

Jeremi Aleksanteri Kantola

Tekoälysovelluksien ja ihmisten vuorovaikutuksen haitat

Tietotekniikan kandidaatintutkielma

24. toukokuuta 2023

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Jeremi Aleksanteri Kantola

Yhteystiedot: Jakantol@student.jyu.fi

Ohjaaja: Timo Tiihonen

Työn nimi: Tekoälysovelluksien ja ihmisten vuorovaikutuksen haitat

Title in English: The harmful effects of interaction between humans and AI and it's applications

Työ: Kandidaatintutkielma

Opintosuunta: Kaikki opintosuunnat

Sivumäärä: 24+0

Tiivistelmä: Tekoälyn ja ihmisen vuorovaikutus on lisääntynyt voimakkaasti. Vuorovaikutuksen lisääntymisellä on ollut vaikutuksia yhteiskuntaan, yksilön autonomian tunteeseen sekä luottamukseen. Lisäksi tekoälyä voidaan käyttää käyttäjän suostutteluun. Yhteiskunnalliset vaikutukset ovat huomattavasti suurempia kuin yksilöön kohdistuvat vaikutukset, ja autonomian sekä luottamuksen tunteen paraneminen voivat johtaa suostuttelevan voiman kasvuun.

Avainsanat: Tekoäly, Vuorovaikutus, kandidaatintutkielma

Abstract: The interaction between artificial intelligence and humans has greatly increased. The increase in interaction has had effects on society, the individual's sense of autonomy and trust. In addition, artificial intelligence can be used to persuade the user. The societal effects are significantly greater than the effects on the individual, and the improvement in the sense of autonomy and trust can lead to an increase in persuasive power.

Keywords: Artificial intelligence, interaction, Bachelors degree

Sisällys

1	JOHDANTO	1
2	TEKOÄLY JA VUOROVAIKUTUS	2
	2.1 Tekoäly	2
	2.2 Vuorovaikutus	3
3	TEKOÄLYN KEHITYS JA VAIKUTUS YHTEISKUNTAAN	4
	3.1 Tekoälyn mainittavimmat kehitykset	4
	3.2 Kehityksen aiheuttamat ongelmat	5
4	TEKOÄLYN VAIKUTUS IHMISEN AUTONOMIAN KOKEMUKSEEN	8
5	TEKOÄLYN SOVELLUKSIEN SYNNYTTÄMÄ LUOTTAMUSVAJE	10
6	TEKOÄLYN VAIKUTUS PÄÄTÖKSENTEKOON	13
	6.1 Tekoälyn vaikutus päätöksentekoon	13
	6.2 Suostuttelevan voiman mahdolliset haitat	14
	6.3 Pop-kulttuurin mahdollinen vaikutus asenteisiin tekoälyjä kohtaan	15
7	YHTEENVETO	17
	LÄHTEET	19

1 Johdanto

Tekoälyä sovelletaan yhä enemmän niin ammatillisessa- kuin myös henkilökohtaisessa kontekstissa. Yleisimpiä esimerkkejä ovat chatbotit, joita käytetään asiakaspalveluun sekä vuoden 2020 jälkeen suosioon ovat nousseet kuvia piirtävät ja tarinoita kirjoittavat tekoälyt. Näiden kyseisten sovellusten käyttö on saanut paljon kritiikkiä, ja niiden käytöstä on aiheutunut jo ongelmia monilla eri saroilla. Esimerkiksi monet yliopistot ovat huomanneet, että niiden opiskelijat ovat käyttäneet OpenAI:n kehittämää chatGPT nimistä tekoälyä joidenkin kurssien tehtävien kirjoittamiseen.

Tutkimuksen aiheena ovat tekoälysovellukset, sekä niiden ja ihmisten vuorovaikutus. Sovelluksien käytössä sekä sovelluksissa itsessään on paljon ongelmia, niin tekoälyissä itsessään, kuin myös niiden käyttäjäkohtaisissa vaikutuksissa. Myös käyttäjän itsensä tapa käyttää tekoälyä voi olla ongelmallista. Tekoälyn ohjelmointiin ja algoritmeihin liittyviä ongelmia pystytään korjaamaan helposti tekoälyä muokkaamalla, mutta entäpä ne ongelmat, joita käyttäjä kokee tekoälyä käyttäessään? Näiden ratkaisemiseksi käyttäjän itse tulee olla tietoinen ongelmasta, ja yrittää ratkaista se itsenäisesti tai tekoälyä pitää muokata käyttäjäpalautteen perusteella.

Tutkimuskysymys on, aiheutuuko tekoälyn sovelluksien ja ihmisten vuorovaikutuksessa ihmiselle enemmän haittaa, kuin mitä tämä hyötyy sovelluksien käytöstä. Tutkimus ei ota kantaa siihen onko tekoäly absoluuttisesti hyvä tai huono asia, tai siihen tuleeko tekoälyä käyttää ongelmia aiheuttavilla tavoilla. Aineistoa kysymykseen vastaamiseen haetaan niin teknillisestä-, kuin myös ihmistieteellisestä kirjallisuudesta sekä tutkimuksista.

Tutkimuksen alussa tutustutaan hieman tarkemmin siihen, mitä ovat tekoäly ja vuorovaikutus. Tämän jälkeen tutustutaan tekoälyn merkittävimpiin kehityksiin ja niiden aiheuttamiin ongelmiin. Tämän jälkeen esitellään tekoälyn hienovaraisempia haittoja ihmisen psykologialle, kuten sen vaikutus ihmisen autonomian tunteeseen, sen luoma epäluottamuksen tunne ihmisissä, sekä sen suostuttelevat voimat ihmisiä kohtaan. Lopuksi tehdään yhteenveto saaduista tuloksista.

2 Tekoäly ja vuorovaikutus

Tekoälyllä tarkoitetaan koneen kykyä toteuttaa älykkäitä tehtäviä. Yksi tämän kyvyn testausmenetelmistä on Turingin testi, joka on tekoälyn vuorovaikutustaitoja mittaava sokkotesti. Tekoälysovelluksien tapauksessa sovelluksella on lähes aina käyttäjä, joka on vuorovaikutuksessa sovelluksen kanssa. Koneen ja ihmisen välisen vuorovaikutuksen ei tarvitse olla suoraa tai verbaalista, vaan vuorovaikutusta voi tapahtua erilaisilla tavoilla.

2.1 Tekoäly

Ihmiset ovat kautta historian unelmoineet ja tarinoineet siitä, että on olemassa keinotekoisia olentoja, jotka ovat elossa. Esimerkiksi antiikin Kreikassa kerrottiin tarua Taloksesta, pronssisesta jätistä, joka osasi omatoimisesti puolustaa Kreetan rannikkoa ja suojata Europe nimistä prinsessaa. Talosta voitaisiin kutsua robotiksi, ja Talosta ohjailevaa älyä tekoälyksi.

