

**Kielten itsenäisen oppimisen vaikeudet – Voiko  
suuria kielimalleja hyödyntävistä keskusteluboteista  
olla apua luonnollisten kielten puhe- ja  
kommunikaatiotaitojen harjoittelussa?**

Tietojenkäsittelytiede  
Tietotekniikan laitos, Teknillinen tiedekunta  
Kandidaatintutkielma

Laatija:  
Santeri Karvola

Maaliskuu 2026

**Kandidaatintutkielma**  
**Tietotekniikan laitos, Teknillinen tiedekunta**  
**Turun yliopisto**

**Tutkinto-ohjelma:** Tietojenkäsittelytiede

**Tekijä:** Santeri Karvola

**Otsikko:** Kielten itsenäisen kielten oppimisen vaikeudet – Voiko suuria kielimalleja hyödyntävistä keskusteluboteista olla apua luonnollisten kielten puhe- ja kommunikaatiotaitojen harjoittelussa?

**Sivumäärä:** 35 sivua

**Päivämäärä:** Maaliskuu 2026

Kielten itsenäisen oppimisen mahdollisuudet ovat kasvaneet valtavasti vuosituhaten alusta. Tällä hetkellä oppijoilla on Internetin myötä käytössään hyvin suuri määrä tietoa sekä useita erilaisia sovelluksia ja käännöspalveluita. Ongelmana itsenäisessä kielten oppimisessa edelleen on kuitenkin ollut puhe- ja kommunikaatiotaitojen kehittäminen. Näiden taitojen harjoitteluun on vaadittu toinen ihminen, joka ihanteellisesti olisi opittavassa kielessä edistyneempi, jotta hän voisi korjata oppijan virheitä. Näin oppija voisi kehittyä kielitaidoissa lähikehityksen vyöhykkeellä ja oppiminen olisi tehokkaampaa. Valitettavasti tällaisen harjoittelukumppanin löytäminen voi olla haastavaa, jolloin tällainen harjoittelu voi jäädä vähäiseksi. Tämä ongelma voisi olla ratkaistavissa tekoälyn kehityksen myötä. Suuria kielimalleja hyödyntävät agentit, kuten ChatGPT, voisivat korvata esimerkiksi keskustelukumppanin tai opettajan roolin puhe- ja kommunikaatiotaitojen oppimisessa. Näin oppijat voisivat saada huomattavasti enemmän harjoitusta sekä toistoa näiden taitojen harjoittelussa.

Tämän tutkimuksen tarkoituksena oli selvittää, voisiko suuria kielimalleja hyödyntävistä agenteista olla hyötyä vieraiden kielten puhe- ja kommunikaatiotaitojen harjoittelussa ja miten niiden hyödyt vertautuvat perinteisiin oppimismenetelmiin. Tutkimus toteutettiin systemaattisena kirjallisuuskatsauksena monitieteisiin tietokantoihin. Katsaukseen valikoitui aineiston läpikäymisen jälkeen 11 vertaisarvioitua artikkelia.

Katsauksen tulokset olivat osittain ristiriitaisia. Kuudessa tutkimuksessa löydettiin keskimäärin suurella efektikoolla tilastollisesti merkitseviä tuloksia, kahdessa tutkimuksessa ei löydetty tilastollisesti merkitseviä tuloksia ja yhdessä tutkimuksessa löydettiin sekä tilastollisesti merkitseviä että tilastollisesti ei-merkitseviä tuloksia. Katsauksen tulos antaa kuitenkin viitteitä siitä, että kielimalliagenteista olisi hyötyä ainakin englannin kielen puhe- ja kommunikaatiotaitojen harjoittelussa. Katsauksen mukaan kielimalliagenttien kanssa harjoittelusta olisi vähintään yhtä paljon hyötyä puhe- ja kommunikaatiotaitojen oppimisessa kuin perinteisillä oppimismenetelmillä harjoittelusta. Tutkimusten koasetelmat olivat kuitenkin erilaisia, joten tarvittaisiin lisää tutkimusta samankaltaisilla tutkimusasetelmilla, jotta saataisiin selville tarkemmin, miten kielimalliagentteja kannattaa hyödyntää vieraiden kielten puhe- ja kommunikaatiotaitojen harjoittelussa.

**Asiasanat:** tekoäly, kielimalliagentit, puhetaidot, kommunikaatiotaidot, kielten oppiminen

# Sisällysluettelo

<b>1</b>	<b>Johdanto</b>	<b>5</b>
1.1	Aiempi tutkimus	6
1.2	Tutkimuskysymykset	7
1.3	Tutkielman rakenne	7
<b>2</b>	<b>Tausta</b>	<b>9</b>
2.1	<b>Tekoälyn mekanismit</b>	<b>9</b>
2.1.1	Tekoälyn määritelmä	9
2.1.2	Koneoppiminen	9
2.1.3	Neuroverkot ja syväoppiminen	9
2.1.4	Luonnollisen kielen prosessointi	10
2.1.5	Suuret kielimallit	11
2.1.6	Suuria kielimalleja hyödyntävät keskustelubotit	11
2.2	<b>Oppimisen mekanismit</b>	<b>12</b>
2.2.1	Kielten oppimisen mekanismit	12
2.2.2	Lähikehityksen vyöhyke	13
<b>3</b>	<b>Menetelmät</b>	<b>14</b>
<b>4</b>	<b>Tulokset</b>	<b>19</b>
4.1	<b>Tutkimusaineisto</b>	<b>19</b>
4.2	<b>Replika-sovellusta hyödyntäneet tutkimukset</b>	<b>26</b>
4.3	<b>Tutkimusasetelmat</b>	<b>26</b>
4.3.1	Muuttujat ja termit	26
4.3.2	Mittarit	27
4.3.3	Suuren kielimallin käyttö	28
4.3.4	Kielimalliagentin kanssa kommunikoinnin toteutus	29
4.3.5	Kielimalliagentin rooli	30
4.3.6	Kielimalliagentin vertailun kohde kontrolliryhmässä	30
4.4	<b>Tuloksien yhteenveto</b>	<b>31</b>
<b>5</b>	<b>Pohdinta</b>	<b>33</b>
5.1	<b>Tutkimusasetelmat</b>	<b>33</b>
5.2	<b>Muuttujat ja mittarit</b>	<b>34</b>
5.3	<b>Puheentunnistus</b>	<b>34</b>

<b>5.4</b>	<b>Vertailun kohde</b>	<b>34</b>
<b>5.5</b>	<b>Tutkitut kielimallit</b>	<b>36</b>
<b>5.6</b>	<b>Kielimalliagentin rooli</b>	<b>37</b>
<b>6</b>	<b>Yhteenveto</b>	<b>38</b>
	<b>Lähteet</b>	<b>40</b>

# 1 Johdanto

Teknologisen kehityksen vuoksi itsenäiseen kielten oppimiseen on tullut valtavasti lisää työkaluja ja mahdollisuuksia. Esimerkiksi Internetin avulla on aiempaa helpompaa löytää erilaisia materiaaleja liittyen kielten oppimiseen. Lisäksi varsinkin erilaiset käännössovellukset, kuten Google Kääntäjä ovat kehittyneet valtavasti viimeisten kahdenkymmenen vuoden aikana (Schäferhoff, 2024). Näiden lisäksi on tullut kielten oppimiseen tarkoitettuja sovelluksia, kuten Duolingo, Babbel ja WordDive.

Oppikirjojen, videoiden ja kielten oppimiseen tarkoitettujen sovellusten avulla on voinut harjoitella sanastoa, luetun ymmärtämistä, kuullun ymmärtämistä ja kielioppia. Näiden avulla on voinut harjoitella myös ääntämistä esimerkiksi oppikirjojen ääntämisohjeilla tai kuuntelemalla ja yrittämällä toistaa kuullun äänten, sanan tai virkkeen. Myös esimerkiksi Duolingo-sovelluksessa on ollut mahdollista harjoitella tekstin tuottamista, mutta yleisesti näissä on pitänyt kääntää valmis teksti toiselle kielelle. Ongelmana itsenäisessä kielten oppimisessa on kuitenkin edelleen puhe- ja kommunikaatiotaitojen harjoittelu, sillä ilman keskustelukumppania näiden taitojen harjoittelu on vaikeaa (Yahaya ym., 2021; Veneziano, 2010). Kieltä käytetään muiden ihmisten kanssa kommunikointiin, jolloin uuden tekstin ja puheen tuottaminen on keskeisessä roolissa. Keskustelutilanteissa yleisesti korostuu puhutun kielen tuottamisen taito sekä taito osata reagoida keskustelukumppanin vastauksiin. Keskustelutilanteissa puhetta tai tekstiä täytyy pystyä tuottamaan sujuvasti ja nopeasti, minkä harjoittelu on haastavaa, jos keskustelukumppania ei ole.

Tällaisten puhe- ja kommunikaatiotaitojen harjoitteluun on vaadittu toinen ihminen, jonka kanssa on voinut harjoitella tavallisia keskustelutilanteita. Toinen ihminen on voinut reagoida oppijan vastauksiin, arvioida oppijan tuottaman tekstin tai puheen sopivuutta tilanteeseen sekä korjata oppijan mahdollisia virheitä. Näitä on voinut harjoitella esimerkiksi kielten kursseilla ja luennoilla. Opettaja on voinut seurata oppijoiden keskustelua vierestä, korjata mahdollisia virheitä ja antaa palautetta. Lisäksi oppijat ovat voineet harjoitella keskustelutaitoja toistensa kanssa. Yleisesti ongelmana keskustelun harjoittelussa vertaisten kanssa on se, että kielten kursseilla oppijat ovat samantasoisia kielellisesti, jolloin varsinkin aloittelijoilla voi olla vaikea antaa palautetta vertaisten tuottamasta puheesta tai tekstistä. Näissä tilanteissa olisi hyvä, jos oppija voisi harjoitella puhe- ja kommunikaatiotaitoja jonkun kielellisesti edistyneemmän henkilön kanssa. Valitettavasti edistyneemmälle kielen käyttäjälle alkeellisemmän tason keskusteluista ei ole juurikaan hyötyä kielen harjoittelussa, koska heikomman kielitaidon

omaava oppija ei voi korjata hänen virheitään. Tämän vuoksi tällainen harjoittelu on yleensä mahdollista vain opettajan kanssa harjoitellessa tai opettajan seurattuna. Kuitenkin kursseilla voi olla kymmeniä oppijoita, joten harjoittelu opettajan kanssa voi jäädä vähälle. Tällaisessa tilanteessa paras olisikin, jos pääsisi harjoittelemaan keskustelua paremman kielitaidon omaavan tuutorin tai yksityisopettajan kanssa, mutta tämä voi olla esimerkiksi taloudellisesti haastavaa.

Vuoden 2022 marraskuussa itsenäiseen keskustelun harjoitteluun vaikutti tulevan vaihtoehto, kun OpenAI julkaisi ensimmäisen ChatGPT-versionsa (OpenAI, 2022). OpenAI:n GPT-3.5-kielimallia käyttävä ChatGPT pystyi mallintamaan toisen ihmisen kanssa käytävää keskustelua hyvin uskottavasti (Abdullah ym., 2022). ChatGPT:n jälkeen tuli myös monia muita suuria kielimalleja hyödyntäviä keskustelubotteja, kuten Googlen Gemini ja Anthropicin Claude. Nykyään keskustelubotit hallitsevat eri kieliä, ja esimerkiksi ChatGPT hallitsee keskustelun yli 50 kielellä (OpenAI, 2025). Suuria kielimalleja hyödyntäville keskusteluboteille voi esittää toiveita, että se näytelisi vieraan kielen tuutorin tai opettajan roolia ja korjaisi käyttäjän virheitä keskustelutilanteissa. Tämän vuoksi onkin mielenkiintoista tutkia, voisiko ChatGPT:n ja muiden suuria kielimalleja hyödyntävien keskustelubottien avulla harjoitella puhe- ja kommunikaatiotaitoja vieraalla kielellä. Tässä tutkielmassa käytän jatkossa suuria kielimalleja hyödyntävistä keskusteluboteista myös nimitystä kielimalliagentti.

## 1.1 Aiempi tutkimus

Tekoälyn vaikutuksista kielten oppimiseen on tehty paljonkin tutkimusta (Hana, 2023; Kundu & Bej, 2025; Lyu ym., 2025; Qin & Zhang, 2025; Şahin Kızıl ym., 2025; Tram ym., 2024; Wanli ym., 2025). Kuitenkaan aiempaa kokeellisia tutkimuksia kokoavaa tutkimusta tarkemmin kielimalliagenttien vaikutuksesta kielten oppimiseen puhe- ja kommunikaatiotaitojen kontekstissa ei ole tehty. Esimerkiksi Tram ja kumppanit (2024) tutkivat systemaattisen katsauksen, kyselytutkimuksen ja haastatteluiden avulla ChatGPT:n käyttöä ja hyötyjä englannin kielen opiskelussa. Tässä tutkimuksessa ei käsitelty tarkemmin puhe- ja kommunikaatiotaitoja, tai koottu yhteen kokeellisia tutkimuksia kielimalliagenttien käytöstä vieraan kielen tekstin tuottamisessa.

Kielimalliagentit ovat hyvin uusia, mikä voi vaikuttaa siihen, että tutkimusta tästä aiheesta ei ole vielä kovin paljoa. Tämänkin takia on mielenkiintoista selvittää, että voivatko kielimalliagentit toimia puhe- ja kommunikaatiotaitojen harjoittelukumppanina, ja ovatko suuret kielimallit tarpeeksi kehittyneitä, että kielimalliagenttien kanssa käytävä keskustelu

vieraalla kielellä voisi toimia korvikkeena toisen ihmisen kanssa käytävälle puhe- ja kommunikaatiotaitojen harjoittelulle.

## 1.2 Tutkimuskysymykset

Koska tekoäly on hyvin laaja käsite ja eri tekoälyagentit ovat hyvin eri tasoisia eri tehtävissä, ei ole mielekäästä verrata kaikkia tekoälyagentteja toisiinsa. Myös tarkemmin generatiivisissa tekoälyissä ja chatboteissa on paljon eroja. Esimerkiksi jo vuonna 1961 kehitetty ELIZA on generatiivista tekoälyä (Foote, 2024), mutta sitä ei ole järkevää ryhmittää kielimalliagenttien kanssa, kun tutkitaan agentin kykyä toimia kielten opettajana tai tuutorina. Ei siis ole järkevää ryhmittää kaikkia generatiivisiakaan tekoälyagentteja keskenään, joten tässä tutkimuksessa tarkastelu rajattiin suurilla kielimalleilla hyödyntäviin keskustelubotteihin. Tämän systemaattisen kirjallisuuskatsauksen tutkimuskysymyksinä onkin:

- 1) Voiko suurilla kielimalleilla hyödyntävistä keskusteluboteista olla hyötyä vieraiden kielten puhe- ja kommunikaatiotaitojen harjoittelussa?
- 2) Miten kielimalliagenttien hyödyt vertautuvat perinteisiin oppimismenetelmiin?

## 1.3 Tutkielman rakenne

Tutkielman luvussa 2 esitellään taustoja ja tärkeitä käsitteitä. Luvussa käydään läpi tutkimuksen osalta keskeiset tekoälyn osa-alueet, kuten koneoppiminen, neuroverkot, syväoppiminen, luonnollisen kielen prosessointi sekä suuret kielimallit. Luvun lopussa käsitellään vielä kielimalliagentteja sekä suurten kielimallien ja kielimalliagenttien eroja.

