja parandada individuaalset diagnostikat (Euroopa Parlament, 2023). Lisaks on välja töötatud hädaabi kõnedele vastamise programm, mis suudab tuvastada südame seiskumist kiiremini kui inimene (Euroopa

Parlament, 2023). Uuringud Lõuna-Koreas on näitavad, et tehisintellekti abil tehtud diagnoos oli parem varajase rinnavähi avastamisel (91%) kui radioloogide diagnoos (74%) (Kim jt, 2020). Samuti kasutatakse süvaõppe algoritme kopsupõletiku tuvastamiseks rindkere radiograafial, saavutades tundlikkuse ja spetsiifilisuse vastavalt 96% ja 64%, võrreldes radioloogidega, kelle tulemused olid vastavalt 50% ja 73%. (Becker jt, 2022).

Sarnaselt kiire areng toimub ka transpordisektoris. Kui 2005. aastal võitis Stanfordini ülikooli meeskond autonoomsete (isesõitvate) sõidukite võistluse, tõestades esimest korda autonoomsete sõidukite teostatavust (Thrun jt, 2006). Suurele saavutusele järgnesid tohutud arengud autonoomsete sõidukite turul. Järgnevate aastate jooksul hakkas autotööstus investeerima märkimisväärselt autonoomsete sõidukite tehnoloogiasse, juhtrolli mängisid suured ettevõtted nagu Tesla, Google ja Uber (Hamza jt, 2023). Tehisintellekti, eriti masinõppe ja arvuti poolse esemete nägemise integreerimine on olnud oluline, et võimaldada sõidukitel teha reaajas otsuseid ja navigeerida keerukates keskkondades (Hamza jt, 2023).

Tehisintellekti kasutatakse üha enam ka hariduses. Näiteks Tartu Ülikool julgustab kasutama tehisintellekti ja tekstiroboteid õppetöös, et toetada õppimist ja õpetamist ning arendada õpi- ja tööoskusi (Tartu Ülikool, 2024). Sarnasel arvamusel on ka Tallinna Ülikool, kus loodi õppematerjal tehisintellekti õppetöös kasutamiseks (Maljukov, 2023). Tartu Ülikooli õppija saab kasutada tekstirobotit järgmistes olukordades:

- iseseisva töö tegemisel, et paluda selgitada mõisteid, küsida tagasisidet, esitada endale kontrollküsimusi;
- kirjatöö alustamise tõrkest ehk nn tühja paberi hirmust ülesaamisel;
- ajurünnaku abilisenä;
- programmeerimise abivahendina;
- teksti toimetamisel ja tõlkimisel;
- kriitilise mõtlemise arendamisel, hinnates tekstiroboti väljundit;
- mahukast materjalist esmase ülevaate saamiseks (Tartu Ülikool, 2024).

Meelelahutuses on tehisintellekt kohal sageli märkamatu. Valdavas määras kasutatakse tehisintellekti erinevate algoritmide kujul, kuid inimesed sageli ei taju neid, sest need töötavad taustal, näiteks Facebooki fotode sildistamises, Google'i otsingus, Netflix'i või

Amazon'i soovitusel (Dafoe ja Zhang, 2019). See aitab inimestele pakkuda sisu, millest nad ise huvitatud on.

Lisaks on sotsiaalmeedia abil hakatud levitama laialdaselt tehisintellekti tehnoloogia abil loodud sisu. Näiteks 2016. aasta uuring tuvastas, et 9–15% Twitteri aktiivsetest kontodest olid tekstirobotid (Varol jt, 2017). Generatiivse tehisintellekti arendamine tõi kaasa võimsamad tööriistad: ChatGPT-d kasutatakse Twitteri pealkirjade loomiseks, Instagram pakub pilditaustade muutmise ja kunstiliste filtrite tööriistu, ning YouTube'is on hakanud ilmuma TI-ga loodud virtuaalseid sisuloojaid, kelle videod on täielikult TI põhised (Hua jt, 2024).

Äri- ja tööstusmaailmas on tehisintellekt muutunud asendamatuks tööriistaks, mille abil saab näiteks ennustada äriliste tegevuste tulemusi ja lihtsustada otsuste tegemist (Baby jt, 2017). Spark Cognitioni looja ja tegevjuht Amir Husain nimetab TI tarkvara teiseks tulemiseks, mis suudab ise otsuseid langetada ja tegutseda ootamatutes olukordades (Bharadiya jt, 2023). Tehisintellekti kui tööriista kasutatakse mitmes sektoris, alates küllastajate ja töötajate abistamisest ettevõtte territooriumil kuni tuuleturbiinide jälgimiseni, et ennustada hooldusvajadusi (Balasundaram ja Venkatagiri, 2020).

Kõigi arengutega kaasnevad aga ka tõsised eetikaga ja privaatsusega seotud küsimused. Tehisintellekti süsteemid võivad käituda ebaõiglaselt mitmel põhjusel: ühiskondlikel, tehnilistel või nende kahe kombinatsioonil (Bird jt, 2020). Näiteks võib TI peegeldada ühiskondlikke kallutusi, mis kajastuvad andmestikes, millega neid treenitakse või arendustiimide poolt tehtud eeldustest ja otsustest kogu arenduse ja kasutuselevõttule järgneva elutsükli jooksul (Bird jt, 2020). Teistel juhtudel tuleneb ebaõiglus tehnilistest teguritest, näiteks andmestike piirangutest: mõnede inimrühmade kohta on liiga vähe andmeid või süsteemi kasutajaliidesed ei arvesta teatud rühmade vajadustega (Bird jt, 2020).

Indias viidi läbi uuring intervjuuerides 20 tehisintellekti kasutajat, sealhulgas spetsialiste, arendajaid, akadeemikuid ja lõppkasutajaid, kes esindasid erinevaid valdkondi, nagu tervishoid, rahandus, haridus ja transport (Kamila ja Jasrotia, 2023). Osalejad väljendasid muret, et tehisintellekti süsteemid võivad jätkata olemasolevaid eelarvamusi ja diskrimineerimist, eriti sellistes tundlikes valdkondades nagu töölevõtmine, laenamine ja õiguskaitse (Kamila ja Jasrotia, 2023). Kuna TI-süsteeme treenitakse ajalooliste andmete põhjal, võivad nad pärida ja taastoota andmekogumite varasemaid kallutusi, mille tulemusena

võivad otsused ebaõiglaselt mõjutada teatud elanikkonda, suurendades sotsiaal-majanduslikke ebavõrdsusi ja soodustades diskrimineerivat käitumist (Kamila ja Jasrotia, 2023).

Sama uuringu osalejad võtsid sõna ka privaatsuse teemal. Osalejad väljendasid muret, et tehisintellekti süsteemid võivad isikuandmeid väärkasutada (Kamila ja Jasrotia, 2023). Nad rõhutasid vajadust tugevate andmeturbe meetmete järele ja rohkem teavet selle kohta, kuidas TI andmeid kogub, salvestab ja kasutab (Kamila ja Jasrotia, 2023). Kuna TI-tehnoloogiad muutuvad üha integreeritumaks erinevatesse ühiskonna valdkondadesse, suureneb oht, et isikuandmeid võidakse ära kasutada või vääralt käsitleda, mis on sage mureallikas andmekaitse tõhususe ja läbipaistvuse juures (Kamila ja Jasrotia, 2023).

Kuna tehisintellekti tehnoloogilised võimed arenevad, kasvab ka avalikkuse mure privaatsuse ja turvalisuse osas (Ioannou ja Tussyadiah, 2021; Tussyadiah jt, 2019). TI seadmed võivad hõlpsasti koguda ja salvestada klientide isiklikku teavet, näiteks biomeetrilisi andmeid, identifikaatoreid ja käitumislogisid (Ioannou jt, 2021), vahepeal isegi ilma klientide teadmisseta – piisab vaid sellest, kui klient tehisintellektiga suhtleb (Manikonda jt, 2018). Kliendid kardavad enim, et nende isiklikku teavet võidakse kasutada ja levitada loata või et nende informatsioon võib sattuda küberkurjategijate kätte (Davenport jt, 2020), nagu oli Apotheka juhtumi puhul, kui kurjategijad said kätte 700 000 kliendi andmed (Põlendik, 2024). Samuti näiteks Jaapanis asuv Henn-na hotell pidi tugevdama oma robotite turvameetmeid pärast külaliste privaatsuse rikkumisega seotud vahejuhtumit (Hertzfeld, 2019).

Lõpuks kerkib esile küsimus vastutusest. Küsimusele, kes vastutab tehisintellekti otsuste ja vigade eest, saab läheneda väga mitmet moodi. Peamiselt oleneb see kontekstist. Inimese või tehisintellekti tehtud otsuste puhul peetakse eelnevate uuringute põhjal sageli osaliselt vastutavaks kas otsustajat või TI arendajat (Tolmeijer jt, 2022). Isegi kui arendaja tegevus võib tahtlikult või tahtmatult piirata otsustaja tegutsemisvõimalusi (Muller ja Weisz, 2022). Moraalsetes dilemmades võivad inimesed hinnata inimesi ja masinaid ebavõrdselt (Hidalgo jt, 2021). Näiteks hinnatakse inimesi nende kavatsuste ja TI-d tulemuste põhjal, kusjuures inimeste tegevust juhuslikes olukordades vabandatakse tihtipeale sagedamini kui masinate tegevust (Hidalgo jt, 2021). Lisaks võib AI-de tajutud õiglus ja usaldusväärsus sõltuda kasutatavast terminoloogiast (nt algoritm, arvutiprogramm, tehisintellekt), mis mõjutab süsteemi edukust ja tulemusi (Langer jt, 2022).

1.2 Faktikontroll

Esimesed spetsiaalsed faktikontrolli osakonnad loodi 1920. aastatel ajakirjades Time ja The New Yorker. See praktika levis järk-järgult ka teistesse riiklikutesse ajakirjadesse, mis konkureerisid üha enam teiste ajalehtedega ja omavahel, püüdes pakkuda usaldusväärset ja põhjalikku ülevaadet maailmast (Angeletti ja Oliva, 2010; Silverman, 2007; Sivek ja Bloyd-Peshkin, 2019). Eesti ajakirjanduseetika koodeksi järgi peab ajakirjaniku või toimetuse poolt avaldatav uudismaterjal põhinema tõestataval ja tõenditega tagatud faktilisel informatsioonil (Meedialiit, 1998). Lisaks peavad uudised, arvamused ja oletused olema selgelt eristatavad. (Meedialiit, 1998). Rääkides faktikontrollist ajakirjanduslikust vaatepunktis tuleb pöörata tähelepanu nendele väidetele, mida me saame kontrollida.

Faktikontrolli populaarsus on viimastel aastatel kõvasti kasvanud (vt Joonis 2). Duke Reporter's Labi andmetel on aktiivsete faktikontrolli organisatsioonide arv kasvanud alates aastast 2009, kui Politifact võitis Pulitzeri preemia fakti kontrollimise eest. Sellel ajal oli 17 aktiivset veebilehte, mis tegelesid väidete kontrollimisega (Stencel, 2024). Aastal 2024 on aktiivseid organisatsioone 439 (Stencel, 2024).



Joonis 2. Aktiivsed faktikontrolliga tegelevate veebilehtede arv maailmas. *Allikas: Stencel, 2024; autori tõlgitud ja kujundatud.*

Eestis tegutsevad faktikontrollimisega valitsusvälised organisatsioonid, nt Eesti Päevalehe faktikontroll, kus teevad kaastööd Delfi Meedia uuriva toimetuse liikmed ning Eesti Väitlusseltsi meeskond, kes tuvastavad valeinformatsiooni ja toovad sellele laiema auditooriumi ees tähelepanu (Turk ja Kose, 2023). Eesti Päevalehe faktikontroll jälgib Eesti avalikus inforuumis avaldatud väidete täpsust, kus faktikontrolli tehes jälgitakse väite faktipõhisust, faktikontrolli läbipaistvust ja erapooletust (Eesti Päevaleht, 2018). Selline protsess võtab aset aga pärast informatsiooni avaldamist, kus Eesti Väitlusseltsi ja Päevalehe faktikontrolörid jälgivad poliitikute ja teiste arvamusi liidrite meedias väljaöeldut ning kontrollivad kahtlust äratavaid või küsimusi tekitavaid väiteid (Eesti Väitlusselts, i.a)

Sellist lähenemist on avalikult kritiseerinud ajakirjandusuurija Marju Himma, kelle sõnul jõuab valeinfo sellisel kujul läbi ajakirjanduse suurema auditooriumi ette (Himma, 2021). Himma kirjeldas, et ajakirjandusteoreetiliselt liigitatakse faktikontroll nii öelda lavataguseks tegevuseks, mis tähendab, et info tõesust tuleb kontrollida enne avaldamist, mitte kõigi ees, mis toob rambivalgusesse valeinfo (Himma, 2021).

Sellegipoolest kipuvad paratamatult avalikkuse ette jõudma vale või eksitavat informatsiooni jagavad teosed, mistõttu keskendub käesolev uurimistöo olukordadele, kus faktikontroll toimub alles pärast väite avaldamist. Selline lähenemine on kooskõlas Eesti Päevalehe faktikontrolli praktika ja põhimõtetega, mille kohaselt kontrollitakse väiteid sõltumata nende algallikast, sealhulgas ka sotsiaalmeedia postitusi ja avalikus inforuumis avaldatud väljaütlemisi, juhul kui need on mõjukad, vastuolulised või võivad avalikkust eksitada (Eesti Päevaleht, 2018). Eesti Päevalehe faktikontrolli meeskond lähtub Poynteri Rahvusvaheline faktikontrolli võrgustiku (IFCN) põhimõtetest, mis seisavad teabe usaldusväärsuse ja desinformatsiooni vastu võitlemise eest (Eesti Päevaleht, 2018; International Fact-Checking Network, i.a). Euroopa komisjon määratleb desinformatsiooni kui vale, ebatäpset või eksitavat teavet, mida levitatakse teadlikult ning mille üks eesmärke on ühiskonnas eksisteerivaid lõhesid suurendada (Euroopa komisjon, 2018; Wong, 2024).

1.3 Generatiivsete tehisintellektide tuntud mudelite ülevaate

Hiljutised edusammud suurtes keelemudelites (ingl. *Large Language Models*, ehk LLM) on viinud mitmete võimekate keelemudelite loomiseni. Ühed kõige populaarsemad on OpenAI ChatGPT 4o, Anthropic Claude 3.7 Sonnet, Google Deepmind Gemini 2.0 Flash, xAI Grok 3 ja DeepSeek V3. Käesolev ülevaade tutvustab keelemudeleid järgmiste tunnuste alusel:

- 1) Teadmiste treenimise viimane kuupäev (ingl. *knowledge cutoff date*)
- 2) Võimekus teha veebiotsinguid või pääseda ligi internetis leiduvatele allikatele
- 3) Mudeli kättesaadavus
- 4) Piirangud info kogumisel
- 5) Üldine tutvustus

Tabel 1. Võrdlus juhtivate keelemudelite teadmiste lõppkuupäevade, kättesaadavuse ja internetiühenduse kohta.

Keelemudel	Teadmiste lõppkuupäev	Kättesaadavus	Ligipääs internetile
DeepSeek V3	Juuli 2024	Avatud lähtekoodiga. Kättesaadav kõigile.	Jah, kui on rakenduse siseselt lubatud.
Claude 3.7 Sonnet	Oktoober 2024	Kättesaadav kõigile. Tasuta ja tasulised versioonid.	Jah, vaid tasulises versioonis ja USA piirkonnas.
ChatGPT 4o	Oktoober 2023	Kättesaadav kõigile. Tasuta ja tasulised versioonid	Jah, kui mudelil on vaja ajakohast informatsiooni.
Gemini 2.0 Flash	Juuni 2024	Kättesaadav kõigile. Tasuta ja tasulised versioonid.	Jah, kui mudelil on vaja ajakohast informatsiooni.
Grok 3	November 2024	Saadaval X platvormi tasulistele kasutajatele ja Grok veebilehe kaudu.	Jah (reaalajas ligipääs X platvormi sotsiaalmeedia andmetele ja internetile)

Iga nimetatud keelemudel on arendatud erineva ettevõtte poolt (nt Claude, ChatGPT, Grok, jne), mille treenimisel on katsetatud erinevaid algoritme. mistõttu põhinevad iga keelemudeli teadmised erineval treeningul ja samal põhjusel on neid ka huvitav võrrelda.

OpenAI mudel ChatGPT 4o

Töö kirjutamise hetkel ei ole OpenAI ChatGPT 4o mudeli kohta avaldatud teadusartiklit, mis kinnitaks ametlikult selle treeningandmete lõppkuupäeva. Avalikult kättesaadavate allikate põhjal ja keelemudeli enda vastusest võib siiski eeldada, et mudeli treenitud teadmised ulatuvad ligikaudu 2023. aasta oktoobrini. (OpenAI, i.a). See aga ei tähenda, et mudel ei suudaks leida informatsiooni, mis on avaldatud pärast viimast treeningandmete uuendamise kuupäeva. Interneti ligipääsuga keelemudelitel, nagu ChatGPT 4o, on võimalik kasutada vajaliku informatsiooni leidmiseks reaajas leiduvaid allikaid (Xu ja Li, 2024). Viimane treeningandmete uuendamise kuupäev on relevantne kasutaja jaoks vaid juhul, kui mudelil puudub internetile ligipääs, sest siis saab keelemudel ligi vaid andmetele, mis on avaldatud enne lõppkuupäeva. Samuti on keelemudelitel tavaks hallutsineerida või teisisõnu vastuseid välja mõelda. Näiteks kui keelemudel ei suuda enda väite toetamiseks õiget allikat leida, mõtleb ta selle välja või väidab, et sai informatsiooni ühest allikast, kust seda ei leidu (Griot jt, 2025). Filipiinidel läbiviidud uuring näitab, et nii mõnedki õpetajad ja õpilased kasutavad ChatGPT keelemudelit alternatiivse või ainukese otsingumootorina (Lima jt, 2025).

