

Kohti eettisesti toimivaa tekoälyn kehittämis- ympäristöä

Integroiva kirjallisuuskatsaus

LAB-ammattikorkeakoulu

Sairaanhoidaja (YAMK), Asiakkuusjohtaminen sosiaali- ja terveyspalveluissa

2022

Jonna Sirviö

Tiivistelmä

Tekijä(t) Sirviö, Jonna	Julkaisun laji Opinnäytetyö, YAMK	Valmistumisaika 2022
	Sivumäärä 49+6	
Työn nimi Kohti eettisesti toimivaa tekoälyn kehittämisympäristöä LAB WellTech		
Tutkinto Sosiaali- ja terveysalan ylempi ammattikorkeakoulututkinto, Sairaanhoidtaja (YAMK)		
Toimeksiantajan nimi, titteli ja organisaatio LAB ammattikorkeakoulu Oy		
<p>Tiivistelmä</p> <p>Tekoälyä käytetään jo nyt runsaasti ihmisten arjessa. Tekoälyn hyödyntäminen tulee lisääntymään tulevaisuudessa mikä herättää pohtimaan tekoälyn eettisyyttä ja periaatteita suunnittelussa ja kehittämisessä, jotta kehittämistyö olisi vastuullisesti kestävä. Tämän opinnäytetyön tarkoituksena on selvittää, millaisia tekoälyn eettisiä periaatteita on kansainvälisesti tunnustettu ja miten nämä periaatteet tulisi ottaa huomioon, jotta tekoälyn kehittäminen olisi vastuullisesti kestävä.</p> <p>Tämä opinnäytetyö toteutettiin integroivana kirjallisuuskatsauksena. Tietokantahaut toteutettiin LAB-ammattikorkeakoulun LUT-Primo ja Google Scholar sähköiseen aineistoon. Hakuprosessissa valikoitui 15 tutkimusta ja tieteellistä artikkelia. Sisällön analyysi suoritettiin induktiivisena sisällönanalyysinä.</p> <p>Aineiston analyysissä nousi esiin kolme yläluokkaa, joita tarkastellaan tekoälyn eettisen kehittämisen kannalta. Nämä yläluokat ovat, tekoälyn suunnittelun ja kehittämisen näkökulma, yhteiskunnallinen näkökulma ja ihmisyyden / ihmiskeskeisyyden näkökulma.</p> <p>Tulosten mukaan tekoälyn eettisyyden periaatteilta puuttuu kansainvälinen tunnustus, joka ohjaa kehittämistyötä globaalisti ja lisäksi periaatteita tulkitaan eritavoin, joka lisää eriarvoisuutta. Suunnittelu ja kehittämistyön tulee olla monialaista, jolloin voidaan välttää mahdollisia eettisiä ristiriitoja ja riskejä. Lisäksi yhteiskunnalliset eroavaisuudet haastavat eettiset periaatteet käytännössä.</p>		
Asiasanat Tekoäly, eettisyys, periaatteet, ohjeet		

Abstract

Author(s) Sirviö, Jonna	Type of Publication Master's thesis	Published 2022
	Number of Pages 49+6	
Title of Publication Towards an ethically working development environment for artificial intelligence		
Name of Degree Master's Degree Programme in Social- and Health Care Development and Management		
Name, title and organization of the client LAB University of Applied Sciences		
<p>Abstract</p> <p>Artificial intelligence is already widely used in people's daily lives. The utilization of artificial intelligence will increase in the future, which raises the question of the ethics and principles of artificial intelligence in design and development, so that development work is responsibly sustainable. The purpose of this thesis is to find out what ethical principles of artificial intelligence have been identified internationally and how these principles should be taken into account in order for the development of artificial intelligence to be responsibly sustainable.</p> <p>This thesis was implemented as an integrative literature review. Database searches were performed on LAB University of Applied Sciences' LUT-Primo and Google Scholar electronic data. 15 research and scientific articles were selected during the application process. Content analysis was performed as an inductive content analysis.</p> <p>The analysis of the data revealed three upper classes, which are examined in terms of the ethical development of artificial intelligence. These upper classes are, the perspective of the design and development of artificial intelligence, the societal perspective and the humanity / human focus perspective.</p> <p>According to the results, the ethical principles of artificial intelligence lack international recognition, which guides development work globally, and in addition, the principles are interpreted differently, which increases inequality. Design and development work must be multidisciplinary, so that potential ethical conflicts and risks can be avoided. In addition, societal differences challenge ethical principles in practice.</p>		
Keywords Artificial intelligence, ethics, principles, guidelines		

Sisällys

1	Johdanto.....	4
2	Tutkimuksen lähtökohdat	6
2.1	Tavoitteet ja tarkoitus.....	6
2.2	LAB- ammattikorkeakoulu.....	7
2.3	Tekoölypaja- hanke ja LAB WellTech	7
3	Tekoöly.....	9
3.1	Tekoöly käsitteenä.....	9
3.2	Tekoöly, etiikka ja eettiset periaatteet	12
4	Tekoölyn suunnittelu ja kehittäminen	17
4.1	Vastuullinen tutkimus- ja innovaatiotoiminta.....	17
4.2	Etiikkapesu	18
5	Tutkimuksen toteuttaminen.....	19
5.1	Integroiva kirjallisuuskatsaus	19
5.2	Tutkimusprosessi.....	20
5.3	Hakustrategian suunnittelu ja toteutus	21
5.4	Aineiston laadun arviointi ja esittely	26
5.5	Aineiston analyysi.....	30
6	Tulokset.....	34
6.1	Tekoölyn suunnittelun ja kehittämisen näkökulma	34
6.2	Yhteiskunnallinen näkökulma	38
6.3	Ihmisyiden näkökulma	40
7	Johtopäätökset ja pohdinta	42
7.1	Integroidun kirjallisuuskatsauksen arviointi ja pohdinta	42
7.2	Eettisyys ja luotettavuus	43
7.3	Johtopäätökset ja jatkotutkimusaiheet	44
	Lähteet	46

Liitteet

Liite 1. Aineiston esittely ja pisteytys

1 Johdanto

Tekoälyn tekniikat muokkaavat yhteiskuntaamme. Ne vaikuttavat jo jokapäiväiseen elämäämme. Samaan aikaan monet oikeudelliset ja yhteiskunnalliset kysymykset ovat paljastaneet näiden tekniikoiden potentiaalın tuottaa myös ei-toivottuja vaikutuksia. Algoritmit voivat lisätä ja vahvistaa jo olemassa olevia harhoja. Ne voivat uhata turvallisuuttamme, manipuloida meitä ja niillä voi olla vaarallisia seurauksia. Tekoäly nostettiin yhdeksi hallituksen kärkihankkeeksi vuonna 2017 elinkeinoministeri Mika Lintilän toimesta. (Työ- ja elinkeinoministeriö 2019.) Sanna Marinin hallitusohjelma jatkaa korostamalla eettisten ja sosiaalisten vaikutuksien ja yhdenvertaisuuden sekä sääntelyn tärkeyttä taloudellisesti kestäväen data- ja tekoälypolitiikan kehittämisessä. (Valtioneuvosto 2019, 73, 81.)

Eettinen ja vastuullinen tekoäly on noussut tärkeäksi puheenaiheeksi uusia tekoälysovelluksia kehittäessä. Euroopan komission tekoälyä käsittelevä korkean tason asiantuntijaryhmä (AI HLEG) julkaisi vuonna 2018 ensimmäisen asiakirjan koskien luotettavan tekoälyn eettisten ohjeiden perustaksi. Keskusteleminen aiheesta ei kuitenkaan riitä, Gartner nostaa esille teknologiatrendien strategioissa, että vastuullisen tekoälyn tarkoitus onkin siirtyä periaatteista ja julistuksista käytännön toimintaan organisaatioissa kuin yhteiskuntatasolla. (Burke 2020, 11.)

Tekoälyn hyödyntämiseen liittyy monia eettisiä kysymyksiä. Tällä hetkellä on keskitytty lähinnä yksityisyyden suojaamiseen, inhimilliseen arvokkuuteen ja turvallisuuden huomioimiseen. Tekoälyn käyttöönotosta näyttää nousevan monenlaisia haasteita, joita ei aina ennalta osata arvioida. Tekoälyn hyödyntäminen ei ole vain teknillinen haaste, vaan aina myös taloudellinen, sosiaalinen ja eettinen. Tekoälyn etiikka voidaan jakaa kolmeen alakategoriaan, jotka edustavat vastuusuhteita. Kategoriat ovat; dataetiikka, hyödyntämisen etiikka ja ohjelmoinnin etiikka. Ohjelmoinnin etiikka ja dataetiikka ovat alakäsitteitä hyödyntämisen etiikalle. Hyödyntämisen etiikka vastaakin kysymykseen, mitä ja millaisia arvoja hyödyntämisellä haetaan ja tavoitellaan sekä sen, kuinka se vaikuttaa itse yksilöön, yhteisöön ja koko yhteiskuntaan. (Koivisto, ym. 2019.)

Tekoäly on joukko eri teknologioita, joissa yhdistyvät algoritmit, data ja laskentateho. Kun haluamme rakentaa eettisesti luotettavaa tekoälyä on otettava huomioon tekoälyn mahdollisesti merkittävä vaikutus yhteiskuntaamme. Olennaisen tärkeää on, tekoäly perustuu arvoihimme, perusoikeuksiimme ja ihmisarvoihimme sekä yksityisyyden suojaamiseen. Näistä syistä meidän on laajasti tutkittava myös tekoälyjärjestelmien eettisiä, sosiaalisia ja oikeudellisia näkökulmia. Tekoälyn etiikassa kysytäänkin yleisesti – miten voimme kehittää

ja käyttää tekoälytekniikkaa eettisesti hyväksyttävällä ja kestäväällä tavalla. Mitkä ovat ohjaavat eettiset ja moraaliset periaatteet, joita meidän tulisi noudattaa. (Euroopan komissio, 2020; Ethics of AI, 2020.)

Tämän integroivan kirjallisuuskatsauksen tavoitteena on kuvata, millaisia tekoälyn eettisiä periaatteita on laadittu ja millaisia tekoälyn eettisiä periaatteita tulisi soveltaa tekoälyn kestäväen kehittämistyön tueksi. Tämän opinnäytetyön perusteella tullaan luomaan LAB- ammattikorkeakoulun LAB WellTech ja Tekoälypaja- hankkeen kehittämisympäristöön tekoälyn eettinen kehittämisprosessi, sivuten tietoturvaan vain eettisistä näkökulmista sekä laatia toiminnalle eettinen arviointimalli luotettavan tekoälyn varmistamiseksi.

2 Tutkimuksen lähtökohdat

2.1 Tavoitteet ja tarkoitus

Integroivan kirjallisuuskatsauksen tarve nousi esille keväällä 2020. Toimin LAB-ammattikorkeakoulussa TKI-asiantuntija Tekoälypaja- hankkeessa. Tehtävän kuvaani kuuluu kehitettävien tekoälysovellusten eettisten ja sosiaalisten vaikutusten arvioniti. Hankkeen päämääränä on luoda kehittämisympäristö tekoälyn kehittämiselle. Hankkeen edetessä tiivistimme yhteistyötä muiden teknologiahankkeiden kanssa ja aloitimme kehittämään suurempaa kehittämisympäristöä LAB WellTech, joka keskittyy hyvinvoinnin palveluinnovaatioiden kokonaisuuden ja hyvinvointiteknologian ympärille. Tekoäly ei kuitenkaan ole vain hyvinvoinnin palveluinnovaatioiden käytettävissä. Tekoälytiimi siirtyy toimimaan monialaisesti tulevaan ympäristöön. Kehittämisympäristöä suunnitellessa tiedostimme, että tekoälyn monialainen kehittäminen ja uusien sovellusten luominen tarvitsee lisää tutkimuksellista kehittämistä eettisen ja luotettavan tekoälyn kehittämisen tueksi.

Tämän integroivan kirjallisuuskatsauksen tavoitteena on selvittää, millaisia tekoälyn etiikan periaatteita on luotu, sivuten tietoturva ja lainsäädäntöä vain eettisistä näkökulmista sekä kartoittaa mitä periaatteita tekoälyn kehittämisessä tulisi soveltaa, jotta tekoälyn kehittäminen on kestävä

Integroivan kirjallisuuskatsauksen tarkoituksena on edistää eettistä toimintaa tekoälyn eri työvaiheissa ja tukea henkilöstön tietämystä, osaamista, sekä edistää tekoälyn turvallista ja vastuullista käyttöä. Kehittämisympäristössä tulee työskentelemään eri koulutustaustan ja -tarpeiden omaavia henkilöitä, ja näin tuetaan sekä tehostetaan monialaista yhteistyötä.

Integroivassa kirjallisuuskatsauksessa tarkastellaan myös, mitä eettisen tekoälyn edistämiseksi on jo tehty kansallisella ja kansainvälisellä tasolla. Työssä tutustutaan eri toimijoiden laatimiin tekoälyn eettisiin periaatteisiin, niitä tarkasteleviin julkaisuihin ja käyttökokemuksiin.

Tutkimuskysymykset ovat:

1. Millaisia tekoälyn eettisiä periaatteita on laadittu kehittämistyön tueksi?
2. Miten tekoälyn eettisiä periaatteita tulisi soveltaa LAB WellTech kehittämisympäristössä, jotta se olisi kehittämistyön kannalta kestävä?

2.2 LAB- ammattikorkeakoulu

LAB- ammattikorkeakoulu syntyi 1.1.2020 Lahden ja Saimaan ammattikorkeakoulujen yhdistymisen myötä. Yhdistymisen tavoitteena on ollut vahvistaa ja uudistaa korkeakoulutusta ja työelämää. LAB- ammattikorkeakoulu toimii työelämän innovaatiokouluna, joka tarjoaa opiskelun ja työelämän parhaat puolet. LAB-ammattikorkeakoulun strategiana on luoda parempaa työelämää ja pyrkiä rakentavaksi innovaatiokorkeakouluksi. LAB- ammattikorkeakoululla on kolme kampusta, Lahdessa, Lappeenrannassa ja verkossa. Opiskelijoita on yli 8500, opettajia ja TKI- asiantuntijoita yli 500. LAB- ammattikorkeakoulu ja LUT- yliopisto muodostavat yhdessä kahden itsenäisen korkeakoulun konsernin. Tämän konsernin tarkoituksena on, että molemmat osapuolet hyötyvät yhdessä tekemisestä. (LAB-ammattikorkeakoulu, 2020.)

LAB- ammattikorkeakoulun strategiana on kasvava kansainvälinen TKI-toiminta. TKI-toiminnalla haetaan ratkaisuja kestävään kehitykseen, ihmisten hyvinvointiin, materiaalien käyttöön ja luodaan käyttäjälähtöisiä innovaatioita, palveluja sekä tuotteita. TKI-toiminta muodostuu neljän vahvuusalueen ympärille, joille jokaiselle on laadittu vahvuusalueen tiekartat. Nämä tiekartat toimivat tulevien vuosien toimintasuunnitelmina. Vahvuusalueita LAB:ssä on neljä: hyvinvointi; kestävät hyvinvoinnin palveluinnovaatiot ja ratkaisut, kiertotalous; kestävä materiaalikierro, innovaatiot; innovaatioiden kaupallistaminen ja design; kestävien ratkaisujen muotoilu. LABin TKI-toiminnan tavoitteena on alueelliseen kasvuun tähtäävien työpaikkojen lisääminen ja korkeakoulutettujen määrän lisääminen työelämässä. (LAB-ammattikorkeakoulu, 2020.)

2.3 Tekoälypaja- hanke ja LAB WellTech

Tekoälypaja englanniksi AI Mill hanke on LAB- ammattikorkeakoulun hallinnoima tekoälyyn ja sensoriteknologiaan keskittyvä kehittämishanke. Hankkeessa yhdistyy hyvinvointi, sosiaali- ja terveysala ja kiertotalous. Pää tavoitteena hankkeella on edistää Päijät-Hämeen alueen hyvinvointialan ja kiertotalousalan pienten ja keskisuurten yritysten liiketoimintaa ja sensoriteknologian hyödyntämistä. (LAB-ammattikorkeakoulu, 2019.)

Hanke kehittää yhdessä alueen yritysten kanssa heidän tarpeitansa tukevaa tekoälyn hyödyntämiseen innovaatio- ja tuotekehitysympäristöä ja mahdollistaa tuotekehitys- ja neuvonta- ja ohjauspalveluja Tekoälypajan yritysten tueksi. Lisäksi luodaan tekoälyä kehittävä ja yritysten toimintaa edistävä monialainen yhteistyöverkosto. Hankkeen aikana kehitetään ja rakennetaan erilaisia tekoälyyn ja sensoriteknologiaan pohjautuvia sovelluksia ja ratkaisuja demonstraatiomalleiksi ja pilotoitavaksi yrityksissä. (LAB-ammattikorkeakoulu, 2019.)

Tekoölypajan toimintamalli lisää alueen tekoölyn soveltamismahdollisuuksien osaamista. Hanke edistää Päijät- Hämeen alueen hyvinvointi- ja sote-alan ja kiertotalouden aloilla toimivien pienten ja keskisuurten yritysten kasvua ja kehittämistä. Se tarjoaa myös uusia tapoja liiketoimintamallien käynnistymiselle. Hanke antaa hyvinvointi- ja sote-alalla sekä kiertotalouden parissa toimiville yrityksille tukea ja neuvoa uusien palveluinnovaatioiden kehittämiseen. (LAB-ammattikorkeakoulu, 2019.)

LAB WellTech kehittämissympäristön toiminnan käynnistyi Päijät-Hämeessä ja Etelä-Karjalassa 1.1. 2021. LAB WellTech (kuvio 1.) on matalan kynnyksen hyvinvointiteknologian testi-, kehittämis- ja pilotointiympäristö sote- ja teknologia-alan julkisille toimijoille, yrityksille ja start-upeille, joka toimii hyvinvoinnin vahvuusalueen alla. LAB WellTech on helposti lähestyttävä taho, joka tarjoaa moniammatillisia testbed-palveluita tuotekehityksen koko kaarella, mahdollisuus kehittää ideoita ja tuoteaihoita, testata ja pilotoida aidoissa sote-alan käyttäjä ympäristöissä tai simulaatioympäristöissä. Kehittämissympäristön rakentamisen taustalla ovat LABn hankkeet HyTeLab, Elsatestbed (nämä hankkeet päättyivät 31.12.2020), Tekoölypaja ja Konenäön uudet sovellukset sote- ja hyvinvointialalla. (LAB-ammattikorkeakoulu, 2021.)

LAB WellTech kehittämissympäristö tarjoaa ja mahdollistaa teknologia-alan yrityksille pääsyn yhteistyöhön ja yhteiskehittämiseen sote-alan toimijoiden kanssa. Matalan kynnyksen testbed-toimintaa start-up yrityksille mahdollistaa verkostoitumisen sote- ja teknologia-alan toimijoiden välillä ja lisää sote-alan toimijoiden tietoisuutta tarjolla olevista teknologisista ratkaisuista sekä teknologian kehittäjien tuntemusta sote-alan tarpeista. (LAB-ammattikorkeakoulu, 2021.)



KUVIO 1. LAB WellTech Hyvinvoinnin ja terveyden edistämisen innovaatioalusta (mukaan LAB WellTech, 2021.)

3 Tekoäly

3.1 Tekoäly käsitteenä

Tekoälystä englannin kielen termi Artificial Intelligence, lyhenne AI, käydään paljon keskustelua, muuta määritelmänä se ei ole yksiselitteinen. Usein tekoälystä keskustellessa käytetään myös termejä keinoäly, syväoppiminen ja koneoppiminen. Pohjimmiltaan tekoälyn tekniikka on matematiikkaa, ohjelmointia ja tilastotiedettä. (Kananen & Puolitaival, 2019, 27.) Merilehdon (2018, 18) mukaan tekoälyn määritelmänä, voidaan pitää, että kone suorittaa toimintoja, jotka ihmisten tekeminä olisi älykäästä. Hän muistuttaa, että tekoälyä on kehitettävä siten, ettei se rajoitu ihmisen älylliseen tasoon a tulisi pystyä itsenäiseen oppimiseen, ennakoimiseen ja kykenee päättämään ihmistä paremmin. (Merilehto 2018, 18.)

Aluksi tekoäly kehitettiin ja luotiin vastaamaan ihmisen ajattelua ja päättelykykyä. Tekoäly jaetaan (kuvio 2.) kapeaan eli heikkoon, vahvaan eli yleiseen ja super- tekoölyyn. Vahvan ja super- tekoälyn rakentamisen sijaan on vielä tällä hetkellä keskitetty kapeaan eli heikkoon tekoölyyn, joka suoriutuu yksittäisistä ja sille määritellyistä tehtävistä. Vahvan tekoälyn mahdollisuuksia pyritään jatkuvasti kehittämään, sen edellytyksenä on, että tekoäly pystyy oppimaan itsenäisesti, ilman ohjausta ja yleistämään oppimansa. Neuroverkot toimivat ja oppivat jo hyvin autonomisesti ja syvät neuroverkot osaavat jo osittain yleistämään. Kuitenkin vahvan tekoälyn kehittäminen on vielä vuosikymmenten päässä. (Merilehto 2018, 18, 23–25.)