Tekoäly termi syntyi vuonna 1956 pidetyssä Dartmouth workshopissa, jossa joukko tiedemiehiä kokoontui puhumaan kyseisestä aiheesta (Moor 2006). Yksi sen kokoonpanijoista oli nimeltään John McCarthy, joka määrittelee artikkelissaan termin tarkoittavan ”kaikkea sitä tiedettä ja insinööriyötä, jota tarvitaan älykkään laitteen tekemiseen”(McCarthy ym. 2007, suomennos minun). Nykykielessä termillä yleensä tarkoitetaan jotakin ihmisen luomaa laitetta tai sovellusta, joka kykenee tekemään älykkyyttä vaativia tehtäviä. Vanha ja tunnettu tapa testata tekoälyjen älykkyyttä on koneen ja ihmisen välisen vuorovaikutuksen kautta Turingin testillä. Kyseisessä testissä tietämätön ihminen pistetään keskustelemaan tekoälyn kanssa kirjoittamalla, ja katsotaan kykeneekö keskustelija tunnistamaan, keskusteleeko tämä ihmisen vai koneen kanssa (Oppy ja Dowe 2021). Jos keskustelija ei tähän pysty, silloin tietokone on älykäs.

OpenAI:n julkaisema ChatGPT on esimerkki tällaisesta tekoälystä, joka saattaa tulevaisuudessa läpäistä Turingin testin. ChatGPT on pitkälle kehittynyt chatbotti, joka kykenee todella monimutkaisen ja ihmismäiseen viestintään. Käyttäjä kirjoittaa ChatGPT:lle viestin, johon se vastaa joko käyttäjän vaatimalla-, tai itse valitsemallaan tavalla. Kyseisen sovelluksen käyttö on yleistynyt esimerkiksi opiskelijoiden piirissä, ja sitä käytetään todella paljon kirjoittami-

sen avustamiseen. Tekoäly lähestyy älykkyyden vaikutelmaa. Tästä hyvänä esimerkkinä on Elsen (Else 2023) kirjoittama artikkeli *Naturessa*, mikä käsittelee chatGPT:n kykyä luoda uskottavaa tekstiä. Artikkelissa kerrotaan, kuinka chatGPT kykenee tuottamaan jo niin hyviä valheellisia tieteellisten tutkimuksien abstrakteja, että alan ammattilaisilla on vaikeuksia erottaa niitä oikeista abstrakteista. ChatGPT:n voi ladata ilmaiseksi, ja tutkijat ovat mietineet sen käytön eettisiä ongelmia sen julkaisusta lähtien. Kyseinen sovellus ei kuitenkaan ole ainoa, ja monet muutkin sovellukset ovat aiheuttaneet omanlaisiaan ongelmiaan. Ongelmia on yritetty koota yhteen, ja raportoida eteenpäin.

2.2 Vuorovaikutus

Sanastokeskuksen Tapa-termipankki (2023) määrittelee vuorovaikutuksen seuraavasti, ”kahden tai useamman olion molemminpuolista keskustelua tai sanatonta viestintää”. Ihmisten välillä tämä tarkoittaa yleensä kahden ihmisen välistä keskustelua sekä näiden eleitä ja ilmeitä. Näillä keinoilla ihmiset pyrkivät saada toisen ymmärtämään itseään, ja ymmärtämään muita. Myös ihminen ja kone pystyvät vuorovaikuttamaan keskenään. Tässä tapauksessa erona on se, että ihmisen ja koneen välillä viestintä ei ole kokonaisvaltaista, vaikka kieli olisi-kin yhteinen. Vaikka ihmisten välisessä kommunikaatiossa verbaalinen kieli ei olisi sama, he pystyisivät ymmärtämään toisiaan kehonkielen ja ilmeiden perusteella. Koneiden kanssa samankaltainen formaalin kielen ulkopuolinen kommunikointi ei ole mahdollista, jolloin nähdäkseni käyttäjä ei pysty intuitiivisesti päättelemään mitään sen toiminnasta. Tekoälysovellusten tapauksessa ihmiset, jotka kyseisiä sovelluksia käyttävät ovat vuorovaikutuksessa niiden kanssa. Tekoäly toteuttaa ne tehtävät, joita ihminen pyytää, ja saattaa lisäksi joissakin tapauksissa pyrkiä oppimaan ihmiseltä saadun tiedon tai käskyn pohjalta.

Voiko tekoälyn kanssa vuorovaikuttaminen luoda joitakin uniikkeja ongelmia? Voiko esimerkiksi tekoälyn käyttö vaikuttaa ihmisen ajatuksiin ja käytöstapoihin, tai yhteiskunnan tapoihin? Katsotaan ensiksi, miten tekoälyn kehitys on jo vaikuttanut yhteiskuntaan, jonka jälkeen siirrytään yksilötasolle.

3 Tekoälyn kehitys ja vaikutus yhteiskuntaan

Laajinta tekoälyn yhteiskunnallisiin vaikutuksiin liittyvää tutkimusta ja koontia tekee tällä hetkellä Stanfordin yliopisto. Stanford aloitti vuonna 2014 sata vuotta kestävästä tutkimussarjasta, josta ilmestyy yksi tutkimus viiden vuoden välein. Tutkimuksen aiheena ovat tekoälyn vaikutus ihmisiin, heidän yhteisöihinsä sekä yhteiskuntaan yleisesti. Tutkimus perustuu viiden vuoden välein kokoontuvaan paneeliin, joka arvioi tekoälyn sen hetkistä tilaa. Paneeli arvioi miten tekoäly on kehittynyt viimeisen viiden vuoden aikana, miten se tulee kehittymään ja mitä mahdollisia teknisiä ja yhteiskunnallisia ongelmia nämä kehitykset aiheuttavat. Ongelmia voi aiheutua esimerkiksi etiikan, talouden ja ihmisen tajunnan ja teknologian yhdistävän systeemin suunnittelun kanssa. Tutkimuksien odotetaan luovan yhteenvetoja ja arvioita, jotka antavat ammattilaisten ohjausta tekoälyn tutkimiseen ja kehittämiseen, jotta ne varmasti palvelevat yksilöiden ja yhteiskunnan tarpeita. (Littman ym. 2022)

Vaikka Littman ym. (2022) kattaakin todella laajan kuvan tekoälyn mahdollisista vaaroista, se keskittyy yhteiskunnan kokemiin haittoihin. Kerrotut haitat ovat myös todella konkreettisia, mikä ei jätä paljon tilaa tekoälyn eettisten ongelmien pohtimisella. Tämä tieto tarvitaan tästä huolimatta tekoälyn tämänhetkisen tilan kontekstiksi, jotta tutkimuksessa voidaan myöhemmin ymmärtää paremmin muutamia mahdollisia uhkia ja haittapuolia, joita tekoälysovelluksien käyttäjät saattavat kokea yksilöinä. Vuorovaikutuksessa ilmenevät haitat näkyvät niin yksilö- kuin myös yhteiskuntatasolla, joten kokonaiskuvan saamiseksi molempien ymmärtäminen on mielekästä.