ChatGPT 4o on kättesaadav nii tasuta variandina kui ka Plus kasutajatele, kellel on viis korda kõrgemad sõnumi limiidid, ehk kui tavakasutaja saab esitada keelemudelile 10-20 päringut päevas, siis Plus kasutaja saab esitada ligikaudu 100 päringut päevas (OpenAI, 2024).

Claude 3.7 Sonnet

Claude 3.7 Sonnet on tehisintellekti keelemudel, mille on loonud ettevõtte nimega Anthropic. Mudeli õpetamisel on kasutatud suurel hulgal andmeid, sealhulgas avalikult kättesaadavat teavet internetist ja sisemiselt loodud treeningmaterjali (Anthropic, 2025a). Claude 3.7 Sonnet keelemudelit ei ole õpetatud kasutajate vestluste põhjal, vaid täiesti eraldi kogutud ja puhastatud andmekogumite alusel (Anthropic, 2025a). Kuigi Claude 3.7 Sonnet on treenitud internetis avalikult kättesaadava teabe põhjal kuni 2024. aasta novembrini, on selle teadmiste lõppkuupäev 2024. aasta oktoobri lõpp (Anthropic, 2025a). See tähendab, et mudeli

teadmistebaas on kõige ulatuslikum ja usaldusväärsem just sündmuste ja teabe osas, mis ulatuvad kuni 2024. aasta oktoobrini (Anthropic, 2025a).

Anthropic avaldas 2025. aasta alguses Claude 3.7 Sonneti System Card, mis tutvustab nende Claude perekonna uut mudelit (Anthropic, 2025b). System Card on sarnane valgele raamatule (ingl. *white paper*), mis on loodud mudeli struktuuri näitamiseks. Käesolev dokument viitab probleemile, et mudeli poolt esitatud mõtlemiskäik ei pruugi peegelduda mudeli lõplikus esitatud vastuses (Anthropic, 2025b). Seda nähtust on laiemalt käsitletud ka teistes uuringutes, kust on selgunud, et isegi hästi põhjendatud vastused võivad olla mitteusutavad, mis ei näita, kuidas mudel tegelikult vastuseni jõudis (Tutek jt, 2025).

Claude 3.7 Sonneti eripära on tema võime tegutseda laiendatud mõtlemise režiimis (ingl. *extended thinking mode*), kus keelemudel lahendab probleeme samm-sammult ja näitab kasutajale oma mõtlemiskäiku enne lõpliku vastuse andmist, mis suurendab läbipaistvust ja usaldusväarsust (Anthropic, 2025c). Alates 2025. aasta märtsist saab Claude 3.7 Sonnet mudel kasutada otsingute tegemiseks küll interneti, kuid seda vaid tasulises versioonis ja ainult USA-s (AI Eesti, 2025).

Grok 3

Grok on xAI ehk Elon Muski tehisintellektiettevõtte loodud vestlusrobot, mis esindab uuenduslikku lähenemist vestlusele põhinevatele tehinsintellektidele, rõhutades reaalses andmetöötlust, huumoriga rikastatud suhtlust ning keerukat matemaatilist mõtlemist. (Wangsa jt, 2024). xAI poolt loodud Grok 3 kasutab kõige uuemaid andmed ja arenenud järeldusvõimet (xAI, 2025). Oluliseks piiranguks on ligipääsetavus, kuna Grok on hetkel saadaval vaid X-i tasulistele kasutajatele, mis piirab selle laiemat kasutuselevõttu (Mohammed jt, 2025). Lisaks on Groki puhul märgatud võimalikke kallutusi vastustes (probleem, mis esineb paljudes keelemudelites), kuid eriti murettekitav on see just Groki puhul, arvestades selle seotust reaalses kasutajate poolt loodud sisuga (Mohammed jt, 2025).

Grok 3 viimane treeningandmete uuendamise kuupäev oli 2024. aasta novembris, kuid samuti omab ligipääsu internetile ja on tuntud X platvormil (endise nimega Twitter) kui faktikontrolli abimehena (xAI, i.a). Grok AI, mille on arendanud xAI, eristub tänu oma reaalses ligipääsule sotsiaalmeedia andmetele, eriti X platvormilt (Adinath ja Smiju, 2025).

Siiski võib selle sõltuvus X platvormi andmetest põhjustada teabepõhist kallutatust, kuna mudeli treening on piiratud just selle võrgustiku kontekstiga (Souza ja Weigang, 2025).

Google Gemini 2.0 Flash

Gemini 2.0 Flash keelemudeli teadmised ulatuvad 2024. aasta juunini (Google Cloud, i.a). See tähendab, et mudeli algteadmised piirduvad selle kuupäevani kogutud andmetega. Samas on mudelil võimekus ka otsida täiendavat teavet internetist, kasutades reaaliajast ligipääsu allikatele (Ramachandran, 2024). Seda kõike pakub mudel enda kasutajatele tasuta, kuid nende kõige võimsam mudel Google Gemini 2.0 Pro on kasutajatele saadav tasulisena (Mallick & Kilpatrick, 2025).

Gemini, ChatGPT ja Claude keelemudelitega on läbi viidud uuring, mille põhjal toetuvad mitmed keelemudelid väite kontrollimisel enda algteadmistele ja ei kasuta enda oskust toetuda koheselt reaaliajast interneti andmetele (Xu ja Li, 2024). Google tiimi poolt läbi viidud uuringu tulemused näitavad, et Gemini 1.5 Flash oli võrreldes teiste tehisintellekti mudelitega, (GPT-4 Turbo, GPT-3.5 Turbo, Claude 3 Sonnet/Haiku/Opus) kõige kiirem vastuse esitamise poolest ja kõige täpsem faktide esitamisel võrreldes varasemate Gemini mudelitega (1.0 Pro, 1.0 Ultra, 1.5 Pro) (Georgiev jt, 2024).

Google Gemini 2.0 Flash on Google Gemini perekonna kõige uuem tehisintellekti mudel, mis tugineb 1.5 Flashi edule, pakkudes täiustatud jõudlust sama kiire reageerimisajaga, mis suudab ka loomupäraselt kasutada tööriistu, nagu Google Search, koodi täitmine ja kasutaja poolt määratud funktsioone (Hassabis, 2024).

DeepSeek V3

Alates DeepSeek versiooni R1 ilmumisest 20. jaanuaril 2025 on DeepSeek tõusnud Apple Store'i allalaadimiste tippu ja üllatanud investoreid ning põhjustanud mõne tehnoloogia aktsia, nagu Nvidia, languse (Gao jt, 2025). DeepSeeki teeb eriliseks see, et mitmetes võrdlustestides saavutas see sama häid või isegi paremaid tulemusi nagu OpenAI ChatGPT, kuid palju väiksema treeningukuluga (Gao jt, 2025). Veebruaris 2025 läbi viidud uuring näitab, et DeepSeek mudel v3 on võrreldes teiste tehisintellektidega eriti tõhusam ülesannete puhul, mis nõuavad palju teadmisi (Aydin jt, 2025).

Kolmanda osapoolte andmetel piirduvad DeepSeek V3 mudeli teadmised 2024. juuliga. (PromptHub, i.a; Knostic, 2025). Ametlik kinnitus selle väite jaoks puudub, kuid Knosticu meeskond kasutas DeepSeeki piirangutest mööda hiilimiseks meetodit, mille tulemusena ütles DeepSeek neile ise, mis ajast mudeli viimased andmed pärinevad (Knostic, 2025). DeepSeek V3 on avatud lähtekoodiga mudel (DeepSeek, 2025). Täielikult avatud lähtekoodiga mudelid pakuvad piiramatut ligipääsu koodile ja dokumentatsioonile, võimaldades selle muutmist ja edasis levitamist (Vake jt, 2025). Kuid praktikas ühtegi täielikult avatud lähtekoodiga mudelit ei eksisteeri (Vake jt, 2025). Sellegipoolest on DeepSeek V3 keelemudel laialdaselt kõigile kättesaadav ja kasutatav (vt Tabel 1).

2. UURIMISKÜSIMUSED

Milline on tänapäeva tehisintellekti keelemudelite võimekus iseseisva, täiemahulise faktikontrolli teostamisel?

- Millisel määral suudavad erinevad keelemudelid eristada fakte ja väljamõeldisi?
- Kui täpselt suudavad erinevad keelemudelid kontrollida sama teavet võrreldes inimese faktikontrolliga?
- Millised on keelemudelite tugevused ja nõrkused faktikontrolli osas (nt täpsus, kontekst, valdkond, läbipaistvus, allikad)?

3. EMPIIRILISE UURINGU KAVA

3.1 Kasutatavad andmekogumis- ja analüüsimeetodid

Uurimistöös kasutatakse analüüsimeetodina võrdlevat analüüsi, mille eesmärk on hinnata viie erineva tehisintellekti keelemudeli võimekust teostada faktikontrolli. Võrdlev analüüs on valitud, kuna see võimaldab esile tuua mudelite tugevused ja nõrkused erinevates kategooriates (nt faktidele toetumine, usaldusväärsete allikate kasutamine ja kontrolli põhjalikkus). Analüüsi aluseks on viie tehisintellekti vastused samale väitele. Võrdlevat analüüsi saab kasutada, sest kõik viis tehisintellekti keelemudelit erinevad üksteisest treening andmete põhjal. Meetod keskendub sarnasuste ja erinevuste süstemaatilisele vaatlemisele ning kategooriate kaupa võrdlemisele. Autor hindab keelemudeleid autori poolt loodud hindamismaatriksil, mille alusel määrab autor iga kategooria hinnangu iseseisvalt.

Andmete kogumine toimub tehisintellekti keelemudelitega ja eksperdiga suhtlemise kaudu. Ekspert jagab autorile väite, mida keelemudelid kontrollivad. Keelemudelitele esitatakse ühesugune prompt väite kontrollimiseks, mis annab keelemudelile juhendi vastamiseks.

Andmekogumis- ja analüüsimeetodi metodoloogilised piirangud

Nende meetodite puhul esinevad mõned metodoloogilised piirangud:

- 1) Tehisintellektide valik ei kata kõiki saadaval olevaid mudeleid ja ei anna head ülevaadet või arusaama, kuidas uurimuses mitte kaasatud tehisintellektid ülesandega hakkama saavad;
- 2) Valimis on suuremate tehiskeelte perekondade (nt OpenAI, Anthropic, Google) tehisintellektid, mistõttu jäävad väiksemad tehisintellektid uurimata;
- 3) Tehisintellekti mudelid arenevad kiiresti, mistõttu võib uuringu hetkeseis kaotada kiiresti oma asjakohasuse;
- 4) Keelemudelid ei anna järjepidevaid vastuseid, mistõttu võib järgmine uurija saavutada teistsugused tulemused;

- 5) Autori iseseisev hinnang igas kategoorias, ehk otsus selle kohta, kas tehisintellekti vastus oli parem või halvem kui eksperdi oma, võib sisaldada subjektiivseid tõlgendusi ja seetõttu erineda mõne teise hindaja arvamustest.

Uuringus osalevad vaid viie suurema keelemudeli (OpenAI, Google, Anthropic jt) esindajad, mistõttu ei saa tulemusi üldistada kõikidele turul olevatele mudelitele. Eriti jäävad välja väiksemad, regionaalsed või valdkonnale spetsialiseerunud mudelid.

3.2 Plaanitav valim

Valimisse kuuluvad 5 tehisintellekti keelemudelit, mis on treenitud erinevate andmete põhjal (Minaee jt, 2024). Valimisse jõudnud keelemudelid on kõik valitud vastavalt nende faktipõhise täpsuse, mille mõõtmiseks kasutati Hughes Hallucination Evaluation Modelit (HEMMi) ja nende populaarsuse tõttu (Hughes, 2024) (Minaee jt, 2024):

- 1) ChatGPT 4o;
- 2) Gemini 2.0 Flash;
- 3) Claude 3.7 Sonnet;
- 4) DeepSeek V3;
- 5) Grok 3.

HEMM mõõdab, kui sageli keelemudelid „hallutsineerivad,” ehk lisavad vastusesse infot, mida algallikas tegelikult ei sisalda või teisisõnu kasutavad väiteid, mis ei põhine tõenditel (Hughes, 2024).

3.3 Algne hindamismaatriks

Esialgse hindamismaatriksi loomise vajadus tekkis töö algfaasis, kui selgus, et tehisintellekti keelemudelite faktikontrolle ei ole võimalik hinnata vaid üldmulje või lihtsustatud kriteeriumite alusel. Sellest tulenevalt tekkis vajadus struktureeritud ja kategooriapõhise hindamissüsteemi järele, mis võimaldaks analüüsida vastuseid mitme kategooria alusel (vt Tabel 2).

Tabel 2. Tehisintellekti faktikontrolli algne hindamismatriks.

Kategooria	Kirjeldus	Hinnang (1-5)	Kommentaariid
Faktide täpsus ja allikate usaldusväärsus	Kas vastus tugineb faktidele ja usaldusväärsetele allikatele?		
Konteksti mõistmine ja loogiline ülesehitus	Kas vastus võtab arvesse konteksti ja on loogiliselt ülesehitatud?		
Viidete kasutamine ja läbipaistvus	Kas vastus viitab mitmele kontrollitavale allikale ja otsuseni jõudmine on arusaadav?		
Erapooletus ja käsitluse põhjalikkus	Kas vastus on esitatud erapooletu seisukohast ja teema on põhjalikult käsitletud?		
Asjakohasus ja kõlblikkus	Kas vastus põhineb asjakohastel andmetel ning on praktiliselt kasutatav?		
Faktikontrolli tulemus	Kas fakt sai kontrollitud ja tulemus oli korrektne?		

Allikas: autori koostatud; Nieminen ja Sankari (2021) ning Barua (2024) alusel

Selleks, et hinnata kuidas tehisintellekt faktikontrolliga hakkama saab, on vaja luua hindamismatriks. Valimile esitatakse eelnevalt faktikontrolli organisatsiooni (Eesti Päevalehe faktikontroll) poolt kontrollitud fakt, seejärel on tehisintellekti ülesandeks välja selgitada, kas artikli sisu on tõsi või vale. Tehisintellekti vastused langevad kategooriasse *tõsi*, *vale*, *pole kindel*, *pole fakt* (DeVerna, 2023). Tõeks või valeks loetakse tehisintellekti vastus siis, kui ta väidab et fakt on absoluutselt tõsi või absoluutselt vale. Ebakindlate väidete puhul ei anna tehisintellekt konkreetset vastust, ei oska vastata või esitatakse umbmäärane vastus (DeVerna, 2023). Samuti peab tehisintellekt selgitama, miks selline vastus esitati. Tehisintellekti vastuseid hinnatakse iga kategooria alusel skaalal 1-5, seejärel kogutakse kokku kõikide kategooriate tulemused ja võrreldakse nende erinevusi (Tabel 2).

1 - Väga halb (kummalegi küsimusele ei saa vastavas kategoorias jaatavalt vastata);

- 2 - Halb (ühele küsimusele ei saa vastavas kategoorias jaatavalt vastata ja teise vastus vajab lisa kommentaari);
- 3 - Keskpärane (ühele küsimusele ei saa vastavas kategoorias jaatavalt vastata ja teine vastus ei vaja parandusi);
- 4 - Hea (mõlemale küsimusele saab vastata jaatavalt kuid tekivad üksikud lisaküsimused või kommentaarid);
- 5 - Väga hea (igale kategooria küsimusele saab vastata jaatavalt ning lisaküsimusi või kommentaare ei teki).

3.4 Meetodikatsetus

Uurimistöö eesmärgiks on hinnata tehisintellekti mudelite võimekust teostada täiemahulist faktikontrolli. Selleks viiakse läbi eksperiment, mille raames võrreldakse erinevate tehisintellekti keelemudelite poolt koostatud faktikontrolle, mis põhinevad samal väitel. Keelemudelite vastuseid kõrvutatakse Delfi Meedia uuriva toimetuse liikme poolt läbiviidud faktikontrolliga, mis põhineb samuti samal väitel. Analüüsimeetodina kasutatakse võrdlevat analüüsi.

Võrdlevat analüüsi kasutatakse kahe tasandiliselt. Esmalt võrreldakse erinevate keelemudelite vastuseid samale väitele, mis aitab meil mõista mudelite tugevusi ja nõrkusi. Teisalt võrreldakse mudelite vastuseid eksperthinnanguga. Eksperdi hinnang on kaasatud selleks, et hinnata, kas tehisintellekti vastus vastab faktikontrolli standarditele. Hindamismaatriks võimaldab hinnata tehisintellekti vastuseid struktureeritult mitme kategooria kaupa. Eksperdi poolt teostatud faktikontroll ei ole otseseks hindamise aluseks, kuid selle kasutamine aitab suurendada hindamise usaldusväärsust ja võimaldab kriitilisemat analüüsi. Hindamise viib läbi uuringu läbiviija, tuginedes selgelt määratletud kriteeriumitele, kuid tehes seda iseseisvalt, mitte tuginedes vaid eksperdi vastustele. Selline lähenemine võimaldab ka juhtudel, kus tehisintellekt ületab eksperti, anda tehisintellektile positiivse hinnangu.