KUVIO 2. Tekoälyn kehitysaskleet (mukaillen, Ailisto ym. 2018).

Filosofi Ollila (2018, 8) ehdottaa, että voisimme puhua heikosta tekoälystä termillä ”koneoppiminen”. Täytyy kuitenkin ymmärtää, ettei kaikki tekoäly ole koneoppimista, eikä kaikki koneoppimiseen liittyvät alalajit pidä sisällään tekoälyn ominaisuuksia, mikäli ajattelemme, että tekoälyn älykkyys on automaattista, itseoppimisista, adaptiivisuutta tai korkeaa suorituskykyä. Eri käsitteiden selventämiseksi oleelliset tekoälyn käyttö- ja osaamisalueet on hyvä avata. Ailisto ym. (2018,1) jakaa tekoälyn kymmeneen tieteellisteknologiseen alueeseen, tässä työssä tarkastellaan tarkemmin data-analyysia, koneoppimista, syväoppimista ja neuroverkkoja.

Data-analytiikan osaamis pohja luodaan datatieteessä. Datalla tarkoitetaan kerättyä tietoa, joka on kerättyä tietoa, jostakin lähteestä. Data-analytiikalla tarkoitetaan datan analysointia ja puhdistamista tiedoksi ja siitä vielä johtopäätöksiksi. Data-analytiikkaan kuuluu kokonainen sarja toimintoja alkaen datan esikäsittelystä, visualisointiin, erittelyyn ja tulkintaan. Datat hallintaan ja esikäsittelyyn kuuluu myös datan yhdistäminen eri datalähteistä, muokkaaminen, suodattaminen, tarkistaminen, puhdistaminen ja annotointi (jossa kerättyyn aineistoon liitetään kuvaamiseen, luokitteluun ja jäsentelyyn liittyviä tietoa) sekä datan hankinta ja sen tallentaminen. (Ailisto ym. 2018, 8.)

Data-analytiikka auttaa tekemään saatuun tietoon eli dataan perustuvia päätöksentekoa. Data-analytiikan eri sovellukset ovat joko pieniä tilastoanalyyssejä tai isoja tekoälyä ja koneoppimista soveltavia ja hyödyntäviä. Se auttaa vähentämään ihmisistä johtuvia virheitä, joita esiintyy rutiininomaisessa työssä ja sen avulla pystytään visualisoimaan tieto ihmiselle ymmärrettäväksi. Data-analytiikalla voidaan selvittää asiayhteyksien välisiä suhteita ja eri tekijöitä, jotka vaikuttavat työn tehokkuuteen tai laatuun. Data-analytiikka mahdollistaa löytämään parhaat ratkaisut päätöksentekemisen tueksi. (Harjamäki, 2018, 41.)

Ailisto ym. (2018, 46) määrittelee koneoppisen englannin kielen termi machine learning tietotekniikan osa-alueeksi, jossa käytetään yleisesti tilastotieteiden menetelmiä. Nämä menetelmät antavat tietokoneelle kyvyn oppia datasta. Toisin sanoen voidaan todeta, että menetelmät parantavat suorituskykyä tehtävän suorittamisessa ilman täsmällistä ohjelmointia. Roos (2018) mukaan koneoppiminen voidaan jakaa kahteen osa-alueeseen. Koneeseen, jolla tarkoitetaan tietokonetta tai tietokoneohjelmaa sekä oppimiseen, jolla tarkoitetaan kykyä oppia erilaisten esimerkkien kautta ja avulla. Ohjelmoijan ei tarvitse koneoppimisen näkökulmasta luoda algoritmeja tai sääntöjä jokaisessa oppimisen vaiheessa, vaan kone oppii itsenäisesti toimimaan esimerkkien, pohjatietojen ja erinäisten tapahtumien avulla. Koneoppimisella on rajattomasti potentiaalia ja käyttömahdollisuuksia. Ratkaiseva asia koneoppimiselle on datalähde. Tällä hetkellä dataa on saatavilla rajattomasti eri tietolähteistä,

terveydenhuolto, teollisuus, organisaatioiden tietolähteet, pankit ja sosiaalinen media. (Roos 2018.)

Algoritmit ovat koneoppimisen perusta. Opetusalgoritmit oppivat ja kehittyvät useista toistuvista tapahtumista. Tällainen vaihe vaiheelta tapahtuva oppiminen tarvitsee käytettäväksi dataa ja mitä enemmän sitä on saatavilla, sitä paremmin se kykenee luomaan ennustuksia. (Merilehto 2018, 27–28.) Oppimista kutsutaan prosessiksi, jossa opetusalgoritmien kokemus muuttuu tiedoksi. Algoritmit rakennetaan jäljittelemään oppimisprosessia, jotta ne voivat luoda annetusta tiedosta kokemuksia. Tämän algoritmi kykenee muuttamaan tiedoksi, jota voidaan käyttää erillisissä tehtävissä hyödyksi. (Shalev-Shwartz & Ben-David, 2014, 1.)

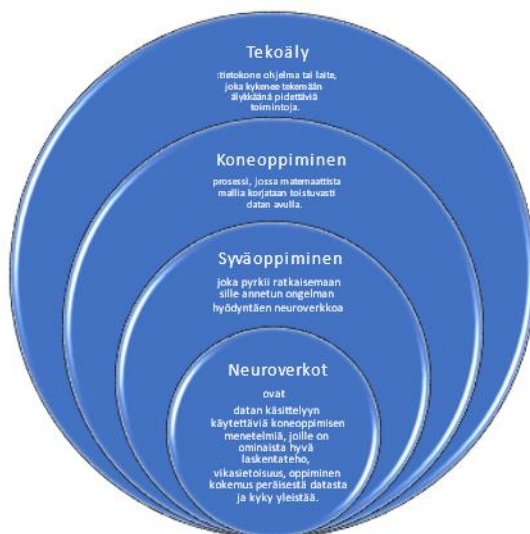
Koneoppiminen ja sen tuoma hyöty data-analyysissä syntyy koneoppimisen mahdollisuudesta nopeaan oppiseen ja toimintaan ja näin tuottaa monimutkaisista data-aineistoista järkeviä päätelmiä, joihin ihminen ei kykenisi. Digitaalista dataa on saatavilla paljon ja datan seassa on hyödyllistä tietoa. Tämän tiedon järjestely vaatii ihmistä tehokkaamman toimijan sitä ymmärtämään. Tietokoneet sekä ohjelmat kehittyvät jatkuvasti ja hyödyntävät laajaa muistia. Prosessointinopeuden kehittyminen avaa uusia mahdollisuuksia datan hyödyntämiseen koko ajan. (Shalev-Shwartz & Ben-David, 2014, 3–4.)

Syväoppiminen (Deep learning) ja neuroverkot kuuluvat koneoppimisen osa-alueisiin. Syväoppiminen perustuu laskentatehoon, sen kehitykseen ja datan kasvavaan määrään. Tämä mahdollistaa sen laajan soveltamisen. Neuroverkot ovat kerroksellisia matemaattisista yksiköistä eli neuroneista koostuvia verkkoja. Neuronit ovat yhteydessä toisiinsa ja kukin neuroni havainnoi sille annettua asiaa sekä välittää prosessoimansa tiedon eteenpäin. Neuroverkkojen kerroksilla on oma tehtävä ja mitä useampia kerroksia neuroverkossa on, sitä syvemmästä verkosta on kysymys. Tästä muodostuu termi syväoppiminen. Mitä enemmän neuroverkoilla on opetusdataa käytettävissä, sitä paremmin ne toimivat. (Merilehto 2018, 20–57, 163.)

Goodfellow ym. (2018, 1–16) määrittelevät syväoppimisen järjestelmäksi, joka kykenee koneoppimisen avulla oppimaan kokemuksista ja ymmärtämään ympäristöä rakentamalla käsitteitä opetusdatasta sekä muodostamaan näistä hierarkkisia käsitteitä. Syväoppiminen hyödyntää neuroverkkoja. Sen toimintamallit jäljittelevät biologista ihmisen kaltaista oppimista. Biologinen tapa lähestyä syväoppimista ja sen kehittämistä motivoivana tapana voidaan jakaa kahteen osaan. Ensimmäisessä osassa ihmisten ja eläinten aivot tarjoavat todisteita, että biologinen älykkyys on mahdollista ja samanlainen järjestelmä on melko yksinkertainen mallintaa koneen avulla. Toiseksi koneoppimismallit voivat ehkä antaa meille paremman ymmärryksen ihmisälyn toiminnasta ja biologista aivoista, joista ihmisäly

muodostuu. Nykyaikaisiin syväoppimismenetelmiin vaikuttavat kuitenkin monet alat, jotka eivät välttämättä ole neurotieteestä. Merkittävässä osassa ovat esimerkiksi matematiikan algebra, todennäköisyyslaskenta, informaatioteoria ja numeerinen optimointi.

Kuviossa 3 on visualisoitu tekoäly, koneoppiminen, syväoppiminen ja neuroverkot. Tekoäly on osa tietojenkäsittelytiedettä, koneoppiminen osa tekoälyä, syväoppiminen taas osa koneoppimista ja neuroverkot syväoppimista. Kaikki edellä mainitut ovat osa datatiedettä. (Elements of AI, 2018.)



KUVIO 3. Tekoälyn ympyrä, jossa esitetään tekoälyn keskeiset tekniikat ja käsitteet. (muokailen Elements of AI, 2018.)

3.2 Tekoäly, etiikka ja eettiset periaatteet

Etiikka ja eettisyys ymmärretään oikean ja väärän sekä hyvän ja pahan toisistaan erottamisena. Käsitteet etiikka ja moraaliksi sekoittuu yleensä arjessa synonyymeiksi toisilleen, vaikka ne eivät täysin samaa tarkoita. Etiikka on oma tieteenala, jonka tarkoituksena on tutkia moraalialia. (Laaksonen & Ollila 2017, 150; Leikas 2008, 59.)

Etiikka tieteenalana kiteytyy elämän säilyttämiseen. Se on ihmisen moraalista toimijuutta tutkiva ala, joka sisältää siveysopin, eettiset normit ja moraalisäännöt. Moraalin määrittelynä pidetään hyvän elämän tavoittelua, siihen liittyviä arvoja, käytäntöjä ja eettisiä tapoja toimia. Eettisyydellä tarkoitetaan myöskin ihmisen omien tekojen sekä arvojen hyväksyttävyyttä ja tuomittavuutta, tämä pitää sisällä myös yksilön toiminnan tavoitteet ja pyrkimyksen hyväksyttävään ja oikeaan toimintaan. Eettisesti käyttäytyvä ihminen jättää toimissaan

huomioimatta omia etuja ja käyttäytyy epäitsekkäästi. Epäeettinen ihminen puolestaan jättää huomiotta yhteisesti sovitut käytännöt, eivätkä edistä yhteistä hyvää teoillaan ja valinnoillaan. (Sivistyssanakirja 2002; Kotimaisten kielten keskus 2021; Ollila 2019, 31–33.)

Yleisesti ottaen etiikkaa voidaan pitää taitona, jonka opettaminen on vaikeaa. Voimme kuitenkin opiskella etiikan eri teorioita, mutta tämä on kuitenkin vasta alku eettiselle ajattelulle. Etiikkaa tutkiessa voimme kuitenkin oppia elämäkatsomuksesta ja elämästä. (Heinimäki 2012, 81–82.) Ominaista etiikalle on, ettei se anna valmiita eettisiä vastauksia, joita ei voisi näkökulmien mukaan kyseenalaistaa tai kritisoida. Lähtökohtana etiikalle on, että emme ole yksin maailmassa, teoissamme tulee huomioida myös muut ihmiset. (Heinimäki 2012, 21–22; 51; Kotkavirta & Nyysönen 1996, 16.)

Hiekkänen (2017) kirjoittaa artikkelissaan digitalisaatioon liittyvästä eettisestä keskustelusta. Artikkelin mukaan eettinen keskustelu on unohdettu eri tekniikanalan ammattilaisten keskuudessa, kun kehitetään digitaalisia sovelluksia ja järjestelmiä. Hiekkänen nostaa esille, että digitalisaatio ja siihen liittyvien eri mahdollisuuksien ja hyödynnettävyyden rinnalla, tulisi keskusteluun nostaa myös eettinen ulottuvuus. (Hiekkänen, 2017.) Digitaalisen yhteiskunnan infrastruktuuri ja rakenneosat toteutuvat jo nyt. Tulisi kehittää hallinnollinen järjestelmä, jonka tehtävänä olisi käsitellä syntyviä eettisiä ja sosiaalisia haasteita yhteiskunnallisesti siksi, että digitalisoinnin sosiaalisia ja eettisiä kysymyksiä ei ole selkeää ymmärtää. (Royackers ym. 2018.)

Puhuttaessa tekoälystä siinä on aina mukana moraalisia ja eettisiä kysymyksiä, sillä tekoäly tekniikan ja teknologian alana tavoittelee vaikuttamaan ja muuttamaan maailmaa. Kysymykset tekoälyn ympärillä vaihtelevat luonteeltaan teknisistä, (virheiltä välttyminen), taloudellisiin (työttömyys, vaurauksien keskittyminen) yhteiskunnallisiin (ihmisen käyttäytyminen, vuorovaikutus) ja kyberturvallisuuteen (kyberhyökkäykset) ja moraalifilosofiaan (koneiden oikeudet ja tietoisuus). (Ailisto ym. 2018, 21–22.)

Tekoälyn etiikka ei tarkoita sen synnyttämää etiikkaa, sillä tarkoitetaan tekoälyn kehitykseen ja käyttämiseen tarkoitettua sovellettavissa olevaa etiikkaa, jonka avulla tekoälyä kyetään tarkastelemaan ja tutkimaan eri näkökulmista. Tekoälyn etiikka tarkoittaaakin tekoälyn käytötarkoitusten tuloksien ja toteutusten sekä eettisyyttä. (Jääskeläinen 2019, 81; Ollila 2019, 11, 146.) Tekoälyyn kytkettyjä moraalien ja etiikan osa-alueita ovat esimerkiksi seurausetiikka, hyve-etiikka, soveltavat etiikan alat, teknologioiden etiikka ja sodankäynnin etiikka. Tekoälyn käyttäminen luo uusia haasteita ja näkökulmia, koska esimerkiksi teknologian etiikka perustui aikaisemmin teknologiaa suunnittelevan ihmisen omaan toimijuuteen, eikä siihen, että moraalisen ja eettisen päätöksen tekisi kone, joka toimii autonomisesti. (Ailisto ym. 2018, 21–22.)

Ollila (2019, 13, 52–53) kyseenalaistaa, riittääkö tähän mennessä laadittu etiikka selvittämään tekoälyyn liittyvät ongelmat, koska tekoälynominaisuuksista etenkin autonomisuus ja mukautumiskyvykyys luovat etiikalle haasteita. Anna Seppänen muistuttaa, ettei tekoälyn eettiset kysymykset ole osa teknologiaan liitettävä etiikkaa. Tekoäly kasvaa eri aloilla ja kaikkialla, myöskin moraalisen toiminnan alueilla, kuten hoiva- ja hoitotyössä ja demokratiassa. Teko-älyn etiikkaa ei voida pitää vain etiikkana, sillä moniulotteiset etiikan haasteet ja kysymykset ovat vielä muotoutumassa. (Seppänen 29.10.2020.)

Tekoälyn ja teknologian etiikassa laajemmin on välttämätöntä keskustella avoimesti ja päättää mitä pidämme oikeana ja mikä on väärin, millaisia arvoja edistetään ja mitä normeja noudatetaan sekä täsmentää eri periaatteiden tarkoitus teknologioiden suunnittelussa ja käytössä sekä altistaa etiikka julkisesti käytävälle keskustelulle. Tällaisessa tapauksessa meidän on pakko miettiä myös niitä uhkia, joita eettiset periaatteet voivat aiheuttaa ja mitä ne voivat aiheuttaa teknologioille. (Koivisto ym. 2019, 10.)

Coeckelbergh (2021,1, 68–69, 53–60) toteaa tekoälyn etiikasta, että kyse on nykyisyyden ja tulevaisuuden haasteista yhteiskunnassamme ja sen vaikutuksista ihmisiin, demokratiin ja elämään. Tekoälyn etiikka liittyy oleellisesti päätöksen tekoon ja ihmisiin meidän tulee puuttua eettisiin haasteisiin heti. On ruvetta etsimään käytännönläheisiä haasteita ja asioita, jos halutaan, että tekoälyn etiikka ei ole vain tarkastuslista. Coeckelbergh muistuttaa meitä myös siitä, että tekoälyjärjestelmät ovat ennen kaikkea työkaluja ihmisten tavoitteiden saavuttamiseksi. Koska eettiset normit voivat kuitenkin vaihdella paikasta toiseen ja henkilöstä toiseen, on tärkeää, että näiden järjestelmien suunnittelijat muistavat, että päätöksenteko ei ole vain mustavalkoista. Tekoälyn on kyettävä ymmärtämään ja jäljittelemään sen kulttuurin moraalikieltä, jossa sitä käytetään.

2019 keväällä yhteensä 42 valtiota allekirjoittivat OECD:n periaatteet tekoälylle. Nämä ovat ensimmäiset luotettavaa tekoälyä koskevat hallitustenväliset poliittiset suositukset. (OECD 2019a.) Valtiot sopivat toimivansa tavoitteellisesti, jotta tekoälyjärjestelmät suunnitellaan ratkaisuiltaan turvallisiksi, kestäviksi, luotettaviksi ja oikeudenmukaisiksi. OECD:n suositukset pitävät sisällään viisi arvopohjiin nojautuvaa periaatetta, joiden tarkoitus on luoda luotettava ja kestävä tekoälyn käyttö ja käyttöönotto. Nämä periaatteet muodostuvat osallistavan ja kestävä kasvun sekä hyvinvoinnin edistämisen, oikeudenmukaisuuden ja ihmiskeskeisten arvojen ja normien kunnioittamisen, tekoälyjärjestelmien avoimuuden, selitettävyyden, turvallisuuden ja vakauden sekä tekoälyn kehittäjän, käyttöönottajajan ja kehittäjän vastuun. Näiden suositusten tarkoituksena on tukea Euroopan komission luomaa ohjeistusta luotettavaan ja eettiseen tekoälyyn. Euroopan komission ohjeistuksen vahvistamisen lisäksi

näistä OECD:n suosituksista löytyy myös samankaltaisuuksia maailman laajuisten järjestöjen laatimiin suosituksiin. (OECD 2019b.)

Koivisto, ym. (2019) nostaa tekoälyn periaateisiin

Itsemääräämisoikeuden kunnioittaminen on yksilön ihmisarvon ja vapauden periaate. Tärkeää on, että ihminen saa ilmaista vapaasti omia tunteitaan, arvojaan, omia valintojaan ja päättää itse tehdyistä valinnoistaan. Itsemääräämisoikeus tarkoittaa, että emme saa mitätöidä henkilön itsemääräämisoikeutta pakottamalla, uhkaamalla tai muulla vapautta rajoittavalla tavalla, vaikka hänen tekemä päätös meistä tuntuisi olevan huono, eikä se mielestämme edistäisi inhimillistä hyvää. Tämä pitää sisällään myös älä vahingoita toisia vahingoittamisen periaatteen sekä toiminnan- ja valinnanvapauden kunnioittamisen.

Yhteisen hyvän, yhteisöllisyyden ja oikeudenmukaisuuden periaatteet vaativat huomiota. Oikeudenmukaisuuden periaate pyrkii katsomaan, kuinka yhteisössä yhteisön jäsenten tulisi toimia ja miten haitat ja edut tulisi jakaantua yhteisössä. Oikeudenmukaisuus koskee siis myös sitä, mitkä oikeudet ja velvollisuudet yhteisön jäsenillä on. Oikeudenmukaisuus voidaan ymmärtää hyvin eri tavoin riippuen minkä näkökulman yhteisö on valinnut.

Tasa-arvoisuusperiaatteen mukaan etiikka ei saa edistää kenenkään omaa hyvää esimerkiksi rikkaita köyhien kustannuksella, erinomaisia heikkojen kustannuksella, miehiä naisten kustannuksella, valkoihoisia värillisten kustannuksella, nuoria henkilöitä vanhojen henkilöiden kustannuksella, terveitä vammaisten kustannuksella jne. Riippumatta henkilön sosiaalisesta statuksesta tai taustasta, jokaista ihmistä tulee kohdella samalla tavalla etujen ja haittojen arvioinnissa.

Kun arvioidaan teknologian yhteyttä yksilöön ja yhteisöön sekä näiden suhdetta toisiinsa nousee esiin päällekkäisiä periaatteita. Leikas (2008) määrittelee tekoälyn periaatteet myös kolmeen kategoriaan, joissa huomio kiinnittyy eri käyttäjien näkökulmiin, jotka ovat tässä opinnäytetyössä keskeisiä. (Leikas 2008.)

