

Hannemari Kuusisto

# TEKOÄLYN VAIKUTUS KÄSITYKSIIN VUOROVAIKUTUSKUMPPANISTA

Informaatioteknologian ja viestinnän tiedekunta  
Kandidaattitutkielma  
Kesäkuu 2022

# TIIVISTELMÄ

Hannemari Kuusisto: Tekoälyn vaikutus käsityksiin vuorovaikutuskumppanista  
Kandidaattitutkielma  
Tampereen yliopisto  
Tietojenkäsittelytieteiden tutkinto-ohjelma  
Kesäkuu 2022

---

Suuri osa ihmisten välisestä vuorovaikutuksesta tapahtuu nykyään teknologian välityksellä. Teknologiavälitteistä vuorovaikutusta on jo vuosia pyritty nopeuttamaan ja tehostamaan tekoälyn avulla. Tekoälyä hyödynnetään ihmisten välisessä vuorovaikutuksessa tänä päivänä jo monin tavoin, esimerkiksi sähköpostisovellusten älykkäissä vastausehdotuksissa. Tekoälyteknologiat kehittyvät jatkuvasti ja tulevaisuudessa tekoälyä voidaan mahdollisesti hyödyntää laajemminkin esimerkiksi automaattiseen profiilinluontiin. Tekoälyavusteisesta vuorovaikutuksesta on kyse silloin, kun tekoäly muokkaa tai täydentää ihmisen kirjoittamia viestejä tai tuottaa kokonaan uusia viestejä lähettäjän puolesta ihmisten välisessä keskustelussa. Tekoälyn osallisuuden laajuus voi vaihdella pienten kirjoitusvirheiden korjaamisesta kokonaisten tekstikappaleiden tuottoon. Tekoälyä voidaan hyödyntää eri medioissa kuten video-, ääni- tai tekstimuotoisessa vuorovaikutuksessa, ja se voi toimia joko viestin lähettäjän tai vastaanottajan apuna.

Tässä tutkielmassa keskitytään tekstipohjaiseen tekoälyavusteiseen vuorovaikutukseen ja selvitetään kirjallisuuskatsauksen avulla, millaisia vaikutuksia tekoälyllä on käsityksiin vuorovaikutuskumppanista. Tarkastellaan siis tilanteita, joissa viestin laatija eli lähettäjä käyttää kirjoittamisen apuna tekoälyä, ja tutkitaan, millainen vaikutus sillä on vastaanottajan käsityksiin lähettäjistä. Kirjallisuuskatsauksessa on mukana seitsemän vuosina 2019–2022 julkaistua tutkimusta, joissa käsitellään automaattista profiilinluontia, älykkäitä vastausehdotuksia ja tekoälyä sähköpostiviestin kirjoittamisen apuna. Tutkimuksissa selvitetään tekoälyn vaikutusta vastaanottajan käsityksiin lähettäjän luotettavuudesta, viehättävyydestä, sosiaalisesta vetovoimasta, lämmöstä sekä pätevydestä.

Tulosten perusteella tekoälyllä on enemmän negatiivisia kuin positiivisia vaikutuksia viestin vastaanottajan käsityksiin lähettäjistä. Erityisesti viestin lähettäjän koettuun luotettavuuteen tekoälyn käyttö näyttäisi vaikuttavan yleensä negatiivisesti. Mielipiteisiin lähettäjän viehättävyydestä tekoälyllä ei ollut vaikutusta, mutta käsityksiin tämän sosiaalisesta vetovoimasta sillä oli. Koetun lämmön ja pätevyden suhteen tulokset olivat ristiriitaisia. Tutkimusten tuloksien vertailu osoittautui hankalaksi, koska käsityksien mittaamiseen käytettiin erilaisia mittareita.

Avainsanat: tekoäly, vuorovaikutus, tekoälyavusteinen vuorovaikutus, teknologiavälitteinen vuorovaikutus

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

# SISÄLLYSLUETTELO

<b>1</b>	<b>Johdanto .....</b>	<b>1</b>
<b>2</b>	<b>Tekoälyavusteinen vuorovaikutus .....</b>	<b>2</b>
<b>3</b>	<b>Tutkimusmenetelmä .....</b>	<b>5</b>
<b>4</b>	<b>Tekoälyn vaikutus käsityksiin vuorovaikutuskumppanista .....</b>	<b>7</b>
4.1	Tutkimusten tarkastelukohteita	7
4.2	Tekoälyn vaikutus koettuun luotettavuuteen	10
4.3	Tekoälyn vaikutus käsityksiin muista ominaisuuksista	12
<b>5</b>	<b>Keskustelu .....</b>	<b>13</b>
<b>6</b>	<b>Yhteenveto .....</b>	<b>16</b>
	<b>Lähdeluettelo.....</b>	<b>17</b>

## 1 Johdanto

Ihmisten välinen vuorovaikutus tapahtuu yhä enemmän ja enemmän teknologian välityksellä. Tätä kutsutaan nimellä teknologiavälitteinen vuorovaikutus (Computer-Mediated Communication) ja se voi tapahtua tietokoneen, älypuhelimien tai jonkin muun teknologian välityksellä video-, ääni- tai tekstimuodossa (Carr, 2020). Tekoälyn kehittymisen ja yleistymisen myötä teknologiavälitteisen vuorovaikutuksen yhdeksi osaluokaksi on muodostunut tekoälyavusteinen vuorovaikutus (Artificial Intelligence-Mediated Communication).

Tekoälyavusteinen vuorovaikutus onnistuu luonnollisen kielen käsittelyn eli NLP:n (Natural Language Processing) avulla. Tietokone tarvitsee NLP-tekniikoita, jotta se voi viestiä ihmisten kanssa ja oppia ihmisten kirjoittamasta tekstistä (Russell & Norvig, 2022). Yksi tekoälyavusteisessa vuorovaikutuksessa yleisesti käytetty NLP-tekniikka on kielimallien hyödyntäminen. Kielimallit opetetaan tekstiaineistojen avulla. Ne ovat kehittyneet teknisesti kyvykkäiksi, mutta ne myös heijastavat opetusaineistoissa esiintyviä vääristymiä (Zhang ja muut, 2022).

Tekoälyavusteisessa vuorovaikutuksessa tekoäly voi muokata, täydentää tai tuottaa viestejä ihmisen puolesta (Hancock ja muut, 2020). Yksi esimerkki tästä on joissain sähköpostisovelluksissa ja pikaviestimissä käytössä olevat älykkäät vastausehdotukset. Esimerkiksi Gmailissa niitä on hyödynnetty jo useita vuosia, ja jo vuonna 2016 10 % sähköpostiviesteistä kirjoitettiin älykkäiden vastausehdotusten avulla (Kannan ja muut, 2016).

Tekoälyn ja ihmisen välinen vuorovaikutus rajataan pois tekoälyavusteisen vuorovaikutuksen määritelmästä, koska siinä botti tai muu älykäs virtuaaliagentti ei edusta toista ihmistä (Hancock ja muut, 2020). Tässä tutkielmassa ei siis tarkastella nettipalveluissa yleistyneitä chatbotteja. Tekoäly voi olla ihmisten välisessä vuorovaikutuksessa apuna myös muilla tavoin, esimerkiksi ehdottamalla keskusteluaiheita, mutta mukaan on otettu vain tutkimukset, joissa tekoälyä hyödynnetään muokkaamaan tai tuottamaan tekstiä ihmisen puolesta.

Tekoäly tuo teknologiavälitteiseen vuorovaikutukseen uuden ulottuvuuden, jonka vaikutuksia on syytä tutkia. Tekoälyavusteisen vuorovaikutuksen yleistymässä on tärkeää ymmärtää, miten se vaikuttaa käsityksiin vuorovaikutuskumppanista. Tekoälyn osallisuudella itsessään voi olla ennalta-arvaamattomia vaikutuksia vuorovaikutukseen. Lisäksi tekoälyn tuottama kieli voi erota ihmisten tuottamasta kielestä, mikä voi myös vaikuttaa käsityksiin vuorovaikutuskumppanista.