Andmete kogumine toimub tehisintellekti keelemudelitega (DeepSeek V3, Gemini 2.0 Flash, Grok 3, Claude 3.7 Sonnet ja ChatGPT 4o) suhtlemise kaudu. Keelemudelitele esitatakse ühesugune sisend (ingl. *prompt*), mis annab keelemudelile juhendi vastamiseks.

3.4.1 Pilootuuring

Enne lõpliku uuringu läbiviimist viidi läbi pilootuuringud, et katsetada faktikontrolli hindamismaatriksit. Uuringute eesmärk oli tuvastada hindamismaatriksi ja prompti puudused. Esimeses pilootuuringus esitati ChatGPT 4o keelemudelile järgmine sisend: „teosta minu jaoks põhjalik faktikontroll faktile, et Eestis on Euroopa kalleim elektrihind.”

ChatGPT 4o vastas järgnevalt (ChatGPT 4o, 2025):

The claim that Estonia has the highest electricity prices in Europe is not accurate based on the latest available data. In 2023, several European countries experienced high electricity prices, but the highest rates were observed in different countries depending on specific conditions and the time frame considered.

For the first half of 2023, the highest electricity prices for household consumers, when measured in purchasing power standards (PPS), were found in countries like the Netherlands and Germany. The Netherlands particularly saw a significant spike in electricity prices due to the reversal of tax subsidies and an increase in energy taxes ([European Commission](#)). In contrast, Estonia did not rank at the top for the highest electricity prices during this period.

Additionally, the latest Eurostat data from the first half of 2024 indicates that Germany, Ireland, and Denmark had some of the highest electricity prices for household consumers. Prices in these countries remained significantly high due to factors like high taxes and levies on electricity ([Qery](#)).

It's important to note that electricity prices can fluctuate widely across Europe due to various factors, including government policies, taxation levels, and shifts in supply and demand. While Estonia does experience high electricity prices relative to some other countries, it does not consistently hold the position of having the highest prices in Europe.

Keelemudeli esitatud vastust hinnati vastavalt pilootuuringu hindamismaatriksi kategooriatele (vt Tabel 3). Esimeses hindamismaatriksis kasutati Likerti skaalat vahemikus 1–5, kus number 1 on väga halb ja 5 on väga hea. Skaala puhul tekkis probleem, et number 1 ei näita lugejale piisavalt hästi, et tegu oli väga kehva vastusega. Kategooriad olid sõnastatud kahe

kriteeriumi kaupa, (nt faktide täpsus ja allikate usaldusvärsus), et tabel liiga pikaks ei läheks. See aga eksitas lugejat, sest tabelist ei saa välja lugeda näiteks kui usaldusväärased olid allikad ja kui täpsed olid faktid.

Tabel 3. ChatGPT 4o faktikontrolli pilootuuringu esimese hindamismatriksi näidistulemused.

Kategooria	Kirjeldus	Hinnang (1-5)	Kommentaariid
Faktide täpsus ja allikate usaldusvärsus	Kas vastus tugineb faktidele ja usaldusväärsetele allikatele?	3	Viimane lõik ei sisalda viidet, mistõttu loetakse vastus hinnanguks mitte faktiks.
Konteksti mõistmine ja loogiline ülesehitus	Kas vastus võtab arvesse konteksti ja on loogiliselt ülesehitatud?	4	Viimane lõik ei ole faktikontrolli vaatepunktist oluline.
Viidete kasutamine ja läbipaistvus	Kas vastus viitab mitmele kontrollitavale allikale ja otsuseni jõudmine on arusaadav?	3	Vastusest ei saa välja lugeda, kuidas on jõutud viimases lõigus esitatud järeldusteni.
Erapooletus ja käsitluse põhjalikkus	Kas vastus on esitatud erapooletu seisukohast ja teema on põhjalikult käsitletud?	3	Vastuses räägitakse vaid hinnast kodutarbijatele ja oleks saanud esile tuua rohkem arvulisi näitajaid, mis otsust enamgi toetaks.
Asjakohasus ja kõlblikkus	Kas vastus põhineb asjakohastel andmetel ning on praktiliselt kasutatav?	3	Faktikontroll ei ole praktiliselt kasutatav, sest viimane lõik ei sisalda kontrollitavaid fakte, pole läbipaistvalt esitatud ega kontekstis oluline.
Faktikontrolli tulemus	Kas fakt sai kontrollitud ja tulemus oli korrektne?	5	Tulemus oli korrektne, kuid fakti saab põhjalikumalt kontrollida.

Allikas: autori koostatud; Nieminen ja Sankari (2021) ning Barua (2024) alusel.

Järgmises uuringus parandati eelnevalt esile toodud vead. Iga kategooria sisaldas edaspidi vaid ühte kriteeriumit ning skaala laiendati vahemiku 0–5 peale. Google Gemini 2.0 Flash

keelemudelile esitati prompt: „teosta minu jaoks põhjalik faktikontroll faktile, et Ukraina müüb poole USA relvaabist maha Mehhiko narkokartellidele.”

Gemini 2.0 Flashi vastust hinnati uuendatud hindamismaatriksi alusel ja tulemused tulid realistlikumad ja näitasid paremini tehisintellekti tegelikku taset. Sellegipoolest oli tunne, et skaala 0–5 ei illustreeri piisavalt hästi tulemusi. Järgmine probleem esines keelemudelitele esitatud sisendis. Esiteks mainiti sisendis, et teosta minu jaoks faktikontroll faktile, mis võib keelemudelile vihjata, et tegu on juba faktiga mitte väitega. Samuti puuduvad sisendis konkreetsed juhised, mida tehisintellektilt oodatakse. Väga lühikese sisendi puhul esitas mõni keelemudel (nt ChatGPT 4o) enda faktikontroll ka inglise keeles ja kasutati ingliskeelseid viiteid, sest polnud teada, et oodatakse eestikeelset vastust.

Selle jaoks muudeti prompti vastavalt promptide loomise juhendile (Giray, 2023). Uus prompt sisaldas detailsemaid juhiseid ja konkreetsemat tööülesannet, mis peaks aitama keelemudelil sisendile paremini vastata. Prompt on koostatud nii, et keelemudel esitaks neutraalse, ajakirjanduslikule faktikontrollile keskendunud, faktidega toetatud ja viidatud faktikontrolli. Uurimistöös kasutati järgmist prompti:

„Tee põhjalik ja põhjendatud faktikontroll (nimi) esitatud väitele, et (väide). Kasuta faktikontrollis usaldusväärseid ja autoriteetseid allikaid. Palun esita oma vastus koos viidetega allikatele, mida kasutasid oma järelduse tegemiseks.”

Sõnastus on valitud eesmärgiga suunata mudelit andma sisukat, faktidel põhinevat ja kontrollitavate allikatega põhjendatud vastust. Täpsed juhised sisendis aitavad tagada, et mudel ei jääks nipsisõnaliseks, väldiks ebausaldusväärseid ja eksitavaid viiteid ja esitaks kontrollitavaid andmeid.

Kui hinnatakse ainult tehisintellekti tööd 0–5 skaalal, aga eksperdi tulemused samas mõõtkavas puuduvad, siis ei saa otseselt hinnata, kas tehisintellekt suudaks teostada täiemahulist faktikontrolli. Selle kitsaskoha ületamiseks kaaluti lahendusena kasutada väiteid, mille kohta on juba olemas ekspertide poolt koostatud faktikontrollid. Pilootuuringus läbiviidud faktikontrollid põhinesid juba teostatud faktikontrollidel, mis olid avaldatud Eesti Päevalehe faktikontrolli rubriigis (Vunš, 2024; Vunš, 2025).

Pilootuuringu põhjal tuvastatud probleemid viisid hindamismaatriksi ümberkujundamiseni. Uus hindamismaatriks põhineb samadel kategooriatel, kuid skaalal -2–2. See skaala aitab

lugejal ja hindajal paremini eristada uuringu tulemusi, vastavalt kas tehisintellekt sooritas antud kategoorias parema või halvema tulemuse kui eksperdid. Sisulisi tekstilisi vastuseid ei saa uuringu läbiviija hinnangul täiesti objektiivselt hinnata, mistõttu pürgitakse objektiivsuse poole, täielikku objektiivsust saavutamata.

Järgmised tulemused esitati skaalal -2 kuni 2:

- 2 – Keelemudeli vastus kategoorias oli oluliselt parem kui eksperdil
- 1 – Keelemudeli vastus kategoorias oli parem kui eksperdil
- 0 – Keelemudeli ja eksperdi vastus kategoorias olid samal tasemel
- -1 – Keelemudeli vastus kategoorias oli halvem kui eksperdil
- -2 – Keelemudeli vastus kategoorias oli oluliselt halvem kui eksperdil

ChatGPT 4o keelemudelile esitati uuenduste põhjal järgmine sisend, mis põhines taaskord eksperdi poolt teostatud faktikontrollil (Vunš, 2025b): „Tee põhjalik ja põhjendatud faktikontroll Jaanus Härmsa väitele, et 30 000 Saksamaa tuulikut katavad riigi energiavajadusest vaid 3,2 protsenti. Kasuta faktikontrollis usaldusväärseid ja autoriteetseid allikaid. Palun esita oma vastus koos viidetega allikatele, mida kasutasid oma järelduse tegemiseks.”

Tabel 4. ChatGPT 4o faktikontrolli pilootuuringu viimase hindamismatriksi näidistulemused.

Kategooria	Kirjeldus	Hinnang (-2– 2)	Kommentaariid
Faktidele toetumine	Kas tehisintellekti vastus tugines faktidele rohkem kui ekspert?	-1	Ekspertid tuginesid veel lisaks kirjalikele allikatele ka graafikutele
Allikate usaldusvärsus	Kas tehisintellekti kasutatud allikad olid usaldusväärsemad kui eksperdi omad?	0	Mõlema poolt esitatud allikad olid usaldusväärsed
Kiirus	Kas tehisintellekt jõudis vastuseni kiiremini kui ekspert?	2	Tehisintellekt jõudis vastuseni paari sekundiga
Viidete kasutamine	Kas tehisintellekt viitas rohkematele ja paremini kontrollitavatele allikatele kui ekspert?	0	Tehisintellekt viitas küll paremini kontrollitavatele allikatele kuid eksperdid viitasid rohkematele (mida sai vähese vaevaga kontrollida)
Läbipaistvus	Kas tehisintellekti otsustusprotsess oli selgem ja läbipaistvam kui eksperdil?	-1	Eksperti otsustusprotsess oli tunduvalt selgem, kui tehisintellektil.
Erapooletus	Kas tehisintellekt esitas vastuse erapooletumalt kui ekspert?	0	Mõlemad kontrollisid fakti erapooletu seisukohast.
Käsitluse põhjalikkus	Kas tehisintellekt käsitles teemat põhjalikumalt kui ekspert?	-1	Tehisintellekt ei selgitanud primaarenergia ja lõppenergia erinevusi.
Asjakohasus	Kas tehisintellekti kasutatud andmed olid asjakohasemad kui eksperdi omad?	0	Mõlemad kasutasid asjakohaseid andmeid.
Faktikontrolli tulemus	Kas tehisintellekt jõudis õigema ja täpsema järelduseni kui ekspert?	0	Mõlemad jõudsid sama tulemuseni.

Allikas: autori koostatud; Nieminen ja Sankari (2021) ning Barua (2024) alusel

Sel korral saavutati siiani kõige paremini keelemudeli tulemusi illustreerivad tulemused, mis näitavad, et tehisintellekt jääb eksperdile alla (vt Tabel 4). Tabelit analüüsid ja mitmeid eeluuringuid tehes selgus, et kategooria „kiirus” on ebavajalik, sest selles kategoorias saavutab keelemudel alati parema tulemuse kui ekspert, ning eksperdi kiirust väite kontrollimisel ei saa kontrollida. Samuti on kategooria „asjakohasus” sõnastus eksitav, mistõttu muudetakse kategooria uueks nimeks „ajakohasus.” Varem ekspertide poolt teostatud faktikontrollide kasutamine kujunes probleemseks, sest internetiühendusega tehisintellekti mudelid kipuvad kasutama oma vastustes allikana just neidsamu ekspertide poolt teostatud faktikontrolle, kuna need käsitlevad täpselt samu väiteid. See aga ei anna meie uurimusküsimusele vastust, kas tehisintellekt suudaks tegutseda faktikontrollina.

Sellegipoolest nõuab skaala -2–2 kasutamiseks eksperdi poolt kontrollitud väidet, kuid eksperdi hinnang ei tohi olla internetis avaldatud enne kui tehisintellekt on otsuseni jõudnud. Selle jaoks tuleb ühendust võtta mõne faktikontrollijaga ja kordineerida sama väite faktikontroll. Seejärel peab uuringu läbiviija sooritama faktikontrolli iga tehisintellektiga, enne kui eksperdi hinnang avalikustatakse.

3.4.2 Uuring

Pilootuuringu põhjal tuvastatud probleemide ja kitsaskohtade lahendamise järel alustatakse uuringuga. Uuringu eesmärk on hinnata, kuidas erinevat tehisintellekti keelemudelid (DeepSeek V3, Gemini 2.0 Flash, Grok 3, Claude 3.7 Sonnet ja ChatGPT 4o) suudavad läbi viia faktikontrolli väite osas, mille kohta eksperthinnangut veel avalikult ei eksisteeri. Tulemuste hindamiseks kasutatakse täiendatud hindamismaatriksit, mis sisaldab selgelt defineeritud kategooriaid ja skaalat -2–2, kus 2 näitab, et keelemudel ületas eksperdi ja -2, et keelemudel jäi selgelt eksperdile alla.

Kategooriad defineeritult:

- **faktidele toetumine** – hinnatakse keelemudeli võimekust toetuda enda väidete ja järelduste esitamisel faktidele (sh viidatud allikate sisule);
- **allikate usaldusvärsus** – hinnatakse keelemudeli võimekust kasutada usaldusväärseid allikaid (nt ametlik statistika, teadusartiklid, meediaväljaanded);

- **viidete kasutamine** – hinnatakse keelemudeli võimekust kasutada viiteid arusaadavalt ja paigutada neid sobivatesse kohtadesse;
- **läbipaistvus** – hinnatakse keelemudeli võimekust esitada jälgitav, loogiline ja kergesti mõistetav mõttekäik, mille järgi saab lugeja aru, kuidas järeldusteni jõuti ;
- **erapooletus** – hinnatakse keelemudeli võimekust esitada väiteid erapooletu seisukohast, vältides emotsionaalseid või kallutatud sõnumeid;
- **käsitluse põhjalikkus** – hinnatakse keelemudeli võimekust käsitleda teemat põhjalikult, lisades vajalikku konteksti;
- **ajakohasus** – hinnatakse keelemudeli võimekust kasutada ajakohaseid viiteid, mis kajastavad väite kontrollimise ajahetkel kehtivat infot;
- **faktikontrolli tulemus** – hinnatakse keelemudeli sõnastust järelduse tegemisel ja järelduse seost esitatud tõendite ja allikatega.

Keelemudelite esitatud sisend on standardiseeritud, ehk kõikidele keelemudelitele esitatakse samasugune sisend, mis sisaldab selgeid juhiseid (vt ptk 3.4.1). Ekspert avaldab oma hinnangu pärast seda, kui kõik tehisintellekti keelemudelid on vastused andnud. Seejärel hindab uuringu läbiviija faktikontrolli tulemusi ja kannab need tabelisse, misjärel tehakse järeldus tehisintellekti võimekusest teostada täiemahulist faktikontrolli.

4. UURIMISTÖÖ TULEMUSED

Selles peatükis kontrollivad faktikontrolli keelemudelid Sandra Lauri Facebooki videos väljaöeldud väidet, mida võrreldakse eksperdi kontrolliga. Sandra Laur on London King's College'i rahvusvaheliste suhete tudeng, Keskerakonna liige, Euroopa Parlamendi praktikant ja endine Tartu Linna Noortevolikogu esimees, kellel on Facebookis ligi 2,5 tuhat jälgijat (Laur, i.a). Enda postitatud videos mainis Laur, et Tartus on vaid üks varjend, kuhu raketirünnaku korral varjuda. (Laur, 2025). Väidet kontrollib lisaks keelemudelitele ka Delfi faktikontrollimise ja valeinfo reporter Mari-Liis Somelar. Tehisintellektidelt adekvaatse tulemuse saamiseks viib Somelar enda kontrolli läbi pärast seda, kui uuringu läbiviija on tehisintellektidelt enda vastused saanud. Seejärel saab uuringu läbiviija võrrelda eksperdi hinnangut ja tehisintellektide vastust.

Prompt, mis esitati tehisintellekti keelemudelitele, et nad saaksid läbi viia täiemahulise faktikontrolli:

Tee põhjalik ja põhjendatud faktikontroll Sandra Lauri poolt esitatud väitele, et 'Tartu linnas on vaid üks varjend kuhu raketirünnaku korral varjuda'. Kasuta faktikontrollis usaldusväärseid ja autoriteetseid allikaid. Palun esita oma vastus koos viidetega allikatele, mida kasutasid oma järelduse tegemiseks.

