

Tekoälyn luoma sisältö sosiaalisessa mediassa - Käyttäjän luottamusta heikentävät tekijät

Markkinoinnin kandiditutkielma

Laatija:

Minja Riihimaa

2205822

Ohjaaja:

Aino Halinen-Kaila

3.5.2025

Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck -järjestelmällä

Kandidaatintutkielma

Tutkinto-ohjelman nimi, oppiaine *Aino Halinen-Kaila*

Etunimi Sukunimi *Minja Riihimaa*

Tutkielman otsikko *Tekoälyn luoma sisältö sosiaalisessa mediassa -Käyttäjän luottamusta heikentävät tekijät*

Sivumäärät: 34, 1

Tämä tutkielma käsittelee tekoälyn luomaa sisältöä sosiaalisessa mediassa ja sen vaikutuksia käyttäjän luottamukseen. Tutkielman tarkoitus on kuvata, miten tekoälyn tuottama sisältö voi heikentää kuluttajan luottamusta sosiaaliseen mediaan. Tähän pyritään vastaamaan kolmella tutkimuskysymyksellä:

(1) Miten tekoälyä käytetään sosiaalisen median sisällöntuotannossa?

(2) Mistä luottamus rakentuu sosiaalisessa mediassa?

sekä (3) Miten tekoälyn tuottama sisältö vaikuttaa kuluttajan luottamukseen?

Tutkielma on kirjallisuuskatsaus, jossa sovelletaan Mayerin ym. (1995) sekä Lalotin ja Bertrammin (2025) luottamuksen muodostumisen malleja. Mayerin ym. (1995) ABI-malli (kyvykyys, hyväntahtoisuus, rehellisyys) toimii taustana kaikelle ja Lalotin ja Bertrammin malli on kohdennettu tekoälyn luotettavuuden muodostumiseen. Tekoälyä voi käyttää sosiaalisessa mediassa moneen asiaan, mutta tutkielmassa käydään läpi kolme käyttötapaa: huijauksien tekeminen, väärän tiedon levittäminen ja markkinointi. Käyttömuodoista tarkastellaan syvävärennöksiä, virtuaalivaikuttajia ja sosiaalisia botteja. Tässä tutkielmassa määriteltiin käyttäjän luottamuksen sosiaalista mediaa kohtaa muodostuvan: 1) luottamuksesta instituutioihin, 2) luottamuksesta muihin ihmisiin, 3) luottamuksesta tekoölyyn ja 4) luottamuksesta sosiaalisen median sisältöön. Henkilökohtainen luottamustaipumus vaikuttaa näiden kaikkien takana yksilöllisesti. Tekoälyn tuottaman sisällön vaikutukset luottamukseen riippuvat pitkälti siitä, miten ja mihin tarkoitukseen tekoälyä on käytetty. Instituutioiden rooli luottamuksen ylläpitämisessä on tärkeä. Tutkielma tuo esiin tärkeän näkökulman tekoälyn rooliin sosiaalisessa mediassa ja sen vaikutuksiin käyttäjän luottamukseen. Tutkielman rajoituksena on lähteinä käytettyjen tutkimusten pirstaleisuus. Luottamusteorioiden yhteen tuomisessa sosiaalisen median kontekstissa on selkeä tutkimusaukko ja tätä voisi tutkia tulevaisuudessa tarkemmin.

Sisällysluettelo

1	Johdanto	6
1.1	Tutkimuksen tarkoitus	8
1.2	Tutkimuksen rakenne	8
2	Tekoälyn luoman sisällön muodot ja käyttötavat sosiaalisessa mediassa	9
2.1	Sisällön muodot	9
2.1.1	Syväväärennökset	9
2.1.2	Virtuaaliset vaikuttajat	10
2.1.3	Sosiaaliset botit	11
2.2	Tekoälyn käyttötavat sosiaalisessa mediassa	12
2.2.1	Huijaukset	12
2.2.2	Väärän tiedon levittäminen	12
2.2.3	Markkinointi	13
3	Luottamuksen rakentuminen sosiaalisessa mediassa	14
3.1	Luottamuksen tyypit	14
3.1.1	Henkilökohtainen luottamustaipumus	15
3.1.2	Interpersoonallinen luottamus	15
3.1.3	Institutionaalinen luottamus	16
3.2	Luottamuksen kohteet	16
3.2.1	Ihminen ja instituutiot	17
3.2.2	Sosiaalisen median sisältö	17
3.2.3	Tekoäly -teknologia vai luoja?	18
3.3	Luottamuksen saajan piirteet	18
3.3.1	ABI-malli: kyvykkyys, hyväntahtoisuus ja rehellisyys	19
3.3.2	Tekoälyn ihmismäisyys	19
4	Tekoälyn tuottaman sisällön vaikutus käyttäjän luottamukseen	21
4.1	Tekoälyn eri muotojen vaikutus luottamukseen	21
4.1.1	Syväväärennösten vaikutus luottamukseen	21
4.1.2	Virtuaalivaikuttajien vaikutus luottamukseen	22
4.1.3	Sosiaalisten bottien vaikutus luottamukseen	24
4.2	Tekoälyn objektiivisuuden haasteet	24
4.3	Instituutioiden rooli ja haasteet tekoälyn valvonnassa	25

5 Johtopäätökset	27
6 Yhteenveto	29
Lähteet	31
Liitteet	37
Liite 1. Tekoälyn käyttö	37

KUVIOT

Kuvio 1. Luottamuksen muodostuminen sosiaaliseen mediaan

1 Johdanto

Tekoäly kehittyi nopeasti ja se on jo laajassa julkisessa käytössä. Kenen tahansa on helppo käyttää sitä erilaisten työkalujen avulla kuten ChatGPT, joka pystyy generoimaan tekstiä tai DALL-E, joka generoi kuvia (Feuerriegel ym. 2024). Vaikka tekoälyä voidaan käyttää hyödyksi, se mahdollistaa myös väärän tiedon levittämisen, sekä huijaukset, joita on hyvin hankala erottaa huijauksiksi. Helsingin Sanomat (5.3.2025) uutisoi verkossa tapahtuvien tekoälyä hyödyntävien huijausten räjähdysmäisestä noususta. Huijauksia oli yli 100 miljoonan edestä. Yleisin huijausmuoto, tietojenkalastelu, oli kasvanut 161% ja sijoitushuijaukset ovat kasvaneet 24% vuodesta 2023. Ei ole siis yllättävää, että mediassa käydään keskustelua siitä, miten erottaa tekoälyn tuottama sisältö ihmisen tuottamasta sisällöstä. On vain ajan kysymys, milloin tekoäly oppii peittämään erot, joita on jo nyt vaikea huomata. Useassa lähteessä (Vaccari & Chadwick 2020; Karnouskos 2020; Harris 2021; Mustak ym. 2023) osoitetaan huolta kansalaisten luottamuksen murenemisesta varsinkin instituutioita kohtaan tekoälysisällön ja tekoälyavusteisten huijausten yleistyessä. Jotkut tutkijat pelkäävät tekoälyteknologian, varsinkin syvävääreennysten aiheuttavan ”todellisuuden romahduksen” ja ”informaation maailmanlopun” (Harris 2021). Tutkimusaihe on siis ajankohtainen. Markkinoinnin näkökulmasta on tärkeää ymmärtää, miten ilmiö vaikuttaa kuluttajan luottamukseen sosiaalista mediaa kohtaan, sillä sosiaalisen median markkinointiin käytettävä budjetti kasvaa vuosittain (Statista 2024). Sosiaaliseen mediaan käytettävän markkinointibudjetin odotetaan kasvavan yli neljään miljardiin 2029 mennessä (Statista 2024), minkä vuoksi on hyvä ymmärtää, miten kuluttajan suhtautuminen sosiaalisen median sisältöön muuttuu. Lisäksi, tekoälyä käyttävän markkinoijan on hyvä ymmärtää tekoälyn luoman sisällön luotettavuuteen vaikuttavat tekijät.

Luottamusta tekoälyyn ja tekoälyn tuottaman sisällön vaikutusta luottamukseen on tutkittu paljon viime vuosina. Syvävääreännöksiä on tutkittu varsinkin poliittisesta näkökulmasta ja terveyden (esim. rokotteet) näkökulmasta. Laaja-alainen analyysi tekoälyn eri muotojen vaikutuksista luottamukseen kuitenkin puuttuu. Myös luottamuksen eri tyyppisiä on sovellettu sosiaaliseen mediaan, mutta yhtenäistä katsausta luottamuksen muodostumisesta sosiaalisessa mediassa ei ole. Tekoälyn luotettavuuden yhteydessä käydään keskustelua luottamuksen kohteesta. Pohditaan, kohdistuuko luottamus teknologiaan vai tekoälyn luojaan. Tutkijat ovat eri mieltä, joten aihe vaatii jatkotutkimusta. Tässä kirjallisuuskatsauksessa käytetään Mayerin ym. (1995) määrittelemää luottamuksen mallia luottamuksen saajan luotettavuuteen

vaikuttavista piirteistä (ABI-malli: kyvykkyys, hyväntahtoisuus ja rehellisyys). Käytössä on myös Lalotin ja Bertramin (2025) malli tekoälyn luotettavuuden muodostumisesta.

1.1 Tutkimuksen tarkoitus

Tutkimuksen tarkoitus on kuvata, miten tekoälyn tuottama sisältö voi heikentää kuluttajien luottamusta sosiaaliseen mediaan. Tutkimus koostuu kolmesta osaongelmasta:

(1) Miten tekoälyä käytetään sosiaalisen median sisällöntuotannossa?

(2) Mistä luottamus rakentuu sosiaalisessa mediassa?

sekä (3) Miten tekoälyn tuottama sisältö vaikuttaa kuluttajan luottamukseen?

1.2 Tutkimuksen rakenne

Tutkimus toteutetaan kirjallisuuskatsauksena. Tutkielmassa tarkastellaan tekoälyn erilaisia käyttötapoja sosiaalisessa mediassa sekä luottamusta institutionaalisen, interpersoonallisen sekä yksilöllisen luottamustaipumuksen näkökulmasta. Sovellamme myös niihin sopivia luottamusteorioita kuten Mayerin ym. (1995) institutionaalisen luottamuksen ABI-mallia, sekä Lalotin & Bertramin (2025) mallia luottamuksen muodostumisesta tekoälyyn. Näitä malleja sovelletaan sosiaalisen median ympäristössä. Tutkielma koostuu johdannosta, kolmesta asiakappaleesta sekä johtopäätöksestä ja yhteenvedosta. Johdannossa pohjustetaan ja esitellään tutkielman aihe. Ensimmäisessä asiakappaleessa käydään läpi tekoälyn eri muotoja sekä käyttötapoja sosiaalisessa mediassa. Toisessa asiakappaleessa käsitellään luottamuksen rakentumista sosiaalisessa mediassa. Kolmannessa asiakappaleessa käsitellään, miten tekoälyn tuottama sisältö vaikuttaa kuluttajan luottamukseen.