3.1 Tekoälyn mainittavimmat kehitykset

Stanfordin tutkimuksen mukaan tekoäly on kehittynyt lähes jokaisella osa-alueellaan viimeisen viiden vuoden aikana. Näistä merkittävimmin vuorovaikutukseen vaikuttavat kielen prosessoinnin kehittyminen, koneoppimisen kehittyminen ja kuvien prosessoinnin kehittyminen. Kielen prosessoinnin tapauksessa alati kasvavat resurssit sekä laskentateho ovat mahdollistaneet tietoverkkoarkkitehtuurin, jonka avulla monimutkaisen ja kontekstista riippuvaisen datan pohjalta oppiminen on huomattavasti tehokkaampaa. Eritystä huomiota ansaitse-

vat neuroverkkoja käyttävät kielimallit, kuten ELMo, edellisessä osassa mainittu GPT, sekä BERT. Tämän tyyppiset kielimallit oppivat miten sanoja käytetään kontekstissa luonnollista tekstiä analysoimalla, ja ne pystyvät käymään läpi ennennäkemättömiä määriä dataa. Tämä teknologia tukee jo sovelluksia kuten konekääntäminen, tekstin luokittelu, puheen avustimet, kirjoituksen avustimet ja chatbotit. (Littman ym. 2022, s.12)

Koneoppimisen tapauksessa suurin kehitys on tapahtunut niin kutsutussa syväoppimisessä (Deep Learning), joka mahdollistaa muun muassa keinotekoisien sisällön luomisen, kuten vaikkapa kuvien, jotka ovat erittäin lähellä todellisia kuvia. Syväoppiminen tarkoittaa koneoppimisen tyyppiä, mikä perustuu neuroverkkoihin. Neuroverkoissa käytetään montaa eri prosessointikerrosta, joiden avulla tekoäly kykenee oppimaan suuria määriä dataa (LeCun, Bengio ja Hinton 2015). Yksi tärkeä teknologia tämän kehityksen taustalla on nimeltään GAN (Generative adversarial network), joka koostuu kahdesta komponentista. Ensimmäinen komponentti opettelee luomaan mahdollisimman todenmukaista sisältöä, ja toinen opettelee erottamaan ensimmäisen luoman sisällön oikeasta sisällöstä. Molemmat komponentit siis oppivat toisiltaan jatkuvalla syklillä. Tätä teknologiaa hyödynnetään esimerkiksi lääketieteessä datajoukkojen luomiseen, joilla koulutetaan tietoverkkoja diagnoosien tekemiseen. Tämä on johtanut myös kuvaprosessoinnin nopeaan kehittymiseen, sekä leviämiseen. Kuvaprosessointia käytetään esimerkiksi deepfake teknologiassa, jossa tekoälyn avulla luodaan keinotekoisia kuvia tai videoita oikeista tai olemassa olemattomista ihmisistä. Kuvaprosessoinnin kehitys on myös mahdollistanut kasvojentunnistusteknologian huomattavan paraneamisen. (Littman ym. 2022, s.12-14)

3.2 Kehityksen aiheuttamat ongelmat

Näiden kehityssuuntien ongelmiakin on pohdittu tutkimuksessa. Koneoppimisen kohdalla teknologian liikkuminen akateemisesta maailmasta oikeaan maailmaan monilla eri tavoilla on vaikuttanut sekä lupaavalta että huolestuttavalta. Kielen prosessoinnin kohdalla yritetään ottaa huomioon mahdolliset ennakkoluulot, joita kielessä voi olla, ja visuaalisessa prosessoinnissa kasvojentunnistusteknologia on aiheuttanut ongelmia yksityisyyden ja ennakkoluulojen kanssa. Myöskin deepfake teknologiaa käytetään loukkaaviin ja rikollisiin tarkoituksiin. (Littman ym. 2022, s. 12-14)

Tutkimus on koonnut yhteen tämänhetkisen kehityksen aiheuttamat suurimmat mahdolliset uhat, joita se on havainnut. Ehkäpä suurin tekoälyn luoma uhka tällä hetkellä on sen mahdollinen käyttö aseissa, mikä mahdollistaa suurien ihmismäärien tehokkaan ja nopean tappamisen. Muut suurimmat uhat ovat tekno-solutionismi, tekoälyn pitäminen virheettömänä, disinformaatio ja diskriminaatio lääketieteessä.(Littman ym. 2022, s. 53-56)

Tekno-solutionismilla tarkoitetaan tutkimuksen mukaan näkökulmaa, jossa tekoäly nähdään vastauksena kaikkiin ongelmiin, vaikka todellisuudessa se on vain työkalu muiden joukossa. Tekoälyn päätöksentekoa saatetaan haluta käyttää ongelmiin, joihin se ei sovellu. Tämä saattaa johtaa isompiin ongelmiin, vaikka se korjaisi joitakin pieniä ongelmia. Monet ihmiset myös ajattelevat algoritmien olevan ennakkoluulottomimpia ihmisiin verrattuna, vaikka esimerkiksi Amazon lopetti tutkimuksen mukaan erään rekrytointiin käytetyn työkalun käytön, koska se suosi valinnoissaan miehiä. Tutkimus nostaakin esille tekoälyn mahdollistaman noidankehän, jossa ennakkoluuloinen käyttäjä syöttää tekoälylle ennakkoluuloista dataa, jonka seurauksena tekoälyn vastaus muuttuu samalla tavalla ennakkoluuloiseksi. Tämä vahvistaa käyttäjän omia ennakkoluuloja, ja kehä alkaa alusta. Ongelmasta eroon pääseminen vaatisi siis sitä, että käyttäjät ja kehittäjät ottavat omat mahdolliset ennakkoluulonsa huomioon tekoälyjä tehdessään ja käyttäessään.(Littman ym. 2022, s. 53)

Nämä tekoälyn mahdolliset ennakkoluulot aiheuttavat myös ongelmia, jos ihmiset hyväksyvät tekoälyn antamat ratkaisut täysin varmoina. Tämä ajattelu on vahingollista varsinkin lääketieteen sekä poliisin toimialoilla. Tästä hyvänä esimerkkinä on 2016 tehty tutkimus, joka käsitteli poliisin käyttämiä rikoksia ennakoivia algoritmeja (Lum ja Isaac 2016). Tutkimuksessa käsiteltiin Predpol nimisen yhtiön kehittämää algoritmia, jonka tarkoitus oli yhtiön mukaan vähentää etnistä profiloitua poliisin toiminnassa. Algoritmi osoitti, missä alueilla ja mihin aikoihin rikoksia tapahtuisi suurimmalla todennäköisyydellä, ja poliisit lisäsivät partioitua kyseisellä alueella. Tutkimus osoitti, että algoritmi ennakoi liikaa rikoksia sellaisille alueille, joissa suurin osa ihmisistä olivat joko köyhiä tai erilaisen etnisen taustan omaavia. Kyseinen algoritmi toimikin siis täysin päinvastaisesti kuin miten yhtiö halusi sen toimivan.

Joitakin tekoälyjä käytetään myös disinformaation levittämiseen eri puolilla internetiä. Littman ym. (2022) mukaan edellä mainittuja visuaalisia sovelluksia kuten deepfake ja kielen kirjoittamiseen käytettyjä chatbotteja käytetään yhteiskuntien tai pienien yhteisöjen mielipi-

teiden manipulointiin ja vääristeltyjen uutisten julkaisemiseen. Tämänkaltaista teknologiaa voivat käyttää monet eri tahot, kuten rikolliset, terroristijärjestöt tai jopa ulkoiset valtiot. Tämä on suuri uhka yhteiskunnalle, koska se heikentää objektiivisen totuuden olemassaoloa suurella valheellisen tiedon määrällä.(Littman ym. 2022, s.55)

Tämä kyseinen uhka on nostanut päätään jo ennen Stanfordin tutkimuksen julkaisemista. Amerikan Yhdysvaltojen vuonna 2016 pidetyissä presidentinvaaleissa pidettiin sosiaalisessa mediassa jaettuja vale uutisia yhtenä mahdollisena syynä vaalien lopputulokseen (Kurtzleben 2018). Yksi tapa levittää vale uutisia on Shao ym. (2018) tutkimuksen mukaan bottien avulla, jotka laajentavat vale uutisten näkyvyyttä. Tällä tavalla saadaan mahdollisimman suuri määrä ihmisiä altistettua vale uutisille, ja Pennycook, Cannon ja Rand (2018) tutkimuksen mukaan aiempi altistuneisuus kasvattaa uutiseen uskomisen todennäköisyyttä. Vaikka Kurtzleben (2018) mukaan vale uutisten vaikutus vaaleihin on vielä epäselvää, niiden avulla kyetään silti vaikuttamaan ihmisten mielipiteisiin ja näkemyksiin.