Teknologian käyttäjän näkökulmasta:

Ihmisarvokkuus ja sen loukkaamattomuus, yksilön oikeuksien kunnioittaminen ja itsemääräämisoikeus. Yksilön tietoinen suostumus teknologiaan, oikeus kieltäytyä, luottamus teknologiaan, pätevyys, yhdenvertaiset mahdollisuudet kaikille, tasa-arvo, yksilöllisyys, yksilön osallistuminen, yksilön suojaaminen ja valvonta sekä turvallisuus, saavutettavuus, vahingoittaminen, valinnanvapaus, vapaaehtoisuus ja yksityisyyden suoja.

Yhteisön näkökulmasta:

Hyödyn tasapuolinen jakautuminen, kulttuurien moninaisuus, yhteistyö, yhteisesti sovitut sopimukset, syrjintä ja teknologian sosiaalinen ja yhteiskunnallinen vaikuttavuus.

Teknologian kehittäjän näkökulmasta tekoälyjärjestelmien turvallisuus, valvonta, yksimielisyys, pätevyys, vastuullisuus, tekijänoikeuksien kunnioitus ja ymmärrys.

Latvanen (2019) mukaan tekoälyjärjestelmien eettiseen vaikuttavuus arviointiin kannattaisi soveltaa joustavaa mallia, missä osa periaatteista tunnistettaisiin ryhmiksi ja tapauskohtaisesti käyttöyhteyksien mukaan. Ryhmitetyt periaatteet voisivat liittyä esimerkiksi turvallisuuden ja yksityisyydensuojaan, kun taas esimerkiksi hallintoon ja talouteen liittyvät periaatteet voitaisiin käsitellä tapauskohtaisesti. Tekoälyn eettinen ja kestävä käyttäminen vaatii yhteiskunnalta hyväksytyjen eettisten periaatteiden tietoisuutta, jonka jälkeen on mahdollista ottaa käyttöön periaatteet sääntelyyn ja tekoälyjärjestelmien suunnitteluun.

Tekoäly teknologioina tuo esille uusia tekoälyjärjestelmien toimintaan ja käytettävyyteen liittyviä eettisiä ongelmia. Niitä käsitellään seuraavassa luvussa

4 Tekoälyn suunnittelu ja kehittäminen

4.1 Vastuullinen tutkimus- ja innovaatiotoiminta

Tekoälyteknologioita suunnitellessa, kehittäessä ja käyttöönottamisessa lähestymistapana voidaan pitää mahdollisuutta tehdä hyvää sekä mahdollisuutena asemoitua eettisesti ja vastuullisesti toimijana. Teknologia voi tuottaa epävarmuutta yksilöllisellä ja yhteiskunnallisella tasolla. Onkin tärkeää pyrkiä edistämään eettisiä käytäntöjä ja sääntöjä tekoälyn kehittämisessä ja käyttöönottamisessa. (Ojanen ym. 2019, 11; ETAIROS, 2019.)

Vastuullinen tutkimus- ja innovaatiotoiminta on noussut esille erityisesti Euroopan Komission toimesta. RRI (Responsible Research and Innovation) ajattelun taustalla on teknologian ja tieteen yhdessä muodostamien riskien ja erilaisten tutkimustulosten hyväksyttävyydestä yhteiskunnallisesti. Vastuullinen tutkimus- innovaatiotoiminta määritellään vuorovaikutteiseksi, läpinäkyväksi prosessiksi. Prosessin kautta yhteiskunnallisten toimijoiden ja innovoijien näkemykset tulevat avoimiksi toisilleen, teknologian eettisyyden hyväksyttävyydestä, yhteiskunnallisesta tuottavuudesta ja kestävydestä. Yhteiskunnallinen vuorovaikutus toimijoiden kesken ja on keskeistä. Ihmisoikeuksien ja eettisyyden kunnioittaminen on vastuullisen tutkimus- ja innovaatiotoiminnan kulmakivi, muita kulmakiviä ovat tasa-arvo, toimijoiden osallistaminen sekä avoimuus. Sen periaatteet korostavat innovaatioiden yhteiskehittämistä ajatuksella, että ottamalla mukaan käyttäjät ja muut kehittämiseen osallistuvat toimijat kehittäminen tapahtuu yhteiskunnallisesti ja syntyy vakuuttavia innovaatioita. Useat teknologiset innovaatiot pyrkivät muutokseen yhteiskunnassa tai yhteiskunnallisten haasteiden ratkaisemiseen. Tällöin on tärkeää, että vastuullisuus tulee jakaa kehittämisessä mukana olleiden kesken ja että kaikista toimijoista tulee vastuullisia vaikutusten, tulosten ja seurausten suhteen. (Koivisto ym. 2019, 16.)

Institute of Electrical and Electronics Engineers (IEEE) on julkaissut usean asiantuntijan kirjoittaman A Vision for Prioritizing Human Well-being with Autonomous and Intelligent System- julkaisun autonomisten ja älykkäiden tekoälyjärjestelmien eettisestä ja arvopohjaisesta suunnittelusta, kehittämisestä ja toteuttamisesta. EADv2 -julkaisu on kattavin, väestölähteen ja maailmanlaajuinen tutkielma, joka koskee toistaiseksi saatavilla olevaa itsenäisten ja älykkäiden järjestelmien etiikkaa. Yli 250 maailmanlaajuisen poikkitieteellisen tutkijan luoma EADv2 koostuu yli sadasta käytännöllisestä suosituksesta, joita teknologian kehittäjät, päätöksentekijät ja tutkijat voivat hyödyntää. Tämän asiakirjan tarkoitus on käynnistää julkaista keskustelua, kuinka voimme luoda eettisiä ja sosiaalisia toteutuksia älykkäille järjestelmille ja mukauttaa ne määrättyihin arvoihin ja eettisiin periaatteisiin. (IEEE, 2019.)

AlgorithmWatch, joka on kansainvälinen voittoa tavoittelematon tutkimus- ja vaikuttamisjärjestö, seuraa ja listaa kansainvälisesti tekoälyn kehitystä ja julkaisee sivuillaan aiheesta laadittuja dokumentteja. Se on sitoutunut tarkkailemaan, purkamaan ja analysoimaan automaattisia päätöksentekojärjestelmiä ja niiden vaikutuksia yhteiskuntaan ja ihmisen käyttäytymiseen ja osoittamaan eettisiä ristiriitoja. (AlgorithmWatch, 2021.)

Leikas (2021) mukaan tekoälyn soveltaminen ja hyödyntäminen voi synnyttää uudenlaista epäarvoisuutta jo olemassa oleviin yhteiskunnan rakenteisiin. Tekoälyn kehittäminen on suurelta osin myös poliittinen kysymys, jolla on vaikutuksia lain ja vallan kysymyksissä. Tästä voi aiheutua riskejä sen väärinkäytölle, joilla on pitkäaikaisia vaikutuksia ihmiskuntaan. Kansallisesti tekoälyn vaikutuksia on tutkittu vähän, koska olemme vielä suunnittelun tasolla eikä käyttöönottoa ole laajamittaisesti vielä toteutettu. Usein käyttöönoton hidastaneena on osaamisen puute, varovaisuus ja tiedon puuttuminen kuinka tekoälyä voidaan hyödyntää vastuullisesti ja eettisesti. Tämän takia uusien teknologioiden käyttöönotto ja hyödynnyttäminen julkisessa hallinnossa ja kaupallisissa innovaatioissa on hidasta. Lisäksi kunnallisella tasolla hidastaviksi tekijöiksi katsotaan sääntelyä, ymmärryksen, koulutuksen ja osaamiseen puutetta tekoälystä. (Leikas, 2021.)

4.2 Etiikkapesu

Tekoälyn eettinen pesu terminä on noussut tutkijoiden keskuudesta. Tekoälyn nopea kehittyminen ja eettisten periaatteiden laatiminen on ajanut monet organisaatiot luomaan omia eettisiä periaatteita ja lupauksia. Usein organisaatioilta puuttuu toteutustapoja ja arviointikäytäntöjä, joiden avulla eettisiä ohjeita voidaan käyttää kokonaisvaltaiseen ja tiedostavaan tekoälyn kehittämiseen. Tätä kutsutaan tekoälyn keskusteluissa etiikan näkökulmasta etiikkapesuksi eli organisaation julkisuuskuvan katteettomaksi kiillottamiseksi. (Ethics of AI, 2020.)

Bietti (2021) mukaan eettinen pesu on nousussa, mikä herättää huolta tutkijoiden ja teknisen yhteisön keskuudessa. Tästä syystä sitä on alettu tarkastelemaan kriittisesti etiikan ja tekniikan suunnitellussa ja kehittämisessä. Eettisessä pesussa etiikan kysymykset yksinkertaistetaan kapeiksi teknologisoiduiksi käsitteiksi ja eettiset kysymykset ratkaistaan teknisesti toteutettavilla ratkaisuilla, jotka estävät laajempien sosioteknisten ongelmien näkyvyyden. (Ethics of AI, 2020.)

5 Tutkimuksen toteuttaminen

5.1 Integroiva kirjallisuuskatsaus

Kirjallisuuskatsaus tarkoittaa tutkimustekniikka, missä tutkitaan jo julkaistuja tutkimuksia. Sen on täytettävä tieteen tutkimukselle asetetut yleiset vaatimukset, esimerkiksi, kriittisyys, objektiivisyys, itsekorjaavuus ja julkisuus. Kirjallisuuskatsaus on tutkimuksen tekemistä tutkimuksesta, jossa koostetaan tutkimuksien tuloksia. Näin saadaan tietoa siitä, kuinka paljon on jo olemassa tutkimustietoa aiheesta, millaisista eri näkökulmista tutkimusta on tehty ja millaisin menetelmin. Nämä tulokset toimivat perustana uusien tutkimusten tuloksille. Tavoitteena kirjallisuuskatsauksessa on kehittää jo olemassa olevaa teoreettista tietoa, arvioida sitä ja luoda uutta tutkimustietoa. Kirjallisuuskatsauksen avulla voidaan luoda kokonaiskuva valitusta asiakokonaisuudesta ja siten pyrkiä havaitsemaan ja tunnistamaan ongelmia. Sen keinoin hahmotetaan aiheen kokonaisuutta. Kirjallisuuskatsaukseen kuuluu teoreettinen viitekehys, missä määritellään sen keskeiset käsitteet. Teoreettinen viitekehys perustuu systemaattiseen tiedonhakuun. (Hirsjärvi ym. 2009, 121; Salminen 2011, 1–3.)

Kuvaileva kirjallisuuskatsaus jaetaan kahteen kategoriaan: integroituun ja narratiiviseen kirjallisuuskatsaukseen. Integroitu kirjallisuuskatsaus eroaa narratiivisesta kirjallisuuskatsauksesta siten, että se on jäsennellympää ja muistuttaa joiltain osin systemaattista kirjallisuuskatsausta, Narratiivinen kirjallisuuskatsaus on rakenteeltaan kuvailevampi. Integroitu kirjallisuuskatsaus antaa tekijälleen mahdollisuuden laaja-alaisempaan aiheen tarkasteluun kuin systemaattinen kirjallisuuskatsaus. Integroiva kirjallisuuskatsaus kohdentuu tutkimustuloksiin ja sisältää johtopäätöksiä sekä loogisia seuraussuhteita. Integroiva kirjallisuuskatsaus tuottaa usein melko syvällisen ja laajan ymmärryksen tutkimuksen kohteena olevasta aiheesta. Integroiva kirjallisuuskatsausprosessi käsittää viisi eri vaihetta: tutkimuskysymyksen nimeäminen, tutkimusaineiston keruun, aineiston laadun arvioinnin, aineiston analysoinnin ja tulkinnan sekä tulosten esittämisen. (Stolt ym. 2016, 13, 107.)

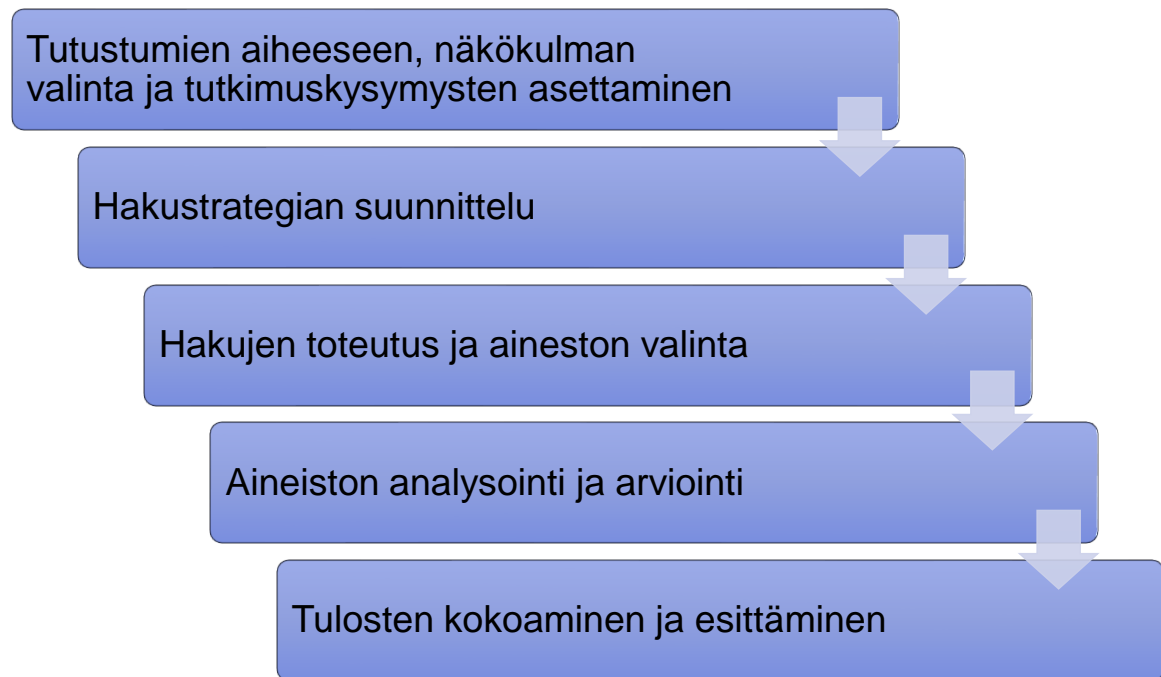
Integroidussa kirjallisuuskatsauksessa ei rajoituta käyttämään vain alkuperäistutkimuksia, vaan tavoitteena on muodostaa laaja-alaisesti kattava käsitys tutkittavasta aihepiiristä. Integroitu kirjallisuuskatsaus antaa mahdollisuuden luoda uutta käsitystä aihepiiristä ja edistää käytännön työhön soveltuvan tietoperustan muodostumista. Kuvaileva kirjallisuuskatsaus antaa tutkijalle mahdollisuuden asettaa tutkimuskysymyksen väljäksi, mikä mahdollistaa useampien näkökulmien huomioon ottamisen. Kuvaileva kirjallisuuskatsaus mahdollistaa nykyisten tietojen ja käytäntöjen arvioimisen, joka puolestaan mahdollistaa kehittämisen. (Coughlan & Cronin 2017, 15; Kangasniemi ym. 2013, 295–296; Onwuegbuzie & Frels 2016, 28–29.)

Integroitu kirjallisuuskatsaus eroaa systemaattisesta kirjallisuuskatsauksesta siten, että siinä on mahdollista kuvata tutkittavaa ongelmaa ja ilmiötä laaja-alaisemmin. Sen tarkoituksena on tehdä edeltävistä tutkimuksista yhteenveto, joka on koottu toisiinsa liittyvistä tutkimuksista. Integroidussa kirjallisuuskatsauksessa aineistoa seulotaan ja valikoidaan monipuolisesti, joka mahdollistaa isomman otoksen, jota voidaan arvioida kriittisesti. Tarkoituksena on tuottaa uutta tietoa aiemmin tehdyistä tutkimuksista. (Salminen 2011, 8.)

5.2 Tutkimusprosessi

Keskeinen rooli laadullisessa tutkimuksessa on tutkimuksen aineiston tuottamisella ja siitä tehdyllä analyysillä. Tutkimusprosessia aloittaessa ideoidaan ja suunnitellaan sitä, mitä tutkitaan, minkä takia tutkitaan ja kuinka tutkimus toteutetaan. Tutkimukseen valitun aineiston tulee olla relevantti tutkimuksen tarkoitukseen. Tämän alkuprosessin jälkeen tulee pohtia, millaista tietoa aiheesta etsitään ja mitä aiheesta halutaan tietää. Laadullinen tutkimus antaa mahdollisuuden asettaa tutkimuksen yleisluontoisemmaksi ja näin ollen tutkimukselle ei tarvitse määrittää tarkkoja tutkimuksellisia ongelmia. Tutkimuksessa on kuitenkin tärkeää löytää tutkimusongelmaan vastaukset tutkimuksellisella keinolla. On mahdollista, että alun perin valitut tutkimusongelmat muuttuvat tutkimuksen edetessä. Asetetun tutkimusongelman tulee olla yksiselitteinen ja ymmärrettävä. (Gunter & Hasanen 2021.)

Tämä opinnäytetyön prosessi alkoi tammikuussa 2021. Tutkittavan ongelman muotoilu selkeäksi kokonaisuudeksi ja tämän myötä tutkimuskysymysten asettaminen veivät alussa aikaa. Aiheeseen tutustumisen myötä tutkimuskysymykset tarkentuivat. Elokuussa 2021 informaation avulla tutustuttiin hakustrategiaan, hakusanoihin ja tehtiin koehakuja. Elokuussa 2021 toteutettiin varsinainen tiedonhaku. Hakuprosessi kuvataan tässä opinnäytetyössä vaiheittain. Aineiston valinta, arviointi ja analysointi, sekä tulosten kokoaminen ja yhteenveto tapahtuivat syyskuussa 2021. Kirjallisuuskatsauksen tutkimusprosessi toteutettiin Cooperin mallin mukaan (Cooper 1989, 14.) Tätä mallinnusta ovat myöhemmin päivittäneet Evans ja Whitemore & Knafel. (Whitemore & Knafel 2005, 549; Evans 2007, 139.) Tutkimusprosessi kulki viidessä eri vaiheessa. Tutkimusprosessi ja sen eteneminen on kuvattu vaiheittain kuviossa 4.



KUVIO 4. Tutkimusprosessin eteneminen (Mukaillen Cooper 1989, Evans 2007, Whitmore & Knafi 2005)

5.3 Hakustrategian suunnittelu ja toteutus

Hakustrategian suunnittelulla on tärkeä osa tutkimuksen onnistumisen kannalta. Huolimattomasti suunniteltu hakustrategia ja epäasianmukaisesti haettu tietopohja saa aikaan epäluotettavan tutkimustuloksen. Jokainen haku ja hakutulosten rajausta tulee dokumentoida huolellisesti tutkimuksen läpinäkyvyyden takaamiseksi. Tutkimuksessa käytetyt hakujen sanastot ja tietokannat tulee tulla esiin tutkimuksessa. Tutkimuksen sisäänotto- ja poissulkukriteerit kuvataan selkeästi. (Whittemore & Knafi 2005, 548–549.)

Tämän opinnäytetyön hakustrategian suunnittelu alkoi tutkimuskysymysten asetteluun perusteella hakusanojen valikointiin. Ennen tutkimuksen varsinaisia tietokantahakuja olin yhteydessä LAB-ammattikorkeakoulun informaattikkoon varmistaakseni tutkimuksessa käytettävien tietokantojen käyttöä ja tutustuin tutkimuksen kannalta oleellisiin tietokantoihin, jotka ovat LAB-ammattikorkeakoulun opiskelijoiden käytössä. Ennen aineiston keruuta perehdyin teoriaan kirjallisuuden ja internetin kautta. Valikoiduista hakusanoista muodostettiin kirjallisuuskatsauksessa käytettävät hakulauseet. Hakusanoja määrittäessä huomioitiin teoriaosuuden lähdemateriaaleissa ilmennyt aiheen termistö sekä koehakujen perusteella hakusanojen tuottamat tulokset. Asiasanat tarkastettiin FINTO ja MeSH- tietokannoista. Nämä esitellään taulukossa 1. Analysoitavan aineiston keruu sisältää pääasialliset haut sekä

tarkoituksenmukaisen kirjallisuuden valintaprosessin. Kirjallisuuskatsauksen luotettavuuden kannalta järjestelmällinen tiedon haku on tutkimuksen kannalta tärkeää. (Stolt ym. 2016, 25.)

Yläkäsite	Ohjaustermit
Tekoäly (YSA, YSO, MeSH) Artificial Intelligence (en) (MeSH)	keinoäly (YSA, YSO, MeSH) AI (Artificial Intelligence), Computational Intelligence, Computer Reasoning, Intelligence, Computational Intelligence, Machine, Machine Intelligence, Reasoning, Computer (MeSH)
Etiikka (YSA, YSO, MeSH) Ethics (en) (MeSH)	moraalifilosofia, siveysoppi (YSA, YSO) eettisyys , moraalifilosofia, moraalikäsite, moraalikäsitteet, siveysoppi (MeSH)
Periaatteet (YSA, YSO) Principles (en) (YSA, YSO)	arvot, asenteet, mielipiteet (YSO)
Ohjeet (YSA, YSO, MeSH) Guidelines (en) (YSA, YSO, MeSH)	ohjeistot (YSA, YSO) ohjeet aiheena, ohjeistukset aiheena (MeSH)
Eettisyys (YSA, YSO) Ethically	
Luottamus (YSA, YSO) Trust (en) (YSA, YSO, MeSH)	

Taulukko 1. Hakusanojen ja fraasien muodostaminen

Hakulausekkeiden muodostaminen aloitettiin hakusanojen ja -fraasien yhdistämisellä. Alustavien hakujen perusteella muodostui käsitys siitä, millaisilla hakulausekkeilla saavutetaan toivottuja lopputuloksia ja jotka vastaavat tutkimuskysymyksiin. Hakulauseet on kuvattu taulukko 2.

