

Anni Ohrimovitsch

KONEOPPIMISEN EETTISYYS TERVEYDENHUOLLOSSA

Kandidaatintyö
Johtamisen ja talouden tiedekunta
Tarkastaja: Ilona Ilvonen
06/2023

TIIVISTELMÄ

Anni Ohrimovitsch: Koneoppimisen eettisyys terveydenhuollossa
The Ethics of Machine Learning in Healthcare
Kandidaatintyö
Tampereen yliopisto
Tietojohdamisen tutkinto-ohjelma
Kesäkuu 2023

Tekoäly kehittyä jatkuvasti ja se on ajankohtainen ilmiö. Tekoälyä hyödynnetään päivittäin monessa yhteydessä ja sen tarkoitus on koneena tai ohjelmana oppia ja tehdä päätöksiä lähes ihmisen tavoin. Tämän tutkimuksen aiheena on koneoppimisen hyödyntäminen terveydenhuollossa ja sen eettisyys. Tutkimus keskittyy tekoälyn osa-alueeseen koneoppimiseen, koska tekoälyn hyödyntäminen on kasvanut terveydenhuollossa ja lääketieteessä vasta koneoppimisen kehityksen myötä. Hyödyntämisen tavoite lääketieteessä on terveydenhuollon palvelun parantaminen. Koneoppimisen algoritmeilla on erityisiä vahvuuksia luokituksessa, kuvioden tunnistamisessa, ennustamisessa, järjestelmän parametrien optimoinnissa ja havainnoidun datan monimutkaisten dynamiikan mallien rakentamisessa. Lääketieteellisissä diagnoosissa pyritään määrittämään sairaudet, jotka selittävät potilaan oireet. Diagnosoinnin suorittaa lääkäri ja kun diagnosointiin käytetään tukena koneoppimistyökalua, sen perustoimintaperiaate on tunnistaa sairaudet, jotka korreloivat vahvasti potilaan oireiden kanssa. Tutkimuksen tavoitteena on tunnistaa eettisyys koneoppimisen hyödyntämisessä ja kehittämisessä terveydenhuollossa ja tutkia koneoppimisen nykytilaa etiikan näkökulmasta terveydenhuollon diagnostiikassa.

Eettisyyttä käsiteltäessä käsitteet oikeasta ja väärästä, oikeudenmukaisuudesta ja epäoikeudenmukaisuudesta, ja hyvästä ja pahasta nousevat esiin. Tekoälyn (AI) etiikka seuraa teknologian muutosta, ja sen vaikutusta yksilöön sekä yhteiskunnan ja talouden muutoksiin. Sen tarkoitus on verrata tekoälyn etuja eettisiin huolenaiheisiin. Eettisyyttä tarkastellaan tässä tutkimuksessa Euroopan komission tekoälyn eettisten ohjeiden avulla. Niissä määritellään eettisten periaatteiden ja arvojen noudattaminen, johon kuuluvat ihmisen itsemääräämisoikeuden kunnioittaminen, vahinkojen välttämisen periaate, oikeudenmukaisuuden periaate ja selitettävyyden periaate.

Tutkimusmenetelmänä toimi kuvaileva kirjallisuustutkimus. Aineisto kerättiin kirjallisuuskatsauksella, jonka avulla etsittiin ajankohtaista tutkimusta koneoppimisesta terveydenhuollossa. Tutkimuksen perusteella löydettiin keskeisiä teemoja eettisten periaatteiden ja arvojen noudattamiseen. Jokaisessa eettisyyden periaatteessa esiintyi kolme keskeistä teemaa, jotka löydettiin aineiston perusteella. Ammatillaisen rooli koneoppimisen hyödyntämisessä ja käyttöönotossa, asiakkaan tietoisuus datan käytöstä ja koneoppimisen luotettavuus esiintyivät ihmisen itsemääräämisoikeuden periaatetta tukien. Koneoppimisalgoritmin vastuullinen kehitys ja hyödyntäminen, ammatillaisen mahdollisuus vaikuttaa koneoppimisalgoritmin toimintaan, ja koneoppimisen hyödyntäminen ammatillaisen toimesta diagnosoinnin tukena esiintyivät vahinkojen välttämisen periaatteeseen vaikuttaen. Koneoppimisen vahvuuksien ja heikkouksien tunnistaminen ja tiedostaminen, ja ammatillaisen velvollisuudet tukivat oikeudenmukaisuuden periaatetta. Datan laatu algoritmin kehityksessä, koneoppimisalgoritmin toimivuuden varmistaminen ja tarkka toiminnan selvitys, ja koneoppimisalgoritmin hyödyntäminen päätöksen teon tukena tukivat selitettävyyden periaatetta. Kirjallisuuskatsauksen perusteella eettisyys näkyi sekä algoritmin kehittämisessä, käyttöönotossa ja toiminnassa että ammatillaisen toiminnassa. Tutkimuksessa lopuksi todettiin, että diagnoosi yhdessä ammatillaisen ja koneoppimisalgoritmin toimesta tuottaa eettisyyden kannalta merkityksellisimmän tuloksen ja että eettisyyden tarkastelu on yhä tärkeämpää, kun koneoppimisen toiminta ja mahdollisuudet kehittyvät.

Avainsanat: Koneoppiminen, terveydenhuolto, lääketieteellinen diagnostiikka, etiikka, tekoälyn etiikka

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

SISÄLLYSLUETTELO

1. JOHDANTO	1
1.1 Tutkimusaiheen tausta	1
1.2 Tutkimuskysymykset	2
1.3 Tutkimuksen rajaus ja näkökulma	2
1.4 Tutkimuksen rakenne	4
2. TUTKIMUKSEN TOTEUTUS	6
2.1 Tutkimusmenetelmä	6
2.2 Tutkimusaineisto	6
3. KONEOPPIMINEN JA SEN EETTISYYS	8
3.1 Koneoppiminen	8
3.2 Vastuu koneoppimisessa	9
3.3 Koneoppimisen eettisyys	10
4. KONEOPPIMINEN TERVEYDENHUOLLOSSA	12
4.1 Koneoppimisen nykytila terveydenhuollossa	12
4.2 Data ja koneoppiminen terveydenhuollossa	13
4.3 Diagnostiikka ja koneoppiminen	15
4.4 Eettisyys ja koneoppiminen terveydenhuollossa	17
5. YHTEENVETO	25
5.1 Tulokset	25
5.2 Päätelmät ja tutkimuksen arviointi	26
LÄHTEET	28

1. JOHDANTO

1.1 Tutkimusaiheen tausta

Tekoälyä hyödynnetään nykyään päivittäin monessa eri yhteydessä, kuten erilaisissa palveluissa, sosiaalisessa mediassa sekä puhelimissa ja tietoliikenteessä. Tekoälyllä tarkoitetaan konetta tai ohjelmistoa, joka kykenee oppimaan ja tekemään järkeviä päätöksiä tehtävissä ja tilanteissa lähes ihmisen tavoin. (Siukonen & Neittaanmäki 2019, 2)

Tässä tutkimuksessa aiheena on yhden tekoälyn osa-alueen, koneoppimisen hyödyntäminen terveydenhuollon aloilla. Tutkimus keskittyy hyödyntämiseen ja kehittämiseen eettisyyden näkökulmasta. Koneoppimisen avulla tapahtuva päätöksenteko ei ole yksiselitteistä. Kirjassa AI Ethics esitellään tekoälyn erilaisia eettisiä tilanteita. Esimerkki haastavasta päätöksenteosta on itseohjautuvat autot ja niiden päätöksenteko tilanteessa, jossa on kaksi huonoa lopputulemaa ja itseohjautuvan auton tulee päättää niistä toinen. (Coeckelbergh 2020, s. 5)

Tekoälyn ja koneoppimisen eettisyyttä on tärkeää ja mielenkiintoista tutkia terveydenhuollossa ja lääketieteessä, sillä päätöksenteon ja tulosten algoritmien tulisi olla tasarvoisia kaikille terveydenhuollon asiakkaille. Koneoppiminen kehittyy jatkuvasti sekä itsenäisesti dataa kerätessään että algoritmin tekijän avulla, jolloin eettisyyttä voi tarkastella monesta näkökulmasta.

Tekoälyn innovointiin liittyy sekä haittoja että hyötyjä. King et al. (2020) käsittelevät artikkelissaan haittoja, hyötyjä sekä mahdollisia riskejä, kuten rikollisuutta, tekoälyn innovoinnissa. Koneoppimisen käyttöönoton ja hyödyntämisen tavoite lääketieteessä on parantaa terveydenhuollon palvelua. Laadun merkitys on suuri, ja siten voi pohtia, kuinka palvelun laatua testataan ja parantaako se asiakkaan palvelua. Palvelun parantamisen ja terveydenhuollon organisaation taloudellisten voittojen vastakkainasettelu saattaa esimerkiksi vaikuttaa koneoppimisen algoritmien toimintaan. (Char et al. 2020)

1.2 Tutkimuskysymykset

Tässä kirjallisuustutkimuksessa pyritään tutkimaan tekoälyn ja tarkemmin koneoppimisen nykytilaa terveydenhuollossa etiikan näkökulmasta sekä etiikan vaikutusta koneoppimisen kehittämiseen ja käyttöönottoon. Tutkimusongelma on tekoälyn eettisyys terveydenhuollossa.

Päätutkimuskysymys on seuraava:

- kuinka eettinen ajattelu näkyy koneoppimisen käyttöönotossa ja hyödyntämisessä terveydenhuollossa?

Alatutkimuskysymyksiä on seuraavat kysymykset:

- Mitä on koneoppiminen ja sen eettisyys?
- Kuinka koneoppimista hyödynnetään terveydenhuollossa, ja erityisesti terveydenhuollon diagnostiikassa?

1.3 Tutkimuksen rajaus ja näkökulma

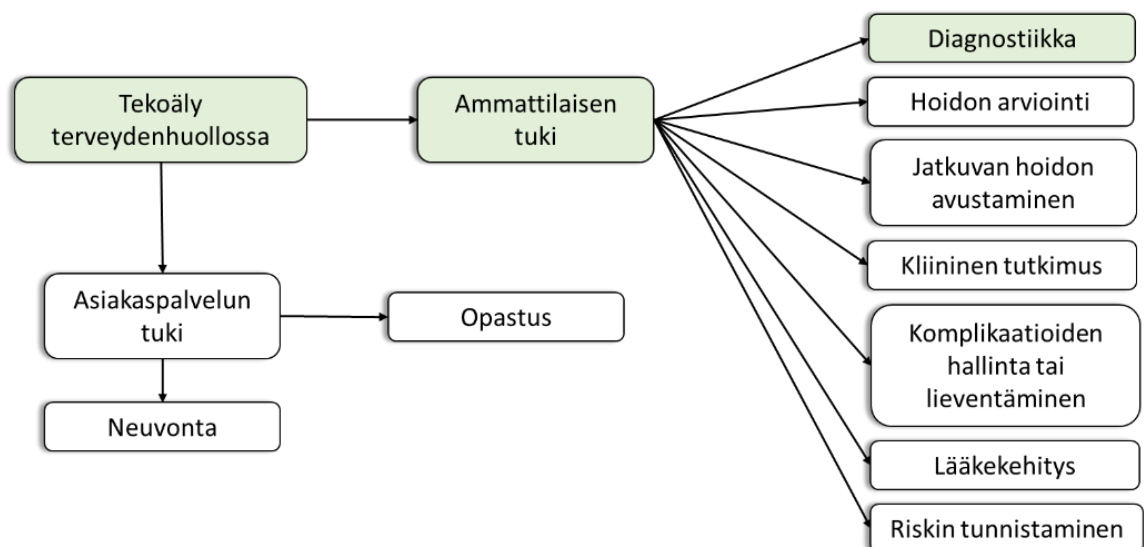
Koneoppiminen, terveydenhuolto ja eettisyys ovat laajoja käsitteitä. Koneoppimiseen kuuluu kolme vaihetta. Ensin koneoppimisen algoritmi koulutetaan datalla ja sen laatua tarkkaillaan, seuraavana tapahtuu algoritmin muuttuminen eli itsenäinen oppiminen ja viimeisenä algoritmi tuottaa päätöksen syötetyn datan avulla (Chandramouli et al. 2018, 1.4.1). Tämä päätös luo vastakkainasettelun koneen ja ihmisen välille, ja muodostaa kysymyksen, kuinka vastuu jaetaan päätöksien seurauksissa, kun siinä hyödynnetään itse oppivaa algoritmia.