Käsityksien muodostumista on tutkittu teknologiavälitteisen vuorovaikutuksen osalta. Teknologiavälitteisessä vuorovaikutuksessa vuorovaikutuskumppanista saadaan vähemmän informaatiota kuin kasvotusten tapahtuvassa vuorovaikutuksessa. Tämä on

eritoten totta tekstipohjaisessa vuorovaikutuksessa. Tämä johtaa siihen, että teknologiavälitteisessä vuorovaikutuksessa viestin vastaanottajat usein luovat liioiteltuja käsityksiä vuorovaikutuskumppanista vähäisen informaation perusteella. (Walther & Tong, 2014)

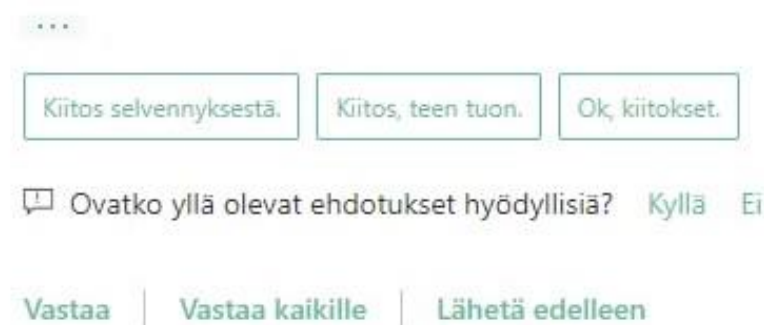
Tämän tutkielman tavoitteena on selvittää kirjallisuuskatsauksen avulla, miten tekoälyn osallisuus vaikuttaa käsityksiin vuorovaikutuskumppanista. Keskitytään selvittämään, miten tekoälyn käyttö lähettäjän puolella vaikuttaa vastaanottajan käsityksiin lähettäjästä. Kirjallisuuskatsauksessa on etsitty uusia enintään viisi vuotta vanhoja tutkimuksia, joissa tekoälyä on tutkittu tekstipohjaisessa vuorovaikutuksessa. Löydetyissä lähdeartikkeleissa tutkittiin tekoälyn vaikutusta käsityksiin luotettavuudesta, viehättävyydestä, sosiaalisesta vetovoimasta, lämmöstä sekä pätevydestä. Tekoäly vaikuttaa ainakin koettuun luotettavuuteen ja lähes kaikissa lähdeartikkeleissa sillä havaittiin olevan negatiivinen vaikutus. Vain yhdessä lähdeartikkelissa havaittiin positiivinen vaikutus. Vaikutuksen sosiaaliseen vetovoimaan havaittiin olevan negatiivinen. Vaikutusta lähettäjän viehättävyyteen ei havaittu ja lämmön sekä pätevyden suhteen tulokset olivat ristiriitaisia.

Tutkielman luvussa 2 avataan tekoälyavusteisen vuorovaikutuksen määritelmää sekä sen nykyisiä ja tulevia sovelluskohteita. Käydään myös läpi tekoälyavusteisen vuorovaikutuksen ulottuvuudet ja miten ne soveltuvat löytämiini lähdeartikkeleihin. Tämän jälkeen on helpompi ymmärtää termejä, joita käytetään myöhemmissä luvuissa. Luvussa 3 kuvataan tutkimusmenetelmää ja avataan hieman, miten löydetyissä tutkimuksissa on tutkittu tekoälyn vaikutuksia. Luku 4 on jaettu kolmeen osaan. Ensimmäinen kerrotaan lähdeartikkeleiden valitsemista tarkastelukohteista. Sen jälkeen käydään läpi tuloksia koetusta luotettavuudesta. Lopuksi käydään läpi muihin ominaisuuksiin liittyvät tulokset. Luvussa 5 pohditaan tuloksia ja niiden merkityksiä sekä tekoälyavusteisen vuorovaikutuksen määritelmää. Lopuksi luvussa 6 tehdään yhteenveto tutkimuksen tuloksista.

## **2 Tekoälyavusteinen vuorovaikutus**

Tekoälyavusteinen vuorovaikutus on ihmisten välistä vuorovaikutusta, jossa älykäs agentti voi muokata, täydentää tai tuottaa ihmisen puolesta viestejä vuorovaikutustavoitteiden saavuttamiseksi (Hancock ja muut, 2020). Vuorovaikutus voi tapahtua kahden ihmisen välillä tai yhdeltä ihmiseltä monelle (Jakesch ja muut, 2019). Tavoitteet, joita vuorovaikutuksella pyritään saavuttamaan voivat olla monenlaisia, mutta yleisiä tavoitteita ovat esimerkiksi pyrkimys näyttää viehättävältä, luotettavalta, miellyttävältä tai pätevältä (Hancock ja muut, 2020). Tekoälyä voidaan mahdollisesti hyödyntää näiden tavoitteiden saavuttamisessa, mutta tekoälyn osallisuudella itsellään voi olla ennalta-arvaamattomia vaikutuksia tavoitteiden saavuttamiseen.

Tekoälyä käytetään esimerkiksi älykkäiden vastaustenehdotusten tuottamiseen, automaattiseen tekstintäydennykseen ja automaattisiin vastauksiin. Yksi jo laajassa käytössä oleva tekoälyavusteisen vuorovaikutuksen tyyppi on älykkäät vastausehdotukset (Smart Reply). Monet pikaviestimet sekä Gmail ja Outlook hyödyntävät tekoälyä tuottamaan automaattisesti lyhyitä vastausehdotuksia. Käyttäjälle tarjotaan kolme eri vastausvaihtoehtoa, joista voi valita yhden; kuvassa 1 on esimerkki Outlookin vastausehdotuksista. Vastausta pystyy halutessaan vielä muokkaamaan ennen lähetystä. Vastausehdotuksien avulla vastaaminen onnistuu parilla napautuksella, mikä säästää aikaa etenkin mobiilikäytössä. (Microsoft, 2022; Kannan ja muut, 2016)



Kuva 1. Esimerkki Outlookin vastausehdotuksista. Kuvakaappaus suomenkielisen Outlookin selainversiosta. (12.4.2022)

Tekoälyjärjestelmä tuottaa vastausehdotukset vastaanotetun viestin perusteella. Se käsittelee vastaanotetun viestin ja valikoi siihen todennäköisimmin sopivat vastaukset. Järjestelmä hyödyntää kielimallia, joka on koulutettu aidoilla ihmisten kirjoittamilla viesteillä. Sen tuottamat ehdotukset karsitaan niin, että käyttäjälle näytetään monipuolinen joukko vastausehdotuksia, joiden merkityksissä ei ole päällekkäisyyksiä ja jotka eivät sisällä kielellisiä virheitä. Vastausehdotukset luodaan vain, jos viestiin on mahdollista vastata lyhyesti ja vastaaminen on tarpeellista ylipäätään. (Kannan ja muut, 2016)

Hancockin ja muiden (2020) mukaan tulevaisuudessa tekoälyä voidaan hyödyntää myös kokonaisten nettiprofiilien luomiseen. Tekoäly voisi kerätä tarvittavan tiedon käyttäjän muusta sosiaalisesta mediasta. Automaattinen profiilinluonti ei vielä ole käytössä, joten löytämissäni tutkimuksissa käytettiin ihmisten kirjoittamia profiileja.

Tekoälyavusteista vuorovaikutusta on hyödyllistä tarkastella Hancockin ja muiden (2020) määrittelemien ulottuvuuksien avulla. Heidän mukaansa tekoälyavusteisella vuorovaikutuksella on viisi ulottuvuutta, jotka ovat laajuus (magnitude), mediatyyppi (media type), optimoinnin tavoite (optimization goal), autonomia (autonomy) ja roolin suuntautuminen (role orientation). Ulottuvuuksien määritelmät löytyvät taulukosta 1. Tutkielmaan valittuja lähteitä tarkastellaan niiden avulla taulukon jälkeen.