Luvussa 3 esitellään tarkemmin tämän tutkielman kirjallisuushaun menetelmät ja kuvataan tarkemmin kirjallisuushaun prosessia. Luvussa eritellään valitut hakulausekkeet, hakujen kohteena olleet tietokannat, sisäänotto- ja poissulkukriteerit sekä tuloksien läpikäyminen.

Luvussa 4 käsitellään systemaattisesta kirjallisuuskatsauksesta löytyneiden tutkimusten tulokset. Ensin esitellään yleisellä tasolla tutkimusaineisto sekä valittujen tutkimusten tutkimusasetelmat. Tämän jälkeen Taulukossa 1 esitellään tarkemmin katsaukseen sisällytetyt tutkimukset. Taulukon jälkeen käsitellään kokoavasti sisällytettyjen tutkimusten muuttujia ja termejä, mittareita, kielimalliagentin käyttöä, agentin kanssa kommunikointia, agentin roolia sekä vertailun kohdetta tutkimuksissa. Tämän jälkeen tehdään vielä yhteenveto tutkimuksen tuloksista ja mahdollisista vaikutuksista kielten oppimiseen puhe- ja kommunikaatiotaitojen

näkökulmasta sekä lasketaan tilastollisesti merkitsevien tulosten puuttuvat efektikoot, mikäli ne ovat laskettavissa.

Luvussa 5 pohditaan luvussa 4 esiteltyjä tuloksia eri näkökulmista katsaukseen valittujen tutkimusten tutkimusasetelmien perusteella. Tuloksia tulkitaan esimerkiksi kielimallin version ja kielimalliagentin roolin kautta sekä ryhmitettynä sen perusteella, mihin kontrolliryhmän kielten opetuksen osa-alueeseen kielimalliagentin kanssa käytävää harjoittelua verrattiin. Lisäksi luvussa esitetään jatkotutkimustarpeita. Lopulta luvussa 6 tehdään koko tutkielman kattava yhteenveto.

## 2 Tausta

### 2.1 Tekoölyn mekanismit

#### 2.1.1 Tekoölyn määritelmä

Tekoölyn määrittely on ollut koko tekoölyn historian ajan hyvin kiistelty (Caluori, 2024). Käytän tässä tutkielmassa tekoölyn määritelmää kirjasta *Artificial Intelligence: A Modern Approach* (Russell ym., 2010). Kirja määrittelee tekoölyn näin: "We define AI as the study of agents that receive percepts from the environment and perform actions." (s. viii). Vapaasti suomennettuna kirja määrittää tekoölyn alana, joka tutkii agenteja, jotka havainnoivat ympäristöään ja suorittavat toimintoja. Myös tässä tekoölyn määritelmässä käyttämänsä agentti-termin kirja määrittelee niin, että agentti on jokin asia, joka havainnoi ympäristöään ja toimii siellä. Lisäksi kirja määrittelee, että rationaalinen agentti toimii tavoitteellisesti siten, että se saavuttaisi parhaan mahdollisen lopputuloksen. Agentin havainnoiman ympäristön ei tarvitse olla fyysinen ympäristö, vaan esimerkiksi videopelissä toimiva tekoölyagentti havainnoi omaa ympäristöään, eli videopelin virtuaalista maailmaa.

#### 2.1.2 Koneoppiminen

França ja kumppanit (2021) määrittävät koneoppimisen tekoölyn osa-alueena, jossa tekoölyagenttia on mahdollista opettaa suorittamaan tietty tehtävä. Tätä varten on oltava jokin data-aineisto, josta koneoppiva agentti voi esimerkiksi tutkia korrelaatioita ja yhteyksiä eri datan välillä. Koneoppimisteknologioita on hyvin montaa eri muotoa (Russell ym., 2010). Näitä ovat esimerkiksi päätöspuut ja tukivektorikoneet. Tässä tutkielmassa tulemme kuitenkin keskittymään koneoppimisteknologioista vain neuroverkkoihin.

#### 2.1.3 Neuroverkot ja syväoppiminen

Neuroverkot ovat koneoppimisteknologioita, jotka koostuvat aivojen neuroverkkojen kaltaisista neuroneista (Russell ym., 2010). Kuten aivojen neuronit, keinotekoisien neuroverkkojen neuronit ovat yhteydessä toisiinsa. Ne toimivat siten, että ensin ne saavat jonkin syötteen, jonka jälkeen syöte syötetään neuronin aktivaatiofunktioon, joka määrittää neuronin tulosteen suuruuden. Syötteenä voi olla joko aiempien neuronien tulosteet tai ensimmäisten neuronien tapauksessa koko neuroverkon ensimmäinen syöte, esimerkiksi mustavalkoisen kuvasyötteen tietyn pikselin väriarvo. Neuroverkossa on usein rinnakkaisesti useampi neuroni.

Tässä esimerkissä voisi olla vaikkapa sadan pikselin kokoinen mustavalkoinen kuva. Tämä voitaisiin syöttää sellaiselle neuroverkolle, jossa on sata neuronua ensimmäisessä kerroksessa. Tällöin kullekin neuronille varattaisiin oma pikseli, ja kukin neuroni saisi syötteekseen tämän pikselin väriarvon. Tämän jälkeen ensimmäisen kerroksen neuronit muuntavat oman aktivaatiofunktionsa avulla saamansa syötteen tulosteeksi ja lähettävät sen eteenpäin seuraavalle kerrokselle. Seuraava kerros voi olla joko piilokerros tai tulostekerros. Tulostekerros on neuroverkon viimeinen kerros, jonka neuronien aktivaatiolla päätellään neuroverkon saama tulos. Jos tämän neuroverkon tarkoituksena olisi selvittää, esittääkö kuva neliötä, kolmiota vai ympyrää, olisi tulostekerroksessa kolme neuronua kutakin mahdollista tulosta kohden. Piilokerroksilla tarkoitetaan syötekerroksen ja tulostekerroksen välissä olevia neuronikerroksia (Zhang, 2018).

Mikäli neuroverkossa on useampi piilokerros, kutsutaan sitä syväoppivaksi neuroverkoksi (Mohanasundaram ym., 2019). Syväoppivissa neuroverkoissa ensimmäisen kerroksen neuronit lähettävät siis omat tulosteensa syötteenä ensimmäiselle piilokerrokselle, jossa jokainen neuroni taas muuntaa oman aktivaatiofunktionsa mukaisesti syötteen tulosteeksi ja lähettää sen eteenpäin seuraavalle piilokerrokselle. Syväoppivissa neuroverkoissa, kuten konvoluutioneuroverkoissa, piilokerrokset ovat yleensä erikoistuneita johonkin tiettyyn tehtävään, jolloin saadaan tehostettua neuroverkon toimintaa (Holdsworth & Scapicchio, 2024). Jos neuroverkko toimii hyvin, niin esimerkkinä tulostekerroksen aktivaatiosta voidaan päätellä mitä syötteenä syötetty sadan pikselin kuva esittää. Jos neuroverkko tunnistaa kuvan väärin, se voi muuttaa vastavirta-algoritmin avulla neuronien syötteenä saamien arvojen painokertoimia, jotta tulostoneuronien aktivaatio muuttuisi oikeampaan suuntaan. Näin syväoppivat neuroverkot oppivat (Holdsworth & Scapicchio, 2024).

#### 2.1.4 Luonnollisen kielen prosessointi

Luonnollisen kielen prosessointi tarkoittaa luonnollisen kielen analysointia tietokoneen avulla (Nyrhinen, 2019). Näitä ovat muun muassa tekstin muuttaminen puheeksi, puheen muuttaminen tekstiksi sekä kirjoitetun tekstin analysointi. Strykerin ja Holdsworthin (2024) mukaan luonnollisen kielen prosessoinnilla pyritään opettamaan tietokoneita ymmärtämään ja tuottamaan luonnollista kieltä. Nykyään suurin osa luonnollisen kielen prosessoinnista tapahtuu syväoppivien neuroverkkojen avulla. Muun muassa suuret kielimallit hyödyntävät luonnollisen kielen prosessointia.

### 2.1.5 Suuret kielimallit

Suuret kielimallit pohjautuvat valtavalla määrällä tekstiä koulutettuihin luonnollisen kielen prosessointiin erikoistuviin syväoppiviin neuroverkkoihin (Stryker, 2023). Esimerkiksi suurten kielimallien koulutukseen käytettävässä *Common Crawl* -data-aineistossa on lähes biljoona sanaa (Brown ym., 2020). Suuret kielimallit koulutetaan siten, että koulutusvaiheessa niiden tehtävä on ennustaa, mikä sana tai sanan osa, eli *token*, tulee seuraavaksi lauseeseen (Stryker, 2023). Jos koulutusdatassa on esimerkiksi lause: "Vietin vapaapäiväni rannalla makoillen.", kielimallille voidaan näyttää koulutusvaiheessa pätkä "Vietin vapaapäiväni rannalla", jonka jälkeen kielimallin pitäisi osata ennustaa, että seuraava sana on "makoillen". Kielimallin tulosten jälkeen vastavirta-algoritmi muuttaa kielimallin neuroverkon neuroneja samalla periaatteella kuin aiemmin esitelty yksinkertainen muodontunnistukseen tarkoitettu neuroverkko. Koska dataa on valtava määrä, oppii kielimalli tällä yksinkertaisella tavalla tuottamaan merkityksellistä tekstiä (Stryker, 2023).

Esikoulutus (*pre-training*) tarkoittaa ohjaamattoman suuren kielimallin oppimisen muotoa, jossa mallille ei kerrota, mitä sen pitäisi osata syötteestä tunnistaa (Belcic & Stryker, 2024). Näin malli oppii itse tunnistamaan merkkamattomasta datasta yhteyksiä ja samankaltaisuuksia. Tällöin malli pystyy esimerkiksi vastaamaan täysin uuteen merkkamattomaan käyttäjän syötteeseen merkityksellisellä tavalla.

Lähes kaikki suuret kielimallit hyödyntävät syväoppivan neuroverkon luonnollisen kielen prosessointiin erikoistunutta arkkitehtuuria Muunnin (*Transformer*), joka hyödyntää huomiomekanismia (*self-attention*) (Stryker, 2023). Koulutusvaiheessa muunnin muuntaa sanat vektoreiksi, jotka se asettaa hyvin moniulotteiseen vektoriavaruuteen (Bergmann & Stryker, 2025). Näiden vektoreiden välisestä etäisyydestä muunnin voi päätellä semanttisia yhteyksiä eri *tokenien*, eli sanojen tai sanan osien, välillä. Mikäli esimerkiksi sanoihin "puolukka" ja "pilvi" liittyvien vektorien välinen etäisyys on suuri, muunnin ei tulkitse sanojen olevan semanttisesti läheisiä, jolloin se ei kohdenna huomiotaan samassa syötteessä esiintyvään sanaan "pilvi", kun se ennustaa, mikä sana tulee seuraavana lauseeseen "Puolukka on...".

### 2.1.6 Suuria kielimalleja hyödyntävät keskustelubotit

Suuret kielimallit ja niitä hyödyntävät keskustelubotit ovat eri asia, vaikka ne voivatkin arkikielessä mennä sekaisin. Esimerkiksi termejä GPT ja ChatGPT voidaan käyttää monesti samassa yhteydessä, mutta ne tarkoittavat kuitenkin eri asiaa. ChatGPT on OpenAI:n luoma

sovellus, jonka kautta voi kommunikoida eri GPT-mallien kanssa (OpenAI, 2022). ChatGPT siis tarjoaa käyttäjille käyttöliittymän GPT-mallien hyödyntämiseen (Belcic & Stryker, 2025). GPT-mallit suuria kielimalleja, joita myös ChatGPT hyödyntää. GPT-malleja voi hyödyntää myös omissa sovelluksissa ChatGPT:n ulkopuolella hyödyntämällä OpenAI:n API:ta (OpenAI, ei pvm.). Suuret kielimallit eivät siis suoraan tarkoita samaa kuin niitä hyödyntävät keskustelubotit.

## 2.2 Oppimisen mekanismit

### 2.2.1 Kielten oppimisen mekanismit

Kielten, kuten muidenkin taitojen, oppimisessa toisto on erittäin tärkeää (Randall, 2007). Ensimmäisen kielen oppimisen kohdalla lapsi on jatkuvasti altistuneena kieltä käyttävään ympäristöön, jolloin toistoa tapahtuu jatkuvasti. On toki mahdollista oppia myös muita kieliä vastaavalla tyylillä, esimerkiksi muuttamalla opittavaa kieltä puhuvaan maahan. Monella ei kuitenkaan ole mahdollisuutta altistaa itseään uuden opittavan kielen ympäristöön eikä meillä yleensä ole aikaa tällaiselle intensiiviselle kielen altistukselle, joten joudumme turvautumaan toisiin tapoihin oppia muita kieliä (Randall, 2007).

Koska lapsena ensimmäisen kielen oppiminen ei ole kovin strukturoitua, vaan lapsi oppii kieltä epämääräisesti seuraamalla, voidaan toisten kielten oppimisessa hyödyntää järjestelmällisyyttä, jolloin kielten oppiminen tehostuu (Randall, 2007). Edelleen kielten oppiminen vaatii toistoa, mutta näin voidaan kohdistaa toisto tehokkaasti opittaviin asioihin, jolloin kielen oppimiseen tarvittavan ajan määrä vähenee huomattavasti. Lisäksi toisen kielen oppimisessa ihminen voi hyödyntää jo oppimaansa kieltä. Esimerkiksi suomenkielisen ihmisen ei tarvitse *bear*-sanan kuultuaan enää oppia yhdistämään, mitkä kaikki karhun ominaisuudet muodostavat karhun käsitteen, vaan hän voi yhdistää *bear*-sanan suomenkieliseen sanaan "karhu", josta hänellä on jo representaatio.

Näitä järjestelmällisiä tapoja oppia kieltä voi hyödyntää esimerkiksi itseopiskelussa oppikirjojen ja erilaisten sovellusten kautta tai esimerkiksi oppilaitosten vieraiden kielten kursseilla. Oppikirjoista tai kursseilta ei saavuteta samaa altistusta kuin lapsena ensimmäisen kielen oppimisessa, mutta nämä ovat strukturoituja ja fokusoituja tapoja oppia vierasta kieltä, joten näilläkin tavoilla on mahdollista oppia sujuvaksi vieraan kielen käyttäjäksi. Kielten kursseilla päästään harjoittelemaan kieltä opettajalähtöisesti, jolloin kielten oppiminen tapahtuu niin kutsutulla lähikehityksen vyöhykkeellä (Vygotski, 1978).

## 2.2.2 Lähikehityksen vyöhyke

Lev Vygotski esitteli teorian lähikehityksen vyöhykkeestä kirjassaan *Mind in society : the development of higher psychological processes* (Vygotski, 1978). Lähikehityksen vyöhykkeellä tarkoitetaan Vygotskin mukaan sitä taitojen kehityksen vaihetta, jota ihminen ei vielä itsenäisesti osaa, mutta johon ihminen on kykeneväinen jonkun taidon paremmin hallitsevan tuella. Tämän teorian mukaan ihminen pystyy siis parempiin suorituksiin silloin, kun häntä on tukemassa jokin taidon paremmin hallitseva ihminen. Näin myös oppiminen on tehokkaampaa.