2 Tekoälyn luoman sisällön muodot ja käyttötavat sosiaalisessa mediassa

Tekoälyä, joka pystyy luomaan uutta sisältöä, kutsutaan generatiiviseksi tekoälyksi. Generatiivinen tekoäly voi luoda sisältöä, kuten tekstiä, kuvia, ääntä tai videoita (Feuerriegel ym. 2024). Tässä tutkielmassa käytetään kuitenkin termiä tekoäly. Tekoälyä käytetään tuottavuuden ja työn tehokkuuden parantamiseen monilla eri sektoreilla, myös sosiaalisen median sisältötuotannossa (Hewapathirana ym. 2024). Kun sosiaalisen median sisällön tarve kasvaa (Statista 2024), moni turvautuu tekoälyn apuun sisältötuotannossa. Tekoälyteknologia on myös laajassa julkisessa käytössä. Tässä kappaleessa käsitellään tekoälyn muotoja ja käyttötapoja sosiaalisessa mediassa. Syvävääreännökset, virtuaalivaikuttajat ja sosiaaliset botit ovat tekoälyn käyttömuotoja sosiaalisessa mediassa. Käyttötapoja taas huijaukset, väärän tiedon levittäminen ja markkinointi.

2.1 Sisällön muodot

Tekoällysisällön muotoja on tutkittu paljon erikseen, mutta ei yhdessä. Tässä tutkimuksessa käyttömuodot on määritelty eri tutkimuksista, joissa tarkastellaan tekoällysisältöä (esim. Gorwa & Guilbeault 2020; Conti & Tricomi 2022; Mustak 2023). Tässä tutkimuksessa käsitellään syvävääreännöksiä, virtuaalivaikuttajia ja sosiaalisia botteja.

2.1.1 Syvävääreännökset

”Deepfake” eli syvävääreännös tarkoittaa tekoällyllä luotua kuvaa, ääntä ja/tai videota, joka vaikuttaa aidolta. Deepfake-termi yhdistää sanat ”epäaito” ja ”syväoppiminen”, joka viittaa tekoälyn oppimismenetelmään (Mustak ym. 2023). Syvävääreännös on ilmiö, jossa tekoälyn avulla luodaan väärennettyjä digitaalisia sisältöjä, kuten kuvia, tekstiä, videoita ja ääntä, jotka voidaan syväoppimisen avulla tehdä erittäin realistisiksi (Karnouskos 2020). Mustak ym. (2023) kuvasivat syvävääreännöksiä vielä tarkemmin niin, että niissä näytetään jotain, mitä ei ole tapahtunut oikeasti. AI-teknologia, jolla syvävääreännöksiä luodaan, on helppokäyttöinen ja se on laajasti julkisessa käytössä (Westerlund 2019). Vuonna 2020 julkaistussa artikkelissa kerrotaan visuaalisten syvävääreännösten toimivan niin, että toisen henkilön kasvot siirretään

toiseen kuvaan tai videoon. Artikkelissa kuitenkin todetaan tekoälyteknologian kehittyvän jatkuvasti (Kietzmann ym. 2020). Westerlundin (2019) mukaan voidaan luokitella neljä pääryhmää, jotka tuottavat syvävääreännöksiä: harrastelijat, jotka tekevät syvävääreännöksiä pääasiassa viihteen vuoksi esimerkiksi meemejä, poliittiset tekijät kuten vieraan maan hallitus tai aktivistit, muut ilkeiden tekijät kuten internethuijarit sekä lailliset toimijat kuten elokuva-alan toimijat. Syvävääreännöksillä on mahdollisuus vaikuttaa yhteiskuntaan niin positiivisesti, kuin negatiivisesti, riippuen sen käytön tarkoituksesta.

Tunnetuin syvävääreännös on Karnouskoksen (2020) mukaan video Barack Obamasta varoittamassa syvävääreännöksistä. Tämä syvävääreännös oli Buzzfeedin luoma testi, jolla pyrittiin selvittämään tunnistavatko ihmiset syvävääreännöksen. Syvävääreännöksiä käytetään paljon mm. värentämään poliitikkojen puheita ja luomaan pornografista sisältöä esimerkiksi kuuluisista vaikuttajista tai julkisuuden henkilöistä. Sosiaalinen media antaa alustan syvävääreännöksille ja mahdollistaa niiden nopean leviämisen. (Karnouskoksen 2020.) Syvävääreännöksiä on jo nyt hankala erottaa aidoista videoista ja teknologian kehitys tulee tekemään siitä entistä vaikeampaa ja kenties jopa mahdotonta havaita ihmissilmillä. Hankalan erottamisen takia niitä voidaan käyttää kuluttajan huijaamiseen mm. poliittisessa tai markkinointitarkoituksessa. (Mustak ym. 2023.) Kun aidon ja epäaidon sisällön rajat sumenevat, saattaa tämä aiheuttaa luottamuksen puutetta sosiaalisen median käyttäjissä. Tätä vahvistavat sosiaalisen median algoritmit, jotka puskevat sisältöä käyttäjälle sen perusteella, millaisesta sisällöstä katsoja pitää, tällöin katsoja ei välttämättä ole yhtä kriittinen näkemästään sisällöstä, koska se vahvistaa tämän jo olemassa olevia ajatuksia ja mielipiteitä, mikä taas edistää syvävääreännöksiä vaikuttavuutta (Westerlund 2019). Sama toimii toisinpäin: sisältö, joka on aito, mutta ei sovi katsojan jo olemassa oleviin uskomuksiin, nähdään epäaitona (Westerlund 2019). Westerlund (2019) kuvaa artikkelissaan syvävääreännösten isoimpien uhkien olevan väärän tiedon leviäminen ja sen aiheuttama paine journalisteille, käyttäjien luottamuksen heikentyminen viranomaisilta tulevaan tietoon sekä tietoturvariskit niin organisaatioille kuin tavallisille kansalaisille.

2.1.2 Virtuaaliset vaikuttajat

Virtuaalivaikuttaja tai tekoälyvaikuttaja on henkilö tai asia, joka on luotu tietokoneohjelmistolla. Virtuaalivaikuttaja pystyy vaikuttamaan toisiin pääasiassa markkinoinnin tai yhteistöiden kautta tai osallistumalla kampanjoihin ja sitä kulutetaan yksinomaan virtuaalisesti. (Conti & Tricomi 2022). Toisen määritelmän mukaan virtuaalivaikuttajat ovat fiktiivisiä

tietokoneohjelmoituja yksilöitä, joilla on ihmismäisiä ominaisuuksia, piirteitä ja persoonallisuksia (Thomas & Fowler 2021). Virtuaalivaikuttajat pyrkivät jäljittelemään ihmisvaikuttajia myös ilmeillä, eleillä, käyttäytymisellä sekä vuorovaikutuksessa sosiaalisen median käyttäjien kanssa, esimerkiksi vastaamalla kommentteihin (Gutuleac ym. 2024). Virtuaalisten vaikuttajien juuret ovat jo 1980-luvulla, jolloin luotiin maailman ensimmäiseksi virtuaaliseksi vaikuttajaksi luokiteltu televisiohahmo Max Headroom, joka teki yhteistyötä mm. Coca-Colan kanssa (Conti & Tricomi 2022). Nykyiseen muotoonsa virtuaalivaikuttajat ovat muotoutuneet noin 2016, jolloin luotiin Lil Miquela, yksi tämän hetken tunnetuimmista tekoälyn avulla luoduista virtuaalivaikuttajista, joka luo sisältöä Instagramiin ja TikTokkiin. Eniten seuratuilla virtuaalivaikuttajilla on yhteenlaskettuna jo miljoonia seuraajia Instagramissa, tosin on arvioitu, että vain noin 60% VV seuraajista on aitoja. Isot brändit kuten Prada, Calvin Klein, Ikea, Chanel, Porsche ja Balmain ovat jo tehneet yhteistyötä isoimpien virtuaalivaikuttajien kanssa. (Conti & Tricomi 2022.) Monissa eri lähteissä kuvataan virtuaalivaikuttajien olevan vuorovaikutuksen mittareiden osalta samalla tasolla oikeiden vaikuttajien kanssa (Conti & Tricomi 2022; Gutuleac ym. 2024). Tunnettuja virtuaalivaikuttajia on kuitenkin vielä kahdella kädellä laskettava määrä ja ero oikeisiin ihmisvaikuttajiin on huomattava mm. seuraajamäärän suhteen, joten näitä on hankala vertailla. Kuka vain voi luoda virtuaalisen vaikuttajan, mutta käytännössä kaikki isoimmat virtuaalivaikuttajat ovat yritysten luomia ja omistamia ja virtuaalivaikuttajan yhteistöistä tienaamat rahat kuuluvat ne luoneille yrityksille. (Conti & Tricomi 2022.)

2.1.3 Sosiaaliset botit

Epäaidot sosiaalisen median käyttäjät tai automatisoidut sosiaalisen median tilit, eli puhekielessä ”botit” eivät ole uusi juttu ja niitä kutsutaan akateemisissa tutkimuksissa eri nimillä, kuten ”chatbot”, ”spam bot”, ”sosiaalinen bot”, ”poliittinen bot”, ”botnet” tai ”sybil”, joka on suosituin tietojenkäsittelytieteilijöiden teksteissä. ”Botti” on suosittu termi puhekielessä ja muiden alojen akateemikkojen keskuudessa. (Gorwa & Guilbeault 2020.) Näillä termeillä viitataan erilaisiin botteihin internetissä, joilla on eri tarkoitukset. Botteja, jotka tuottavat automaattisesti sisältöä (esim. tekstiä) ja on tarkoitettu kommunikoimaan ihmisten kanssa sosiaalisessa mediassa, kutsutaan sosiaalisiksi boteiksi (Gorwa & Guilbeault 2020). Sosiaalisia botteja, joiden tarkoitus on poliittinen häirintä, kutsutaan poliittisiksi boteiksi, mutta tässä

kirjallisuuskatsauksessa automatisoituja sosiaalisen median tilejä kutsutaan vain sosiaalisiksi boteiksi tai boteiksi käyttötarkoituksesta riippumatta.

2.2 Tekoälyn käyttötavat sosiaalisessa mediassa

Tekoälyn käyttötavat ovat moninaiset ja sitä voi käyttää niin positiiviseen kuin negatiiviseenkin tarkoitukseen. Tekoälyn käyttötapoja on useita, mutta tutkituimmat käyttötavat sosiaalisessa mediassa ovat huijaukset, väärän tiedon leviäminen sekä markkinointi (esim. Conti & Tricomi 2022; Mustak ym. 2023).