Tekoäly	("artificial intelligence" OR "AI" OR "tekoäly")	
Etiikka	("etiikka" OR "ethics")	AND
Ohjeita rajaava	("periaatteet" OR "principles" OR "ohjeet" OR "guidelines" "luotettava" OR "trustworthy")	

Taulukko 2. Hakulausekkeiden muodostus.

Tutkimuksen varsinaisten hakujen teossa päätettiin käyttää LAB-ammattikorkeakoulun LAB Primo- järjestelmän aineistohakua, EBSCO Academic Search Elite, EBSCO Applied Science & Technology Source ja Google Scholar. Artikkelien ja tutkimusten valinta perustui kirjallisuusanalyysiin, joka toteutettiin kahdessa vaiheessa. Ensimmäisellä kerralla etsittiin edellä mainituista tietokannoista osumia tai vertaisarvioituja artikkeleja hakusanoilla ”AI” OR ”Artificial intelligence” AND ethics, ”AI principles”, Artificial intelligence principles” ja ”AI guidelines”, ”Artificial intelligence guidelines”.

Aineistoa haettiin aluksi tässä integroivassa kirjallisuuskatsauksessa tieteellisistä julkaisuista ja tutkimuksista, joiden julkaisuvuosi oli välillä 2015–2021. Tutkittava aihealue kehitty nopeasti ja aineisto uudelleen rajattiin vuosiin 2018–2021. Aineiston rajaaminen oli mahdollista, sillä tutkittavaa aineistoa löytyi runsaasti. Hakulausekkeiden käyttäminen osoittautui liian monipuoliseksi ja aineiston määrä kasvoi liian suureksi. Koehauissa ei hakutermeillä löytynyt yhtään suomenkielistä artikkelia tai tutkimusta päätettiin lopullisesta hausta jättää suomenkieliset hakutermit tekoäly, etiikka, periaatteet ja ohjeet jättää pois ja rajata haku koskemaan vain englanninkielistä aineistoa. Hakulausekkeet pilkottiin osiin ja varsinaiset haut aiheesta tehtiin ”AI ethics”, ”Artificial intelligence ethics”, AI principles”, Artificial intelligence principles”, ”AI guidelines”, Artificial intelligence guidelines” ja ”AI trustworthy” ja Artificial intelligence trustworthy”. Näin toimien osumien määrä saatiin pidettyä maltillisena. Aineistohaku rajattiin vertaisarvioituihin julkaisuihin, jotka olivat saatavilla kokonaisuudessaan. Mukaan katsaukseen otettiin kansainväliset, että suomessa tehdyt tutkimusjulkaisut.

Tietokantahaun ja manuaalisesti tehdyssä tiedon haussa aineiston ensimmäisenä kriteerinä oli otsikon ja tiivistelmän perusteella tehty valinta, jonka jälkeen tutkimusaineistoa lajiteltiin sisäänotto- ja poissulkukriteerien avulla. (Taulukko 3.) Tutkimusaineistoa valitessa kiinnitettiin huomiota tekstiin, että siinä selkeästi käsiteltiin tekoälyn etiikkaa, tekoälyn periaatteita ja tekoälyn ohjeita. Tutkimusaineisto kerättiin tieteellisistä artikkeleista ja tutkimuksista, jotka olivat vertaisarvioitu.

Sisäänottokriteerit	Poissulkukriteerit
<ol style="list-style-type: none"> 1. Kyseessä tieteellinen vertaisarvioitu artikkeli tai tutkimus 2. Tutkimuksessa on kuvattu tekoälyn eettisiä ohjeita tai periaatteita. 	<ol style="list-style-type: none"> 1. Kyseessä ei ole vertaisarvioitu artikkeli tai tutkimus. 2. Tutkimuksessa ei käsitellä tekoälyn eettisiä ohjeita tai periaatteita.

3. Tutkimuksen julkaisukieli on suomi tai englanti.	3. Tutkimuksen julkaisukieli on muu kuin suomi tai englanti.
4. Tutkimuksen julkaisuvuosi 2018–2021.	4. Tutkimus on julkaistu ennen vuotta 2018.
5. Tutkimus on saatavilla LAB-ammattikorkeakoulun kirjaston tietokantojen kautta.	5. Tutkimusta ei ole saatavana kokotekstinä LAB-ammattikorkeakoulun kirjaston tietojen kautta.

Taulukko 3. Sisäänotto- ja poissulkukriteerit

Ratkaisevana tekijänä valinnassa ei ole ollut yksittäinen artikkelin tai tutkimuksen yksityiskohtien syvyys, vaan havaittavissa oleva aikomus luokitella ja kartoittaa väitteitä tekoälyn etiikan alalla. Hakusanojen tuli myös löytyä otsikosta tai abstraktista.

Tutkimukseen suoritettu haku antoi tuloksia 724, jotka käytiin läpi otsikko tasolla. Ensimmäisen haun luvut sisältävät myös mahdolliset kaksoiskappaleet, jotka karsittiin pois otsikon luku vaiheessa. Otsikoiden rajaaminen tuotti vaikeutta, sillä otsikoiden tasoista ei voinut päätellä aineistoa, vaan karsiminen tehtiin vasta tiivistelmien lukemisen jälkeen. Otsikoiden perusteella aineistoa karsittiin ja tiivistelmien lukemisen jälkeen tutkittavaa aineistoa kertyi 26 kappaletta. Tutkimukseen valikoitunut materiaali koostui englanninkielisistä aineistoista, jotka luettiin ja käännettiin suomenkieliseksi aineistoksi.

Kaikkiaan hakujen perusteella käytiin 257 tutkimusta ja vertaisarvioitua artikkelia läpi otsikotasolla, kolmesta eri tietolähteestä haettuna. Otsikotason rajauksen ja tiivistelmän luvun jälkeen jäljellä oli 105 tutkimusta. Tiivistelmien perusteella materiaali karsittiin 31 tutkimukseen, jotka luettiin ja analysoitiin tarkemmin. Näistä materiaaleista mukaan kirjallisuuskatsaukseen valittiin 15 tutkimusta ja vertaisarvioitua artikkelia. ESBCO Adademic search Elite x, ESBCO Applied Sciences & Technology source x ja x Google Scholarista. Jokainen haku suoritettiin ensin ESBCO tietokantoihin ja sen jälkeen Google Scholariin, jonka jälkeen löytyneet duplikaatit poistettiin. Tämän takia ESCBO Applied Sciences & Technology source tietokannasta valittiin katsaukseen eniten tutkimuksia ja vertaisarvioituja. Kirjallisuuskatsauksen hakuprosessi kuvattu taulukossa 4.

Tietokanta	Hakusanat	Kaikki osumat	Rajaus otsikon ja tiivistelmän peruusteella ja dublikaattien poisto	Lopullinen valinta
EBSCO Academic Search Elite	AI Ethics	24		
	Artificial intelligence Ethics	12	7	4
	AI principles	8		
	Artificial intelligence principles	6	2	2
	AI guidelines	8		
	Artificial intelligence guidelines	5	0	0
	AI Trustworthy	14		
	Artificial intelligence trustworthy	17	2	1
EBSCO Applied Sciences & Technology Source	AI ethics	20		
	Artificial intelligence ethics	8	2	1
	AI principles	9		
	Artificial intelligence principles	5	0	0
	AI guidelines	3		
	Artificial intelligence guidelines	1	0	0
	AI Trustworthy	12		
	Artificial intelligence trustworthy	13	2	1
Google Scholar	AI ethics	29		
	Artificial intelligence ethics	24	5	3
	AI principles	8		
	Artificial intelligence principles	10	1	0
	AI guidelines	9		
	Artificial intelligence principles	11	2	2
	AI Trustworthy	12		
	Artificial intelligence trustworthy	13	3	1
		281	26	15

Taulukko 4. Hakuprosessin eteneminen.

Kun mahdollista materiaalia aineistoksi löytyy, on tutkija arvioitava materiaalin hyödyllisyys. Hakuvaiheessa on tärkeää arvioida, vastaako artikkelissa kuvattu tutkimus asetettuihin tutkimuskysymyksiin. Mikäli tutkimuksia löytyy, mutta ne eivät vastaa tutkimukselle asetettuihin sisäänotto- ja poissulkukriteerit ovat toimimattomia. Tällaisessa tapauksessa kirjallisuuskatsauksen tulos on vaarassa jäädä merkityksettömäksi. (Flinkman & Salanterä 2007, 92.)

Kirjallisuuskatsaukseen otettaville tutkimuksille ja artikkeleille tulee määritellä tarkat sisäänottamisen ja poissulkemisen kriteeristö. Kriteerien tulee perustua aiemmin luotuihin ja laadittuihin tutkimuskysymyksiin. Sisäänottamisen kriteereissä pystytään rajaamaan alkuperäistutkimusten kohderyhmää, tutkimusmenetelmiä, laatuun vaikuttavia tekijöitä ja tuloksia. Kriteerien toimivuutta on hyvä tarkastella koehakujen avulla ennen varsinaisia hakuja. Kirjallisuuskatsaukseen sisään otettavien ja poissuljettavien aineistojen valinta on vaihteleva tapahtuva prosessi, joka perustuu siihen, vastaako koe hauissa löydetty tutkimukset asetettuja sisäänottamisen kriteerejä. Vastaavuuksia tulee tarkastella otsikon ja abstraktin eli tiivistelmän sekä kokonaisen tekstin tasolla. (Whittemore & Knafel 2005, 548; Stolt & Routasalo 2007, 59.)

Aineiston keräämisessä käytettiin sisäänotto- ja poissulkukriteereitä. Kriteerit auttavat luotettavan aineiston tunnistamisesta ja lisäävät katsauksen luotettavuutta sekä varmistavat, että katsaus pysyy rajatussa aiheessa. Lisäksi sisäänottokriteereinä oli, että tutkimusta käsittelevä teksti on luettavissa kokonaisuudessaan ilmaiseksi ja se on julkaistu suomen- tai englannin kielellä sekä teksti tulee olla vertaisarvioitu. Analysoitavan aineiston keruu sisältää pääasialliset haut sekä tarkoituksenmukaisen kirjallisuuden valintaprosessin. Kirjallisuuskatsauksen luottavuuden kannalta järjestelmällinen tiedon haku on tutkimuksen kannalta tärkeää. (Stolt ym. 2016, 25–26.)

5.4 Aineiston laadun arviointi ja esittely

Tutkimukseen valitun aineiston valikoinnissa tulee huomioida alkuperäisen tutkimuksen vastaavuus tutkimuskysymyksiin. Valitun aineiston tulee olla riittävä tutkimuskysymyksen vastaamiseen ja laajuuteen. Kuvaileva kirjallisuuskatsaus eroaa systemaattisesta kirjallisuuskatsauksesta myös siten, että aineisto ei rajoitu vain hakukriteereihin vaan niistä voidaan poiketa. Tärkeää on, että aineisto vastaa asetettuihin tutkimuskysymyksiin. (Kangasniemi, ym. 2013, 296.)

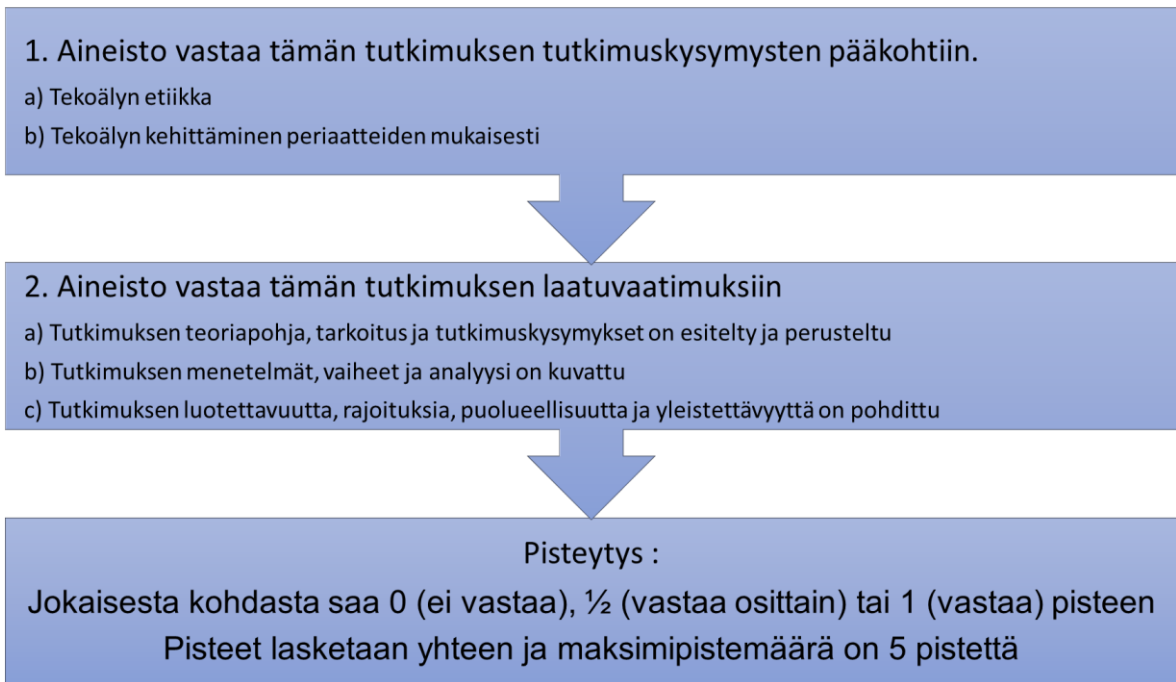
Integroivaan kirjallisuuskatsaukseen sisältyy aineistoon valikoitujen alkuperäistutkimusten arviointi. Sillä kyetään kasvattamaan kirjallisuuskatsauksen luotettavuutta ja vaikuttavuutta. (Kontio & Johansson 2007, 101.) Näin pystytään myös arvioida, paljonko painoarvoa

kirjallisuuskatsauksen päätelmille voidaan antaa. Integroivan kirjallisuuskatsauksen materiaaliin voidaan sisällyttää eri menetelmillä tehtyjä tutkimuksia, jonka takia tutkimuksen laadun arvioiminen on haastava. Laadun arvioinnin varmistamiseen on kehitetty erilaisia menetelmiä. Kun kyseessä on empiirinen tutkimus, voidaan laadun arviointi kirjoittaa avoimesti auki ja kuvata heikkoudet ja vahvuudet sekä arvioida luotettavuutta, toistettavuutta ja tutkimusaineiston otantaa edustavuutta sekä kokoa ja laatua. (Flinkman & Salanterä 2007, 92–93.)

Aineiston laadun arvioinnin toteuttamiseen on olemassa monia valmiita saatavia tarkistuslistoja ja arviointiasteikkoja. Tarvittaessa tutkimuksessa voidaan käyttää myös itse kehitettyä arviointimenetelmää ja laatua arvioitaessa voidaan käyttää erilaisia arviointimenetelmiä yhdessä. Tärkeintä on, että asteikko, jota käytetään, sopii kerätylle aineistolle ja sitä käytetään luotettavasti ja järjestelmällisesti. (Kontio & Johansson 2007, 102–103.) Integroivassa kirjallisuuskatsauksessa aineiston arviointia suoritettiin myös hakutoimintoja tehdessä seuraavalla tavalla:

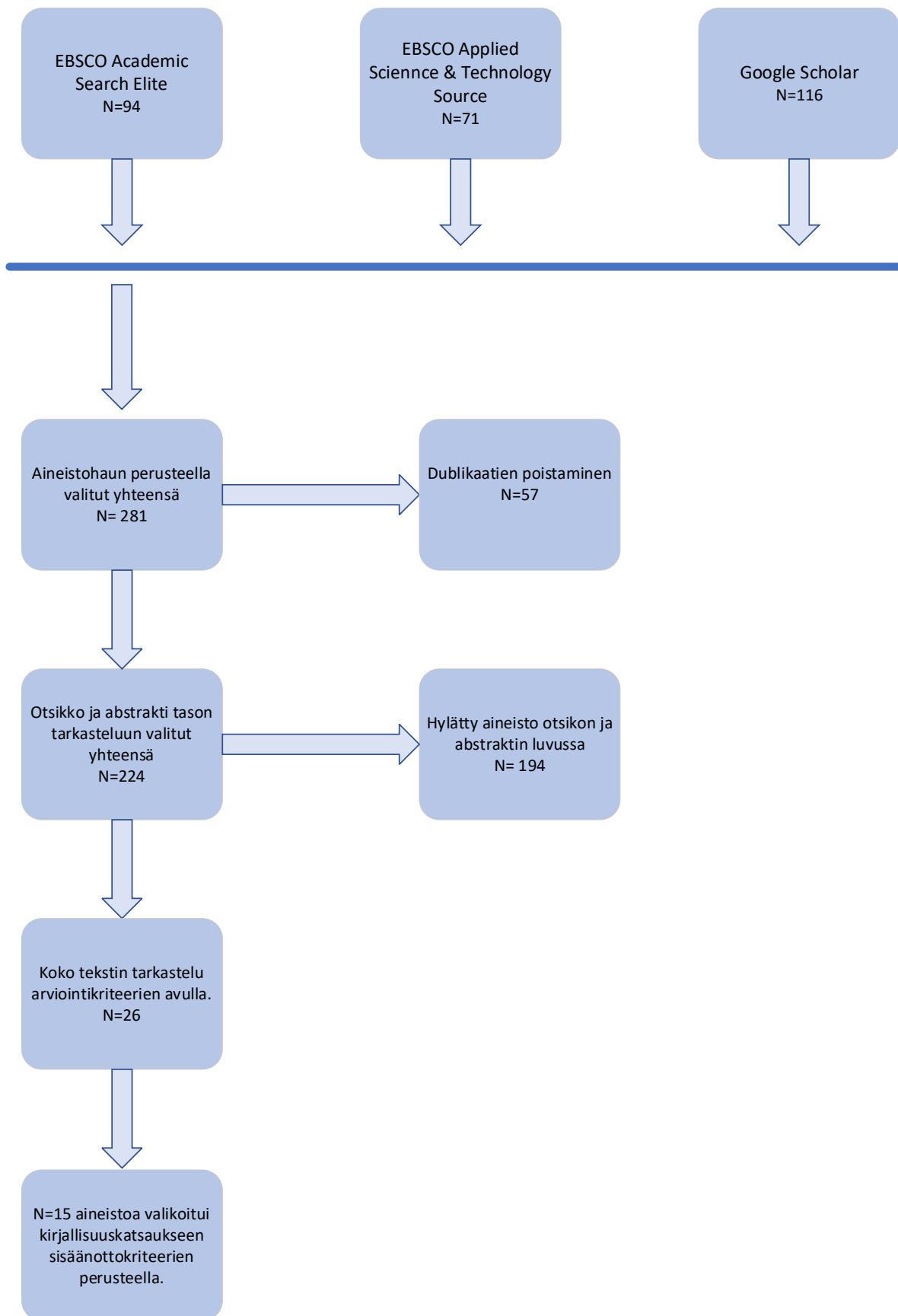
- Sisäänottokriteerien täytyminen
- Kaksoiskappaleiden eli duplikaattien poistaminen
- Tutkimuksen yleistettävyys (liian suppea)
- Tutkimuksen vastaavuus aiheeseen ja tutkimuskysymyksiin otsikon ja tiivistelmän perusteella
-

Tähän opinnäytetyöhön ei hauissa löytynyt sopivaa ja valmista arviointimenetelmää, joten arvioitavan aineiston tarkistuslista laadittiin itse aineiston pohjalta. Aineiston arvioinnin kriteerit ja pisteytys esitetty kuviossa 5.



KUVIO 5. Kirjallisuuskatsaukseen valitun aineiston arvioinnissa käytetty pisteytys (Soveltaen Whittemore & Knafel 2005, 549–550; Cooper 1989, 90–94)

Kirjallisuuskatsaukseen valittujen tutkimusten ja vertaisarvioitujen artikkelien arviointitaulukko esitellään liitteessä 1. Kuviossa 6 esitellään hakuprosessin eteneminen.