Taulukko 1. Tekoölyavusteisen vuorovaikutuksen ulottuvuudet (taulukko muokattu lähteestä (Hancock ja muut, 2020)).

<b>Ulottuvuus</b>	<b>Määritelmä</b>	<b>Esimerkki</b>
Laajuus	Tekoölyn viestiin tekemien muutosten laajuus	Kirjoitusvirheiden korjaus vs. täysin uuden viestin luominen
Mediatyyppi	Media, jossa tekoöly toimii (esim. teksti, ääni, video)	Tekstivastausten ehdottaminen vs. ulkomuodon muokkaaminen videolla
Optimoinnin tavoite	Tavoite, jota varten tekoöly optimoi viestejä	Näyttääkseen viehättävältä, luotettavalta, humoristiselta, hallitsevalta jne.
Autonomia	Missä määrin tekoöly voi toimia viesteissä ilman lähettäjän valvontaa	Lähettäjä valitsee tekoölyn ehdottamien viestien väliltä vs. tekoöly hoitaa suurimman osan viestinnästä
Roolin suuntautuminen	Rooli, jonka puolesta tekoöly toimii (esim. lähettäjä tai vastaanottaja)	Lähettäjä: tekoöly ehdottaa viestejä vastaustehokkuuden parantamiseksi vs. vastaanottaja: tekoöly arvioi vastaanotetun viestin perusteella esimerkiksi valehteleeko lähettäjä

Tässä tutkielmassa tarkasteltavissa tutkimuksissa mediatyyppi on kaikissa teksti ja roolin suuntautuminen lähettäjä. Aihe on rajattu tekstipohjaisen vuorovaikutuksen tarkasteluun, koska luonnollisen kielen käsittely on teknisesti hyvin kehittynyttä ja se on jo yleisessä käytössä. Tutkimuksissa tekoöly toimii lähettäjän puolesta ja lähettäjän ominaisuuksia arvioiva henkilö on vastaanottajan roolissa.

Tutkimuksissa tekoölyn tekemien muutosten laajuus on enimmäkseen suuri. Vastausehdotusten ja profiilinluonnin kohdalla tekoöly tuottaa kokonaan uuden tekstin. Laajuudessa on silti tässäkin isot erot: profiilien kohdalla tekstin pituus on huomattavasti pidempi kuin vastausehdotusten, ja lähettäjä voi vielä täydentää viestiä vastausehdotuksen valinnan jälkeen. Yhdenkin tutkimuksen sisällä laajuus voi vaihdella eri testiryhmillä. Esimerkiksi Liun ja muiden (2022) tutkimuksessa sähköposti oli merkitty joko älykkään automaattitäydennyksen avulla kirjoitetuksi, jolloin laajuus olisi pienempi, tai täysin kehittyneen tekoölyjärjestelmän tuottamaksi, jolloin laajuus vastaa automaattista profiilinluontia. Weussin ja muiden (2022) tutkimuksessa tekoöly oli myös sähköpostiviestin kirjoittamisen apuna, mutta sen tekemien muutosten laajuutta ei tiedetä.

Vastausehdotusten tapauksessa tekoölyn autonomia ei ole kovinkaan suuri, koska lähettäjä voi valita käyttääkö vastausehdotuksia ja mahdollisesti vielä muokata niitä. Profiilien ja sähköpostiviestien kohdalla jää epäselväksi, millainen autonomia tekoölyllä on.

Käsitellyissä tutkimuksissa tekoölyllä ei pyritty optimoimaan viestejä. Tulevaisuudessa tekoölyllä voidaan mahdollisesti yrittää optimoida nettideittailuprofiilin viehättävyys vähäisellä vaivalla (Wu & Kelly, 2020).

### 3 Tutkimusmenetelmä

Tämän tutkielman tutkimusmenetelmä on kirjallisuuskatsaus. Kirjallisuushauissa on käytetty ACM-tietokantaa ja Andor-hakupalvelua. Hakuja tehtiin myös muihin tietokantoihin, mutta niistä ei löytynyt sopivia lähdeartikkeleita. Haut on tehty tammi- ja helmikuussa 2022. Hakusanoina käytettiin termejä ”artificial intelligence”, ”machine learning”, ”AI-mediated communication” ja ”computer-mediated communication”. Näillä tehtiin hakuja eri yhdistelmillä.

Lisäksi lähteitä etsittiin käymällä läpi löytyneiden tutkimusten lähdeluetteloita ja Google Scholarista niihin viitanneita tutkimuksia. Hakutuloksia tarkasteltiin otsikon ja tiivistelmän perusteella. Mukaan otettiin lähdeartikkelit, joissa käsiteltiin tekoälyn vaikutuksia vastaanottajan käsityksiin lähettäjistä tekstipohjaisessa vuorovaikutuksessa. Hakutulos karsittiin myös, jos tekoälyä ei hyödynnetty tekstin tuottamiseen vaan esimerkiksi keskusteluhuoneiden organisoimiseen tai keskusteluaiheiden ehdottamiseen.

Kirjallisuuskatsaukseen valitut lähdeartikkelit on esitetty taulukossa 2 lajiteltuna käytetyn tekoälyavusteisen sovellustyyppin mukaan. Taulukon jälkeen on kuva 2, joka havainnollistaa tutkimusten vuorovaikutustilanteita. Ensimmäisessä kuvassa lähettäjä kirjoittaa profiilin tai sähköpostiviesti tekoälyn avulla ja vastaanottaja lukee tekstin. Toisessa kuvassa kaksi henkilöä on vuorovaikutuksessa pikaviestisovelluksen välityksellä. Myös tässä tilanteessa lähettäjä hyödyntää tekoälyä viestien kirjoittamisessa. Taulukkoon on merkitty, kumpaa asetelmaa kussakin tutkimuksessa on käytetty.

Taulukossa näkyy myös, käytettiinkö tutkimuksessa aitoa tekoälyä vai simuloitiinko tekoälyn käyttö jollain tavalla. Osassa lähdeartikkeleista ei oikeasti käytetty tekoälyä tuottamaan arvioitavaa tekstiä. Automaattista profiilinluontia ja sähköpostiviestejä tarkastelevissa lähdeartikkeleissa teksti oli ihmisen kirjoittama ja arvioijille uskoteltiin eri menetelmin, että se on tekoälyn tuottama.



Taulukko 2. Tarkastellut tekoölyavusteisen vuorovaikutuksen sovellustyypit ja kirjallisuuskatsauksen päälähteet.

Tekoölyavusteisen vuorovaikutuksen sovellustyyppi	Tekoölyn muoto	Lähdeartikkelit
Älykkäät vastausehdotukset (Kuva 2b)	Aito tekoöly	Hohenstein & Jung, 2020
	Aito tekoöly	Mieczkowski ja muut, 2021
Automaattinen profiilinluonti (Kuva 2a)	Simuloitu tekoöly	Bruzzese ja muut, 2020
	Simuloitu tekoöly	Jakesch ja muut, 2019
	Simuloitu tekoöly	Wu & Kelly, 2020
Sähköpostiviestin kirjoittamisen apuna (Kuva 2b)	Simuloitu tekoöly	Liu ja muut, 2022
	Simuloitu tekoöly	Weiss ja muut, 2022



Kuva 2. Tekoöly automaattisessa profiilinluonnissa tai sähköpostiviestin kirjoittamisen apuna sekä pikaviestintätilanteessa.

Ainoastaan älykkäitä vastausehdotuksia tarkastelevissa lähdeartikkeleissa käytettiin pikaviestintäsovellusta, joka hyödynsi aitoa tekoölyjärjestelmää ehdotusten tuottamiseen. Muissa tutkimuksissa tekoölyn osallisuus simuloitiin. Weiss ja muut (2022) selvittivät, miten tekoöly vaikuttaa käsityksiin työnhakijasta. He pyysivät osallistujia arvioimaan fiktiivisen ehdokkaan profiilin ja hakemussähköpostin. Sähköpostissa oli merkintä siitä, että hakija on käyttänyt tekoölyä viestin kirjoittamiseen. Arvioijalle jäi siis epäselväksi, miten tarkalleen tekoölyä oli hyödynnetty.