2.2.1 Huijaukset

Tekoälyn kanssa tehdyt huijaukset ovat lisääntyneet, siitä on uutisoinut mm. Helsingin sanomat 5.3.2025. Huijauksia tehdään usein mm. julkisuuden henkilöiden kasvoilla. Esimerkkinä Sointu Borgin kryptohuijaustapaus (Seiska 13.12.2024). Huijauksesta oli tehty Iltalehden artikkelin näköinen, jossa Sointu tai muu julkisuuden henkilö kertoo tienanneensa paljon rahaa sijoittamalla kryptoihin tai johonkin muuhun. Sitten ”lehtiartikkelissa” on linkki, joka vie huijaussivustolle. Muita huijauksia mitä näkyy sosiaalisessa mediassa usein, liittyy siihen, että käytetään asiantuntijan (esim. lääkäri) kasvoja mainostamaan jotain tuotetta. Tällaiset huijaukset tehdään usein näyttämään joko podcast formaatilta tai ”green screeniltä” eli asiantuntijan kasvot puhumassa toisen videon tai kuvan päällä. Käyttäjät ”Dr Kajan Rajanin” sekä ”Doctor Youn” ovat ottaneet Tiktokissa (Tiktok 13.4.2025; Tiktok 30.3.2025) kantaa siihen, että heidän kasvojaan on käytetty ilman lupaa mainostamaan lisäravinne tuotetta. Myös tämän tyyppiset markkinointi huijaukset yleistyvät. Myös tekoälypohjaisia positiivisia arvosteluja käytetään markkinointi huijauksena vaikuttamaan potentiaalisen asiakkaan mielikuvaan tuotteesta (Mustak ym. 2022).

2.2.2 Väärän tiedon levittäminen

Kenties vaarallisin tekoälyn käyttötapa on väärän tiedon levittäminen. Monet tutkijat pelkäävät väärän tiedon leviämistä ja sen aiheuttaman epätiedon vaikutuksia luottamukseen. (Vaccari & Chadwick 2020; Karnousko, 2020; Harris 2021; Mustak ym. 2023.) Väärää tietoa voidaan levittää esimerkiksi viihdyttämistarkoituksessa tai katsojien ja huomion ja vuorovaikutuksen saamiseksi sosiaalisessa mediassa. Tunteita herättävät tarinat, joista saadaan

uskottavampia kuvien ja videoiden avulla usein saavat paljon huomiota sosiaalisessa mediassa. Tämä on vielä suhteellisen harmitonta, sillä väärä tieto ei aiheuta erityistä harmia katsojalle. Haitallista tai jopa vaarallista on poliittisen väärän informaation sekä uutisten leviämisen (Gorwa & Guilbeault 2020; Khaund ym. 2022). Poliittinen häirintä tulee usein muista maista ja sen tarkoitus on vaikuttaa vaalien tuloksiin ja kansalaisten luottamukseen (Gorwa & Guilbeault 2020; Khaund ym. 2022).

2.2.3 Markkinointi

Tekoälyllä tehdyt mainokset ja markkinointikampanjat ovat herättäneet keskustelua sosiaalisessa mediassa. Vastaanotto on ollut niin positiivista, kuin negatiivistakin. Viimeisimpänä keskustelua on herättänyt pikamuotiyritys H&M:n aikomus luoda digitaaliset versiot 30 eri mallista (BBC 27.3.2025). Myös Cola-Cola on julkaissut tekoälyllä luotuja joulumainoksia (Forbes 15.12.2024). Tekoälyn avulla yritykset voivat personoida mainoksia, joka johtaa korkeaan vuorovaikutukseen (Shaikh ym. 2025). Tekoälyn luoma mainos on myös kustannustehokasta yritykselle (Shaikh ym. 2025), mikä saattaa osaltaan selittää yritysten suuren kiinnostuksen tekoälyn käyttöön mainonnassa. Virtuaalivaikuttajista kertovassa kappaleessa mainittiin niiden osallistuminen isojen yritysten markkinointikampanjoihin. Virtuaalisten vaikuttajien hyödyiksi brändin näkökulmasta nähdään työskentelyn joustavuus, brändin turvallisuus (kontrolloitava, ei skandaaleita), eksklusiivisuus (brändin itse luoma kasvo ei työskentele muille) ja innovatiivisuus. Haitoiksi nähdään epäaitous, virtuaalivaikuttajat eivät ole yhtä samaistuttavia, kuin ihmisvaikuttajat sekä epärealistiset odotukset ja korkea hinta. (Conti & Tricoli 2022.)

3 Luottamuksen rakentuminen sosiaalisessa mediassa

Luottamusta katsotaan eri tieteenaloilla eri näkökulmista. Ekonomit näkevät luottamuksen joko laskelmallisena tai institutionaalisenä, psykologit lähestyvät asiaa luottamuksen osapuolten ominaisuuksien näkökulmasta ja sosiologit toimivat olettamuksella, että sosiaalisen vuorovaikutuksen ominaispiirteet pysyvät samana (Rousseau ym. 1998). Rousseau ym. (1998) etsivät yhteneväisyyksiä eri tieteenalojen luottamuksen määritelmistä ja he päätyivät määrittelemään luottamuksen seuraavasti: ”Luottamus on psykologinen tila, joka koostuu aikomuksesta hyväksyä haavoittuvuus perustuen myönteisiin odotuksiin toisen aikomuksista tai käyttäytymisestä”. Tämä määritelmä on relevantti huolimatta siitä, onko luottamuksen osapuolet yksilöitä vai instituutio ja yksilö. Luottamukselle ei kuitenkaan ole mitään virallista yhtä määritelmää. Tässä kappaleessa käsitellään luottamuksen tyyppiä, luottamuksen kohteita sekä luottamuksen saajan ominaisuuksia.

3.1 Luottamuksen tyypit

Luottamuksen tyyppiä on useita ja niiden määritelmä sekä käyttötapa vaihtelee lähteen mukaan. Eri tutkimusten perusteella (Mayer ym. 1995; Ayaburi & Treku 2020; Lalot & Bertram 2025) käsitellään sosiaalisen median yhteydessä kolmea tyyppiä: 1) Henkilökohtaista luottamustaipumusta 2) interpersoonallista, eli ihmisten välistä luottamusta, sekä 3) institutionaalista luottamusta, missä luottamuksen saaja ei ole ihminen, vaan organisaatio tai muu instituutio tai systeemi. Henkilökohtainen luottamustaipumus on osana useita luottamusmalleja (esim. Mayer ym. 1995 ja Lalot & Bertram 2025) ja se on tärkeä syy sen takana, miksi yksilö luottaa tai ei luota. Luottamustaipumus on tärkeä varsinkin tarkastellessa luottamusta tuntemattomiin ihmisiin. Interpersoonallista luottamusta käsitellään siksi, että sosiaalisessa mediassa käyttäjien välinen vuorovaikutus on usein interpersoonallista, vaikka vuorovaikutus ei olekaan fyysistä. Institutionaalinen luottamus on mukana käsittelyssä, koska instituutioiden luotettavuuden on todettu vaikuttavan sosiaalisen median käyttäytymiseen (Ayaburi & Treku 2020; Zoonen ym. 2024). Jotkut lähteet esittävät (esim. Freitag & Traunmuller 2009) henkilökohtaisen luottamuksen olevan interpersoonallisen luottamuksen alla, mutta tässä tutkimuksessa käsittelemme niitä erikseen, sillä henkilökohtainen luottamustaipumus vaikuttaa vahvasti myös siihen, miten yksilö luottaa instituutioihin.

3.1.1 Henkilökohtainen luottamustaipumus

Henkilökohtainen luottamustaipumus tarkoittaa yleistä taipumusta luottaa toisiin, jopa tuntemattomiin, ja se riippuu yksilön yleisestä maailmankuvasta ja asenteesta muita kohtaan. Psykologisen näkökulman mukaan se on luonteenpiirre, joka ei muutu, vaikka luottamus petettäisiin (Freitag & Traunmuller 2009). Fu ym. (2024) mukaan, taipumuksen muodostuminen alkaa jo varhaislapsuudessa huoltaja-lapsisuhteen perusteella. Jos huoltaja vastaa lapsen tarpeisiin kiintymys suhteen muodostumisen aikana, saa lapsi jo tässä vaiheessa perusluottamuksen muita kohtaan. Tämä antaa pohjan luottamustaipumukselle, joka kehittyy vielä myöhemmässäkin iässä yksilön kokemusten perusteella. Tämä taipumus kertoo miksi jotkut yksilöt ovat luontaisesti toisia luottavaisempia. Fu ym. (2024.)

Henkilökohtainen luottamustaipumus on tärkeässä osassa, kun tarkastellaan luottamusta tuntemattomiin ihmisiin. Luottamus tuntemattomiin ihmisiin on tämän tutkielman kannalta tärkeä käsite, sillä luottamusta käsitellään sosiaalisen median miljöössä, missä yksilöt eivät usein tunne toisiaan. Myös sosiaalisessa mediassa yksilöt voivat muodostaa oikean elämän kaltaisia suhteita (Xu ym. 2012), mutta tässä tutkielmassa keskitytään kuitenkin pääasiassa tuntemattomien välisiin suhteisiin.

3.1.2 Interpersoonallinen luottamus

Interpersoonallinen eli henkilöiden välinen luottamus koostuu kognitiivisesta ja affektiivisesta luottamuksesta (Xu ym. 2012; Wang ym. 2016; Massey ym. 2019). Kognitiivinen luottamus muodostuu luottamuksen saajasta valmiiksi tiedetystä informaatiosta ja päätelmästä hänen kyvykkyydestään, luotettavuudestaan ja johdonmukaisuudestaan (Xu ym. 2012). Xu:n ym. (2012) dimensiot kyvykkyys, luotettavuus ja johdonmukaisuus ovat hyvin samanlaisia kuin Mayerin ym. (1995) kyvykkyys, hyväntahtoisuus ja rehellisyys dimensiot, jonka vuoksi ABI-malli sopii käsiteltäväksi myös interpersoonallisen luottamuksen kontekstissa. Affektiivinen luottamus tarkoittaa tunteisiin perustuvaa luottamusta ja se vahvistuu, kun luottamuksen osapuolet osoittavat aitoa huolta toisistaan (Massey ym. 2019). Affektiivista luottamusta on esimerkiksi identifikaatio eli se, että luottamuksen antaja kokee yhteenkuuluvuuden tunnetta luottamuksen saajan kanssa, kuten samat arvot tai kulttuurinen tausta. Identifikaatio voi auttaa luottamuksen muodostumisessa myös tuntematonta henkilöä kohtaan. Näiden lisäksi luottamusta parantaa toistuva positiivinen vuorovaikutus, esimerkiksi vuorovaikutus vaikuttajan tai

toisen käyttäjän kanssa. Sekä kognitiivisen että affektiivisen luottamuksen on huomattu refleктоivan osallistuvien osapuolten suhdetta myös virtuaalisessa ympäristössä (Xu ym. 2012). Henkilöiden välinen luottamus on kuitenkin toissijainen tekijä tarkastellessa luottamusta sosiaaliseen mediaan kokonaisuutena. Tutkimukset, jotka käsittelevät luottamusta heikentäviä tekijöitä sosiaalisessa mediassa eivät keskity käyttäjien välisiin suhteisiin (esim. Vaccari & Chadwick 2020; Ahmed 2023; Mustak ym. 2023). Tämän vuoksi interpersoonallisen luottamuksen käsittely tässä tutkimuksessa jää taka-alalle.