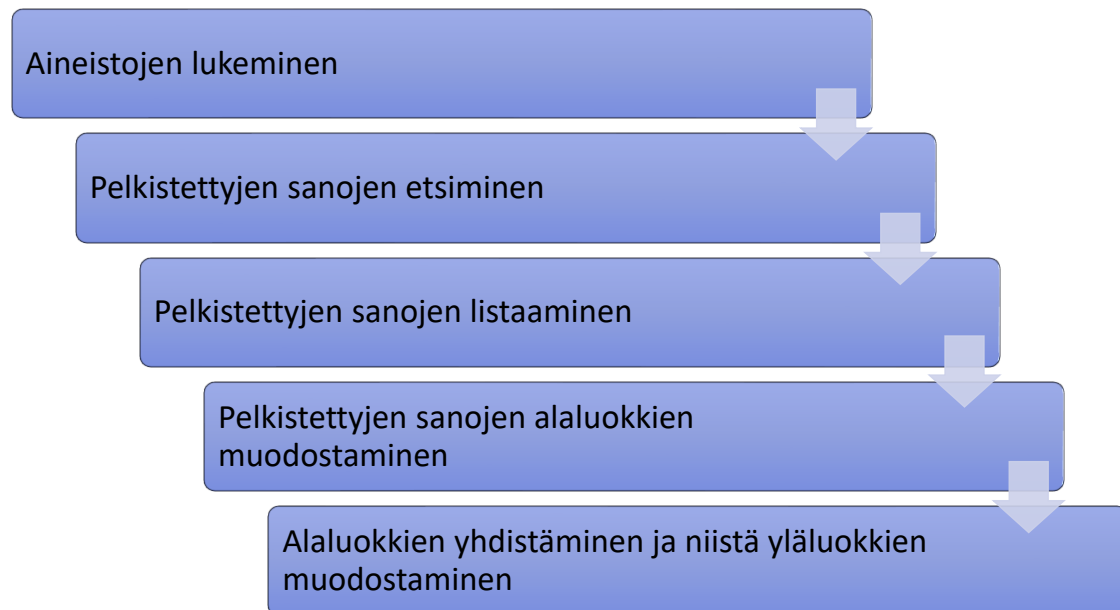


KUVIO 6. Aineiston hakuprosessi.

5.5 Aineiston analyysi

Sisältöanalyysi on menetelmä, jolla voidaan kuvata tarkasti tutkittavaa ilmiötä. Sisällönanalyysin avulla muutetaan järjestelmällisesti suuri aineiston määrä ja teksti tuottamalla selkeä, järjestelmällinen ja kompakti kuvaus tutkittavasta aineistosta. Laadullisen sisällönanalyysin avulla yritetään tuottaa yhtenäistä mielenkiintoista tietoa, jotta tutkittavasta ilmiöstä saadaan tuotettua luotettavia johtopäätöksiä, jotka ovat johdonmukaisia. (Erlingsson & Brysiewicz 2017, 94; Kyngäs ym. 2011, 139; Tuomi & Sarajärvi 2009, 95–117.) Sisältöanalyysin avulla on mahdollista tutkia materiaalia ja analysoida aineistoa objektiivisesti sekä järjestelmällisesti. Sisällönanalyysin avulla on mahdollista, että aineistoa voidaan luokitella, luoda käsitteitä ja käsittekarttoja, joiden avulla kuvataan tutkittavana olevaa ilmiötä. Tekota-
poja sisällönanalyysille ovat teoriaohjaava, teorialähtöinen ja aineistolähtöinen analyysi. Aineistolähtöisesti tehty sisällönanalyysi vaatii tekijältä aineiston hyvää hallintaa. (Kyngäs ym. 2011, 139; Kangasniemi ym. 2013.) Tämän opinnäytetyön tutkimusaineiston analysoinnissa käytettiin aineistolähtöistä sisällönanalyysiä. Analyysin ensimmäinen vaihe oli luoda taulukkomuotoinen yhteenveto valittujen tutkimusten sisällöstä, kuten tutkimuksen tarkoituksesta sen menetelmistä ja tuloksista. Yhteenvetoaineisto on esitelty liitteessä 1.

Aineistolähtöisen sisällönanalyysin tavoitteena on luoda tutkimusaineistosta teoreettinen kokonaisuus. Analysointia ohjaavat tutkimuksen tarkoitus, tehtävänkuvaus ja tutkimusaineisto. Edeltävä teoretieto aiheesta ei saa vaikuttaa tutkimuksen lopputulokseen. (Tuomi & Sarajärvi 2018, 108.) Tässä opinnäytetyössä sisällönanalyysi toteutettiin Tuomen ja Sarajärven (2018, 122) ohjeistuksen mukaan, aineiston analyysi jaettiin kolmeen omaan vaiheeseen; aineiston pelkistämiseen, aineisto ryhmittelyyn ja viimeiseksi teoreettisten käsitteiden luomiseen. Kuviossa 7 on kuvattu aineiston analysoinnin prosessi.



KUVIO 7. Sisällönanalyysi ja eteneminen (mukaillen Tuomi & Sarajärvi 2018, 123.)

Pelkistämisvaihe eli ensimmäinen vaihe, jossa aineistosta rajattiin pois opinnäytetyölle merkityksettömät tiedot ulos ja etsittiin asetettuihin tutkimuskysymyksiin kytkeytyviä ilmaisuja. Ryhmittelyn vaiheessa löydetty pelkistetyt ilmaukset tarkastettiin huolellisesti ja etsittiin niistä yhteen liittyviä ja/tai erilaisia käsitteitä. Toisiaan vastaavista käsitteistä muodostettiin ryhmiä ja niistä uudelleen luokkia yhdistelyllä. Alaluokkien luokittelu suoritettiin ensin, joista yhdistelyn avulla muotoutui pääluokat. Käsitteiden muodostuksen tarkoitus on käsitteellistää alkuperäinen tutkimusaineisto ja siksi tässäkin integroidussa kirjallisuuskatsauksessa jatkettiin eri luokkien yhdistämistä, kunnes ilmauksista muodostui teoreettisia käsitteitä ja johtopäätöksiä. (Tuomi & Sarajärvi 2018, 122–127.)

Aineiston analyysi toteutettiin aineistolähtöisenä eli induktiivisena aineistoanalyysinä. Aineisto käytiin huolellisesti läpi ja etsittiin tutkimuskysymyksiin liittyviä ilmaisuja ja sanoja. Ensimmäisenä ilmaisut ja sanat alleviivattiin eriväreillä. Tutkimukset koostuivat englanninkielisistä alkuperäistutkimuksista ja julkaisuista, joten poimitut ilmaisut käännettiin suomenkieliseksi ja kirjoitettiin erivärisille post-it lapuille. Vaikka analyysiin valikoitunut aineisto sisälsi tutkimustyyppiltään erilaisia alkuperäisperäistutkimuksia ja artikkeleita, tehtiin analyysi aineistolle samalla tavalla. Seuraavaksi suoritettiin pelkistäminen eli ilmaisujen tiivistäminen. Tämä toteutettiin löytyneiden ilmaisujen perusteella useamman kerran. Ilmaisujen tiivistämistä tehtiin myös post-it – lapuille kirjoittamisen yhteydessä. Lappuja yhdistettiin toisiinsa löytyneiden ilmaisujen perusteella ja luomalla yhteinen ilmaisu tai sana.

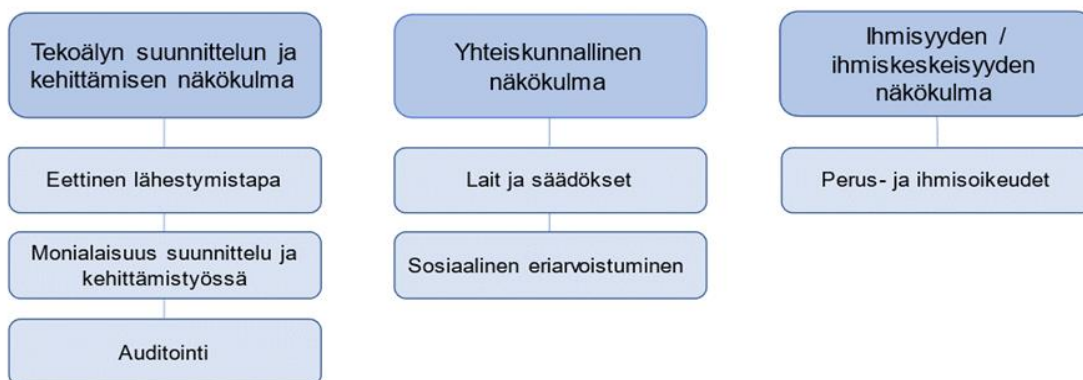
Ilmaisujen pelkistämässä aineistosta tehtiin huomio, että useat ilmaiset vastasivat suurelta osalta kumpaakin esitettyyn tutkimuskysymykseen. Aineisto jaettiin yhteen ryhmään, jonka jälkeen ilmaisia tarkasteltiin tutkimuskysymysten perusteella, jotka olivat;

1. Millaisia tekoälyn eettisiä periaatteita on laadittu kehittämistyön tueksi?
2. Miten tekoälyn eettisiä periaatteita tulisi soveltaa LAB WellTech kehittämissympäristössä, jotta se olisi kehittämistyön kannalta kestävä?

Kahdeksassa tutkimuksessa nousi ilmaisu ”eettiset periaatteet” liittyivät ensimmäiseen tutkimuskysymykseen ja ”vastuullinen kehittäminen” liittyi toiseen kysymykseen. Post-it-lappuille kirjoitetut ilmaiset olivat helposti siirrettävästi ja yhdisteltävissä. Tässä vaiheessa poistettiin vielä ilmaisia, jotka eivät vastanneet tutkimuskysymyksiin. Ryhmittely vaiheen jälkeen aloitettiin pääkäsitteiden muodostaminen. Tässä vaiheessa havaittiin, pääkäsitteitä tutkimuskysymyksille muodostuu kolme. Kolmen pääryhmän sisälle havaittiin muodostuvan tarvetta sisäiselle ryhmittelylle ja näin muodostui pääryhmien sisälle pienempiä ryhmiä. Nämä ryhmät muodostuivat yhdisteltäessä toisiinsa liittyviä lappuja, joissa ilmeni samankaltaisia ilmaisia. Yläluokkien nimitykset laadittiin tutkimuskysymysten perusteella ja alaluokat kuvaillen ilmaisia. Näin toimien aineistosta pyrittiin erottelemaan samankaltaisuu- det, teemat ja ilmaisujen väliset suhteet. (Flinkman & Sahanterä 2007, 96.)

Analysoinnin perusteella aineistosta nousi kolme yläluokkaa tekoälyn suunnittelun ja kehittämisen tueksi. Yläluokiksi muodostui; tekoälyn suunnittelu ja kehittäminen, yhteiskunnallinen vastuu ja ihmiskeskeisyys. Nämä yläluokat linkittyvät osin vahvasti toisiinsa. Alaluokat on jaoteltu siten, mihin olen niiden parhaiten katsonut sopivan. Alaluokkia muodostui yhteensä kuusi. (kuvio 4).

Tekoälyn eettisyys kehittämistyössä



KUVIO 8. Sisällön analyysin tulokset.

Ensimmäinen yläluokka muodostui tekoälyn suunnittelun ja kehittämisen näkökulmaa ilmaisevista aineistoista. Tämän yläluokan alle muodostui kolme alaluokkaa, jotka muodostuivat sen mukaan missä ilmaisun vaikutus koettiin olevan. Alaluokiksi muodostui eettinen lähestymistapa, monialaisuus suunnittelu ja kehittämistyössä ja auditointi. Toinen yläluokka muodostui yhteiskunnallisen näkökulman huomioon ottamisesta tekoälyn eettisyyttä tarkasteltaessa. Alaluokkia toiseen yläluokkaan muodostui kaksi, lait ja säädökset, sosiaalinen eriarvoistuminen. Kolmanteen yläluokkaan muodostui ihmisyyden ja ihmiskeskeisyyttä tarkasteleva näkökulma, jonka alaluokan muodostaa perus- ja ihmisoikeudet. Tämän integroidun kirjallisuuskatsauksen tulokset esitellään tarkemmin kappaleessa 6.

6 Tulokset

6.1 Tekoälyn suunnittelun ja kehittämisen näkökulma

Eettinen lähestymistapa

Tekoälyn eettiseen lähestymistapaan vaikuttavat useat eri tekijät. Eettisiä periaatteita on luotu tutkijayhteisöissä, organisaatioissa ja hallinnollisilla tasoilla. Yksi eurooppalaisista eettisen tekoälyn periaatteista on korkean tason asiantuntijaryhmä (AI HLEG), joka julkaisi 2019 asiakirjan luotettavan tekoälyn saavuttamiseksi. Asiakirja mainitsee seitsemän keskeistä vaatimusta tekoälylle: inhimillinen toiminta ja valvonta, tekninen kestävyys ja turvallisuus, yksityisyys ja tietojen hallinta, avoimuus, monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus sekä ympäristön ja yhteiskunnan hyvinvointi ja vastuullisuus. Kuitenkin yksimielisyys siitä, mitä periaatteet tarkoittavat, keitä ne koskevat ja miten niitä tulisi toteuttaa aiheuttaa haasteita. Tämän katsotaan johtuvan siitä, että periaatteet ovat epämääräisiä ja vaikeasti tulkittavia. Myös tekoälyn globaalisuus ja erimaiden kulttuurisidonnaisuus aiheuttaa periaatteiden tulkinnalle haasteita. (Ayling & Chapman 2021; Floridi ym 2020; Mora-Cantallops ym 2021; Iason 2020; Hagendorff 2020; Kazim & Koshiyama 2021.)

Yleisesti eettisissä ohjeissa esitetään hyvin laajoja ja kattavia periaatteita, joiden oletetaan toteutuvan laajasti monipuolisissa tieteellisissä, teknisissä ja taloudellisissa käytännöissä ja maantieteellisesti hajautetuissa tutkijoiden ja kehittäjien ryhmissä, joilla on erilaiset prioriteetit ja tehtävät. Etiikan katsotaan toimivan etäällä käytännöistä, joihin sillä tahdotaan vaikuttaa kuten insinöörit, jotka suunnittelevat tekoälyjärjestelmiä. Kuitenkin juuri eettinen lähestymistapa antaisi organisaatioille mahdollisuuden hyödyntää sosiaalista arvoa, jonka avulla voitaisiin ennaltaehkäistä sosiaalisesti hyväksymättömiä toimintatapoja. Tekoälyn eettinen lähestymistapa tulisi nähdä suunnittelussa ja kehittämisessä positiivisena ennakkovaroitussuunnitelmana riskeiltä, jotka saattaisivat vaarantaa organisaatioita ja hallintoa. (Floridi ym 2020; Hagendorff 2020; Khan ym. 2021.)

Tutkimuksessa selvisi, että tekoälyn eettisyyden periaatteet vaihtelevat ja samaa periaatetta kuvataan usealla eri sanalla, mutta niistä voidaan nostaa yhteneväisyyksiä. Useat asiakirjat sisältävät selittävyyden, läpinäkyvyyden, jäljitettävyyden, vastuullisuuden ja oikeudenmukaisuuden. Selittävyyden tarkkasanaamuoto vaihtelee periaatteissa. Joissakin periaatteissa se ilmaistaan ”avoimuus” ja ”ymmärrettävyys”. Selitettävyyttä pidetään moraalisenä periaatteena, jonka pitäisi auttaa ymmärtämään algoritmien hyväksyttäviä käyttötapoja. Selitettävyyden ja läpinäkyvyyden periaatteita pidetään yhtenä ratkaisuna, jolla voidaan tarjota ratkaisua ”mustan laatikon” ongelmalle. Läpinäkyvyyden periaatteen katsotaan olevan hyvin lähellä selitettävyyttä, vaikka ne eivät olekaan synonyymeja keskenään.

Jäljitettävyyden periaatetta pidetään useassa julkaisussa yhtenä tekoälyn keskeisimmistä periaatteista ja vaatimuksista. Tämän periaatteen katsotaan liittyvän vahvasti tarpeeseen vaatia täydellistä selvitystä tekoälymallin tuottamiin tietoihin ja sen prosesseihin. Jäljitettävyys on siis avaintekijänä tekoälyn läpinäkyvyyden kannalta. (Mora-Cantalopps ym. 2021; Khan ym. 2021; Kazim & Koshiyama 2021; Robert ym. 2020; Robins 2019.)

Vastuullisuuden periaatteen katsotaan olevan yksi tärkeimmistä tekoäly tutkimuksen peruspilareista ja siksi onkin välttämätön ottaa huomioon siinä, millaista tekoälyä kehitetään, miten ja mihin sitä käytetään. Tähän periaatteeseen kuuluu, että tiedetään, kuka on tehnyt päätökset, miten päätökset tehtiin, mitä järjestelmiä ja työkaluja käytettiin ja kuinka sitä mitataan ja seurataan. Tällä tavalla toimien voidaan varmistaa, että vastuu säilyy suunnittelijoilla. Vastuullisuuden katsotaan myös olevan keskeinen asia oikeudellisen vastuullisuuden kannalta (Floridi ym; 2018, Khan ym. 2021; Kazim & Koshiyama 2021; Virginia 2018.)

Oikeudenmukaisuuden periaate haastaa tutkijoita ja kehittäjiä kansainvälisesti, sillä sen katsotaan liittyvän pitkälti paikalliseen lainsäädäntöön ja täten sen yhtenäistäminen on ongelmallista. Tästä syystä tutkijat pitävätkin tärkeänä, että pyritään maailmanlaajuisiin tekoälyperiaatteisiin, jotta voitaisiin rakentaa rajoituksia tekoälylle ja sen käytölle. Oikeudenmukaisuuden periaatetta tulee tarkastella ihmisten tasa-arvon ja ihmisoikeuksien eettisenä periaatteena. (Iason 2020; Khan ym. 2021; Kazim & Koshiyama 2021.)

Tekoälyn etiikasta puhuminen ei siis yksissään riitä. Eettisyyden periaatteet ovat vasta ensimmäinen askel, jossa vasta yleinen eettinen käytäntö on ratkaisevassa asemassa. Tutkijat näkevät aihetta syvään huoleen tekoälyjärjestelmien lisääntyvän eriarvoisuuden mallista, joita on ollut havaittavissa tähän mennessä. Tekoäly uskotaan olevan kuitenkin myös voima, joka vähentää jakautumista ja luo tasa-arvoisemman maailman. "Tekoäly voi auttaa kuromaan umpeen digitaalista kuilua ja luomaan osallistavan yhteiskunnan". (Hagerty & Rubinov 2019.)

Monialaisuus suunnittelu ja kehittämistyössä

Tutkimusten ja artikkelien mukaan tekoälyn suunnitteluun ja kehittämiseen liittyy monia erilaisia tekijöitä ja eettisesti ratkaisemattomia ongelmia. Tekoälyn eettisyyden tarkasteluun on viime vuosina laadittu useita erilaisia työkaluja suunnittelijoille, mutta niitä on moitittu joko liian joustaviksi tai joustamattomiksi, jonka takia on ollut mahdollista sortua tekoälyn eettiseen pesuun. Lisäksi teorian ja käytännön välinen kuilu on vielä merkittävä. Merkittävässä asemassa on myös maantieteellinen sijainti ja kulttuurisidonnaisuus, jotka poikkeavat toisistaan maanosien välillä ja maanosien sisällä. Tutkimuksissa pidettiin tärkeänä yhteisten eettisten periaatteiden ja työkalujen luontia sekä luotettavaa ja avointa auditointia. Eettiset näkökohdat tuleekin ottaa huomioon tekoälysovellusten suunnittelun alkuvaiheessa.

(Eötvös 2021; Hauer 2020; Morley ym. 2021; Mökander & Floridi 2021; Floridi ym. 2020; Shneiderman 2020; Rossi 2020.)

Tekoälyn suunnittelussa ja kehittämisessä tulee kiinnittää erityisesti huomioita siihen, että ne eivät aiheuttaisi ihmiselle haittaa. Suunnittelu- ja kehittämistyössä tulee huomioida, ettei tekoäly lisää ihmisten eriarvoisuutta tai vähennä sosiaalista vaikutusvaltaa. Lähtökohtana tulee aina olla kunnioitus ihmisen autonomiaa kohtaan. Tiedostettu ongelma on myös, että usein ohjelmoijalta tai tekoälyn kehittäjältä puuttuu tarvittava eettinen asiantuntemus, joka voi johtaa väärin eettisiin oletuksiin. Tämän takia monialainen suunnittelutyö tulisi olla lähtökohtaisesti yksi vaade suunnittelu- ja kehittämistyölle. Mikäli tekoälyä kehitetään suunnitellusti ja harkiten, se tarjoaa mahdollisuuden parantaa ihmisen toimijuuden mahdollisuuksia. Tekoälyjärjestelmät voisivat tehokkaasti suunniteltuina vahvistaa yhteisiä moraalijärjestelmiä. Suunnittelu, kehittäminen ja niiden hallinta vaatii vankkaa ihmisoikeuskehystä, siitä kuinka kehitystyö on vastuullisesti kestävä. (Floridi ym.; 2020; Fukura-Parr & Gibbons 2021; Hauer 2020.)