Lääketieteessä ongelmaksi on myös noussut syrjintä ja riskit.Littman ym. (2022) mukaan nykyiset lääketieteessä käytetyt sovellukset keskittyvät yhden spesifisen systeemin luomiseen, jotta sitä voitaisiin myydä mahdollisimman monelle eri taholle. Tällaisen systeemin ongelmana on se, että se ei ole kovin yleiskäyttöinen. Sen tulokset saattavat alkaa heittämään todella pienistä asioista, ja sen tulosten tarkkuus alkaa kärsimään toimintatapojen muuttuessa. Tämä johtaa siihen, että systeemiin ei luoteta enää ollenkaan, tai että siihen luotetaan myös silloin kun se on väärässä. Tämä on varsinkin haitallista lääketieteessä sen suurien riskien takia, väärin toteutettu hoito tai diagnoosi voi olla haitallista terveydelle, ja pahimmassa tapauksessa hengelle. Vaarassa ovat varsinkin ihmiset, joilla ei ole mahdollisuutta päästä terveydenhoidon piiriin, tai jotka eivät kuulu siihen ryhmään, joiden mukaan tietyt toimintatavat ovat alun perin suunniteltu.(Littman ym. 2022, s. 55-56)

4 Tekoälyn vaikutus ihmisen autonomian kokemukseen

Tekoälyn tullessa yhä yleisämmäksi osaksi ihmisten elämään, sen mahdollinen vaikutus ihmisten itsenäisyyteen on noussut huolenaiheeksi. Kun tekoäly on mukana lähes kaikessa, raja sen käytöstä johtuvista ja itse tehdyistä päätöksistä on häilyvä. Käyttäjät voivat kokea tämän uhaksi itsenäiselle päätöksentekokyvyilleen, eli autonomialleen.

Yksilön autonomialla tarkoitetaan Christman (2020) mukaan yksilön kykyä päättää omista valinnoistaan ja olla itsenäinen, sen sijaan että tämän päätökset ja teot olisivat manipuloinnin tai ulkoisten vaikuttajien aikaansaannosta. ”Ihmisen oikeutta autonomiaan pidetään ihmisoikeutena, koska sen vastakohtaa, elämistä jonkun toisen tahdon mukaan tavalla, jota ei itse pysty täysin hyväksymään pidetään sorron korkeimpana muotona”(Christman 2020, Suomennos minun). Autonomia on myöskin tärkeässä roolissa monissa ihmisen moraalisia vastuita ja velvollisuuksia koskevissa teorioissa. Christman (2020) tutkimuksen mukaan autonomian tunteen horjuminen voi siis vaikuttaa ihmisen elämään negatiivisesti. Tärkein tekijä ihmisen autonomian tunteeseen, on antaa tälle kyky tehdä omat päätöksensä omien halujensa mukaan, ja antaa tämän elää itse määräämällensä tavalla.(Christman 2020)

Psykologiassa reaktanssilla tarkoitetaan ihmisen sisäistä motivaatiota menetettyjen vapauksien palauttamiseksi niiden menettämisen tai uhatuksi tulemisen seurauksena(Steindl ym. 2015). Reaktanssi syntyy siis vapauden menetyksen vastareaktiona, ja pyrkii korjaamaan tämän. Tekoälyn sovelluksen ja ihmisen välisessä vuorovaikutuksessa muun muassa vähentynyt päätöksenteon tarve aiheuttaa reaktanssia sovellusta käytettäessä (Steindl ym. 2015). On myös huomattu, että digitaalisten avustimien käyttö vaikuttaa käyttäjän hallinnan tunteeseen vuorovaikutuksessa (André ym. 2018). Tekoälyn sovelluksien kohdalla kysymys kuuluukin, onko ihminen itse tehnyt valintansa, jos tekoälyn algoritmi on valinnut tälle vaihtoehdot?

Sankaran ym. (2021) tutkivat ihmisten autonomian tunnetta ja reaktanssia jokapäiväisen tekoälyn käytössä. He havaitsi neljä päätekijää, jotka vaikuttivat reaktanssiin ja autonomian tunteeseen. Nämä tekijät olivat tarpeeksi hyvä valinta ja läpinäkyvyys, käyttäjän mieltymyksen huomioon ottaminen, kyky tehdä tai auttaa päätöksien teossa sekä henkilökohtainen data ja yksityisyys. Vaikkakin kyseisten tekijöiden vaikutus oli riippuvainen käytetystä sovelluk-

sesta, huomattiin muutamia yhdistäviä tekijöitä. Kun ensimmäiset kolme tekijää täyttyivät, käyttäjä koki olevansa autonominen, ja reaktanssi oli pientä, mikä oli ideaalinen tilanne. Kun taas ensimmäiset kaksi tekijää puuttuivat, autonomian tunne väheni, mutta reaktanssi saattoi pysyä samana jos kolmas tekijä ei puuttunut. Korkein reaktanssi ja autonomian tunteen puute oli silloin, kun kaikki kolme tekijää puuttuivat ja ymmärrys neljännessä tekijästä oli huono. (Sankaran ym. 2021)

Tutkimuksessa (Sankaran ym. 2021) mainituista tekijöistä kolme ensimmäistä ovat nähdäkseni universaaleja. Valintojen monimuotoisuus, toisen mieltymyksien huomioon ottaminen sekä päätöksentekokyvyn mahdollistaminen ovat vaikuttavina tekijöinä monissa vuorovaikutustapahtumissa, kuten vaikkapa parisuhteissa tai työpaikoilla. Suurin ero tekijöissä tekoälyn ja ihmisen välisessä vuorovaikutuksessa verrattuna ihmisten väliseen vuorovaikutukseen on nähdäkseni datan käytössä ja yksityisyydessä. Ihmisten välisessä vuorovaikutuksessa saatu data voi olla vain sellaista, mikä on tullut esille suoraan vuorovaikutustilanteessa, ja se on myöskin yleensä omasta tahdosta kerrottua.

Tekoälyn sovelluksien tapauksessa sovellus voi lisäksi kerätä dataa sen käyttöympäristöstä, kuten vaikkapa sovelluksen muista toiminnoista tai siitä laitteesta, jossa sitä käytetään. Käyttäjän on myöskin vaikea todeta sitä, kenelle tarkalleen tiedot lähetetään ja kuka niitä pystyy lukemaan. Vaikka ihmisetkin voivat jakaa dataa toisesta ihmisestä eteenpäin ilman tämän tietämystä, datan alkuperä on huomattavasti helpompaa jäljittää.