3.1.3 Institutionaalinen luottamus

Luottamusta, jossa luottamuksen toinen osapuoli ei ole henkilö, vaan esimerkiksi organisaatio, valtio tai systeemit, kutsutaan institutionaaliseksi luottamukseksi (Mayer ym. 1995; Rousseau ym. 1998; Kaasa & Andriani 2022). Institutionaalinen luottamus tarkoittaa uskoa siihen, että yhteiskunnalliset rakenteet ovat rakennettu niin, että ne parantavat yksilön mahdollisuuksia onnistua pyrkimyksissään (McKnight ym. 2002).

3.2 Luottamuksen kohteet

Kun käsitellään käyttäjän luottamusta sosiaaliseen mediaan ja sen sisältöä kohtaan, kohteina voidaan pitää ihmistä, instituutioita, sosiaalisen median sisältöä sekä tekoälyä. Luottamuksen kohteet on määritelty yhdistelemällä usean luottamusta käsittelevän tutkimuksen löydöksiä. Instituutiot ovat erillään sosiaalisen median ympäristöstä, mutta instituutioiden toiminta vaikuttaa sosiaalisen median ympäristön luotettavuuteen ja turvallisuuteen (Ayaburi & Treku 2020), minkä vuoksi se on määritelty luottamuksen kohteeksi. Ihminen puolestaan on määritetty kohteeksi, sillä sosiaalinen media on käyttäjien välistä vuorovaikutusta (Xu ym. 2012). Eri tekoälyn luoman sisällön muotoja, kuten syvävääreännöksiä, virtuaalivaikuttajia ja sosiaalisia botteja, on tutkittu paljon luottamuksen yhteydessä (esim. Khaundin ym. 2022; Mustak ym. 2023; Gutuleac ym. 2024). Näille yhteistä on, että ne on luotu tekoälyllä. Lalotin ja Bertramin (2025) mukaan tekoälyn ihmistyessä tekoäly itse on luottamuksen kohteena. Tekoäly ja sen luomaa sisältöä käsitellään erikseen, koska niiden luotettavuuteen vaikuttaa osittain eri tekijät ja siksi, että tekoäly voidaan nähdä ihmismäisenä, mutta sisältöä ei nähdä ihmismäisenä, paitsi virtuaalivaikuttajien tapauksessa.

3.2.1 Ihminen ja instituutiot

Interpersoonallisessa luottamuksessa ihminen on luottamuksen kohteena. Sosiaalisessa mediassa esimerkiksi luottamus muihin käyttäjiin, käytännössä tuntemattomiin. Instituutioiden tapauksessa taustalla on usein ihminen ja luottamus instituutioon saattaa riippua yksilön vuorovaikutuksesta, vaikka instituution asiakaspalvelijan kanssa. Instituutioiden kohdalla kuitenkin luottamuksen kohteena on yleisesti jokin epäinhimillinen (Mayer ym. 1995; Rousseau ym. 1998; Kaasa & Andriani 2022). Sosiaalisen median luotettavuudessa voidaan ajatella valtion ja sosiaalisen median alustojen omistajayritysten olevan luottamuksen kohteina (Ayaburi & Treku 2020). Tekoälyn tapauksessa käydään keskustelua siitä, onko luottamuksen kohteena tekoäly vai ihminen (lisää kappaleessa 3.2.3).

3.2.2 Sosiaalisen median sisältö

Sosiaalisen median (kuten Facebook, Instagram ja Tiktok) sisältö on nopeatempoista ja informaatiota on paljon, mikä tarkoittaa, että yhden julkaisun käsittelyaika on pieni. Tilanteissa, joissa informaatiota on paljon, tiedonkäsittelytapa vaihtuu heuristiseksi tiedonkäsittelyksi. Heuristinen tiedonkäsittely vie vähemmän aikaa, mutta ei huomioi kaikkea saatavilla olevaa tietoa. Heuristisen tiedonkäsittelyn tavoitteena on käsitellä paljon tietoa nopeasti ja tehokkaasti. (Ma ym. 2020.) Yleisimmät sosiaalisen median yhteydessä käsitellyt heuristiikat ovat vahvistusharha, auktoriteetti sekä suositteluheuristiikka (Lin ym. 2016; Kim 2018; Westerland 2019; Ma ym. 2020) Metzgerin ja Flanagin (2013) mukaan vahvistusharhaheuristiikka tarkoittaa sitä, että käyttäjä pitää todennäköisemmin luotettavana sisältöä, joka sopii heidän olemassa oleviin uskomuksiinsa. Auktoriteettiheuristiikka sen sijaan olettaa käyttäjän luottavan enemmän informaatioon, jonka jakaja on luotettava auktoriteetti, kuten vaikka viranomainen (Metzger & Flanagin 2013; Lin ym. 2016) Suositteluheuristiikka puolestaan viittaa siihen, että käyttäjä uskoo informaatiota, jos muutkin uskovat siihen (Metzger & Flanagin 2013). Sosiaalisen median yhteydessä tämä voi tarkoittaa mm. tykkäyksiä, kommentteja ja jakoja (Kim 2018). Kaikkien näiden heuristiikkojen vaikutus on vahvistettu mm. Linin ym. (2016), Kimin (2018) ja Man ym. (2020) tutkimuksissa.

3.2.3 Tekoäly -teknologia vai luoja?

Lalot & Bertram (2025) pyrkivät selvittämään luottamusta tekoälyyn tutkimalla tekoälyyn perustuvia chatbotteja. Heidän tutkimuspaperissaan nousee esiin tärkeä ongelma, kun tutkitaan luottamusta tekoälyyn: kohdistuuko luottamus tekoälyä itseään kohtaan vai sen luoja? Aikaisempi näkemys on ollut, että luottamus teknologiaan on erillinen luottamuksesta ihmisiin. Kuitenkin useat uudemmat tutkimukset viittaavat siihen, että tekoälyteknologia on niin ihmismäistä, että luottamusta tekoälyyn tulisi tarkastella intersoonallisen luottamuksen näkökulmasta (Gilliath ym. 2021; Lalot & Bertram 2025). Tutkijat eivät ole yksimielisiä siitä, koskeeko luottamus tekoälyä itseään vai sen luoja, mutta moni on sitä mieltä, että mitä ihmismäisemmäksi ja autonomisemmaksi tekoäly kehittyy, sitä vähemmän käyttäjät kiinnittävät huomiota tekoälyn luojaan ja enemmän itse tekoälyyn (Saffarizadeh ym. 2024; Lalot & Bertram 2025). Tekoälyn tekijän negatiivinen julkisuus saattaa kuitenkin olla poikkeus. ChatGPT:n omistava yritys on saanut negatiivista julkisuutta epäeettisten käytäntöjen vuoksi ja epäillään, että tämä saattaa vaikuttaa ChatGPT:n suosioon tulevaisuudessa. Aiheesta tarvitaan kuitenkin lisää tutkimusta. (Lalot & Bertram 2025.) Tässä tutkimuksessa seurataan Saffarizadehin ym. (2024) sekä Lalotin ja Bertramin (2025) löydöksiä ja oletetaan itse tekoälyn olevan luottamuksen kohteena.

3.3 Luottamuksen saajan piirteet

Luottamusta käsittelevissä tieteellisissä julkaisuissa oli nähtävissä samankaltainen kaava: luottamuksen saajan piirteet olivat merkittävässä roolissa monissa luottamusmalleissa riippumatta siitä, kuka tai mikä on luottamuksen kohteena. Luotettavuus perustuu luottamuksen saajan havaittuihin ominaisuuksiin (Mayer ym. 1995; Lalot & Bertram 2025). Lisäksi, luottamusta herättäviä piirteitä kuvailtiin usein samankaltaisesti, mutta käytetyt termit saattoivat vaihdella. Monessa mallissa, riippumatta luottamuksen kohteesta, oli käytetty pohjana Mayerin ym. (1995) ABI-mallia luottamuksen dimensioista. Mallin suosion vuoksi myös tässä tutkielmassa käytetään pohjana Mayerin ym. (1995) mallia, kun käsittelyssä on luottamusta herättävät piirteet. Lalotin ja Bertramin (2025) mukaan tekoälyn ollessa luottamuksen saajana, ABI-mallin lisäksi tärkeä vaikuttava tekijä siinä, miten käyttäjä kokee sen luotettavuuden, on tekoälyn ihmismäisyys.

3.3.1 ABI-malli: kyvykkyys, hyvántahtoisuus ja rehellisyys

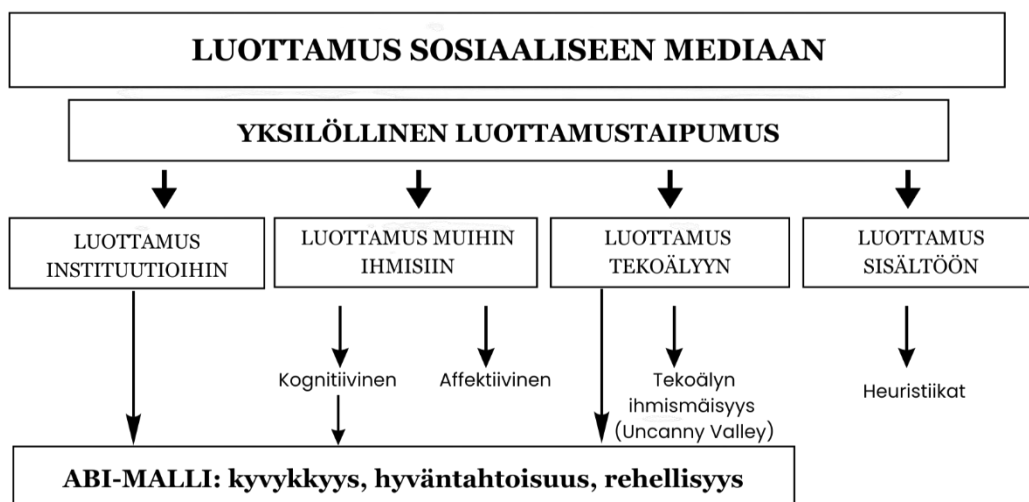
Mayer ym. (1995) kehittivät luottamuksen ABI-mallin (ability, benovelence, integrity), jonka mukaan luottamukseen vaikuttaa luottamuksen saajan kolme piirrettä: kyvykkyys, hyvántahtoisuus sekä rehellisyys. Tätä mallia on sittemmin käytetty laajasti pohjana luottamuksen tarkastelussa. Kyvykkyys viittaa luottamuksen saajan taitoihin, pätevyyteen ja ominaisuuksiin jollakin tietyllä alueella. Hyvántahtoisuus on luottamuksen antajan näkemys luottamuksen saajan positiivisista aikomuksista. Tämä viittaa siihen, että luottamuksen osapuolilla on jo ennestään jonkinlaista kanssakäymistä. Rehellisyys viittaa luottamuksen antajan näkemykseen siitä, miten hyvin luottamuksen saaja pitää kiinni periaatteista. Esimerkiksi johdonmukaisuuden puute voidaan nähdä rehellisyyttä alentavana. Toiminnan läpinäkyvyys lisää havaittua rehellisyyden tunnetta. (Mayer ym. 1995.)