Allikas: autori koostatud; OpenAI mudeli ChatGPT 4o ja Giray (2023) alusel, et tagada tehisintellektidele võimalikult arusaadav juhised

Tabel 5. Faktikontrolli hindamismatriks

Kategooria	Kirjeldus	Hinnang (-2- 2)	Kommentaariid
Faktidele toetumine	Kas tehisintellekti vastus tugines faktidele rohkem kui ekspert?		
Allikate usaldusväarsus	Kas tehisintellekti kasutatud allikad olid usaldusväärsemad kui eksperdi omad?		
Viidete kasutamine	Kas tehisintellekt viitas rohkematele ja paremini kontrollitavatele allikatele kui ekspert?		
Läbipaistvus	Kas tehisintellekti otsustusprotsess oli selgem ja läbipaistvam kui eksperdil?		
Erapooletus	Kas tehisintellekt esitas vastuse erapooletumalt kui ekspert?		
Käsitluse põhjalikkus	Kas tehisintellekt käsitles teemat põhjalikumalt kui ekspert?		
Aktuaalsus	Kas tehisintellekti kasutatud andmed olid aktuaalsemad kui eksperdi omad?		
Faktikontrolli tulemus	Kas tehisintellekt jõudis õigema ja täpsema järelduseni kui ekspert?		

Allikas: autori koostatud; Nieminen ja Sankari (2021) ning Barua (2024) alusel

4.1 Faktidele toetumine

Esimesena kontrollitakse tehisintellekti keelemudelite võimekust toetuda faktidele. Kategoorias hinnati peamiselt keelemudelite võimekust toetuda väiteid esitades kontrollitavatele faktidele, mitte oletustele või hinnangutele. DeepSeek V3 mainis enda kontrollis, et Tartus on mitu avalikku varjendit, kuid tegelikult on Tartus 24 varjumiskohta, mitte varjendit, mistõttu ei põhine see väide faktil. Samuti viitab keelemudel Tartu Kaubamaja keldrikorrusele kui varjendile ja Mihkel Kärmasese kui kaitseministeeriumi nõunikule, kuid seda infot ei ole võimalik keelemudeli poolt esitatud viidete puhul kontrollida (vt Lisa 1). Nende andmete põhjal sai DeepSeek V3 faktidele toetumise kategoorias hinnanguks –2.

„Kaitseekspert Mihkel Kärmas (Kaitseministeeriumi nõunik) on rõhutanud, et Eestis on varjendivõrgustiku arendamine prioriteet, eriti pärast 2022. aasta Ukrainasõda” (DeepSeek V3, 2025).

Gemini 2.0 Flash ei suutnud faktidega ümber lükata väidet, et Tartus on vaid üks varjend, vaid mainis, et linnas on mitmeid olemasolevaid keldreid ja maa-aluseid ruume, mida saab kasutada varjumiskohtadena (vt Lisa 2). Gemini 2.0 Flash sai faktidele toetumise kategoorias hindeks –2. Samuti ei erista keelemudel 3.7 Sonnet mõisteid „varjend” ja „varjumiskoht,” mis on selle väite kontekstis olulised mõisted. Sellegipoolest toob keelemudel välja selle, et Tartus on rohkem kui üks varjumiskoht, mis on kooskõlas eksperdi hinnanguga. Lisaks sellele on väljatood konkreetsed hooned, kus varjumiskohad asuvad, mis toetab eelöeldud väidet. Nende tulemuste põhjal saab keelemudel 3.7 Sonnet faktidele toetumise kategoorias hinnanguks –1 (vt Tabel 6).

Tabel 6. Faktikontrolli hindamismatriks – faktidele toetumine

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Faktidele toetumine	–2	–2	–2	–1	0

OpenAI ChatGPT 4o ja xAI keelemudel Grok 3 said faktidele toetumisega paremini hakkama kui eelnevad keelemudelid. Sellegipoolest eksisid nii Grok 3 kui ka ChatGPT 4o enda

faktikontrollis avalike varjumiskohtade arvuga, kus Grok 3 väitis, et neid on 25 ja ChatGPT 4o kirjutas, et neid on Tartu linnas 23 tükki, kuid tegelikult on neid hetkeseisuga 24.

„Uuringud viitavad, et Tartus on vähemalt 25 avalikku varjumiskohta, sealhulgas koolides, linna hoonetes, eraettevõtete hoonetes ja ülikooli hoonetes” (Grok 3, 2025).

Grok 3 puhul pole teada, kust selline informatsioon koguti, sest kontrollis on viidatud Tartu linnavalitsuse lehele, kus on esitatud loetelu, kuhu kuulub 24 avaliku varjumiskohta ja Päästeameti veebileht, mille kohaselt on Tartu linnas 23 avaliku varjumiskohta (vt Lisa 3). ChatGPT 4o mõttekäik on selgem, sest on teada, et keelemudel sai enda informatsiooni artiklist, mis on avaldatud selle aasta alguses.

4.2 Usaldusväärsed allikad

Usaldusväärsete allikate kategoorias hinnati peamiselt keelemudelite oskust kasutada usaldusväärseid allikaid. Usaldusväärseid allikaid hinnati kolme erineva kriteeriumi põhjal: autor, väljaanne ja allika tüüp. Autori puhul kontrolliti, milline on tema erialane kvalifikatsioon ja töökoht. Oluline oli mõista, kas autor kirjutab teksti kasu püüdmiseks või laiema avalikkuse informeerimiseks. Selle jaoks tuli uurida, millises väljaandes on artikkel avaldatud ja millist tüüpi allikaga on tegu. Kui artikkel on ilmunud näiteks Delfi uudisteportaalis, kuid on märgitud sisuturundusena, ei ole tegu usaldusväärse allikaga. Usaldusväärsete allikate hindamiseks on kasutatud Tartu ülikooli Akadeemiliste tekstide kirjutamise teavikut.

Kategoorias hinnati keelemudelite võimekust kasutada usaldusväärseid ja erapooletu seisukohaga allikaid. DeepSeek V3, Gemini 2.0 Flash ja Claude 3.7 Sonneti puhul oli kõige suuremaks probleemiks internetile ligipääsu puudumine. DeepSeek V3 andmed on värskendatud 2024. aasta juunini, Gemini 2.0 Flash andmed 2023. aasta alguseni ja Claude 3.7 Sonneti andmed 2024. aasta juunini. Kõige rohkem tekitas muret see, et eelmainitud kolmest keelemudelitest vaid Claude 3.7 Sonnet andis kontrollijale teada, et selle mudeli andmed ei ole tänapäevani uuendatud, mis võib faktikontrollijat eksitada.

DeepSeek V3 oli nendest kolmest ainukene, kes lisas enda viidetele hüperlingid, et kontrollija saaks allika usaldusvärsust mugavamini kontrollida. Kuid ükski viidatud link ei viinud kontrollijat info allikani, vaid pigem kodulehe avalehele või vigasele leheküljele. DeepSeek viitas enda faktikontrollis Päästeametile, Tartu linnavalitsuse kodulehele, ERR-i artiklile ja Riigi Teatajale, mis on kõik usaldusväärsed allikad, kuid mõnda neist ei eksisteerinud, mistõttu sai keelemudel allikate usaldusvärsuse kategoorias hinnanguks –2 (vt Tabel 7). Viidete kasutamist hinnatakse eelolevas peatükis (vt ptk 4.3).

Näide allikast, mida ei eksisteeri:

*„Allikas: [ERR uudis "Eesti suured linnad laiendavad varjendivõrgustikke"](#) (2023)”
(DeepSeek V3, 2025).*

Gemini keelemudel 2.0 Flash ei viidanud enda faktikontrollis mitte ühelegi allikale vaid andis suuniseid kuidas infot leida. Suunised olid pealiskaudsed ja ei suunanud kontrollijat info allikani. Mainitud allikaid (Päästeamet, Tartu linnavalitsus, Eesti meedia) ei saa kontrollida, mistõttu ei saa allikate usaldusvärsust positiivselt hinnata. Gemini 2.0 Flash sai allikate usaldusvärsuse kategoorias hinnanguks –2. Sarnaselt Geminile ei viidanud ühelegi allikale Claude keelemudel 3.7 Sonnet, mis jagas samuti vaid suuniseid, kust infot leida võib. Selle põhjal said kõik kolm eelnevalt mainitud keelemudelit hinnanguks –2. Hinnang –2 ei tähenda siin kontekstis aga seda, et kõik keelemudelid olid samal tasemel, vaid eksperdi esitatud allikad olid palju usaldusväärsemad. Kolmest keelemudelist esitas kõige usaldusväärsemaid allikaid DeepSeek V3, sest need olid esitatud kõige konkreetsemalt.

Näide allikast, mis on esitatud liiga üldiselt või umbkaudselt, mis ei luba kontrollida allika usaldusvärsust:

„Viide: Erinevad uudisportaalid (nt ERR Uudised, Postimees, Delfi) on avaldanud artikleid varjumiskohtade kohta. (Konkreetsete artiklite otsimine ja viitamine reaalses ei ole võimalik, kuid need on üldteadaolevad infoallikad.)” (Google Gemini 2.0 Flash, 2025).

Tabel 7. Faktikontrolli hindamismatriks – allikate usaldusväärsus

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Allikate usaldusväärsus	-2	-2	0	-2	-1

ChatGPT 4o ja Grok 3 kasutasid faktikontrollis palju usaldusväärsemaid allikaid kui eelnevad keelemudelid. Faktidele toetumise kategoorias võrreldi nende kahe keelemudelit oskust leida faktipõhine info (vt ptk 4.1) Mõlemad keelemudelid kasutasid usaldusväärseid allikaid vaatamata sellele, et info oli ChatGPT 4o allika puhul aegunud ja Grok 3 ei kasutanud allikast leiduvat informiooni. Ajakohaste allikate kohta jagatakse eraldi hinnang (vt ptk 4.7). Sellegipoolest viitasid keelemudelid enda allikate nimekirjas korrektsetele alamehtedele, mis viisid otse informatsiooni allikani. Kuigi ChatGPT 4o kasutas muuhulgas ka Reformierakonna veebilehe arvamust, mida ei saa poliitiliselt neutraalseks lugeda, mistõttu kaotab keelemudel punkti.

4.3 Viidete kasutamine

Viidete kasutamise kategoorias hinnati keelemudelite oskust viidata enda allikatele korrektselt ja järgitavalt. Korrektsuse all ei mõelda konkreetset viitamise stiili (APA, MLA, Chicago), vaid seda kuidas on viide lisatud, sõnastatud ja kasutatud. Viide peab kindlasti olema järgitav, et peale klikates saaks edasi liikuda viidatud sisule.

Näide korrektsest ja järgitavast viitest:

„Tartu Postimees: „Tartus märgistati esimene avalik varjumiskoht” – 28.06.2022 (ülevaade Tartu varjumiskohtadest) tartu.postimees.ee” (ChatGPT 4o, 2025).

Kategoorias saavutasid sarnaseid tulemusi DeepSeek V3, Gemini 2.0 Flash ja Claude 3.7 Sonnet. Kuigi DeepSeek V3 lisas enda töösse hüperlinke, polnud need korrektselt esitatud, järgitavad ega täpsete pealkirjadega, mistõttu polnud võimalik allikaid kontrollida. Sarnased puudused esinesid ka Gemini 2.0 Flash ja Claude 3.7 Sonneti puhul, kus tihti ei olnud lisatud isegi avaldamise aastat või kuupäeva. Seevastu Grok 3 ja ChatGPT 4o esitasid oma

järeldustes selged ja kontrollitavad viited. Grok 3 puhul ilmnes siiski probleem tekstisestest viidetega, mis esitati kontrollijale vigaste URL-idenä. Nende tähelepanekute põhjal said DeepSeek V3, Gemini 2.0 ja Claude 3.7 Sonnet hindeks –2, Grok 3 –1 ning ChatGPT 4o hindeks 1, sest võrreldes eksperdiga esitas ChatGPT 4o allikaid paremini, lisades faktikontrolli lõppu allikate nimekirja (vt Tabel 8).

Tabel 8. Faktikontrolli hindamismatriks – viidete kasutamine

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Viidete kasutamine	–2	–2	–1	–2	1

Siinkohal tuleb märkida, et keelemudelitel on võimalus koguda punkte vaid siis, kui keelemudel suudab kindlas kategoorias ületada eksperti. Eksperti ületamine ei ole lihtne, sest Eesti Päevalehe faktikontrolli eksperdi töö on üldjuhul tehtud täpsusega ja vastab sellele, kuidas rahvusvahelise faktikontrolli võrgustiku (IFCN) põhimõtete koodeks ette näeb (International Fact-Checking Network, i.a). Keelemudel kogub punkte, kui tehakse midagi mida pole nõutud, kuid mis aitab lugejal/kontrollijal paremini faktikontrolli mõista. Viidete kasutamise kategooria on mõtte selgitamise jaoks hea näide, sest DeepSeek V3, Gemini 2.0 Flash ja Claude 3.7 Sonnet jäid eksperdile alla, kuid ChatGPT 4o suutis lisada allikate nimekirja, mis aitas kontrollijal info allikaid mugavamalt kontrollida. Grok 3 oleks peaaegu sooritanud sarnase tulemuse, kuid kontrollija leidis teksti seest vigaseid URL-e. Kui tekstisestest viiteid poleks üldse olnud oleks punktisumma olnud samaväärne eksperdiga.

Näide vigasest tekstisestest viitest:

„Eesti Päästeameti veebileht ([invalid url, do not cite]) annab üldist infot varjumiskohtade kohta, sealhulgas soovitusi, kuidas ohu korral käituda” (Grok 3, 2025).

Vigaste URLide, pealkirjade või info jagamine eksitab kontrollijat palju rohkem kui nende puudumine. See rõhutab, et tehisintellektide puhul pole oluline vaid viite olemasolu, vaid ka nende korrektsus ja kontrollitavus.

4.4 Läbipaistvus

Läbipaistvuse kategoorias oli kõige rohkem märgata hallutsineerimist ja kontrollija eksitamist. Läbipaistvuse all mõistetakse töös keelemudeli võimet esitada informatsiooni arusaadavalt, jälgedega ja argumenteeritult. See tähendab seda, et faktikontrolli läbipaistvuse mõõtmisel hinnatakse, kas keelemudel on enda mõttekäiku selgitanud, lisanud faktidele või väidetele konteksti ja kasutanud viidatud allikaid tekstis.

Peatükis on selgunud, et kõige rohkem jäävad enda mõtete põhjendamisega teistele alla esimesed kolm keelemudelit – DeepSeek V3, Gemini 2.0 Flash ja Claude 3.7 Sonnet. Nende eelmainitud keelemudelite faktikontrolli metoodika põhjendamine piirdus kas veebilehe nime või allika üldnimetuse jagamisega. Üldiste viidete jagamine ei anna lugejale märku, kust on täpsemalt info saadud. Selle põhjal ei saa lugeja informatsiooni usaldusväärseks lugeda ja faktikontroll ei tundu seeläbi adekvaatne.

„Allikas: Tartu Linnavalitsuse pressiteade (2022)” (DeepSeek V3, 2025).

Sarnaseid üldisi viiteid kasutasid kõik kolm eelmainitud keelemudelit, mispärast saavad kõik kolm keelemudelit ühiselt hindeks –2. Grok 3 keelemudeli põhjal on hinnangu jagamine keerulisem, sest keelemudel on kõige põhjalikumalt põhjendanud enda mõttekäiku, kuid selle põhjal pole endiselt kohati võimalik mõista, kuidas järelduseni jõuti. Keelemudel on suutnud väga detailselt põhjendada, mida ta uuris ja miks ta uuris. Faktikontrollis on isegi olemas metodoloogia ja allikate valiku peatükk, mis peaks justkui lugejale näitama, kuidas järeldusteni jõuti.

„Faktikontrolli läbiviimiseks kasutati mitmeid allikaid, sealhulgas Tartu linna ametlikku veebilehte, mis pakub kõige ajakohasemat infot kohalike varjumiskohtade kohta. (...) Tartu Postimehe artiklid, et mõista varjumiskohtade ajalugu ja arengut. (...) Kontrolliti ka Eesti Päästeameti veebilehte, mis annab üldist infot varjumiskohtade kohta Eestis” (Grok 3, 2025).

Kõik lõigus mainitud on väga hea selgitus ja kirjeldus allikate valiku kohta, mistõttu tekib küsimus, kust on saadud väljamõeldud arvud (nt 26 avalikku varjumiskohta tartus) – hinnanguks –1. ChatGPT 4o on suutnud enda mõttekäiku kõige läbipaistvamalt põhjendada, sest nad on kasutanud korrektseid tekstisiseseid viiteid nagu kasutas faktikontrolli ekspert.

Samuti on aru saada, kuidas on jõutud vastuseni, et Tartus on 23 varjendit, sest nad viitavad allikale, kus on seda mainitud. Seetõttu saab ChatGPT 4o hinnanguks 0 (vt Tabel 9).

Tabel 9. Faktikontrolli hindamismatriks – läbipaistvus

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Läbipaistvus	-2	-2	-1	-2	0

Läbipaistvuse hindamisel on kõige keerulisem osa keelemudeli uskumine. Kui eelnevalt on kõik keelemudelid öelnud, et nemad said oma informatsiooni Tartu linna veebilehelt, kuid nad ei ole lisanud selle väite toetamiseks konkreetset viidet, pole seda võimalik kontrollida. Kõik keelemudelid kasutasid Tartu linna veebilehte allikana, kuid ainult ChatGPT 4o ja Grok 3 suutsid näidata, et nad on seda kasutanud.

4.5 Erapooletus

Erapooletuse kategoorias üllatasid kõik keelemudelid (v.a ChatGPT 4o) enda teadmispõhise ja hinnangute vaese kontrolliga. ChatGPT 4o kasutas faktikontrollis Reformierakonna veebilehe artiklit, mis toob alla nende faktikontrolli usaldusväarsuse. Olgugi, et Reformierakond on Tartus võimul ja võib jagada Tartu eelarve kohta usaldusväärset informatsiooni, ei saa allikat lugeda usaldusväärseks allikaks.