Tekoälyn turvallinen suunnittelutyö vaatii erilaisia hyväksytyjä suunnittelumenetelmiä, jotka kykenevät tukemaan eettisten vaikutusten analysointia ja arviointia. Lisäksi tekoälyn suunnittelun ja kehittämisen tueksi tulee luoda yhtenäisiä käytäntöjä, standardeja ja sertifiointiprosesseja. Näiden käytänteiden avulla voitaisiin varmistaa kehittäjien ja käyttäjien eheys. Yhtenä tärkeimmistä tutkimussuunnista on kehittää objektiivisia valvonnan ja autonomian tasojen mittareita, jotka on sidottu tekoälyn erilaisiin tehtäviin. Toisena korostetaan tarvetta kehittää sellaisia järjestelmiä, joilla kiinnitetään huomioita tekoälyn harhaan, toisin sanoen vinoumaan. Vinoumien havaitseminen on suunnitteluvaiheessa tärkeää, sillä niiden lieventämisellä on suuri vaikutus tekoälyn luottamuksen saavuttamisessa. Käytännön haasteeksi nousee kehittää tarpeeksi tehokkaita ja hyviä malleja, joiden avulla voidaan luoda luotettavia käytäntöjä turvallisuuteen ja sosiaalisiin rakenteisiin. Onnistunut suunnittelu johtaa vääjäämättä turvallisiin ja luotettaviin tekoälyjärjestelmiin. Lähdekoodin tarkastaminen ja sisäiset luotettavuuden sekä kestävyys tarkastamiset ovat jo yleisesti käytännössä olevia mekanismeja erilaisten ohjelmistojen suunnitteluissa. Tällaiset organisaatioiden sisäiset tarkastukset pienentävät mahdollisia ulkoisiin tarkastuksiin liittyviä riskejä ja pakottavat organisaatiot pohtimaan omia oletuksia ja valintoja. (Virginia 2018; Robert ym. 2020; Rossi 2020; Shneiderman 2020.) Shneiderman 2020 korostaa tutkimuksessaan, että luotettavan ja turvallisen tekoälyn suunnittelun ja kehittämistyön tueksi tarvitaan luotettavuutta tukevia teknisiä käytänteitä ja johtamisstrategioita, jotka luovat yhtenäistä turvallisuuskulttuuria ja riippumattomia valvontarakenteita, jotka osaltaan tukevat tekoälyn kohdistuvaa luottamusta.

Vaikutusten ja riskien arviointi tulee suunnitella hyvin ja tarkasti sekä kiinnittää huomiota käytettäviin mittareihin, joiden avulla päästään haluttuun lopputulokseen. Mittareita on jo kehitetty useita; teknologian mittareista ihmiskeskeisiin mittareihin. Mittareiden valintaan vaikuttaa suuresti tekoälyjärjestelmän päämäärä ja lait sekä säädökset. Lisäksi vaikutusten ja riskien arvioinnin tulee olla säännöllistä ja koskea vähintään kolmea vaihetta suunnittelun ja kehittämisen aikana. Ensin validoinnin avulla tarkastellaan, onko järjestelmä oikeuskelpoinen, varmennuksen avulla haetaan, kehitetäänkö järjestelmää oikealla tavalla ja viimeiseksi arvioidaan mittareiden avulla järjestelmää kokonaisvaltaisesti. (Ayling & Chapman 2021; Eötvös 2021; Mora-Cantalops ym. 2021; Morley ym. 2021; Rossi 2020.)

Käyttäjien tietoisuutta tarvitaan, jotta voidaan luoda sidosryhmiä ja auttaa hallituksia ja organisaatioita ymmärtämään eettisen tekoälyn tärkeyden. Tämä voisi puolestaan luoda myös sääntelypainetta ohjeellisena ohjenuorana toimia eettisten tekoälyjärjestelmien kehittämiseksi ja toteuttamiseksi. Tähän tarvitaan maailmanlaajuisia koordinoituja toimia tekoälyn oikeudenmukaisuuteen, luottamukseen ja etiikkaan liittyvien ongelmien ratkaisemiseksi. (Rossi 2020; Robert ym. 2020.)

Auditointi

Tekoälyn eettisestä auditoinnista uskotaan tulevan lähitulevaisuudessa tärkeä väline operatiivisen tekoälyn etiikassa ja sitä on ehdotettu keinoksi tutkia tekoälyä ja algoritmien aiheuttamia vinoumia. Vaikka vielä ei standardeja olekaan luotu, on erilaisia lähestymistapoja tekoälyn eettisyyteen olemassa. Etiikkaan perustuva tekoälyn auditoinnin katsotaan edistävän hyviä käytänteitä ja hallintoa. Auditointi voidaan suunnata esimerkiksi päätöksen takana olevaan lähdekoodiin ja vaikutusten tarkistamiseen, itse teknologiaan, selitettävyyteen ja kestävyYTEEN. Etiikkaan perustuvan auditoinnin tulee olla jatkuva prosessi ja sen avulla on seurattava ja arvioitava järjestelmää kokonaisvaltaisesti. (Kamiz & Koshiyama 2021; Morley ym. 2021; Mökander & Floridi 2021.)

Tekoälyn etiikkaan kohdistuvan auditoinnin on täytettävä sille ennalta määrätyt erityisvaatimukset, mutta itse auditoinnin ei tarvitse olla vaikea toteuttaa. Kuitenkin tärkeänä pidetään, että se tulee suorittaa järjestelmällisesti ja että siihen sisältyy tarkka dokumentointi menetelmistä ja logiikka auditoinnin suorittamisesta. Tekoälyn etiikkaan perustuva auditointi ei vielä ole laajasti käytössä, eikä sille ole määritelty selkeää viitekehystä. Syyksi tähän perustellaan, että se lisää kustannuksia, jotka rasittavat yhteiskuntia ja siltä puuttuu vielä sitä tukevat säädökset ja lait. Lisäksi tekoälyn etiikkaan perustuva tarkastelu periaatteineen on laaja ja osin ristiriitainen. Myös lainsäädännölliset ja teknologian globaali luonne vaikuttaa auditointiin. Auditoinnilla on paljon potentiaalia parantaa ja täydentää muita jo olemassa olevia työkaluja ja mittareita menetelmineen. Auditointiin ja vaikutustenarviointiin kuuluukin

mittareiden luominen päätösten jäljittämiseksi ja seurantaan varten, jolloin teknologiat ovat käytettävissä todentamista varten. Tekoälyn etiikkaan perustuva auditointi ei tule korvaamaan eettisen pohdinnan tarvetta vaan nämä ovat toisiaan tukevia menetelmiä. (Mökander & Floridi 2021; Kamiz & Kosiyama 2021.)

6.2 Yhteiskunnallinen näkökulma

Lait ja säädökset

Tekoälyn eettiset ja yhteiskunnalliset vaikutukset ovat herättäneet teollisuuden, akateemisen tutkijoiden ja suuren yleisön mielenkiinnon. Useat valtiot ovat kehittäneet omia tekoälystrategioita, jotka sisältävät ehdotuksia uusiksi tekoälyn liittyviksi säädöksiksi ja laeiksi. Voidaan kuitenkin todeta, että tekoälyn globaalisuus haastaa ja tulee tulevaisuudessakin haastamaan kansalliset lait ja niiden käyttöalueet. Tekoäly on tuote, joka ylittää maantieteelliset rajat ja siksi on yhteiskunnallisesti vaikea ymmärtää, kuinka paikallisia ratkaisuja, voidaan luotettavasti kehittää tehokkaasti käyttöön. (Eötvös 2021; Morley ym. 2021; Mökander & Floridi 2021; Robert ym. 2020.)

Euroopan unionin sisällä tekoälyn suunnittelua ja kehittämistä lakien ja säädösten lisäksi säätelee myös yleinen tietosuoja-asetus (GDPR), joka tutkimuksissa on tulkittu tekoälyn aihealueessa ”oikeudeksi saada selityksiä”, jos automatisoidut päätökset ovat vaikuttaneet merkittävästi ihmiseen. Usein puhutaan tekoälypolitiikasta, jonka avulla voidaan varmistaa kansallinen ja kansainvälinen periaatejärjestelmä liittyen tekoälyn suunnitteluun, käyttöön ja rajoittamiseen. Lakien ja säädösten lisäksi tekoälyn periaatteet tulisi nähdä toisiaan täydentävinä elementteinä. Lain säädännön hitaus ja teknologian nopea kehittymisen takia eettiset periaatteet tulee nähdä suunnittelun perustana, jonka avulla eettiset ongelmat havaitaan nopeammin. Lakeja ja säädöksiä onkin vahvistettava, jotta niitä voidaan käyttää yleisen edun puolustamiseen ja tekoälyn suunnitteluun, kehittämiseen ja käyttöönottoon osallistuvien yksityisten ja julkisten elinten vastuun määrittelyyn. (Robert ym. 2020; Robins 2019; Virginia 2018.)

Floridi (2020) nostaa keskusteluun mukaan vastuuvakuutusmekanismien kehittämisen, jota voitaisiin vaatia erillisenä liitteenä tietyiltä tekoälyjärjestelmiltä EU:ssa ja muilla markkinoilla. Hänen mukaansa näin toimien voidaan varmistaa, että tekoälyjärjestelmä on luotettava. Tulevaisuudessa tällä järjestelyllä tulisi olemaan kilpailuetua markkinoilla. Voidaan todeta, että lain ja säädösten valossa tekoälyn etiikka ei vielä toimi sen koko potentiaalilla monisakaan tapauksissa, koska periaatteet ovat osin jopa ristiriitaisia ja hajautettuja eikä eettisistä säännöistä poikkeamisille ole asetettu seurauksia. Etiikkaa käytetään usein markkinoitustrategiana ilman siihen liittyvää arviointia ja pohdintaa. Tekoälyn eettisten ohjeiden

lukemisella ei ole merkittävää vaikutusta järjestelmien kehittäjien päätöksen tekoon. Käytännössä lakeihin ja säädöksiin verrattuna sitä pidetään usein vieraana teknisten huolien lisäosana ja ilman sitovaa kehystä se on helppo asettaa teknisen yhteisön ulkopuolelle. (Floridi 2020; Hagendorff 2020; Robert ym. 2020.)

Sosiaalinen eriarvoistuminen

Tekoälyn nopean yleistymisen kiihtyessä jokapäiväisessä elämässämme ja teknologian tarjoamat mahdollisuudet ovat kasvattaneet yleistä huolta riskeistä, joita ne aiheuttavat yksilöille ja yhteiskunnalle. Mikäli erilaisia tekoälyjärjestelmiä on suunniteltu, kehitetty tai käytetty väärin, niillä voi olla erittäin vakavia seurauksia sekä yksilöille että yhteiskunnalle. Kaikkien laadittujen eettisten periaatteiden tulee olla kulttuurisesti herkkiä sille erityiselle kontekstille, johon se on sosiaalisesti tarkoitettu. Tällainen toimintamalli edellyttäisi kulttuurikon­tekstin lisäksi myös siihen liittyvien eri rajoitusten ymmärtämistä tietyn tekoälyjärjestelmän kanssa. Toisaalta paras tapaus voisi olla kehittää tekoäly, joka tunnistaa tietyn kulttuurin kontekstiin ja mukautuu sen mukaan. Se ei yksinkertaisesti ehkä ole mahdollista erilaisista syistä, teknologian aiheuttamista rajoituksista ei-teknologiaan, eli sosiaalisiin, poliittisiin jne. liittyviin kysymyksiin. Siksi, ymmärrystä mahdollisista kompromisseista kulttuurisesti herkkyyden ja johdonmukaisuuden välillä ovat perusteltuja. Lähtökohtaisesti tekoälyn tulisi olla hyödyllinen yhteiskunnalle. Hyöty, kestävyys, solidaarisuus sekä inhimillisten arvojen edistäminen, tulisi nostaa arvoiksi, jotka edistävät ja lisäävät ihmisten hyvinvointia. Tekoäly tarjoaa lukemattomia mahdollisuuksia parantaa ja lisätä yksilöiden ja koko yhteiskunnan hyvinvointia ja tasa-arvoa. (Fukura-Parr & Gibbons 2021; Iason 2020; Robert ym. 2020.)

Arvot ovat hyvin riippuvaisia sosiokulttuurisesta kontekstista ja ne ovat usein viitteellisiä pohdiskeluprosesseja, mikä tarkoittaa, että tulemme tarvitsemaan erilaisia menetelmiä kaikkien sidosryhmien ja kulttuurien arvojen tuomiseksi esiin ja näiden arvojen selkeäksi tekemisen tulisi johtaa parempaan yhteiskuntaan. Lisäksi se lisäisi ymmärrystä ja luottamusta keinotekoi­siin autonomisiin järjestelmiin. Toisin sanoen tekoälyn päättelyn tulisi pysyä ottamaan huomioon yhteiskunnalliset arvot sekä moraaliset ja eettiset näkökohdat, jossa punnitaan eri sidosryhmien arvoja eri monikulttuurisissa yhteyksissä, selittää sen perustelut ja takaa avoimuuden. (Rossi 2020; Virginia 2018.)

Laadittavissa tekoälyyn liittyvissä ohjeissa olisi korostettava sen tuomia mahdollisuuksia lisätä sosioekonomista eriarvoisuutta kulttuurillisesti eikä vain katsoa tekoälyn syrjintää yksilön kannalta. Onkin kiinnitetty huomiota siihen, että tekoälyn teknologioiden käyttämisessä on jätetty nämä kulttuurisidonnaisten vaikutusten tarkastelu vähälle. Maantieteelliset erot ovat kuitenkin merkittäviä ja tekoäly voi pahentaa sosiaalista eriarvoisuutta ja

sosiaalisten erojen lisääntymistä. Eri valtioiden oikeudelliset ja taloudelliset vastuuseen liittyvät kysymykset vaihtelevat myös suuresti kansojen välillä ja niiden sisällä. Mikäli haluamme ajatella tekoälyn etiikkaa universaalina tulisi sen sisältöä kehittää ja siihen sisällyttää vaihtelevia ja joustavia standardeja kaikkien yhteiskuntien kehityksen mukaisesti. (Fukura-Parr & Gibbon 2021; Hagerty & Rubinov 2019; Robert ym. 2020; Morley ym. 2021.)

Eettisten vaikutusten sosiaalinen ymmärtäminen vaatii ymmärrystä, miten koko järjestelmä toimii mukaan lukien se mitä voi tapahtua, kun järjestelmä on jo otettu käyttöön ja sitä käytävää yhteisöt tai muut tahot muuhun tarkoitukseen kuin alkuperäisen kehittäjän ilmoitettuun tarkoitukseen. (Morley ym. 2021.)

6.3 Ihmisyden näkökulma

Ihmisoikeudet

Ihmisoikeudet pitävät sisällään; yksilön yksityisyyden, vapauden ja autonomian, luottamuksen ja ihmisarvon. Tekoälyllä on jo suuria vaikutuksia yksilöön ja ihmisoikeuksiin. Jo syntynyt yksimielisyys "eettisestä tekoälystä" on ongelmallinen, koska siitä katsotaan puuttuvan kansainvälinen ihmisoikeuslaki ja se painottaa vain vähän yksilön omaa vastuuta ja osallistumista. Olemme jo huomanneet, että tekoälyn on todettu tuottavan epäreiluja, puolueellisia ja epäeettisiä päätöksiä. Esimerkkinä voidaan todeta viimeaikaisia tapauksia, joihin on liittynyt tekoälyjärjestelmien suorittamaa seulontaa niin, että sen on osoitettu olevan puolueellinen mm. naisten tai muiden vähemmistöjen palkkaamisessa. Uudet erilaiset teknologiat voivat siis lisätä yksilön sosiaalista eriarvoisuutta, uhata yksilön ihmisoikeuksia, helpottaa valtioiden autoritaarisia käytäntöjä, hallita ja kaupallistaa yksilön yksityisiä tietoja tai luoda uusia järjestelmiä, jotka riistävät ihmisiltä oman päätöksentekovallan. Kansainvälisesti sovitut ihmisoikeudet tarjoavat vahvemmat puitteet, koska ne ovat laillinen joukko sovittuja arvoja ja normeja ja niitä valvotaan kansallisten ja kansainvälisten oikeusmekanismien avulla. Tekoälyn mahdollisten eettisten riskien ennustaminen nousee siksi tärkeäksi asiaksi, sillä sen avulla voidaan ennustaa ja minimoida vaikutuksia ihmisarvoon ja ihmisoikeuksiin. (Floridi 2020; Fukura-Parr & Gibbons 2021; Robbins 2019; Robert ym. 2020.)

Yksi tärkeimmistä eroista eettisen lähestymistavan ja ihmisoikeuslähestymistavan välillä on tapa, jolla vastuuta ja vallan epäsymmetriaa käsitellään. Ihmisoikeudet eivät itsessään ole vain joukko arvoja, vaan ne on suunniteltu toimimaan mielivaltaisen vallan rajoittimena. Tekoälyn eettinen tarkastelu ja viitekehysten käyttö lisää yksilöiden luottamusta tekoälyyn ja sen avulla voidaan yhtenäistää ihmisoikeuksia. Tämän vuoksi on tärkeää saada luotua universaalinen tekoälyn eettinen viitekehys, joka perustuu oikean ja väärän normeihin. Vastuullisuus, osallistuminen, mukaan lukien siihen liittyvä tiedon avoimuuden vaatimus ja

vastuu ovat toisiinsa liittyviä ihmisoikeuksiin perustuvia tärkeitä periaatteita. Ihmisoikeudet ja etiikka eivät siis sulje toisiaan pois vaan täydentävät toisiaan. Ihmisoikeuksien todellinen lisäarvo piilee vastuullisuudesta sen täydessä merkityksessä. Tämän vuoksi on tärkeää tuoda ihmisoikeudelliset sääntelyt ja lait tarkasteluun ja painottaa vastuullisuutta, jotka kuvaavat mitä "vastuullinen tekoäly" tarkoittaa. (Fukura-Parr & Gibbons 2021; Hagerty & Rubinov 2019; Iason 2020.)

Iason (2020) herätteleekin kysymyksellään; ”mitä ihmisoikeuksia tekoälyn tulisi yhtenäistää? Tulisiko meidän huolehtia vain kielteisistä velvollisuuksista, olla vahingoittamatta ihmisiä? vai myös positiivisista velvoitteista varmistaaksemme, että yksilöt voivat saada elintärkeitä tavaroita ja palveluita? Tämän kysymyksen edessä näyttäisi, että olemme erilaisten kompromissien edessä. (Iason 2020.)

7 Johtopäätökset ja pohdinta

7.1 Integroidun kirjallisuuskatsauksen arviointi ja pohdinta

Integroidun kirjallisuuskatsauksen keskeinen osa arviointi. Arviointia tulee tehdä kaikissa prossin vaiheissa, vaikka se kuvataan yleensä omana vaiheena. Kirjallisuuskatsauksen aikana tehtävän arvioinnin tarkoituksena on ohjata työtä oikeaan suuntaan. Kirjallisuuskatsauksen lopussa arvioinnin tehtävänä on osoittaa, kuinka hyvin onnistuttiin ja miten sille asetetut tavoitteet saavutettiin. Kirjallisuuskatsauksen arvioinnin kohteena on usein tavoitteet, suunnittelutyö, toteutus, lopputulos ja näiden suhteet. (Ojasalo ym. 2015, 47.)

Tämän opinnäytetyön tavoitteena oli integroidun kirjallisuuskatsauksen avulla selvittää, millaisia tekoälyn eettisiä periaatteita on laadittu ja kuinka tekoälyn eettinen kehittäminen tulisi suunnitella, jotta se on kestävä. Tuloksia tullaan hyödyntämään LAB WellTech-kehittämisympäristön tekoälyn suunnittelu-, kehittämis- ja arviointityössä. Tämän perusteella luodaan tekoälyn periaatteista ja mittareista opas kehittämisympäristössä työskentelevien tueksi. Kehittämisympäristö toimii monialaisesti, joka mahdollistaa monoammatillisen eettisen arvioinnin. Opinnäytetyötä käytetään myös henkilöstön tiedon lisäämiseen aiheesta.

Valittu aihe on ajankohtainen ja kehittämistyön kannalta tärkeä, sillä tekoälyn lisääntyvä hyödyntäminen ihmisten keskuudessa lisääntyä vauhtia. LAB WellTech-kehittämisympäristössä kehitetään ja arvioidaan hyvinvointiteknologiaan liittyviä tuotteita ja usein tekoälyllä on osuutta tuotteiden toiminnassa. Siksi tämä opinnäytetyö tulee vastaamaan käytännön kehittämistyön haasteisiin ja antaa työkaluja vastata tekoälyn eettisyyden haasteisiin.

Integroitu kirjallisuuskatsaus eteni sykleissä ja kesti pitempään kuin alun perin oli ajatus. Tutkimuksen suunnittelu aloitettiin vuonna 2020 ja tarkentui 2021 maaliskuussa integroituun kirjallisuuskatsaukseen. Aineiston haku suoritettiin loka-marraskuussa 2021, sisälönanalyysi suoritettiin maaliskuun ja toukokuun aikana 2022.

Integroidun kirjallisuuskatsauksen tulee perustua aiheesta jo olemassa olevaan tietopohjaan. (Ojasalo ym. (2015, 21) Tämän opinnäytetyön tekemisessä hyödynnettiin teorian ja tutkimustulosten tietoa koko prossin ajan. Teoriatietopohja tukee hyvin tutkimustuloksia.