Palataan vielä aiemmin esiteltyyn esimerkkiin lapsen ensimmäisen kielen oppimisesta. Lapsi kuulee jatkuvasti opittavaa kieltä ja saa harjoitella sen käyttöä lähikehityksen vyöhykkeellä useita vuosia, kun kieltä paremmin osaavat vanhemmat ihmiset toimivat lapsen tukena. Toisen kielen kohdalla lähikehityksen vyöhykettä on vaikeampi saavuttaa, sillä ihminen ei enää ole muiden, paremmin kieltä osaavien, ympäröivänä. Kuitenkin esimerkiksi sanastoa opetellessa tätä edistyneemmän ohjaajan roolia voi edustaa esimerkiksi kurssikirja tai jokin kielten opetteluun tarkoitettu sovellus. Puhe- ja kommunikaatiotaitojen harjoittelun kohdalla tämä on vaikeaa, sillä paremmin kieltä osaavan keskustelukumppanin löytäminen voi olla haastavaa tai kallista. Kielten kursseilla puhe- ja kommunikaatiotaitoja voi harjoitella esimerkiksi kurssitoverin tai opettajan kanssa, mutta tällöin toistomäärä voi jäädä vähäiseksi, jos keskustelua pääsee harjoittelemaan vain muutaman kerran viikossa. Tästä syystä tämän tutkielman on tarkoitus tutkia, voisivatko kielimalliagentit vastata tätä edistyneemmän ohjaajan roolia, ja auttaa ihmisiä kielten puhe- ja kommunikaatiotaitojen oppimisessa.

### 3 Menetelmät

Toteutin tämän tutkimuksen systemaattisena kirjallisuuskatsauksena. Tein kirjallisuushaun useana erillisenä hakuna kohdistuen käytetyimpiin suuriin kielimalleihin. Tein ensimmäisen OpenAI:n GPT-malleihin kohdistuvan haun 05.07.2025 seuraaviin tietokantoihin: APA PsycInfo (EBSCO), ACM Digital Library, IEEE Xplore, Web of Science, Scopus, SpringerLink, ProQuest Central, Wiley Online Library, Sage Journals sekä Taylor & Francis Online. Koska kandidaatintutkielman aihe koskee montaa eri tieteenalaa, yritin käyttää mahdollisimman montaa monitieteistä tietokantaa kuitenkin keskittyen psykologian ja tietojenkäsittelytieteen tärkeimpiin tietokantoihin. Näihin kaikkiin tietokantoihin käytin samaa hakulauseketta:

(ChatGPT OR GPT) AND ("learning languages" OR "language learning" OR "second language acquisition" OR "SLA") AND ("communication skill" OR "communication skills" OR "conversation skill" OR "conversation skills" OR "oral communication" OR "oral proficiency" OR "speaking proficiency") AND ("empirical study" OR "experimental study" OR "intervention" OR "control group" OR "comparison group") AND ("traditional methods" OR "traditional method" OR "human conversation" OR "peer" OR "peers" OR "native speaker" OR "native speakers" OR "another student" OR "another students" OR "other student" OR "other students")

Kohdistin hakulausekkeen kokoteksteihin kaikissa tietokannoissa, paitsi Web of Science -tietokannassa, sillä siellä ei pystynyt hakemaan tuloksia hakulausekkeella kokoteksteistä. Tein näin siksi, että tutkimuksessa käytettyä suurta kielimallia ei välttämättä ole mainittuna nimeltä julkaisun otsikossa tai abstraktissa. Tästä syystä haku pelkkään abstraktiin ja otsikkoon ei löytäisi kaikkia julkaisuja, joissa suurta kielimallia olisi kuitenkin käytetty.

Yhteensä valituista tietokannoista löydettiin ensimmäisellä hakulausekkeella 935 tulosta, joista ACM Digital Librarystä löytyi 150 tulosta, Scopusesta löytyi 140, SpringerLinkistä 59, ProQuest Centralista 509, Wiley Online Librarystä 30, Sage Journalsista 19 sekä Taylor & Francis Onlinesta 28. APA PsycInfosta ja IEEE Xploresta ei löytynyt yhtään tulosta. Web of Science -tietokannassa ei pystynyt hakemaan hakulausekkeella kokoteksteistä, joten Web of Sciencestakaan ei tullut tällä hakulausekkeella yhtään tulosta. Yritin hakea tuloksia tällä hakulausekkeella myös ScienceDirectistä, mutta hakulauseke oli liian monimutkainen tähän tietokantaan. Seuraavaksi kävin Rayyan-ohjelmalla (Ouzzani ym., 2016) hakutulosten

duplikaatit läpi. Rayyan merkitsi näistä tuloksista 125 mahdolliseksi duplikaateiksi. Kävin nämä läpi manuaalisesti ja lopulta näiden joukosta poistin metatietojen perusteella 70 duplikaatteina. Jäljelle jäi läpikäytäväksi 865 artikkelia.

Tämän systemaattisen kirjallisuuskatsauksen sisäänottokriteerit olivat seuraavat:

- 1) Artikkelin on vertaisarvioitu tutkimusartikkeli tai yksittäinen vertaisarvioitu konferenssijulkaisu.
- 2) Tutkimuksessa on tutkittu suurta kielimallia hyödyntävän keskustelubotin käyttöä kommunikaatio- tai puhetaidon kehittämiseen vieraan tai toisen kielen oppimisessa.
- 3) Tutkimus on suoritettu kokeellisena tai kvasikokeellisena tutkimuksena.
- 4) Tutkimuksessa yksi tai useampi koeryhmä on harjoitellut puhe- tai kommunikaatiotaitoja vieraalla kielellä käyttäen suurta kielimallia hyödyntävää keskustelubottia.
- 5) Tutkimuksessa on verrattu koeryhmää tai koeryhmiä kontrolliryhmään, jossa puhe- tai kommunikaatiotaitoja vieraalla tai toisella kielellä on harjoiteltu toisten ihmisten kanssa ilman keskustelubotin apua.
- 6) Tutkimuksessa on mitattu kommunikaatio- tai puhetaitoa siihen soveltuvalla mittarilla sekä pretest-posttest-mallilla.
- 7) Tutkimus on saatavilla englannin kielellä.

Katsauksen poissulkukriteerit olivat seuraavat:

- 1) Tuloksen kokoteksti ei ollut saatavilla ilmaiseksi tai Turun yliopiston kautta.
- 2) Tulos oli kirja tai konferenssijulkaisujen kokoelma.
- 3) Tutkimuksessa ei tutkittu kommunikaatio- tai puhetaitoa vieraalla tai toisella kielellä.
- 4) Artikkelin ei ollut vertaisarvioitu.
- 5) Tutkimuksessa ei tutkittu suurten kielimallien käyttöä.
- 6) Tutkimus ei ollut kokeellinen tai kvasikokeellinen.
- 7) Tutkimuksessa koeryhmä ei ollut harjoitellut puhe- tai kommunikaatiotaitoja suurta kielimallia hyödyntävää keskustelubottia käyttäen.
- 8) Tutkimuksessa kontrolliryhmä ei harjoitellut kieltä toisten ihmisten kanssa tai kontrolliryhmäkin käytti keskustelubottia apuna.
- 9) Tutkimuksessa kielitaitoa ei mitattu ennen ja jälkeen intervention.
- 10) Tutkimuksessa ei kuvattu, miten kielitaitoa mitattiin.
- 11) Tutkimus ei ollut saatavilla englannin kielellä.

Kävin seuraavaksi läpi sisäänottokriteerit huomioiden Rayyan-ohjelman avulla ensimmäisen haun tulosten otsikot ja abstraktit. Kokonaistekstitarkasteluun valitsin tutkimukset, jotka voisivat otsikon ja abstraktin perusteella täyttää sisäänottokriteerit. Otsikkojen ja abstraktien tarkastelun jälkeen kokotekstitarkasteluun jäi yhteensä 107 artikkelia. Kävin tämän jälkeen läpi nämä 107 kokotekstiä sisäänotto- sekä poissulkukriteerit huomioiden. Ensimmäisen haun kokotekstitarkastelun jälkeen jäljelle jäi 11 artikkelia. Kävin tämän jälkeen vielä lumipallotekniikalla läpi 11 löydetyn artikkelin lähdeluettelot sekä kokotekstitarkasteluun valittujen systemaattisissa katsauksissa mukana olleet artikkelit. Näistä ei löytynyt uusia sisäänottokriteerejä täyttäviä artikkeleita.

Tämän jälkeen kävin 13.12.2025 läpi muut tunnetut suuret kielimallit. Toteutin haut samoin kuin ensimmäisessä haussa, kohdistuen haun samoihin tietokantoihin sekä kokoteksteihin niissä tietokannoissa, joissa se oli mahdollista. Kävin muut suuret kielimallit läpi seuraavalla hakulausekkeella kaikissa tietokannoissa:

[Kielimalli] AND LLM AND ("learning languages" OR "language learning" OR "second language acquisition" OR "SLA") AND ("communication skill" OR "communication skills" OR "conversation skill" OR "conversation skills" OR "oral communication" OR "oral proficiency" OR "speaking proficiency") AND ("empirical study" OR "experimental study" OR "intervention" OR "control group" OR "comparison group") AND ("traditional methods" OR "traditional method" OR "human conversation" OR "peer" OR "peers" OR "native speaker" OR "native speakers" OR "another student" OR "another students" OR "other student" OR "other students")

Kävin tällä hakulausekkeella läpi nämä kielimallit, sijoittaen ”[Kielimalli]”-kohdan tilalle: Claude, Gemini, Mistral, DeepSeek, Gemma, Llama, “Command R+”, “Microsoft Phi” sekä “Phi-3 OR Phi-4”. Muiden kielimallien nimet olivat osittain yleisessä käytössä, joten lisähaussa hakulausekkeen termi ”LLM” oli tarpeellinen lisärajausena.

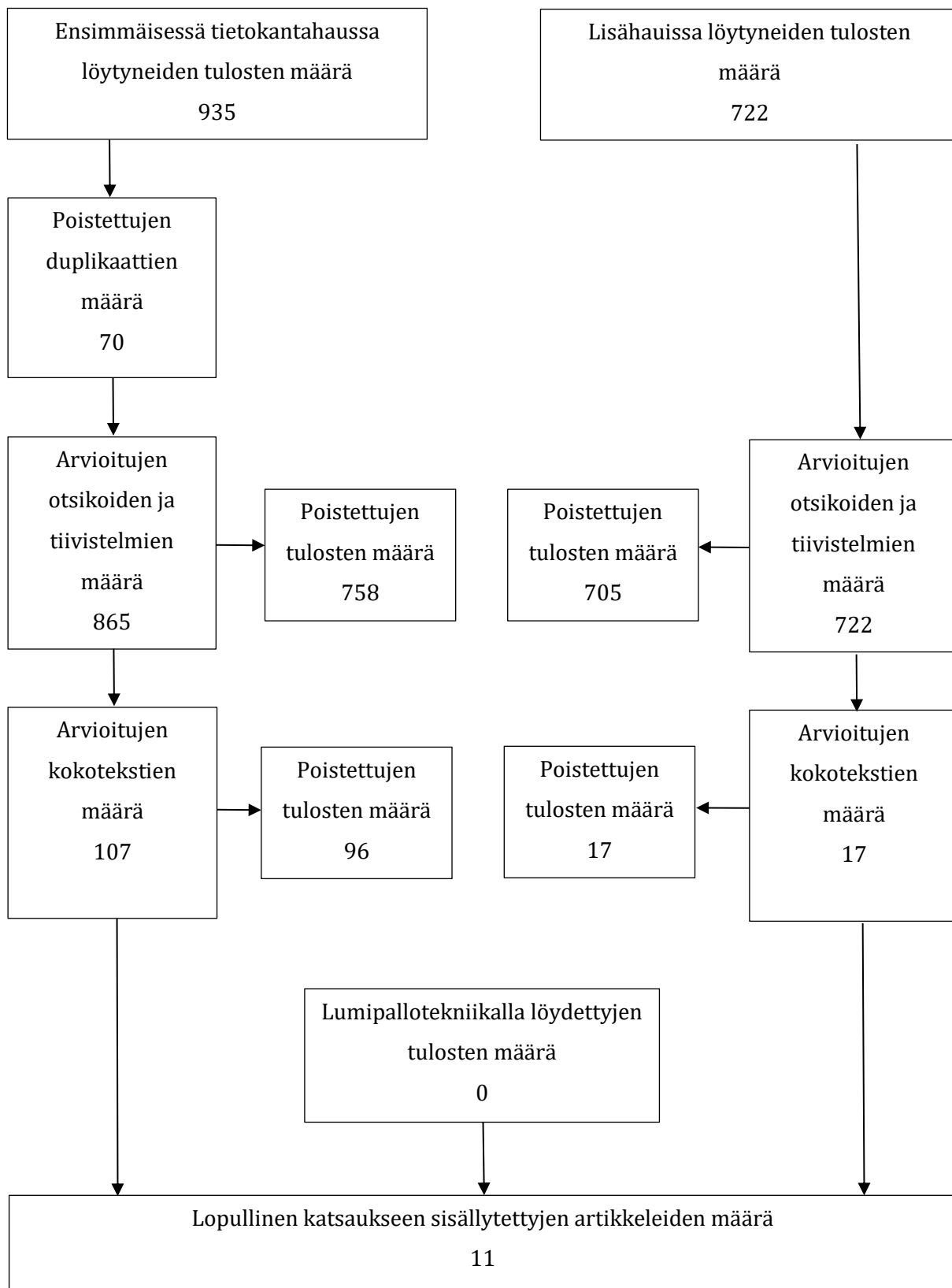
Yhteensä valituista tietokannoista löydettiin kielimallien lisähauissa 656 tulosta, joista ACM Digital Librarystä löytyi 435 tulosta, joista kaikki olivat konferenssijulkaisujen kokoelmia, Scopusesta löytyi 9 tulosta, SpringerLinkistä 32, ProQuest Centralista 96, Wiley Online Librarystä 42, Sage Journalsista 30 sekä Taylor & Francis Onlinesta 12. APA PsycInfosta, IEEE Xploresta ja Web of Sciencesta ei löytynyt yhtään tulosta. Nämäkin hakulausekkeet olivat liian monimutkaisia ScienceDirect-tietokantaan.

Näiden lisäksi ensimmäisestä hausta nousi esiin kolme erillistä kielten harjoittelussa käytettyä sovellusta: EAP Talk, CoolE bot ja Replika. Koska artikkelissa voidaan mainita pelkkä sovellus ilman sen hyödyntämää teknologiaa, tein vielä 13.12.2025 kolmannen hakuvaiheen, jossa kävin läpi aiempien hakujen tapaan samat tietokannat seuraavalla hakulausekkeella:

[Sovellus] AND ("learning languages" OR "language learning" OR "second language acquisition" OR "SLA") AND ("communication skill" OR "communication skills" OR "conversation skill" OR "conversation skills" OR "oral communication" OR "oral proficiency" OR "speaking proficiency") AND ("empirical study" OR "experimental study" OR "intervention" OR "control group" OR "comparison group") AND ("traditional methods" OR "traditional method" OR "human conversation" OR "peer" OR "peers" OR "native speaker" OR "native speakers" OR "another student" OR "another students" OR "other student" OR "other students")

Yhteensä valituista tietokannoista löydettiin sovellusten lisähauissa 66 tulosta, joista ACM Digital Librarystä löytyi 22 tulosta, Scopusesta löytyi 16, SpringerLinkistä 10, ProQuest Centralista 6, Wiley Online Librarystä 5 sekä Taylor & Francis Onlinesta 7. APA PsycInfosta, Sage Journalsista, IEEE Xploresta ja Web of Sciencesta ei löytynyt yhtään tulosta. Kuten aiemmin, nämäkin hakulausekkeet olivat liian monimutkaisia ScienceDirect-tietokantaan.