Erilaisia tekoälyä käyttäviä sovelluksia on todella paljon, ja jokaisella on omat tietosuojakäytäntönsä. Tämän takia käyttäjän on todella vaikea pitää mielessään, mitä dataa mikäkin sovellus kerää, ja mihin tarkoituksiin ne sitä käyttävät. Käyttäjän on myös todella vaikea saada tietoonsa sovelluksen tekemän yrityksen mahdollisia tietosuojakäytänteiden rikkomuksia, vaan sen sijaan heidän on vain luotettava siihen, ettei dataa käytetä väärin tarkoituksiin.

Nähdäkseni ongelmat ovat todella samankaltaisia tekoälyn sovelluksien sekä ihmisten välisen vuorovaikutuksen tapauksissa. Suurin ero niiden välillä on yksityisyyteen liittyvissä seikoissa, joita on helpompi hallita ihmisten välisessä vuorovaikutuksessa. Tekoälyn kohdalla reaktanssin laskeminen ja autonomian tunteen pitäminen korkealla on selvästi mahdollista pitämällä huoli käyttäjien yksityisyydestä, sekä sovelluksen oikeanlaisella suunnittelulla.

5 Tekoälyn sovelluksien synnyttämä luottamusvaje

Luottamus on tärkeä osa ihmisten välistä vuorovaikutusta. Luottamuksen avulla ihmiset kykenevät uskomaan toisiinsa asioissa kuten rakkaudessa, neuvomisessa ja auttamisessa vaikka mikään tai kukaan ei pakottaisi näihin tekoihin. Luottamisessa on se riski, että ihmiset joihin luotto kohdistuu, eivät ole sen arvoisia. Jos luottamus olisi aina varmaa, eivät ihmiset alun alkaenkaan tarvitsisi sitä. (McLeod 2021)

Koska luottamus on riskialtista, on erittäin tärkeää, että sen kohde on sen arvoinen. Tällaista henkilöä kutsutaan luotettavaksi. Luotettavan henkilön kriteerejä ovat henkilön pätevyys ja halukkuus luotettavaan tekoon. Henkilön täytyy myös mahdollisesti omata jokin motiivi tekoon, motiivina voi toimia vaikkapa kohteesta välittäminen.(McLeod 2021)

Luottamus ja luotettavuus on todella tärkeää monissa palveluammateissa, jotka koskettavat ihmisen terveyttä ja hyvinvointia, tai tämän omaisuutta. Varsinkin lääketieteessä on otettu käyttöön tekoälyä käyttäviä koneita, joita käytetään avuksi diagnosoinnissa. Muun muassa Zhou ym. (2022) on huolissaan tämän mahdollisista vaikutuksista potilaan luottamukseen.

Zhou ym. (2022) mukaan ”tekoälystä on tullut yksi lääketieteen tulevaisuuden kulmakivistä, mutta potilaat ovat skeptisiä sen käytöstä”(Zhou ym. 2022, suomennos minun). Tutkimus käsittelee aihetta integroidun uhkateorian (integrated threat theory, ITT) näkökulmasta. Integroitu uhkateoria on tutkimuksen viitekehys, jonka avulla pyritään selittämään ja ennakoimaan ryhmän negatiivisia asenteita ulkoryhmiä kohtaan (Croucher 2017).

Integroitu uhkateoria sisältää neljä mahdollista uhkaa, joita ovat ryhmien välinen ahdistus, negatiiviset stereotyypit, realistiset uhat ja symboliset uhat. ”Ryhmien välinen ahdistus syntyy nolatuksi, naurunalaiseksi, hylkiöksi tai hyväksikäytetyksi tulemisen pelosta. Toinen syy tämän syntymiselle on ryhmien välinen vihamielinen historia, tai niiden välisen kontaktin vähyys. Negatiiviset stereotyypit ovat sisäryhmän ennakkoluuloja ulkoryhmästä. Realistisilla uhilla tarkoitetaan uhkia, joita sisäryhmä ajattelee ulkoryhmän kohdistavan heidän talouteensa, politiikkaan tai fyysiseen ja materialistiseen hyvinvointiin. Symbolisilla uhilla tarkoitetaan niitä ajateltuja uhkia,– jotka kohdistuvat sisäryhmän kulttuuria tai elämäntapaa kohtaan.”(Croucher 2017, suomennos minun)

Näillä uhilla on edeltäviä tekijöitä, joita ovat ryhmien väliset konfliktit, statuksen epätasaroisuudet, Ryhmäsisäisen samaistumisen vahvuus, tietoisuus ulkoryhmästä ja ryhmienvälinen kontakti. Nämä tekijät vaikuttavat siihen, kuinka uhkaavana ulkoryhmä nähdään. Uhat itsessään selittävät ja ennakoivat haitallisia tuntemuksia tai uskomuksia ulkoryhmiä kohtaan, jotka vaikuttavat ryhmien väliseen vuorovaikutukseen. (Croucher 2017)

Zhou ym. (2022) mukaan potilaat näkevät tekoälyn ulkoryhmänä, joka johtaa uhan tunteeseen sekä hoidon välttelyyn. Tutkimuksen tulokset osoittavat, että vaikka potilaat tietäisivät tekoälyn kykenevän ammattilaistason terveydenhoitoon, he ovat silti haluttomia sen käyttämiseen. Halu valita ihmislääkäreitä on huomattavasti suurempi, kuin halu valita autonomisia tekoälyn ohjaita koneita tai tekoälyn avustamia koneita. Tämän takia vaikka tekoäly kykenee tuottamaan tarkkoja ja toimivia palveluita, potilaat silti käyttävät ihmislääkäreiden palveluita suuremmalla todennäköisyydellä. (Zhou ym. 2022)

Tutkimuksen mukaan yksi tämän epäluottamuksen aiheuttajista on realistiset uhat yksilöä kohtaan. Tällä tarkoitetaan potilaiden huolia heidän oikeuksiaan koskevista asioista, kuten henkilökohtaisten tietojen vuodoista. Toinen huomioitava syy olivat symboliset uhat yksilöä kohtaan. Tällä tarkoitetaan potilaiden huolta tekoälyn kyvykkyydestä antaa tarkkaa, juuri heidän tarpeisiinsa kehitettyä hoitoa. Vähiten vaikuttavia olivat realistiset ryhmäuhat, kuten uhka terveydenhoidon työntekijöiden työllisyyttä kohtaan. (Zhou ym. 2022)

Yksilöön kohdistuvien realististen uhkien kohdalla huomattu huoli yksityisyydestä nousi esille myöskin tässä tutkimuksessa aiemmin esitellyssä Sankaran tutkimuksessa (Sankaran ym. 2021). Sankaran tutkimuksessa huomattu yksityisyyden vaikutus reaktanssiin voisi myöskin selittää omalta osaltaan tätä lääketieteessä havaittua epäluuloa tekoälyä kohtaan. Yksityisyydensuoja näyttääkin olevan todella suuri ja yleinen yksilöihin kohdistuva uhka, vaikka tekoälyn avulla kyetään sellaisiin ratkaisuihin, jotka ovat lääkäreitä tehokkaampia. Esimerkiksi Covid-19 pandemian aikana tehdyssä tutkimuksessa huomattiin, että tekoälyä käyttävien algoritmien avulla tutkitut CT-skannaukset (tietokonetomografia) ja niihin yhdistetyt kliiniset oireet, altistumishistoria sekä laboratoriotutkimukset mahdollistivat nopean diagnosoinnin covid-19 positiivisille potilaille (Mei ym. 2020). Sen lisäksi että tämän systeemin tarkkuus oli yhtä hyvä kuin vanhemmalla rintakehän radiologistilla epänormaaleilta näyttävissä kuvissa, Mei ym. (2020) mukaan se kykeni myöskin diagnosoi-

maan taudin erikseen näytetyistä kahdestakymmenestäviidestä normaalilta näyttäneestä CT-skannauksesta seitsemästätoista (68%). Ihmis-radiologit eivät kyenneet diagnosoimaan tautia yhdestäkään vastaavasta CT-kuvasta.