Mayerin ym. (1995) malli sopii sovellettavaksi niin interpersoonallisen luottamuksen, institutionaalisen luottamuksen kuin myös tekoölyyn luottamisen tarkastelussa. Lalot ja Bertram (2025) määrittelivät mallin, jossa tekoölyn luotettavuuteen vaikutti tekoölyn kyvykkyys, hyvántahtoisuus ja rehellisyys (ABI-malli), tekoölyn ihmismäisyys sekä yksilön luottamustaipumus. Kuitenkin tutkimuksen tuloksissa kävi ilmi kyvykkyuden ja rehellisyyden olevan hyvántahtoisuutta tärkeämpiä dimensioita, kun tarkastellaan tekoölyyn kohdistuvaa luottamusta. ABI-mallia sovellettiin tekoölyn ihmismäisyyden vuoksi. Aikaisemmin teknologian luottamusta on tutkittu PPP-mallin (purpose, process ja performance) avulla. PPP-mallin dimensiot viittaavat enemmän teknologian tehokkuuteen ja toimintakykyyn. (Lalot & Bertram, 2025.)

3.3.2 Tekoölyn ihmismäisyys

Tekoölyn ihmismäisyys on tärkeä piirre sen luotettavuuden kannalta (Troshani ym. 2021; Lou ym. 2022; Gutuleac ym. 2024; Lalot & Bertram 2025). Tällaisia inhimillisiä piirteitä ovat esimerkiksi ääni, ulkonäkö, käytös ja tunteiden ilmaiseminen. Kohtelias käytös, aloitteellisuus keskustelussa kuten kysymysten esittäminen sekä ymmärryksen ja myötätunnon ilmaiseminen lisää käyttäjän luottamusta tekoölyä kohtaan. Myös nimen antaminen tekoölylle inhimillistää sitä ja lisää luottamusta. (Troshani ym. 2021.) Ihmismäisyys voi kuitenkin mennä liian pitkälle, jolloin käyttäjälle tulee epä mukava ja kammottava olo (”Uncanny Valley” -efekti, Mori 1970), ja tämä vähentää luottamusta (Troshani ym. 2021; Lou ym. 2022; Gutuleac ym. 2024).

Alla oleva Kuvio 1 havainnollistaa kappaleessa kolme läpikäytyä asiaa. Kuvio havainnollistaa luottamuksen muodostumisesta sosiaaliseen mediaan. Yksilön luottamustaipumus vaikuttaa kaikkiin alla oleviin luottamuksen kohteisiin (luottamus instituutioihin, muihin ihmisiin, tekoölyyn sekä sisältöön). Kohteiden alla on luotettavuuteen vaikuttavia asioita ja piirteitä. Kuvio on kooste useista erillisistä tutkimusartikkeleista. Yksilön luottamustaipumus sekä ABI-malli (tai jokin hyvin samanlainen malli) ovat toistuva teema useissa luottamukseen liittyvissä artikkeleissa (esim. Mayer ym. 1995; Xu ym. 2012; Ayaburi & Treku 2020; Lalot & Bertram 2025). Tekoölyn luotettavuus Kuviossa 1 perustuu Lalotin ja Bertramin (2025) malliin. Luottamus sosiaalisen median sisältöön on erillinen ABI-mallista, sillä vaikka luottamus voi kohdistua sisällön tekijään, silloin kyseessä on interpersoonallisesta luottamuksesta eli luottamuksesta toisiin ihmisiin.



Kuvio 1. Luottamuksen muodostuminen sosiaaliseen mediaan

4 Tekoälyn tuottaman sisällön vaikutus käyttäjän luottamukseen

Vaikka luottamuksesta ei olekaan universaalia määritelmää, tutkijat ovat yhtä mieltä siitä, että luottamiseen liittyy aina riski (Wang ym. 2016). Riski viittaa luottamuksen antajalle koituvan haittaan, mikäli luottamuksen saaja ei käyttäydykään odotetulla tavalla. Heikko tai heikentynyt luottamus on kuitenkin eri asia kuin täysi epäluottamus (Kang & Park 2017). Heikentynyttä luottamusta voi kuvata luottamusta signaloivien elementtien puuttumisena, kuten passiivisuudella, epäröinnillä sekä varmuuden ja uskon puutteella (Kang & Park 2017). Tässä kappaleessa käydään läpi tekoälyn eri muotojen vaikutusta luottamukseen, tekoälyn objektiivisuuden haasteita sekä instituutioiden roolia ja haasteita tekoälyn valvonnassa.

4.1 Tekoälyn eri muotojen vaikutus luottamukseen

Syväväärennosten, virtuaalivaikuttajien ja sosiaalisten bottien vaikutus sosiaalisen median käyttäjän luottamukseen.

4.1.1 Syvävääreännosten vaikutus luottamukseen

Visuaalinen efekti väärän tiedon yhteydessä parantaa sen uskottavuutta, mitä yksityiskohtaisempi ja selkeämpi, sitä positiivisempi vaikutus sillä on uskottavuuteen (Vaccari & Chadwick 2020; Hwang ym. 2021). Syvävääreännöksillä on siis tärkeä rooli väärän informaation uskottavuuden parantamisessa, sillä ne mahdollistavat aidolta tuntuvan visuaalisen efektin käytön informaation jakamisen yhteydessä. Ahmedin (2023) mukaan syvävääreännökset ovat erityisen hyviä väärän tiedon levittämiseen, sillä ne hyödyntävät realismiheuristiikkaa (asian on oltava totta, koska se näyttää siltä) sekä siksi, että syvävääreännöksien mahdollisuus heikentää myös aidon tiedon uskottavuutta. Tämä mahdollistaa esimerkiksi julkisuuden henkilöille sen, että he voivat kieltää heidät epäedullisessa valossa näyttävät kuvat tai videot. Vaccari ja Chadwick (2020) spekuloiivat mahdollisuutta siihen, että ihmisille tulee ajatus siitä, että totuudelta puuttuu pohja, sillä minkään aitoudesta ei voi olla varma.

Vaccari ja Chadwick (2020) tutkivat syvävääreännosten vaikutusta epävarmuuden tunteeseen ja tulokset viittasivat, harhaanjohtavalle syvävääreännökselle altistuminen lisää epävarmuuden tunnetta sosiaalisessa mediassa näytettäviä uutisia kohtaan. Kuitenkin opettavainen

syvävääreännös, jonka tarkoituksena oli varoittaa katsojia syvävääreännöksen vaaroista, paransi luottamusta sosiaalisen median uutisia kohtaan. Tutkimuksessa käytettiin BuzzFeedin opetus tarkoitukseen luomaa syvävääreännöstä Barack Obamasta, jossa ”Obama” kutsuu Trumpia ”dipshit” nimellä, jonka jälkeen hän varottaa syvävääreännöksistä. Osalle näytettiin leikattu video, missä näkyy vain Obama kutsumassa Trumpia ”dipshitiksi” ja osalle näytettiin koko video. Tutkimus toteutettiin Isossa-Britanniassa kysely muodossa ja vastaajia oli yli 2000. Myös toisessa Yhdysvalloissa tehdyssä tutkimuksessa (Ahmed 2023) todettiin jatkuvan syvävääreännöksille altistumisen ja niiden jakamisen epähuomiossa, sekä pelkästään syvävääreännöksistä huolehtimisen lisäävän skeptisyyttä sosiaalisen median uutisia kohtaan. Myös kognitiivisten kykyjen todettiin vaikuttavan siihen, kuinka helposti käyttäjä on syvävääreännösten huijattavissa. Korkeat kognitiiviset kyvyt omaava on skeptisempi, kuin matalat kyvyt omaava. Tässä tutkimuksessa oli mukana noin tuhat vastaajaa Yhdysvalloista, tutkimus toteutettiin kyselyn muodossa.

Kun luottamusta tutkitaan ABI-mallin (Mayer ym. 1995) pohjalta, yksi dimensio oli rehellisyys. Tutkimusten tulokset vahvistavat oletusta siitä, että syvävääreännösten käyttö epärehellisessä mielessä tai jos käyttäjä on kokenut tulleen huijatuksi esimerkiksi jakaessaan väärää tietoa heikentää käyttäjän luottamusta sosiaaliseen mediaan.

On kuitenkin mahdollista, että alentunut luottamus ei ole pelkästään negatiivinen asia. Hie-man vähentyneellä luottamuksella tai epävarmuudella on havaittu olevan jopa positiivisia vaikutuksia. Pieni epävarmuus on tervettä ja auttaa pitämään käyttäjän tietoisena esimerkiksi väärän sisällön mahdollisuudesta ja täten vähentää yksilön riskiä tulla huijatuksi (Vaccari & Chadwick 2020; Ahmed 2023). Kuitenkin, jos skeptisyys muuttuu täydeksi epäluottamukseksi, jolloin käyttäjä ei luota edes mediassa näkemäänsä oikeisiin uutisiin, tilanne voi olla haitallinen (Ahmed 2023). Medialukutaidon todettiin vähentävän syvävääreännösten negatiivisia vaikutuksia (Hwang ym. 2021).

4.1.2 Virtuaalivaikuttajien vaikutus luottamukseen

Luottamusta virtuaalivaikuttajiin on tutkittu pääasiassa markkinoinnin yhteydessä, sillä se on virtuaalivaikuttajien pääasiallinen käyttötapa. Virtuaalivaikuttajien luotettavuudelle on aikaisempien tutkimusten mukaan tärkeää, että virtuaalivaikuttajalla on ihmismäisiä ominaisuuksia tiettyyn pisteeseen asti (Lou ym. 2022). Monessa (Lou ym. 2022; Gutuleac ym. 2024)

virtuaalivaikuttajiin liittyvissä tutkimuksissa ja artikkeleissa mainittiin Morin (1970) ”Uncanny Valley” -teoria, jonka perusteella voidaan olettaa sosiaalisen median käyttäjien negatiivinen reaktio virtuaalivaikuttajiin, jos niiden ulkomuoto muistuttaa liian läheisesti ihmistä tai siitä tulee liian realistisen näköinen ja oloinen. Niiltä kuitenkin puuttuu jokin ihmisyydelle keskeinen piirre tai ominaisuus, luoden ”Uncanny Valley” -efektin. Tämä siis tarkoittaa sitä, että virtuaalivaikuttajan ihmismäinen olemus voi mennä rajan yli, jolloin käyttäjän negatiiviset tunteet virtuaalivaikuttajaa kohtaan ylittävät kiinnostuksen. Gutuleac ym. (2024) halusivat tutkia ilmiötä modernien virtuaalivaikuttajien suhteen ja löydöt mukailivat Morin (1970) teoriaa. Ihmismäisempi virtuaalivaikuttaja herättää katsojassa oudoksumisen ja kammon tunteita enemmän, kuin epärealistisempi virtuaalivaikuttaja, mikä johti haluttomuuteen seurata virtuaalivaikuttajaa. Kammutuksen tuntemuksia pystyi kuitenkin vähentämään, jos kuvissa tai videoissa oli mukana oikea tunnettu ihminen.