Tabel 10. Faktikontrolli hindamismatriks – erapooletus

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Erapooletus	1	1	1	1	-1

Ühestki teisest keelemudeli poolt teostatud faktikontrollist ei saanud välja lugeda hinnanguid Sandra Lauri isiksuse kohta ega ei mainitud ühtegi poliitilist parteid. Keskenduti puhtalt väitele ja selle tõesuse kontrollimisele. Ekspert on enda faktikontrollis jaganud aga konteksti mainides, et Sandra Laur on Eesti Keskerakonna liige. Samuti otsustas ekspert kasutada enda faktikontrollis Sandra Lauri tsitaati:

„Laur räägib ka nõukogudeaegsetest varjenditest, mida Eestis on omajagu, aga mille seisukord ei ole hea või on hindamata. Seistes ühe varjendi ees Tartus, nendib Laur: „27 aasta jooksul, mil Reformierakond on Tartus võimul olnud, pole linn pidanud vajalikuks varjendeid korda teha.” (Somelar, 2025).

Tsitaadile oleks võinud koheselt järgneda ka selgitus, miks pole linn varjendeid korda teinud. Eksperti faktikontrollis selgitus küll lisatakse, kuid mitte kohe tsitaadi järgi. See pole otseselt eksperti enda kallutatuse näitaja, kuid valik väidet faktikontrollis kasutada oli ikkagi eksperti otsus. Kõige erapooletum variant oleks olnud poliitiliselt kallutatud tsitaadist hoidumine. Teise märkusena kasutas ekspert faktikontrollis lauset:

„Ehkki ta eristab mõisteid „varjend” ja „varjumiskoht,” on nende tähendus tema videos sassis” (Somelar, 2025).

Sõna „sassis” kasutamist võib käsitleda kui tõlgenduslikku hinnangut, kuna seda väidet ei toeta otseselt ükski fakt. Selle analüüsi põhjal saab ChatGPT 4o erapooletuse kategoorias hindeks -1 ja teised keelemudelid 1 (vt Tabel 10). Hinnangul tuleb silmas pidada, et “1” ei tähenda, et eksperti faktikontroll oli kallutatud, vaid et tehisintellekti vastused olid konkreetses vormistusviisis erapooletumad.

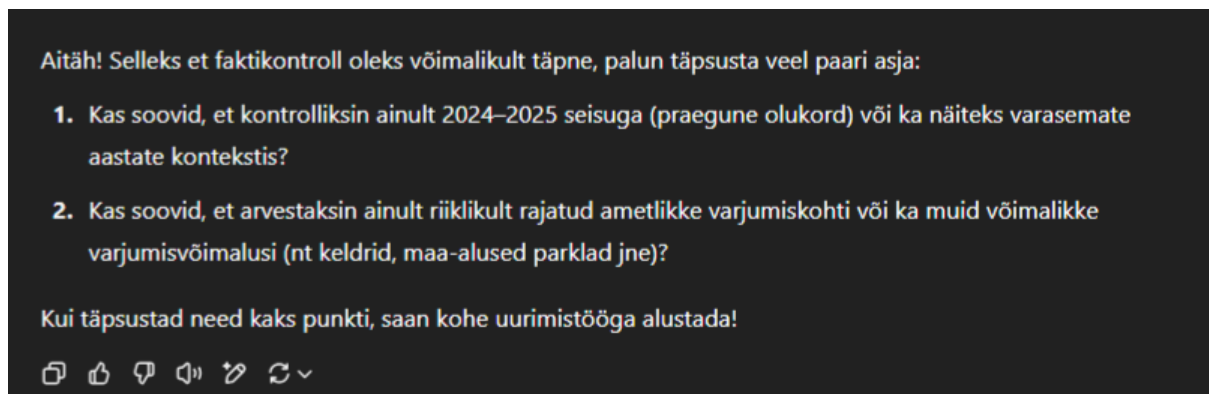
4.6 Põhjalikkus

Põhjalikkuse all peetakse hindamismaatriksis silmas seda, kui mitmekülgsest keelemudel või ekspert käsitleb kontrollitavat väidet. Arvesse on võetud vastuse pikkust, selgituse mitmekesisust, faktide ja konteksti seletamist ning visuaalsete elementide (nt tabelid) kasutamist.

Kõige põhjalikuma ja struktureerituma faktikontrolli esitas Grok 3 keelemudel. Grok 3 eristus teistest mudelitest eelkõige selle poolest, et keelemudeli vastuses oli kasutatud konkreetset loetelu varjumiskohtade jaotuse kohta kui teistes faktikontrollides olid esindatud vaid mõned varjumiskohtade asukohad. Isegi kui mõned nimetatud arvud ei vastanud täielikult ametlikele

või viidatud allikatele, näitas mudel siiski soovi liigitada olemasolevat teavet ning esitada seda lugejale võimalikult loetavalt. Faktikontroll koosnes kümnest alapeatükist, mis kõik sisaldasid kasulikku informatsiooni. Lisaks pakkus Grok 3 laiemat konteksti, kirjeldades ajalugu ja faktikontrolli metoodikat.

Järgmisena paistis põhjalikkuse poolest silma OpenAI keelemudel ChatGPT 4o. See mudel oli teistest keelemudelitest ainus, kes küsis enne promptile vastamist täpsustusi (vt joonis 3). Tänu täpsustustele suutis keelemudel tuvastada Sandra Lauri isiku ja väite allika, kus oli räägitud ka Tartu linna eelarvest. Faktikontroll sisaldas viiteid Tartu eelarvele, varjumiskohtade arvule ning viitas konkreetsetele hoonetele. Lisaks selgitas ChatGPT 4o ka poliitilist ja seaduslikku tausta. Kuigi ChatGPT 4o ei kasutanud faktikontrollis visuaalseid lahendusi nagu Grok 3, oli selle vastuse ülesehitus siiski väga informatiivne ning toetatud allikatega.



Joonis 3. ChatGPT 4o vastus promptile. (Allikas: ChatGPT 4o, 2025)

Kõige lühema ja sisuliselt nõrgeima faktikontrolli esitas Claude 3.7 Sonnet (vt Lisa 4). Keelemudeli vastus jäi pinnapealseks ega läinud süvitsi nagu eelmainitud keelemudelid. Faktikontrollist puudusid konkreetsed faktid, numbrid, statistika ja taust. Vastus oli konkreetne ja minimaalne, mis oli pigem kokkuvõtlik kui selgitav.

Tabel 11. Faktikontrolli hindamismatriks – põhjalikkus

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Põhjalikkus	-2	-2	-1	-2	-1

Ülejäänud keelemudelite, DeepSeek V3 ja Gemini 2.0, vastused olid sarnasel tasemel. Mõlemad pakkusid üldist teavet, kuid ei eristunud eriti Claude 3.7 faktikontrollist. DeepSeek V3 ja Gemini 2.0 suutsid küll esitada mitu alapeatükki, kuid need oli üpriski sisutühjad ja ei käsitletud teemat süvitsi, mistõttu said nad kõik sarnase hinnangu (vt Tabel 11). Sellegipoolest oli eksperdi poolt tehtud töö põhjalikum, sest kaasatud olid suulised allikad, keda keelemudelid ei kasutanud.

4.7 Aktuaalsus

Aktuaalsuse kategooria eesmärk on hinnata, kui võrd ajakohastele ja aktuaalsetele allikatele tuginesid keelemudelid faktikontrolli koostamisel. Arvesse võeti, kas kasutati kõige värskemad informatsiooni, mis selle väite puhul oli 2025. aasta kevad. Kuna varjumiskohtade arv muutub pidevalt, oli oluline, et keelemudelid esitaksid kõige aktuaalsemat informatsiooni. Kategoorias ei suutnud ükski keelemudel esitada aktuaalsemaid allikaid kui ekspert. Ekspert kasutas kõige värskemaid andmeid, kaasates faktikontrolli suulised allikad Tartu linnavalitsusest. Ekspert kajastas ka hiljutist arengut, mille kohaselt on Tartu linnal 24 varjumiskohta, mitte 23, nagu varasemates allikates mainitud.

DeepSeek V3, Gemini 2.0 Flash ja Claude 3.7 Sonnet saavutasid kategoorias maksimaalselt negatiivse hinde, sest neil puudusid allikatel aastaarvud või kasutasid aegunud informatsiooni. Grok 3 kasutas lisaks ajakohastele allikatele ka aegunud ja ebavajalikke allikaid. ChatGPT 4o kasutas varjumiskohtade määramiseks selle aasta alguses avaldatud allikat, kuid see pole enam aktuaalne, sest varjumiskohtade arv on nüüdseks muutunud (vt Lisa 5).

Kolm keelemudelit, DeepSeek V3, Gemini 2.0 Flash ja Claude 3.7 Sonnet, saavutasid aktuaalsuse kategoorias maksimaalselt negatiivse hinde. Nende faktikontrollis kasutati aegunud allikaid või puudus ajaline määratlus täielikult. See tuleneb ilmselt sellest, et keelemudelite enda andmed on aegunud.

„Märkus: Minu viimane teadmiste uuendamine oli 2024. aasta oktoobris, seega on võimalik, et pärast seda kuupäeva on toimunud muudatusi varjendite arvus või nende seisundis” (Claude 3.7 Sonnet, 2025).

Grok 3 keelemudel kasutas ajaliselt sobivaid viiteid, kuid ei suutnud neist välja lugeda vajalikku informatsiooni. Lisaks on allikate nimekirjas viidatud 2003. aasta artiklile, kuid seda kasutati vaid ajaloolise konteksti lisamiseks, mistõttu sai Grok 3 väga hästi hakkama ajakohaste allikate leidmisega (vt Tabel 12).

ChatGPT 4o tugines varjumiskohtade määramisel artiklile, mis oli avaldatud 2025. aasta alguses ja viitas seeläbi 23 varjumiskohale. ERR-i artikkel, kust ChatGPT 4o enda andmed sai, oli faktikontrolli teostamise ajaks aegunud. Ametlikult oli lisandunud veel üks varjumiskoht Lutsu tänaval, mis oli puudu ka Päästeameti nimekirjast. Selline allikate kasutus näitab, et ChatGPT 4o ei kasutanud ajakohast informatsiooni, mille tulemusena ei jaga keelemudel faktikontrollis korrektset informatsiooni.

Tabel 12. Faktikontrolli hindamismatriks – aktuaalsus

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Aktuaalsus	-2	-2	0	-2	-1

Aktuaalsuse puudumine võib vähendada faktikontrolli usaldusväarsust, kuna väited võivad põhineda aegunud või vahepeal muutunud asjaoludel.

4.8 Faktikontrolli tulemused

Faktikontrolli tulemuste kategooria eesmärk oli võrrelda, kuidas keelemudelid sõnastasid enda hinnangu väite tõesusele. Enda hinnangu pidid keelemudelid jagama väitele: „Tartu linnas on vaid üks varjend, kuhu sõjalises olukorras varjuda.” Eksperti faktikontrolli otsus oli eksitav. See tähendas, et väide sisaldas tõeseid elemente (Tartus on üks varjend), kuid see andis vale üldmulje, et Tartus oleks justkui ainult üks koht, kuhu saab sõjalise olukorras varjuda. Eksperti selgituse kohaselt saab sõjalise ohu korral varjuda ka varjumiskohtadesse, mis on küll leebemate nõuetega, kuid siiski mõeldud elanikkonna kaitseks. Sandra Lauri

sõnastus eksitab, kuna ta vastandab varjendit ja varjumiskohta, mis tähendaks, et varjumiskohad pole sõjalise olukorra puhul kasutatavad.

Keelemudelite otsused olid sisuliselt sarnased, kuid sõnastuselt erinesid nad eksperdi otsusest:

- DeepSeek V3 andis väitele hinnangu **vale**. Samas ei teinud mudel vahet varjendil ja varjumiskohal, mis tähendab et järeldus põhines ebatäpsusel. Samuti ei ole Sandra Lauri väide täiesti vale, sest Tartus on tõepoolest hetkel vaid üks varjend.
- Gemini 2.0 Flash andis väitele hinnangu **faktiliselt ebatäpne**, viidates et Tartus on mitmeid varjumiskohti. See vastab ka sisuliselt eksperdi hinnangule, kuigi sõnastus on erinev.
- Grok 3 andis väitele hinnangu **tõenäoliselt vale**, selgitades et korrektset varjumiskohtade arvu polnud võimalik leida. Selline järeldus ei aita lugejal mõista faktikontrolli tõesust, sest lõplik hinnang puudub.
- Claude 3.7 Sonnet andis väitele kaks hinnangut, **eksitav** ja **väär**, kuid lisas juurde, et Tartus on mitmeid varjumiskohti ja ka varjendeid. Viimane väide ei vasta tõe, kuna Tartus neid mitmeid ei ole. Claude keelemudeli järeldus põhines faktivigadel.
- ChatGPT 4o andis väitele hinnangu **vale**. Põhjendades seda sellega, et Tartus on mitmeid avalikke varjumiskohti, kuhu saab sõjalise olukorra puhul varjuda. Keelemudeli põhjendus oli küll kooskõlas eksperdi arusaamaga, kuid järelduse sõnastamisel kasutati karmimat keelt ning ei eristatud piisavalt varjendit ja varjumiskohta.

Kokkuvõttes võib öelda, et ainult Google Gemini 2.0 Flash jõudis sisuliselt sama otsuseni kui eksper, kirjeldades otsust kui faktiliselt ebatäpset (vt Tabel 13).

Tabel 13. Faktikontrolli hindamismaatriks – faktikontrolli tulemus

Kategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o

Faktikontrolli tulemus	-1	0	-1	-1	-1
------------------------	----	---	----	----	----

Ülejäänud keelemudelid, DeepSeek V3, Grok 3, Claude 3.7 Sonnet ja ChatGPT 4o, kaldusid eksperdi otsusest küll veidi kõrvale, kuid mitte ülemäära.

4.9 Järeldused

Antud uurimistöö eesmärk oli võrrelda erinevate tehisintellekti keelemudeli faktikontrolli kvaliteeti ekspedi poolt teostatud faktikontrolliga. Prooviti selgeks teha, millised on tehisintellekti võimekused ja puudujäägid. Analüüs viidi läbi kaheksas kategoorias: faktidele toetumine, allikate usaldusväarsus, viidete kasutamine, läbipaistvus, erapooletus, käsitluse põhjalikkus, aktuaalsus ja faktikontrolli tulemused. Tulemuste põhjal saab järeldada, et kuigi tehisintellekti keelemudelid saavad esitada sisulisi, eestikeelseid, kontrollitavaid ja usaldusväärseid tekste, jäävad nad siiski mitmes aspektis eksperdist maha.

Kõige paremaid tulemusi näitas ChatGPT 4o, mille faktikontroll oli struktureeritud, sisutihe ning põhjendatud. Siiski esines probleeme erapooletuse ja aktuaalsuse kategooriates, kus oli kasutatud aegunud allikat ja Reformierakonna veebilehelt leitavat arvamulugu. Sellegipoolest leiab uuringu põhjal ChatGPT 4o keelemudel rakendust viidete kasutamisel, läbipaistvuse näitamisel ja faktidele toetumisel, mis võib aidata eksperte faktikontrolli läbiviimisel. Keelemudelile järgnes ühe punktiga Grok 3, mis suutis esitada kõikidest keelemudelitest kõige põhjalikuma ning kõige usaldusväärsemate ajakohaste allikatega faktikontrolli. Samas oli Grok 3 nõrgem faktuaalsuses, kus mudel ei kasutanud viidatud allikatest leitavat informatsiooni vaid mõtles enda vastuse välja.

Google Gemini 2.0 Flash oli ainult keelemudel, kelle otsus väite tõesusele vastas sisuliselt eksperdi hinnangule, mistõttu võib mudelid pidada järelduse täpsuse poolest edukaimaks. Paraku puudus faktikontrollist põhjalikkus ja ajakohasus, mis piirab selle kasutust faktikontrollina. Anthropic Claude 3.7 Sonnet ja DeepSeek V3 esitasid kvaliteedilt kõige kehvemad faktikontrollid. Claude mudel ei põhjendanud enda mõttekäiku ja kasutas segamini mõisteid varjend ja varjumiskoht. DeepSeek V3 aga viitas väljamõeldud allikatele ja pealkirjadele, mida ei eksisteeri. Seetõttu paiknevad nad lõplikus hindamistabelis viimaste seas (vt Tabel 14).

Tabel 14. Keelemudelite võimekus teostada faktikontrolli võrreldes eksperdiga

Hindamiskategooria	DeepSeek V3	Gemini 2.0	Grok 3	Claude 3.7	ChatGPT 4o
Faktidele toetumine	-2	-2	-2	-1	0
Allikate usaldusväärsus	-2	-2	0	-2	-1
Viidete kasutamine	-2	-2	-1	-2	1
Läbipaistvus	-2	-2	-1	-2	0
Erapooletus	1	1	1	1	-1
Käsitluse põhjalikkus	-2	-2	-1	-2	-1
Aktuaalsus	-2	-2	0	-2	-1
Faktikontrolli tulemus	-1	0	-1	-1	-1
Kogusumma	-12	-11	-5	-11	-4

Kokkuvõtlikult võib järeldada, et tehisintellekti keelemudelid suudavad pakkuda faktikontrolli struktuuri ja üldise sisu tasemel arvestatavat tuge. Kuid jäävad eksperdile kõvasti alla usaldusväärsete allikate hankimisega ja ajakohase info leidmisel. Keelemudelite kasutamine võiks sellisel juhul olla eksperdi tööd toetav, mitte asendav tööriist.