Kävin seuraavaksi manuaalisesti läpi toisen ja kolmannen hakuvaiheen tulosten otsikot ja abstraktit sisäänotto- ja poissulkukriteerit huomioiden. Vertasin otsikoita ja abstrakteja ensimmäisen haun tuloksien otsikoihin ja abstrakteihin, jonka jälkeen kokotekstitarkasteluun jäi yhteensä 17 uutta artikkelia, joita ei löytynyt ensimmäisessä haussa. Lisähakujen kokotekstitarkastelun jälkeen jäljelle ei jäänyt yhtään uutta sisäänottokriteerit täyttävää artikkelia. Lopulta tähän katsaukseen valikoitui yhteensä 11 artikkelia. Kuvassa 1 esitetään vuokaavio tulosten läpikäymisestä.



**Kuva 1.** Vuokaavio tietokantahausta ja tulosten läpikäymisestä.

## 4 Tulokset

### 4.1 Tutkimusaineisto

Tämä systemaattinen kirjallisuuskatsaus pyrki selvittämään, voisiko kielimalliagenteista olla hyötyä kielten oppimisessa puhutun kielen ja kommunikaatiotaidon näkökulmasta, ja miten kielimalliagenttien käyttö vertautuu perinteisiin oppimismenetelmiin. Katsaukseen valikoitui lopulta 11 artikkelia sekä niiden 11 tutkimusta. Näistä 11 tutkimusartikkelista yhdeksän tutki OpenAI:n GPT-malleja hyödyntävien keskustelubottien vaikutuksia (Celik ym., 2025; Chen ym., 2025; Feng & Wang, 2025; Tai & Chen, 2024; Wang ym., 2024; Wang, 2025; Yang ym., 2025; Zheng ym., 2025; Zhou ym., 2025) ja kaksi tutki Replika-nimisen sovelluksen vaikutuksia puhe- ja kommunikaatiotaitoihin (Shalaan, 2025; Wu ym., 2025).

Kaikki artikkelit ovat julkaistu aikavälillä 2024–2025. Suurin osa tutkimuksista oli toteutettu Kiinassa (7), yksi Taiwanissa, yksi Irakissa, yksi Saudi-Arabiassa ja yksi Indonesiassa. Kaikki tutkimukset olivat joko kokeellisia (4) tai kvasikokeellisia (7) tutkimuksia. Joissain mukaan otetuissa tutkimusartikkeleissa oli kokeellisten tutkimusten lisäksi mukana myös kvalitatiivisia tutkimuksia sekä kyselytutkimuksia, mutta ne keskittyivät esimerkiksi tutkittavien kokemukseen kielimalliagentin hyödyistä tai tutkittavien käsitykseen omasta vieraan kielen taidostaan. Kaikissa tutkimuksissa vieraana kielenä tutkittiin englantia. Tutkimusten otoskokojen vaihteluväli oli 44–120 koehenkilöä. Kaikissa paitsi yhdessä tutkimuksessa koehenkilöt olivat yliopiston opiskelijoita. Tässä yhdessä tutkimuksessa tutkittiin kuudesluokkalaisia, noin 12–13-vuotiaita oppilaita (Tai & Chen, 2024). Kaikissa tutkimuksissa ei ollut raportoitu koehenkilöiden ikä- tai sukupuolijakaumaa, joten katsaukseen sisällytetyjen tutkittavien ikä- tai sukupuolijakauma ei ole tiedossa. Katsaukseen sisällytetyt tutkimukset esitellään Taulukossa 1.

**Taulukko 1.** *Katsaukseen valitut tutkimukset*

Tekijät, julkaisuvuosi, maa	Otos	Tutkittu kieli, tutkimusasetelma ja intervention kesto	Muuttujat ja mittarit	Mitattava vaikutus	Tulokset
Celik ym., 2025, Irak	44 yliopiston ”Advanced English” -kurssin opiskelijaa (18–28-vuotiaita)	Englanti, Kokeellinen (1 koeryhmä, n=22, ja 1 kontrolliryhmä, n=22), 8 viikkoa	Puhetaito, EFL asiantuntijoiden luoma puhetaidon koe (pretest-posttest)	ChatGPT:n (versiota ei mainittu) kanssa käytävän puhevälitteisen keskustelun vaikutus englannin puhetaitoon.	Kontrolliryhmän pretest-tulokset: M=53,41; SD=14,75 Koeryhmän pretest-tulokset: M=54,32; SD=14,16 Kontrolliryhmän posttest-tulokset: M=57,05; SD=16,15 Koeryhmän posttest-tulokset: M=80,00; SD=10,69
Chen ym., 2025, Kiina	53 yliopiston englannin kurssin opiskelijaa (ikää ei raportoitu)	Englanti, Kvasikokeellinen (1 koeryhmä, n=29, ja 1 kontrolliryhmä, n=24), 2 viikkoa	Puhetaito, puhutut vasaukset kysymyksiin, jotka arvioitu IELTS Speaking Band:in avulla (pretest-posttest)	GPT-4 API:a käytävän virtuaalisen hahmon kanssa käytävän puhevälitteisen keskustelun vaikutus englannin puhetaitoon.	Keskivertoero pretestin ja posttestin tuloksien välillä: Kontrolliryhmä: 0,310 Koeryhmä: 0,429 Ryhmien ero ei ollut tilastollisesti merkitsevä (F = 0,019, p = 0,892, $\eta_p^2 = 0,001$ ).

Tekijät, julkaisuvuosi, maa	Otos	Tutkittu kieli, tutkimusasetelma ja intervention kesto	Muuttujat ja mittarit	Mitattava vaikutus	Tulokset
Feng & Wang, 2025, Kiina	120 yliopiston ”College English (Comprehensive English)” -kurssin opiskelijaa (ikää ei raportoitu)	Englanti, Kvasikokeellinen (1 koeryhmä, n=41, ja 1 kontrolliryhmä, n=42), 16 viikkoa	Kuullun- ja luetun-ymmärtäminen sekä puhe-, kirjoitus- ja kääntämistaito, Edellisen lukukauden ”College English” -kurssin arvosana ja kuluneen ”College English” -kurssin arvosana (pretest-posttest) sekä CET-4 ja CET-6 kokeet (pretest-posttest)	ChatGPT:n (versiota ei mainittu) kanssa englannin kielen harjoittelun vaikutus englannin kielen taitoon eri kielen osa-alueilla.	<p>Ei tilastollisia tuloksia erikseen puhetaidosta. Tutkimuksessa mainittiin sanallisesti, että ChatGPT:n käyttö paransi merkittävästi opiskelijoiden suullisia taitoja sekä kirjoitustaitoja. ”College English” -kurssien arvosanat:</p> <p>Kontrolliryhmän pretest-tulokset (n=42): M=73,25±8,49  Koeryhmän pretest-tulokset (n=41): M=72,84±8,52  Kontrolliryhmän posttest-tulokset (n=42): M=74,13±8,66  Koeryhmän posttest-tulokset (n=41): M=87,36±9,24</p> <p>CET-4-kokeiden tulokset:  Kontrolliryhmän pretest-tulokset (n=15): M=408,17±27,33  Koeryhmän pretest-tulokset (n=13): M=405,52±26,75  Kontrolliryhmän posttest-tulokset (n=19): M=421,46±30,25  Koeryhmän posttest-tulokset (n=37): M=506,41±43,16</p> <p>CET-6-kokeiden tulokset:  Kontrolliryhmän pretest-tulokset (n=3): M=387,11±29,14  Koeryhmän pretest-tulokset (n=4): M=386,39±28,77  Kontrolliryhmän posttest-tulokset (n=7): M=413,52±31,42  Koeryhmän posttest-tulokset (n=23): M=487,36±40,83</p>
Shalaan, 2025, Saudi-Arabia	120 englannin kurssin opiskelijaa, (18–35-vuotiaita)	Englanti, kokeellinen (1 koeryhmä, n=60, ja 1 kontrolliryhmä, n=60), 12 viikkoa	Puhetaito, kielitieteilijöiden pisteyttämä strukturoitu suullinen haastattelu	Replika-sovelluksen kanssa keskustelun vaikutus englannin kielen puhetaitoon.	<p>Kontrolliryhmän pretest-tulokset: M=58; SD=14  Koeryhmän pretest-tulokset: M=55; SD=15  Kontrolliryhmän posttest-tulokset: M=67; SD=12  Koeryhmän posttest-tulokset: M=71; SD=13</p> <p>Ryhmiä posttest-tuloksien ero oli tilastollisesti merkitsevä:  <math>F(1,118) = 18,81; p &lt; 0,01</math>.</p>

Tekijät, julkaisuvuosi, maa	Otos	Tutkittu kieli, tutkimusasetelma ja intervention kesto	Muuttujat ja mittarit	Mitattava vaikutus	Tulokset
Tai & Chen, 2024, Taiwan	85 kuudesluokkalaista oppilasta (noin 12–13-vuotiaita)	Englanti, kvasikokeellinen (2 koeryhmää, n=28 ja n=28 sekä 1 kontrolliryhmä, n=29) (1. koeryhmä käytti keskusteluagenttia yksin (I-Bot-ryhmä) ja 2. koeryhmä parin kanssa (P-Bot-ryhmä)), 3 viikkoa	Puhetaito, Puhetaidon osuudet General English Proficiency Test (GEPT)) Kids -testistä (pretest-posttest)	ChatGPT:tä (versiota ei mainittu) hyödyntävän keskusteluagentin (CoolE Bot) kanssa käytävän keskustelun vaikutus englannin puhetaitoon.	Kontrolliryhmän pretest-tulokset: M=28,16; SD=10,90 I-Bot-ryhmän pretest-tulokset: M=31,57; SD=14,62 P-Bot-ryhmän pretest-tulokset: M=24,86; SD=10,08  Kontrolliryhmän posttest-tulokset: M=49,34; SD=11,82 I-Bot-ryhmän posttest-tulokset: M=63,21; SD=16,50 P-Bot-ryhmän posttest-tulokset: M=62,96; SD=10,89  Post hoc Scheffé -testien avulla löydettiin tilastollisesti merkitsevä ero koeryhmien ja kontrolliryhmän välillä. Kahden koeryhmän välillä ei ollut merkitsevää eroa.
Wang ym., 2024, Kiina	99 yliopiston englannin kurssin opiskelijaa, (19–23-vuotiaita)	Englanti, kvasikokeellinen (2 koeryhmää, n=33 ja n=33 sekä 1 kontrolliryhmä, n=33) (1. koeryhmä käytti Typebot-keskusteluagenttia (Typebot-ryhmä) ja 2. koeryhmä D-ID Agent -keskusteluagenttia (D-ID-ryhmä)), 6 viikkoa	Puhetaito, kahden englannin opettajan arvioimat testit (IELTS Speaking test:in ensimmäisestä osan kaltaiset 9 arkielämään liittyvää kysymystä sekä 3 minuutin puhuttu esitelmä) (pretest-posttest)	GPT-3.5 API:tä hyödyntävien keskusteluagenttien (Typebot ja D-ID Agent) kanssa käytävän keskustelun vaikutus englannin puhetaitoon.	Kontrolliryhmän pretest-tulokset: M=5,49; SD=1,20 Typebot-ryhmän pretest-tulokset: M=5,24; SD=1,68 D-ID Agent -ryhmän pretest-tulokset: M=5,18; SD=2,02  Kontrolliryhmän posttest-tulokset: M=5,76; SD=1,35 Typebot-ryhmän posttest-tulokset: M=5,61; SD=2,03 D-ID Agent -ryhmän posttest-tulokset: M=6,21; SD=2,06  Kummankaan koeryhmän ja kontrolliryhmän posttest:ien välillä ei ollut tilastollisesti merkitsevää eroa (Typebot-ryhmä: b = 0,00; SE = 0,61; z = 0,00; p > 0,05 ja D-ID Agent -ryhmä: b = 0,67; SE = 0,61; z = 1,10; p > 0,05)

Tekijät, julkaisuvuosi, maa	Otos	Tutkittu kieli, tutkimusasetelma ja intervention kesto	Muuttujat ja mittarit	Mitattava vaikutus	Tulokset
Wang, 2025, Kiina	68 yliopiston englannin kurssin opiskelijaa (ikää ei raportoitu)	Englanti, kokeellinen (1 koeryhmä, n=33, ja 1 kontrolliryhmä, n=35), 12 viikkoa	Suullinen kommunikaatiotaito, 20 minuutin keskustelutehtävä (Test of English for International Communication) vertaisen kanssa, jonka opettaja arvioi (pretest-posttest)	ChatGPT-4:ltä saadun palautteen vaikutus opiskelijoiden puhuttuun kommunikaatiotaitoon.	Kontrolliryhmän pretest-tuloksia ei raportoitu. Koeryhmän pretest-tulokset: M=2,51; SD=0,03 Kontrolliryhmän posttest-tulokset: M=1,27; SD=0,52 Koeryhmän posttest-tulokset: M=3,66; SD=0,38  Ryhmien posttest-tuloksien ero oli tilastollisesti merkitsevä (p=0.000). Lisäksi koeryhmä pretest- ja posttest-tuloksien ero oli tilastollisesti merkitsevä (p=0.000).
Wu ym., 2025, Indonesia	75 yliopiston ”Casual Conversation in English” -kurssin opiskelijaa (18–20-vuotiaita)	Englanti, kvasikokeellinen (1 koeryhmä, n=39, ja 1 kontrolliryhmä, n=36), 6 viikkoa	Puhetaito, puhutut vasaukset kysymyksiin, jotka arvioitu kahden opettajan toimesta IELTS Speaking Band:in avulla (pretest-posttest)	Replika-sovelluksen kanssa keskustelun vaikutus englannin kielen puhetaitoon neljällä osaluella: sujuvuus ja johdonmukaisuus (SJ), sanavaraston laajuus (SVL), kieliopin laajuus ja tarkkuus (KLT) sekä ääntäminen (Ä).	Kontrolliryhmän (CG) ja koeryhmän (EG) pretest-tulokset: CG-SJ: M=5,861; SD=0,351   EG-SJ: M=5,628; SD=0,732 CG-SVL: M=5,861; SD=0,351   EG-SVL: M=5,513; SD=0,730 CG-KLT: M=5,833; SD=0,378   EG-KLT: M=5,628; SD=0,686 CG-Ä: M=5,889; SD=0,433   EG-Ä: M=5,755; SD=0,457  Kontrolliryhmän (CG) ja koeryhmän (EG) posttest-tulokset: CG-SJ: M=6,028; SD=0,266   EG-SJ: M=6,308; SD=0,635 CG-SVL: M=5,944; SD=0,232   EG-SVL: M=6,359; SD=0,678 CG-KLT: M=6,028; SD=0,446   EG-KLT: M=6,054; SD=0,762 CG-Ä: M=6,472; SD=0,506   EG-Ä: M=6,603; SD=0,620  Ryhmien posttest-tuloksien välillä oli tilastollisesti merkitsevät erot: SJ: F = 29,862; p = 0,000; $\eta_p^2 = 0,293$ SVL: F = 97,784; p = 0,000; $\eta_p^2 = 0,576$ KLT: F = 4,287; p = 0,056; $\eta_p^2 = 0,056$ SJ: F = 8,258; p = 0,005; $\eta_p^2 = 0,103$