Liu ja muut (2022) lisäsivät sähköpostiviestin oheen samaan tapaan merkinnän tekoölyn osallisuudesta. He kertoivat kuitenkin merkinnässä, että sähköposti oli kirjoitettu joko älykkään automaattisen täydennyksen avulla tai että se oli kokonaan kehittyneen tekoölyjärjestelmän tuottama.

Profiileja arvioiville osallistujille uskoteltiin, että profiilit olivat tekoälyn automaattisesti luomia. Heille näytettiin tekaistu demovideo, jossa tekoäly luo profiilin Facebook-tilin perusteella. Lisäksi profiileissa oli muistutuksena merkintä siitä, että ne olivat tekoälyn luomia.

#### **4 Tekoälyn vaikutus käsityksiin vuorovaikutuskumppanista**

Luottamus on perustavanlaatuinen osa ihmisten vuorovaikutusta ja sitä on tutkittu jo paljon teknologiavälitteisen vuorovaikutuksen osalta. Niin Airbnb-isäntien kuin treffisovelluksen käyttäjienkin on tärkeää pystyä viestimään omasta luotettavuudestaan nettiprofiilissa, jotta he voivat päästä haluttuun lopputulokseen. Sosiaalinen vetovoima, viehättävyys, lämpö ja pätevyys ovat ominaisuuksia, joita vuorovaikutuksessa yleisimmin pyritään osoittamaan. Lähdeartikkeleissa nämä ominaisuudet on valittu tarkastelukohteiksi. Niissä pyritään siis selvittämään, miten tekoälyn osallisuus vaikuttaa viestin vastaanottajan käsityksiin viestin lähettäjän ominaisuuksista. Aikaisemmissa tutkimuksissa lämmön ja pätevyyden on huomattu olevan tärkeimpiä tekijöitä vaikutelman muodostumisessa (Mieczkowski ja muut, 2021).

Tekoälyn osallisuus itsessään voi vaikuttaa viestin lähettäjistä koettuihin ominaisuuksiin. Lisäksi tekoälyn tuottama kieli, ainakin nykyisellä teknologialla, eroaa ihmisten tuottamasta kielestä, joten sillä voi olla myös vaikutuksia. Yksi havaittu ilmiö on vastausehdotusten positiivisuustaipumus (Hohenstein & Jung, 2018; Mieczkowski ja muut, 2021). Positiivisuustaipumuksesta ja sen mahdollisesta vaikutuksesta käsityksiin vuorovaikutuskumppanista kerrotaan kohdassa 4.3.

Tässä luvussa määritellään ensin tarkemmin kirjallisuuskatsaukseen valittujen tutkimusten tarkastelukohteet. Sen jälkeen käydään läpi tutkimusten tuloksia. Tarkastelukohteet on jaettu kahteen kategoriaan, koettu luotettavuus ja muut ominaisuudet. Tekoälyn vaikutusta koettuun luotettavuuteen on tutkittu eniten, joten se käsitellään ensin omassa kohdassaan. Tämän jälkeen käsitellään muut ominaisuudet, joita ovat viehättävyys, sosiaalinen vetovoima, lämpö ja pätevyys.

##### **4.1 Tutkimusten tarkastelukohteita**

Tässä kohdassa määritellään ensin, kenen ominaisuuksia kussakin lähdeartikkelissa on arvioitu. Sitten selvitetään, miten eri ominaisuudet eli tarkastelukohteet on niissä määritelty. Lähdeartikkeleiden eri tarkastelukohteet on esitelty taulukossa 3. Taulukossa lähdeartikkelit on jaoteltu sen mukaan, minkä perusteella kohdehenkilön ominaisuuksia on arvioitu.

Taulukko 3. Lähdeartikkeleiden tarkastelukohteita tekoölyavusteisen vuorovaikutuksen sovellustyyppin mukaan jaoteltuna.

Lähdeartikkelit	Luotettavuus	Viehättävyyys	Sosiaalinen veto-voima	Lämpö	Pätevyys
<b>Profiilin perusteella</b>					
Bruzzese ja muut, 2020	x				
Jakesch ja muut, 2019	x				
Wu & Kelly, 2020	x	x			
<b>Keskustelun perusteella</b>					
Hohenstein & Jung, 2020	x				
Mieczkowski ja muut, 2021			x	x	x
<b>Sähköpostin perusteella</b>					
Liu ja muut, 2022	x				
Weiss ja muut, 2022			x	x	x

Lähdeartikkeleissa kukin ominaisuus määritellään kohdehenkilön ominaisuutena. Kohdehenkilö on viestin lähettäjä ja arvioija on viestin vastaanottaja. Automaattista profiilinluontia tarkastelevissa tutkimuksissa kohdehenkilö on profiilin edustama Airbnb-isäntä (Bruzzese ja muut, 2020; Jakesch ja muut, 2019) tai online-treffijärjestelmään profiilin tehnyt käyttäjä (Wu & Kelly, 2020). Kahdessa tutkimuksessa tekoölyn vaikutusta selvitettiin keskustelun perusteella. Näissä kaksi henkilöä kävi keskustelun pikaviestintäsovelluksen välityksellä. Hohensteinin ja Jungin (2020) tutkimuksessa osallistujat keskustelivat toisen henkilön kanssa, jonka he luulivat olevan toinen osallistuja. Todellisuudessa keskustelukumppani oli tutkijoiden liittolainen, joka ei käyttänyt sovelluksen vastausehdotuksia. Osallistujia ei käsketty käyttämään vastausehdotuksia, mutta heille kerrottiin, että sovellus tuottaa tekoölyn avulla vastausehdotuksia, joita he voivat käyttää keskustelussa. Osallistujat siis olettivat, että myös vuorovaikutuskumppani käyttää vastausehdotuksia. Kohdehenkilö, jonka ominaisuuksia osallistuja arvioi, oli siis liittolainen. Mieczkowskiin ja muiden (2021) tutkimuksessa molemmat keskusteluun osallistuvista henkilöistä olivat osallistujia. Toiselle osallistujalle annettiin ohjaajan rooli ja toiselle vastaajan rooli. Ohjaajaa ohjeistettiin käyttämään vastausehdotusta aina kun niitä oli tarjolla. Vastaajilla taas ei ollut vastausehdotuksia käytössä eivätkä he tiesivät, että heidän parinsa käyttivät sellaisia. Kohdehenkilö tässä tapauksessa oli siis ohjaaja ja arvioija vastaaja.

Luotettavuus määritellään hieman eri tavoin lähdeartikkeleissa. Kolmessa tutkimuksessa (Bruzzese ja muut, 2020; Jakesch ja muut, 2019; Liu ja muut, 2022) luotettavuutta mitattiin aikaisemmissa tutkimuksissa käytetyllä mallilla, jossa

luotettavuus muodostuu kohdehenkilön kyvykkyydestä (ability), hyväntahtoisuudesta (benevolence) ja rehellisyydestä (integrity). Näistä kaikista oli muodostettu tilanteeseen sopivat väitteet osallistujalle vastattavaksi. Wun ja Kellyn (2020) tutkimuksessa osallistujilta kysyttiin suoraan, oliko kohdehenkilö heidän mielestään luotettava.