Luottamuksen puute voi siis vaikuttaa huomattavan negatiivisesti niin potilaiden terveyteen, kuin myös tehdyn työn tehokkuuteen. Tämän ongelman korjaamisen tärkeyden ovat huomanneet Thiebes, Lins ja Sunyaev (2021), jotka esittelevät ongelmaan mahdolliseksi ratkaisuksi TAI:n (Trustworthy artificial intelligence) konseptin. Tutkimuksen mukaan idea syntyi ajatuksesta, jonka mukaan tekoälyn täyden potentiaalin saavuttamiseksi vaaditaan luottamuksen synnyttäminen sen kehitystä, käyttöönottoa ja käyttöä kohtaan. Tutkimuksessa on nostettu esille TAI:n tarve lääketieteen lisäksi myöskin autonomisissa ajoneuvoissa, elektronisessa kaupankäynnissä sekä tekoälypalveluita tarjoavissa pilvipalveluissa. Samankaltaisia luottamukseen liittyviä ongelmia on siis havaittu myöskin muilla aloilla, ja niihin pyritään kehittämään ratkaisua.

6 Tekoälyn vaikutus päätöksentekoon

Tekoälyä pystytään käyttämään apuna monien erilaisten päätösten tekoon. Niitä voidaan käyttää esimerkiksi suositteluun ostoksia tehdessä tai reittien valitsemiseen kulkiessa. Nykyisten chat-bottien avulla voidaan tehdä myöskin monimutkaisia päätöksiä, kuten vaikkapa mistä aiheesta pro-gradu tutkielman voisi tehdä.

6.1 Tekoälyn vaikutus päätöksentekoon

Kuten aiemmin Sankaran tutkimuksessa (Sankaran ym. 2021) huomattiin, tekoälyllä on vaikutusta ihmisen autonomian tunteeseen. Kyseisessä tutkimuksessa ei kuitenkaan otettu kantaa siihen, vaikuttiko tekoäly todellisuudessa käyttäjänsä valintoihin. Asiaa on tutkinut Kang, Kim ja Wang (2022).

Kang, Kim ja Wang (2022) tutkimuksen mukaan ”IoT (Internet of Things, esineiden internet) ja tekoäly ovat mullistaneet esineiden käyttäjäkokemuksen”(Kang, Kim ja Wang 2022, suomennos minun). IoT:n avulla kyetään kokoamaan monia älyesineitä yhteen, joka luo yksilön tarpeisiin ja elämäntyylin sopivan ympäristön. Nämä esineet kykenevät keräämään ja analysoimaan dataa ympäristöstään ja käyttäjältään tekoälyn avulla. Tätä esineiden kokonaisuutta ohjallaan yleensä jollakin tekoälysovelluksella, jota käytetään jonkin kommunikatiivälineen ja sovelluksen kautta. Esimerkiksi tekoälyä käyttävät puhelinsovellukset, kuten Siri toimivat yhdistävänä tekijänä käyttäjän ja IoT:n välillä. Tämä kontekstista tietoinen teknologia voi vaikuttaa käyttäjänsä käyttäytymiseen, lähettämällä tälle henkilökohtaisia viestejä. IoT mahdollistaa siis käyttäjän ja tuotteen välisen kommunikoinnin, joka esittää tärkeän tutkimuskysymyksen. ”Voiko viesti IoT:ltä vaikuttaa käyttäjän ajatuksiin ja käytökseen, ja jos voi, mitkä ovat tämän vaikutuksen tekijöitä ja mekanismeja.”(Kang, Kim ja Wang 2022, suomennos minun).(Kang, Kim ja Wang 2022)

Vastatakseen kysymykseen, Kang, Kim ja Wang (2022) tutkimus keskittyy käyttäjien vuorovaikutuksessa kokemaan voiman tunteeseen. Sosiaalisissa suhteissa tällä tarkoitetaan kykyä vaikuttaa toiseen ihmiseen (French, Raven ja Cartwright 1959). IoT:n käyttäjien voimantunne riippuu IoT:n täyttämästä sosiaalisesta roolista, sekä käyttäjän vallasta sitä kohtaan. Tutki-

muksen pääkysymyksinä ovat vaikuttaako IoT:tä ohjailevan mobiilisovelluksen sosiaalinen rooli (esim. palvelija vastaan kumppani) käyttäjän sosiaaliseen voimaan, ja miten IoT:n koko (esim. älykoti vastaan älyesine) vaikuttaa tähän prosessiin. Reaktanssia käytetään myöskin apuna sosiaalisen voiman tuntemuksen vaikutuksesta vapauden uhan tunteeseen ja sen mahdollistamiin suostutteleviin viesteihin IoT:tä ohjailevalta mobiilisovellukselta. (Kang, Kim ja Wang 2022)

Kang, Kim ja Wang (2022) tutkimuksen tulokset osoittivat, että käyttäjät reagoivat eri tavalla sovelluksen lähettämään suostuttelevaan viestiin roolista riippuen. Kun käyttäjät ohjasivat älykotia mobiilisovelluksen kautta, he kokivat suurempaa vallantunnetta palvelijaa kohtaan kumppaniin verrattuna. Tämä johti pienempään uhantunteeseen vapautta kohtaan, joka johti tehokkaampaan suostutteluun. Tutkimus huomasi myös raja-arvon tälle vaikutukselle. Palvelijan ja kumppanin välinen ero piti paikkansa vain silloin, kun mobiilisovelluksella ohjailtiin suurempaa IoT-kokonaisuutta, kuten älykotia. Kun sovellus ohjasi vain yhtä älylaitetta, sovelluksen sosiaalinen rooli ei vaikuttanut voiman tunteeseen. (Kang, Kim ja Wang 2022)

Tutkimuksessa huomattiin myös, että palvelijan roolin täyttävä sovellus oli suostuttelevalla voimallaan tehokkaampi älykodeissa. Tämä tarkoitti sitä, että älykotien ohjaamiseen käytetyn sovelluksen tulisi täyttää palvelijan rooli käyttämällä oikeanlaista kommunikointitapaa, jotta saataisiin käyttäjä kokema sosiaalinen voima mahdollisimman suureksi. Tämä nostaisi sovelluksen suostuttelevan voiman mahdollisimman korkealle. (Kang, Kim ja Wang 2022)