5. DISKUSSIOON

Uurimistulemusi lugedes tuleb silmas pidada, et uuring on läbi viidud ühe konkreetse väite põhjal, mis ei pruugi kajastada kõige täpsemaid tulemusi tehisintellekti kasutamise kohta faktikontrollis. Võimalik, et mõne teise teema puhul (nt majandus või tervishoid) oleksid keelemudelite tugevused ja nõrkused avaldunud teisiti, mistõttu võivad suurema valimi puhul tulemused varieeruda. Töös esitatud hindamismaatriksi alusel (vt Tabel 5) hindamine võib tuua erinevate hindajate puhul sama väite hindamisel erinevaid tulemusi. See tähendab seda, et iga hindaja annab tehisintellektile enda subjektiivse hinnangu kui hästi sai tehisintellekt võrreldes eksperdiga hakkama. Siin töös on kasutatud negatiivseid hinnanguid eesmärgiga, et näidata kas keelemudel oli palju kehvem või natuke kehvem kui ekspert. Ehk siis juhul, kui kolm keelemudelit saavad samas kategoorias hinnanguks -2, ei tähenda see seda, et nad olid kategoorias võrdsel tasemel, vaid nad olid kõik kategoorias palju kehvemad kui ekspert. Täpsuse lisamiseks saab skaalat laiendada (nt -5 kuni 5), et näidata rohkem erinevusi keelemudelite vahel. Selle töö eesmärk ei olnud aga keelemudeleid võrrelda, vaid uurida millisel määral suudaks tehisintellekt teostada iseseisvalt täiemahulist faktikontrolli.

Keelemudelite kasutamisel on oluline teada, milliste keelemudelitega oleks kõige parem faktikontrolli teostada, sest keelemudeleid on maailmas palju ja paljudel neist on kindel otstarve. Valimit koostades lähtuti sellest, et iga keelemudel oleks erinevast perekonnast. Tehisintellekti keelemudelite juures viidatakse perekonnale kui keelemudeli tehnoloogiale (nt ChatGPT 4o ja GPT-3.5 on samast OpenAI perekonnast). Seda tehti selleks, et saada paremat arusaama, kuidas erinevad tehnoloogiad sama ülesandega hakkama saavad. Hindaja teadlikkus erinevatest keelemudelistest ja nende maine kohta võis alateadlikult mõjutada hinnangute kujunemist. Näiteks hinnati ChatGPT 4o vastuseid erapooletuse kategoorias algselt heaks, kuid hiljem selgus, et mudel oli kasutanud faktikontrollis allikat, mis pärines poliitilise partei veebilehelt (vt ptk 4.5). Samuti oleks saanud laiemalt uurida faktikontrolli sõnastust, tooni ja loetavust, ehkki ükski neist ei olnud eraldi kategooria. Neid kõiki uuriti pealiskaudselt põhjalikkuse kategoorias (vt ptk 4.6).

Autorile teadaolevalt Eesti Päevalehe poolt koostatud faktikontrolli lõpuks ei avaldatud, kuna võeti vastu toimetuskollegiumi otsus artikkel mitte avaldada, sest Sandra Lauri video puhul, mida kontrolliti, oli piir vale, eksitava, konteksti puudumise ja tõe vahel üsnagi hägune. Uurimistöö

autor sai artiklile ligipääsu privaatse hüperlingi kaudu, mille põhjal uuring läbi viidi. Uuringu tulemusi artikli mitte avaldamine ei mõjutanud.

KOKKUVÕTE

Bakalaureusetöö eesmärk oli hinnata tehisintellekti keelemudelite võimekust teostada täiemahulist faktikontrolli. Selle jaoks võrreldi keelemudelite poolt teostatud faktikontrolle eksperdi poolt teostatud faktikontrolliga. Töö keskendus küsimusele, kas ja mil määral suudavad tehisintellekti mudelid (DeepSeek V3, Gemini 2.0 Flash, Grok 3, Claude 3.7 Sonnet ja ChatGPT 4o) toetuda faktidele, pakkuda usaldusväärseid ja aktuaalseid allikaid, kasutada viiteid, teha kontroll läbipaistvalt ja erapooletult ning esitada töö põhinev järeldus, mis oleks võrreldav eksperdi faktikontrolliga.

Uuring viidi läbi kahes etapis. Esmalt viidi läbi pilootuuring, mille eesmärgiks oli testida autori loodud hindamismaatriksit ning täiustada keelemudelitele esitatavat sisendit. Pilootuuring võimaldas tuvastada mitmeid puudusi senistes hindamismeetodites, mille tulemusena loodi täiustatud maatriks skaalal -2 kuni 2 . Seejärel viidi läbi uuring, mille käigus võrreldi tehisintellekti keelemudelite vastuseid Delfi uuriva toimetuse eksperdi poolt koostatud faktikontrolliga.

Tulemused näitasid, et tänapäeva tehisintellekti keelemudelite võimekused ei ole piisavalt head, et teostada iseseisvalt täiemahulist faktikontrolli. Igal keelemudelil olid oma plussid ja miinused, mis takistasid või toetasid faktikontrolli teostamist. Keelemudelid suutsid eristada fakte ja väljamõeldisi ning jõudsid kõik järeldusele, et väide ei vasta tõele. Vastused keelemudelite vahel enamasti varieerusid, sest iga keelemudel põhines erineval treeningul (vt ptk 1.3), kuid esines ka mitmeid sarnasusi ülesande lähenemisele. Enamasti ei olnud niivõrd tõhusad mudelid, mis ei esitanud faktikontrollis konkreetseid allikaid. Internetile ligipääs ja ligipääsu kasutamine mõjutas kõige rohkem mudelite hinnangut, sest faktikontroll põhineb kontrollitavatel väidetel.

Autorile teadaolevalt on tegemist esimese tööga Eestis, kus on välja töötatud spetsiaalne hindamismaatriks tehisintellekti poolt koostatud faktikontrollide võrdlemiseks eksperdi poolt teostatud faktikontrolliga. Uuring näitas, et tehisintellekt ei suuda iseseisvalt ja usaldusväärselt faktikontrolli teostada, küll aga võib see olla kasulik tööriist ja abiline ekspertidele. Sellest tulenevalt pakub töö aluse edasisteks uuringuteks, kus tehisintellekti keelemudelite faktikontrolli võimekust saab veelgi täpsemalt uurida (nt rohkemate väidete ja mudelite põhjal), samuti täpsustada, millistes valdkondades on tehisintellekt usaldusväärne ja millistes mitte.

SUMMARY

„Generative AI's Ability to Verify Claims Compared to Delfi Fact-Checking”

The aim of this bachelor’s thesis was to assess the capability of artificial intelligence language models to carry out comprehensive fact-checking. For this purpose, fact-checks conducted by language models were compared to those done by experts. The thesis focused on the question of whether and to what extent AI models (DeepSeek V3, Gemini 2.0 Flash, Grok 3, Claude 3.7 Sonnet, and ChatGPT 4o) are able to rely on facts, provide reliable and up-to-date sources, use references appropriately, conduct checks transparently and impartially, and present a conclusion based on their analysis that is comparable to that of an expert fact-checker.

The research was carried out in two stages. First, a pilot study was conducted with the aim of testing the author’s self-developed evaluation matrix and refining the input prompts presented to the language models. The pilot study revealed several shortcomings in previous assessment methods, which led to the creation of an improved matrix using a –2 to 2 scale. This was followed by the main study, in which responses generated by the AI models were compared with a fact-check composed by an investigative journalist from Delfi’s editorial team.

The results showed that the current capabilities of AI language models are not yet sufficient to independently carry out comprehensive fact-checking. Each model had its own strengths and weaknesses that either hindered or supported the execution of fact-checking. The models were able to distinguish between facts and fabrications, and all concluded that the given claim was false. The responses generally varied between models due to differences in their training (see Chapter 1.3), although there were also notable similarities in how they approached the task. Models that did not provide specific sources generally underperformed in the fact-check. Internet access and how it was utilized had the greatest influence on the evaluation of the models.

To the author’s knowledge, this is the first work to develop a dedicated evaluation matrix for comparing AI-generated fact-checks to expert assessments. The thesis showed that AI cannot independently and reliably perform fact-checking, but it may serve as a useful tool and assistant for human experts. As such, this work provides a foundation for further studies into the fact-checking capabilities of AI language models (e.g., using a broader set of claims and

models) and for identifying in which domains AI can be considered reliable and in which it cannot.

KASUTATUD KIRJANDUS

Adinath, D. R., ja Smiju, I. S. (2025, 19. märts). *Manus AI, Gemini, Grok AI, DeepSeek, and ChatGPT: A Comparative Analysis of Advancements in NLP*. SSRN. <https://dx.doi.org/10.2139/ssrn.5185131>

AI Eesti. (2025, 22. märts). *Claude sai (lõpuks) ligipääsu internetile*. AI Eesti. Kasutatud 20.04.2025, <https://aieesti.ee/claude-sai-lopuks-ligipaasu-internetile/>

Angeletti, N. ja Oliva, A. (2010). *Time: The illustrated history of the world's most influential magazine*. New York, NY: Random House.

Anthropic. (2025a). *Claude (versioon 3.7 Sonnet)*. Kasutatud 22.04.2025, <https://claude.ai>

Anthropic. (2025b). *Claude 3.7 Sonnet System Card*. Anthropic. <https://assets.anthropic.com/m/785e231869ea8b3b/original/claude-3-7-sonnet-system-card.pdf>

Anthropic. (2025c, 20. märts). *Claude can now search the web*. Anthropic. <https://www.anthropic.com/news/web-search>

Aydın, Ö., Karaarslan, E., Erenay, F. S. ja Bačanin, N. (2025). Generative AI in academic writing: A comparison of DeepSeek, Qwen, ChatGPT, Gemini, Llama, Mistral, and Gemma. *arXiv*. <https://arxiv.org/abs/2503.04765>

Balasundaram, S. ja Venkatagiri, S. (2020). A structured approach to implementing Robotic Process Automation in HR. *Journal of Physics Conference Series*, 1427(1), 012008. <https://doi.org/10.1088/1742-6596/1427/1/012008>

Barua, M. (2024). Assessing the Performance of ChatGPT in Answering Patients' Questions Regarding Congenital Bicuspid Aortic Valve. *Cureus*. <https://doi.org/10.7759/cureus.72293>

Becker, J., Decker, J. A., Römmele, C., Kahn, M., Messmann, H., Wehler, M., Schwarz, F.,

Kroencke, T., ja Scheurig-Muenkler, C. (2022). Artificial Intelligence-Based detection of pneumonia in chest radiographs. *Diagnostics*, 12(6), 1465. <https://doi.org/10.3390/diagnostics12061465>

Bharadiya, J. P., Thomas, R. K., ja Ahmed, F. (2023). Rise of artificial intelligence in business and industry. *Journal of Engineering Research and Reports*, 25(3), 85–103. <https://doi.org/10.9734/jerr/2023/v25i3893>

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., Walker, K. (2020). Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft*. https://www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_WhitePaper-2020-09-22.pdf

C. J. Baby., F. A. Khan., J. N. Swathi. Home automation using IoT and a chatbot using natural language processing," *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vellore, India, 2017, 1-6, <https://doi:10.1109/IPACT.2017.8245185>

Davenport, T., Guha, A., Grewal, D. ja Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42. <https://doi.org/10.1007/s11747-019-00696-0>

DeepSeek. (2025). *DeepSeek (version v3)*. Kasutatud 22.04.2025, <https://deepseek.com>

DeepSeek-AI. (2025, 18. veebruar). *DeepSeek-V3 Technical Report*. arXiv. <https://arxiv.org/abs/2412.19437>

Kim, H.-E., Kim, H. H., Han, B.-K., Kim, K. H., Han, K., Nam, H., Lee, E. H., ja Kim, E.-K. (2020). Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study. *Lancet Digital Health*, 2(3), e138–e148. [https://doi.org/10.1016/s2589-7500\(20\)30003-0](https://doi.org/10.1016/s2589-7500(20)30003-0)

DeVerna, M. R., Yan, H. Y., Yang, K.-C. ja Menczer, F. (2023). *Artificial intelligence is ineffective and potentially harmful for fact checking*. arXiv. <https://doi.org/10.48550/arXiv.2308.10800>

Eetikakoodeks. (1998). Eesti Meediaettevõtete Liit. Kasutatud 05.01.2025, <https://meedialiit.ee/eetikakoodeks/>

European External Action Service. (2018). Action plan against disinformation. European Commission, High Representative of The Union for Foreign Affairs and Security Policy.

Faktikontroll. (i.a). Eesti Väitlusselts. <https://argument.ee/uhiskond/faktikontroll/>

Gao, T., Jin, J., Ke, Z. T. ja Moryoussef, G. (2025). A comparison of DeepSeek and other LLMs. *arXiv*. <https://arxiv.org/abs/2502.03688>

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY: Basic Books.

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>

Google. (2025). *Gemini (versioon 2.0 Flash)*. Kasutatud 22.04.2025, <https://gemini.google.com>

Google Cloud. (i.a). *Gemini 2.0 Flash*. Google Cloud. Kasutatud 22.04.2025, <https://cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-0-flash>

Griot, M., Hemptinne, C., Vanderdonckt, J., ja Yuksel, D. (2025). Large Language Models lack essential metacognition for reliable medical reasoning. *Nature Communications*, 16, 642. <https://doi.org/10.1038/s41467-024-55628-6>

Hamza, G., Es-Sadek, M. Z. ja Taher, Y. (2023). Artificial Intelligence in Self-Driving: Study of Advanced Current Applications. *International Journal of Engineering Trends and Technology*, 71(8), 225–242. <https://doi.org/10.14445/22315381/ijett-v71i8p220>

Hassabis, D. (2024). Introducing Gemini 2.0: our new AI model for the agentic era. *Google*. 11. detsember. Kasutatud 05.01.2025,

[https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024/#c-
eo-message](https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024/#c-
eo-message)

Hertzfeld, E. (2019). Study: 80 percent of consumers positive about chatbots. *Hotel Management*. 4. märts. Kasutatud 05.01.2025, <https://search.proquest.com/docview/2187686591?accountid=158798>

Hidalgo, C. A., Orghian, D., Albo-Canals, J., De Almeida, F. ja Martin, N. (2021). How humans judge Machines. In *The MIT Press eBooks*. <https://doi.org/10.7551/mitpress/13373.001.0001>

Himma, M. (2021). Marju Himma: faktikontroll ajakirjandusžanrina on ohtlik. *ERR*. 6. detsember. Kasutatud 05.01.2025, <https://www.err.ee/1608426839/marju-himma-faktikontroll-ajakirjanduszanrina-on-ohtlik>

Hua, Y., Niu, S., Cai, J., Chilton, B, L., Heuer, H., Wohn, Y, D. (2024). Generative AI in User-Generated Content. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3613905.3636315>

Hughes, S. (2024). „Hughes hallucination evaluation model (hhem) leaderboard,” Kasutatud 30.12.2024, <https://huggingface.co/spaces/vectara/Hallucination-evaluation-leaderboard>

Ioannou, A. ja Tussyadiah, I. (2021). Privacy and surveillance attitudes during health crises: Acceptance of surveillance and privacy protection behaviours. *Technology in Society*, 67, 101774. <https://doi.org/10.1016/j.techsoc.2021.101774>

Kamila, M. K. ja Jasrotia, S. S. (2023). Ethical issues in the development of artificial intelligence: recognizing the risks. *International Journal of Ethics and Systems*. <https://doi.org/10.1108/ijoes-05-2023-0107>

Kaplan, A. ja Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of Artificial Intelligence. *Business Horizons*, 62(1), 15–25. <https://doi.org/10.1016/j.bushor.2018.08.004>

Knostic Team. (2025, 3. veebruar). *DeepSeek's cutoff date is July 2024: We extracted DeepSeek's system prompt.* Knostic. Kasutatud 20.05.2025, <https://www.knostic.ai/blog/exposing-deepseek-system-system-prompts>

Koit, M. (2015). Tehisintellekt – mis oli, mis on uus ja mis tuleb? *Sirp*. 6. veebruar. Kasutatud 05.01.2025, <https://www.sirp.ee/s1-artiklid/c21-teadus/tehisintellekt-mis-oli-mis-on-uus-ja-mis-tuleb/>

Langemets, M. (Toim.) (2009). Eesti keele seletav sõnaraamat. Tallinn: Eesti keele sihtasutus. Kasutatud 30.12.2025.