Conti ja Tricomi (2022) kysyivät 360 eri ikäisiltä ja eri maista olevilta vastaajilta kysymyksiä liittyen virtuaalisiin vaikuttajiin. Tavoitteena oli tutkia ihmisten asenteita virtuaalivaikuttajia kohtaan. Vain 16% seuraa edes yhtä virtuaalivaikuttajaa ja syyn kerrottiin pääasiassa olevan uteliaisuus virtuaalivaikuttajia kohtaan, eikä niinkään yhteisön luominen tai parasosiaaliset suhteet. Toistaiseksi vain 12% vastaajista luottaisi virtuaalivaikuttajaan yhtä paljon tai enemmän kuin ihmisvaikuttajaan. Samankaltaisia tuloksia saivat Lou ym. (2022) toteuttamassaan pienkokeessa, jossa he haastattelivat syvällisesti ihmisiä, jotka ovat seuranneet vähintään yhtä virtuaalivaikuttajaa. Otos oli pieni, vain 26 ihmistä Singaporesta. Tutkijat kertoivat jättäneensä otoksen pieneksi, sillä haastatteluista ei enää ilmennyt uusia havaintoja. Haastateltujen mukaan he eivät ole ostaneet mitään virtuaalivaikuttajan suosituksesta, sillä he kokevat virtuaalivaikuttajat epäluotettavammiksi, kuin oikeat vaikuttajat. Syyksi virtuaalivaikuttajien seuraamiseksi annettiin uteliaisuus virtuaalivaikuttajia kohtaan, teknologian kehitys, viihdyttävä, sekä/tai esteettinen sisältö. Vain pieni osa kertoi syyksi sosiaalisen vuorovaikutuksen, jos virtuaalivaikuttaja omaa samankaltaisia mielenkiinnon kohteita ja/tai kannattaa samoja arvoja.

Vaikka käyttäjät luottavatkin virtuaalivaikuttajien suosituksiin vähemmän, kuin ihmisvaikuttajien suosituksiin, eivät virtuaalivaikuttajat ole suuri uhka yleiselle luottamukselle. Virtuaalivaikuttajien profiileihin ja julkaisuihin on tarkoin merkitty se, että he eivät ole aitoja, jolloin sosiaalisen median käyttäjä on tietoinen vaikuttajan epäaitoudesta. Tällöin käyttäjää ei johdeta harhaan ja voidaan olettaa, että tämä vaikuta käyttäjän luottamukseen sosiaalista mediaa kohtaan.

4.1.3 Sosiaalisten bottien vaikutus luottamukseen

Sosiaaliset botit levittävät vihaa tai väärää informaatiota ja isona armeijana ne kykenevät vaikuttamaan käyttäjien mielipiteisiin tykkäämällä, jakamalla ja kommentoimalla julkaisuihin (Khaund ym. 2022). Sosiaalisten bottien poliittinen häirintä eri maissa on jo tuttua viime vuosikymmenen ajalta. Eri lähteissä kuvataan poliittisen häirinnän tulevan ulkomailta (esim. Yhdysvaltojen presidentinvaaleissa 2016 uskottiin olevan Venäjältä lähtöisten sosiaalisten bottien häirintää ja vaikutusta) ja sen tarkoituksena on jakaa kansaa ja levittää väärää tietoa (Gorwa & Guilbeault 2020; Khaund ym. 2022). Khaundin ym. (2022) mukaan sosiaalisilla boteilla oli suuri rooli mm. Covid-19-pandemian aikana useilla sosiaalisen median alustoilla. Samassa lähteessä kerrottiin sosiaalisten bottien isoimpia häirinnän kohteita olevan politiikka ja terveys. Esimerkiksi Twitter (nyk. X) on tunnettu laajasta bottiaktiivisuudestaan, mutta bottien aktiivisuus ei ole rajattu vain Twitteriin, vaan niitä on kaikilla sosiaalisen median alustoilla (Gorwa & Guilbeault 2020), eikä sitä ole yrityksistä huolimatta saatu kuriin.

Voidaan siis todeta, että sosiaaliset botit ovat huomionarvoinen uhka: ne levittävät epäuskoa sosiaalisen median alustoilla ja harhauttavat käyttäjien huomion pois tärkeistä aiheista. Syvävääreennösteknologian kehittyessä voi olla mahdollista luoda vieläkin uskottavampia epäaitoja käyttäjiä eri sosiaalisen median alustoille. Sosiaalisen botin tilillä voisi olla tekoälyllä luotuja kuvia tai videoita, jotka saavat sen näyttämään aidolta ihmiseltä, jolloin epäaidon käyttäjän tunnistaminen olisi entistä vaikeampaa sosiaalisen median käyttäjille. Tämän lisäksi sosiaalisten bottien kehittyessä, ne oppivat välttämään algoritmeja ja muita työkaluja, joiden tarkoitus on havaita bottien toimintaa (Khaundin ym. 2022) tehden niiden havaitsemisesta entistä vaikeampaa.

4.2 Tekoälyn objektiivisuuden haasteet

Tutkijat ovat varoittaneet, että tekoäly osoittaa puolueellisuutta sen perusteella, millaista materiaalia on käytetty sen kouluttamiseen (Lalot & Bertram 2025). Yle on uutisoinut Venäjän propagandan saastuttavan tekoälymalleja (jutussa käytetty esimerkkinä ChatGPT:tä). ChatGPT oppii netissä olevasta informaatiosta, mikä tarkoittaa, että jos netissä on väärää tietoa, myös tekoäly oppii väärin. Tätä venäläinen propaganda käyttää hyväkseen, Ylen mukaan

heillä on useita propagandasivustoja eri kielillä. (Yle 12.4.2025.) Tekoälymallit vaikuttavat myös olevan esimerkiksi poliittisesti puolueellisia, jolloin tekoäly saatetaan nähdä epärehellisenä, erityisesti yksilöiden toimesta, joilla on vastakkainen mielipide. Tekoälyn on myös raportoitu myötäilevän käyttäjää, joka antaa epäjohdonmukaisen ja epärehellisen kuvan. (Lalot & Bertram 2025.) Instagram-käyttäjät ovat huomanneet ChatGPT:n uuden mallin 4o myötäilevän käyttäjää jopa vaarallisen paljon (Instagram julkaisu 28.4.2025). Käyttäjät testasivat kuinka pitkälle ChatGPT myötäilisi heidän sanomaansa, ja he kuvailivat kuinka ChatGPT myötäili ja vahvisti mitä vain he sanoivat. Esimerkkeiksi oli nostettu erään käyttäjän yritysidea ”paskaa tikun päässä”, jota ChatGPT kuvaili nerokkaaksi. Toinen käyttäjä kertoi ChatGPT:lle olevansa seuraava messias ja Jumalan valittu. Tähän ChatGPT oli vastannut syillä, miksi se uskoo käyttäjän olevan oikeassa. (Instagram 28.4.2025.) Tällaisella myötäilyllä on potentiaalia olla hyvin vaarallista ja jopa eristää käyttäjää omasta lähipiiristään. Tekoälyn tapauksessa tärkeimmät luotettavuuden muodostumisen dimensiot ovat kyvykkyys ja rehellisyys (Lalot & Bertram 2025). Tekoälyn puolueellisuus ja tiedon saastumisen mahdollisuus viittaavat tekoälyn kyvyttömyyteen. Myötäileminen viittaa epärehellisyyteen.

4.3 Instituutioiden rooli ja haasteet tekoälyn valvonnassa

Instituutioilla kuten valtiolla ja mediayrityksillä on suuri rooli, mihin suuntaan käyttäjien luottamus sosiaalista mediaa ja sen sisältöä kohtaan kehittyy. Sosiaalisen median käyttäjät tarvitsevat luottamusta valtioon, ja järjestelmiin, joiden on tarkoitus suojella heitä. Lait ja niiden valvonta ovat tärkeässä osassa ihmisten luottamuksen tukemisessa. EU on jo säännellyt lakeja tekoälyn rajoittamiseksi, ja sosiaalisessa mediassa tulee merkitä, jos sisällön tekemisessä on käytetty tekoälyä (European parliament 8.6.2023). Tässä tulee esiin kaksi ongelmaa: 1) ovatko rangaistukset tarpeeksi haittaa aiheuttavia, jotta haitta ylittää huijaamisesta saadun hyödyn, ja 2) lakien ja säännösten valvominen on vaikeaa, sillä sosiaalinen media ulottuu maiden rajojen yli ja esimerkiksi Suomen viranomaisten valta ei ulotu muihin maihin. Sääntöjen valvominen jää täysin sosiaalisen median alustojen harteille. Sosiaalisen median alustat ovat taistelleet sosiaalisten bottien kanssa vuosia, eikä vaikuta siltä, että sosiaalisen median alustat ovat kyenneet poistamaan tätä ongelmaa (Orabi ym. 2020). Syvävääreännösten tunnistamisen hankaloituminen tulee aiheuttamaan samankaltaisia ongelmia sosiaalisen median alustoille. Sosiaalisen median alustoilla tulisi olla motivaatio kehittää teknologioita, jotka kykenevät tunnistamaan tekoälypohjaiset sisällöt ja käyttäjät. Tällä hetkellä on kyseenalaista,

onko viranomaisilla ja sosiaalisen median alustoilla tarpeeksi kyvykkyyttä pitämään sosiaalisen median käyttäjät turvassa tekoälyn luoman sisällön haitoilta.

Käyttäjät arvioivat yritysten ja viranomaisten hyvántahtoisuutta, eli sitä, uskovatko he näiden tahojen toimivan heidän etujensa mukaisesti, vai meneekö esimerkiksi teknologiayrityksillä raha ja oma etu käyttäjien edun edelle. Maassa tai alueella, jossa luottamus viranomaisiin on korkea, on todennäköisemmin korkeampi luottamus myös tekoälyyn. Reddit-yhteisön perustaja on myöntänyt luoneensa epäaitoja käyttäjiä Reddit-alustalle, jotta näyttäisi siltä, että alusta olisi aktiivisempi, kuin se on (Ars Technica 22.6.2012; The Economic Times 20.3.2025). Vaikka ne eivät olleet sosiaalisia botteja, vaan perustaja loi tilejä manuaalisesti, tämä herättää kysymyksen siitä, hyötyvätkö sosiaalisen median alustojen omistajat epäaitoista käyttäjistä ja olisiko tilanne sama myös muun haitallisen sisällön kanssa.