Tässä integroidussa kirjallisuuskatsauksessa päästiin tutkimukselle asetettuihin tavoitteisiin. Kuvattiin tekoälyn suunnitteluun ja kehittämiseen keskeisesti vaikuttavat eettiset periaatteet sekä määriteltiin millaiset tekoälyn eettiset periaatteet tukevat vastuullista ja kestävä tekoälyn suunnittelu- ja kehittämistyötä.

7.2 Eettisyys ja luotettavuus

Opinnäytetyön kaikissa vaiheissa eettiset näkökohdat on otettava huomioon ja noudatettava tutkimuseettisen neuvottelukunnan tieteelliselle tutkimukselle laatimia ohjeita. Tutkimus voi olla eettisesti hyväksyttävä ja luotettava kun siinä noudatetaan hyviä tieteellisiä käytänteitä. Näitä käytänteitä ovat; huolellisuus, rehellisyys ja tarkkuus jokaisessa tutkimustoiminnan vaiheessa. (Tutkimuseettinen neuvottelukunta 2012, 6.)

Arene (2020,5) painottaa ohjeistuksessaan ammattikorkeakoulujen opinnäytetöiden eettisten suositusten mukaan, että ylemmän korkeakoulun opinnäytetyön tekijän tulee hallita hyvä tieteellinen käytäntö, asiakokonaisuudet ja kehittämishankkeeseen liitoksissa oleva lainsäädäntö, ihmisiin liittyvät eettiset erityispiirteet sekä ammattialaan liittyvät eettiset normit ja käytännöt soveltuvin osin. (Arene 2020, 5.)

Tutkimustoiminnan aloittamisen edellytyksenä on, että kaikkien osallistuvien osapuolten kanssa on tehty sopimus. Tässä sopimuksessa selvennetään vastuu ja velvollisuus periaatteeneineen. Lisäksi tarvittavien tutkimuslupien hakemisesta on huolehdittava. (Tutkimuseettinen neuvottelukunta 2012, 6.) Kehittämishankkeen alussa solmitaan kirjallinen yhteistyösopimus työelämän edustajan ja opin-näytetyön tekijän kanssa. Sopimuksessa sovitaan kehittämishankkeen tausta ja tarkoitus sekä opinnäytetyön aihe ja ajallinen kokonaiskesto. Lisäksi sovitaan muiden mahdollisten kulujen korvaamisesta ja salassapito velvoitteista sekä tietojen sisällyttämisestä.

Tutkimuslupa tutkimukseen hankittiin LAB-ammattikorkeakoulu Oy ja tutkimus toteutettiin integroivana kirjallisuuskatsauksena. Kaikki tutkimuksen prosessit kirjoitettiin auki, jotta tutkimuksen kulku on hyvin havaittavissa ja mahdollistaa aineisto on pyritty tuomaan tarkasti esille, jotta tutkimus on mahdollisimman hyvin uudelleen toteutettavissa. Tässä kirjallisuuskatsauksessa tavoite ja tarkoitus on esitelty selvitettynä sekä tutkimuskysymykset kriteereinä esitelty. Tutkimukseen valikoitunut aineisto edustaa aiemmin tehtyjä tutkimuksia ja aiheeseen liittyviä tieteellisiä vertaisarvioituja artikkeleja. Aineisto haettiin LAB Primo-tietokirjaston hakukonetta ja Google Scholar-tietokantaa käyttäen, joka poistaa tutkimukseen liittyviä tutkimuksellisia riskejä. Tutkimuksen heikkouksena voidaan pitää, ettei aiheeseen löytyviä alkuperäistutkimuksia löytynyt tai tutkimus ei ollut vapaasti luettavissa, vaikka tutkimuskysymyksiin tehtiin omat laajat haut hakukoneista. Tutkimukseen valikoitu aiheisto analysoitiin tutkimuskysymysten ja näkökulmien ohjaamana.

Hakutermien muodostaminen tuotti alkuvaiheessa lukuisia tuloksia, joten hakutermien muodostamiseen ja koehakuihin käytettiin runsaasti aikaa. Hakutermien muodostamiseen käytettiin LUT-tiedekirjaston informaation palveluita, jossa käytiin läpi mitä hakukonetta

tutkimuksessa olisi hyvä käyttää ja millaisilla hakusanoilla saadaan kirjallisuuskatsaukseen parhaat mahdolliset tulokset. Hakukriteerit ohjasivat aineiston valintaa. Aineistoksi valittiin vertaisarvioidut alkuperäistutkimukset tai artikkelit. Aineistosta jätettiin pois kirjoittajan omaa mielipidettä, ilman tutkimuspohjaa olevat julkaisut. Yhteensä hakuosumia tuli xxx. Ensimmäisessä aineiston karsinnassa luettiin tutkimukset otsikon ja tiivistelmän perusteella niin, että ne, jotka eivät vastanneet tutkimuskysymyksiin ja kirjallisuuskatsauksen aiheeseen poistettiin. Otsikko ja tiivistelmätason karsimisen jälkeen tutkimuksia ja vertaisarvioituja artikkeleita oli xxx. Nämä tutkimukset luettiin kokonaisuudessaan, jonka jälkeen tutkimukseen valikoitui xxx tutkimusta ja vertaisarvioitua artikkelia. Sisäänottokriteerit ja arvointikriteerit ohjasivat aineiston määrittelyssä. Kaikki tutkimukseen valikoituneet tutkimukset olivat englanninkielisiä ja kansainvälisiä sekä tutkimusten aineisto kattava tutkijan mielestä tutkittavaan aiheeseen.

7.3 Johtopäätökset ja jatkotutkimusaiheet

Tämän tutkimuksen tuloksia voidaan hyödyntää eettisen tekoälyn kehittämistyössä. Käytännössä tekoälyn etiikkaa pidetään usein vieraana tai teknisten aiheiden lisäosana, mutta tutkimusten mukaan sitä ei tule pitää erillisenä osana vaan liittää tekoälyn kehittämisen ja käytön koko elinkaareen. Tutkimuksen tavoitteena oli integroivan kirjallisuuskatsauksen avulla selvittää ja kuvata tämän hetken tekoälyn eettisyyden periaatteiden tilannekuva ja kuinka periaatteet tulisi ottaa huomioon tekoälyn suunnittelussa ja kehittämisessä, jotta kehittämis työ on kestävä. Eettinen tekoäly onkin soveltavan etiikan haara ja on kiinnostunut siitä, kuinka tekoälyjärjestelmät vaikuttavat ihmisiin. Miten järjestelmiä kehitetään, kuinka prosessit etenevät, kuinka päätöksenteon logiikka toimii, kuinka tehtävät jakautuvat ja kuka tekee päätöksen, miten ja missä määrin käsitellään vaikutuksia ja miten riskejä ja haittoja mitataan.

Tutkimustulosten mukaan tekoäly tekniikkana on tehokas, jolla on useita myönteisiä vaikutuksia meihin ja elämäämme. Toisaalta sillä on paljon havaittuja haittavaikutuksia, jotka aiheuttavat kestävä tekoälyn suunnitteluun haasteita. Näiden ongelmien havaitsemiseen on kehitetty jo useita mittareita ja säädetty standardeja ja oikeilla työkaluilla voimmekin arvioida sen tuomia hyötyjä ja haittoja. Haasteita tekoälyn eettisyyden suunnitteluun ja ihmisiin kohdistuvien vaikutusten arviointiin tuo se, että mittareita on useita ja niistä oikean mittarin valinta voi olla vaikeaa ilman tarkempaa tietoa mittarin toiminnasta. Meidän on kuitenkin pyrittävä rakentaa järjestelmä, joka tukee luotettavaa tekoälyä ja luottamusta tekniikkaan ja sen kehittäjiin. Haittojen, selittävyyden, tietojenkäsittelyn tulee olla avointa ja suunnittelun että kehittämisen valintojen käsittelyn vastuullista ja avointa. Mikäli kykenemme lisäämään tekoälyn tuottamiin vaikutuksiin vankat ja todistetut vinouman havaitsemis- ja

lievennysominaisuudet sekä kyvyn selittää miten sen päätökset tehdään, voimme kuroa umpeen luottamuksen tutkimuksessa ilmi tullutta kuilua ja näin luoda mahdollisuus yhteiskunnallisille eduille, joissa myös ihminen käyttäjä on tärkeässä asemassa.

Useat korkean tason periaatteet, jotka ohjaavat tekoälyä kohti myönteisiä vaikutuksia olisikin aika ottaa konkreettisesti käyttöön ja luoda niille vahvat käyttöönottomekanismit. Tällä hetkellä tekoälyn periaatteellisia ohjeita on useita, joita Hagerdoff 2020 tutkimuksessaan vertaili. Hänen mukaansa periaatteet tarvitsevat vielä globaalia tarkastelua yhteisten periaatteiden luomiseen, jota hän pitää vaikeana, muttei kuitenkaan mahdottomana toteuttaa kulttuurien eroavaisuuksien takia. Tekoälyn eettiset periaatteet ja eettinen kestävä kehittäminen vaatii siis kokonaisvaltaista ja monialaista lähestymistapaa, sillä vain silloin voidaan rakentaa ja luoda järjestelmä, johon kuuluvat tekoälyn kehittäjät, tekoälyn käyttäjät ja päättäjät. Tällaisella järjestelmällä voisimme varmistaa, että tulevat ongelmat tunnistetaan, niistä keskustellaan yhteisesti ja ratkaistaan organisaatiollisella sekä poliittisella tasolla. Näin toimien löytäisimme parhaat ratkaisut, jotka voisivat johtaa kokonaisvaltaiseen luotettavaan tekoälyyn.

Voidaankin todeta, että yleinen tekoälyn hyväksyntä ja käyttöönotto on mahdollista silloin kun hyödyt nähdään merkityksellisinä ja hyväksytään mahdolliset riskit, jotka ovat kuitenkin ehkäistävissä, vähennettävissä tai niihin voidaan varautua riskienhallinnan avulla. Yleiset asenteet tekoälyä kohtaan riippuvat yksilön, kehittäjän ja yhteiskunnan yhteisestä sitoutumisesta tekoälyteknologioiden kehittämiseen sekä avoimuudesta.

Kirjallisuuskatsaus antoi kattavan kuvauksen tämänhetkisestä tekoälyn etiikan ja sen kehittämisen tilanteesta. Etairos (Ethical AI for the Governance of the Society) -hanke, joka tutkii ja etsii erilaisia tapoja, kuinka voidaan ottaa huomioon eettiset kysymykset tekoälyn kehittämisessä ehdottaa, että suomeen tulisi perustaa tekoälypoliittinen neuvottelukunta, luoda tekoälylle oma kasvuohjelma, vahvistaa monialaista yhteistyötä, tehostaa tekoälyn vaikutusten arvioinnin ja ennakkoinnin toteutusta ja lisätä koulutusta, jolla vahvistetaan myös osaamista.

Tutkimuksen tuloksista on hyötyä tekoälyn eettisyyden arvioinnissa, tekoälyn kestävä kehittäminen suunnittelussa ja tietoisuuden lisäämisessä. Tässä opinnäytetyössä tarkasteltiin kansainvälisiä tutkimuksia tekoälyn etiikasta, sen periaatteista ja kehittämisestä. Tutkimukset osoittivat, että kansainvälisiä ohjeita ja suosituksia tulisi yhtenäistää ja luoda globaalit periaatteet. On myös luotu monia erilaisia arviointi ja auditointi työkaluja, jotka vaatisivat vielä tarkempaa selvittämistä. Tekoälyn vaikutukset ihmisiin vaativat vielä lisää tutkimusta ja tarkastelua yksilö- ja yhteiskuntatasolla, tästä aiheesta jatkotutkimuksien tekeminen on ajankohtaista.

Lähteet

AI HLEG 2019. Luotettavaa tekoälyä koskevat eettiset ohjeet. Saatavissa: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMIT-TEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_FI.pdf.

AI HLEG 2020. The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Viitattu 1.2.2021. Saatavissa: <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

Ailisto, H (toim.), Neuvonen, A., Nyman, H., Halén, M., & Seppälä, T. 2019. Tekoälyn kokonaiskuva ja kansallinen osaamiskartoitus – loppuraportti. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja. 4/2019. Viitattu 2.2.2021. Saatavissa: Tekoälyn kokonaiskuva ja kansallinen osaamiskartoitus – loppuraportti. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja. 4/2019. Saatavissa: <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160925/46-2018-Tekoalyn%20kokonaiskuva.pdf>

Ailisto, H. (toim.), Heikkilä, E. Helaakoski, H., Neuvonen, A. & Seppälä, T. 2018. Tekoälyn kokonaiskuva ja osaamiskartoitus. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 46/2018. Viitattu 13.1.2021. Saatavissa: <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160925/46-2018-Tekoalyn%20kokonaiskuva.pdf>

AlgorithmWatch. Viitattu 15.1.2021. Saatavissa: <https://algorithmwatch.org/en/>

Arene 2020. Ammattikoulujen opinnäytetöiden eettiset suositukset. Ammattikorkeakoulujen rehtorineuvosto Arene ry. Viitattu 27.1.2021. Saatavissa: <http://www.arene.fi/wp-content/uploads/Raportit/2020/AMMATTIKORKEAKOULU-JEN%20OPINN%C3%84YTET%C3%96IDEN%20EETTISET%20SUOSITUKSET%202020.pdf? t=1578480382>

Aylin, J. & Chapman, A. 2021. Putting AI ethics to work: are the tools fit for purpose? 2021. Springer Link.

Bansal, G., Lutge, C., & Robert Jr. 2019. ICIS 2019 SIGHCI Workshop Panel Report: Human-Computer Interaction Challenges and Opportunities for Fair, Trustworthy and Ethical Artificial Intelligence. AIS Transactions on Human-Computer Interaction. 2019. Vol. 12, 97-107.

Bietti, E. 2021. From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics. Journal of Social Computing. Viitattu 27.5.2022. Saatavissa: <https://ieeexplore.ieee.org/document/9684746>

- Coughlan, M., & Cronin, P. 2017. Doing a literature review in nursing, health and social care. Viitattu 28.6.2021. Saatavissa: https://books.google.fi/books?hl=fi&lr=&id=-900DQAAQ-BAJ&oi=fnd&pg=PP1&dq=Coughlan+%26+Cronin+2017&ots=zMg3CzF1Kq&sig=AuR3jzbGXfsh5-Latn2I7aMJKuY&redir_esc=y#v=onepage&q=Coughlan%20%26%20Cronin%202017&f=false
- Coeckelbergh, M. 2021. Tekoälyn etiikka. Helsinki: Terra Cognita.
- Elements of AI, 2018. Verkkokurssi. Viitattu 14.1.2021. Saatavissa: <https://course.elementsofai.com/fi>.
- Erlingsson, C. & Brysiewicz, P. 2017. A hands on guide to doing content analysis. African Journal of Emergency Medicine 7: 93–99.
- Ethics of AI, 2020. Verkkokurssi. Viitattu 4.3.2021. Saatavissa: <https://ethics-of-ai.mooc.fi/>
- ETAİROS, 2019. ETAİROS- tutkimushanke. Viitattu 16.4.2021. Saatavissa: <https://etairos.fi/>
- Euroopan komissio, 2020. Valkoinen kirja, Tekoälystä –Eurooppalainen lähestymistapa huippuosaamiseen ja luottamukseen. Viitattu 3.3.2021. Saatavissa: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_fi.pdf
- Flinkman, M. & Salanterä, S. 2007. Integroitu katsaus –eri metodeilla tehdyn tutkimuksen yhdistäminen katsauksessa. Teoksessa Johansson, K., Axelin, A., Stolt, M. & Ääri, R-L. (toim.) Systemaattinen kirjallisuuskatsaus ja sen tekeminen. Hoitotieteen laitoksen julkaisu A: 51/2007. Turku: Turun yliopisto, 84–100.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, V. 2018. AI4People— An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds & Machines. 2018. Vol. 28, 689-707.
- Fukuda-Parr, S. & Gibbons, E. 2021. Emerging Consensus on ‘Ethical AI’: Human Rights Critique of Stakeholder Guidelines. Global Policy. 2021. Vol. 12.
- Goodfellow, I.; Bengio, Y. & Courville, A. 2018. Deep Learning. Cambridge, Massachusetts: The MIT Press. Viitattu 3.3.2021. Saatavissa: <http://www.deeplearningbook.org/>
- Gunter, K., & Hasanen, K. 2021. Laadullisen tutkimuksen verkkokäsikirja. Tutkimuksen suunnittelu. Viitattu: 15.4.2021. Saatavissa: <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvali/laadullisen-tutkimuksen-prosessi/tutkimuksen-suunnittelu/>

- Hagendorff, T. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines*. 2020. Vol. 30, 99-120.
- Hagerty, A. & Rubinov, I. 2019. Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. Cornell University .2019. ArXiv.
- Hauer, T. 2020. Machine Ethics, Allostery and Philosophical Anti-Dualism: Will AI Ever Make Ethically Autonomous Decisions? *Society*. 2020, Vol. 57, 425-433.
- Hirsjärvi, S. Remes, P. & Sajavaara, P. 2009. Tutki ja kirjoita. 15. uudistettu painos. Helsinki.
- Harjamäki, J. 2018. Selvitys Data-analytiikan nykytilasta ja data-analytiikan hyödyntämisestä Satakunnassa. Tampereen Teknollinen yliopisto. Viitattu 14.1.2021. Saatavissa: <http://www.datatiede.fi/wp-content/uploads/2018/10/Data-analytiikan-selvitys-Julkaisuversio-2-2018-10-30.pdf>
- Heinimäki, J. 2012. Taskuetiikka. Tallinna: Raamatutrukikoda
- Hiekkanen, K. 2017. Digitalisaatio haastaa rakentajiensa arvot – tekoälyä koskeva eettinen keskustelu on olematonta, laki laahaa perässä. Viitattu 1.2.2021. Saatavissa: <https://www.tekniikkatalous.fi/blogit/digitalisaatio-haastaa-rakentajiensa-arvot-tekoalya-koskeva-eettinen-keskustelu-on-olematonta-laki-laahaa-perassa/8675ec60-9ab5-303c-9691-797e968fc47b>
- Iason, G. 2020. Artificial Intelligence, Values, and Alignment. *Minds & Machines*. 2020. Vol. 30, 411-437.
- IEEE (2019). Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. Viitattu 2.4.2021. Saatavissa: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2_feedback_rfi_responses.pdf
- Johansson, K., Axelin A., Stolt, M., Ääri R–L. 2007. Systemaattinen kirjallisuuskatsaus ja sen tekeminen. Turun yliopisto: Digipaino.
- Jääskeläinen, A. 2019. Mitä tapahtuu huomenna, kun tekoäly poistaa järjettömyydet? Helsinki: WSOY.
- Kangasniemi, M., Utriainen, K., Ahonen, S., Pietilä, A-M., Jääskeläinen, P. & Liikanen, E. 2013. Kuvaileva kirjallisuuskatsaus: eteneminen tutkimuskysymyksestä jäsenettyyn tietoon. *Hoitotiede* 25 (4), 291–301.
- Kananen, H., Puolitaival, H., 2019. Tekoäly – Bisneksen uudet työkalut. Liettua: BALTO

- Kazim, E. & Koshiyama A. 2021. A high-level overview of AI ethics. Patterns. 2021. Vol. 2
- Khan, A., Akbar, M., Badshah, S., Khan, B., Liang, P., Niazi, M. & Waseem, M. 2021. Ethics of AI: A Systematic Literature Review of Principles and Challenges. Cornell University. 2021. ArXiv.
- Koivisto, R., Leikas, J., Auvinen, H., Vakkuri, V., Saariluoma, P., Hakkarainen, J., Koulu, R. 2019. Tekoäly viranomaistoiminnassa – eettiset kysymykset ja yhteiskunnallinen hyväksyttävyyys. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2019:14. Viitattu 8.2.2021. Saatavissa: <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf>
- Kotimaisten kielten keskus 2021. Kielitoimiston sanakirja verkossa. Viitattu 5.4.2021 Saatavissa: <https://www.kielitoimistonsanakirja.fi/netmot.exe?motportal=80>.
- Kotkavirta, J. & Nyssönen, S. 1996. Ajatus etiikka. Porvoo: Sanoma Pro Oy.
- Kyngäs, H., Elo, S., Pölkki, T., Kääriäinen, M. & Kanste, O. 2011. Sisällönanalyysi Suomalaisessa hoitotieteellisessä tutkimuksessa. Hoitotiede 23 (2), 138–148.
- LAB-ammattikorkeakoulu. 2019. Viitattu 13.1.2021. Saatavissa: <https://www.lab.fi/fi/projekti/tekoalypaja>
- LAB-ammattikorkeakoulu. 2020. Viitattu 15.1.2021. Saatavissa: <https://lab.fi/fi/info/tietoameista>
- LAB-ammattikorkeakoulu. 2021. Viitattu 8.2.2021. Julkaisematon aineisto.
- Laaksonen, H. & Ollila, S. 2017. Lähijohtamisen perusteet terveydenhuollossa. Helsinki: Edita.
- Latvanen, M. (2019). Tekoäly ja tekoälyn etiikka-muutamia perusasioita. Kuntamarkkinat 2019. Viitattu 5.4.2021. Saatavissa: https://www.kuntaliitto.fi/sites/default/files/media/file/Teko%C3%A4ly%20ja%20teko%C3%A4lyn%20etiikka_esitykset.pdf
- Leikas, J. 2018. Ikääntyvät, teknologia ja etiikka – Näkökulmia ihmisen ja teknologian vuorovaikutukseen ja – suunnitteluun. Viitattu 1.2.2021. Saatavissa: <https://www.vttresearch.com/sites/default/files/pdf/workingpapers/2008/W110.pdf>
- Leikas, J. 2021. Tekoäly raportti. Eduskunnan tulevaisuusvaliokunta. Viitattu 27.5.2022. Saatavissa: <https://www.eduskunta.fi/FI/vaski/JulkaaisuMetatieto/Documents/EDK-2021-AK-405729.pdf>