Langer, M., Hunsicker, T., Feldkamp, T., König, C. J. ja Grgić-Hlača, N. (2022). „Look! It's a computer program! It's an algorithm! it's AI!": Does terminology affect human perceptions and evaluations of algorithmic Decision-Making systems? *CHI Conference on Human Factors in Computing Systems*, 1–28. <https://doi.org/10.1145/3491102.3517527>

Laur, S. (2025, 12. aprill). *Sandra Laur väidab, et Tartu pole sõjaks valmis* [Facebook Reel]. Kasutatud 10.04.2025, <https://www.facebook.com/reel/2363739964007168>

Laur, S. (i.a.). *[Facebooki profiil]*. Facebook. Kasutatud 10.04.2025, <https://www.facebook.com/profile.php?id=61566873944997>

Lima, A. J., Senados, J., Senturias, M., ja Simagala, M. (2025). Shaping the Future of AI in Education: Insights From Pre-Service Science Teachers' Knowledge, Attitudes, and Perceptions. *14th International Conference on Educational and Information Technology (ICEIT)* (lk 380–385). IEEE. <http://dx.doi.org/10.1109/ICEIT64364.2025.10975974>

Maljukov, M. (2023). Tallinna Ülikoolis loodi õppematerjal tehisintellekti õppetöös kasutamiseks. *Tallinna Ülikool*. 20. september. Kasutatud 05.01.2025, <https://www.tlu.ee/uudised/tallinna-ulikoolis-loodi-oppematerjal-tehisintellekti-oppetoos-kasutamiseks>

Mallick, S. B., ja Kilpatrick, L. (2025, 5. veebruar). *The Gemini 2 family expands*. Google Developers Blog. Kasutatud 20.04.2025, <https://developers.googleblog.com/en/gemini-2-family-expands/>

Manikonda, L., Deotale, A., ja Kambhampati, S. (2018). What's up with Privacy? User Preferences and Privacy Concerns in Intelligent Personal Assistants. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, New York, NY, USA, 229–235. <https://doi.org/10.1145/3278721.3278773>

McCarthy, J., Minsky, M. L., Rochester, N., ja Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

Meie põhimõtted. (2018). Eesti Päevaleht. <https://epl.delfi.ee/artikkel/91897223/meie-pohimotted>

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X. ja Gao, J. (2024). Large Language Models: a survey. *arXiv*. <https://arxiv.org/abs/2402.06196>

Minsky, M. L. (1968). *Semantic information processing*. Cambridge, MA: MIT Press.

Mis on tehisintellekt ja kuidas seda kasutatakse? (2023). *Euroopa Parlament*. Kasutatud 30.12.2025, <https://www.europarl.europa.eu/topics/et/article/20200827STO85804/mis-on-tehisintellekt-ja-kuidas-seda-kasutatakse>

Mohammed, A. A. Q., Mudhsh, B. A., Bin-Hady, W. R. A. ja Al-Tamimi, A. S. (2025). *DeepSeek and Grok in the spotlight after ChatGPT in English education: A review study*. *Journal of English Studies in Arabia Felix*, 4(1), 13–22. <https://www.europarl.europa.eu/topics/et/article/20200827STO85804/mis-on-tehisintellekt-ja-kuidas-seda-kasutatakse>

Muller, M., ja Weisz, J. (2022). *Extending a Human-AI Collaboration Framework with Dynamism and Sociality*. *Proceedings of the 2022 Symposium on Human-Computer*

Interaction for Work (CHIWORK '22). (Artikkel 10). ACM.
<https://doi.org/10.1145/3533406.3533407>

Nieminen, S. ja Sankari, V. (2021). Checking PolitiFact's Fact-Checks. *Journalism Studies*, 22(3), 358–378. <https://doi.org/10.1080/1461670x.2021.1873818>

OpenAI. (2024, 13. mai). *Hello GPT-4o*. OpenAI. Kasutatud 05.01.2025, <https://openai.com/index/hello-gpt-4o/>

OpenAI. (2025). *ChatGPT (versioon 4o)*. Kasutatud 05.01.2025, <https://openai.com/chatgpt>

OpenAI. (i.a.). *GPT-4o*. OpenAI Platform. Kasutatud 15.03.2025, <https://platform.openai.com/docs/models/gpt-4o>

Põlendik, K. (2024). Kelmid varastasid 700 000 Apotheka kliendi andmed. *ERR*. 4. aprill. Kasutatud 05.01.2025, <https://www.err.ee/1609302060/kelmid-varastasid-700-000-apotheka-kliendi-andmed>

PromptHub. (i.a.). *DeepSeek V3*. PromptHub. Kasutatud 20.04.2025, <https://www.prompthub.us/models/deepseek-v3>

Ramachandran, Anand. (2024). Unveiling Google's Gemini 2.0: A Comprehensive Study of its Multimodal AI Design, Advanced Architecture, and Real-World Applications. https://www.researchgate.net/publication/387089907_Unveiling_Google's_Gemini_2_0_A_Comprehensive_Study_of_its_Multimodal_AI_Design_Advanced_Architecture_and_Real-World_Applications

Salecha, M. (2016). Story of ELIZA, the first chatbot developed in 1966. *Analytics India Magazine*. 5. oktoober. Kasutatud 05.01.2025, <https://analyticsindiamag.com/ai-features/story-eliza-first-chatbot-developed-1966/>

Silverman, C. (2007). *Regret the error: How media mistakes pollute the press and imperil free speech*. New York, NY: Union Square Press.

Sivek, S. C. ja Bloyd-Peshkin, S. (2019). Where do facts matter? The digital paradox in magazines' fact-checking processes. *Journalism Practice*, 13(8), 998–1002. <https://doi.org/10.1080/17512786.2019.1643767>

Souza, M. E. d. C., ja Weigang, L. (2025). *Grok, Gemini, ChatGPT and DeepSeek: Comparison and Applications in Conversational Artificial Intelligence*. Zenodo. <https://doi.org/10.5281/zenodo.14885243>

Stencel, M. (2024). With half the planet going to the polls in 2024, fact-checking sputters - Duke Reporters' Lab. *Duke Reporters' Lab*. 30. mai. Kasutatud 30.12.2024, <https://reporterslab.org/with-half-the-planet-going-to-the-polls-in-2024-fact-checking-sputters/>

Tartu Ülikooli suunis tehisintellekti kasutamiseks õppetöös. (i.a.). *Tartu Ülikool*. Kasutatud 30.12.2024, <https://ut.ee/et/sisu/tartu-ulikooli-suunis-tehisintellekti-kasutamiseks-oppetoos>

Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., Tanzer, G., Vincent, D., Pan, Z., Wang, S., Mariooryad, S., Ding, Y., Geng, X., Alcober, F., Frostig, R., Omernick, M., Walker, L., Paduraru, C., Sorokin, C., Zoph, B., ja Vinyals, O. (2024, 6. märts). *Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context*. (arXiv preprint arXiv:2403.05530). <https://arxiv.org/abs/2403.05530>

Thrun, S., Montemerlo, M., Hendrik Dahlkamp ja Mahoney, P. (2006). Stanley: The robot that won the DARPA Grand Challenge. *ResearchGate*, 23, 661–692. https://www.researchgate.net/publication/220648006_Stanley_The_robot_that_won_the_DARPA_Grand_Challenge

TI-Hüpe. (2025). *Kellele ja millal?* Kasutatud 20.04.2025, <https://tihuue.ee/millal>

Tolmeijer, S., Christen, M., Kandul, S., Kneer, M. ja Bernstein, A. (2022). Capable but amoral? Comparing AI and human expert collaboration in ethical decision making. *CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3491102.3517732>

Toosi, A., Bottino, A. G., Saboury, B., Siegel, E. ja Rahmim, A. (2021). A brief history of AI: How to Prevent Another Winter (A critical Review). *PET Clinics*, 16(4), 449–469. <https://doi.org/10.1016/j.cpet.2021.07.001>

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>

Turk, K. ja Kose, K. (2023). Monitoring media pluralism in the digital era : application of the media pluralism monitor in the European Union, Albania, Montenegro, the Republic of North Macedonia, Serbia and Turkey in the year 2022. Country report : Estonia. *Cadmus EUI*. <https://doi.org/10.2870/74195>

Tussyadiah, I. ja Miller, G. (2019). Nudged by a robot: Responses to agency and feedback. *Annals of Tourism Research*, 78, 102752. <https://doi.org/10.1016/j.annals.2019.102752>

Tutek, M., Hashemi Chaleshtori, F., Marasovic, A., ja Belinkov, Y. (2025, 20. veebruar). *Measuring Faithfulness of Chains of Thought by Unlearning Reasoning Steps*. arXiv. <https://doi.org/10.48550/arXiv.2502.14829>

Vake, D., Šinik, B., Vičić, J., ja Tošić, A. (2025, 5. märts). *Is Open Source the Future of AI? A Data-Driven Approach*. *Applied Sciences*, 15(5), 2790. <https://doi.org/10.3390/app15052790>

Varol, O., Ferrara, E., Davis, C., Menczer, F. ja Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 280–289. <https://doi.org/10.1609/icwsm.v11i1.14871>

Varol, O., Ferrara, E., Davis, C., Menczer, F. ja Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 280–289. <https://doi.org/10.1609/icwsm.v11i1.14871>

Vunš, M. (2024). Faktikontroll: Eestis pole kaugeltki Euroopa kõrgeim elektrihind. *Eesti Päevaleht*, 30. oktoober. Kasutatud 05.01.2025,

<https://epi.delfi.ee/artikkel/120332923/faktikontroll-eestis-pole-kaugeltki-euroopa-korgeim-el-ektrihind>

Vunš, M. (2025b). Faktikontroll: Saksamaa suurim energiaallikas on tuul. Tuuleenergia osakaalu näitamine kogu primaarenergiavajaduse kaudu on teadlikult eksitav. *Eesti Päevaleht*. 23. jaanuar. Kasutatud 20.04.2025, <https://epi.delfi.ee/artikkel/120351870/faktikontroll-saksamaa-suurim-energiaallikas-on-tuul-tuuleenergia-osakaalu-naitamine-kogu-primaarenergiavajaduse-kaudu-on-teadlikult-eksitav>

Wangsa, K., Karim, S., Gide, E. ja Elkhodr, M. (2024). *A systematic review and comprehensive analysis of pioneering AI chatbot models from education to healthcare: ChatGPT, Bard, Llama, Ernie and Grok*. *Future Internet*, 16(219). <https://doi.org/10.3390/fi16070219>

Wong, O. (2024). Gender-based disinformation: A tool of hybrid warfare. *Policy Briefs*, 9(1). Kasutatud 23.05.2025, <https://www.queensu.ca/cidp/publications/policy-briefs/gender-based-disinformation-tool-hybrid-warfare>

xAI. (2025, 17. veebruar). *Grok 3 Beta — The Age of Reasoning Agents*. Kasutatud 20.05.2025, <https://x.ai/news/grok-3>

xAI. (2025). *Grok (versioon 3)*. Kasutatud 22.04.2025, <https://grok.x.ai>

xAI. (i.a). *Models and Pricing*. xAI Docs. Kasutatud, 20.05.2025, <https://docs.x.ai/docs/models#models-and-pricing>

Xu, R., & Li, G. (2024). *A Comparative Study of Offline Models and Online LLMs in Fake News Detection*. arXiv. <https://doi.org/10.48550/arXiv.2409.03067>

Zhang, B., Dafoe, A. (2019). *Artificial intelligence: American Attitudes and Trends*. Center for the Governance of AI, Future of Humanity Institute, University of Oxford, Oxford. Kasutatud 30.12.2025, <https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/>

LISAD

Lisa 1. DeepSeek V3 faktikontroll

Faktikontroll väitele: "Tartu linnas on vaid üks varjend, kuhu raketirünnaku korral varjuda"

Väite analüüs:

Sandra Lauri väide, et Tartus on vaid üks varjend raketirünnaku korral, on **vale**. Tartu linnas on tuvastatud mitu avalikku varjendit ja kaitsealust ruumi, mis vastavad riiklikele ohutusstandarditele. Lisaks on paljudel hoonetel (koolid, avalikud institutsioonid, elumajad) tugevdatud keldrid või muud ruumid, mida saab kriisiolukorras kasutada.

1. Riiklikud ja kohalikud varjendipoliitika

Eesti Vabariigi **Päästeamet** (Rescue Board) on selgitanud, et Eestis puudub seni ühtne avalik varjendite register turvakaalutlustel, kuid kriisiolukorras suunatakse inimesi lähimastesse sobivatesse kohtadesse. Varjenditeks võivad olla nii spetsiaalselt ehitatud bunkrid kui ka tugevdatud hoonete osad (nt keldrid).

- Allikas: Päästeameti kaitsealuste ruumide juhend (2023).

Tartu Linnavalitsus on viimastel aastatel **kaitsealuseid ruume laiendanud**, sealhulgas koostöös koolide ja avalike hoonete omanikega. Näiteks on Tartu Tähtvere Põhikooli renoveerimisel rajatud ka kaitsealune ruum.

- Allikas: Tartu Linnavalitsuse pressiteade (2022).
-

2. Avalikud varjendid Tartus

Päästeameti andmetel on Tartus **mitu avalikku varjendit**, mille asukohad avalikustatakse kriisiolukorras. Näiteks:

- **Tartu Kaubamaja keldrikorrus** (kohandatud kaitseks),
- **Tartu Ülikooli hoonete all olevad ruumid**,
- **Linnaraamatukogu keldrid**,
- **Spordihoonete alused** (nt Tartu Tamme Staadioni ruumid).

Lisaks on **korterelamute tugevdatud keldrid** märgistatud kui potentsiaalsed varjendid.

- Allikas: Päästeameti kriisijuhtimise osakond (2023).
-

3. Ekspertide kommentaarid

Kaitseekspert **Mihkel Kärmas** (Kaitseministeeriumi nõunik) on rõhutanud, et Eestis on varjendivõrgustiku arendamine prioriteet, eriti pärast 2022. aasta Ukrainasõda. Tartu on üks linnu, kus varjendite arv on viimase kahe aastaga märkimisväärselt kasvanud.

- Allikas: ERR uudis "Eesti suured linnad laiendavad varjendivõrgustikke" (2023).
-

4. Statistika ja turvameetmed

Riigikogu 2022. aasta **kaitsealuste ruumide seaduse muudatus** kohustab uusi avalikke hooned projekteerima kaitsealuseid ruume. Tartus on selle alusel alates 2023. aastast ehitatud mitu uut hoonet (nt Tartu Kunstimuseumi laiendus), mis sisaldavad avalikke varjendeid.

- Allikas: Riigi Teataja: Kaitsealuste ruumide seadus (2022).
-

Kokkuvõte

Väide, et Tartus on vaid üks varjend, on **faktiliselt vale**. Tartu linnas on tuvastatud mitu avalikku ja eraomandis olevat varjendit, mis vastavad Päästeameti nõuetele.

Varjendivõrgustikku laiendatakse pidevalt, eriti sõjategevuse tõttu Ukrainas. Siiski võib avalike varjendite täpne arv olla turvakaalutlustel piiratud avalikkusele.

Soovitus: Kriisiolukorras järgige Päästeameti ja kohalike võimude juhiseid, mis suunavad teid lähimasse varjendisse.

Lisa 2: Gemini 2.0 Flash faktikontroll

Muidugi, ma uurin seda Sandra Lauri väidet põhjalikult ja esitan oma faktikontrolli tulemused koos usaldusväärsete allikatega.

Faktikontroll: Sandra Lauri väide, et Tartu linnas on vaid üks varjend kuhu raketirünnaku korral varjuda, **ei ole täpne.**

Põhjendus ja allikad:

- Päästeameti seisukoht:** Päästeamet on korduvalt rõhutanud, et Eestis ei ole spetsiaalselt raketirünnakute eest ehitatud varjendeid elanikkonnale laialdaselt ette nähtud. Olemasolevad varjumiskohad on peamiselt kohandatud keldrid ja maa-alused ruumid, mis pakuvad kaitset otseste plahvatuste ja suuremate purustuste eest.
 - Viide:** Päästeameti ametlikud pressiteated ja avalikud esinemised kriisiolukorraks valmistumise teemal. (Kahjuks ei ole mul reaalses juurdepääsu konkreetsetele pressiteadetele, kuid Päästeameti veebileht ja meediakanalid on peamised infoallikad sel teemal.)
- Kohalike omavalitsuste tegevus:** Tartu linnavalitsus, sarnaselt teiste Eesti omavalitsustega, on tegelenud olemasolevate sobivate ruumide kaardistamisega, mida saaks kriisiolukorras varjumiseks kasutada. Need ei ole spetsiaalselt ehitatud varjendid, vaid olemasolevad keldrid ja maa-alused korrused, mis on hinnatud sobivaks ajutiseks kaitseks.
 - Viide:** Tartu linnavalitsuse avalikud dokumendid, kriisikomisjoni protokollid ja teated elanikkonnale kriisiolukorraks valmistumise kohta. (Sarnaselt Päästeametiga, konkreetsete dokumentidele reaalses juurdepääs puudub, kuid Tartu linna veebileht on oluline infoallikas.)

3. **Meedia kajastus:** Eesti meedia on korduvalt kajastanud varjumiskohtade teemat, tuues välja, et üle Eesti on kaardistatud mitmeid potentsiaalseid varjumiskohti, sealhulgas Tartus. Need artiklid on tuginenud Päästeameti ja kohalike omavalitsuste infole.
- **Viide:** Erinevad uudisportaalid (nt ERR Uudised, Postimees, Delfi) on avaldanud artikleid varjumiskohtade kohta. (Konkreetsete artiklite otsimine ja viitamine reaalajas ei ole võimalik, kuid need on üldteadaolevad infoallikad.)

Järeldus:

Kuigi Tartus ei pruugi olla ühtainsat spetsiaalselt ehitatud raketivarjendit, on linnas **mitmeid olemasolevaid keldreid ja maa-aluseid ruume**, mis on kaardistatud potentsiaalsete varjumiskohtadena kriisiolukorras. Sandra Lauri väide, et on vaid üks selline koht, ei vasta olemasolevale informatsioonile Päästeameti ja Tartu linnavalitsuse poolt. Oluline on mõista, et Eestis ei ole laialdaselt rajatud spetsiaalseid varjendeid, vaid kasutatakse olemasolevat taristut parimal võimalikul viisil elanikkonna kaitseks.

Seega, Sandra Lauri väide on **faktiliselt ebatäpne**.