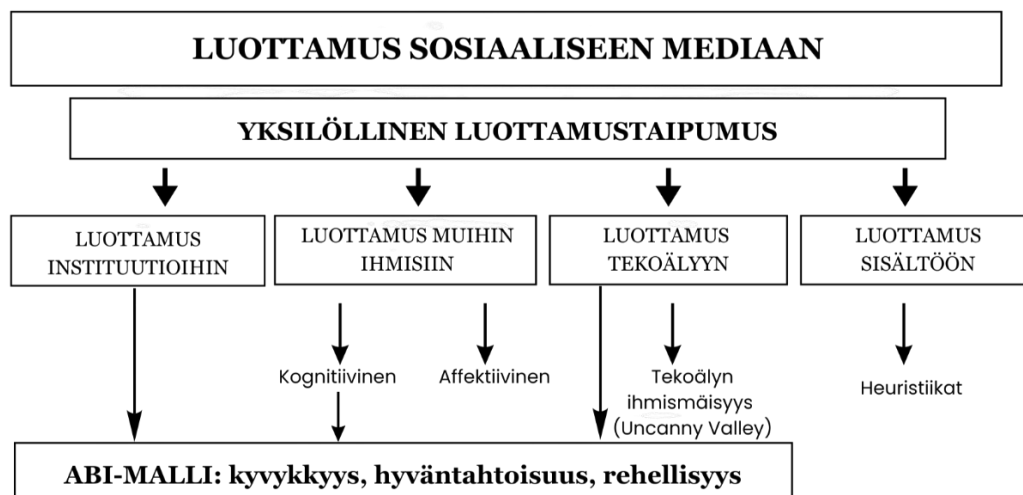
Myös johdonmukainen ja avoin toiminta lisää luottamusta. Se viestii rehellisyydestä eli siitä, että luottamuksen antaja voi luottaa toisen osapuolen pitävän lupauksensa ja toimivan oikeudenmukaisesti (Wang ym. 2016). Facebookin käyttäjien turvallisuudesta keskusteltiin vuonna 2014, kun käyttäjien tietoja päätyi väärille tahoille. Ayaburi ja Treku (2020) tutkivat tapausta ja Facebookin reaktion vaikutuksia käyttäjän luottamukseen. Tulokset osoittivat johdonmukaisen ja rehellisen viestinnän olevan tärkeämpiä elementtejä luottamuksen palautumiselle.

5 Johtopäätökset

Tutkielman tarkoituksena oli tarkastella, miten tekoälyn tuottama sisältö voi heikentää kuluttajien luottamusta sosiaaliseen mediaan. Tutkimuskysymyksen pyrittiin vastaamaan käyttämällä kolmea osaongelmaa: (1) Miten tekoälyä käytetään sosiaalisen median sisällöntuotannossa? (2) Mistä luottamus rakentuu sosiaalisessa mediassa? sekä (3) Miten tekoälyn tuottama sisältö vaikuttaa kuluttajan luottamukseen?

Tekoälyä käytetään moniin eri tarkoituksiin, ja sen käyttötavat muuttuvat ja lisääntyvät jatkuvasti. Tekoälyn käyttömuotoja ovat syvävärennökset, virtuaalivaikuttajat ja sosiaaliset botit. Se, mihin tarkoitukseen näitä käytetään, riippuu käyttäjästä. Pahaa haluava yksilö tai organisaatio voi väärinkäyttää tekoälyä väärän tiedon levittämisessä tai huijaamisessa, mutta tekoälystä saa apua myös sisällöntuotantoon tai markkinointiin. Tekoälypohjainen markkinointi on saanut osakseen paljon huomiota, niin negatiivista, kuin positiivistakin.

Tässä tutkielmassa määriteltiin käyttäjän luottamuksen sosiaalista mediaa kohtaan muodostuvan: 1) luottamuksesta instituutioihin, 2) luottamuksesta muihin ihmisiin, 3) luottamuksesta tekoölyyn ja 4) luottamuksesta sosiaalisen median sisältöön. Henkilökohtainen luottamustaipeumus vaikuttaa siihen, kuinka todennäköisesti yksilö luottaa näihin. Luottamuksen saajan piirteet ovat toinen tärkeä luottamuksen muodostumiseen vaikuttava tekijä. Tekoälyn tapauksessa ihmismäisyydellä on suuri rooli. Sosiaalisen median sisällön tapauksessa yhden julkaisun katsomisaika on lyhyt, jolloin mielen oikotiet, eli heuristiikat ovat suuressa roolissa käyttäjän luottamuksen muodostumisessa. Kuviossa 1 on havainnollistettu luottamuksen muodostumista sosiaaliseen mediaan.



Kuvio 1. Luottamuksen muodostuminen sosiaaliseen mediaan

Tekoälyn käyttötarkoitus vaikuttaa eniten siihen, miten se tulee vaikuttamaan käyttäjien luottamukseen. Tekoälyteknologiaa on käytetty epärehellisin aikomuksin sosiaalisessa mediassa, erityisesti huijauksiin ja väärän tiedon levittämiseen, jolloin luottamus laskee. Luottamukseen vaikuttaa myös pelkkä mahdollisuus huijatuksi tulemisesta. Tekoälyn puolueellisuus ja saastuminen vähentää luottamusta tekoälyyn. Sosiaalisen median alustojen rooli luottamuksen säilyttämisessä on tärkeä, sillä käyttäjien tulee uskoa niiden kykyyn suojella heitä sekä valvoa sääntöjä, joilla vähennetään tekoälyn väärinkäytön haittoja. Toimijoiden tulee myös tiedottaa ja toimia johdonmukaisesti ja rehellisesti, sekä pitää käyttäjien edut omien etujen edellä.

Markkinoinnin kannalta on tärkeää, että yritykset ovat avoimia tekoälyn käytöstään ja merkitsevät sisällön selkeästi. Tekoälyteknologialla on potentiaalia säästää pieniltä yrityksiltä suuria summia rahaa markkinointikuluissa. Virtuaalivaikuttajat ovat näyttäneet potentiaalia markkinoinnissa huomion herättäjinä, mutta eivät niinkään myyntipotentiaalin osalta. Sosiaalisen median käyttäjät ovat jopa uteliaita tekoälyteknologian suhteen, kuten virtuaalivaikuttajat ovat todistaneet, ja tekoälyn käyttö esimerkiksi mainoskampanjoissa tuo yritykselle paljon näkyvyyttä. Epäeettiset käytännöt kuten tekoälyn käytön salaaminen puolestaan vähentävät luottamusta. Tekoälyä käyttävien markkinoijien tulee kiinnittää erityistä huomiota sisällön heuristiikkoja vahvistaviin elementteihin sekä sisällön rehellisyyteen.

Tutkimuksen rajoituksena voidaan pitää, että luottamuksen muodostumista sosiaaliseen mediaan on käsitelty vain osissa, eikä kokonaisuudessaan. Luottamusteorioiden sopivuutta on testattu sosiaalisessa mediassa, mutta yhtenäinen tutkimus eri osa-alueista, jotka vaikuttavat luottamuksen muodostumiseen sosiaalisessa mediassa tuntuu puuttuvan. Tekoälyn käyttöä on tutkittu sosiaalisessa mediassa pääasiassa negatiivisesta näkökulmasta, kuten poliittinen manipulaatio. Tämän takia ymmärrys tekoälyn käytöstä sosiaalisessa mediassa voi olla vääristynyt. Luottamuksen kohde tekoälyn yhteydessä kaipaa vielä lisätutkimuksia, sillä siitä ei olla päästy yhteisymmärrykseen tutkijoiden kesken.

6 Yhteenveto

Tässä tutkielmassa käsiteltiin tekoälyn luomaa sisältöä sosiaalisessa mediassa ja sen vaikutusta sosiaalisen median käyttäjän luottamukseen sosiaalista mediaa ja sen sisältöä kohtaan. Tutkimuksen tarkoituksena oli kuvata, miten tekoälyn tuottama sisältö voi heikentää kuluttajien luottamusta sosiaaliseen mediaan. Tutkimukseen pyrittiin vastaamaan kolmen osaongelman avulla: (1) Miten tekoälyä käytetään sosiaalisen median sisällöntuotannossa?

(2) Mistä luottamus rakentuu sosiaalisessa mediassa?

Sekä (3) Miten tekoälyn tuottama sisältö vaikuttaa kuluttajan luottamukseen?

Osaongelmiin pyrittiin vastaamaan soveltamalla Mayerin ym. (1995) mallia luottamuksen saajan piirteistä (ABI-malli) sekä Lalotin ja Bertramin (2025) mallia luottamuksesta tekoälyyn. Tutkimuksessa tarkasteltiin tekoälyn eri muotoja ja käyttötapoja sosiaalisessa mediassa sekä erilaisia luottamuksen tyyppisiä sosiaalisen median miljöössä. Luottamuksen yhteydessä käytiin läpi luottamuksen kohteita sosiaalisessa mediassa sekä niiden luottamusta herättäviä piirteitä. Lopuksi tarkasteltiin tekoälyn muotojen vaikutuksia käyttäjän luottamukseen sekä instituutioiden roolia sosiaalisen median luotettavuudessa.

Tutkielmassa yhdistettiin luottamukseen, sosiaaliseen mediaan ja tekoälyyn liittyvää kirjallisuutta ja näiden avulla määrittyi, että luottamus sosiaaliseen mediaan muodostuu 1) luottamuksesta instituutioihin, 2) luottamuksesta muihin ihmisiin, 3) luottamuksesta tekoälyyn ja 4) luottamuksesta sosiaalisen median sisältöön. Näiden takana vaikuttaa yksilön henkilökohtainen luottamustaipumus, jonka muodostuminen alkaa jo varhaislapsuudessa, ja joka ei muutu. Tekoälyn piirteet kyvykkyys ja rehellisyys sekä tekoälyn ihmismäisyys ovat tärkeimmät tekoälyn luotettavuuteen vaikuttavat tekijät. Tutkimuksessa kävi ilmi, että suurin potentiaalinen uhka luottamukselle ovat syväväärrennökset ja sosiaaliset botit, sillä niitä väärinkäytetään usein mm. huijaamiseen tai väärän tiedon levittämiseen. Virtuaalivaikuttajilla ei vaikuttanut olevan vaikutusta yleiseen epäluottamukseen sosiaalista mediaa kohtaan, vaikka niihin luotetaankin ihmisvaikuttajia vähemmän. Tekoälyn potentiaalinen saastuminen viittaa, ettei algoritmi ole tarpeeksi kyvykäs tunnistamaan väärää tietoa ja propagandaa, mikä vähentää luottamusta tekoälyä kohtaan. Instituutioiden rooli on merkittävä yleisen luottamuksen ylläpitämisessä sosiaalista mediaa kohtaan tekoälyn väärinkäyttäjien yleistyessä.

Tutkielma tuo yhteen luottamukseen ja tekoölyyn liittyvää kirjallisuutta sosiaalisen median kontekstissa. Tutkielman tulisi auttaa markkinoijia ymmärtämään tekoölyn käytön vaikutuksia käyttäjien luottamukseen sosiaalisen median ympäristössä. Tutkielma voi auttaa markkinoijia arvioimaan omaa tekoölyn käyttöään ja auttaa välttämään luottamusta heikentävien tekijöiden käyttöä. Tutkielma voi myös auttaa ymmärtämään laajempaa uhkaa tekoölysisällön vaikutuksista sosiaalisen median luotettavuuteen.