Ensimmäisenä rajauksena tutkimukselle toimii tekoälyn yksi osa-alue **koneoppiminen**. Koneoppimista pidetään tekoälyn keskeisenä osaamisalueena ja syväoppimista koneoppimisen osa-alueena (Ailisto et al. 2018, s. 1) (Kuva 1). Tutkimuksen toisena rajauksena toimii **koneoppimisen eettisen vastuun tarkastelu**. Terveydenhuollossa ja lääketieteessä tekoälyn hyödyntäminen on laajentunut vasta koneoppimisen ja syväoppimisen tulon myötä (Kaul et al. 2020), ja siten tarkastelu ja lähteiden haku painottuu koneoppimiseen.



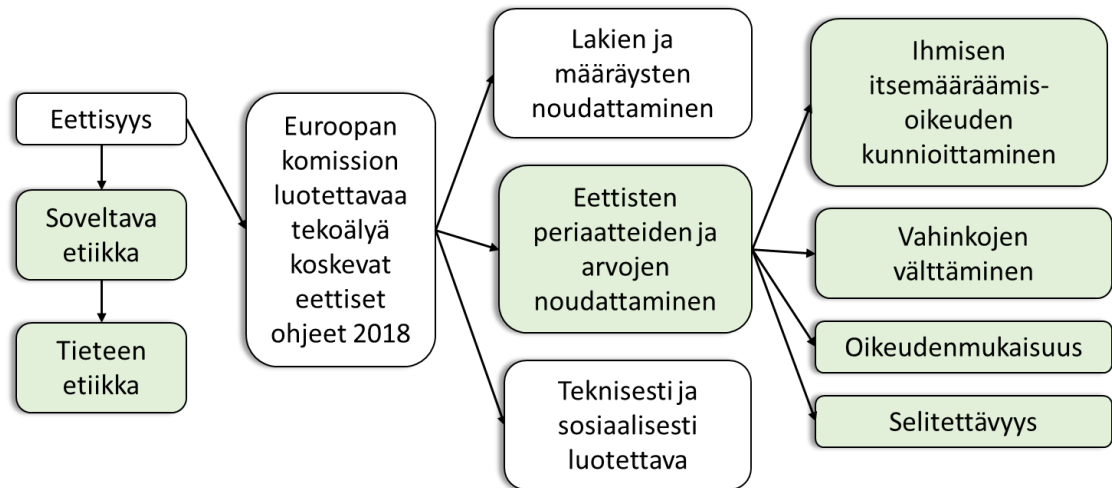
Kuva 1. Koneoppiminen tekoälyn osa-alueena ja syväoppiminen koneoppimisen osa-alueena.

Tekoälyä voidaan hyödyntää terveydenhuollossa sekä ammattilaisen että asiakaspalvelun tukena (kuva 2). Asiakaspalvelua voidaan tukea esimerkiksi yleisellä neuvonnalla ja opastuksella. Koneoppimista voidaan hyödyntää terveydenhuollossa ammattilaisen tukena eri tavoin, kuten diagnostiikassa, hoidon arvioinnissa, jatkuvan hoidon avustamisessa, kliinisessä tutkimuksessa, komplikaatioiden hallinnassa tai lieventämisessä, lääkekehityksessä ja riskin tunnistamisessa (Neittaanmäki et al. 2019, s. 121–128) sekä ammattilaista voidaan tukea esimerkiksi nopeammalla tiedonhakuprosessilla (Hunter 2019). Tutkimuksen kolmantena rajauksena toimii koneoppimisen käyttö ammatillisen päätöksenteon tukena, ja jotta tutkimuksesta saadaan syvällisempi, koneoppimista terveydenhuollossa tarkastellaan syvemmin **diagnostiikan** avulla (kuva 2).



Kuva 2. Tekoälyn mahdollinen hyödyntäminen terveydenhuollossa

Etiikka on haastava käsite yhdistää koneoppimiseen, sillä perinteisesti etiikka hakee ihmisen moraaliseen toiminnalle teoreettisia perusteluja (Pietarinen 2015). Jos etiikalla tutkitaan ihmisen toimintaa, tekoäly ei voi itsessään olla eettistä tai epäeettistä. Tällöin tekoälyn etiikka käsitteenä voi olla harhaanjohtava. Tässä tutkimuksessa tekoälyä tarkastellaan tieteen etiikalla, joka on määritelty tekoälyn eettisyyden tyypiksi (Roth 2016, s. 92) (Kuva 3).



Kuva 3. Eettisyyden näkökulman rajausta tutkimukseen

Neljäntenä rajauksena tässä työssä toimii eettisyyden näkökulma Euroopan komission luotettavan tekoälyn eettisten ohjeiden mukaan. Rajausta on jatkettu **eettisten periaatteiden ja arvojen noudattamiseen**. (Kuva 3) Siihen kuuluu neljä osa-alueita: ihmisen itsemääräämisoikeuden kunnioittaminen, vahinkojen välttäminen, oikeudenmukaisuus ja selitettävyyys. (AI HLEG 2019, s. 14)

1.4 Tutkimuksen rakenne

Johdanto-osiossa käydään läpi tutkimusaiheen tausta, tutkimuskysymykset, rajausta ja näkökulmat sekä tutkimuksen rakenne. Seuraavaksi tutkimuksessa esitetään tutkimuksen toteutus kertomalla tutkimusmenetelmästä ja -aineistosta.

Tutkimuksen teoriaosuus alkaa kolmannessa kappaleessa. Kolmannessa kappaleessa pyritään vastaamaan ensimmäiseen alatutkimuskysymykseen: *Mitä on koneoppiminen ja sen eettisyys?* Kappaleessa käsitellään koneoppimista yleisesti sekä koneoppimisen ennustavia analyysejä ja syväoppimista. Tämän jälkeen siirrytään vastuun jakautumisen haasteeseen. Vastuunjaon jälkeen esitellään tarkemmin Euroopan komission asiantuntijaryhmän luoman luotettavan tekoälyn eettisiä ohjeita ja tarkemmin eettisten periaatteiden ja arvojen noudattamisen ohjeet.

Neljännessä kappaleessa pohditaan koneoppimisen nykytilaa ja datan merkitystä terveydenhuollossa. Tämän jälkeen syvennyttään koneoppimiseen terveydenhuollon diagnostiikassa. Tämän kohdan avulla pyritään vastaamaan toiseen alatutkimuskysymykseen: *Kuinka koneoppimista hyödynnetään terveydenhuollon diagnostiikassa?* Neljännen kappaleen lopuksi, kohdassa 4.4, yhdistetään eri teoriaosuuksia, jotta voidaan vastata niiden avulla tutkimuksen päätutkimuskysymykseen: *Kuinka eettinen ajattelu näkyy koneoppimisen käyttöönotossa ja hyödyntämisessä terveydenhuollossa?* Viidennessä osassa tutkimusta on yhteenveto, jossa pohditaan tutkimuksen keskeisiä tuloksia ja jatkotutkimusmahdollisuuksia ja arvioidaan tutkimuksen laatua.

2. TUTKIMUKSEN TOTEUTUS

2.1 Tutkimusmenetelmä

Tutkimus toteutettiin kuvailevana, integroivana kirjallisuuskatsauksena. Kirjallisuuskatsauksessa tehtiin yleiskatsaus aiheesta ilman tarkempia sääntöjä (Salminen 2011, s. 6). Kuvaileva kirjallisuuskatsaus voidaan toteuttaa joko narratiivisesti tai integroivasti. Tässä tutkimuksessa tehtiin integroiva kirjallisuuskatsaus, jossa pyrittiin kuvaamaan tutkittava asia monipuolisesti (Salminen 2011, s. 8). Tutkimuksessa kuvailtiin ilmiön ominaisuuksia niiden etiikan rajojen sisällä, jotka esiteltiin kohdassa tutkimuksen rajaus ja näkökulma. Aluksi asetettiin tutkimusongelma, josta edelleen asetettiin tutkimuskysymykset. Tutkimusongelma ja -kysymykset esitettiin kohdassa 1.2 Tutkimuskysymykset, ja niihin vastattiin tällä kandidaatin työllä.

Tutkimusaineistoa haettiin tietokannoista. Aineiston valinnan jälkeen aineisto arvioitiin ja analysoitiin. Kirjallisuuskatsauksesta tehtiin tulkinta löydetyistä aineistoista ja esiteltiin tulokset. Aineiston läpikäynnissä hyödynnettiin myös systemaattisen kirjallisuuskatsauksen piirteitä. (Salminen 2011, s. 8–9)

2.2 Tutkimusaineisto

Tietokantana käytetään pääasiassa Andoria. Sen lisäksi tietoa haetaan Google Scholarista ja erilaisista lääketieteen ja tietotekniikan tietokannoista. Hakulausekkeet ovat tutkimuksen alussa valittu siten, että saadaan yleisiä käsitteitä ja asioita perusteltua. Tutkimuksen teoriaosan syventyessä hakulausekkeilla pyritään löytämään eri sanoilla ja synonyymeillä pienempi joukko hakutuloksia juuri tutkittavaan aiheeseen liittyen ja rajataan hakua eri haun kohde -rajauksilla. Pääasialliset aineistotyyppit ovat artikkelit ja kirjat, mutta myös muita aineistotyyppejä tutkitaan, kuten asiakirjoja tai lakeja. Aineisto valitaan lukemalla läpi tiivistelmiä, johdantoja tai päätelmiä aineistosta sekä tutkimalla avainsanoja. Tutkimuksen avainsanoja ovat koneoppiminen, terveydenhuolto, lääketieteellinen diagnostiikka, etiikka ja tekoälyn etiikka. Haussa toimivat esimerkiksi seuraavanlaiset lauseet, joissa käytettävät käsitteet valitaan haluttujen tulosten mukaan sekä kokeilemalla, millä termistöllä tulokset ovat osuvimpia:

- Nimeke sisältää (ML OR "machine learning" OR "koneoppimi*") AND Aihe sisältää ("responsibility" OR "vastuu" OR "Accountability")
- Nimeke sisältää ("Machine Learning") AND ("Health Care" OR "Healthcare") AND Kaikki kentät sisältää ("golden standard" OR "state of art" OR "state-of-art")

Monipuolisuutta tuloksiin saadaan hyödyntämällä sekä lyhenteitä että varsinaista käsitettä ja yhdistämällä ne esimerkiksi käsiteltävään aiheeseen. Tutkimusaineiston hakulauseet ja hakuihin tehdyt rajaukset sekä tulokset eri tietokannoista kerätään Exceeliin. Tutkimuksessa tehdyn yhden haun esimerkki Andor-tietokantaan on esitelty Taulukossa 1. Samankaltaisia hakuja toteutetaan myös muihin tietokantoihin ja hakuja toistetaan yhdistämällä eri avainsanoja ja hakulauseita tarkennetaan tarvittaessa hakujen edetessä.