Hohensteinin ja Jungin (2020) tutkimuksessa luottamus kytkeytyy attribuutioon. Attribuutiolla tarkoitetaan ihmisten pyrkimystä kuvitella tahallisuutta omaan ja muiden käytökseen (Maselli & Altrocchi, 1969). Ihmisillä on taipumus yliarvioida käytöksen tahallisuutta. Teknologiavälitteisessä vuorovaikutuksessa tämä on yleistä, koska käytössä ei ole aina samoja eleitä ja ilmeitä kuin kasvotusten keskustelussa (Walther & Tong, 2014). Kun vuorovaikutuskumppani koetaan luotettavana, hänen tekojensa takana nähdään yleensä olevan positiivinen motivaatio. Jos esimerkiksi hänellä kestää kauan vastata viestiin, syyksi oletetaan tekninen vika tai muu ulkoinen tekijä eikä asenne häntä kohtaan muutu. (Hohenstein ja Jung, 2020)

Koska Hohenstein ja Jung (2020) näkevät luottamuksen attribuutionaalisen prosessin, heidän tutkimuksessaan vuorovaikutustilanne manipuloitiin olemaan joko onnistunut tai epäonnistunut. Näin he pystyivät mittaamaan, kuinka paljon osallistujat attribuivat vuorovaikutustilanteen tuloksesta kullekin osapuolelle. He pyysivät osallistujia arvioimaan prosentteina, kuinka paljon kukin vuorovaikutuksen osapuoli on vastuussa vuorovaikutustilanteen lopputuloksesta. Vastuu jaettiin osallistujan itsensä, vuorovaikutuskumppanin ja tekoälyn välillä. Lisäksi osallistujat arvioivat vuorovaikutuskumppanin ja tekoälyn luotettavuutta aikaisemmasta tutkimuksesta tiivistetyllä viiden kohdan luotettavuusasteikolla (Wheless & Grotz, 1977).

Mieczkowski ja muut (2021) sekä Weiss ja muut (2022) tutkivat tekoälyn vaikutusta kohdehenkilön sosiaaliseen vetovoimaan (social attraction). Näissä tutkimuksissa osallistujien käsityksiä selvitettiin samalla sosiaalista vetovoimaa mittaavalla asteikolla. Osallistujilta kysyttiin esimerkiksi, voisivatko he kuvitella olevansa ystäviä kohdehenkilön kanssa ja haluaisivatko he käydä ystävällismielisen keskustelun kohdehenkilön kanssa.

Wu ja Kelly (2020) selvittivät tutkimuksessaan kohdehenkilön viehättävyyttä. Myös heidän tutkimuksessaan osallistujilta kysyttiin, haluaisivatko he käydä ystävällismielisen keskustelun kohdehenkilön kanssa. Sen lisäksi heidän tutkimuksessaan osallistujia pyydettiin arvioimaan kohdehenkilön yleinen viehättävyys ja romanttinen vetovoima.

Kohdehenkilön lämpöä ja pätevyyttä pyydettiin arvioimaan molemmissa näitä tarkastelevissa tutkimuksissa asteikolla yhdestä (ei ollenkaan) viiteen (erittäin) kahdeksan eri ominaisuuden suhteen, joista muodostettiin kohdehenkilön lämpö- sekä pätevyyslukema. Lämmön kohdalla osallistujaa pyydettiin arvioimaan kohdehenkilön vilpittömyys, hyväluontoisuus, lämpö ja suvaitsevaisuus. Pätevyys muodostettiin älykkyyden, itsevarmuuden, kilpailuhenkisyyden ja itsenäisyyden arvioista. Weissin ja

muiden (2022) kyselyssä näiden neljän lisäksi mukana oli myös suora kysymys pätevyydestä.

#### **4.2 Tekoälyn vaikutus koettuun luotettavuuteen**

Tekoälyn osallisuudella voidaan todeta olevan vaikutus vuorovaikutuskumppanin koettuun luotettavuuteen. Jakeschin ja muiden (2019) sekä Wun ja Kellyn (2020) tutkimusten mukaan tekoälyllä voi olla negatiivinen vaikutus koettuun luotettavuuteen, kun sitä käytetään automaattiseen profiilinluontiin. Jakesch ja muut (2019) tutkivat kokevatko ihmiset sellaisen Airbnb-isännän vähemmän luotettavaksi, jonka profiili on tekoälyn kirjoittama. Heidän ensimmäisessä kokeessaan osallistujat jaettiin kahteen ryhmään, vertailuryhmään ja testiryhmään. Testiryhmän osallistujat luulivat, että heidän arvioimansa profiilit olivat tekoälyn kirjoittamia. Eroa koetussa luotettavuudessa ei havaittu näiden ryhmien välillä. Myöhemmässä kokeessa testiryhmät arvioivat sekalaista kokoelmaa profiileja, jossa oli sekä ihmisten että oletetusti tekoälyn kirjoittamia profiileja. Tässä koeasetelmassa tekoälyprofiilien luotettavuusarviot olivat huomattavasti alhaisempia.

Tämä eroaa Wun ja Kellyn (2020) tuloksista tutkimuksessa, jossa he testasivat tekoälyn vaikutusta luottamukseen nettideittailuprofiileissa. Testiryhmälle näytettiin pelkästään profiileja, jotka oli merkitty tekoälyn kirjoittamiksi, mutta osallistujien kokema luottamus oli silti vähäisempää kuin vertailuryhmässä, jossa osallistujille näytettiin vain ihmisten kirjoittamia profiileja. Wun ja Kellyn (2020) mukaan luotettavuus saattaa olla tärkeämpää nettideittailussa kuin Airbnb-kohteen vuokrauksessa, mikä voisi selittää erot.

Myös Liu ja muut (2022) totesivat, että tekoälyllä on negatiivinen vaikutus vuorovaikutuskumppanin luotettavuuteen. Heidän tutkimuksessaan osallistujat arvioivat lähettäjän luotettavuutta sähköpostiviestin perusteella. He vertasivat arvioita viesteistä, jotka oli merkitty joko kokonaan ihmisen kirjoittamaksi, ihmisen ja tekoälyn yhdessä kirjoittamaksi tai kokonaan tekoälyn kirjoittamaksi. Lisäksi mukana oli vertailuryhmä, jonka arvioimissa viesteissä ei ollut merkintää. Tekoälyn osallisuuden laajuus siis vaihteli eri ryhmillä. Mitä laajempi tekoälyn osallisuus oli, sitä alhaisempi luotettavuusarvio viestin perusteella annettiin.

Toisaalta tekoälyllä voi olla myös positiivinen vaikutus vuorovaikutuskumppanin koettuun luotettavuuteen tekoälyavusteisessa vuorovaikutuksessa. Hohenstein ja Jung (2020) tutkivat tekoälyn vaikutusta koettuun luottamukseen ja attribuutioon. Tutkimuksessa osallistujat käyttivät joko normaalia viestintäsovellusta tai tekoälyavusteista viestintäsovellusta. Tutkimuksen tulokset viittaavat siihen, että epäonnistuneissa vuorovaikutustilanteissa tekoäly toimii eräänlaisena syntipukkina, jonka vastuulle epäonnistuminen voidaan asettaa. Onnistuneessa vuorovaikutustilanteessa tekoälyä ei pidetty vaikuttavana tekijänä. Tekoälyn osallisuus

lisäsi kuitenkin luottamusta keskustelutoveriin niin onnistuneissa kuin epäonnistuneissakin vuorovaikutustilanteissa.

Bruzzese ja muut (2020) selvittivät tutkimuksessaan, onko luottamusindikaattorilla vaikutusta tekoälyn avulla kirjoitettujen profiilien edustamien isäntien koettuun luotettavuuteen. Luottamusindikaattori (confidence indicator) on algoritmin tuottama tekstiarvio sen luottamuksesta sen lähdetietoihin sekä perustelut sille, miksi tiedot ovat luotettavia tai vähemmän luotettavia. Luottamusindikaattoreita ei ole aikaisemmin tutkittu tekoölyavusteisen vuorovaikutuksen kontekstissa, mutta niistä on saatu lupaavia tutkimustuloksia käyttäjän luottamuksesta tekoälyn päätöksiin muissa konteksteissa. Airbnb-profiilien yhteydessä alhainen luottamusindikaattori vaikutti koettuun luotettavuuteen negatiivisesti, mutta korkealla indikaattorilla ei ollut vaikutusta.