Lähteet

- Ahmed, S. (2023) Navigating the maze: Deepfakes, cognitive ability, and social media new skepticism. *New media & society*, Vol. 25 (5), 1108-1129.
- Ayaburi, E. W. – Treku, D. N. (2020) Effect of penitence on social media trust and privacy concerns: The case of Facebook. *International Journal of Information Management*, Vol. 50, 171-181.
- Ars Technica 22.6.2012 Reddit founders made hundreds of fake profiles so site looked popular. <https://arstechnica.com/information-technology/2012/06/reddit-founders-made-hundreds-of-fake-profiles-so-site-looked-popular/?utm_source=chatgpt.com>, haettu 2.5.2025.
- BBC 27.3.2025 H&M to use digital clones of models in ads and social media. <<https://www.bbc.com/news/articles/c3vvg73xnde0>>, haettu 3.5.2025.
- Conti, M. – Tricomi, P. P. P. (2022) Virtual Influencers in Online Social Media. *IEEE Communications Magazine*. Vol. 60 (8), 1-13.
- European Parliament 8.6.2023 EU AI Act: first regulation on artificial intelligence. <<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>>, haettu 2.5.2025.
- Feuerriegel, S. – Hartmann, J. – Janiesch, C. (2024) Generative AI. *Business & Information Systems Engineering*, Vol. 66 (1), 111-126.
- Forbes 16.12.2024 Coca Cola's AI-Generated Ad Controversy, Explained. <<https://www.forbes.com/sites/danidiplacido/2024/11/16/coca-colas-ai-generated-ad-controversy-explained/>>, haettu 3.5.2025.
- Freitag, M. – Traummüller, R. (2009) Spheres of trust: An empirical analysis of the foundations of particularized and generalized trust. *European Journal of Political Research*, Vol. 48 (6), 782-803.

- Fu, C. – Yang, S. – Zhai, M. – Yong, T. – Zheng, C. – Ma, X. – Hou, G. – Su, P. (2024) The component and structure of interpersonal trust. *Heliyon*, Vol. 10 (9), e30071.
- Gillian, W. – Bertsch, A. – Sawe, E. – Lee, D. – Wolfe, T. – Mayer, J. – Engel, J. – Uepati, N. F. (2018) Who trusts social media? *Computers in Human Behaviour*, Vol. 81, 303-3115.
- Gilliath, O. – Ai, T. – Branicky, M., S. – Keshmiri, S. – Davison, R. B. (2021) Attachment and trust in artificial intelligence. *Computers in Human Behaviour*, Vol. 115 (1), 106607.
- Gorwa, R. – Guilbeault, D. (2020) Unpacking the Social Media Bot: A Typology to Guide Research and Policy. *Policy & Internet*, Vol. 12 (2), 225-248.
- Gutuleac, R. – Baima, G. – Rizzo, C. – Bresciani, S. (2024) Will virtual influencers overcome the uncanny valley? The moderating role of social cues. *Psychology & Marketing*, Vol. 41 (7), 1419–1431.
- Harris, K. R. (2021) Video on demand: what deepfakes do and how they harm. *Synthese*, Vol. 199, 13373-13391.
- Hewapathirana, N. – Piyumali, M. – Bandara, S. U. (2024) A Conceptual Review on Effectiveness of AI-generated Content on Customer Experience: Special Reference to Social Media Platforms. 15th *International conference on business and information*, ISSN 2961-5283.
- Helsingin Sanomat 5.3.2025 Tekoälyhuijaukset räjähtäneet käsiin -Yksi somealusta yli muiden. <<https://www.hs.fi/alueet/art-2000011065648.html>>, haettu 13.4.2025.
- Hwang, Y. H. – Ryu, Y. J. – Jeong, S. (2021) Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education. *Cyberpsychology, Behaviour, and Social Networking*, Vol. 24 (3), 188-193.

Instagram julkaisu 28.4.2025 <https://www.instagram.com/p/DI-FWDzNeAi/?img_index=8&igsh=MWhiOXJoZDJscnRwMA==>, haettu 4.5.2025.

Kang, M. – Park, Y. E. (2017) Exploring trust and distrust as conceptually and empirically distinct constructs: Association with symmetrical communication in public engagement across four pairings of trust and distrust. *Journal of public relations research*, Vol. 29 (2), 114-135.

Khaund, T. – Kirdemir, B. – Agarwal, N. – Liu, H. – Morstatter, F. (2022) Social Bots and Their Coordination Durin Online Campaigns: A Survey. *IEEE Transactions on Computational Social Systems*, Vol. 9 (2), 530-545.

Karnouskos, S. (2020) Artificial Intelligence in Digital Media: The Era of Deepfakes. *IEEE transactions on technology and society*, Vol. 1 (3), 138-147.

Kietzmann, J. – Lee, W. L. – McCarthy, P. I. – Kietzmann, T. C. (2020) Deepfakes: Trick or treat? *Business horizon*, Vol. 63 (2), 135-146.

Kim, J. W. (2018) They liked and shared; Effects of social media virality metrics on perceptions of message influence and behavioral intentions. *Computers in Human Behaviour*, Vol. 84, 153-161.

Lalot, F. – Bertram, A. (2025) When the Bot Walks the Talk: Investigating the Foundations of Trust in an Artificial Intelligence (AI) Chatbot. *Journal of Experimental Psychology: General*, Vol. 154 (2), 533-55.

Lin, X. – Spence, P. R. – Lachlan, K. A. (2016) Social media and credibility indicators: The effect of influence cues. *Computers in human behaviour*, Vol. 63, 264-271.

Lou, C. – Kiew, S. T. J. – Chen, T. – Lee, T. Y. M. – Ong, J. E. C. – Phua, Z. (2022) Authentically Fake? How Consumers Respond to the Influence of Virtual Influencers. *Journal of Advertising*, Vol. 52 (4), 540–557.

- Massey, G. – Wang, P. Z. – Kyngdon, A. S. (2019) Conceptualizing and modeling interpersonal trust in exchange relationships: The effects of incomplete model specification. *Industrial marketing management*, Vol 76, 60-71.
- Ma, C. – Au, N. – Ren, L. (2020) Biased minds experience improved decision-making speed and confidence on social media: a heuristic approach. *Information Technology & Tourism*, Vol. 22 (4), 593-634.
- Mayer, R. C. – Davis, J. H. – Schoorman, D. (1995) An Interactive Model of Organizational Trust. *The Academy of Management Review*, Vol. 20 (3), 709-734.
- Metzger, M. J – Flanagin, A. J. (2013) Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, Vol. 59 (B), 210-220.
- McKnight, H. D. – Choudhury, V. – Kacmar, C. (2002) Developing and Validating Trust Measures for e-commerce: An Integrative Typology. *Information Systems Research*, Vol. 13 (3), 334-359.
- Mori, M. (1970) The Uncanny Valley. *Energy*, Vol. 7 (4), 33-35.
- Mustak, M. – Salminen, J. – Mäntymäki, M. –Rahman, A. – Dwivedi, Y. K. (2023) Deep-fakes: Deceptions, mitigations, and opportunities. *Journal of Business Research*, Vol. 154, 113368.
- Rousseau, D. M. – Sitkin, S. B. – Burt, R. S. – Camerer, C. (1998) Not so different after all: A crossdiscipline view of trust. *The Academy of Management Review*, Vol. 23 (3), 393-404.
- Saffarizadeh, K. – Keil, M. – Maruping, L. (2024) Relationship between trust in the AI creator and trust in AI systems: the crucial role of AI alignment and steerability. *Journal of management information systems*, Vol. 41 (3), 645-681.

Seiska 13.12.2024 Sointu Borg ärähtää: “En ole rikastunut kryptoilla!” <
<https://www.seiska.fi/kotimaa/sointu-borg-arahtaa-en-ole-rikastunut-kryptoilla/1552493>>, haettu 3.5.2025

Statista (2024) Social Media Advertising -Worldwide.
 <https://www.statista.com/outlook/amo/advertising/social-media-advertising/worldwide?currency=USD#revenueSocialMedia_112314>, haettu
 13.4.2025.

Shaik, A. – Fatima, N. – Bibi, M. – Awan, A. N. (2025) The Role of Artificial Intelligence in Personalizing Social Media Advertising: Benefits, Challenges, and Future Directions. *International Journal of Trends and Innovations in Business & Social Sciences*, Vol.3 (1), 104-114.

The Economic times 20.3.2025 Reddit ‘faked its first users’: Resurfaced video shows co-founder Alexis Ohanian admitting 99% of early submissions were fabricated.
 <https://economictimes.indiatimes.com/magazines/panache/reddit-faked-its-first-users-resurfaced-video-shows-co-founder-alexis-ohanian-admitting-99-of-early-submissions-were-fabricated/articleshow/119263428.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst>, haettu 2.5.2025.

Thomas, V. L. – Fowler, K. (2021) “Close Encounters of the AI Kind: Use of AI Influencers as Brand Endorsers.” *Journal of Advertising*, Vol. 50 (1), 11–25.

Tiktok @dr.karanr 13.4.2025, sosiaalisen median julkaisu
 <<https://vm.tiktok.com/ZNdjnTSUE/>>, haettu 3.5.2025.

Tiktok @doctoryoun 30.3.2025, sosiaalisen median julkaisu
 <<https://vm.tiktok.com/ZNdjnwpoE/>>, haettu 3.5.2025.

Troshani, I. – Hill, S. R. – Sherman, C. – Arthur, D. (2021) Do we trust in AI? Role of anthropomorphism and intelligence. *Journal of computer information systems*, Vol. 61 (5), 481-491.

- Vaccari, C. – Chadwick, A. (2020) Deepfakes and Disinformation: Exploring the impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, Vol. 6 (1).
- Wang, Y. – Min, Q. – Han, S. (2016) Understanding the effects of trust and risk on individual behavior toward social media platforms: A meta-analysis of the empirical evidence. *Computers in Human Behavior*, Vol.56, 34-44.
- Westerlund, M. (2019) The emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, Vol. 9 (11).
- Xu, B. – Li, D. – Shao, B. (2012) Knowledge Sharing in Virtual Communities: A Study of Citizenship Behavior and Its Social-Relational Antecedents. *Journal of Human-Computer Interaction*, Vol. 28 (5), 347-359.
- Yle 12.4.2025. Amerikkalaiset tutkijat varoittavat: Venäjän propagandan vyöry on saastuttanut tekoälyn vastaukset. <https://yle.fi/a/74-20155223?utm_source=social-media-share&utm_medium=social&utm_campaign=ylefiap>, haettu 1.5.2025.
- Zoonen, W. V. – Luoma-aho, V. – Lievonen, M. (2024) Trust but verify? Examining the role of trust in institutions in the spread of unverified information on social media. *Computers in Human Behaviour*, Vol. 150, 107992.

Liitteet

Liite 1. Tekoälyn käyttö

Tekoälysovellukset, joita on hyödynnetty ovat ScopusAI ja Chatgpt. ScopusAI-työkalua on käytetty lähteiden etsimiseen. Komentona käytetty esimerkiksi ”luottamusteoria” tai ”luottamuksen rakentuminen”. Chatgpt:tä on käytetty pääasiassa teorioiden selittämiseen tai lauseiden kääntämiseen. Lukiessa luottamuksesta, artikkeleissa saattoi olla luottamukseen liittyviä teorioita, jotka olivat tuntemattomia. Komento esimerkki: selitä ABI-teoria. Toinen käyttötapa oli vaikeiden lauseiden kääntäminen suomeksi. Komento: käänä lause x. Viimeiseksi ChatGPT:tä on käytetty tarkistamaan valmiin tutkielmatyön kirjoitusvirheet.