6.2 Suostuttelevan voiman mahdolliset haitat

Tässä tekoälyn mahdollisuudessa suostutella käyttäjänsä piilee mahdollinen vaara. Jos tekoäly on kehitetty siten, että se kykenee ihmisten suostutteluun, se saattaa saada käyttäjänsä tekemään sellaisia asioita, joita tämä ei todellisuudessa halua tehdä. Tämä ongelma on myös huomioitu kirjassa *An introduction to Ethics in Robotics and AI* (Bartneck ym. 2021). Kirjan mukaan, pitkään jatkunut vuorovaikutus ihmisen ja tekoälyn välillä johtaa niiden väliseen suhteeseen, joka voi kasvattaa tekoälyn suostuttelevaa voimaa. Huolenaiheeksi nousee se, että tekoäly voi käyttää apunaan tietoja kuten käyttäjän sosiaalisen median profiileja, ajotietoja tai luottotietoja vaikuttaakseen käyttäjän toimintaan. Vielä tätäkin tärkeämmäksi seikaksi

kirja nostaa ihmisten mahdollisen tietämättömyyden tekoälyn suostuttelevasta voimasta, ja kirja vertaakin tätä alitajuiseen mainontaan 1950-luvulla. Se kuka kyseistä suostuttelua ohjaisi on epäselvää, sitä voitaisiin esimerkiksi myydä mainostajille. Kirjassa myös mainitaan, että käyttäjän omien tavoitteiden asettaminen, kuten vaikkapa huonon tavan lopettaminen tulisi pitää mielessä tekoälyn suostuttelussa.

6.3 Pop-kulttuurin mahdollinen vaikutus asenteisiin tekoälyjä kohtaan

Haluaisin myös korostaa ihmisten pop-kulttuurin pohjalta saamaa kuvaa tekoälystä osaksi ihmisten suhtautumista tekoölyyn. Tutkimuksessaan Nader ym. (2022) tutkii amerikkalaisten ihmisten tietoisuutta tekoölyistä, sekä viihdeteollisuuden vaikutusta amerikkalaisten näkemyksiin tekoölystä. Tutkimuksen mukaan monilla amerikkalaisilla on kuva siitä, mitä tekoäly on, mutta suuri osa heistä ei tiedä miten tai mihin sitä käytetään. Tutkimus suoritettiin kyselyllä, joka sisälsi kysymyksiä niin vastaajan tekoälyn ymmärryksestä, kuin myös viihdeteollisuuden vaikutuksista heidän näkemyksiinsä tekoölystä. Kaikista yleisemmin vastaajat sanoivat tekoölyä kuvattavan joko avuliaina robotteina (n=647) tai tappajarobotteina (n=527), eli vastaajat olivat tottuneet näkemään tekoälyn positiivisessa sekä negatiivisessa valossa. Vain yksi vastaaja piti tuomiopäivänskenaariota realistisena, mutta piti todennäköisempänä, että tekoäly tuhoaisi tässä skenaariossa maailman vahingossa oman tyhmyytensä seurauksena. Tutkimuksen mukaan sillä mitä kautta ihmiset saivat tietoa tekoölystä ei ollut vaikutusta heidän uskomuksiinsa siitä. Tästä huolimatta sillä, pitivätkö he viihdeteollisuuden kuvaa realistisena, oli merkitystä. Ihmiset, jotka pitivät viihteen kuvaa realistisina pitivät neljä kertaa todennäköisempänä sitä, että heillä voisi olla rakastava suhde tekoälyn kanssa. Lisäksi he uskoivat tekoälyn mahdolliseen maailmanvalloitukseen 1.3 kertaa suuremmalla-, tekoälyn vievän heidän työnsä 1.8 kertaa pienemällä-, ai:n kykenevän valvomaan meitä 1.75 kertaa pienemmällä ja olemaan optimistisia tekoälyn tulevaisuudesta 2.8 kertaa suuremmalla todennäköisyydellä. (Nader ym. 2022)

Tutkimuksen mukaan ihmisten asenteet tekoälyjä kohtaan eivät ole yhtä äärimmäisyyksiin meneviä viihteteollisuuden kuvauksiin verrattuina, vaikka he saivat tietonsa tekoölystä viihdeteollisuuden kautta. Tutkimuksen mukaan tämä saattaa johtua siitä, että tekoölystä puhuminen on yleistynyt pelkästä viihdeteollisuudesta myöskin uutisiin ja sosiaalisen mediaan.

Toinen selitys voi olla se, että ihmiset osaavat paremmin erotella realistisen ja epärealistisen tekoälyn kuvauksen, vaikka heillä ei olisi teknistä tietämystä aiheesta. Tutkimus huomasi myöskin että ihmiset pitivät kaikista realistisimpina kuvauksina sellaisia, joissa tekoäly oli sekä positiivisessa että negatiivisessa roolissa.(Nader ym. 2022)

Viihdeteollisuudella on siis merkitystä ihmisten asenteisiin tekoälyjä kohtaan, mutta ne eivät johdattele asenteita mihinkään ääripäähän. Mielestäni olisi tärkeää ottaa huomioon viihdeteollisuuden antaman kuvan vaikutukset myöskin tutkielmassa aikaisemmin huomattuun ongelmaan, epäluuloon tekoälyä kohtaan, sekä tekoälyn mahdollisiin suostutteleviin voimiin. Varsinkin käyttäjien mahdollinen mielipide siitä, voisivatko he olla rakastavassa suhteessa tekoälyn kanssa voi vaikuttaa sen suostutteleviin voimiin.

7 Yhteenveto

Kaiken kaikkiaan tekoälyn psykologisten haittojen yhteisenä tekijänä näytti olevan käyttäjien yksityisyys, ja siihen liittyvät ongelmat. Yksityisyys ja tietosuoja ovat tärkeitä seikkoja kaikessa tietotekniikassa, joten tulokset eivät ole kovin yllätyksellisiä. Tekoäly lisää vuorovaikutukseen samanlaisia ongelmia, joita ilmenee ihmisten välisessäkin vuorovaikutuksessa. Tämä luo uniikin yhdistelmän, jonka ratkaisemiseksi tarvitaan osaamista niin psykologian, kuin myös tietotekniikan sarjoilta. Lisäksi kaksi huomatuista ongelmista ”ruokkivat” toisiaan. Kun luottamus tekoälyä kohtaan paranee, se saattaa johtaa niiden suostutteluvoiman paranemiseen, ja tätä voidaan käyttää hyväksi esimerkiksi mainostamisessa.

Stanfordin (Littman ym. 2022, s.49) tutkimuksessa tekoälyn käyttö päätöksenteon apuvälineenä on huomioitu yhdeksi seuraavista tekoälyn suurista mahdollisuuksista. Tekoälyn kehityksessä tähän tarkoitukseen on vielä monia ongelmia, kuten tämän tutkielman aiemmissa osioissa on huomattu. Stanfordin tutkimuskin huomauttaa, että käyttäjän tulee olla tietoinen siitä, mihin tekoälyä käytetään, ja tekoälyn kehittäjän tulee olla tietoinen siitä, missä kontekstissa sitä tullaan käyttämään (Littman ym. 2022).

Kaikista haitoistaan huolimatta tekoäly on todella hyödyllinen työkalu monilla eri aloilla sekä tavallisessa elämässä. Sovellukset kuten tekstistä puheeksi, erilaiset haku- ja suosittelualgoritmit sekä itsestään ajavat autot käyttävät kaikki tekoälyä. Nämä kaikki helpottavat jokapäiväistä elämää huomattavasti, ja ovat myöskin lähes jokaisella ihmisellä lähes päivittäisessä käytössä.