Tekijät, julkaisuvuosi, maa	Otos	Tutkittu kieli, tutkimusasetelma ja intervention kesto	Muuttujat ja mittarit	Mitattava vaikutus	Tulokset
Yang ym., 2025, Kiina	52 yliopiston englannin kurssin opiskelijaa (ikää ei raportoitu)	Englanti, kvasikokeellinen (1 koeryhmä, n=26, ja 1 kontrolliryhmä, n=28), 9 viikkoa	Improvisoidun puheen taito, Puhetaito kahden englannin opettajan arvioimana perustuen kolmeen skaalaan: China's standards of English language Ability (CSE), The Common European Framework of Reference for Languages (CEFR) ja IELTS Speaking Band Descriptors (pretest-posttest)	ChatGPT-3.5:ltä saadun palautteen vaikutus opiskelijoiden improvisoidun puheen taitoon neljällä osa-alueella: ääntämisen tarkkuus (ÄT), ilmaisun sujuvuus (IS), ilmaisun täydellisyys (IT) ja ilmaisun oikeellisuus (IO).	Ryhmien väliset pretest-tuloksien erot eivät olleet merkitseviä eikä niitä raportoitu erikseen. Kontrolliryhmän posttest-tulokset (keskihajontaa ei raportoitu): ÄT: M=17,58 IS: M=12,69 IT: M=18,50 IO: M=13,00  Koeryhmän posttest-tulokset ja ero kontrolliryhmään: ÄT: M=24,08 (F=5,687; t=5,683; p=0.000) IS: M=15,54 (F=0,327; t= 3,665; p=0.001) IT: M=23,38 (F=0,350; t=3,767; p=0.002) IO: M=16,46 (F=8,872; t=4,356; p=0.000)
Zheng ym., 2025, Kiina	83 yliopiston englannin kurssin opiskelijaa (18–19-vuotiaita)	Englanti, kokeellinen (1 koeryhmä, n=41, ja 1 kontrolliryhmä, n=42), 4 viikkoa	Puhetaito, kahden opettajan arvioimat Fulcherin (2014) perusteella tehdyt testit, joissa mitattiin tarkkuutta, sujuvuutta, täydellisyyttä sekä ääntämistä (pretest-posttest)	GPT-4 API:ta käyttävän sovelluksen kanssa käytävän keskustelun vaikutus englannin kielen puhetaitoon.	Kontrolliryhmän pretest-tulokset: M=87,17; SD=4,10 Koeryhmän pretest-tulokset: M=87,36; SD=3,76 Kontrolliryhmän posttest-tulokset: M=88,36; SD=4,67 Koeryhmän posttest-tulokset: M=90,44; SD=4,33  Kontrolliryhmän posttest-tulokset kuukauden kuluttua interventiosta: M=87,46; SD=3,41 Koeryhmän posttest-tulokset kuukauden kuluttua interventiosta: M=89,34; SD=3,21  Ryhmien posttest-tuloksien välillä oli tilastollisesti merkitsevä ero (p=0,023).

Tekijät, julkaisuvuosi, maa	Otos	Tutkittu kieli, tutkimusasetelma ja intervention kesto	Muuttujat ja mittarit	Mitattava vaikutus	Tulokset
Zhou ym., 2025, Kiina	72 yliopiston englannin kurssin opiskelijaa (ikää ei raportoitu)	Englanti, kvasikokeellinen (1 koeryhmä, n=36, ja 1 kontrolliryhmä, n=36), 8 viikkoa	Puhetaito, kahden opettajan arvioimat (1 koeryhmä, n=36, puhetaidon osuudet CET-4-testistä (pretest-posttest))	ChatGPT-4:n kanssa käytävän keskustelun vaikutus englannin kielen puhetaitoon kolmella osa-alueella: kielellinen tarkkuus ja monimutkaisuus (KTM), puheen pituus ja sujuvuus (PPS) sekä kielen joustavuus ja sopivuus (KJS).	Kontrolliryhmän pretest-tulokset: KTM: M=2,71; SD=0,44 PPS: M=2,57; SD=0,41; KJS: M=2,76; SD=0,40 Yhteensä: M=8,03; SD=1,13 Koeryhmän pretest-tulokset: KTM: M=2,51; SD=0,47 PPS: M=2,43; SD=0,43; KJS: M=2,74; SD=0,34 Yhteensä: M=7,69; SD=1,04 Kontrolliryhmän posttest-tulokset: KTM: M=2,66; SD=0,40 PPS: M=2,51; SD=0,38; KJS: M=2,88; SD=0,35 Yhteensä: M=8,06; SD=0,92 Koeryhmän posttest-tulokset: KTM: M=2,64; SD=0,40 PPS: M=3,07; SD=0,36; KJS: M=2,89; SD=0,35 Yhteensä: M=8,67; SD=1,04 Ryhmien posttest-tuloksien erot olivat tilastollisesti merkitseviä puheen pituuden ja sujuvuuden ( $p<0,001$ ) sekä yhteispisteiden ( $p=0,015$ ) suhteen. Ryhmien posttest-tuloksien erot eivät olleet merkitseviä kielellisen tarkkuuden ja monimutkaisuuden ( $p=0,920$ ) sekä kielen joustavuuden ja sopivuuden ( $p=0,733$ ) suhteen.

EFL = Englanti vieraana kielenä (*English as a foreign language*)

IELTS = Kansainvälinen englannin kielen testijärjestelmä (*International English Language Testing System*)

CET-4 = Kiinan opetusministeriön standardoitu testi: Yliopiston englannin testi 4 (*College English Test Band 4*)

CET-6 = Kiinan opetusministeriön standardoitu testi: Yliopiston englannin testi 6 (*College English Test Band 6*)

M = Keskiarvo (*mean*)

SD = Keskihajonta (*standard deviation*)

SJ = Sujuvuus ja johdonmukaisuus (*fluency and coherence*)

SVL = Sanavaraston laajuus (*lexical resource*)

KLT = Kieliopin laajuus ja tarkkuus (*grammatical range and accuracy*)

Ä = Ääntäminen (*pronunciation*)

ÄT = Ääntämisen tarkkuus (*pronunciation accuracy*)

IS = Ilmaisun sujuvuus (*expression fluency*)

IT = Ilmaisun täydellisyys (*expression completeness*)

IO = Ilmaisun oikeellisuus (*expression correctness*)

KTM = Kielellinen tarkkuus ja monimutkaisuus (*linguistic accuracy and complexity*)

PPS = Puheen pituus ja sujuvuus (*utterance length and fluency*)

KJS = Kielen joustavuus ja sopivuus (*language flexibility and appropriateness*)

## 4.2 Replika-sovellusta hyödyntäneet tutkimukset

Replika-sovellusta hyödyntäneet tutkimusartikkelit eivät maininneet, mikä teknologia Replika-sovelluksen taustalla on. Replika-sovelluksen omien verkkosivujen mukaisesti Replika-sovellus toimii ”yhdistämällä kehittyneen neuroverkkopohjaisen koneoppimismallin ja skriptatun dialogisisällön” (Luca Inc, ei pvm., [oma suomennos]). Ei ole varmuutta, että millainen neuroverkkopohjainen koneoppimismalli Replikan taustalla on. Kirjallisuushaussa löytyi myös 24.10.2022 Novitas-ROYAL-lehdessä julkaistu kolmas Replikaa hyödyntävä artikkeli (Çakmak, 2022) Tätä artikkelia ei otettu mukaan katsaukseen, koska artikkeli ei täyttänyt sisäänottokriteerejä, sillä tutkimuksessa ei jaettu osallistujia koe- ja kontrolliryhmiin. Tämä tutkimus kuitenkin osoittaa, että Replikaa on tutkittu jo ennen kuin ensimmäinen ChatGPT-malli julkaistiin, eikä tässä tutkimuksessa ei mainittu Replikan tarkkaa toimintaperiaatetta. Koska en saanut selville, hyödynnetäänkö Replikassa jotain suurta kielimallia, jätin sitä hyödyntäneet artikkelit pois tarkemmista jatkovertailuista, sillä toimintaperiaatteen epäselvyys heikentää vertailtavuutta. Tästä syystä vertailen seuraavissa kappaleissa tarkemmin vain näitä yhdeksää kielimalliagenttien käyttöä tutkineita tutkimuksia.

## 4.3 Tutkimusasetelmat

### 4.3.1 Muuttujat ja termit

Tutkimuksissa käytettiin puhe- ja kommunikaatiotaidoista eri termejä. Tutkimuksissa käytetyt termit olivat suullinen kielitaito (*oral proficiency*), puhetaidot (*speaking skills*), improvisoidun puheen suorituskkyky (*impromptu speaking performance*) sekä kommunikaatiotaidot (*communication skills*). Näitä termejä käytettiin tutkimuksissakin osittain päällekkäisesti, joten tämä muuttujiin ja termeihin jaottelu ei ole täysin tarkka. Kaikki tutkimukset kuitenkin tutkivat puhe- tai kommunikaatiotaitoja.

Kahdessa tutkimuksessa tutkittiin suullista kielitaitoa (*oral proficiency*) (Zheng ym., 2025; Zhou ym., 2025). Zheng ja kumppanit (2025) tutkivat suullisen kielitaidon neljää osa-aluetta: tarkkuus (*accuracy*), sujuvuus (*fluency*), täydellisyys (*completeness*) ja ääntäminen (*pronunciation*). Zhou ja kumppanit (2025) puolestaan tutkivat CET-4 kokeen arviointikriteerien mukaisesti suullisen kielitaidon kolmea osa-aluetta: kielellinen tarkkuus ja monimutkaisuus (*linguistic accuracy and complexity*), puheen pituus ja sujuvuus (*utterance*

*length and fluency*) sekä kielen joustavuus ja sopivuus (*language flexibility and appropriateness*).

Neljässä tutkimuksessa tutkittiin puhetaitoja (*speaking skills*) (Celik ym., 2025; Chen ym., 2025; Tai & Chen, 2024; C. Wang ym., 2024). Celikin ja kumppaneiden (2025) sekä Chenin ja kumppaneiden (2025) tutkimuksissa ei raportoitu mitä puhetaidon osa-alueita tutkittiin. Tai ja Chen (2024) tutkivat puhetaidon osa-alueista sujuvuutta (*fluency*), sisältöä (*content*), ääntämistä (*pronunciation*), kieliopillista sopivuutta (*grammatical appropriateness*) sekä sanaston käyttöä (*lexical usage*). Lisäksi he mainitsivat heidän käyttämänsä testin mittaavan koehenkilöiden kykyä selittää (*narrate*), ilmaista mielipiteitä (*express opinions*) ja vastata asianmukaisesti (*respond appropriately*). C. Wang ja kumppanit (2024) tutkivat muun muassa seuraavia puhetaidon osa-alueita: ääntäminen (*pronunciation*), sanasto (*vocabulary*), kyky kertoa omin sanoin (*ability to paraphrase*), joustavuus (*flexibility*), lauserakenteiden monimutkaisuus (*sentence structure complexness*), ilmeikkyys (*expression*) ja monipuolinen kielen käyttö (*language usage diverseness*).

Yhdessä tutkimuksessa keskityttiin tutkimaan improvisoidun puheen suorituskyykyä (*impromptu speaking performance*) (Yang ym., 2025). Yang ja kumppanit (2025) mittasivat tämän neljää osa-aluetta: ääntämisen tarkkuus (*pronunciation accuracy*), ilmaisun sujuvuus (*expression fluency*), ilmaisun täydellisyys (*expression completeness*) sekä ilmaisun oikeellisuus (*expression correctness*).

Yhdessä tutkimuksessa tutkittiin kommunikaatiotaitoja (*communication skills*) (Y. Wang, 2025). Y. Wang (2025) tutki koehenkilöiden kommunikaatiotaitoja puhetaidon sekä keskustelun aikana tehtyjen virheiden kautta.

Feng ja Wang (2025) eivät tutkineet pääasiallisesti mitään näistä, eivätkä he raportoineet tilastollisia tuloksia näiden muuttujien osalta, sillä he tutkivat englannin kielen taitoja yleisesti. He kuitenkin mainitsivat artikkelissaan, että ChatGPT:n käyttö paransi merkittävästi koehenkilöiden suullista englannin kielen taitoa.

#### 4.3.2 Mittarit

Kahdessa tutkimuksessa käytettiin valmiita testejä (Feng & Wang, 2025; Zhou ym., 2025). Feng ja Wang (2025) käyttivät kielitaidon mittaamiseen *CET-4*- ja *CET-6*-kokeita. Lisäksi Feng ja Wang (2025) käyttivät *College English* -kurssin loppukoetta kielitaidon mittaamiseen.

Näiden testien Cronbachin alfaa ei raportoitu. Zhoun ja kumppaneiden (2025) tutkimuksessa käytettiin myös *CET-4*-testiä, mutta he käyttivät vain tämän testin suullista osuutta.

Neljässä tutkimuksessa käytetyt testit luotiin soveltamalla valmiita testejä (Chen ym., 2025; Tai & Chen, 2024; Yang ym., 2025; Zheng ym., 2025). Tain ja Chenin (2024) tutkimuksessa sovellettiin Yleisen englannin kielitaidon testin lasten versiota (*General English Proficiency Test [GEPT] Kids*). Tutkimuksessa käytetyssä testissä oli kolme osiota: yksinkertaisten lauseiden ääneen lukeminen (*reading simple sentences aloud*), yksinkertaisten lauseiden käyttäminen esineiden ja ihmisten esittelyyn (*using simple sentences to introduce things and people*) ja päivittäiseen elämään liittyviin kysymyksiin vastaaminen (*responding to questions related to daily life*). Chenin ja kumppaneiden (2025) tutkimuksessa sovellettiin *IELTS*-testin puheosion (*International English Language Testing System Speaking Band*) kolmea osiota: sujuvuus ja joustavuus (*fluency and flexibility*), sanaston monipuolisuus (*vocabulary diversity*) sekä kieliopin monipuolisuus ja tarkkuus (*grammar diversity and accuracy*). Yangin ja kumppaneiden (2025) tutkimuksessa kielitaidon mittaaminen pohjautui kolmeen arviointiasteikkoon: Kiinan englannin kielen taidon standardit (*China's Standards of English Language Ability*), Eurooppalaiseen kielitaidon viitekehys (*Common European Framework of Reference for Languages*) sekä Chenin ja kumppaneidenkin (2025) soveltama *IELTS*-testin puheosion asteikko (*IELTS Speaking Band Descriptors*). Zhengin ja kumppaneiden (2025) tutkimuksessa käytetty testi pohjautui Fulcherin (2014) kehittämän testiin.

Kolmessa tutkimuksessa oli käytössä tutkimuksen oma testi kielitaidon mittaamiseen (Celik ym., 2025; C. Wang ym., 2024; Y. Wang, 2025). Celikin ja kumppaneiden (2025) tutkimuksessa englannin kielen oppimisen asiantuntijat loivat testin puhetaidon mittaamiseen. C. Wangin ja kumppaneiden (2024) käyttämä testi koostui kahdesta osasta. Ensimmäinen osa oli samankaltainen kuin Chenin ja kumppaneiden (2025) käyttämä *IELTS*-testin puheosio. Toisessa osassa esitettiin kolmen minuutin suullinen esitelmä liittyen kurssimateriaaliin. Y. Wangin (2025) tutkimuksessa testaamiseen käytettiin opettajien arvioimaa keskustelua koehenkilöiden välillä.

Vain Celik ja kumppanit (2025) ilmoittivat käyttämänsä testin Cronbachin alfan, joka oli 0.89.

#### 4.3.3 Suuren kielimallin käyttö

Viidessä tutkimuksessa höydynnettiin GPT-kielimallia suoraan ChatGPT:n nettiselaimessa tai sovelluksessa (Celik ym., 2025; Feng & Wang, 2025; Y. Wang, 2025; Yang ym., 2025; Zhou

ym., 2025). Neljässä tutkimuksessa oli käytössä eri sovelluksia, joissa kielimallia hyödynnettiin ohjelmointirajapinnan, eli API:n, kautta (Chen ym., 2025; Tai & Chen, 2024; C. Wang ym., 2024; Zheng ym., 2025). Tutkimuksissa oli käytössä eri GPT-kielimallin versioita. Ohjelmointirajapintaa käyttävissä tutkimuksissa C. Wangin ja kumppaneiden (2024) tutkimuksessa oli käytössä kaksi eri keskusteluagenttia Typebot ja D-ID Agent, jotka molemmat hyödynsivät GPT-3.5:n ohjelmointirajapintaa. Kahdessa, Chenin ja kumppaneiden (2025) sekä Zhengin ja kumppaneiden (2025), tutkimuksissa käytettiin heidän itsensä luomaa keskusteluagenttia, jotka hyödynsivät GPT-4:n ohjelmointirajapintaa. Tai ja Chen (2024) käyttivät tutkimuksessaan CoolE bot -nimistä keskusteluagenttia, mutta heidän tutkimuksessaan ei mainittu käytettyä GPT:n versiota, vaan mainittiin CoolE bot -agentin vain käyttävän ChatGPT:tä. ChatGPT:tä suoraan käyttävissä tutkimuksissa Yang ja kumppanit (2025) käyttivät ChatGPT-3.5:tä ja Y. Wang (2025) sekä Zhou ja kumppanit (2025) käyttivät ChatGPT-4:ää. Celikin ja kumppaneiden (2025) sekä Fengin ja Wangin (2025) artikkeleissa ei raportoitu käytettyä ChatGPT:n versiota.

#### 4.3.4 Kielimalliagentin kanssa kommunikoinnin toteutus

Seitsemässä tutkimuksessa kielimalliagentin kanssa kommunikointiin käytettiin pääsääntöisesti jotain puheentunnistussovellusta. Näistä neljä tutkimusta käytti OpenAI:n puheentunnistusta (Celik ym., 2025; Feng & Wang, 2025; C. Wang ym., 2024; Zhou ym., 2025) ja kolme tutkimusta Microsoftin puheentunnistussovelluksia. Kaksi näistä käyttivät Microsoftin *Azure AI Speech* -sovellusta (Tai & Chen, 2024; Zheng ym., 2025) ja yhdessä tutkimuksessa Microsoftin puheentunnistussovellusta ei eritelty (Chen ym., 2025). C. Wang ja kumppanit (2024) käyttivät OpenAI:n puheentunnistuksen lisäksi *EAP Talk* -sovellusta puheentunnistukseen ja puheen harjoitteluun. *EAP Talk* on sovellus, jonka avulla voi harjoitella englannin kielen puhetaitoja (C. Wang & Zou, 2024), mutta *EAP Talkista* ei mainittu, miten sen puheentunnistus toteutettiin ja mitä tekoälysovellusta siinä käytettiin.

Kahdessa tutkimuksessa puheentunnistusta käytettiin joko osittain tai ei ollenkaan. Yang ja kumppanit (2025) muuttivat *iFLYTEK*in *FiF Intelligent Learning System* -sovelluksen avulla puheen tekstiksi ja syöttivät sen ChatGPT-3.5:lle. Tämän jälkeen ChatGPT:n kanssa käyty keskustelu puheen oikeellisuudesta ja puheen kieliopillinen korjaaminen toteutettiin tekstipohjaisesti. Y. Wangin (2025) artikkelissa ei mainittu käytettiinkö puheentunnistusta, mutta artikkelista kuvattu vuorovaikutustapa viittaa siihen, että tutkimuksessa käytettiin tekstipohjaista kommunikaatiota ChatGPT-4:n kanssa.

#### 4.3.5 Kielimalliagentin rooli

Kuudessa tutkimuksessa kielimalliagentin rooli oli toimia kuten tuutorina koehenkilölle. kielimalliagentti toimi keskustelukumppanina, mutta se myös samalla esimerkiksi korjasi koehenkilön virheitä, ehdotti keskusteluaiheita ja toimi tukena koehenkilölle (Celik ym., 2025; Chen ym., 2025; Tai & Chen, 2024; C. Wang ym., 2024; Zheng ym., 2025; Zhou ym., 2025). Yhdessä tutkimuksessa kielimalliagentti toimi tuutorina sekä koehenkilöille että opettajalle (Feng & Wang, 2025). Kahdessa tutkimuksessa kielimalliagentin rooli oli toimia kuin opettajana koehenkilöille. Koehenkilöt saivat kielimalliagentilta opettajan tapaan palautetta keskustelusuorituksesta, mutta agenttia ei hyödynnetty keskustelukumppanina (Y. Wang, 2025; Yang ym., 2025).

#### 4.3.6 Kielimalliagentin vertailun kohde kontrolliryhmässä

Kolmessa tutkimuksessa kielimalliagentti tuli lisänä koeryhmälle, eli kontrolliryhmä ei saanut mitään vastinetta agentille (Celik ym., 2025; Feng & Wang, 2025; Zhou ym., 2025). Kaikissa näissä tutkimuksissa kontrolliryhmä kävi tutkimuksen aikana koulun opetussuunnitelman mukaisen englannin kielen kurssin ilman mitään lisätukea. Celikin ja kumppaneiden (2025) tutkimuksessa myös koeryhmä kävi opetussuunnitelman mukaisen kurssin, mutta tämän lisäksi he harjoittelivat puhetaitoa kotona kielimalliagentin kanssa. Zhoun ja kumppaneiden (2025) tutkimuksessa kielimalliagenttia puolestaan hyödynnettiin kurssin ohessa keskustelukumppanina koehenkilöille. Fengin ja Wangin (2025) tutkimuksessa koeryhmässä kielimalliagenttia käytettiin vielä laajemmin kurssilla muun muassa arvioimisen tukena opettajalle, keskustelukumppanina ja luetun ymmärtämisen harjoittelun apuna koehenkilöille.

Kuudessa tutkimuksessa kielimalliagentin käyttö korvasi jonkin kontrolliryhmän saaman harjoittelun (Chen ym., 2025; Tai & Chen, 2024; C. Wang ym., 2024; Y. Wang, 2025; Yang ym., 2025; Zheng ym., 2025). Näistä kolmessa tutkimuksessa agentin käyttö korvasi luokkatoverin kanssa käytävän keskustelun, eli kontrolliryhmässä koehenkilöt kävivät keskusteluja luokkatovereidensa kanssa ja koeryhmässä he kävivät keskusteluja kielimalliagentin kanssa (Chen ym., 2025; Tai & Chen, 2024; Zheng ym., 2025). Kahdessa tutkimuksessa kielimalliagentti korvasi ihmisopettajan roolin (Y. Wang, 2025; Yang ym., 2025). Y. Wangin (2025) tutkimuksessa kontrolliryhmä sai kommunikaatioaidoistaan palautetta opettajalta, kun taas koeryhmässä koehenkilöt saivat palautetta kielimalliagentilta. Yangin ja kumppaneiden (2025) tutkimuksessa kontrolliryhmä toteutti "opettajan tukemaa

vuorottelevaa harjoittelustrategiaa" (*teacher-supported interleaved training strategy*) ja koeryhmä "tekoälyn tukemaa vuorottelevaa harjoittelustrategiaa" (*AI-supported interleaved training strategy*). Koeryhmä harjoitteli tekoälyn avulla vuorottelevasti kuullun ymmärtämistä ja puhetaitoa sekä luetun ymmärtämistä ja kirjoitustaitoa. Kontrolliryhmä puolestaan harjoitteli näitä opettajajohtoisesti. Yhdessä tutkimuksessa kielimalliagentti korvasi sekä luokkatoverin kanssa käytävän keskustelun että opettajan roolin (C. Wang ym., 2024). Heidän tutkimuksessaan kontrolliryhmä osallistui tavalliseen opetukseen, jossa he opettelivat kieltä pohjautuen oppikirjaan ja opettajan esitelmiin sekä harjoittelivat puhetaitoa luokkatovereidensa kanssa. Tässä tutkimuksessa kumpikaan koeryhmä ei osallistunut tavalliseen opetukseen, vaan harjoittelivat näitä asioita ainoastaan kielimalliagenttien avulla.

Yhdessäkin tutkimuksessa kielimalliagentin ohjausta tai agentin kanssa käytävää keskustelua ei verrattu natiivipuhujalta saatavaan ohjaukseen tai natiivin kanssa käytävään keskusteluun.

#### 4.4 Tuloksien yhteenveto

Viidessä tutkimuksessa löydettiin tilastollisesti merkitsevä positiivinen vaikutus kielimalliagentin käytöllä oppijoiden puhe- tai kommunikaatiotaitoon (Celik ym., 2025; Tai & Chen, 2024; Y. Wang, 2025; Yang ym., 2025; Zheng ym., 2025). Yhdessä tutkimuksessa raportoitiin sanallisesti, että ChatGPT:n käyttö paransi merkittävästi koehenkilöiden suullista kielitaitoa ja ChatGPT:n käytöllä oli tilastollisesti merkitsevä positiivinen vaikutus käytettyjen testien kokonaistuloksiin (Feng & Wang, 2025). Kahdessa tutkimuksessa ei löytynyt tilastollisesti merkitsevää eroa koe- ja kontrolliryhmien posttest-tuloksien välillä (Chen ym., 2025; C. Wang ym., 2024). Lisäksi yhdessä tutkimuksessa löydettiin sekä tilastollisesti merkitseviä että ei-merkitseviä eroja ryhmien posttest-tuloksien välillä (Zhou ym., 2025).

Vain kaksi tutkimusta raportoivat tuloksien efektikoot (C. Wang ym., 2024; Zhou ym., 2025). Muiden tilastollisesti merkitseviä tuloksia saaneiden tutkimusten efektikoot olivat kuitenkin laskettavissa ja käytin efektikokojen raportoimiseen Cohenin d:tä. Jacob Cohenin (1988) esittelemä efektikoon mittari Cohenin d on edelleen käytössä oleva mittari. Cohenin mukaan d:n arvo 0.2 tarkoittaa pientä efektiä, arvo 0.5 keskikokoista efektiä ja arvo 0.8 suurta efektiä.

Laskin efektikoot *psychometrica.de*-sivuston laskureilla (W. Lenhard & A. Lenhard, 2022). Kolmessa tutkimuksessa oli raportoitu pre- ja posttestien tuloksien keskiarvot sekä keskihajonnat (Celik ym., 2025; Tai & Chen, 2024; Zheng ym., 2025), joten näiden laskemiseen käytin sivuston laskuria "3. Effect size for mean differences of groups with unequal sample size

within a pre-post-control design". Celikin ja kumppaneiden (2025) tutkimuksen efektikoko oli 1,497. Tain ja Chenin (2024) tutkimuksessa I-Bot-ryhmän ja kontrolliryhmän välinen efektikoko oli 0,802 ja P-Bot-ryhmän ja kontrolliryhmän välinen efektikoko 1,589. Nämä kaikki efektikoot olivat suuria. Zhengin ja kumppaneiden tutkimuksessa (2025) efektikoko heti intervention jälkeen oli 0,471 ja kuukauden kuluttua interventtiosta 0,447. Nämä molemmat efektikoot ovat Cohenin mukaan pieniä, mutta lähellä keskikoisia efektejä.

Y. Wangin (2025) artikkelissa ei raportoitu kontrolliryhmän pretest-tuloksia, joten laskin tämän tutkimuksen efektikoon *psychometrica.de*-sivuston laskurilla "2. Comparison of groups with different sample size (Cohen's d, Hedges' g)". Tämän tutkimuksen efektikoko oli 5,224, joka on hyvin suuri.

Yangin ja kumppaneiden (2025) tutkimuksessa raportoitiin varianssianalyysin F-arvot, joten Cohenin d:n arvo laskettiin *psychometrica.de*-sivuston laskurilla "7. Calculation of effect sizes from ANOVAs with multiple groups, based on group means". Tutkimuksen efektikokojen suuruudet olivat seuraavat: ÄT:  $d = 0,662$ ; IS:  $d = 0,159$ ; IT:  $d = 0,164$  ja IO:  $d = 0,827$ . Nämä efektikoot olivat järjestyksessä keskikokoinen, hyvin pieni, hyvin pieni ja suuri.

Zhoun ja kumppaneiden (2025) tutkimuksessa efektikoot raportoitiin r-arvoina. Tämän tutkimuksen osien efektikoot ja p-arvot olivat seuraavat: KTM:  $r = 0,013$  ja  $p=0,920$ , KJS:  $r = 0,043$  ja  $p=0,733$ , PPS:  $r = 0,717$  ja  $p < 0,001$  sekä yhteispisteet:  $r = 0,316$  ja  $p = 0,015$ . Näiden osien efektikokojen suuruudet olivat järjestyksessä mitättömän pieni, mitättömän pieni, suuri sekä keskisuuri.

Keskimääräisesti tilastollisesti merkitseviä tuloksia saaneiden tutkimusten efektikoot olivat suuria.

## 5 Pohdinta

### 5.1 Tutkimusasetelmat

Olin miettinyt jo keväällä 2024 tekeväni tästä aiheesta kandidaatintutkielman psykologian alalla, mutta silloin tutkimusta ei ollut vielä tarpeeksi kiinnostavaan katsaukseen. Nämäkin kaikki tässä haussa löytyneet tutkimukset olivat julkaistu vuosina 2024 ja 2025, joten vasta nyt onnistuin löytämään tarpeeksi aihetta tutkineita tutkimuksia. Marraskuussa 2022 OpenAI:n julkaisema ChatGPT-3 tuli ensimmäisenä suurta kielimallia hyödyntävänä keskustelubottina valtavirran saataville. Uusi tekniikka yhdistettynä tuoreeseen tutkimusaiheeseen voi hyvin selittää sitä, miksi kaikissa yhdeksässä tarkemmassa analyysissä olleessa tutkimuksessa oli käytössä OpenAI:n GPT-kielimallit, vaikka kirjallisuushaun kautta oli tarkoitus löytää myös muita suuria kielimalleja, kuten Geminiä, hyödyntäneet tutkimukset.