Lisa 3. Grok 3 faktikontroll

Otsene vastus

- Sandra Lauri väide, et Tartus on vaid üks varjend raketirünnaku korral, on tõenäoliselt vale, kuna linnal on mitu avalikku varjumiskohta.
- Uuringud viitavad, et Tartus on vähemalt 25 avalikku varjumiskohta, sealhulgas koolides, linna hoonetes, eraettevõtete hoonetes ja ülikooli hoonetes.
- Täpne arv võib varieeruda, kuid tõendid toetavad mitme varjumiskoha olemasolu, mitte üheainsa.

Taust

Tartu linnavalitsuse andmetel on avalikke varjumiskohti paigutatud erinevatesse linnaosadesse, et tagada elanike turvalisus sõjaliste ohtude, nagu raketirünnakud, korral. Need kohad on tähistatud ja kättesaadavad vajaduse korral.

Allikad ja usaldusvärsus

Informatsioon põhineb Tartu linna ametlikul veebilehel, mis on usaldusväärne allikas kohalike teenuste kohta. Lisaks toetavad seda ka varasemad uudised, mis näitavad, et varjumiskohtade arv on aja jooksul suurenenud.

Ülevaade ja detailne analüüs

See ülevaade annab põhjaliku ülevaate Sandra Lauri väite faktikontrollist, et Tartus on vaid üks varjend raketirünnaku korral. Analüüs tugineb usaldusväärsetele ja autoriteetsetele allikatele, nagu Tartu linna ametlik veebileht, uudisteportaalid ja Eesti Päästeameti informatsioon.

Meetodoloogia ja allikate valik

Faktikontrolli läbiviimiseks kasutati mitmeid allikaid, sealhulgas Tartu linna ametlikku veebilehte, mis pakub kõige ajakohasemat infot kohalike varjumiskohtade kohta. Lisaks analüüsiti vanemaid uudiseid, nagu Tartu Postimehe artiklid, et mõista varjumiskohtade ajalugu ja arengut. Kuna teema on seotud tsiviilkaitsega, kontrolliti ka Eesti Päästeameti veebilehte, mis annab üldist infot varjumiskohtade kohta Eestis.

Varjumiskohtade olemasolu Tartus

Tartu linna ametliku veebilehe kohaselt ([invalid url, do not cite]) on linnas mitu avalikku varjumiskohta, mis on mõeldud erinevate ohtude, sealhulgas sõjaliste rünnakute, korral. Need kohad on jaotatud erinevatesse kategooriatesse, nagu näitab allolev tabel, mis põhineb veebilehe andmetel:

Kategooria	Arv	Lisainfo
Koolid	12	Varjumiskohad asuvad koolide hoonetes, tagades õpilaste turvalisuse.
Linna hooned	6	Linna omandis olevad hooned, nagu vallamajad, on varjumiskohtadeks.
Erahooned	3	Eraettevõtete hooned, mis on määratud avalikeks varjumiskohtadeks.
Ülikooli hooned	4	Tartu Ülikooli hooned, mis pakuvad varjumisvõimalusi.

Kokku annab see vähemalt 25 avalikku varjumiskohta, kuigi täpne arv ei ole veebilehel otse välja toodud. Kaart, mis on kättesaadav aadressil ([\[\[invalid url, do not cite\]\]](#)), näitab nende asukohti, mis kinnitab mitme varjumiskoha olemasolu.

Ajalooline kontekst

2003. aasta artikkel Tartu Postimeesest ([\[\[invalid url, do not cite\]\]](#)) mainib, et tollal oli Tartus vaid kaks avalikku pommivarjendit (Pargi 8 ja Hermanni 1a), kuna kolmas (Tamme puiestee 7a) oli lammutatud. See artikkel on aga vana ja ei kajasta tänapäeva olukorda. Uuemad andmed näitavad, et varjumiskohtade arv on suurenenud, eriti pärast Ukraina sõja algust, mil Eestis on tähelepanu pööratud tsiviilkaitse tugevdamisele. Näiteks 2022. aasta uudis BNN-s ([\[\[invalid url, do not cite\]\]](#)) mainib, et Tallinnas märgiti 23 võimalikku varjumiskohta, ja plaaniti laiendada seda teistesse suurtesse linnadesse, sealhulgas Tartusse, järgmisel aastal. See viitab, et Tartus on tõenäoliselt veelgi rohkem varjumiskohti täna.

Tsiviilkaitse ja lisainformatsioon

Eesti Päästeameti veebileht ([\[\[invalid url, do not cite\]\]](#)) annab üldist infot varjumiskohtade kohta, sealhulgas soovitusi, kuidas ohu korral käituda. Kuigi täpset arvu Tartu kohta seal ei leidu, kinnitab see, et varjumiskohad on osa riiklikust tsiviilkaitse süsteemist. Lisaks on olevalmis.ee veebileht ([\[\[invalid url, do not cite\]\]](#)) ressurss, mis pakub juhiseid varjumiseks, kuid seal ei ole konkreetset infot Tartu kohta.

Sandra Lauri ja väite kontekst

Sandra Lauri kohta ei leitud konkreetset infot, kes ta on või millal ta väite esitas. See võib viidata, et ta on kohalik tegelane või et nimi on valesti märgitud. Kuna väide on konkreetne, analüüsiti seda Tartu linna ametlike andmete põhjal, mis näitavad, et väide ei vasta tõele.

Järeldus ja piirangud

Tõendid toetavad, et Tartus on mitu avalikku varjumiskohta, mitte vaid üks, nagu Sandra Lauri väitis. Täpne arv võib olla suurem, kuna andmed põhinevad veebilehel toodud kategooriatel, ja kaart võib sisaldada täiendavaid kohti. Kuna teema on seotud riikliku julgeolekuga, võib osa infot olla piiratud, kuid avalikult kättesaadav informatsioon näitab selgelt mitme varjumiskoha olemasolu.

Võtmeallikad

- Tartu linna valitsuse veebileht varjumiskohtade kohta: <https://tartu.ee/et/varjumine>
- Tartu Postimehe artikkel pommivarjenditest (2003):
<https://tartu.postimees.ee/2009597/tartu-pommivarjendid-on-varjusurmas>
- BNN uudis Tallinna varjumiskohtade märgistamisest (2022):
<https://bnn-news.com/tallinn-installs-street-signs-informing-residents-of-bomb-shelters-235509>
- Eesti Päästeameti veebileht tsiviilkaitse kohta: <https://www.rescue.ee/en>
- Olevalmis.ee veebileht varjumise juhiste kohta: <https://www.olevalmis.ee/varjumine>

Lisa 4. Claude 3.7 Sonnet faktikontroll

Faktikontroll väitele Tartu varjendite kohta

Väide

Sandra Lauri väitis, et "Tartu linnas on vaid üks varjend kuhu raketirünnaku korral varjuda".

Hinnang

Väär. Tartu linnas on rohkem kui üks varjend või varjumiskoht, kuhu raketirünnaku korral varjuda.

Faktid ja analüüs

Tartu linnas on mitu erinevat varjumiskohta, mis on dokumenteeritud mitmes usaldusväärses allikas:

1. **Tartu linna koduleht** - Tartu linn on avalikustanud info mitme avaliku varjumiskoha kohta, sealhulgas:
 - Jaan Poska Gümnaasiumi varjend
 - Oskar Lutsu nimelise Tartu Linnaraamatukogu varjend
 - Tartu Ülikooli raamatukogu varjend
 - Lisaks mitmed avalikud varjumiskohad ja -alad
2. **Eesti Päästeamet** - on kaardistanud avalikke varjumiskohti üle Eesti, sealhulgas mitmed Tartu linnas asuvad varjumiskohad.
3. **Tartu Postimees** on kajastanud 2022-2023. aastal mitmes artiklis Tartu varjendite kaardistamist ja renoveerimist. Artiklite põhjal on selge, et Tartus on mitu erinevat varjumiskohta.
4. **Riigi Kinnisvara AS** on teinud inventuuri varjendite osas ning selle tulemusena on teada, et Tartus on mitmeid nõukogude ajast pärit varjendeid, mida on võimalik kasutada.

Lisaks tavapärasele varjenditele on Päästeameti juhiste kohaselt võimalik kasutada ka muid sobivaid kohti nagu maa-aluseid parklaid, keldrikorruseid, metroojaamu ja muid sügavamaid maa-aluseid rajatise ajutise varjumiskohana.

Järeldus

Sandra Lauri väide, et Tartu linnas on vaid üks varjend, on ekslik. Avalikult kättesaadava info põhjal on Tartu linnas mitu varjumiskohta ja varjendit, kuhu raketirünnaku korral varjuda.

Allikad

- Tartu linna ametlik koduleht
- Eesti Päästeameti avalik info varjumiskohtade kohta
- Tartu Postimehe artiklid varjendite teemal (2022-2023)
- Riigi Kinnisvara AS info riigile kuuluvate varjendite kohta

Märkus: Minu viimane teadmiste uuendamine oli 2024. aasta oktoobris, seega on võimalik, et pärast seda kuupäeva on toimunud muudatusi varjendite arvus või nende seisundis.

Lisa 5. ChatGPT 4o faktikontroll

Väide 1: *Tartu linnas on vaid üks koht, kuhu sõjalises olukorras varjuda.*

Faktikontroll: See väide on vale. Tegelikuses on Tartus **märgistatud mitukümmend avalikku varjumiskohta**, mitte ainult üks. Pärast Ukraina sõja puhkemist on Tartu linn koostöös Päästeametiga kaardistanud ja tähistanud terve rea hoonete keldreid ja maa-aluseid ruume, kuhu elanikud saavad ohuolukorras peitu minnatartu.postimees.eetartu.ee. Juba 2022. aastal teatati esimese avaliku varjumiskoha märgistamisest (Herbert Masingu kooli kelder) ning kokku valiti välja **18 varjumiskohta** üle linnatartu.postimees.ee. Hiljem lisandus neid veel – näiteks on **Tartu Kaubamaja ja Kvartali keskuse maa-alused parklad** nüüd avalikud varjumiskohadartu.postimees.ee. 2025. aasta kevadeks polnud Tartus küll ühtegi spetsiaalselt ehitatud pommivarjendit, kuid **avalikke varjumiskohti oli kokku juba 23**err.ee. Need paiknevad erinevates linnaosades (sh koolide keldrites, muuseumide ja spordihoonete all ning mõne eraettevõtte maa-alustes ruumides) ja on tähistatud tsiviilkaitse märgigatartu.eetartu.ee. Seega on vale väita, et Tartus leiduks ainult *üks* varjumispaik – tegelikult on neid paarikümne ringis, mis on valmis elanikele sõjalise ohu (või muu hädaolukorra) korral varju pakkumaerr.eetartu.postimees.ee.

Allikad: Tartu linnavalitsuse ja Päästeameti info kinnitab mitmete varjumiskohtade olemasolu. Näiteks ERR uudis märgib, et „*Tartus pole praegu ühtegi varjendit, varjumiskohti on linnas kokku aga 23.*”err.ee. Tartu linna kodulehel on avaldatud nimekiri avalikest varjumiskohtadest (sh koolide keldrid, raamatukogu, turuhoone, Kaubamaja parkla jne)tartu.ee. Ka Tartu Postimees on kajastanud, kuidas juba 2022. aastal alustati varjumiskohtade märgistamist üle linnatartu.postimees.ee ning 2024. aastal sõlmiti lepingud, et lisada varjumiskohtadena suured maa-alused parklad kesklinnas (Kaubamaja, Kvartal)tartu.postimees.ee. Need faktid kinnitavad üheselt, et **Tartus on rohkem kui üks varjumiskoht olemas.**

Väide 2: Tartu ~300-miljonilisest eelarvest on kriiside haldamiseks ette nähtud vaid 400 000 eurot, millest enamus kulub väiksemate kriiside (nt tormikahjustuste) kõrvaldamiseks.

Faktikontroll: See väide on peaaegu täpne. Tartu linna 2024. aasta kogueelarve maht on umbes **305 miljonit eurot**tartu.ee, millest kriisireguleerimisele ja elanikkonnakaitsele on kavandatud väga väike osakaal – ligikaudu **400 000 eurot aastas**, s.o umbes **0,13% linna eelarvestreform.ee. Selle numbri on kinnitanud ka Tartu linna esindajad. Näiteks Tartu volikogu liige Mihkel Lees on märkinud, et **käesoleval aastal panustab Tartu elanikkonnakaitsele ~400 000 €**, mis moodustab 0,13% linna ~300-miljonilisest eelarvestreform.ee. Seega on 400 000 € suurusjärguna õige.**

Täpsustada tuleb, **milleks** see raha kulub. Tartu linna eelarve seletuskiri näitab, et kriisireguleerimise kulud sisaldavad peamiselt elanikkonnakaitse korralduse ja valmisoleku tegevusi: näiteks kriisimeeskonna väljaõpe, teavitustegevus, kriisiinfokeskuste varustuse planeerimine, varugeneraatorite ja kütusevaru soetamine jmstartu.ee. Need on suhteliselt väikesed summad (nt 2024. a eelarves ca 76 000 € konkreetseid kriisivalmiduse tegevusitartu.ee), samuti finantseeritakse linna kriisikoordinaatorite ametikohti. Ülejäänud osa sellest ~400 000 €-st on tõepoolest mõeldud **otseste kriiside lahendamiseks ja kahjude likvideerimiseks**, mis praktikas tähendab just **igapäevasemate hädaolukordade** (näiteks tugeva tormi tekitatud kahjud, üleujutused, tulekahjude järeltegevused jmt) katteks. Linn peab olema valmis koristama tormimurdu, taastama taristut ja tagama olulised teenused häirete järel – paraku tuleb seda teha väga napi eelarvega. Nagu Lees oma arvamuses osutas, on 400 000 eurot sellises kontekstis „*olematu summa*”tartu.postimees.ee, mistõttu on arutletud, et edaspidi tuleks kriisivalmidusse investeerida kordades rohkem (eesmärgiks on välja käidud isegi 0,5% linna eelarvest ehk ~1,5 miljonit € aastas)reform.ee.

Allikad: Tartu linna 2024. aasta eelarvedokument kinnitab eelarve kogumahtu (~305 mln €) ja viitab kriisireguleerimise kulude mahuletartu.eetartu.ee. Samuti on teemat kajastanud ERR ja Tartu linna esindajad: siseministeeriumi haldusalas rõhutati hiljuti, et kohalike omavalitsuste panus elanikkonnakaitsele on seni olnud väga väikereform.ee. Mihkel Lees on

toonud konkreetselt välja 400 000 € suuruse summa ning selle tühisuse 300 mln € eelarve kõrvalreform.eetartu.postimees.ee. Kuigi detailset jaotusprotsenti (kui suur osa sellest rahast läheb just tormikahjude likvideerimisele) avalikest allikatest ei leidu, saab olemasoleva info põhjal järeldada, et **valdav enamus Tartu kriisieelarvest kulubki hädaolukordade operatiivseks lahendamiseks** (nagu loodusõnnetuste tagajärgede kõrvaldamine) ning väga vähe jääb üle uute varjumisvõimaluste loomiseks või muude suuremahuliste elanikkonnakaitse meetmete tarbekstartu.eereform.ee.

Kokkuvõte: Sandra Lauri väljaöeldud numbrid ja üldine mure on faktiliselt mõistetavad – **Tartus on tõesti väga vähe ressursse elanikkonna varjumiseks sõja korral.** Kuid väide, et peidus oleks vaid üks koht, on liialdus: ametlike andmete järgi on Tartus üle 20 avaliku varjumiskohaerr.ee. Linna kriisivalmiduse rahastust (ca 400 000 € aastas) võib pidada marginaalseks, ja see raha kulub suures osas jooksvalt tekkivate kriiside lahendamisele, mistõttu strateegilisteks investeeringuteks (nt uuteks varjenditeks) seda peaaegu ei jätkugireform.eetartu.ee. Kõnealused faktid on kinnitatud usaldusväärsetes allikates (Tartu linn, ERR, eelarvedokumendid) ning viitavad vajadusele suuremateks investeeringuteks elanikkonnakaitse, nagu ka Laur ja teised on tähelepanu juhtinud.

Viidatud allikad:

- Tartu Postimees: „**Tartus märgistati esimene avalik varjumiskoht**” – 28.06.2022 (ülevaade Tartu varjumiskohtadest)tartu.postimees.ee
- ERR Uudised: „**Tartus rajab linn esimese varjendi (Siuru kultuurikeskuse keldrisse)**” – 28.04.2025 (märgib, et seni **0 varjendit, 23 varjumiskohta** Tartus)err.ee
- Tartu linna koduleht: „**Kriisideks valmisolek – Avalikud varjumiskohad**” (loetleb kõik Tartu avalikud varjumiskohad, sh koolide keldrid, parkimismajad jne)tartu.ee

- Tartu linna 2024 eelarve seletuskiri (lõik **kriisireguleerimise kuludest** ja nende kasutusest)tartu.ee
- ERR (Arvamus): **Mihkel Lees** – „Elanikkonnakaitsele peab rohkem raha leidma”
– märts 2025 (toob välja, et **Tartu panustab ~0,13% eelarvest ≈ 400 000 €** kriisivalmidusse)reform.ee.

Lisa 6. Lihtlitsents lõputöö reprodutseerimiseks

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Volli Soasepp ,
(*autori nimi*)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

Generatiivse tehisintellekti rakenduste esmane kasutuselevõtt Eesti kommunikatsioonijuhtide töös ,
(*lõputöö pealkiri*)

mille juhendaja(d) on Diana Poudel ,
(*juhendaja nimi*)

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;

2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi 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.

Volli Soasepp
23.05.2025