Taulukossa 4 on esitetty kaikkien luotettavuutta tarkastelleiden lähdeartikkeleiden tulokset.

Taulukko 4. Tekoälyn vaikutus koettuun luotettavuuteen.

Lähdeartikkelit	Positiivinen vaikutus	Negatiivinen vaikutus	Huomioita
<b>Profiilin perusteella</b>			
Jakesch ja muut, 2019		x	Vain kun tekoölyavusteiset ihmisten kirjoittamien seassa
Bruzzese ja muut, 2020		x	Luottamusindikaattori
Wu & Kelly, 2020		x	
<b>Keskustelun perusteella</b>			
Hohenstein & Jung, 2020	x		Epäonnistuminen nähtiin tekoälyn syynä
<b>Sähköpostin perusteella</b>			
Liu ja muut, 2022		x	

Hohensteinin ja Jungin (2020) tutkimus oli ainut, jossa tekoälyn osallisuudella oli positiivinen vaikutus lähettäjän koettuun luotettavuuteen. Heidän tutkimuksensa eroaa muista myös siinä, että he käyttivät aidosti tekoälyn tuottamia vastausehdotuksia, kun taas muissa profiili tai sähköpostiviesti oli oikeasti ihmisen kirjoittama. Tekoälyn osallisuus oli muissa tutkimuksissa ilmaistu viestin oheen lisätyllä tekstillä. Luotettavuutta mitattiin myös erilaisilla kysymyksillä kuin muissa tutkimuksissa, joten tulosten vertailu on haastavaa.

### 4.3 Tekoölyn vaikutus käsityksiin muista ominaisuuksista

Koetun luotettavuuden lisäksi tutkimuksissa selvitettiin käsityksiä kohdehenkilön viehättävyydestä, sosiaalisesta vetovoimasta, lämmöstä ja pätevyyydestä.

Aikaisemman tutkimuksen mukaan kieli on vahvasti yhteydessä muodostuviin käsityksiin ihmissuhteissa. Tekoölyn tuottama kieli ei vielä vastaa täysin ihmisen tuottamaa kieltä ja tekstin tuottamiseen hyödynnettävät kielimallit saattavat heijastaa niiden koulutukseen käytetyssä datassa olevia vääristymiä. Yksi tällainen vääristymä on vastausehdotusten positiivisuustaipumus. Jo Kannan ja muut (2016) huomasivat, että vastausehdotuksia tuottava tekoölyjärjestelmä suosii vahvasti positiivisia vastauksia, ja he muokkasivat järjestelmää niin, että positiivisten vastausehdotusten lisäksi ainakin yksi ehdotus on negatiivinen. Hohenstein ja Jung (2018) huomasivat kuitenkin ylivoimaisesti suurimman osan vastausehdotuksista olevan yhä positiivisia ja Mieczkowski ja muut (2021) vahvistivat tämän löydön.

Mieczkowski ja muut (2021) tutkivat voiko mahdollinen positiivisuustaipumus vaikuttaa ihmisten tunneilmaisuihin ja siihen, millaisin kielen rakentein älykkäisiin vastauksiin vastataan. Heidän tutkimuksessaan selvitettiin tekoölyn vaikutusta koettuun lämpöön ja sosiaaliseen vetovoimaan. Tekoölyllä ei ollut vaikutusta keskustelutovereiden kokemaan lämpöön toisiaan kohtaan, mutta tulokset viittaavat siihen, että sillä saattaa olla negatiivinen vaikutus sosiaaliseen vetovoimaan. Mahdollinen selitys tälle on, että tekoölyn positiivisuustaipumus heikentää vuorovaikutuskumppanin sosiaalista vetovoimaa, koska liian positiiviset vastaukset vaikuttavat kummallisilta. (Mieczkowski ja muut, 2021) Tätä tukee Robertsonin ja muiden (2021) tutkimus, jossa pyrittiin kuvailemaan, mikä voi tehdä sähköpostin vastausehdotuksista ongelmallisia. He havaitsivat, että ehdotusten negatiivisuutta lisättiin ja hyperpositiivisten ehdotusten positiivisuutta vähennettiin. Tämä viittaisi siihen, että positiiviset vastaukset nähdään joskus epäsoveliaina.

Weiss ja muut (2022) tutkivat millainen vaikutus tekoölyn käytöllä on työnhaussa. He tarkastelivat, miten tekoölyn käyttö vaikutti osallistujien vaikutelmaan työnhakijoiden lämmöstä, pätevyyydestä ja sosiaalisesta vetovoimasta. Tutkimuksessa osallistujia pyydettiin arvioimaan työnhakija sähköpostiviestin perusteella, jossa oli merkintä siitä, että viestin kirjoittamisessa oli hyödynnetty tekoölyä. Toisin kuin Mieczkowskiin ja muiden (2021) tutkimuksessa, tekoölyä käyttäneet kohdehenkilöt arvioitiin huomattavasti vähemmän lämpimiksi kuin muut kohdehenkilöt. Tekoölyn käyttö vaikutti myös negatiivisesti kohdehenkilön sosiaaliseen vetovoimaan. Tutkimuksessa verrattiin tekoölyä hyödyntävää ryhmää ja ihmiseltä apua saaneiden ryhmää sekä vertailuryhmää. Tekoölyä hyödyntäneet ja ihmiseltä apua saaneet arvioitiin vähemmän päteviksi kuin vertailuryhmä. Nämä ryhmät arvioitiin muita alhaisemmin myös, kun osallistujilta kysyttiin, olisivatko he valmiita palkkaamaan kohdehenkilön. Tämä viittaisi siihen, että

minkä tahansa avun tarvitseminen antaa kohdehenkilöstä negatiivisen kuvan eikä pelkästään tekoälyn hyödyntäminen.

Wun ja Kellyn (2020) tutkimuksessa osallistujat arvioivat kohdehenkilöä nettideittailuprofiilin perusteella. Tekoälyllä ei havaittu olevan vaikutusta kohdehenkilön viehättävyyteen.

Lähdeartikkeleiden tulokset on koottu taulukkoon 5.

Taulukko 5. Tekoälyn vaikutus kohdehenkilön koettuun viehättävyyteen, sosiaaliseen vetovoimaan, lämpöön sekä pätevyys.

Lähdeartikkelit	Viehättävyys	Sosiaalinen vetovoima	Lämpö	Pätevyys
<b>Profiilin perusteella</b>				
Wu & Kelly, 2020	Ei vaikutusta			
<b>Keskustelun perusteella</b>				
Mieczkowski ja muut, 2021		Negatiivinen vaikutus	Ei vaikutusta	Ei vaikutusta
<b>Sähköpostin perusteella</b>				
Weiss ja muut, 2022		Negatiivinen vaikutus	Negatiivinen vaikutus	Negatiivinen vaikutus (ei eroa ihmisavusteiseen)

Kuten taulukosta näkee, osa lähdeartikkeleiden tuloksista on ristiriitaisia keskenään. Käsitystä kohdehenkilön lämmöstä mitattiin molemmissa tutkimuksissa samalla asteikolla, mutta Weiss ja muut (2022) havaitsivat tekoälyllä olevan negatiivinen vaikutus kohdehenkilön koettuun lämpöön, kun taas Mieczkowski ja muut (2021) eivät havainneet minkäänlaista vaikutusta koettuun lämpöön.

Sosiaaliseen vetovoimaan sen sijaan havaittiin molemmissa tutkimuksissa negatiivinen vaikutus. Positiivista vaikutusta tekoälyllä ei ollut mihinkään näistä ominaisuuksista.