Lisäksi suurin osa tämä katsauksen tutkimuksista on tehty Aasiassa ja tutkittavana kielenä on kaikissa tutkimuksissa ollut englanti. Hakulauseke ei varsinaisesti rajoittanut hakua vain englannin kieltä tutkiviin tutkimuksiin tai Aasiassa tehtyihin tutkimuksiin, joten vaikuttaisi siltä, että tutkimusta ei ole juuri tehty muualla tai muilla kielillä. Toki tutkimuksia lukiessani huomasin, että hyvin monessa tutkimuksessa mainitaan termi EFL, eli *English as a Foreign Language*. Myös esimerkiksi ranskan kielen opiskelusta voidaan käyttää vastaavaa englanninkielistä termiä, FFL, eli *French as a Foreign Language*. Tulevissa tutkimuksissa kannattaisikin selvittää mahdolliset vastaavat termit ja ottaa ne erikseen huomioon hakuvaiheessa.

Replika-sovellusta tutkineita tutkimuksia ei otettu tarkempaan vertailuun Replikan toimintaperiaatteen epäselvyyden vuoksi. Molemmat tutkimukset kuitenkin saivat tulokseksi, että Replikan käytöstä olisi perinteistä opetusta enemmän hyötyä englannin kielen puhetaitojen harjoittelussa.

GPT-kielimalleja hyödyntäviä keskustelubotteja tutkineet tutkimukset saivat keskimäärin suuria efektikokoja kielimalliagenttien kanssa harjoittelun vaikutuksesta koehenkilöiden puhe- ja kommunikaatiotaitoihin. Kaksi tutkimusta eivät kuitenkaan löytäneet tilastollisesti merkitsevää eroa koe- ja kontrolliryhmien välillä (Chen ym., 2025; C. Wang ym., 2024) sekä yksi tutkimus löysi sekä tilastollisesti merkitseviä että ei-merkitseviä tuloksia (Zhou ym., 2025). Näissä tutkimuksissa oli paljon asetelmallisia eroja, joten onkin tarpeellista vertailla tuloksia eri näkökulmista.

## 5.2 Muuttujat ja mittarit

Vain kahdessa tutkimuksessa käytettiin samaa mittaria kielitaidon testaamiseen (Feng & Wang, 2025; Zhou ym., 2025). Zhou ja kumppanit (2025) kuitenkin käyttivät vain *CET-4*-kokeen suullista osuutta ja Feng ja Wang (2025) käyttivät lisäksi myös *CET-6*-koetta sekä suoritettun englannin kielen kurssin arvosanaa. Fengin ja Wangin (2025) tutkimuksessa ei myöskään eritelty kokeiden eri osioiden tuloksia, joten näidenkään kahden tutkimuksen tuloksia ei ole mielekästä vertailla käytetyn mittarin perusteella. Kaikissa muissa tutkimuksissa oli käytössä eri mittarit kielitaidon mittaamiseen. C. Wang ja kumppanit (2024) sekä Chen ja kumppanit (2025) sovelsivat samaa testiä (*IELTS*) omissa mittareissaan, mutta heidän käyttämänsä mittarit olivat silti toisistaan erillisiä. Tämän vuoksi tutkimuksia ei ole mielekästä vertailla käytettyjen mittareiden perusteella. Koska mitattavien muuttujien, eli kielitaidon, termit olivat osittain päällekkäisiä ei ole mielekästä vertailla tutkimuksia näidenkään perusteella. Tutkimukset eivät myöskään määritelleet käytettyjä termejä kovinkaan tarkasti, joten näidenkään kautta tuloksien vertailu ei olisi mielekästä.

## 5.3 Puheentunnistus

Ensimmäisenä tarkastellaan puheentunnistusta, eli sitä, miten puhe muutettiin kielimalliagentille tekstimuotoon. Tämän tutkimuksen kannalta on kuitenkin tärkeintä, että puheentunnistus on sellaisella tasolla, että se muuntaa puheen oikein tekstiksi niin, että kirjoitettu teksti vastaa puhuttua. Missään tutkimuksessa ei mainittu, että tutkimuksessa käytetty puheentunnistus olisi vaikuttanut negatiivisesti tuloksiin, joten voidaan olettaa, että puheentunnistus toimi riittävällä tasolla, jolloin se ei vaikuttanut olennaisesti tutkimustuloksiin.

## 5.4 Vertailun kohde

Tutkimuksissa yksi merkittävimpiä tekijöitä kielimalliagentin vaikuttavuuteen on se, että mihin koeryhmän käyttämää kielimalliagenttia verrattiin kontrolliryhmässä. Tutkimuksessa on voitu verrata kielimalliagentin kanssa käytyä keskustelua paremmin kieltä osaavan ihmisen, kuten tuutorin, kanssa käytyyn keskusteluun. Tällöin tilastollisesti ei-merkitsevä tulos ei tarkoita, ettei kielimalliagentista ollut hyötyä. Tämä tarkoittaisi vain, että koeryhmän ja kontrolliryhmän välillä ei ollut tilastollisesti merkitseviä eroja, eli kielimalliagentin kanssa keskustelusta on ollut tutkimuksen mukaan yhtä paljon hyötyä kuin ihmisen kanssa keskustelusta. Tämä vastaisi johdannossa esitettyyn kysymykseen, että kielimalliagentit voisivat toimia kielten oppimisessa korvikkeena ihmisille, ja ne voisivat tuoda toimivan lisän kielten kommunikaation opiskeluun.

Niistä viidestä tutkimuksesta, joissa löydettiin tilastollisesti merkitsevä positiivinen vaikutus suurella efektikoolla, kielimalliagentti tuli yhdessä tutkimuksessa lisänä koeryhmälle (Celik ym., 2025), jolloin koeryhmä sai enemmän harjoitusta kuin kontrolliryhmä, mikä vaikuttaa tulokseen. Kuitenkin näistä tutkimuksista neljässä kielimalliagentti korvasi jonkin kontrolliryhmän saaman opetuksen osa-alueen (Tai & Chen, 2024; Y. Wang, 2025; Yang ym., 2025; Zheng ym., 2025). Kahdessa tutkimuksessa kielimalliagentti korvasi opettajan, joten näiden tutkimusten perusteella kielimalliagentti voisi olla jopa parempi antamaan palautetta kuin ihmisopettajat (Y. Wang, 2025; Yang ym., 2025). Kahdessa muussa tutkimuksessa kielimalliagentti korvasi luokkatoverin kanssa käydyn keskustelun (Tai & Chen, 2024; Zheng ym., 2025). Näiden tutkimusten perusteella kielimalliagenttien kanssa käytävä keskustelu voisi olla hyödyllisempää kuin vertaisten oppijoiden kanssa käyty keskustelu. Tämä tukisi aiempaa pohdintaa siitä, että kielimalliagentti voisi toimia edistyneemmän ohjaajan roolissa oppijalle. Näin oppija voisi kielimalliagentin avulla toimia lähikehityksen vyöhykkeellä ja saavuttaa parempia oppimistuloksia kuin samantasoisien henkilön kanssa harjoittellessa.

Myös Fengin ja Wangin (2025) tutkimuksessa löydettiin, että kielimalliagentti paransi merkittävästi koehenkilöiden suullisia kielitaitoja, vaikka tutkimuksessa ei erikseen mainittu tilastollisia tuloksia suullisista kielitaidoista. Kuten Celikin ja kumppaneiden (2025) tutkimuksessa, myös tässä tutkimuksessa kielimalliagentti tuli lisänä koeryhmälle, jolloin heidänkin tulokseensa on vaikuttanut koeryhmän saaman lisääntyneen harjoittelun määrä.

Kahdessa tutkimuksessa koe- ja kontrolliryhmien välillä ei ollut tilastollisesti merkitsevää eroa (Chen ym., 2025; C. Wang ym., 2024). Näissä molemmissa tutkimuksissa kielimalliagentti korvasi jonkin kontrolliryhmän saaman opetuksen osa-alueen. Chenin ja kumppaneiden (2025) tutkimuksessa kielimalliagentti korvasi luokkatovereiden kanssa käytävän keskustelun samoin kuin positiiviset tulokset saaneissa Zhengin ja kumppaneiden (2025) sekä Tain ja Chenin (2024) tutkimuksissa. Chenin ja kumppaneiden (2025) tutkimuksen tulokset poikkeavat näiden kahden tutkimuksen tuloksista, vaikka kielimalliagentin roolin ja kontrolliryhmän saaman opetuksen vertailu on samankaltaista. Tarvittaisiin vielä lisää tutkimusta tällä kielimalliagentin roolilla ja vertailun kohteella, jotta voitaisiin selvittää, mistä tällainen ero syntyi. C. Wangin ja kumppaneiden (2024) tutkimuksessa puolestaan koeryhmät eivät osallistuneet ollenkaan tavalliseen opetukseen. Missään muussa tutkimuksessa näin ei ollut, joten tämä voi olla yksi syy, miksi tässä tutkimuksessa kielimalliagentti tuotti yhtä hyviä tuloksia kuin tavallinen opetus. Voi olla, että tavallista opetusta parempiin tuloksiin vaaditaan kielimalliagentin kanssa käydyn opetuksen lisäksi myös tavallista opetusta. Tästäkin tarvittaisiin lisää tutkimusta.

Zhoun ja kumppaneiden (2025) tutkimuksessa puolestaan löydettiin sekä tilastollisesti merkitseviä että ei-merkitseviä tuloksia. Kuten Celikin ja kumppaneiden (2025) sekä Fengin ja Wangin (2025) tutkimuksissa, tässäkin tutkimuksessa kielimalliagentti tuli lisänä koeryhmälle, eikä kontrolliryhmä saanut mitään vastinetta. Toisin kuin näissä kahdessa tutkimuksessa, tässä tutkimuksessa löydettiin myös tilastollisesti ei-merkitseviä tuloksia. Kielellinen tarkkuus ja monimuotoisuus sekä kielen joustavuus ja sopivuus pysyivät samalla tasolla kontrolliryhmään verrattuna. Kuitenkin puheen pituus ja sujuvuus -osiossa koeryhmä sai kontrolliryhmää paremmat tulokset suurella efektikoolla. Tämän myötä myös näiden kolmen osion yhteistulokset olivat tilastollisesti merkitsevästi ja keskipituisella efektikoolla suuremmat koeryhmällä kuin kontrolliryhmällä. Tämä voi tarkoittaa sitä, että kielimalliagenteista olisi enemmän apua puhutun kielen tietyissä osa-alueissa kuin toisissa. Toisaalta Yangin ja kumppaneiden (2025) tutkimuksessa tutkittiin myös monia puhutun kielen osa-alueita erikseen, ja he löysivät kaikkien puhutun kielen osa-alueilla tilastollisesti merkitsevät tulokset, mutta näissäkin oli suuria eroja efektikoissa. Yangin ja kumppaneiden (2025) tutkimuksessa ilmaisun sujuvuus -osiossa efektikoko oli hyvin pieni, vaikka Zhoun ja kumppaneiden (2025) tutkimuksessa vastaavan osion efektikoko oli suuri. Toisaalta Yangin ja kumppaneiden (2025) tutkimuksessa kielimalliagentti korvasi opettajan tukeman harjoittelustrategian, joten tämäkin voi vaikuttaa tämän osa-alueen efektikokojen eroon. Yangin ja kumppaneiden (2025) tutkimuksessa myös esimerkiksi ilmaisun oikeellisuus -osiossa efektikoko oli suuri, kun taas Zhoun ja kumppaneiden (2025) tutkimuksessa kielen joustavuus ja sopivuus -osion tulokset eivät olleet tilastollisesti merkitsevästi eroavia koe- ja kontrolliryhmien välillä. Tarvittaisiin myös lisää tutkimusta kielimalliagentin vaikuttavuudesta puhe- ja kommunikaatiotaitojen eri osa-alueisiin.

## 5.5 Tutkitut kielimallit

Kielimalleista GPT-4-versiota käytettiin yhteensä neljässä tutkimuksessa (Chen ym., 2025; Y. Wang, 2025; Zheng ym., 2025; Zhou ym., 2025). Näistä Chen ja kumppanit (2025) sekä Zheng ja kumppanit (2025) käyttivät OpenAI:n API:ta ja Zhou ja kumppanit (2025) sekä Y. Wang (2025) käyttivät ChatGPT-4:ää. Kaikissa näissä neljässä tutkimuksessa saatiin positiivisia tuloksia ja efektikokoja. Zhoun ja kumppaneiden (2025) tutkimuksessa saatiin toki myös tilastollisesti ei-merkitseviä tuloksia.

GPT-3.5-versiota käytettiin yhteensä kahdessa tutkimuksessa (C. Wang ym., 2024; Yang ym., 2025). C. Wang ja kumppanit (2024) käyttivät OpenAI:n API:ta ja Yang ja kumppanit (2025)

ChatGPT-3.5:ttä. Yangin ja kumppaneiden (2025) sekä Zhoun ja kumppaneiden (2025) tutkimuksissa ilmenneet erot voivat johtua ainakin osittain kielimallin eri versioista. Lisäksi C. Wangin ja kumppaneiden (2024) tutkimus oli Celikin ja kumppaneiden (2025) tutkimuksen lisäksi ainoa, joka saanut ei löytänyt tilastollisesti merkitsevää tulosta. Celikin ja kumppaneiden (2025) tutkimuksessa ei raportoitu käytettyä ChatGPT:n versiota, joten tutkimusten ei-merkitsevien tuloksien löytyminen voi myös johtua GPT-mallin vanhemmasta versiosta. GPT-4 on kehittyneempi malli kuin GPT-3.5 ja OpenAI:n mukaan se on "luotettavampi, luovampi sekä kyvykkäämpi hallitsemaan monimutkaisempia ohjeita kuin GPT-3.5" (OpenAI, 2023, [oma suomennos]).

Celikin ja kumppaneiden (2025) lisäksi myöskään Feng ja Wang (2025) sekä Tai ja Chen (2024) eivät maininneet käyttämänsä GPT-mallin versiota. Molemmat näistä saivat kielimalliagentin käytöstä positiivisia tuloksia, joten olisi hyvä saada vielä lisää tutkimusta, joissa vertaillaan eri kielimalleja samassa tutkimuksessa.

## 5.6 Kielimalliagentin rooli

Niissä kahdessa tutkimuksessa, joissa kielimalliagentin roolina oli toimia opettajana, saatiin molemmissa tutkimuksissa positiivisia tuloksia (Y. Wang, 2025; Yang ym., 2025). Myös Fengin ja Wangin (2025) tutkimuksessa, jossa kielimalliagentti toimi sekä tuutorina että opettajana, saatiin positiivisia tuloksia. Näiden kahden tutkimuksen perusteella kielimalliagentista on hyötyä puhe- ja kommunikaatiotaitojen oppimisessa, kun se toimii opettajan roolissa.

Niissä kuudessa tutkimuksessa, joissa kielimalliagentin rooli oli toimia kuten tuutorina koehenkilölle, tulokset olivat ristiriitaisia. Kolmessa tutkimuksessa saatiin positiivisia tuloksia (Celik ym., 2025; Tai & Chen, 2024; Zheng ym., 2025), kahdessa tutkimuksessa ei saatu tilastollisesti merkitseviä tuloksia (Chen ym., 2025; C. Wang ym., 2024) ja Zhoun ja kumppaneiden (2025) tutkimuksessa saatiin sekä positiivisia että tilastollisesti ei-merkitseviä tuloksia. Kuitenkaan tämän tutkielman perusteella ei voi vielä arvioida, että kuinka paljon kielimalliagentin rooli vaikuttaa sen hyödyllisyyteen. Kuuden tutkimuksen tulosten ristiriitaisuus saattaa viitata siihen, että tutkimustulosten erot eivät johdu suurelta osin kielimalliagentin roolista, vaan siihen vaikuttaa enemmän muut tutkimusasetelmalliset tekijät. Olisikin hyvä saada lisää tutkimusta, jossa verrattaisiin, mihin kielimalliagenttia käytetään, ja millaista roolia kielimalliagentti edustaa muuten samanlaisilla koeasetelmilla.

## 6 Yhteenveto

Tämän tutkielman tarkoituksena oli selvittää, voisiko kielimalliagenteista olla hyötyä vieraan kielen puhe- ja kommunikaatiotaitojen harjoittelussa sekä, miten kielimalliagenttien avulla harjoittelu vertautuu perinteisiin oppimismenetelmiin. Tutkielmassa esiteltiin suuriin kielimalleihin, keskustelubotteihin ja yleisesti tekoölyyn liittyvät peruskäsitteet sekä kielen oppimisen psykologisia mekanismeja. Tutkimuksen alussa tuotiin esille, että kielimalliagenteista voisi olla hyötyä kielten oppimisessa, kun ne voisivat toimia edistyneemmän ohjaajan roolissa oppijalle, jolloin hän voisi kehittyä lähikehityksen vyöhykkeellä. Tällöin kielten puhe- ja kommunikaatiotaitojen oppimisesta tulisi saavutettavampaa, kun suuremmalle osalle ihmisistä näiden taitojen päivittäinen tehokas harjoittelu olisi mahdollista.

Tutkimus toteutettiin systemaattisena kirjallisuuskatsauksena. Kirjallisuuskatsauksesta saatiin yhteensä 1657 tulosta, joista lopulliseen tarkasteluun päätyi 11 artikkelia. Katsaukseen otettiin mukaan vain kokeellisia tai kvasikokeellisia, vertaisarvioituja tutkimuksia. Kahdessa tutkimuksessa tutkittiin Replika-sovelluksen käyttöä, mutta niitä ei otettu tarkempaan analyysiin, sillä ei saatu selville, mikä teknologia Replikan taustalla on. Muissa yhdeksässä tutkimuksessa tutkittiin OpenAI:n GPT-kielimalleja hyödyntäviä keskustelubotteja.

Tarkemmin vertailtujen tutkimusten tutkimusasetelmat erosivat toisistaan monella tavalla, kuten mitä GPT-kielimallin versiota käytettiin, oliko tutkimuksissa käytössä ChatGPT vai jokin muu sovellus, mihin kielimalliagenttia verrattiin, mikä oli kielimalliagentin rooli puhe- ja kommunikaatiotaitojen harjoittelussa sekä mitä mittareita kielitaidon mittaamiseen käytettiin. Tutkimustulokset olivat hieman ristiriitaisia, mutta pääosin positiivisia. Kahdessa tutkimuksessa ei saatu tilastollisesti merkitseviä tuloksia, yhdessä tutkimuksessa saatiin sekä tilastollisesti merkitseviä että ei-merkitseviä tuloksia ja viidessä tutkimuksessa saatiin tilastollisesti merkitseviä, keskimäärin suurella efektikoolla olevia tuloksia. Yhdessä tutkimuksessa kerrottiin sanallisesti, että kielimalliagentin käyttö paransi merkittävästi koehenkilöiden suullisia kielitaitoja. Kaikissa tutkimuksissa tutkittiin englantia vieraana kielenä.

Tämä tutkielma antaakin alustavaa näyttöä sille, että kielimalliagenteista olisi hyötyä ainakin englannin kielen puhe- ja kommunikaatiotaitojen oppimisessa, ja niitä voitaisiin hyödyntää sekä perinteisten oppimismenetelmien tukena että jonkin perinteisen oppimismenetelmän osa-

alueen korvikkeena. Tutkimusten perusteella kielimalliagenttien hyödyt olisivat vähintään samantasoiset verrattuna perinteisillä oppimismenetelmillä saatuihin hyötyihin puhe- ja kommunikaatiotaitojen oppimisessa. Tarvittaisiin kuitenkin lisää tutkimusta aiheesta, jotta voitaisiin selvittää tarkemmin, mikä rooli kielimalliagentilla tulisi olla puhe- ja kommunikaatiotaitojen harjoittelussa, ja missä tilanteessa kielimalliagenttia kannattaisi hyödyntää. Lisäksi olisi tärkeää tutkia kielimalliagenttien käyttöä myös muiden kuin englannin kielen puhe- ja kommunikaatiotaitojen harjoittelussa.

## Lähteet

- Abdullah, M., Madain, A., & Jararweh, Y. (2022). ChatGPT: Fundamentals, Applications and Social Impacts. *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 1–8.  
<https://doi.org/10.1109/SNAMS58071.2022.10062688>
- Belcic, I., & Stryker, C. (2024, syyskuuta 18). *What is GPT (generative pre-trained transformer)?* | IBM. <https://www.ibm.com/think/topics/gpt>
- Belcic, I., & Stryker, C. (2025, huhtikuuta 19). *What is ChatGPT?* | IBM.  
<https://www.ibm.com/think/topics/chatgpt>
- Bergmann, D., & Stryker, C. (2025, maaliskuuta 28). *What is a Transformer Model?* | IBM.  
<https://www.ibm.com/think/topics/transformer-model>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv.  
<https://doi.org/10.48550/arXiv.2005.14165>
- Caluori, L. (2024). Hey Alexa, why are you called intelligent? An empirical investigation on definitions of AI. *AI & SOCIETY*, 39(4), 1905–1919. <https://doi.org/10.1007/s00146-023-01643-y>
- Çakmak, F. (2022). Chatbot-Human Interaction and Its Effects on EFL Students' L2 Speaking Performance and Anxiety. *Novitas-ROYAL (Research on Youth and Language)*, 16(2), 113–131.
- Celik, B., Yildiz, Y., & Kara, S. (2025). Using ChatGPT as a virtual speaking tutor to boost EFL learners' speaking self-efficacy. *Australian Journal of Applied Linguistics*, 8(1), 102418. <https://doi.org/10.29140/ajal.v8n1.102418>

- Chen, A., Jia, J., Li, Y., & Fu, L. (2025). Investigating the Effect of Role-Play Activity With GenAI Agent on EFL Students' Speaking Performance. *Journal of Educational Computing Research*, 63(1), 99–125. <https://doi.org/10.1177/07356331241299058>
- Cohen Ventura, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Cumming, G. (2011). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Taylor & Francis Group.  
<http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=957018>
- Feng, Y., & Wang, X. (2025). Exploring the Development of Chinese College Students' Proficiency in English through ChatGPT: An Experimental Study. *Proceedings of the 2024 16th International Conference on Education Technology and Computers*, 148–154. <https://doi.org/10.1145/3702163.3702184>
- Foot, K. D. (2024, maaliskuuta 5). A Brief History of Generative AI. *DATAVERSITY*.  
<https://www.dataversity.net/a-brief-history-of-generative-ai/>
- França, R. P., Borges Monteiro, A. C., Arthur, R., & Iano, Y. (2021). Chapter 3—An overview of deep learning in big data, image, and signal processing in the modern digital age. Teoksessa V. Piuri, S. Raj, A. Genovese, & R. Srivastava (Toim.), *Trends in Deep Learning Methodologies* (ss. 63–87). Academic Press.  
<https://doi.org/10.1016/B978-0-12-822226-3.00003-9>
- Fulcher, G. (2014). *Testing Second Language Speaking*. Routledge.  
<https://doi.org/10.4324/9781315837376>
- Hana, V. (2023). AI and AI-powered tools for pronunciation training. *Journal of Language and Cultural Education*, 11(3), 12–24. <https://doi.org/10.2478/jolace-2023-0022>
- Holdsworth, J., & Scapicchio, M. (2024, kesäkuuta 17). *What Is Deep Learning?* | IBM.  
<https://www.ibm.com/think/topics/deep-learning>

- Kundu, A., & Bej, T. (2025). Transforming EFL Teaching with AI: A Systematic Review of Empirical Studies. *International Journal of Artificial Intelligence in Education*, 1–34. <https://doi.org/10.1007/s40593-025-00470-0>
- Lenhard, W., & Lenhard, A. (2022). *Computation of effect sizes*. Psychometrica. [https://www.psychometrica.de/effect\\_size.html](https://www.psychometrica.de/effect_size.html)
- Luka, Inc. (ei pvm.). *The AI companion who cares*. Replika. Haettu 22.12.2025 osoitteesta <https://replika.com/>
- Lyu, B., Lai, C., & Guo, J. (2025). Effectiveness of Chatbots in Improving Language Learning: A Meta-Analysis of Comparative Studies. *International Journal of Applied Linguistics*, 35(2), 834–851. <https://doi.org/10.1111/ijal.12668>
- Mohanasundaram, R., Malhotra, A. S., Arun, R., & Periasamy, P. S. (2019). Chapter 8—Deep Learning and Semi-Supervised and Transfer Learning Algorithms for Medical Imaging. Teoksessa A. K. Sangaiah (Toim.), *Deep Learning and Parallel Computing Environment for Bioengineering Systems* (s. 139–151). Academic Press. <https://doi.org/10.1016/B978-0-12-816718-2.00015-4>
- OpenAI. (ei pvm.). *API Platform*. Noudettu 17. elokuuta 2025, osoitteesta <https://openai.com/api/>
- OpenAI. (2022, marraskuuta 30). *Introducing ChatGPT*. <https://openai.com/index/chatgpt/>
- OpenAI. (2023). *GPT-4*. <https://openai.com/index/gpt-4-research/>
- OpenAI. (2025, elokuuta 15). *How to change your language setting in ChatGPT*. OpenAI Help Center. <https://help.openai.com/en/articles/8357869-how-to-change-your-language-setting-in-chatgpt>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan — a web and mobile app for systematic reviews. *Systematic Reviews*. 5:210. <https://doi.org/10.1186/s13643-016-0384-4>

- Qin, Q., & Zhang, S. (2025). Visualizing the knowledge mapping of artificial intelligence in education: A systematic review. *Education and Information Technologies*, 30(1), 449–483. <https://doi.org/10.1007/s10639-024-13076-1>
- Randall, M. (2007). *Memory, Psychology and Second Language Learning*. John Benjamins Publishing Company.  
<http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=622830>
- Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence: A modern approach* (3rd ed). Prentice Hall.
- Şahin Kızıl, A., Klimova, B., Pikhart, M., & Parmaxi, A. (2025). A Systematic Review of the Recent Research on the Usefulness of Chatbots for Language Education. *Journal of Computer Assisted Learning*, 41(2), e70001. <https://doi.org/10.1111/jcal.70001>
- Schäferhoff, N. (2024, heinäkuuta 5). The History of Google Translate (2004-Today): A Detailed Analysis. *TranslatePress*. <https://translatepress.com/history-of-google-translate/>
- Shaalán, IEAW. (2025). The Impact of Artificial Intelligence and Machine Learning on Linguistic Accuracy, Fluency, and Self-Direction Among Advanced EFL Students. *Theory & Practice in Language Studies (TPLS)*, 15(6).  
<https://doi.org/10.17507/tpls.1506.27>
- Stryker, C. (2023, marraskuuta 2). *What Are Large Language Models (LLMs)? | IBM*.  
<https://www.ibm.com/think/topics/large-language-models>
- Stryker, C., & Holdsworth, J. (2024, elokuuta 11). *What Is NLP (Natural Language Processing)? | IBM*. <https://www.ibm.com/think/topics/natural-language-processing>
- Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, 220, 105112. <https://doi.org/10.1016/j.compedu.2024.105112>

- Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System*, *127*, 103528.  
<https://doi.org/10.1016/j.system.2024.103528>
- Nyrhinen, S. (2019). 6 Tekstianalytiikka. Teoksessa Tuominen, H., Neittaanmäki, P., Niinimäki, E., Pölönen, I., Rautiainen, I., Äyrämö, S., . . . Äyrämö, S. (2019). *Tekoälyn perusteita ja sovelluksia* (s.64–69).  
<https://urn.fi/URN:ISBN:978-951-39-7796-2>
- Veneziano, E. (2010). *Conversation in language development and use: An Introduction*.  
<https://journals.sagepub.com/doi/abs/10.1177/0142723710380531>
- Vygotski, L. S. (with Cole, M.). (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, *127*, 103533. <https://doi.org/10.1016/j.system.2024.103533>
- Wang, Y. (2025). A Study on the Efficacy of ChatGPT-4 in Enhancing Students' English Communication Skills. *Sage Open*, *15*(1).  
<https://doi.org/10.1177/21582440241310644>
- Wanli, Z., Youjun, T., & Xiaomei, M. (2025). From Learning Science to Computer Science: A Scientometric Review of Deeper Learning in Foreign Languages (1993–2024). *SAGE Open*, *15*(1), 21582440251322564.  
<https://doi.org/10.1177/21582440251322564>
- Wu, T. T., Hapsari, I. P., & Huang, Y. M. (2025). Effects of incorporating AI chatbots into think–pair–share activities on EFL speaking anxiety, language enjoyment, and

speaking performance. *Computer Assisted Language Learning*, 1-39.

<https://doi.org/10.1080/09588221.2025.2478271>

Yahaya, R. S., Madzlan, N. A., & Muhammad, M. (2021). ESL Learners' Obstacles in Learning Speaking Skills. *The English Teacher*, 50(3), 173–181.

<https://doi.org/10.52696/WYZY7207>

Yang, G., Wang, Y., Zhang, Y., Yang, M., Zeng, Q., & Song, Z. (2025). An Empirical Study of AI-Supported Interleaved Training Strategy to Improve EFL Students' English Impromptu Speaking Performance, Learning Engagement, Technology Acceptance and Epistemic Network Structure. *The Asia-Pacific Education Researcher*, 34(4), 1519–1540. <https://doi.org/10.1007/s40299-024-00962-y>

Zhang, Z. (2018). Artificial Neural Network. *Teoksessa Multivariate Time Series Analysis in Climate and Environmental Research* (ss. 1–35). Springer, Cham.

[https://doi.org/10.1007/978-3-319-67340-0\\_1](https://doi.org/10.1007/978-3-319-67340-0_1)

Zheng, Y.-B., Zhou, Y.-X., Chen, X.-D., & Ye, X.-D. (2025). The influence of large language models as collaborative dialogue partners on EFL English oral proficiency and foreign language anxiety. *Computer Assisted Language Learning*, 0(0), 1–27.

<https://doi.org/10.1080/09588221.2025.2453191>

Zhou, Q., Hashim, H., & Sulaiman, N. A. (2025). Supporting english speaking practice in higher education: The impact of AI chatbot-integrated mobile-assisted blended learning framework. *Education and Information Technologies*, 30(10), 14629–14660.

<https://doi.org/10.1007/s10639-025-13401-2>