Taulukko 1. *Yhden haun hakulauseet, rajaukset ja tulokset*

HAKULAUSEKE JA LISÄTYT RAJAUKSET	TULOKSIEN MÄÄRÄ
Nimeke sisältää (ML OR "machine learning" OR "koneoppimi*") AND Aihe sisältää ("responsibility" OR "vastuu" OR "Accountability")	84
1. rajaus: Artikkelit	42
2. rajaus: Saatavilla verkossa	36
3. rajaus: Vertaisarvioidut lehdet	26
joista otsikon ja tiivistelmän perusteella valittiin	7
joista vastuu-aiheeseen osui	5
joista diagnostiikka-aiheeseen osui	2

Taulukossa 1. näkyy yksi alkuperäisistä hakulausekkeista ja sen antamien tuloksien määrä sekä eri rajoitusten lisäämisen jälkeinen tuloksien määrä. Haululla etsitään artikkeleita, jotka ovat saatavilla verkossa. Vertaisarvioitujen lehtien tarkasteleminen tuo aineistoon luotettavuutta. Otsikosta ja tiivistelmästä haetaan tähän tutkimukseen liittyviä teemoja. Esimerkiksi Taulukon 1. hausta pyritään löytämään koneoppimiseen liittyviä teemoja, jotka voitaisiin tulkita eettisyyden näkökulmasta. Sen lisäksi hakutuloksista kerätään suoraan diagnostiikkaan ja koneoppimiseen liittyvät artikkelit, jotta niihin voidaan perehtyä tarkemmin. Aineiston valinnan jälkeen hauilla löydetyt aineistot taulukoidaan toiseen taulukkoon hakusanan mukaan ja merkitään aineiston nimi sekä mihin kansioon aineisto on tallennettu Zotero -viitteiden hallinta työkaluun. Aineistojen sisältöä kerätään ajatuskarttoihin ja muistiinpanoihin, ja tutkimukseen sisällytettyjä aineistoja käsitellään lopuksi etiikan näkökulmien avulla. Tämän tutkimuksen analyysi tehdään näiden työkalujen avulla.

3. KONEOPPIMINEN JA SEN EETTISYYS

3.1 Koneoppiminen

Mitchell (1997, s.2) määrittelee kirjassaan koneoppimisen seuraavasti: Tietokoneohjelman sanotaan oppivan kokemuksesta E, kunnioittaen joidenkin tehtävien luokan T ja suorituskyvyn P suhdetta, jos sen suorituskkyky tehtävä T:ssä mitattuna P:llä paranee kokemuksen E kanssa. Tarkemmin koneoppimisessa kuvataan tilastotekniikoiden ja numeeristen algoritmien käyttöä tehtävien suorittamisessa ilman tarkkoja ohjelmoituja ja menettelyllisiä ohjeita. Koneoppimisen algoritmeilla on erityisiä vahvuuksia luokituksessa, kuvioiden tunnistamisessa, ennustamisessa, järjestelmän parametrien optimoinnissa ja havainnoidun datan monimutkaisten dynamiikan mallien rakentamisessa. (Jordan & Mitchell 2015) Chandramouli et al. (2018, 1.4) pohjustavat koneoppimista samalla määritelmällä kuin Mitchell (1997) ja selittävät sen tarkoittavan sitä, että kone kykenee oppimaan, jos se kykenee keräämään kokemusta suorittamalla tiettyjä tehtäviä, ja siten parantamaan omaa suoritustaan samankaltaisissa tehtävissä tulevaisuudessa.

Koneoppimisessa tapahtuva yksityiskohtainen informaatioprosessinen algoritmi ja sen oppimisen ymmärtäminen voi johtaa ihmisten oppimiskykyjen ja haasteiden parempaan ymmärtämiseen. Koneoppimisen algoritmeja hyödynnetään suurilla datapohjilla, kuten potilastiedoilla, tietämyksen saavuttamiseksi. Koneoppimisen yhtenä näkökulmana toimii se, että koneoppiminen etsii laajalta alueelta mahdollisia hypoteeseja, jotka tukevat parhaalla mahdollisella tavalla havaittua dataa ja oppijan aikaisempaa tietämystä. (Mitchell, 1997, s.1,14) Toinen koneoppimisen vahvuus on se, että se tarjoaa yleensä järjestelmille mahdollisuuden oppia ja parantaa kokemuksesta automaattisesti ilman, että niitä ohjelmoidaan erikseen (Sarker et al. 2021).

Koneoppiminen, syväoppiminen ja data-analyysi ovat tekoälyn osa-alueita, jotka osittain kuuluvat käsitteinä toisiinsa. **Koneoppiminen** on tietokoneen kyky oppia datasta ja parantaa sen avulla suorituskkykyä. Työkaluna se soveltuu luokittelemaan ja ennustamaan toimien lopputulosta. (Bishop, 2006) **Syväoppiminen** on runsaalla datalla opetettua neuroverkon hyödyntämistä, jossa eri tehtävien suorittaminen tapahtuu monissa kerroksissa (Lee et al. 2018). Data-analyysillä jalostetaan dataa tiedoksi, ja edelleen hyödynnetään tietoa johtopäätöksissä (Runkler 2012, s. 2).

3.2 Vastuu koneoppimisessa

Kuten eettisyys, myös vastuu on haastava käsite koneoppimisessa, sillä koneoppimisen kehittämiseen liittyy yleensä monia tekijöitä. Tietokonetta ei voi suoraan syyttää esimerkiksi virheistä, sillä se ei ole moraalinen toimija. (Nissenbaum 1996) Datavetoisten algoritmien, kuten koneoppiminen, valvontaan ja arvioimiseen tarvitaan mekanismeja, joilla luodaan vastuuvellisuutta algoritmien tekijöille ja käyttäjille haitallisia seurauksia vastaan (Cooper et al. 2022).

Koneoppimisella on useita hyötyjä, kuten tiedon tehokas etsintä äänen, tekstin ja kuvien kautta, paperityön käsittelyn automatisoituminen, sähköpostien suodatus ja vilpillisten tapahtumien löytäminen kuten väärät luottokorttitapahtumat, mutta myös puutteita muun muassa tautien vakavuustuloksissa tai keinotekoisissa pisteytysjärjestelmissä kuten lainojen hyväksymisessä (Kim & Doshi-Velez 2021). Koneoppimiseen kuuluu monia vastuita, kuten tutkittavuus, oikeudenmukaisuus, turvallisuus ja yksityisyys (Borrellas & Unceta 2021). Kim & Doshi-Velez (2021) määrittelevät koneoppimisen vastuuvellisuuden siten, että tekoälyjärjestelmän tulee käyttäytyä luvatussa tavalla. Koneoppiminen on matemaattisia malleja eli algoritmeja, ja niiden kehitystä. Algoritmit voivat olla monimutkaisia ja tällöin niiden tulkitseminen voi olla vaikeaa. Vaikka algoritmi toimisi paremmin monimutkaisena, niihin liittyy haasteita. Voi olla haastavaa tarkistaa, vahvistaa tai selittää tulokset, varmistaa niiden oikeellisuus, turvallisuus sekä varmistaa käytetyn datan yksityisyys. (Borrellas & Unceta, 2021)

Organisaatioiden ja julkisten laitosten tulisi olla vastuussa koneoppimisen mallien toiminnasta ja seurauksista. Vastuullisuus algoritmista tarkoittaa, että algoritmin tai mallin suorituskyvyn maksimoinnin lisäksi asianomaisten tulisi saada ymmärrettävät selitykset algoritmin tai mallin toiminnalle, reiluiluudelle, turvallisuudelle ja yksityisyydelle. Vastuullisuus tuo koneoppimisen hyödyntämiselle rajoituksia, kunnes esillä oleviin haasteisiin on tarpeeksi ratkaisuja. (Borrellas & Unceta, 2021) Datavetoiset algoritmiset järjestelmät voivat lisätä vastuullisuuden esteitä. Ylläpitostandardit, oikeus ja politiikka, ja tietotekniikka ovat oleellisia osia teknologisessa vastuussa. (Cooper et al. 2022)

Käyttäjän oikeus selitettävyyteen voi tuoda haasteita organisaatiossa, sillä koneoppimisen tekemät päätökset tulisi olla perusteltavissa, jolloin monimutkaisissa malleissa selitettävyyden kustannukset kasvavat ja tällöin kustannukset saattavat ylittää saadut hyödyt. Tämä voi myös hidastaa koneoppimisen innovaatioita, sillä liian kalliita malleja ei haluta käyttää. Toinen haaste on syy-yhteyden todistaminen. Jos koneoppimismallin vahingosta, kuten mallin viallisuudesta tai puolueellisuudesta, ei ole selvyyttä, kuinka vas-

tuu jakaantuu. Tällaisissa tilanteissa vahingonkorvaus on epäselvä, ja voi vaikuttaa käyttäjäkokemukseen negatiivisesti, jolloin koneoppimismallin turvallisuus ja oikeudenmukaisuus tai niiden puute korostuu. (Borrellas & Unceta, 2021)

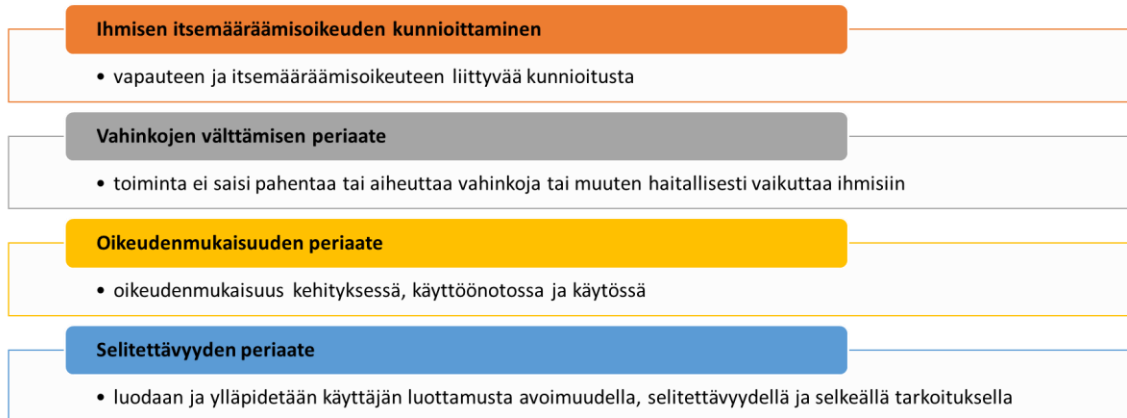
Kim & Doshi-Velez (2021) jakavat koneoppimisen avoimuuden kahteen osa-alueeseen: prosessin avoimuuteen eli tiedon kulun, tiedonkäsittelyn, mallinnusvalinnan ja koulutuksen laatuun ja ohjelmiston avoimuuteen eli julkaisukoodiin. He käsittelevät artikkelissaan avoimuuden lisäksi tulkittavia malleja, post-hoc -tarkistuksia, empiiristä suorituskkyä ja suunnittelun takaamia ominaisuuksia, joilla pystytään eri tilanteissa mittaamaan erilaisia koneoppimisen malleja ja algoritmeja. Koska järjestelmiä, niissä olevia algoritmeja ja niitä mittaavia algoritmityyppjeä on monenlaisia, erilaiset järjestelmät saavat määrittämään arvot aina uudelleen tapauskohtaisesti. Esimerkiksi Richens et al. (2020) selittävät lääketieteellisen diagnosoinnin algoritmin tarkastuksen kerroksellisuutta siten, että ensin tarkastellaan algoritmin tulosten todennäköisyyttä, sitten riittävyttä ja lopuksi testamista. Siten vastuu voi tarkoittaa eri aihe yhteyksissä eri asioita. (Kim & Doshi-Velez 2021)

3.3 Koneoppimisen eettisyys

Eettisiä asioita ovat pohjimmiltaan heijastukset ihmisten toimintaan. Kysymykset, joissa mietitään, miten tulisi toimia ja tehdä tai mitä muut ihmiset odottavat ja mitkä ovat toimintatapojen rajat, ovat eettisiä pohdintoja. Eettisyyttä käsitellessä konseptit oikeasta ja väärästä, oikeudenmukaisuudesta ja epäoikeudenmukaisuudesta, ja hyvästä ja pahasta nousevat esiin. Tosin tietokonetekniikan kehitys on laajentanut eettisyyden käsitystä esimerkiksi tieteen etiikkaan, jossa ei suoraan tarkastella enää pelkästään ihmisen toimia. (Roth 2016, s. x-xi)

Tekoälyn (AI) etiikka seuraa teknologian muutosta, ja sen vaikutusta yksilöön ja yhteiskunnan ja talouden muutoksiin. AI-etiikka tarkastelee tekoälyn etuja verrattuna eettisiin huolenaiheisiin. Hyödyntäessä koneoppimista ja sen eettisyyttä tarkastellessa pohditaan, kenelle kuuluu tekoälyn tekemien päätöksiin vastuu ja kuinka päätöksiin määrä ja laajuus määritellään. Ongelmana voi olla ohjelmiston puolueellisuus tai epäoikeudenmukainen syrjintä. Jos tekoälyn hyödyntämistä pitää parantaa, on mietittävä, kenelle parannus tehdään, esimerkiksi tehdäänkö parannukset asiakkaan vai palveluntarjoajan tarpeisiin. (Coeckelbergh 2020) Kuvassa 4 on tutkimuksen näkökulmana toimivat Euroopan komission eettiset ohjeet koskien eettisten periaatteiden ja arvojen noudattamista.

Tekoälyä koskevat eettiset periaatteet ja arvot



Kuva 4. Tutkimuksen neljä eettisyyden näkökulmaa (AI HLEG 2019, s. 14–16)

Ihmisen itsemääräämisoikeuden kunnioittamisella tarkoitetaan vapauten ja itsemääräämisoikeuteen liittyvää kunnioitusta. Toisin sanoen perusteeton ihmisten holhoaminen, harhaan johtaminen, alistaminen, pakottaminen, ehdollistaminen tai manipulointi ei saisi toteutua. **Vahinkojen välttämisen periaate** tarkoittaa sitä, ettei toiminta saisi pahentaa tai aiheuttaa vahinkoja tai muuten haitallisesti vaikuttaa ihmisiin eli sen tulisi suojella ihmisarvon koskemattomuutta. **Oikeudenmukaisuuden periaatteella** tarkoitetaan oikeudenmukaisuutta kehittämisessä, käyttöönotossa ja käytössä. Valinnanvapautta ei saisi rajoittaa, eikä käyttäjää johtaa harhaan. Tasapaino kilpailevien etujen sekä tavoitteiden välillä tulisi harkita huolellisesti. Päätöksentekoprosessit tulisi olla selitettäviä ja päätöksestä vastaavan olla tunnistettavissa. **Selitettävyyden periaate** tarkoittaa sitä, että luodaan ja ylläpidetään käyttäjän luottamusta. Prosessien tulisi olla avoimia, tarkoituksen selkeä ja ne pitäisi pystyä selittämään kaikille vaikutuksenalaisena oleville. (AI HLEG 2019, s. 14–16)

4. KONEOPPIMINEN TERVEYDENHUOLLOSSA

4.1 Koneoppimisen nykytila terveydenhuollossa

Tekoälystä puhuttiin jo vuonna 1950, mutta monet rajoitteet estivät sen hyväksynnän ja soveltamisen lääketieteessä. Koneoppimisen ja syväoppimisen myötä vasta 2000-luvulla aikaisempiin rajoitteisiin löydettiin ratkaisuja ja tekoälyä aloitettiin hyödyntämään enemmän ja monipuolisemmin terveydenhuollossa. (Kaul et al. 2020) Tekoäly on nopeampaa, puolueettomampaa, melko tarkkaa ja luotettavaa verrattuna perinteisiin menetelmiin monissa tilanteissa (Hedge & Shenoy 2021). Nykyään tekoälyä on mahdollista soveltaa riskien arviointiin, diagnosointiin ja niiden tarkkuuden parantamiseen ja asioiden käsittelyn tehostamiseen. Tekoäly terveydenhuollossa voi parantaa diagnostista tarkkuutta, palveluntarjoajan työnkulun ja kliinisten toimintojen tehokkuutta, helpottaa sairauksien ja terapeuttisten toimien seuraamista ja parantaa menettelytarkkuutta ja potilaan yleisien tuloksien tarkkuutta. (Kaul et al. 2020) Lääketieteen tekoälyn (Artificial intelligence in medicine, AIM) toteuttamisen tunnetuimmat alueet ovat taudin puhkeamisen riskin tunnistaminen, hoidon onnistumisen tai tehokkuuden arviointi, hoidon komplikaatioiden hallinta tai lieventäminen, potilaan jatkuvan hoidon avustaminen sekä kliininen tutkimus ja lääkekehitys (Becker 2019).

Tekoälyä sovelletaan terveydenhuollossa, koska se esimerkiksi helpottaa sairauksien diagnosointia ja hoitoprosesseja, auttaa tekemään lääkekehitystä, säästämään kustannuksia ja parantaa työ- ja asiakastytyvyyttä. Hyödyntämällä digitaalisia järjestelmiä, terveysdataa ja tekoälyjärjestelmiä tavoitellaan hyvinvoinnin lisäämistä ja tautien ennaltaehkäisemistä. Esimerkkejä tekoälyn hyödyntämisestä on hoito ja tutkimus keuhkosairauksissa, kardiologia, unihäiriöiden hoito ja onkologia. (Neittaanmäki et al. 2019 s. 24) Syväoppimisen hyödyt lääketieteessä ovat lääketieteellisen kuvantamisen tarkkuuden parantaminen, johdonmukaisuus ja tehokkuus lääketieteessä. (Kaul et al. 2020)

Terveydenhuollossa käytettynä ammattilaisella on velvollisuus kertoa potilaalle tekoälyn toimintaperiaatteet, edut ja rajoitteet. Lääketieteen tekoälyssä esiintyy monia etuja. Esimerkiksi suuren tietojoukon tutkimisen tai rajaamisen helpottaminen, puheentunnistuksen avulla tiedon muuttaminen digitaaliseen muotoon ja sen tallentaminen sekä tiedonhaun tekeminen tehokkaasti. Sen lisäksi AI-robotit mahdollistavat ainutlaatuinen tarkkuuden leikkauksissa ja VR-mallinnus voi auttaa mielenterveyspotilaita. Internet of Things (IoT)-mallin mahdollistama etäterveysjärjestelmien kehittäminen, terveyssovellukset hy-

vinvoinnin seurantaan sekä mahdollisuus vähentää inhimillisiä virheitä vaikuttamalla oikealla tavalla algoritmin koulutukseen ja suunnitteluun ovat tuoneet esille tekoälyn hyödyllisyyttä. Lisäksi mahdollinen ajan säästäminen päätöksentekotilanteissa, tai mahdollisuus hyödyntää koneoppimista kuvantamisessa ammattilaisten apuna. Koneoppimisen järjestelmät voivat myös työskennellä ympäri vuorokauden, ja siten ovat tärkeä resurssi. (Hedge & Shenoy 2021)

Tekoälyalgoritmit tarvitsevat vielä lisätutkimusta ja hyväksyntää terveydenhuollon alalla. Lisätiedot algoritmien tehokkuudesta, arvosta ja vaikutuksesta potilaan hoitoon ja sen lopputulokseen, mallien kustannustehokkuudesta ammattilaisille ja käytännön kehittämisestä ovat tärkeitä tekoälyn päivittäisessä käytössä. (Kaul et al. 2020) Tekoölyyn liittyvällä tekniikalla, jota hoitotoimenpiteisiin osittain tai kokonaan käytetään voi olla sekä etuja että haittoja potilaiden etujen näkökulmasta tai lääkäreiden ja potilaiden välisissä huolenaiheissa. (Arnold 2021) Käyttöönoton korkeat alkukustannukset voivat olla haaste koneoppimisen hyödyntämiseen terveydenhuollossa, sillä hyödyntäminen vaatii monitieteistä keskustelua. Olennaisena haasteena tekoälyn hyödyntämisessä terveydenhuollossa on ammattilaisten pelko työpaikkojensa menettämisestä. Tämä voi pahimmillaan johtaa terveydenhuollon tekoälyn hyväksymättömyyteen tai haluttomuuteen käyttää sitä. (Hedge & Shenoy 2021) Haasteita voivat olla puolueellisuuden arvioiminen tai oikeudellinen suhde, AIM:n edistymisen ja toimeenpanon edellyttämä jatkuva keskustelu potilashoidossa, ja tarve lääkäreiden aktiiviseen osallistumiseen keskusteluihin. Jotta potilaan hoito hyötyisi AIM:sta, lääkäreillä on vastuu oppia uusia mahdollisuuksia siitä. Haasteena on myös lääkärin ja potilaan välillä olevan suhteen ja hoidon mahdolliset seuraukset ja kuinka lääkärin pitää ottaa ne huomioon. Lääkärien on osallistuttava moraalisisella kekseliäisyydellä vastaamaan kysymyksiin, jotka koskevat oikeudenmukaisuutta, puolueellisuutta ja hyödyllisyyttä ja siihen, kuinka mahdollisesti tekoälyn integrointi terveydenhuoltoon voi potentiaalisesti häiritä potilaan ja lääkärin välistä lopputulemaa ja kuinka välttyttäisiin vahingon tuottamiselta. (Arnold 2021)

Silti on ennustettu, että esimerkiksi seuraavan 30–50 vuoden aikana tekoälyn kehitys nanorobottien avulla edesauttaa syövän geenien manipuloinnissa. Nanorobottien avulla voidaan monitoroida ja valvoa elimiä jatkuvasti. Myös geeniterapia tulee vaikuttamaan lääketieteeseen ja terveydenhuoltoon siten, että se vähentää sairauksien tuomia haittoja ja määrää tai jopa eliminoi ne kokonaan. (Neittaanmäki et al. 2019, s. 121)

4.2 Data ja koneoppiminen terveydenhuollossa

Sen lisäksi, että koneoppiminen tarvitsee laajan laskentatehon, se tarvitsee myös saatavilla olevaa tietoa eli dataa, jota algoritmissa hyödynnetään (Brynjolfsson & McAfee

2017). Algoritmin kehittämisessä hyödynnetään harjoitus- ja testausdataa, jotka tulee kerätä ja valmistella algoritmia varten (Chandramouli et al. 2018, 1.4.1.1, 1.4.1.2 & 1.4.2). Valvotussa koneoppimisessa tulos voi olla vain yhtä hyvä kuin sen opettamiseen käytetty harjoitusdata eli jos harjoitusdata on heikkolaatuista, tällöin algoritmin ennuste ei ole tarkka ja koneoppimisen hyödyntäminen ei ole tehokasta (Chandramouli et al. 2018, 1.5.1 & 1.6). Wichmann et al. (2020) kertovat, että lääketieteellisen koneoppimisen testausaineiston laatu on jopa tärkeämpi kuin koulutusaineiston laatu, koska sitä käytetään suorituskyvyn testaukseen ja viranomaishyväksyntään.

Pohjana lääketieteellisen tekoälyn kehitykselle on digitalisaatio, kliiniset tietokannat ja potilastietojärjestelmät (Kaul et al. 2020). Hedge & Shenoy (2021) huomauttavat, että koneoppimisjärjestelmien tehokkuus lääketieteessä riippuu syötetyn datan määrästä ja laadusta, tekijänoikeussäätelystä, tekniikan luotettavuudesta, kyberturvallisuudesta ja tietosuojasta. Wichmann et al. (2020) toteavat, että datan tarkkuuden tarkastelussa olisi hyvä käydä läpi kaikki aineisto silmämääräisesti tai hyödyntää kliinisten asiantuntijoiden merkitsemiä tietoja tiettyä hanketta varten. Tekoäly vaatii yleensä suuren tietojoukon algoritmin kehittämiseen ja kouluttamiseen (Mesko & Görög 2020; Hedge & Shenoy 2021) ja jos tietomalli koneoppimisessa on virheellistä, tulokset ovat huonoja (Hedge & Shenoy 2021). Tosin datan valmistelussa algoritmia varten tulee ottaa huomioon, että jos dataa valikoi ja rajoittaa liikaa jo algoritmin koulutusvaiheessa, se voi korostaa haastetta ennakoasenteista ja puolueellisuudesta. Monipuolisen harjoitusdatan käyttäminen useilta eri maantieteellisiltä alueilta olisi ihanteellista, mutta sitä ei usein tehdä kliinisessä tekoälyn tutkimuksessa rajoitetun saatavuuden vuoksi. (Wichmann et al. 2020)

Dataan lääketieteellisessä koneoppimisessa voi liittyä monenlaisia haasteita. Sen lisäksi, että Mesko & Görög (2020) kertovat uhasta tietojen tai tekniikan väärinkäytöstä potilaiden toimesta, haasteena voi olla myös se, kuka todella omistaa datan ja kuinka potilastietoja todella käytetään (Morrisey 2014). Lääketieteellisen koneoppimisalgoritmin kehittämisessä datan avulla voidaan, riippuen kliinisestä tilanteesta, josta tiedot dataan hankitaan, tarvita yksittäisten potilaiden suostumus. Kuten kliinisessäkin tutkimuksessa, olemassa olevien tietojen arviointi vaatii yleensä retrospektiivisen tutkimuksen suunnittelun, joka tarvitsee usein suostumuksen paikalliselta eettiseltä komitealta. Esimerkiksi arvioidessa kliinisten tutkimuksien kuvantamisdataa, tarvitaan kirjallinen tietoinen suostumus sekä hyväksyntä tutkijoilta tiedonvaihtoon, joka voi olla terveydenhuollossa haastavaa. (Wichmann et al. 2020) Myös esimerkiksi hakkerit ovat pystyneet manipuloimaan kuvantamisdataa siten, että taudin syyt, synty, vaikutus ja kehitys muuttuvat eikä ammattilaiset ole havainneet niiden lisäämistä tai poistamista (Mirsky et al. 2019).

Wichmann et al. (2020) ehdottavat mahdollisia ratkaisuja haasteisiin liittyen koneoppimisen hyödyntämiseen ja kehittämiseen terveydenhuollon alalla. Datan rajoituksiin ja saatavuuden haasteisiin voisi auttaa yhdistetty oppiminen. Tällöin koneoppimisalgoritmin koulutus tapahtuisi useassa paikassa ja dataa ei tarvitsisi siirtää. Ohjelmisto asennettaisiin paikalliselle palvelimelle, jolloin hallinta eri ammattilaisryhmien toimesta olisi helpompaa ja data ei poistuisi terveydenhuollon laitoksen verkosta. (Wichmann et al. 2020)

4.3 Diagnostiikka ja koneoppiminen

Lääketieteellisessä diagnoosissa pyritään määrittämään sairaudet, jotka selittävät potilaan oireet. Diagnoosin suorittaa terveydenhuollon henkilö, kuten lääkäri. Kun diagnosointiin käytetään koneoppimista, sen perustoimintaperiaate on, että se tunnistaa sairauksia, jotka korreloivat voimakkaasti potilaan oireiden kanssa. (Richens et al. 2020) Terveydenhuollon diagnostiikassa on määriteltynä kolme keskeistä tavoitetta: nopeus, tarkkuus ja kustannustehokkuus (Chi et al. 2020). Tekoälyllä, koneoppimisella ja syväoppimisella on potentiaali parantaa diagnostiikkaa terveydenhuollossa, mutta ne aiheuttavat myös huolta väärinkäytöksistä ja niiden seurauksista esimerkiksi terveydenhuollon sovelluksiin. Tekoäly tai koneoppiminen eivät suoraan ole vastaus itse sairauksien hoitoon, mutta tehokas työkalu erilaisten tietojen yhdistämiseen ja siten diagnosoinnin auttamiseen ja hoidon hallintaan. (Hunter 2019) Sitä käytetäänkin ammattilaisen tukena, sillä tekoäly antaa monipuolista tukea terveydenhuollon ammattilaisille päätöksentekoon ja diagnosointiin (Neittaanmäki et al. 2019). Koneoppimista voidaan hyödyntää diagnostiikassa useammalla eri tavalla, esimerkiksi diagnoositestihin, veren kuvan analysointiin sekä radiologiassa ja kuvantamisessa erilaisten rakenteiden havaitsemiseen diagnoosin saamiseksi (Wichmann et al. 2020; Hunter 2019). Wichmann et al. (2020) huomauttavat, että kuvantamisessa algoritmien kouluttamisen tärkeät näkökulmat ovat tiedonkeruussa ja laadunvalvonnassa.

Richens et al. (2020) yrittävät tunnistaa epäoptimaalisten tai vaarallisten diagnoosien mahdollisuutta eli he tutkivat vaihtoehtoisia algoritmeja ja vertaavat niitä lääkäreiden suorittamiin diagnooseihin. Koneoppimisen soveltamisessa lääketieteellisessä diagnoosissa on tärkeää ottaa huomioon syyperäinen päättely. Myös liian tarkat diagnoosit voivat olla haaste. Väärän diagnoosin riski voi aiheuttaa vakavaa vahinkoa potilaalle ja koneoppimista käytetään monimutkaisten ongelmien päätöksenteon apuna ja tukena. (Richens et al. 2020) Hunter (2019) kertoo, että haasteena voi olla herkkyyden ja tarkkuuden suhde, esimerkiksi Silicon Valley niminen yritys on kehittänyt verestä tehtävän testin mahdolliselle peräsuolen syöväälle, jonka tavoitteena on suurempi tarkkuus. Testi perustuu siihen, että se tunnistaa suurella herkkyydellä veren analyyttejä. Suuri herkkyys

voi johtaa tarkkuuden heikentymiseen lisäämällä mahdollisia vääriä positiivisia tuloksia. Koneoppimisen algoritmi käy läpi analyytit, DNA:n, RNA:n ja proteiinit, jotta se löytäisi erilaisista yhdistelmistä tarkimman testin. Vaikka koneoppimisen myötä ykköstason syövän tunnistamisen herkkyys on noussut 30–40 prosentista jopa 83 prosenttiin, niin diagnosoinnin haasteena on silti se, että jokainen syöpä on erilainen kehittyen ja vastaten hoitoihin eri tavalla. (Hunter 2019) Wichmann et al. (2020) kertovat riskeistä kuvantamisen kerroksellisuudesta, sillä ensimmäinen kuvantaminen ei välttämättä johda diagnosoinnissa suoraan ratkaisuun, vaan kuvantamista tulisi jatkaa myös jollain toisella työkalulla. Tämä taas johtaa haasteeseen, kuinka koneoppimisalgoritmeihin saa mukaan tämän kaltaisen porrastetun diagnosoinnin.

Richens et al. (2020) ehdottavat erilaisia periaatteita algoritmin suunnittelulle, jotta mahdolliset sekaannukset vähenisivät, sillä hyvä diagnoosi tarvitsee johdonmukaisuutta, syy-yhteyttä ja yksinkertaisuutta. Koneoppimista käytetään myös mikroskopiassa, jotta välttäisiin monimutkaisten kalliiden korkean resoluution mikroskooppien käytöltä. Koneoppiminen mahdollistaa elävien biologisten prosessien havainnoinnin reaaliajassa, vaikka optiset tekniikat eivät pysty vastaamaan elektronimikroskopian resoluutioita. Tällöin koneoppimistekniikat luottavat siihen, että matalan resoluution kuvat sisältävät riittävästi tietoa tuottamaan niistä korkeamman resoluution kuvia. Oleellista on laskea toiminnolle virhe, jotta koneoppimisennuste saadaan mahdollisimman lähelle todellista kuvaa. (Hunter 2019)

Johdonmukaisuutta tutkitaan todennäköisyydellä erilaiseen lopputulokseen. Syy-yhteyttä tutkitaan riittävydellä Tämä tarkoittaa, että verrataan algoritmia rakenteelliseen syyperäiseen malliin diagnosoille, joka on luotu ammattilaisten avustuksella. (Richens et al. 2020) Tässäkin on silti merkittävä petosten ja väärinkäytösten riski, kuten Hunter (2019) toteaa, tekoäly ja koneoppiminen vaikuttavat radiologisessa diagnosoinnissa löytämällä kuvioita ja rakenteita kuvista, jotka voisivat viitata kasvaimen, mutta esimerkiksi algoritmin tahallinen kouluttaminen huonolla datalla voi lisätä merkittävästi väärin diagnosoien riskiä.

Algoritmeja testataan vertaamalla odotettua lääketieteellistä haittaa ja algoritmin riittävyttä. Algoritmeja ja lääkäriyhmää verrataan erilaisissa kliinisissä tapauksissa antamalla niille erilaisia diagnostiikkatehtäviä. Lääkäriyhmä luo erilaisia lääketieteellisiä tapauksia ja puolueellisuus pyritään poistamaan useamman lääkärin tarkastuksella. Tällöin saadaan mahdollisimman realistinen diagnoositapaus. (Richens et al. 2020) Myös Hunter (2019) toteaa, että koneoppiminen saattaa tunnistaa helpommin kuvantamisessa

yksityiskohtia, joita ihminen ei tarkastelussaan huomaa ja diagnostiikkaa parantaa lisätyn todellisuuden tekniikka, joka yhdistää todellisiin kuviin digitaalisesti koneoppimisen avulla luodut visuaaliset kuvat.

Richens et al. (2020) varmistavat diagnostiikan oikeellisuutta vertailulla. Vertailua varten lääkärit ja algoritmit määrittelevät omat luettelot mahdollisista taudeista. Eri algoritmeihin syötetään identtiset mallit oireista, jolloin algoritmit ovat vertailtavissa toisiinsa. (Richens et al. 2020) Oikeellisuudessa on haasteena algoritmin kehitysvaiheessa lääketieteen ammattilaisen ja tietojenkäsittelytieteilijöiden erilaisuus, mikä voi johtaa väärin koulutettuun algoritmiin. Tätä haastetta ehdotetaan ratkaistavaksi eri asiantuntijaryhmien osallisuudella, kuten insinöörien, joilla on tekninen osaaminen, lääketieteen opiskelijoiden tai radiologien yhteisellä algoritmien testaamisella, jolloin se voisi edesauttaa kehitystä. Ratkaisu on edelleen epätodennäköinen, sillä se edellyttää eri toimijoiden taustojen tarkastamista turvallisen toiminnan takaamiseksi. Poikkeamia algoritmissa voi esiintyä myös sen eri kehitysvaiheissa, sillä siihen vaikuttavat mantereet, populaation vaihtelu ja tautiesiintyvyyksien eroavaisuudet. (Wichmann et al. 2020)

Wichmann et al. (2020) toteavat, että monikerroksinen turvallisuus on ratkaisevan tärkeää tekoälyn luotettavassa toteuttamisessa kliinisessä työssä. Seuraava vaihe diagnostisissa sovelluksissa on ennustavien tekoälysovellusten kehittäminen, jotka ovat monimutkaisempia ja siten tarvitsevat enemmän tietoa monimutkaisten suhteiden löytämiseksi (Wichmann et al. 2020). Yksi koneoppimisen heikkouksissa lääketieteessä on se, että se on looginen ja ohjelmoitu eli koneoppimisella ei ole emotionaalista yhteyttä, ja tällöin algoritmi ei voi rikkoa sääntöjä esimerkiksi hengen pelastamiseksi (Mesko & Görög 2020). Yksinkertaisissa tapauksissa lääkäreille on tyyppisempää saavuttaa korkeammat tarkkuudet diagnosoinnissa. Haastavammissa tapauksissa algoritmilla on tyyppisempää saavuttaa korkeammat tarkkuudet diagnosoinnissa. Koneoppimisen hyödyntämisessä diagnostiikassa ja sen haasteissa on tärkeää ymmärtää kuinka algoritmi tuottaa ja suorittaa diagnoosin ja kuinka selkeästi se määrittää algoritmin halutun tuloksen. Todennäköisempää on, että lääkärin ja algoritmin yhdistetty diagnoosi on merkittävämpää kuin kumpikaan yksin. (Richens et al. 2020)

4.4 Eettisyys ja koneoppiminen terveydenhuollossa

Tähän tutkimuksen kohtaan kootaan teoriaosuudesta asioita, jotka voidaan yhdistää tekoälyä koskeviin eettisiin periaatteisiin ja arvoihin. Eettiset periaatteet ja arvot käydään yksi kerrallaan läpi. Joitain teoriaosuuden kohtia voi soveltaa useampaan periaatteeseen

tai arvoon. Käsiteltävät tekoälyn eettiset periaatteet ja arvot ovat ihmisen itsemääräämisoikeuden kunnioittaminen, vahinkojen välttämisen periaate, oikeudenmukaisuuden periaate ja selitettävyyden periaate (AI HLEG 2019, s. 14–16).

Ensin pohditaan ihmisen itsemääräämisoikeuden kunnioittamisen periaatteen toteutumista koneoppimisen käyttöönotossa ja hyödyntämisessä terveydenhuollon diagnostiikassa ammattilaisen tukena. Ammattilaisen velvollisuus kertoa potilaalle toimintaperiaatteista (Hedge & Shenoy 2021) edesauttaa harhaan johtamisen välttämistä. Lisätiedoilla arvosta ja vaikutuksesta hoidon lopputulokseen (Enslin et al. 2020) vaikutetaan siihen, että potilas saa kattavasti tietoa diagnosoinnin vaikutuksista, ja siten harhaan johtamisen ja manipuloinnin toteutumisen riski pienenee. Ammattilaisten hyvällä kekseliäisyydellä algoritmien selittämiseen (Arnold 2021) turvataan sekä ammattilaisen, että potilaan itsemääräämisoikeutta.

Kun datan omistamisen ja sen käyttötarkoituksen tuntemus selkenee, myös tarve potilaan suostumukselle täsmentyy (Morrissey 2014; Wichmann et al. 2020). Jos datan omistaminen ja käyttötarkoitus ovat epäselviä, potilaan itsemääräämisoikeus voi heiketä. Algoritmin tarkastaminen useamman lääkärin toimesta (Richens et al. 2020) voi vähentää harhaan johtamista tai manipulointia. Joskus koneoppiminen voi vaikuttaa myös negatiivisesti diagnostiin, kuten vakavuustuloksissa (Kim & Doshi-Velez 2021), jolloin riski harhaan johtamiseen voi kasvaa. Järjestelmän toiminnan varmistaminen (Kim & Doshi-Velez 2021) ehkäisee sekä ammattilaiseen, että potilaaseen kohdistuvaa harhaan johtamisen ja ehdollistamisen riskiä. Algoritmin tai mallin toiminnan, reiluuden, turvallisuuden ja yksityisyyden ymmärtäminen (Borrellas & Unceta 2021) lisää algoritmin käyttäjien itsemääräämisoikeutta. Ihmisen itsemääräämisoikeuden ja vapauden kunnioittamiseen liittyy teoriassa taulukkoon 2. listatut tekijät, jotka on jaettu kolmeen eri teemaan.

Taulukko 2. Ihmisen itsemääräämisoikeuden kunnioittamiseen liittyvät teemat.

Ihmisen itsemääräämisoikeuden kunnioittaminen	
Teema	Aineiston ihmisen itsemääräämisoikeuden kunnioittamiseen liittyviä tekijöitä
Ammattilaisen rooli koneoppimisen hyödyntämisessä ja käyttöönotossa	<ul style="list-style-type: none"> ➤ Ammattilaisen velvollisuus kertoa potilaalle koneoppimisen toimintaperiaatteet, edut ja rajoitteet ➤ Ammattilaisen tiedot arvosta ja vaikutuksesta potilaan hoitoon ja sen lopputulokseen ➤ Ammattilaisten osallistuminen koneoppimisalgoritmiin liittyviin kysymyksiin vastaamiseen
Asiakkaan tietoisuus datan käytöstä	<ul style="list-style-type: none"> ➤ Datan omistaminen ja käyttötarkoitus ➤ Potilaan suostumus datan hyödyntämiseen
Koneoppimisen luotettavuus	<ul style="list-style-type: none"> ➤ Terveydenhuollon ammattilaisen hyödyntäminen datan tarkastuksessa ➤ Algoritmin puutteet tautien vakavuustuloksissa ja lupaus järjestelmän toimivuudesta ➤ Algoritmin tai mallin toiminnan, reiluuden, turvallisuuden ja yksityisyyden tiedostaminen

Ihmisen itsemääräämisoikeuden periaatteen kolme keskeistä tutkimusaineistossa esiintyvää teemaa ovat ammattilaisen rooli koneoppimisen hyödyntämisessä ja käyttöönotossa, asiakkaan tietoisuus datan käytöstä ja koneoppimisen luotettavuus. Ammattilaisella on selkeä rooli koneoppimisen hyödyntämisessä ja käyttöönotossa eettisyyden, ja tarkemmin ihmisen itsemääräämisoikeuden näkökulmasta. Toiminnallaan ja omien työkalujen tuntemuksella ammattilainen voi vähentää perusteetonta ihmisten holhoamista, harhaan johtamista, alistamista, pakottamista, ehdollistamista tai manipulointia. Datalla on merkittävä osuus koneoppimisalgoritmin käyttöönotossa ja hyödyntämisessä, ja siten asiakkaan eli potilaan tietoisuus itseään koskevan datan käytöstä lisää ihmisen itsemääräämisoikeuden toteutumista. Datan hyödyntäminen luvallisesti koneoppimisalgoritmien kouluttamisessa ja käytössä kunnioittaa ihmisen itsemääräämisoikeutta, sillä dataa ei tulisi hyödyntää ehdollistamalla tai pakottamalla asiakasta siihen. Koneoppimisalgoritmin toimivuuden ja oikeellisuuden varmistamisella ihmisen itsemääräämisoikeuden kunnioittaminen paranee, sillä siten pyritään varmistamaan, että harhaan johtamista tai manipulointia algoritmilla tai algoritmin avulla ei toteutuisi.

Seuraavaksi pohditaan vahinkojen välttämisen periaatteen toteutumista koneoppimisen käyttöönotossa ja hyödyntämisessä terveydenhuollon diagnostiikassa ammattilaisen tukena. Valvontaan ja arviointiin käytettävät mekanismit vähentävät vahinkojen tai muuten haitallisten vaikutusten mahdollisuutta potilaisiin ja ammattilaisiin (Cooper et al. 2022). Datan käsittelyssä liiallinen valikointi tai rajoittaminen (Wichmann et al. 2020) voi heikentää vahinkojen välttämisen periaatetta. Tarkkuuden ja herkkyyden tasapainon tarkastelu on koneoppimisessa tärkeää (Hunter 2019), ja niiden tutkiminen ja algoritmin tasapainon

löytyminen edesauttavat vahinkojen välttämistä. Algoritmin tarkastaminen kehityksessä ja käytössä oikean lopputuloksen varmistamiseksi todennäköisyyden avulla (Richens et al. 2020) auttaa ehkäisemään mahdollisia vahinkoja.

Ammattilaisen tuntemus käyttämistään työkaluista kuten koneoppimisalgoritmista (Arnold 2021), auttaa välttymään vahingon tuottamiselta, sillä se kasvattaa myös potilaan tietämystä. Tekoälyn hyväksymättömyys tai haluttomuus hyödyntää sitä ammattilaisen toimesta (Hedge & Shenoy 2021) voi lisätä riskiä vahingon aiheutumiselle, varsinkin jos työkalun hyödyntäminen edesauttaisi diagnosointia. Porrastettu diagnosointi (Wichmann et al. 2020) voi ehkäistä mahdollisia vahinkoja tai muuten haitallisia vaikutuksia diagnosointiin liittyen. Vahinkoja pyritään ehkäisemään myös sillä, että diagnosointi tapahtuu yhdessä ammattilaisen ja algoritmin toimesta (Richens et al. 2020). Koneoppimisen algoritmin koulutus ja suunnittelu vähentävät mahdollisesti haitallisia virheitä (Hedge & Shenoy 2021). Koneoppiminen voi auttaa oikean diagnoosin löydössä ja tällöin väärän diagnoosin riski voi vähentyä (Richens et al. 2020), ja siten vahinkojen välttämisen periaate toteutuu. Vahinkojen välttämisen periaatteeseen liittyä teoriassa taulukkoon 3. listatut tekijät, jotka on jaettu kolmeen eri teemaan.

Taulukko 3. *Vahinkojen välttämisen periaatteeseen liittyvät teemat.*

Vahinkojen välttämisen periaate	
Teema	Aineiston vahinkojen välttämisen periaatteeseen liittyviä tekijöitä
Koneoppimisalgoritmin vastuullinen kehitys ja hyödyntäminen	<ul style="list-style-type: none"> ➤ Koneoppimisen valvonnan ja arvioinnin mekanismit vastuuvollisuuden parantamiseksi ➤ Datan esikäsittelyn määrä ➤ Tarkkuus vastaan herkkyys ➤ Koneoppimisalgoritmin johdonmukaisuus
Ammattilaisen mahdollisuus vaikuttaa koneoppimisalgoritmin toimintaan	<ul style="list-style-type: none"> ➤ Ammattilaisten osallistuminen koneoppimisalgoritmiin liittyviin kysymyksiin vastaamiseen ➤ Ammattilaisten hyväksymättömyys tai haluttomuus käyttää terveydenhuollon tekoälyä
Koneoppimisen hyödyntäminen ammattilaisen toimesta diagnosoinnin tukena	<ul style="list-style-type: none"> ➤ Porrastettu diagnosointi ➤ Yhdistetty diagnoosi yhdessä ammattilaisen ja algoritmin toimesta ➤ Koneoppimisen hyödyntäminen päätöksenteon tukena ratkaistaessa monimutkaisia ongelmia, tavoitteena vähentää väärän diagnoosin riskiä ➤ Mahdollisuus inhimillisten virheiden vähentämiseen

Vahinkojen välttämisen periaatteen kolme keskeistä tutkimusaineistossa esiintyvää teemaa ovat koneoppimisalgoritmin vastuullinen kehitys ja hyödyntäminen, ammattilaisen mahdollisuus vaikuttaa koneoppimisalgoritmin toimintaan ja koneoppimisen hyödyntäminen ammattilaisen toimesta diagnosoinnin tukena. Vahinkojen välttämistä voidaan edesauttaa koneoppimisalgoritmin kehityksessä laadukkaasti datan hyödyntämisellä, koulutusvaiheessa perehtymällä algoritmin vastauksiin huolellisesti ja hyödyntämisessä

valvomalla ja arvioimalla algoritmia säännöllisesti. Näin haitallista vaikutusta ammattilaiseen ja asiakkaaseen saadaan vähennettyä. Vahinkojen välttämisen periaatetta edesauttaa ammattilaisten halukkuus toimia koneoppimisalgoritmin kanssa ja valmiudet ymmärtää sen toimintaa, jotta ammattilaisten ja koneoppimisalgoritmin yhteinen toiminta ei pahentaisi tilannetta tai aiheuttaisi lisää vahinkoa. Vahinkoja pyritään välttämään sillä, että koneoppimisalgoritmia hyödynnetään ammattilaisen toiminnassa nimenomaan päätöksenteon tukena. Koneoppimisalgoritmin virheitä eli haitallista vaikutusta ihmiseen pyritään välttämään käyttämällä diagnosointiin useita työkaluja. Koneoppimisalgoritmi työkaluna voi vähentää inhimillisen virheen mahdollisuutta. Siten ammattilaisen toiminta yhdistettynä koneoppimisalgoritmin kanssa voivat edesauttaa vahinkojen välttämistä.

Seuraavaksi pohditaan oikeudenmukaisuuden periaatteen toteutumista koneoppimisen hyödyntämisessä terveydenhuollon diagnostiikassa ammattilaisen tukena. Tekoäly voi helpottaa diagnosointia, auttaa säästämään kustannuksia ja parantaa työ- ja asiakastytyväisyyttä sekä säästää aikaa päätöksentekotilanteessa (Neittaanmäki et al. 2019, s. 24), mutta toisaalta korkeat alkukustannukset voivat estää algoritmien hyödyntämistä (Hedge & Shenoy 2021). Oikeudenmukaisuuden periaatteen toteutumista lisää tietoisuus algoritmien virhemahdollisuuksista ja niiden vertaaminen inhimillisten virheiden mahdollisuuksiin (Hedge & Shenoy 2021). Koneoppiminen voi olla tehokas työkalu tietojen yhdistämisessä ja diagnosoinnissa ammattilaisen tukena (Hunter 2019). Se voi lisätä diagnosoinnin prosessin monipuolisuutta (Neittaanmäki et al. 2019), mutta ei vie ammattilaiselta valinnan vapautta. Tällä tavoin koneoppimisen käyttö voi lisätä oikeudenmukaisuutta. Oikeudenmukaisuutta ja algoritmin selitettävyyttä voi heikentää sen kehityksessä hyödynnettävän datan populaation ja tautiesiintyvyyksien eroavaisuuksien vaikutukset (Wichmann et al. 2020) sekä se, että koneoppiminen on loogista ja ohjelmoitua, eikä silloin voi rikkoa sääntöjä esimerkiksi hengen pelastamiseksi (Mesko & Görög 2020).

Sekä ammattilaisen tietoisuus, että tiedon jakaminen potilaalle (Hedge & Shenoy 2021; Arnold 2021) edesauttaa oikeudenmukaisuuden periaatteen toteutumista, sillä harhaan johtamisen riski pienenee ja päätöksentekoprosessin selitettävyyksy kasvaa. Oikeudenmukaisuutta lisätään niin, että hyödynnetään algoritmin antamaa diagnoosia yhdistettyä lääkärin tekemään diagnoosiin (Richens et al. 2020). Oikeudenmukaisuuden periaatteen liittyä teoriassa taulukkoon 4. listatut tekijät, jotka on jaettu kolmeen teemaan.

Taulukko 4. Oikeudenmukaisuuden periaatteeseen liittyvät teemat.

Oikeudenmukaisuuden periaate	
Teema	Aineiston oikeudenmukaisuuden periaatteeseen liittyviä tekijöitä
Koneoppimisen vahvuuksien tunnistaminen ja tiedostaminen	<ul style="list-style-type: none"> ➤ Mahdollisuus helpottaa sairauksien diagnosointia ja hoitoprosesseja, auttaa tekemään lääkekehitystä, säästämään kustannuksia ja parantaa työ- ja asiakastytyväisyyttä ➤ Mahdollisuus vähentää inhimillisiä virheitä riippuen algoritmin koulutuksesta ja suunnittelusta ➤ Tehokas työkalu erilaisten tietojen yhdistämiseen ja siten diagnosoinnin auttamiseen ja hoidon hallintaan ➤ Monipuolinen diagnosointi
Koneoppimisen heikkouksien tunnistaminen ja tiedostaminen	<ul style="list-style-type: none"> ➤ Mahdollinen ajan säästäminen päätöksentekotilanteissa vastaan korkeat alkukustannukset ➤ Mantereiden, populaation vaihtelun ja tautiesiintyvyyksien eroavaisuuksien vaikutus algoritmin kehitykseen ➤ Algoritmi ei voi rikkoa sääntöjä esimerkiksi hengen pelastamiseksi ➤ Yhdistetty diagnoosi, ei koneoppimisalgoritmi yksin
Ammattilaisen velvollisuudet	<ul style="list-style-type: none"> ➤ Ammattilaisen velvollisuus kertoa potilaalle koneoppimisen toimintaperiaatteet, edut ja rajoitteet ➤ Ammattilaisten osallistuminen koneoppimisalgoritmiin liittyviin kysymyksiin vastaamiseen ➤ Koneoppimisen hyödyntäminen avuksi monimutkaisiin ongelmiin päätöksenteon tueksi väärän diagnoosin riskin vähentämiseksi

Oikeudenmukaisuuden periaatteen kolme keskeistä tutkimusaineistossa esiintyvää teemaa ovat koneoppimisen vahvuuksien tunnistaminen ja tiedostaminen, koneoppimisen heikkouksien tunnistaminen ja tiedostaminen ja ammattilaisen velvollisuudet. Koneoppimisen vahvuuksien ja heikkouksien tunnistaminen terveydenhuollon diagnostiikassa kehityksessä, käyttönotossa ja käytössä lisäävät oikeudenmukaisuutta. Niiden tunnistaminen ja tiedostaminen selkeyttää tavoitteiden mahdollisuutta ja siten edesauttaa sen vertaamista kilpaileviin etuihin. Ammattilaisten velvollisuuksien tunnistaminen koneoppimisen kehityksessä, käytössä ja käyttönotossa vahvistaa oikeudenmukaisuuden periaatetta terveydenhuollon diagnostiikassa. Ammattilaisen tuntiessa koneoppimistyökalun mahdollisuudet ja rajoitteet päätöksentekotilanteissa, se lisää selitettävyyttä päätöksentekoprosessissa.

Viimeiseksi pohditaan selitettävyyden periaatteen toteutumista koneoppimisen hyödyntämisessä terveydenhuollon diagnostiikassa ammattilaisen tukena. Koneoppimisen algoritmin selitettävyyden alku on jo kehitysvaiheessa (Chandramouli et al. 2018, 1.5.1 & 1.6), ja siksi kehitykseen käytettävän datan tulisi olla laadukasta ja määrällisesti oikeaa, eikä liian valikoitua ja rajoitettua (Chandramouli et al. 2018, 1.5.1 & 1.6; Hedge & Shenoy 2021; Wichmann et al. 2020). Tällä mahdollistetaan tehokas hyödyntäminen sekä tulosten selitettävyyden. Useamman ammattilaisen, lääkärin, hyödyntäminen esimerkiksi datan tarkastuksessa algoritmin kehityksessä edesauttaa prosessin avoimuutta ja helpottaa tarkastusta (Richens et al. 2020), ja siten myös luo ja ylläpitää luottamusta.

Joskus datan saatavuudessa voi olla haasteita (Wichmann et al. 2020), ja siten algoritmin selitettävyyden voi heikentyä. Virheiden laskeminen tuo algoritmille uskottavuutta ja ylläpitää luottamusta sen toimintaan (Hunter 2019). Varmistamalla algoritmin johdonmukaisuutta ja syy-yhteyttä saadaan luotettavuutta algoritmiin sekä varmistetaan algoritmin tarkoituksen toteutuminen (Richens et al. 2020), mikä vahvistaa selitettävyyden periaatteen toteutumista. Vaikka algoritmin kehityksessä hyödynnetään lääkärin apua, lääkärin ja algoritmin yhdistetty diagnoosi on merkittävämpää kuin kumpikaan yksin (Richens et al. 2020), ja tällöin prosessi on selitettävämpi. Porrastettu diagnosointi, jonka avulla algoritmista saadaan varmuutta ammattilaisen toimintaan ja toisin päin (Wichmann et al. 2020), edesauttaa prosessin selitettävyyttä. Selitettävyyden periaatteeseen liittyen liittyä teoriassa taulukkoon 5. listatut tekijät, joka on jaettu kolmeen teemaan.

Taulukko 5. Selitettävyyden periaatteeseen liittyvät teemat.

Selitettävyyden periaate	
Teema	Aineiston selitettävyyden periaatteeseen liittyviä tekijöitä
Datan laatu algoritmin kehityksessä	<ul style="list-style-type: none"> ➢ Koneoppimisalgoritmin koulutuksen laadukas data ➢ Koneoppimisalgoritmin tehokkuus riippuu datan määrästä ja laadusta, tekijänoikeussäätelystä, tekniikan luotettavuudesta, kyberturvallisuudesta ja tietosuojasta ➢ Datan valmistelu liian tarkkaan ja yksityiskohtaisesti voi korostaa haastetta ennakoasenteista ja puolueellisuudesta ➢ Algoritmin kehitykseen hyödynnettävän datan tarkistus useamman ammattilaisen toimesta
Koneoppimisalgoritmin toimivuuden varmistaminen ja tarkka toiminnan selvitys	<ul style="list-style-type: none"> ➢ Datan rajoituksiin ja saatavuuden haasteisiin yhdistetty oppiminen ➢ Virheen laskeminen ➢ Vertailutyökalu koneoppimisalgoritmien kehittämisessä ja testaamisessa ➢ Johdonmukaisuuden tutkiminen todennäköisyydellä ja syy-yhteyden tutkiminen ennusteen riittävydellä
Koneoppimisalgoritmin hyödyntäminen päätöksenteon tukena	<ul style="list-style-type: none"> ➢ Yhdistetty diagnoosi ➢ Syyperäinen päättely koneoppimisalgoritmissa ➢ Porrastettu diagnosointi

Selitettävyyden periaatteen kolme keskeistä tutkimusaineistossa esiintyvää teemaa ovat datan laatu algoritmin kehityksessä, koneoppimisalgoritmin toimivuuden varmistaminen ja tarkka toiminnan selvitys, ja koneoppimisalgoritmin hyödyntäminen päätöksenteon tukena. Ammattilainen terveydenhuollon diagnostiikassa on koneoppimisalgoritmin käyttäjä. Jos algoritmin koulutusvaiheeseen käytetty data on ammattilaiselle perusteltua ja siten algoritmin toiminta ja tulokset perusteltuja, se voi lisätä ammattilaisen luottamusta koneoppimisalgoritmiin. Tämä lisää myös ammattilaisen mahdollisuutta parantaa selitettävyyttä eteenpäin asiakkaalle. Ammattilaisten tietämyksen hyödyntäminen koneoppimisalgoritmiin käytettävän datan tarkistuksessa auttaa perustelemaan koneoppimisalgoritmin toimivuutta ja siten lisää selitettävyyttä. Virheiden laskeminen diagnosoinnin koneoppimisalgoritmissa lisää tulosten luotettavuutta ja voi parantaa lopputuloksen

oikeellisuutta. Prosessi on tällöin varmempi ja tulosten oikeellisuus paremmin selitettävissä. Koneoppimisalgoritmin koulutus useammalla eri datakokonaisuudella lisää luotavuutta ja tulosten selitettävyyttä. Jos algoritmi on johdonmukainen eli todennäköisyys oikeaan lopputulokseen on suurempi ja ennustetarkkuus on riittävä, tämä auttaa selitettävyyden periaatteen toteutumisessa. Selitettävyyttä ja prosessin avoimuutta edesauttaa eri diagnostiikassa hyödynnettyjen algoritmien vertailu ja niiden vertailu ammattilaisen suorittamaan diagnosointiin. Jos ammattilainen rakentaa päätöksensä koneoppimisalgoritmin avulla, päätökset on silti oltava perusteltavissa ammattilaisen toimesta. Tämä lisää selitettävyyttä. Selitettävyyttä lisää diagnosoinnin porrastaminen ammattilaisen hyödyntäessä eri työkaluja koneoppimisen lisäksi päätöksenteon tukena.

5. YHTEENVETO

5.1 Tulokset

Tutkimuksen kirjallisuuskatsauksen perusteella löytyi useampia teemoja koneoppimisen kehittämisessä ja hyödyntämisessä terveydenhuollossa, joita voidaan tulkita eettisyyden avulla. Teemat esiteltiin kappaleessa 4.4, jonka perusteella tehtiin tutkimuksen päätulosten yhteenveto. Yhteenveto on koottu Kuvaan 5, jossa esiintyy eettisyyteen liittyviä teemoja, joita tukee aiempi koneoppimisen ja diagnostiikan tutkimus. Teemoja voidaan tulkita eettisyyden, ja erityisesti tekoälyä koskevien eettisten periaatteiden ja arvojen näkökulmasta.

Ihmisen itsemääräämisoikeuden kunnioittaminen	Vahinkojen välttämisen periaate
<ul style="list-style-type: none"> ➤ Ammatilaisen rooli koneoppimisen hyödyntämisessä ja käyttöönotossa ➤ Asiakkaan tietoisuus datan käytöstä ➤ Koneoppimisen luotettavuus 	<ul style="list-style-type: none"> ➤ Koneoppimisalgoritmin vastuullinen kehitys ja hyödyntäminen ➤ Ammatilaisen mahdollisuus vaikuttaa koneoppimisalgoritmin toimintaan ➤ Koneoppimisen hyödyntäminen ammatilaisen toimesta diagnosoinnin tukena
Oikeudenmukaisuuden periaate	Selitettävyyden periaate
<ul style="list-style-type: none"> ➤ Koneoppimisen vahvuuksien tunnistaminen ja tiedostaminen ➤ Koneoppimisen heikkouksien tunnistaminen ja tiedostaminen ➤ Ammatilaisen velvollisuudet 	<ul style="list-style-type: none"> ➤ Datan laatu algoritmin kehityksessä ➤ Koneoppimisalgoritmin toimivuuden varmistaminen ja tarkka toiminnan selvitys ➤ Koneoppimisalgoritmin hyödyntäminen päätöksenteon tukena

Kuva 5. Teemat, jotka tukevat tekoälyn eettisten periaatteiden toteutumista kirjallisuuskatsauksen perusteella.

Kuvassa 5 on jaettu teemat neljään osa-alueeseen: ihmisen itsemääräämisoikeuden kunnioittamiseen, vahinkojen välttämisen, oikeudenmukaisuuden ja selitettävyyden periaatteeseen (AI HLEG 2019, s. 14–16). Tulokset perustuvat erilaisiin ihmisen tai algoritmin mahdollisiin toimintatapoihin ja vastuisiin, jotka voivat vaikuttaa eettisyyden osa-alueiden toteutumiseen. Ihmisen itsemääräämisoikeuden kunnioittamista tukevat teemat ovat ammatilaisen rooli koneoppimisen hyödyntämisessä ja käyttöönotossa, asiakkaan tietoisuus datan käytöstä, ja koneoppimisen luotettavuus. Vahinkojen välttämisen periaatetta tukevat teemat ovat koneoppimisalgoritmin vastuullinen kehitys ja hyödyntäminen, ammatilaisen mahdollisuus vaikuttaa koneoppimisalgoritmin toimintaan, ja koneoppimisen hyödyntäminen ammatilaisen toimesta diagnosoinnin tukena. Oikeudenmukaisuuden periaatetta tukevat teemat ovat koneoppimisen vahvuuksien tunnistaminen ja

tiedostaminen, koneoppisen heikkouksien tunnistaminen ja tiedostaminen, ja ammattilaisen velvollisuudet. Selitettävyyden periaatetta tukevat teemat ovat datan laatu algoritmin kehityksessä, koneoppimisalgoritmin toimivuuden varmistaminen ja tarkka toiminnan selvitys, ja koneoppimisalgoritmin hyödyntäminen päätöksen teon tukena. Kuvasta 5 huomataan, että samat teemat voivat tukea eri eettisyyden arvoja ja periaatteita, kun niitä tarkastellaan eri näkökulmista.

5.2 Päätelmät ja tutkimuksen arviointi

Tutkimuksen perusteella eettinen ajattelu näkyy koneoppimisen käyttöönotossa ja hyödyntämisessä terveydenhuollon tutkimuksessa. Aiemmissa tutkimuksissa on tunnistettavissa, kuinka algoritmin ja ihmisen rooli diagnostiikassa ovat molemmat merkittäviä terveydenhuollossa. Ammattilaisella, kuten diagnoosia tekevällä lääkäriä, vastuut, velvollisuudet, ymmärrys ja toimintatavat vaikuttavat eettisyyden toteutumiseen. Ammattilaisen diagnoosin ja koneoppimisalgoritmin diagnoosin yhdistäminen voi edesauttaa virheiden ja riskien välttämistä. Onnistunut syvälinen perehtyminen algoritmiin, sen toiminnan tutkiminen ja toimivuuden varmistaminen erilaisilla tavoilla voivat vaikuttaa positiivisesti eettisyyden periaatteiden ja arvojen toteutumiseen. Potilaan suostumuksien huomioonotto ja datan toimivuuden ja laadun varmistus voivat tukea eettisten periaatteiden toteutumista. Koneoppimisen käyttöönotossa ja hyödyntämisessä terveydenhuollon tutkimuksessa on nähtävissä piirteitä, jotka ovat eettisyydelle ominaisia, kuten mahdollisuuksien ja riskien vertaileminen, ja huonojen ja hyvien puolien tutkiminen.

Työn tavoitteena oli löytää vastaus kysymykseen, kuinka eettinen ajattelu näkyy koneoppimisen käyttöönotossa ja hyödyntämisessä terveydenhuollossa, ja tämä tutkimus vastasi siihen metodologisilla valinnoillaan ja aineiston jäsentämisellä. Aineiston jäsentämisessä auttoivat alatutkimuskysymykset. Aineisto toimi työssä johdonmukaisesti, eikä kerätyssä aineistossa ollut ristiriitaisuuksia, mikä lisäsi tulosten luotettavuutta. Jotkut eettisyyteen liittyvät piirteet olivat selkeästi nähtävillä koneoppimisessa terveydenhuollossa, ja jotkut hieman piilossa ja tulkinnanvaraisia. Jotkut dataan liittyvät piirteet saattoivat olla haastavasti löydettävissä. Datan määrittelyyn ja oikeellisuuteen vaikuttaa se, että koneoppimisalgoritmien koulutusdata on arkaluontoista, eikä data ole avointa terveydenhuollossa. Eettisyyden tarkastelu oli tulkinnanvaraista, sillä eri tutkijat saattavat nähdä eettisyyden käsitteen ja toteutumisen eri tavoin. Tämä vaikuttaa suoraan tutkimuksen toistettavuuteen. Eettisyyden tulkitseminen on usein subjektiivista. Tässä tutkimuksessa subjektiivisuutta pyrittiin poistamaan hyödyntämällä Euroopan komission tekoälyyn liittyviä

eettisiä ohjeistuksia, joiden avulla tutkimukseen saatiin eettisyydelle suoraan määritelmät, joita pystyttiin soveltamaan koneoppimisen tarkastelussa, jolloin työn objektiivisuus kasvoi.

Tämä kirjallisuustutkimus pyrki vastaamaan yleisesti eettisyyden esiintymiseen diagnostiikassa, ja siten jatkotutkimukselle on useita vaihtoehtoja. Tutkimusta voidaan jatkaa esimerkiksi tutkimalla eettisyyttä joillain muulla alalla, tai terveydenhuollossa muulla osalueella kuin diagnostiikassa. Myös eettisyyteen liittyvien teemojen, jotka esiteltiin kuvassa 5, tarkempi jatkotutkimus yksittäin tai ryhminä, kuten ammattilaisen vastuut tai algoritmien kehittämisen eettisyys, voisivat tuoda tutkimukseen edelleen syvällisyyttä. Eettisyyden tarkastelu ja tutkimuksen jatkaminen on yhä tärkeämpää, sillä koneoppimisen toiminta ja mahdollisuudet kehittyvät.

LÄHTEET

- AI HLEG. Tekoälyä käsittelevä korkean tason asiantuntijaryhmä. 2019. Luotettavaa tekoälyä koskevat eettiset ohjeet. Euroopan komissio, Bryssel.
- Aillisto, H., Neuvonen, A., Helaakoski, H., Neuvonen, A. & Seppälä, T. 2018. Tekoälyn kokonaiskuva ja osaamiskartoitus. Valtioneuvoston selvitys- ja tutkimustoiminta. 46/2018.
- Arnold, M.H. 2021. Teasing out Artificial Intelligence in Medicine: An Ethical Critique of Artificial Intelligence and Machine Learning in Medicine. *Journal of Bioethical Inquiry*. Vol. 18:1, s. 121–139.
- Becker, A. 2019. Artificial intelligence in medicine: What is it doing for us today? *Health Policy and Technology*. Vol. 8:2, s. 198–205.
- Bishop, C.M. 2006. *Pattern recognition and machine learning*. New York: Springer. Information science and statistics.
- Borrellas, P. & Unceta, I. 2021. The Challenges of Machine Learning and Their Economic Implications. *Entropy*. Vol. 23:3, s. 275.
- Brynjolfsson, E. & McAfee, A. 2017. Artificial intelligence, for real. *Harvard business review*. Vol. 1, s. 1-31.
- Char, D.S., Abramoff, M.D. & Feudtner, C. 2020. Identifying Ethical Considerations for Machine Learning Healthcare Applications. *American journal of bioethics*. Vol. 20:11, s. 7–17.
- Chandramouli, S., Dutt S. & Das A. 2018. *Machine Learning*. Pearson Education India.
- Chi, C.-L., Street, W.N. & Katz, D.A. 2010. A decision support system for cost-effective diagnosis. *Artificial Intelligence in Medicine*. Vol. 50:3, s. 149–161.
- Coeckelbergh, M. 2020. *AI ethics*. Cambridge, Massachusetts: The MIT Press (MIT Press essential knowledge series).
- Cooper, A.F., Moss, E., Laufer, B. & Nissenbaum, H. 2022. Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. *ACM International Conference Proceeding Series*. s. 864-876
- Hegde, P., & Shenoy, M. 2021. Artificial intelligence in medicine and health sciences. *Archives of Medicine and Health Sciences*. Vol. 9:1, s. 145–150.

- Hunter, P. 2019. The advent of AI and deep learning in diagnostics and imaging. *EMBO reports*. Vol. 20:7, s. e48559-e48559.
- Jordan, M.I. & Mitchell, T.M. 2015. Machine learning: Trends, perspectives, and prospects. *Science (American Association for the Advancement of Science)*. Vol. 349:6245, s. 255–260.
- Kaul, V., Enslin, S. & Gross, S.A. 2020. History of artificial intelligence in medicine. *Gastrointestinal Endoscopy*. Vol. 92:4, s. 807–812.
- Kim, B. & Doshi-Velez, F. 2021. Machine Learning Techniques for Accountability. *AI Magazine*. Vol. 42:1, s. 47–52.
- King, T.C., Aggarwal, N., Taddeo, M., & Floridi, L. 2020. Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*. Vol. 26:1, s. 89–120.
- Lee, J.H., Shin, J. & Realff, M.J. 2018. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering*. Vol. 114, s. 111–121.
- Mesko, B. & Görög, M. 2020. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digital Medicine*. Vol. 3:1, s. 126–126.
- Mirsky, Y., Mahler, T., Shelef, I. & Elovici, Y. 2019. CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning. Ithaca: Cornell University Library, arXiv.org. Saatavilla: <https://lib-proxy.tuni.fi/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fworking-papers%2Fct-gan-malicious-tampering-3d-medical-imagery%2Fdocview%2F2166880347%2Fse-2%3Faccountid%3D14242>
- Mitchell, T.M. 1997. *Machine learning*. New York: McGraw-Hill.
- Morrissey, J. 2014. The battle over patient data. How secure your access. *Medical economics*. Vol. 91:20, s. 16-20.
- Neittaanmäki, P., Tuominen, H., Äyrämö, S., & Vähäkainu, P. 2019. *Tekoäly ja terveydenhuolto Suomessa*. Jyväskylän yliopisto. Vol. 1.
- Nissenbaum H. 1996. Accountability in a computerized society. *Science and engineering ethics*. Vol. 2:1, s. 25–42.
- Pietarinen, J. 2015. *Etiikka*. filosofia.fi. Saatavilla: <https://filosofia.fi/fi/ensyklopedia/etiikka>

- Richens, J.G., Lee, C.M. & Johri, S. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*. 11:1, s. 3923–3923
- Roth, J. K. 2016. *Ethics*. John K. Roth (ed.). Pasadena, California: Salem Press.
- Runkler, T.A. 2012. Runkler, T.A. (ed.) *Data Analytics: Models and Algorithms for Intelligent Data Analysis* (1st ed. 2012). Wiesbaden: Vieweg+Teubner Verlag, s. 1–3.
- Salminen, A. 2011. Mikä kirjallisuuskatsaus? Johdatus kirjallisuuskatsauksen tyyppeihin ja hallintotieteellisiin sovelluksiin. Vaasa: Vaasan yliopisto. Saatavilla: http://www.uva.fi/materiaali/pdf/isbn_978-952-476-349-3.pdf
- Sarker I.H., Hoque M.M., Uddin M.K. & Alsanoosy T. *Mobile Data Science and Intelligent Apps: Concepts, AI-Based Modeling and Research Directions*. *Mobile networks and applications*. 26:1, s. 285–303.
- Siukonen, T. & Neittaanmäki, P. 2019. *Mitä tulisi tietää tekoälystä*. Jyväskylä: Docendo.
- Wichmann, J.L., Willeminck M.J. & De Cecco, C.N. (2020) *Artificial Intelligence and Machine Learning in Radiology: Current State and Considerations for Routine Clinical Implementation*. *Investigative radiology*. 55 (9), s. 619–627.