## 5 Keskustelu

Tutkimusten tulokset osoittavat, että tekoälyllä voi olla vaikutus käsityksiin vuorovaikutuskumppanista. Koettuun luotettavuuteen tekoäly näyttäisi vaikuttavan negatiivisesti. Vuorovaikutuskumppanin luotettavuuden arvioinnissa vain yhdessä tutkimuksessa (Hohenstein & Jung, 2020) tekoälyllä oli positiivinen vaikutus koettuun luotettavuuteen. Muissa tutkimuksissa havaittiin negatiivinen vaikutus. Miksi Hohensteinin ja Jungin (2020) tulokset eroavat muista? He olivat ainoita, jotka käyttivät oikeasti tekoälyn tuottamia vastausehdotuksia koetun luotettavuuden tutkimiseen, kun



taas muissa tutkimuksissa käytettiin ihmisten kirjoittamia tekstejä. He käyttivät myös erilaista kyselyä koetun luotettavuuden mittaamiseen, joten tulokset eivät ole täysin verrattavissa.

Myös tekoälyn osallisuuden laajuudella saattaa olla vaikutusta koettuun luotettavuuteen. Vastausehdotuksissa tekoälyn osallisuuden laajuus ei ole yhtä suuri kuin profiilinluonnissa. Vastausehdotus on lyhyt teksti, jota lähettäjä voi vielä muokata ja täydentää. Liun ja muiden (2022) tutkimuksessa luotettavuus väheni sitä mukaa mitä laajempi tekoälyn osallisuus oli. Vaikutus oli kuitenkin myös pienimmällä laajuudella negatiivinen, toisin kuin Hohensteinilla ja Jungilla. Tekoälyn laajuuden vaikutusta luotettavuuteen olisi hyvä tutkia lisää.

Toinen huomionarvoinen asia oli Jakeschin ja muiden (2019) havainto siitä, että tekoäly vaikutti negatiivisesti koettuun luotettavuuteen vain, kun tekoälyavusteiset profiilit arvioitiin ihmisten kirjoittamien seassa. Wun ja Kellyn (2020) tutkimuksessa tekoälyavusteiset profiilit arvioitiin eri ryhmässä kuin ihmisten kirjoittamat, mutta koettu luotettavuus oli silti alhaisempi. Tämä saattaa johtua kontekstista, kuten Wu ja Kelly ehdottavat. Ihmiset kokevat yleensä tekoälyn olevan objektiivinen, eikä kykeneväinen tunteelliseen ilmaisuun. Tekoälyn tuottamat tunneilmaisut voidaan siis nähdä epäaitoina. Aitojen tunteiden ilmaisu on tärkeää nettideittailussa, mutta ei niinkään majapaikan vuokraamisessa.

Koetun luotettavuuden lisäksi tekoälyllä näyttäisi olevan negatiivinen vaikutus myös muihin ominaisuuksiin. Tutkimukset eivät ole kaikista kuitenkaan yhtä mieltä. Viehättävyyteen ei havaittu vaikutusta ja lämmön sekä pätevyuden suhteen tulokset olivat ristiriitaisia. Pätevyuden kohdalla konteksti voi selittää eron tuloksissa. Weiss ja muut (2022) tarkastelivat koettua pätevyyttä työnhaun kontekstissa, jossa se on erittäin tärkeä ominaisuus ja avun saaminen voi vaikuttaa siihen negatiivisesti. Sosiaaliseen vetovoimaan havaittiin negatiivinen vaikutus molemmissa sitä tarkastelleissa tutkimuksissa.

Tutkimustuloksiin liittyy useita varauksia, jotka vaikuttavat niiden yleistettävyyteen. Ensinnäkin lähes kaikissa tutkimuksissa osallistujat olivat länsimaalaisia ja kaikki tutkimukset tehtiin englannin kielellä. Tämä rajoittaa tulosten yleistämistä erityisesti tutkimuksista, joissa käytettiin aitoa tekoälyä tekstin tuottamiseen. Englanninkielisissä vastausehdotuksissa havaittu positiivisuustaipumus ei välttämättä esiinny muissa kielissä. Eri kulttuureissa on myös erilaiset kielenkäyttöön liittyvät normit ja käytänteet sekä erilaisia käsityksiä tekoälystä. Tarvittaisiin siis tutkimusta myös muilla kielillä ja eri kulttuureihin kuuluvilla ihmisillä.

Pitää myös pohtia, miten hyvin tutkimukset ovat vertailtavissa keskenään. Suurimmassa osassa tutkimuksista teksti ei ollut aidosti tekoälyn tuottamaa, vaan ihmisen kirjoittamaa ja tekoälyn osallisuus ainoastaan simuloitiin osallistujille. Liu ja muut (2022)

kertovat halunneensa imitoida tekoälyä sellaisena kuin se on tulevaisuudessa, jolloin sen tuottama teksti ei eroa mitenkään ihmisen tuottamasta tekstistä. Jakesch ja muut (2019) kertovat myös käyttäneensä simuloitua tekoälyä, koska he halusivat ymmärtää ihmisten käsityksiä tekoälyn osallisuudesta eivätkä niinkään reaktioita tekoälyn tuottamaan tekstiin. Näissä tutkimuksissa käsityksiin vuorovaikutuskumppanista vaikuttivat luultavasti arvioijan tekoälyyn liittyvät ennakkoluulot ja mielipiteet. Tuloksia on siis hankala verrata tutkimuksiin, jossa tekoälyn vaikutusta tutkittiin sen tuottaman tekstin kautta.

Hohenstein ja Jung (2020) sekä Mieczkowski ja muut (2021) käyttivät tutkimuksissaan aidosti tekoälyn tuottamia vastausehdotuksia. Hohensteinin ja Jungin tutkimuksessa osallistujat käyttivät itse vastausehdotuksia ja luulivat myös vuorovaikutuskumppanin käyttävän vastausehdotuksia. Tekoälyn tuottamaa tekstiä oli siis mukana keskustelussa, mutta osallistujien arvioihin vuorovaikutuskumppanin luotettavuudesta vaikutti luultavasti enemmän tiedostettu tekoälyn osallisuus. Tämä vastaisi siis jokseenkin tutkimuksia, joissa tekoälyn osallisuus simuloitiin.

Mieczkowskiin ja muiden tutkimuksessa taas vuorovaikutuskumppanin sosiaalista vetovoimaa ja lämpöä arvioinut osallistuja ei ollut edes tietoinen, että pari käytti tekoälyn tuottamia vastausehdotuksia. Negatiivinen vaikutus koettuun sosiaaliseen vetovoimaan johtui siis täysin tekoälyn tuottamasta kielestä. Yksi selitys negatiiviselle vaikutukselle voisi siis olla vastausehdotusten positiivisuustaipumus. Tutkimuksessa vastausehdotuksia käskettiin käyttää aina kun sellaisia oli tarjolla. Vastausehdotuksia on siis pitänyt käyttää, vaikka ne eivät olisi olleet täysin sopivia, mikä on myös voinut vaikuttaa kummalliselta viestin vastaanottajalle ja vaikuttaa negatiivisesti lähettäjän sosiaaliseen vetovoimaan. Vastausehdotusten vaikutusta olisi siis hyvä tutkia aidossa viestintätilanteessa, jossa osallistujat voivat käyttää vastausehdotuksia heille luonnolliseen tapaan.

Tässä tutkielmassa huomattiin, että tekoäly voi vaikuttaa käsityksiin vuorovaikutuskumppanista kahdella eri tavalla. Vaikutus voi syntyä tekoälyn tuottaman kielen rakenteista, jotka eroavat ihmisen tuottamasta kielestä tai ihmisten tekoälyyn liittyvistä ennakkokäsityksistä. Tulevissa tutkimuksissa on tärkeää tehdä selväksi, kumpaa näistä tutkitaan. Havaittiin myös, että tekoälyllä voi olla negatiivinen vaikutus koettuun luotettavuuteen, mikä ei ole toivottua. Jos tekoälyn käytön ilmoittamisesta seuraa vain negatiivisia vaikutuksia, herää kysymys siitä pitäisikö tekoälyn osallisuus tekstin tuotossa jättää kertomatta. Onko kuitenkin eettistä pitää tieto salaisena? Tulevaisuudessa tutkimuksen olisi hyvä selvittää, mistä negatiivinen vaikutus pohjimmiltaan syntyy ja miten tekoälyavusteinen vuorovaikutus voidaan suunnitella niin, että se ei tapahdu.