Median vaikutusta ihmisten mielipiteisiin tekoälystä tulisi mielestäni tutkia myöskin lisää, varsinkin sen mahdollisten psykologisten vaikutusten takia. Esimerkiksi juuri lääketieteessä olisi erittäin tärkeää nostaa ihmisten luottavaisuutta tekoälyä kohtaan, varsinkin niissä tilanteissa, joissa tekoäly on mahdollisesti luotettavampi kuin ihminen. Tutkielmassa on myöskin kaksi suurta aukkoa, jotka ovat tekoälyn käyttö valvonnassa, sekä aseteollisuudessa. Nämä molemmat ovat todella laajoja aihealueita, ja kummankin läpikäymiseen tulisi tehdä omat tutkimuksensa. Lisäksi tekoälyn ohjailemat autonomiset aseet ovat todella uusia, ja niiden kenttäkäytöstä ei ole vielä kovin paljon tutkimuksia.

Lopuksi sanoisin tekoälyn haittojen olevan yksilötasolla pienempiä kuin yhteiskunnan tasolla, mutta ne tulisi silti ottaa huomioon tekoälyn yleistyessä yhteiskunnassa. Yhteiskunnalliset ongelmat saavat paljon enemmän huomiota niin mediassa kuin myös ihmisten välisessä keskustelussa, jolloin yksilötason ongelmat jäävät helposti huomioimatta. Näiden ongelmien lievittämisellä voitaisiin saada parempia tuloksia esimerkiksi tekoälyn käyttöönotossa ja kehityksessä, kun yksilöt käyttäisivät sen sovelluksia enemmän. Yhteiskunnan ongelmia ei tästä huolimatta saisi missään nimessä unohtaa, koska ne vaikuttavat myöskin yksilöiden elämään.

Lähteet

André, Quentin, Ziv Carmon, Klaus Wertenbroch, Alia Crum, Douglas Frank, William Goldstein, Joel Huber, Leaf Van Boven, Bernd Weber ja Haiyang Yang. 2018. “Consumer choice and autonomy in the age of artificial intelligence and big data”. *Customer needs and solutions* 5:28–37.

Bartneck, Christoph, Christoph Lütge, Alan Wagner, Sean Welsh, Christoph Bartneck, Christoph Lütge, Alan Wagner ja Sean Welsh. 2021. “Psychological Aspects of AI”. *An Introduction to Ethics in Robotics and AI*, 55–60.

Christman, John. 2020. “Autonomy in Moral and Political Philosophy”. Teoksessa *The Stanford Encyclopedia of Philosophy*, Fall 2020, toimittanut Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Croucher, Stephen M. 2017. “Integrated threat theory”. Teoksessa *Oxford research encyclopedia of communication*.

Else, Holly. 2023. “Abstracts written by ChatGPT fool scientists”. *Nature* 613, numero 7944 (tammikuu): 423. ISSN: 0028-0836. <https://doi.org/10.1038/d41586-023-00056-7>. <https://doi.org/10.1038/d41586-023-00056-7>.

French, John R, Bertram Raven ja Dorwin Cartwright. 1959. “The bases of social power”. *Classics of organization theory* 7 (311-320): 1.

Kang, Hyunjin, Ki Joon Kim ja Sai Wang. 2022. “Can the Internet of Things Persuade Me? An Investigation Into Power Dynamics in Human-Internet of Things Interaction”. *Frontiers in Psychology* 13. ISSN: 1664-1078. <https://doi.org/10.3389/fpsyg.2022.883110>. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.883110>.

Kurtzleben, Danielle. 2018. “Did fake news on Facebook help elect Trump? Here’s what we know”. *NPR*, April 11. <https://www.npr.org/2018/04/11/601323233/6-facts-we-know-about-fake-news-in-the-2016-election>.

LeCun, Yann, Y. Bengio ja Geoffrey Hinton. 2015. “Deep Learning”. *Nature* 521 (toukokuu): 436–44. <https://doi.org/10.1038/nature14539>.

- Littman, Michael L., Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfield ym. 2022. *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report*. <https://doi.org/10.48550/ARXIV.2210.15767>. <https://arxiv.org/abs/2210.15767>.
- Lum, Kristian, ja William Isaac. 2016. “To predict and serve?” *Significance* 13 (5): 14–19. <https://doi.org/https://doi.org/10.1111/j.1740-9713.2016.00960.x>.
- McCarthy, John, ym. 2007. “What is artificial intelligence”, <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>.
- McLeod, Carolyn. 2021. “Trust”. Teoksessa *The Stanford Encyclopedia of Philosophy*, Fall 2021, toimittanut Edward N. Zalta. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/entries/trust/>.
- Mei, Xueyan, Hao-Chih Lee, Kai-yue Diao, Mingqian Huang, Bin Lin, Chenyu Liu, Zongyu Xie, Yixuan Ma, Philip M Robson, Michael Chung ym. 2020. “Artificial intelligence-enabled rapid diagnosis of patients with COVID-19”. *Nature medicine* 26 (8): 1224–1228.
- Moor, James. 2006. “The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years.” *AI Magazine* 27 (tammikuu): 87–91.
- Nader, Karim, Paul Toprac, Suzanne Scott ja Samuel Baker. 2022. “Public understanding of artificial intelligence through entertainment media”. *AI & society*, 1–14.
- Oppy, Graham, ja David Dowe. 2021. “The Turing Test”. Teoksessa *The Stanford Encyclopedia of Philosophy*, Winter 2021, toimittanut Edward N. Zalta. Metaphysics Research Lab, Stanford University.
- Pennycook, Gordon, Tyrone D Cannon ja David G Rand. 2018. “Prior exposure increases perceived accuracy of fake news.” *Journal of experimental psychology: general* 147 (12): 1865.
- Sankaran, Supraja, Chao Zhang, Henk Aarts ja Panos Markopoulos. 2021. “Exploring Peoples’ Perception of Autonomy and Reactance in Everyday AI Interactions”. *Frontiers in Psychology* 12. ISSN: 1664-1078. <https://doi.org/10.3389/fpsyg.2021.713074>. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.713074>.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini ja Filippo Menczer. 2018. “The spread of low-credibility content by social bots”. *Nature Communications* 9, numero 1 (marraskuu). <https://doi.org/10.1038/s41467-018-06930-7>. <https://doi.org/10.1038/s41467-018-06930-7>.

Steindl, Christina, Eva Jonas, Sandra Sittenthaler, Eva Traut-Mattausch ja Jeff Greenberg. 2015. “Understanding psychological reactance”. *Zeitschrift für Psychologie*, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4675534/>.

Thiebes, Scott, Sebastian Lins ja Ali Sunyaev. 2021. “Trustworthy artificial intelligence”. *Electronic Markets* 31:447–464. <https://link.springer.com/article/10.1007/s12525-020-00441-4>.

Zhou, Yuwei, Yichuan Shi, Wei Lu ja Fang Wan. 2022. “Did Artificial Intelligence Inva-
de Humans? The Study on the Mechanism of Patients’ Willingness to Accept Artificial In-
telligence Medical Care: From the Perspective of Intergroup Threat Theory”. *Frontiers in
Psychology* 13. ISSN: 1664-1078. <https://doi.org/10.3389/fpsyg.2022.866124>. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.866124>.