- Merilehto, A. 2018. Tekoäly: Matkaopas johtajalle. Helsinki: Alma Talent.
- Mora-Cantalops, M., Sánchez-Alonso, S., García-Barriocanal, E. & Sicilia, M-A. 2021. Traceability for Trustworthy AI: A Review of Models and Tools. Big Data Cogn. Comput. 2021. 5(2), 20.
- Morley, J., Elhahal, A., Garcia, F., Kinsey, L., Mökander, J. & Floridi, L. 2021. Ethics as a service: A Pragmatic Operationalisation of AI Ethics. Minds and Machines. 2021. Vol. 31, p239-256.
- Mökander, J., Floridi, L. 2021. Ethics-Based Auditing to Develop Trustworthy AI. Minds & Machines. 2021. Vol. 31, 323-327.
- OECD 2019a. Forty-two countries adopt new OECD Principles on Artificial Intelligence. Press release 22.5.2019. Viitattu 6.3.2021. Saatavissa: <http://www.oecd.org/going-digital/forty-two-count-ries-adopt-new-oecd-principles-on-artificial-intelligence.htm>.
- OECD 2019b. OECD Principles on AI. Viitattu 6.3.2021. Saatavissa: <https://www.oecd.org/going-digital/ai/principles/>
- Ojasalo, K. Moilanen, T & Ritalahti, J. 2015. Kehittämistyön menetelmät: uudenlaista osaamista liiketoimintaan. 3.-4. painos. Helsinki: Sanoma Pro Oy.
- Ojanen, A., Oljakka, N., Salhgren, O., Tuikka, A-M., Vaiste, J. 2019. Opas tekoälyn etiikkaan. Turku AI Society. Viitattu 27.3.2021. Saatavissa: <https://aisociety.fi/fi/opas-tekoalyn-etiikkaan>
- Ollila, M–R. 2019. Tekoälyn etiikkaa. Otava. Helsinki.
- Onwuegbuzie, A., & Frels, R. 2016. 7 steps to a comprehensive literature review. Sage. London.
- Robbins, S. 2019. A Misdirected Principle with a Catch: Explicability for AI. Minds & Machines. 2019. Vol. 29, 495-514.
- Roos, T.2018. Johdatus tekoälyyn. Viitattu 15.1.2021. Saatavissa: https://www.cs.helsinki.fi/webfm_send/932.
- Royackers, L., Timmer, J., Kool, L. & Van Est, R. 2018. Societal and ethical issues of digitization. Ethics Inf Technol 20, 127–142. Viitattu: 3.2.2021. Saatavissa: <https://link.springer.com/article/10.1007/s10676-018-9452-x>

Salonen, A., Eloranta, S., Hautala, T. & Kinos, S. Kehittämistoiminta ja kehittämisen menetelmiä ammatillisessa koulutuksessa. 2017. Turun ammattikorkeakoulu. Viitattu 4.2.2021. Saatavissa: <http://julkaisut.turkuamk.fi/isbn9789522166494.pdf>

Salminen, A. (2011). Mikä kirjallisuuskatsaus? Johdatus kirjallisuuskatsauksen tyypeihin ja hallintotieteellisiin sovelluksiin. Vaasan yliopiston julkaisuja. Julkisohtaminen. Viitattu: 23.4.2021. Saatavissa: https://www.univaasa.fi/materiaali/pdf/isbn_978-952-476-349-3.pdf

Shalev-Shwartz, S., & Ben-David, S. Understanding Machine Learning, 2014, Cambridge University Press.

Seppänen, A. 29.10.2020. Eetikko. Tekoälyn etiikka: ajattelun apuvälineitä eettisesti kestäviin ratkaisuihin. GTalks-webinaari. Viitattu 3.2.2012. Saatavissa: <https://www.youtube.com/watch?v=6qYwKhy4oOQ>.

Shneiderman, B. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human-Computer Interaction. 2020. Vol 36, 495-504

Sivistyssanakirja 2002. WSOY. Helsinki

Stolt, M., Axelin, A. & Suhonen, R. 2016. Kirjallisuuskatsaus hoitotieteessä. Turku: Turun yliopisto

Toikka, T. & Rantanen, T. 2009. Tutkimuksellinen kehittämistoiminta: näkökulmia kehittämisprosessiin osallistamiseen ja tiedontuotantoon. Tampere: Tampere University Press.

Tuomi, J. & Sarajärvi, A. 2018. Laadullinen tutkimus ja sisällön analyysi. Helsinki: Kustannusosakeyhtiö Tammi.

Tutkimuksellinen neuvottelukunta 2012. Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa. Viitattu 18.1.2021. Saatavissa: https://tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf

Whittemore, R. 2008. Rigourin Integrative Reviews. Teoksessa Reviewing Research Evidence for Nursing Practice: Systematic Reviews, 149–156. Toim. Christine Webb & Brenda Ross. Oxford: Blackwell Publishing.

Whittemore, R. & Knafelz, K. 2005. The integrative review: updated methodology. Methodological issues in nursing research. Journal of Advanced Nursing 2005. 52(5), 546-553

Liite 1. Aineiston esittely ja pisteytys.

Tekijä(t), vuosi, tutkimuksen/artikkelin nimi ja maa	Tarkoitus/Tavoite	Menetelmä/Analyysi	Tulokset tiivistettynä	Pisteet
[1] Hagendorff, T. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. Minds & Machines. Vol. 30, 99-120	<p>1. Kartoittaa tekoälyn etiikkaan liittyvien ohjeiden puutteita ja päällekkäisyyksiä.</p> <p>2. Tutkia kuinka eettiset periaatteet toteutuvat käytännössä ja kuinka etiikan vaatimusten tehokkuutta voidaan parantaa.</p>	Systemaattinen kirjallisuuskatseaus, joka toteutettiin kahdessa vaiheessa. Haku- toimintojen kautta valikoitui 22 läntisiä/pohjoisia tekoälyn eettistä ohjeistoa.	Tutkimuksen mukaan tekoälyn etiikka epäonnistuu monissa tapauksissa ja sitä pidetään usein vieraina, ylijäämänä tai jonkinlaisena teknisten huolenaiheiden lisäosana, joka asetetaan teknisen yhteisön ulkopuolelle. Taloudelliset kannustimet ylittävät usein sitoutumisen eettisiin periaatteisiin ja arvoihin. On kiinnitettävä enemmän huomiota tekoälyn, koneoppimisen ja eri menetelmien ja teknologioiden yksityiskohtiin. Arvojen ja teknisten toteutuksien välille on rakennettava konkreettisia välineitä.	5/5
[2] Mökander, J., Floridi, L. 2021. Ethics-Based Auditing to Develop Trustworthy AI. Minds & Machines. Vol. 31, 323-327	<p>1. väite, etiikkaan perustuva auditointi voi parantaa päätöksenteon laatua, lisätä käyttäjien tyytyväisyyttä, vapauttaa kasvupotentiaalia, mahdollistaa lainsäädäntötyön ja lievittää inhimillistä kärsimystä.</p> <p>2. korostaa, nykyisiä parhaita käytäntöjä etiikkaan perustuvan auditoinnin suunnittelun ja toteuttamisen tukemiseksi.</p> <p>3. tunnistaa eettiseen auditointiin liittyvät rajoitteet.</p>	Vertaisarvioitu artikkeli	Päättäjiä rohkaistaan pitämään etiikkaan perustuvaa auditointia olennaisena osana tekoälyn aiheuttamien eettisten riskien hallintaan. Tämä ei tarkoita sitä, että perinteiset noudattamismekanismit olisivat tarpeettomia. Sen sijaan etiikkatekoälyn auditoinnilla on potentiaalia täydentää ja parantaa muita työkaluja ja menetelmiä, kuten ihmisen valvonta, sertifiointi ja sääntely.	2/5

<p>[3] Robbins, S. 2019. A Misdirected Principle with a Catch: Explicability for AI. Minds & Machines. Vol. 29, 495-514</p>	<p>1. väite, että periaatteet, jotka edellyttävät, että tekoäly on selitettävissä, ovat harhaanjohtavia. 2. väite, filosofinen merkitys on, että ominaisuus "selitettävyyden vaatimusta" on sovellettu väärin tekoälyyn.</p>	<p>Vertaisarvioitu artikkeli Analyysi tekijäanalyysilla.</p>	<p>Selitettävyyden vaatimisen ominaisuus koneoppimisen algoritmeille kuuluu itse päätöksille ja toimille – ei toiminnan tai päätöksen suorittavalle taholle. Kun huomio kohdistetaan näihin päätöksiin ja toimiin, voimme päättää paremmin, missä yhteyksissä ja rooleissa meidän tulisi käyttää koneoppimisen-algoritmeja.</p>	<p>2/5</p>
<p>[4] Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, V. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds & Machines. Vol. 28, 689-707</p>	<p>1.kartoittaa tekoälyteknologian tarjoamat mahdollisuudet ja niihin liittyvät riskit ihmisarvon vaalimiseen ja edistämiseen; periaatteet, joiden pitäisi olla tekoälyn käyttöönoton taustalla 2. kartoitus tekoälyn tarjoamista keskeisistä ihmisarvon edistämismahdollisuuksista sekä niihin liittyvistä riskeistä 3. mitä etuja organisaatioille on eettisen lähestymistavan kehittämisessä tekoälyn käyttöönottamisessa 4. muotoilla 5 tekoälyn eettistä periaatetta olemassa oleviin analyyseihin perustuen</p>	<p>Tiedekomitean vertaisarvioitu tieteellinen artikkeli. Kokoo yhteen havainnot, jotka on suunniteltu luomaan perusta hyvälle tekoäly yhteiskunnalle. Koostuu neljästä erillisestä poikkileikkauksesta.</p>	<p>Tämä artikkeli ja siinä esitetyt suositukset – pyrkivät edistämään eettisesti ja sosiaalisesti kestävien tekoälyteknikoiden kehittämistä, suunnittelua ja käyttöönottoa. Tässä artikkelissa tunnistetaan tekoälyn keskeiset mahdollisuudet ja riskit sekä viisi tärkeintä eettistä periaatetta, joiden avulla on muotoiltu 20 konkreettista toimintakohtaa vastaamaan tekoälyn asettamiin eettisiin haasteisiin.</p>	<p>4/5</p>
<p>[5] Iason, G. 2020. Artificial Intelligence, Values, and Alignment. Minds & Machines. Vol. 30, 411-437</p>	<p>Tavoitteena tunnistaa ja tarkastella filosofia kysymyksiä tekoälyn yhteydessä.</p>	<p>Tieteellinen vertaisarvioitu artikkeli Yleiskatsaus kirjallisuuteen, koostuu kolmesta erillisestä</p>	<p>Tekoälyn kehittäminen ja luominen vaikuttaa arvoihin, jonka vuoksi normatiivisia kysymyksiä ei voida kokonaan sulkea pois vaan ne pitäisi ottaa mukaan tutkimuksiin. Tekoälyn periaatteista tulisi</p>	<p>3/5</p>

		poikkileikkaukseta.	tunnistaa ne, joita yleisesti pidetään oikeanlaisina.	
[6] Shneiderman, B. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human-Computer Interaction. Vol 36, 495-504				
[7] Hauer, T. 2020. Machine Ethics, Allostery and Philosophical Anti-Dualism: Will AI Ever Make Ethically Autonomous Decisions? Society. Aug2020, Vol. 57, 425-433.	Artikkelin pääehdotus on, että kysymyksiin, jotka koskevat autonomisten älykkäiden järjestelmien ja tekoälyalustojen etiikkaa ja jotka kehittyvät ajan myötä datasta koneoppimisen kautta, ei voida verrata yleisen etiikan käsitteillä ja ajatuksilla.	Vertaisarvioitu artikkeli. Analyysi tekijän analyysilla.	Artikkeli puolustaa pääargumenttia, että jos tehtävä vaatii jonkinlaista moraalista auktoriteettia ihmisten suorittamissa tehtävissä, niin tehtävän automatisointi eli saman tehtävän siirtäminen autonomisille koneille, alustoille ja tekoälyalgoritmeille merkitsee välttämättä moraalisen osaamisen siirtämistä koneille.	1/5
[8] Aylin, J. & Chapman, A. 2021. Putting AI ethics to work: are the tools fit for purpose? Springer Link.	Tutkimuksessa tarkastellaan eettisiä viitekehyksiä keskittyen niihin, jotka menevät periaatelausekkeita pidemmälle. Tutkimuksessa analysoidaan jo olemassa olevia tekoälyn eettisiä työkaluja ja niiden soveltuvuutta tekoälyn kehittämisessä huomioiden tekoälyn elinkaaren. Tutkimuksessa tuodaan esille vaikutusten arvioinnin ja auditoinnin tarpeellisuus.	Kvalitatiivinen sisällönanalyysi Kirjallisuuskat-saus.	Arviointi- ja hallintotuloksiin soveltuvien työkalujen selventäminen ja hyvin säänneltyjen vaatimustenmukaisuusjärjestelmien täytäntöönpano tekoälyjärjestelmien tuottajille olisi suuri askel kohti tekoälyn eettisten työkalujen tuotantoa motivoivien eettisten periaatteiden ja huolenaiheiden tehokasta toteuttamista. Tutkimuksessa löytyi myös aukkoja monien sidosryhmien sisällyttämisessä tekoälyn eettisten työkalujen prosessiin.	5/5

<p>[9] Khan, A., Akbar, M., Badshah, S., Khan, B., Liang, P., Niaz, M. & Waseem, M. 2021. Ethics of AI: A Systematic Literature Review of Principles and Challenges. Cornell University. ArXiv.</p>	<p>Tutkimuksen tavoitteena on selvittää mitkä ovat tekoälyn etiikan keskeiset periaatteet? Mitä haasteita etiikan omaksumisessa on tekoälyssä?</p>	<p>Systemaattinen kirjallisuuskatsaus.</p>	<p>Tutkimuksessa analysoitiin 22 eri tekoälyn eettistä periaatetta. tekoälyn neljäksi tärkeimmiksi periaatteiksi havaittiin avoimuus, yksityisyys, vastuullisuus ja oikeudenmukaisuus. Lisäksi tutkimus toteaa, että päätöksentekoon johtavien järjestelmien tulee myös olla tietoisia eettisistä periaatteista, jotta tiedetään seuraukset teoista.</p>	<p>5/5</p>
<p>[10] Bansal, G., Lutge, C, & Robert Jr. 2019. ICIS 2019 SIGHCI Workshop Panel Report: Human-Computer Interaction Challenges and Opportunities for Fair, Trustworthy and Ethical Artificial Intelligence. AIS Transactions on Human-Computer Interaction. Vol. 12, 97-107.</p>	<p>Paneelikeskustelun tavoitteena on avata tekoälyn tuomia mahdollisia haasteita ja vaikutuksia, joihin liittyy oikeudenmukaisuus, luottamus ja etiikka. Tarkastelussa neljä aihetta. Reilun, eettisen ja luotettavan tekoälyn suunnittelu ja käyttöönotto. Reilun, eettisen ja luotettavan tekoälyn politiikka ja hallinto. Tekoälyn tarkoituksenmukaiset ja sopimattomat sovellukset. Epäreilun, epäeettisen ja epäluotettavan tekoälyn vaikutukset</p>	<p>Paneelikeskustelun raportti. Ensimmäinen osa jäsenelty. Toinen osa strukturoimaton, joka koostui yleisökysymyksistä.</p>	<p>Paneelikeskustelun tuloksena syntyi käsitys tekoälyn oikeudenmukaisuuden, luotettavuuden ja nykyisestä ja tulevasta tilasta tekoälyn etiikassa. Panelistit ja yleisön jäsenet olivat samaa mieltä myös siitä, että tieteenalan on tehtävä harkittua tutkimusta siitä, miten kehittää ja ottaa tekoälyjärjestelmiä käyttöön.</p>	<p>5/5</p>
<p>[11] Hagerty, A. & Rubinov, I. 2019. Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. Cornell University. ArXiv.</p>	<p>Selvittää kirjallisuuskatsauksen avulla tekoälyn sosiaalisten ja eettisten vaikutusten tilanne maailmanlaajuisesti.</p>	<p>Kirjallisuuskatsaus, joka kattoi yli 800 akateemista artikkelia ja monografiaa usealla kielellä. Laadullinen sisällönanalyysi.</p>	<p>Katsaus ja synteesi analyysi viidellä maailman alueella, erityisesti 14 maassa toteaa, että tekoälyllä voidaan odottaa olevan kriittisiä sosiaalisia vaikutuksia kaikkialla maailmassa. Kulttuurisidonnaisuudella on merkittäviä alueellisia vaihteita.</p>	<p>4/5</p>

<p>[12] Mora-Cantallops, M., Sánchez-Alonso, S., García-Barriocanal, E. & Sicilia, M-A. 2021. Traceability for Trustworthy AI: A Review of Models and Tools. Big Data Cogn. Comput. 5(2), 20.</p>	<p>Tarkoituksena tarjota yleiskatsaus tekoälyn jäljitettävyydestä ja käytössä olevista työkaluista AI HLEG -ohjeiden perusteella.</p>	<p>Analyysi tekijänalyysilla.</p>	<p>Jäljitettävyys on avaintekijä läpinäkyvien tekoälyjärjestelmien toiminnassa ja suunnittelussa. Päättäjiä rohkaistaan pitämään etiikkaan perustuvaa auditointia olennaisena osana tekoälyn kehittämistä.</p>	<p>2/5</p>
<p>[13] Kazim, E. & Koshiyama A. 2021. A high-level overview of AI ethics. Patterns. Vol. 2</p>	<p>Antaa kattava monitieteellinen korkean tason johdanto tekoälyn etiikkaan filosofian, oikeuden ja tietojenkäsittelytieteen pohjalta.</p>	<p>Analyysi tekijänalyysilla.</p>	<p>Tässä artikkelissa tarjottiin käsitteellisiä peruskatsauksia termeistä, kuten tekoäly ja etiikka ja tutkittiin tekoälyn etiikan edeltäjiä, tekniikan etiikkaa, teknologiafilosofiaa sekä tiede- ja teknologiaopintoja. Nostettiin esiin kolme nykyisestä tekoälyn etiikkaan liittyvästä lähestymistavasta, periaatteista, prosesseista ja eettisestä tietoisuudesta. Tekoälyn etiikan muuttamista insinöörikäytännöksi tarkasteltiin ihmiskeskeisen tekoälyn, turvallisuuden, läpinäkyvyyden, oikeudenmukaisuuden ja yksityisyyden kautta.</p>	<p>3/5</p>
<p>[14] Fukuda-Parr, S. & Gibbons, E. 2021. Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines. Global Policy. Vol. 12.</p>	<p>Tarkoituksena käsitellä: ihmisoikeuskritiikkiä, joka kohdistuu etiikan kehukseen nouseviin konsensusnormeihin.</p>	<p>Kirjallisuuskatsoaus. Analysointi 15 ohjeesta.</p>	<p>Tutkimuksen mukaan tiukemman ihmisoikeuslähestymistavan soveltaminen korostaa puutteita eettisten ohjeiden täytäntöönpano kelpoisuudessa. Lisäksi tarkastelut eettiset ohjeet tarjosivat epämääräisen muotoilunsa mukaisesti vähän tai ei ollenkaan, toteutuskeinoja, eivätkä tarjonneet realistisia keinoja tekoälyn suunnittelun, kehittämisen ja</p>	<p>4/5</p>

			käyttöön oton muokkamiseen todellisessa maailmassa.	
[15] Morley, J., Elhahal, A., Garcia, F., Kinsey, L., Mökander, J. & Floridi, L. 2021. Ethics as a service: A Pragmatic Operationalisation of AI Ethics. Minds and Machines. Vol. 31, p239-256.	Tarkoituksena selvittää voiko tekoälyn etiikka olla hyödyllistä tekoälyn kehittäjille? Tähän kysymykseen pyritään vastaamaan tutkimuksessa ja selvittää, miksi periaatteita ja teknisiä työkaluja tarvitsee kehittää edelleen, vaikka niitä onkin jo käytössä.	Systemaattinen kirjallisuuskatsoaus	Tutkimuksen mukaan on mahdollista löytää tekoälyn periaatteiden ja kehittämisen yhdistävä tasapaino. Tuloksissa esitellään oletus, että tekoälyn etiikka hyötyisi mukautetusta lähestymistavasta ja että tämä voi olla saavutettavissa.	4/5