Lopuksi on vielä syytä pohtia, mikä kaikki luetaan tekoälyavusteisen vuorovaikutuksen joukkoon, ja onko tutkielmassa tehty mielekkäitä rajauksia lähteiden

valinnassa. Tekoölyavusteinen vuorovaikutus määritellään ihmisten väliseksi vuorovaikutukseksi, jossa tekoöly voi muokata, tuottaa tai täydentää tekstiä. Tämän ei tarvitse tapahtua kahdenkeskisessä reaaliaikaisessa keskustelussa. Netissä esillä olevan profiilin voi lukea monta sataa ihmistä. Sen tarkoituksena on aloittaa vuorovaikutus kahden ihmisen välillä, mutta näin ei välttämättä aina tapahdu. Tämä ei eroa juurikaan esimerkiksi romaanin kirjoittamisesta tekoölyä hyödyntäen. Voitaisiinko senkin siis ajatella olevan tekoölyavusteista vuorovaikutusta? Mielestäni tekoölyn hyödyntäminen luovan kirjoittamisen apuna ei kuulu kuitenkaan tekoölyavusteiseen vuorovaikutukseen, koska siinä tekstin välittäminen toiselle ei ole ihmishuokkeudesta. Kirjallisuuskatsauksessa tuli vastaan myös tutkimus, jossa tekoölyä hyödyntävää chatbottia käytettiin edistämään keskusteluja ehdottamalla keskusteluaiheita. Tätä voisi mielestäni myös kutsua tekoölyavusteiseksi vuorovaikutukseksi, vaikka tekoöly ei ole mukana ihmisten lähettämien viestien tuottamisessa.

## **6 Yhteenveto**

Tekoölyavusteinen vuorovaikutus on vasta syntymässä oleva tutkimusalue ja tässä tutkielmassa tarkasteltiin uusinta aihepiirin tutkimusta. Kirjallisuuskatsauksen avulla pyrittiin selvittämään, miten tekoöly vaikuttaa ihmisten välisessä vuorovaikutuksessa käsityksiin vuorovaikutuskumppanista. Keskityttiin tarkastelemaan vaikutuksia viestin vastaanottajan käsityksiin lähettäjistä tekstipohjaisessa vuorovaikutuksessa. Tekoöly oli tutkimuksissa joko simuloitu profiilinluonnissa ja sähköpostiviestin kirjoittamisen apuna tai aidosti mukana tuottamassa vastausehdotuksia reaaliaikaisessa viestittelyssä.

Kirjallisuuskatsauksen päälähteissä tarkasteltiin tekoölyn vaikutusta koettuun luotettavuuteen, viehättävyyteen, sosiaaliseen vetovoimaan, lämpöön ja pätevytyteen. Tekoölyn osallisuudella vuorovaikutuksessa vaikuttaa olevan lähinnä negatiivinen vaikutus käsityksiin vuorovaikutuskumppanista tai ei vaikutusta ollenkaan. Huomattiin myös, että tutkimustulokset ovat osittain ristiriitaisia keskenään. Lähes kaikki luotettavuutta tarkastelleet tutkimukset havaitsivat tekoölyllä olevan negatiivinen vaikutus. Yhdessä tutkimuksessa kuitenkin sillä huomattiin olevan positiivinen vaikutus koettuun vuorovaikutuskumppanin luotettavuuteen. Tätä löydöstä ja sen mahdollisia käyttökohteita olisi hyvä tutkia lisää. Tekoölyavusteista vuorovaikutusta tutkittiin eri tavoin eri konteksteissa, mikä voi selittää erot löydöksissä.

## Lähdeluettelo

- Bruzzese, T., Gao, I., Dietz, G., Ding, C., & Romanos, A. (2020). Effect of confidence indicators on trust in AI-generated profiles. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8. <https://doi.org/10.1145/3334480.3382842>
- Carr, C.T. (2020). CMC is dead, long live CMC!: Situating computer-mediated communication scholarship beyond the digital age. *Journal of Computer-Mediated Communication*, 25(1), 9–22. <https://doi.org/10.1093/jcmc/zmz018>
- Hancock, J.T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1), 89–100. <https://doi.org/10.1093/jcmc/zmz022>
- Hohenstein, J. & Jung, M. (2018). AI-supported messaging: An investigation of human-human text conversation with AI support. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018-, 1–6. <https://doi.org/10.1145/3170427.3188487>
- Hohenstein, J. & Jung, M. (2020). AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior*, 106, 106190–. <https://doi.org/10.1016/j.chb.2019.106190>
- Jakesch, M., French, M., Ma, X., Hancock, J.T., & Naaman, M. (2019). AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems CHI '19*, 1–13. <https://doi.org/10.1145/3290605.3300469>
- Kannan, A., Kurach, K., Ravi, S., Kaufmann, T., Tomkins, A., Miklos, B., Corrado, G., Lukacs, L., Ganea, M., Young, P., & Ramavajjala, V. (2016). Smart Reply: Automated response suggestion for email. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 955–964. <https://doi.org/10.1145/2939672.2939801>
- Liu, Y., Mittal, A., Yang, D., & Bruckman, A. (2022). Will AI console me when I lose my pet? Understanding perceptions of AI-mediated email writing. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*, 1–13. <https://doi.org/10.1145/3491102.3517731>
- Maselli, M.D. & Altrocchi, J. (1969). Attribution of intent. *Psychological Bulletin*, 71(6), 445–454. <https://doi.org/10.1037/h0027348>
- Microsoft. (n.d.) Käytä vastausehdotuksia Outlookissa. (Haettu 24.4.2022) <https://support.microsoft.com/fi-fi/office/k%C3%A4yt%C3%A4-vastausehdotuksia-outlookissa-19316194-0434-43ba-a742-6b5890157379#ID0EBD=Web>

- Mieczkowski, H., Hancock, J.T., Naaman, M., Jung, M., & Hohenstein, J. (2021). AI-mediated communication: Language use and interpersonal effects in a referential communication task. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–14. <https://doi.org/10.1145/3449091>
- Robertson, R.E., Olteanu, A., & Diaz, F. (2021). I can't reply with that: Characterizing problematic email reply suggestions. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, 1–18. <https://doi.org/10.1145/3411764.3445557>
- Russell, S. & Norvig, P. (2021). *Artificial Intelligence: a modern approach* (4th edition). Harlow: Pearson.
- Walther, J.B. & Tong, S.T. (2014). Inventing partners in computer-mediated communication: How CMC sustains self-fulfilling prophecies and relational attributions. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol 8531*. (s. 517–527). Springer International Publishing. [https://doi.org/10.1007/978-3-319-07632-4\\_49](https://doi.org/10.1007/978-3-319-07632-4_49)
- Weiss, D., Liu, S.X., Mieczkowski, H., & Hancock, J.T. (2022). Effects of using artificial intelligence on interpersonal perceptions of job applicants. *Cyberpsychology, Behavior and Social Networking*. <https://doi.org/10.1089/cyber.2020.0863>
- Wheless, L.R. & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research*, 3(3), 250–257. <https://doi.org/10.1111/j.1468-2958.1977.tb00523.x>
- Wu, Y., & Kelly, R.M. (2020). Online dating meets artificial intelligence: How the perception of algorithmically generated profile text impacts attractiveness and trust. In *32nd Australian Conference on Human-Computer Interaction (OzCHI '20)*, 444–453. <https://doi.org/10.1145/3441000.3441074>
- Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Sellito, M., Sakhaee, E., Shoham, Y., Clark, J., & Perrault, R. (2022). The AI index 2022 annual report. Stanford University Human-Centered Artificial Intelligence (HAI). (Haettu 24.04.2022) [https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report\\_Master.pdf](https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf)