Timoteus Kivikangas

TEKOÄLYN EETTISET TEEMAT TERVEYDENHUOLLON TEKOÄLYTEKNOLOGIOISSA

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2023

TIIVISTELMÄ

Kivikangas, Timoteus
Tekoälyn eettiset teemat terveydenhuollon tekoälyteknologioissa
Jyväskylä: Jyväskylän yliopisto, 2023, 18 s.
Tietojärjestelmätiede, kandidaatin tutkielma
Ohjaaja: Lampi, Anna

Tekoälyn etiikka on nopeasti kehittyvä ja tärkeä ala, erityisesti kun tekoälyä käytetään yhä enemmän arkaluonteisiin aloihin, kuten terveydenhuoltoon. Jotta tekoälyä voidaan käyttää vastuullisesti ja eettisesti näillä aloilla, on tärkeää ymmärtää, miten tekoäly hyväksytään ja otetaan eettisesti vastaan. Tässä tutkielmassa tehtiin kirjallisuuskatsaus, jotta saatiin syvempi ymmärrys terveydenhuollon tekoälyn käytön eettisistä huolenaiheista. Kirjallisuuskatsaus tehtiin käyttäen Artificial Intelligence in Medicine -lehden artikkeleita, ja tuloksia verrattiin Jobin et al. (2019) tunnistamiin eettisiin teemoihin. Tutkimuksen tulokset viittaavat siihen, että läpinäkyvyys on tärkeä eettinen teema terveydenhuollon tekoälyssä, sillä se on yhteydessä muihin teemoihin kuten luottamukseen, vapauteen ja autonomiaan. Läpinäkyvyyden puute voi heikentää lääkärien päätöksentekoa ja potilaskeskeistä hoitoa. On kuitenkin tärkeää huomata, että tekoälyn eettiset ohjenuorat luodaan usein organisaatioiden omien tarpeiden mukaan eivätkä välttämättä vaikuta laajemmin alaan. Tutkimuksen tulokset voivat auttaa organisaatioita huomioimaan eettiset huolenaiheet suunnitellessaan tekoälyn käyttöönottoa terveydenhuollossa. On tärkeää saavuttaa yhteisymmärrys siitä, miten tekoälyn tulisi toimia eettisesti, jotta sitä voidaan käyttää vastuullisesti ja eettisesti arkaluonteisilla aloilla, kuten terveydenhuollossa.

Asiasanat: Tekoäly, Tekoälyn etiikka, Terveydenhuollon tekoäly

ABSTRACT

Kivikangas, Timoteus
Name of the publication
Jyväskylä: University of Jyväskylä, 2023, 18 pp.
Information Systems, Bachelor's thesis
Supervisor: Lampi, Anna

The ethics of artificial intelligence (AI) is a rapidly growing and important field, particularly as AI is increasingly implemented in sensitive areas such as healthcare. In order to ensure that AI is used ethically and responsibly in these contexts, it is crucial to understand how it is being received and accepted ethically. This study aimed to conduct a literature review to gain a deeper understanding of ethical concerns surrounding the use of AI in healthcare. The review was conducted using articles from the journal Artificial Intelligence in Medicine, and the findings were compared to ethical themes identified in guidelines by Jobin et al. (2019). The results of this study suggest that transparency is a key ethical theme in AI healthcare, as it is linked to trust, freedom, and autonomy. Lack of transparency can hinder physician decision-making and patient-centered care. However, it is important to note that guidelines for ethical AI development are often created by organizations for their own use, and may not necessarily have a broad impact on the field. The findings of this study can help organizations consider ethical concerns when planning the implementation of AI in healthcare settings. It is essential to have a consensus on how AI should function ethically in order to ensure its responsible and ethical use in sensitive areas such as healthcare.

Keywords: Artificial Intelligence (AI), AI ethics, Healthcare AI

KUVIOT

TAULUKOT

SISÄLLYS

1 JOHDANTO

Tässä kandidaatin tutkielmassa halutaan saada näkemys siitä, kuinka tekoälyn eettisiin kysymyksiin ja ohjenuoriin suhtaudutaan terveydenhuollon tekoälyteknologiaratkaisuissa. Ala ja aihe on suhteellisen nuori, ja merkittävät artikkelit on julkaistu vain muutamia vuosia sitten, joten uuden tiedon hankkiminen alasta on tärkeää. Terveydenhuollossa on myös aivan viime vuosina alettu puhua tekoälyteknologian suuremmasta roolista alalla. Viime vuosina terveydenhuollon tekoälyyn liittyvien tutkimusten määrä on noussut merkittävästi (Bentley & Bentley, 2018). Tekoälyä käyttävien järjestelmien kehittyneisyys on viime aikoina lisääntynyt siinä määrin, että niiden suunnittelu ja käyttöönotto vaatii yhä vähemmän ihmisen puuttumista (Arrieta ym., 2020). Eli tekoälyteknologia on kehittynyt koko ajan ihmisistä riippumattomammaksi teknologiaksi. Koska tällaisista järjestelmistä johdetut päätökset vaikuttavat ihmisten elämään, esimerkiksi lääketieteessä, lainsäädännössä tai puolustuksessa, on nousemassa tarve ymmärtää, kuinka kyseiset päätökset tekoälymenetelmillä toteutetaan (Arrietta ym., 2020).

Tässä tutkimuksessa tarkastellaan tekoälyä lääketieteessä ja terveydenhuollossa sekä alalla käytettäviä tekoälyteknologiaratkaisuja eettisen suhtautumisen näkökulmasta. Tutkimuksessa pyritään ensin ymmärtämään ja muodostamaan käsitys aikaisemman kirjallisuuden avulla tämän hetkisistä tekoälyn etiikkaan liittyvistä ohjenuorista, joista on jo olemassa jonkilainen konsensus. Tämän jälkeen tarkastellaan tekoälyyn liittyviä artikkeleita, jotka sijoittuvat terveydenhuollon kontekstiin, ja pyritään tunnistamaan niistä eettisiin ohjenuoriin liittyviä teemoja. Tutkielmassa pyritään muodostamaan näkemys, kuinka nämä eri teemat esiintyvät terveydenhuollon tekoälyn käyttöön liittyvissä artikkeleissa. Tarkasteltavat tutkimuskysymykset ovat :

1. Millaisia eettiseen tekoälyyn liittyviä ohjenuoria on olemassa?
2. Mitkä tekoälyyn liittyvät eettiset teemat ovat esillä terveydenhuollon tekoälyssä?

Näin voidaan saada ymmärrys siitä, mitkä teemat ovat tärkeitä ja kriittisiä terveydenhuollon tekoälyn kehittämisessä.

Tutkimuksessa käy ilmi, että tekoälyn ohjenuorat ovat nykyisellään käytännössä vain organisaatioiden itse tekemiä itselleen, eivätkä ne juuri vaikuta tekoälyn kehittäjiin (Hagendorff, 2020). Ohjenuorista voidaan tunnistaa 11 yleisimmin esiintyvää teemaa : läpinäkyvyys, oikeudenmukaisuus ja reiluus), ei-vahingollinnen, vastuullisuus, yksityisyys, hyödyllisyys, vapaus ja itsenäisyys, luotettavuus, arvo ja arvokkuus, kestävyys, solidaarisuus (Jobin ym. 2019). Tutkimuksessa haluttiin tutkia, kuinka nämä yleisesti esiintyvät tekoälyn eettiset teemat ilmenevät terveydenhuollon tekoälyn tutkimuksissa. Kirjallisuuskatsaus oli suppea katsaus ja sisälsi julkaistut artikkelit viimeisestä kolmesta Artificial Intelligence in Medicine julkaisun kerrasta. Näistä kolmestatoista pystyttiin tunnistamaan teemoja. Yleisin teema oli läpinäkyvyys, johon liittyi myös selitettävyys. Tämä esiintyi useimmin ja usein myös linkittyi muihin yleisiin teemoihin, kuten luotettavuuteen ja oikeudenmukaisuuteen.

Tekoäly on tämän kandidaatin tutkielman yksi keskeisimmistä käsitteistä. Tekoälylle ei kuitenkaan ole olemassa yhtä yleisesti hyväksyttyä määritelmää (Wang, 2019). Teknologian tutkimuskeskus VTT:n raportissa tekoäly määritellään: ”Tekoälyn avulla koneet, laitteet, ohjelmat, järjestelmät ja palvelut voivat toimia tehtävän ja tilanteen mukaisesti järkevällä tavalla” (Ailisto, Heikkilä, Helaakoski, Neuvonen & Seppälä, 2018, s. 7). VTT:n määritelmä mukailee Russel ja Norvigin (2020) antamaa määritelmää tekoälylle. Gartnerin määritelmän mukaan tekoäly käyttää kehittyneitä analyysiin ja logiikkaan perustuvia tekniikoita, mukaan lukien koneoppimista, tapahtumien tulkitsemiseen, päätöksien tukemiseen ja toimintojen tekemiseen (Gartner, 2022).

Etiikka itsessään on laaja ja moniulotteinen tieteenala. Yleisesti ymmärretään, että etiikka on rationaalista ja systemaattista standardien ja oikean sekä väärän tutkimista. Moraali taas ymmärretään yleisesti terminä hyvän ja huonon käsitteelle. (Kazim, 2017.)

Tämä kandidaatin tutkielma toteutetaan kuvailevana kirjallisuuskatsauksena. Kirjallisuuskatsaus sopii kandidaatin tutkielman menetelmäksi, sillä kirjallisuuskatsauksen avulla pyritään luomaan kokonaiskuva asiakokonaisuudesta sekä tunnistamaan mahdollisia ongelmia tutkimuskohteessa (Salminen, 2011). Tutkimuskysymyksen vuoksi tässä kandidaatintutkielmassa aineistot ovat laajoja, joten menetelmä on valittu, koska se ei rajaa aineiston valintaa tiukasti (Salminen, 2011).

Käytetyn kuvailevan kirjallisuuskatsauksen tyyli on narratiivinen katsaus, joka auttaa ajantasaistamaan tutkimustietoa, mutta ei tarjoa varsinaista analyyttistä tulosta (Salminen, 2011). Tämän kandidaatin tutkielman tarkoitus ei ole siis tarjota uutta analyyttistä tulosta, vaan luoda katsaus tutkittavaan aiheeseen ja ajantasaistaa tämänhetkistä olemassa olevaa tutkimusta, joten voidaan todeta, että valittu menetelmä sopii tähän kandidaatintutkielmaan toteutustavaksi.

Tutkielman aineisto on kerätty Google Scholar -ja Scopus-tietokannoista. Hakutermeiksi valikoitiin sopiviksi yleisesti:” AI ethics”,” Artificial intelligence ethics” ja” AI OR Artificial intelligence AND medicine OR healthcare”. Näillä

hakutermeillä löytyvistä artikkeleista on valittu sopivimmat ensin otsikon perusteella, jonka jälkeen vielä tiivistelmän perusteella on valikoitu tutkimuksen kannalta sopivimmat artikkelit. Lisää aineistoja on kerätty myös hakutermeillä löytyneiden artikkeleiden lähdeluetteloista.

Tekoälyn etiikan teemoja terveydenhuollon kontekstissa käsittelevää osiota varten on kerätty aineistoa myös käymällä läpi julkaisua ”Artificial Intelligence in Medicine”. Julkaisu julkaisee nimensä mukaisesti terveydenhuoltoon ja lääketieteeseen liittyviä tekoälyä käsitteleviä artikkeleita. Pyrin valitsemaan vain sellaisia artikkeleita, jotka keskittyvät enemmän tekniseen puoleen, eli tekoälyyn, kuin lääketieteeseen.

Seuraavaksi luvussa kaksi tarkastellaan tekoälyn etiikkaa ja ohjenuoria yleisellä tasolla. Luvussa kolme tarkastellaan ensin terveydenhuollon tekoälyn sovelluksia ja käyttöä tällä hetkellä, jonka jälkeen tarkastellaan eettisiä teemoja tämänhetkisessä tutkimuksessa.

2 TEKOÄLYN ETIIKKA JA OHJENUORAT

Tässä pääluvussa esitellään ensin ensimmäisessä alaluvussa tekoälyn etiikkaa. Siihen liittyvät tekoälyn eettiset ohjenuorat ja niiden teemat, joita käsitellään toisessa alaluvussa.

2.1 Tekoälyn etiikka

Tekoälyn etiikka on ala, joka on noussut vastauksena kasvavalle huolenaiheelle koskien tekoälyn vaikutuksia. Voidaan ajatella, että tekoälyn etiikka on alalaji digitaalisen etiikan alla. (Kazim & Koshiyama, 2021.) Vaikka jo vuonna 2006 on esitetty käsite kone-etiikka (engl. machine ethics), on tekoälyn etiikka silti vielä alana alkuvaiheessa (Siau & Wang. 2020.)

Laitteiden, jotka voivat aiheuttaa vahinkoa ja harmia, tulisi osata erottaa hyvä ja oikea sekä huono ja väärä valinta eli toisin sanoen päätöksiä tekevien tietokoneiden tulisi olla selkeä moraalinen päätöksentekijä (Etzioni & Etzioni. 2017). Koneiden moraalinen päätöksenteko voidaan toteuttaa joko ylhäältä alas (engl. top-down) tai alhaalta ylös (bottom-up) -metodeilla. Ylhäältä alas -metodissa koneisiin implementoidaan moraalisia sääntöjä ohjelmoidessa niitä. Alhaalta ylös -metodissa koneet eivät käytä juuri ollenkaan aikaisempaa teoriaa muuten kuin täsmentääkseen tehtävää järjestelmälle, mutta ei kerro kuinka sen tulisi toimia tilanteessa. Alhaalta ylös -lähestymistavassa tähdätään tavoitteisiin, jotka voidaan määritellä tai olla määrittelemättä teoreettisin termein. (Wallach, Franklin & Allen. 2010.) Wallach ym. (2010) mukaan alhaalta ylös -lähestymistapa oli vielä artikkelin julkaisun aikaan rajallista sen aikaisen tekniikan vuoksi, mutta he korostavat koneoppimisen kehityksen merkitystä tulevaisuuden kannalta. Nykyään koneoppiminen ja neuroverkot ovat kehittyneet paljon vuodesta 2010, joten on hyvin todennäköistä, että nykytekniikka paremmin mahdollistaa alhaalta ylös -lähestymistavan.

Vaikka ylhäältä alas -lähestymistapa voi vaikuttaa selkeämmältä ja helpommalta, vaatii se selvästi sen, että koneita ohjelmoivilla henkilöillä on jokin tietty

käsitys, kuinka koneen tulisi toimia eettisesti. Kuitenkin Hagendorffin (2020) mukaan organisaatioiden luomilla ohjenuorilla on vain vähän vaikutusta kehittäjiin ja kehittämiseen.

2.2 Tekoälyn eettiset ohjenuorat ja teemat

Viimeisen viiden vuoden aikana eri organisaatiot ovat laatineet periaatteita ja ohjenuoria eettisen tekoälyn käytölle (Jobin, Ienca & Vayena. 2019). Vaikka vieläkään ei ole täysin selvää, mistä eettinen tekoäly koostuu (Jobin ym. 2019), on siitä jo jonkinlainen konsensus olemassa. Tekoälyyn liittyvät ohjenuorat muodostuvat tietyistä teemoista. Artikkelissaan Jobin ym. (2019) selvittivät, mitkä eettisen tekoälyn periaatteet ja teemat ovat nousseet eniten esille aiheen tutkimuksissa. Analyysin tuloksena esiin nousi 11 eettisen tekoälyn arvoa ja käytäntöä, joista eettisen tekoälyn ohjenuoria muodostetaan:

- läpinäkyvyys (engl. transparency)
- oikeudenmukaisuus ja reiluus (engl. justice & fairness)
- ei-vahingollinnen (engl. non-maleficience)
- vastuullisuus (engl. responsibility)
- yksityisyys (engl. privacy)
- hyödyllisyys (engl. beneficience)
- vapaus ja itsenäisyys (engl. freedom & autonomy)
- luotettavuus (engl. trust)
- arvo, arvokkuus (engl. dignity)
- kestävyys (engl. sustainability)
- solidaarisuus (engl. solidarity)

Näistä teemoista vastuullisuus, yksityisyys ja oikeudenmukaisuus esiintyvät yhdessä noin 80 prosentissa tekoälyn etiikan ohjenuorissa (Hagendorff. 2020) Tutkimuksessaan Hagendorff (2020) käy läpi 22 suurta tekoälyn etiikan ohjenuoraa. Hänen tarkoituksenaan on käydä läpi puolisystemaattisesti alan ongelmia ja normatiivisia asenteita, jotka osoittavat kuinka tekoälyn etiikan yksityiskohdat liittyvät isompaan kokonaisuuteen. Näihin kaikista yleisimmin mainittuihin näkökulmiin on jo olemassa, tai voidaan toteuttaa, tekninen ratkaisu (Hagendorff, 2020). Koska nämä 11 Jobin ym. (2019) artikkelin löytämää ja muodostamaa teemaa ovat kaikista yleisimpiä, käytän tässä artikkelissa niitä lähtökohtana seuraavaksi tehdyn kirjallisuuskatsauksen tuloksien muodostamisessa. Kuitenkaan Hagerdorffin (2020) mukaan ohjenuorat eivät ole riittävän konkreettisia, jotta niillä olisi merkitystä ja vaikutusta. Lisäämällä konkretiaa ohjenuoriin niillä voisi olla enemmän vaikutusta tekoälyä kehittäviin toimijoihin (Hagendorff, 2020).

3 EETTISTEN TEEMOJEN TUNNISTAMINEN JA HYÖDYNTÄMINEN TERVEYDENHUOLLON TEKOÄLYTEKNOLOGIOISSA

3.1 Tekoäly terveydenhuollossa ja lääketieteessä

Tekoälyn tutkimus terveydenhuollossa on kasvattanut suosiota viime vuosina merkittävästi. Julkaistujen tieteellisten artikkelien, jotka koskevat tekoälyä terveydenhuollossa, määrä on kasvanut vuodesta 2011 lähtien (Bentley & Bentley. 2018). Bentleyn ja Bentleyn (2018) tutkimuksen aikana julkaisujen määrä oli korkeimmillaan. Voidaan myös uskoa, että tutkimus ei ole vähentynyt aivan viimeisten vuosien aikana Bentleyn ja Bentleyn (2018) artikkelin jälkeen.

Terveydenhuollon alalla on olemassa tarve vahvalle laskennalliselle tekoälyn etiikan viitekehykselle (Bali, Garg & Bali. 2019). Myös Hagendorff (2019) kertoo, kuinka yleisesti tekoälyn eettiset ohjenuorat eivät ole tarpeeksi konkreettisia, jotta niistä olisi mahdollisimman paljon hyötyä ja niitä olisi helppo käyttää. Artikkelissa Bali ym. (2019) myös antavat esimerkkejä, kuinka tekoälyä nykypäivänä käytetään terveydenhuollossa, millaisessa käytössä se on hyvä ja miksi sen käytössä terveydenhuollossa on eettisiä ongelmia. Bali ym. (2019) mukaan jo nykyään tekoälyllä on pystytty valitsemaan syöpäpotilaiden lääkehoitoa yhtä tarkasti tai tarkemmin kuin ammattilaisen ihmisen valitsemana. Googlen DeepMind-alustaa on käytetty tunnistamaan terveysriskejä analysoimalla potilaiden mobiiliapplikaation dataa ja lääketieteellisiä kuvauksia. Stanfordin radiologian algoritmi poimi keuhkokuumeen paremmin kuin ihmisradiologit. (Bali ym., 2019).

Bali ym. (2019) kertoo kuinka vuonna 2018 kehitettiin automatisoitu algoritmi tunnistamaan diabeettista retinopatiaa (DR) (Bali ym. 2019). DR on verkkokalvosairaus, joka kehittyy monille diabeetikoille (Tarnanen, Summanen & Komulainen. 2017). Algoritmi suoriutui Amerikan silmälääkäreiden ja verkkokalvon asiantuntijoiden suositusten mukaisesti (Bali ym., 2019).

Bentleyn ja Bentleyn (2018) artikkelissa tekoäly ja IoT (Internet of Things) ovat jaoteltu käytön mukaan yhdeksään kategoriaan, jotka ovat:

- Puettavat laitteet & liitettävyys (engl. Wearables & Connectivity)
- Havaitseminen & hoito (engl. Detection & Treatment)
- Anturiverkot (engl. Sensor Networks)
- Potilashoito (engl. Patient Care)
- muu (engl. Other)
- Ennustavat mallit/päätöksentukijärjestelmät (engl. Predictive Models/Decision Support Systems)
- Järjestelmän hallinta (engl. System Management)
- Neuroverkot (engl. Neural Networks)
- Automaatio / robotiikka (engl. Automation/Robotics)

(Bentley & Bentley. 2018 s. 36)

Tekoälyä käytetään erityisesti potilashoitoon, havaitsemiseen, ennustaviin malleihin, neuroverkkoihin ja automatisointiin. Nämä kategoriat pelkästään esiintyvät noin 40 prosentissa kirjallisuuskatsauksen artikkeleista (Bentley & Bentley. 2018).

3.2 Tekoälyn eettiset teemat ja ohjenuorat terveydenhuollon tekoälyteknologioissa

Viime vuosina tekoälyn innovaatiot ovat johtaneet uusien terveydenhuollon tekoälyteknologioiden kehittämiseen. Vaikka uudet tekoälyteknologiat ovat hyödyllisiä ja parantavat potilaiden kokemusta, on silti varoitettu, että tekoäly voi aiheuttaa ja pahentaa haittoja sekä epäkohtia terveydenhuollossa (Frost, Bosward, Saint James Aquino, Braunack-Mayer, Carter. 2022). Eli on olemassa myös tunnistettu huoli tekoälyn käyttöön liittyen juuri terveydenhuollossa.

Seuraavissa alaluvuissa käsitellään Jobin ym. (2019) tunnistamia tekoälyn eettisiä teemoja terveydenhuollon kontekstissa. Terveydenhuollon tekoälyä käsittelevissä tutkimuksissa esiintyneet eettiset teemat esitetään tiivistetysti taulukossa 1.

3.2.1 Läpinäkyvyys, oikeudenmukaisuus ja reiluus

Läpinäkyvyydellä tarkoitetaan muun muassa pyrkimystä lisätä tekoälyn selitettävyyttä ja tulkittavuutta (Jobin ym., 2019), jotka tulevat esille automatisoidussa päätöksenteossa, sekä tietojen käytön tai soveltamisen tarkoituksen selventämistä.

Läpinäkyvyyttä ja selitettävyyttä on pyritty nostamaan esille terveydenhuollon ja lääketieteen tekoälyssä (Combi ym., 2022; Banham ym., 2022; Naseri ym., 2022; Tokuoka ym., 2022; Butz ym., 2022). Erityisesti läpinäkyvyyttä

tarvitaan perustellessa tekoälyjärjestelmän tekemiä päätöksiä (Naseri ym., 2022). Koneoppimisalgoritmien tekemien ennusteiden heikko selitettävyys ei vain ainoastaan vaikeuta ennusteiden luotettavuutta kliinisessä käytössä, vaan menettää myös mahdollisuuden lääkärin päätöksentekoon ja potilaskeskeiseen hoitoon (Tokuoka ym., 2022).

Tekoälyn sovelluksissa yhä enemmän pyritään käyttämään selitettävää tekoälyä tai koneoppimista (Banham ym., 2022). Selitettävä tekoäly (XAI: Explainable Artificial Intelligence) on termi, jolla tarkoitetaan tekoälyä, joka tuottaa selitettävämpiä ja ymmärrettävämpiä malleja menettämättä tehokkuuttaan. Ihmisten on mahdollista ymmärtää malleja ja luottaa sekä hallita tekoälyä (Arrietta ym., 2019). Tekoälyn käytön yleistyessä terveydenhuollossa lääkäreiden tulisi saada tieto siitä, kuinka tekoäly on päätynyt tiettyyn johtopäätökseen (Combi ym., 2022).

Oikeudenmukaisuus ja reiluus ovat pääasiassa ei-toivotun harha, joka saattaa johtaa syrjintään, estämistä ja monitorointia tekoälyn datajoukoissa. Tähän kuuluu myös vaikutus työntekijöihin ja työmarkkinoihin. (Jobin ym., 2019). Terveydenhuollossa käytettävän tekoälyn antamat ennusteet voivat olla rajoittuneita ja vääristyneitä, mutta se voi antaa uusia näkökulmia, joita ihminen ei tule ajatelleeksi (Combi ym., 2022), joten voidaan päätellä, että tekoälyn heikkouksistakin voi olla oikein käytettynä jonkinlaista hyötyä. Koska terveydenhuollon tekoälyn kehittämisessä ymmärretään epätasapainoisen datajoukon ongelmat, on ongelmaan pyritty kehittämään ratkaisuja menettämättä tekoälyn hyötyjä (Bi ym., 2022). Datajoukkoja tehdessä data tulee segmentoida ja merkitä. Tämä prosessi on altis virheille ja eri ihmisillä on taipumus tehdä prosesseja eri tavalla. Oikeanlaisella tekoälyllä on mahdollista automatisoida prosessi, jolloin on mahdollista poistaa manuaalisen segmentoinnin ei-haluttuja virheitä (Gruber ym., 2022), joten tekoälyllä on myös mahdollisuus positiivisesti vaikuttaa datajoukkojen laatuun.

3.2.2 Ei-vahingollisuus, vastuullisuus ja yksityisyys

Ei-vahingollisuus on yleistä turvallisuutta. Tekoälyn ei tulisi koskaan tuottaa ennakoitavissa olevaa tai tahatonta haittaa ja harmia. Tällä voidaan tarkoittaa myös tiettyjen riskien tai tahallisen väärinkäytön välttämistä ja estämistä. Haitta tarkoittaa tässä usein syrjintää, yksityisyyden loukkaamista tai kehollista harmia. (Jobin ym., 2019). Terveydenhuollossa ja lääketieteessä voidaan luoda tekoälyllä malleja riskitekijöistä ja niiden suhteista tarkoituksena parantaa terveydenhuollon turvallisuutta ja vähentää vahingollisuutta (Butz ym., 2022).

Vastuullisuuteen kuuluu rehellinen toiminta ja vastuun sekä oikeudellisen vastuun jakaminen ja selventäminen. Vastuu tulisi mielellään selventää etukäteen sopimuksia tai oikeussuojakeinoja hyödyntäen. Vastuullisuuskysymyksiin kuuluu myös, voidaanko lääkäreitä pitää vastuullisina tekoälyn päätöksistä ja voiko tekoäly olla itse vastuussa? (Jobin ym., 2019). Selkeästi vastuullisuuteen liittyviä teemoja ei havaittu tarkastelluissa artikkeleissa.

Yksityisyys nähdään arvona, jota kannattaa puolustaa, ja oikeutena olla suojattu. Yleensä tekoälyn kontekstissa tämä näkyy datan suojaamisena ja turvaamisena, jotta yksityisyyden suoja säilyisi. (Jobin ym., 2019). Tekoälyn kehittämisessä käytetään dataa, mutta yksityisyyden suojaamiseksi terveydenhuollossa tekoälyä joudutaan usein kehittämään pienillä datajoukoilla (Hao ym., 2022).

3.2.3 Hyödyllisyys, vapaus ja itsenäisyys

Hyödyllisyyttä on yleinen ihmisten hyvinvoinnin paraneminen, rauhan ja onnellisuuden lisääminen, taloudellinen vaurastuminen ja asiakastarpeeseen vastaaminen. Hyödyllisyyteen liittyen mainitaan ihmisten hyvinvointiin kehitettävistä uusista mittareista ja mittauksista. (Jobin ym., 2019).

Terveydenhuollossa tulee miettiä, tuleeko käyttäjä käyttäneeksi uusia järjestelmiä, jos se vastaa johonkin tarpeeseen eli järjestelmän tulisi tuottaa käyttäjälle jotain käytännön hyötyä (Combi ym., 2022). Tekoälymallien ei tarvitse aina olla tulkittavia ollakseen myös hyödyllisiä. Jos tekoäly antaa oikean ennusteen, jonka avulla lääkärin on helpompi hoitaa potilasta, on tekoäly hyödyllinen ilman selitettävyyttä. (Combi ym., 2022). Terveydenhuollossa tekoälyn hyödyllisyys nousee esille muun muassa tehokkaampana diagnosointina (Jiang ym., 2022). Merkittävä hyöty tekoälystä ja koneoppimisesta on bioinformatiikan suuren datamäärän tutkimisessa (Danaila ym., 2022).

Vapauteen ja itsenäisyyteen liittyvät sananvapaus ja tietoinen itsemääräämisoikeus. Autonomia voidaan nähdä sekä hyvänä että huonona vapaudelle ja itsenäisyydelle. (Jobin ym., 2019). Vaikeasti ymmärrettävä tekoäly voi vaikuttaa lääkärin päätöksentekoon ja vaikeuttaa sekä vähentää potilaskeskeistä hoitoa (Tokuoka ym., 2022). Suomessa potilaskeskeisessä toiminnassa potilas osallistuu hoitoa, itsehoitoa ja päätöksentekoa koskeviin prosesseihin (SFS-EN 17398:2020), joten monimutkainen ja heikosti selitetty tekoälyalgoritmi voi mahdollisesti vaarantaa potilaan vapautta ja itsenäisyyttä päätöksenteossa.

3.2.4 Luotettavuus, arvokkuus, kestävyys ja solidaarisuus

Luotettavuus sisältää käsitteen luotettava tekoäly (engl. trustworthy AI) (Jobin ym., 2019). Euroopan Komission luotettavan tekoälyn periaatteiden mukaan tekoälyn tulisi olla lakeja noudattava, eettinen ja vankka sekä teknisesti että sosiaalisesti (European Comission, 2019). Luotettava tekoäly alleviivaa asiakkaan ja käyttäjän luoton merkitystä (Jobin ym., 2019).

Terveydenhuollossa käytetyn tekoälyn tulisi olla ymmärrettävää, jotta terveydenhuollon henkilökunnan on helpompi luottaa tuloksiin (Combi ym., 2022; Tokuoka ym., 2022). Tekoälyn selitettävyyden on todettu mahdollistavan ihmisen paremman luottamisen tekoälyyn (Arrietta ym., 2019). Kun tekoälyn päätöksenteko esitetään ihmisille ymmärrettävässä muodossa, se lisää luottamusta järjestelmän antamiin tuloksiin (Combi ym., 2022).

Tekoälyalgoritmien tietyissä sovelluksissa lääketieteessä, esimerkiksi kuvantunnistuksessa, algoritmi voi edelleen olla liian altis suurelle vaihtelulle, joten

kokenut ihminen voi olla luotettavampi tekemään oikeita johtopäätöksiä. Vaihtelua pyritään vähentämään tarkalla standardisoinnilla. (Vasiljevic ym., 2022.)

Arvokkuudella tarkoitetaan ihmisarvoa ja ihmisten oikeuksiin liittyviä asioita. Tekoälyn ei tulisi vähentää tai tuhota vaan kunnioittaa, säilyttää ja jopa lisätä ihmisarvoa. (Jobin ym., 2019). Kestävyydellä viitataan usein ympäristön suojelemiseen kehittäessä tekoälyä. Tekoäly luo kestävän järjestelmän, joka prosessoi dataa kestävällä tavalla. (Jobin ym., 2019). Arvokkuutta tai kestävyyttä ei pystytty tunnistamaan kirjallisuuskatsauksessa läpi käydyistä terveydenhuollon tekoälyn artikkeleista.

Solidaarisuus käsitetään yleensä tilanteissa, joissa tekoälyä implementoidaan työmarkkinoiden eri osa-alueille. Tekoälyn ei haluta uhkaavan sosiaalista koheesiota ja halutaan kunnioittavan mahdollisesti haavoittuvia ihmisjoukkoja. (Jobin ym., 2019). Tulevaisuudessa tekoälyä ja koneoppimista hyödyntävät autonomiset teknologiat vähentävät työpaikkoja monilla aloilla (Doraiswamy, Charlotte, & Bodner. 2020). Vaikka tekoäly pystyy tällä hetkellä auttamaan lääkäriä, on epätodennäköistä, että se pystyisi korvaamaan lääkärin ennakoitavissa olevassa tulevaisuudessa (Krittanawong, C., 2018). Kyselytutkimuksessa vain 3,8 prosenttia kyselyyn vastanneista psykiatreista koki, että teknologia voisi viedä heidän työnsä ja 17 prosenttia koki, että tekoäly voisi todennäköisesti korvata ihmislääkärin tarjoaman empaattisen hoidon (Doraiswamy ym., 2020). Kyselyn perusteella melko harva terveydenhuollon alalla kokee, että tekoäly voi korvata ihmisarvoa tai kyetä samaan empatiaan kuin ihminen. Tähän liittyvät artikkelit löytyivät kuitenkin erikseen etsittäessä julkaisusta, koska aivan viimeisimmissä julkaisukerroissa aihetta ei ole juuri käsitelty.

TAULUKKO 1 Eettisten teemojen havainnot artikkeleissa

Teema	**Esiintyy Artikkeleissa**
Läpinäkyvyys	Banham ym., (2022); Butz ym., (2022); Combi ym., (2022); Naseri ym., (2022); Tokuoka ym., (2022)
Oikeudenmukaisuus, reiluus	Bi ym., (2022); Combi ym. (2022); Gruber (2022);
Ei-Vahingollisuus	Butz ym., (2022)
Vastuullisuus	–
Yksityisyys	Hao ym., (2022);
Hyödyllisyys	Combi ym., (2022); Jiang ym., (2022); Danaila ym., (2022)
Vapaus ja itsenäisyys	Tokuoka ym., (2022);
Luotettavuus	Combi ym., (2022); Tokuoka ym., (2022); Vasiljevic ym., (2022)
Arvo, Arvokkuus	–
Kestävyys	–
Solidaarisuus	Doraiswamy ym., (2020); Krittawong, C. (2018)

4 YHTEENVETO

Tekoälyn yleistyminen viime vuosina eri aloilla on nostanut esille siihen liittyviä uhkia ja riskejä. Tekoälyn etiikka on alana vielä nuori ja vain muutamia vuosia vanha. Yksi ala, jossa tekoälyn kehitys ja käyttö on yleistynyt viime vuosina, on terveydenhuolto, mutta terveydenhuollossa tekoälyyn liittyy omia eettisiä ongelmia, joita tässä tutkimuksessa pyrittiin tunnistamaan. Kirjallisuuskatsauksen myötä pyrittiin löytämään ja tunnistamaan terveydenhuollon tekoälyyn liittyviä ongelmia, jotta ne tiedostettaisiin ja niihin voidaan kehittää ratkaisuja. Alalta on myös vasta hyvin vähän tutkimusta.

Tutkimus toteutettiin kirjallisuuskatsauksena, koska se sopii kandidaatin tutkielman menetelmäksi ja tutkimuksessa ei pyritty luomaan uutta teoriaa, vaan muodostamaan käsitys olemassa olevasta tutkimuksesta ja sen luonteesta. Kirjallisuuskatsaus toteutettiin Salmisen (2011) mukaisesti. Aineistoa kerättiin omilla hakutermeillä, mutta suuri määrä lähdeaineistoa kerättiin julkaisusta *Artificial Intelligence in Medicine*, koska julkaisu osoittautui sopivaksi tämän tutkimuksen aiheen vuoksi.

Tutkimuksessa ensin kerättiin tekoälyn etiikan kirjallisuudesta alan tämänhetkinen tieto. Tutkimuksen perustana toimii Jobin ym. (2019) tunnistamat tekoälyn eettisten ohjenuorien teemat. Teemoja on yhteensä 11, joista muodostuu suurin osa tekoälyn ohjenuorista, joita eri organisaatiot ovat viime vuosina tehneet.

Ensimmäinen tutkimuskysymys oli: Millaisia eettiseen tekoälyyn liittyviä ohjenuoria on olemassa? Aiheesta on aikaisempaa tutkimusta, joka jaottelee ohjenuorat eri teemojen mukaan, mitä ne sisältävät. Jobin ym. (2019) on tunnistanut 11 eri teemaa, jotka muodostavat suurimman osan teemoista. Hagendorff (2020) tunnisti, että näistä kolme yleisintä eli vastuullisuus, yksityisyys ja oikeudenmukaisuus esiintyvät noin 80 prosentissa tekoälyn ohjenuorista. Hagendorff (2020) myös kritisoi, että ohjenuorat eivät sisällä tarpeeksi konkreettisia asioita ja jäävät sen vuoksi merkityksettömiksi eivätkä ne vaikuta esimerkiksi tekoälyn kehittäjiin juuri lainkaan. Eli ohjenuorat ovat vain organisaatioiden itse itselleen tekemiä eikä niillä ei ole kovinkaan paljon vaikutusta.

Toinen tutkimuskysymys oli, mitkä tekoälyyn liittyvät eettiset teemat ovat esillä terveydenhuollon tekoälyssä? Tutkimnuksessa tunnistettiin tekoälyn terveydenhuollon artikkeleista Jobin ym. (2019) teemoja. Mitkä sillä alalla ovat esillä? Terveydenhuollon tekoälyn artikkeleissa esiintyi eniten läpinäkyvyys. Siihen liittyi usein käsite XAI eli selitettävä tekoäly (engl. Explainable AI). (Combi ym., 2022; Banham ym., 2022; Naseri ym., 2022; Tokuoka ym., 2022; Butz ym., 2022.) Tekoälystä haluttaisiin luoda ymmärrettävämpi käyttäjille sekä potilaille. Tämä liittyy Jobin ym., (2019) teemoista läpinäkyvyyteen.

Tutkimuksissa esiintyi läpinäkyvyyden jälkeen eniten kolme teemaa, jotka olivat oikeudenmukaisuus ja reiluus (Bi ym., (2022); Combi ym. (2022); Gruber (2022)), hyödyllisyys (Combi ym., (2022); Jiang ym., (2022); Danaila ym., (2022)) sekä luotettavuus (Combi ym., (2022); Tokuoka ym., (2022); Vasiljevic ym.,

(2022)). Tutkimuksessa ei tunnistettu teemoja vastuullisuus, arvo ja arvokkuus sekä kestävyys. Tutkimusta rajoitti kuitenkin kirjallisuuskatsauksen pieni ja suppea otos.

Läpinäkyvyydellä ja selitettävyydellä havaittiin olevan paljon vaikutuksia myös muihin teemoihin. Läpinäkyvyys liittyi Jobin ym. (2019) teemaan vapaus ja itsenäisyys, sillä tekoälyn huono ymmärrettävyys voi johtaa potilaskeskeisen hoidon vähentymiseen (Tokuoka ym., 2022), jolloin se vaikuttaa potilaan vapauteen ja itsenäisyyteen. Koska tekoälyn selitettävyys ja läpinäkyvyys ovat yhteydessä vahvasti myös muihin teemoihin välillisesti tai välittömästi, on sen ongelman ratkaiseminen erityisen tärkeää. Jatkotutkimusta siitä, kuinka tekoälystä tehtäisiin terveydenhuollon henkilökunnalle sekä potilaille ymmärrettävämpi, tarvitaan.

Koska ymmärrettävyys, läpinäkyvyys ja luotettavuus nousi paljon esille terveydenhuollossa, voisi sitä tutkia, että johtuuko tämä siitä, että ala ei ole ns. teknologia-ala ja esiintyykö samanlaista käyttäytymistä teknologian aloilla, joissa ihmiset ovat todennäköisemmin tietoisempia tekoälyteknologiasta? Tutkimusta rajoitti tutkittavan kontekstin eli terveydenhuollon laajuus sekä oman ammattitaidon puute siltä alalta. Myös tutkimuksessa käytetty suppea määrä lähdeaineistoja vaikuttaa tulosten luotettavuuteen ja merkittävyyteen.

LÄHTEET

Ailisto, H., Heikkilä, E., Helaakoski, H., Neuvonen, A., Seppälä, T. (2018). Tekoälyn kokonaiskuva ja osaamiskartoitus. *Selvitys- ja tutkimustoiminna julkaisusarja 46/2018, Valtioneuvoston selvitys- ja tutkimustoiminta*

Arrietta, A, B,. Diaz-Rodriguez, N., Ser, J, D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Consepts, taxonomies, opportunmities and challenges towards responsible AI. *Information Fusion, 58, 81-115.*

Bali, J., Garg, R., Bali, R. T. (2019). Artificial intelligence (AI) in healthcare and biomedical research: Why a strong computational/AI bioethics framework is required? *Indian Journal of Ophthamology, 2019, 1:3-6*

Bentley, R. S., Bentley A. C. (2018). IoT and AI in Healthcare: A Systematic Literature Review. *Issue in Information Systems, Volume 19, Issue 3, pp 33-41, 2018*

Bi, L., Kim, J., Su, T., Fulham, M., Feng, D., Ning, G. (2022). Deep multi-scale resemblance network for the subclass differentation of adrenal masses on computed tomography images. *ArtificialIntelligence in Medicine*. 132.

Butz, R., Schulz, R., Hommersom, A., Van Eekelen, M. (2022). Investigating the understandability of XAI methods for enhanced user experience: When Bayesian network users became detectives. *Artificial Intelligence in Medicine.* 134.

Combi, C., Amico, B., Bellazzi, R., Holzinger, a., Moore, J., Zitnik, M., Holmes, J. (2022). A manifesto on explainability for artificial intelligence in medicine. *Artificial Intelligence in Medicine.* 33

Danaila, V., Avram, S., Buiu, C. (2022). The applications of machine learning in HIV neutralizing antibodies research – A Systematic review. *Artificial Intelligence in Medicine*. 134.

Doraiswamy, P., Blease, C., Bodner, K. (2020). Artificial intelligence and the future of psychiatry: Insights from a global physician survey. *Artificial Intelligence in Medicine*. 102.

Etzioni, A., Etzioni, O. (2017). Incorporating Ethics into Artificial Intelligence. *The Journal of Ethics,* 21, pp. 403-418

EU Rules on international Data transfer, , haettu 7.11.2022

European Comission. (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai , viitattu 8.12.2022

Frost, E, K., Bosward, R., Saint James Aquino, Y., Braunack-Mayer, A., Carter, S, M. (2022) Public Views on ethical issues in healthcare artificial intelligence: protocol for a scoping review. *Systematic reviews (2022) 11:142*

Gartner. (2022). Gartner Glossary: Artificial Intelligence (AI). https://www.gartner.com/en/information-technology/glossary/artificial-intelligence , Viitattu 16.10.2022

Giovanni, B. & Le Moine, O. (2020). Artificial Intelligence in Medicine: Today and Tomorrow. *Frontiers in Medicine. 05 February 2020*

Gruber, N., Galijasevic, M., Regodic, M., Grams, A., Siedentopf, C., Steiger, R., Hammerl, M., Haltmeier, M., Gizewski, E., Janjic, T. (2022). A deep learning pipeline for the automated segementation of posterior limb of internal capsule in preterm neonates. *Artificial Intelligence in Medicine*. 132.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines 30, 99-120 (2020)*

Hao, D., Ashan, M., Salim, T., Duarte-Rojo, A., Esmaeel, D., Zhang, Y., Arefan, D., Wu, S. (2022). A self-training techer-student model with an automatic label grader for abdominal skeletal muscle segementation. *Artificial Intelligence in Medicine*. 132.

Jiang, J., Peng, J., Hu, C., Jian, W., Wang, X., Liu, W. (2022). Breast cancer detection and classification in mammogram using three-stage deep learning framework based on PAA algorirthm. *Artificial Intelligence in Medicine.* 134.

Jobin, A., Ienca, M & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines. *Nature Machine Intelligence 1, 389-399*

Kazim, E. (2017). Kant on conscience: A unified approach to moral self-consciousness, *Brill.*

Kazim, E., Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Pattern*(2) 9.

Krittanawong, C. (2018). The rise of artificial intelligence and the uncertain future for physicians. *European Journal of Internal Medicine*. 48, pp. e13- e14.

Naseri, M., Tabibian, S., Homayounvala, E. (2022). Adaptive and personalized user behavior modeling in complex event processing platforms for remote health monitoring systems. *Artificial Intelligence in Medicine.* 134

Russel, S., Norvig, P. (2020). Artificial Intelligence: A modern Approach 4yh US edition. *Prentice Hall*

Salminen, A. (2011). Mikä kirjallisuuskatsaus?: johdatus kirjallisuuskatsauksen tyyppeihin ja hallintotieteellisiin sovelluksiin. Saatavilla verkossa

SFS-EN 17398:2020 (2021). Potilaan ja asiakkaan osallisuus terveydenhuollossa. Potilaskeskeisen ja asiakaslähtöisen hoidon vähmmäisvaatimukset. *Suomen Standardisoimisliitto SFS ry*. viitattu 16.12.2022.

https://sales.sfs.fi/fi/index/tuoteuutiset/potilaskeskeisenhoidonminimivaatimuksetuudessastandardissa.html.stx

Siau, K., Wang, W. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *Journal on Database Management*, 31(2).

Tarnanen, K., Summanen, P., Komulainen, J. (2017) Diabeettinen retinopatia – diabtekseen liittyvä silmäsairaus. *Duodecim, käypähoito*. Haettu 7.11.2022 osoitteesta https://www.kaypahoito.fi/khp00059

Tokuoka, Y., Yamada, T., Mashiko, D., Ikeda, Z., Kobayashi, T., Yamagata, K., Funahashi, A. (2022). An explainaable deep learning-based algorithm with an attention mechanism for predicting the live birth potential of mouse embryos. *Artificial Intelligence in Medicine.* 134

Vasiljevic, J., Nisar, Z., Feuerhake, F., Wemmer, C., Lampert, T. (2022). CycleGan for virtual stain transfer: Is seeing really believing?. *Artificial Intelligence in Medicine*. 133

Wallach, W., Franklin, S., Allen, C. (2010). A Conceptual and Computational Model of Moral Decision Making in Human and Artificial Agents. *Topics in Cognitive Science*. 2(3), pp. 454-485

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence, 10(2) 1-37, 2019*