

Iiro Kinnunen

**TEKOÄLYN LÄPINÄKYVYYS JULKISELLA
SEKTORILLA**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2020

TIIVISTELMÄ

Kinnunen, Iiro

Tekoälyn läpinäkyvyys julkisella sektorilla

Jyväskylä: Jyväskylän yliopisto, 2020, 32 s.

Tietojärjestelmätiede, Kandidaatin tutkielma

Ohjaaja: Seppänen, Ville

Tämän kandidaatin tutkielman tarkoituksena oli selvittää kirjallisuuskatsauksen avulla, minkälaisia eettisiä kysymyksiä liittyy tekoälyn käyttöön julkisella sektorilla. Näiden kysymysten tarkempaa tarkastelua pohjustetaan tutkielmassa esiteltävillä tekoälyn, etiikan ja julkisen sektorin määritelmillä. Tekoälyn hyödyntäminen lisääntyy jatkuvasti myös julkisella sektorilla muuttaen prosesseja ja toimintatapoja. Tekoäly viranomaistoiminnassa nähdään usein eettisten ongelmien aiheuttajana, mutta se avaa myös useita positiivisia mahdollisuuksia esimerkiksi laadun parantamisen ja tehokkuuden lisäämisen kautta. Oikein käytettynä tekoäly voi edesauttaa eettisyyden ja yhteiskunnallisen hyväksyttävyyden kannalta positiivisia muutoksia. Tekoälyn kasvavan käyttöönoton myötä keskustelua on herättänyt nykylainsäädännön riittävyys, koska virallisia tekoälyyn liittyviä lakeja ei vielä ole säädetty.

Asiasanat: tekoäly, etiikka, kone-etiikka, läpinäkyvyys, julkinen sektori

ABSTRACT

Kinnunen, Iiro

Transparency of Artificial Intelligence in the Public Sector

Jyväskylä: University of Jyväskylä, 2020, 32 pp.

Information Systems Science, Bachelor's Thesis

Supervisor: Seppänen, Ville

The aim of this bachelor's thesis was to find out the ethical issues involved in using artificial intelligence in the public sector. The study is performed through a literature review. For a closer look at these issues, the paper presents definitions of artificial intelligence, ethics and the public sector. Utilization of artificial intelligence is constantly increasing in the public sector, while changing the processes and ways of acting. Artificial intelligence in governmental activities is often seen as a source of ethical problems, but at the same time it opens many positive opportunities, for example through the quality and efficiency improvement. It can also contribute to positive changes in ethical and social acceptability when used properly. Due to the increasing introduction of artificial intelligence, there has been debate about the sufficiency of current legislation, as no official laws have been enacted on artificial intelligence.

Keywords: artificial intelligence, ethics, machine ethics, transparency, public sector

KUVIOT

KUVIO 1 Tekoälyn taustatieteet ja sovellusalueet (Ailisto, 2018).....10

KUVIO 2 Tekoälyn lainsäädännöllinen perusta (Haataja & Latvanen, 2019)26

TAULUKOT

TAULUKKO 1 Tekoälyn tasot julkisella sektorilla (Stirling, 2017).....15

TAULUKKO 2 Tekoälyn etiikan kolme tasoa (Dignum, 2018).....20

TAULUKKO 3 Eettisen tekoälyn periaatteet (The IEEE, 2019)21

TAULUKKO 4 Eettinen toimintamalli julkisella sektorilla (Koivisto ym., 2019).....22

SISÄLLYS

TIIVISTELMÄ

ABSTRACT

KUVIOT

1	JOHDANTO.....	6
2	TEKOÄLY JA ETIIKKA.....	8
	2.1 Tekoäly	8
	2.2 Tekoälyn historia ja kehitys.....	10
	2.3 Etiikka.....	11
3	TEKOÄLY JULKISELLA SEKTORILLA.....	13
	3.1 Julkinen sektori	13
	3.2 Käyttökohteet ja -mahdollisuudet.....	14
	3.3 Tekoälyn hyödyntämisen haasteet.....	17
4	TEKOÄLYN EETTISYYDEN VAATIMUKSET	19
	4.1 Tekoälyn etiikka.....	19
	4.2 Eettiset periaatteet	21
	4.3 Läpinäkyvyys ja vastuut.....	23
	4.4 Lainsäädäntö ja valvonta	24
5	POHDINTA JA YHTEENVETO	27
	LÄHTEET	29

1 JOHDANTO

Perinteistä automaatiota käytetään jo laajasti viranomaistoiminnassa ja julkisella sektorilla. Automaattinen käsittely perustuu selkeisiin, lainsäädännöstä lähteviin laskentamalleihin, joihin ei liity lainkaan harkintaa. Tekoäly puolestaan kykenee suorittamaan myös harkintaa sisältäviä toimintoja, minkä vuoksi sen käyttö voi muuttaa koko viranomaisten palveluprosessia. Suurimmiksi muutoksiksi on tunnistettu esimerkiksi kontaktointitapojen monipuolistuminen, aika- ja paikkariippumattomuus, palveluprosessin nopeutuminen sekä asiakkaan parempi huomioiminen päätöksenteossa. Näiden lisäksi eri virastojen välinen yhteistyö, tarkistusprosessit, perustelujen esittäminen, viranomaisten kouluttaminen ja kansalaisten suhtautuminen viranomaisiin tulevat olettavasti muuttamaan tekoälyn myötä. Uhkana puolestaan nähdään asiakkaan ja viranomaisen heikentynyt kyky kyseenalaistaa asioita, kun kaikki päätöksenteko tapahtuu nopeasti algoritmien tukemana. (Koivisto ym., 2019).

Tekoälyn käyttöönottoon ja hyödyntämiseen sisältyy valtavasti potentiaalia ja muutosvoimaa. Tekoäly voi auttaa meitä ratkaisemaan globaaleja ongelmia, parantamaan hyvinvointia sekä kiihdyttämään talouskasvua (Työ- ja elinkeinoministeriö, 2019). Tekoälysovellusten lisääntyvä käyttö nostaa esiin myös monia eettisiä, yhteiskunnallisia ja juridisia haasteita yhteiskunnan eri tasoilla. Vaikka laajempia käyttöönottoja tehdään vielä harkiten, tuo tekoälyn lisääntyvä hyödyntäminen eteemme useita taloudellisia, koulutuksellisia ja turvallisuuteen liittyviä kysymyksiä, esimerkiksi läpinäkyvyydestä ja vastuista (Valtiovarainministeriö, 2018).

Tutkielmassa on pyritty keräämään tarkasteluun mahdollisimman korkealaatuisia lähteitä ja niitä valittaessa on otettu huomioon julkaisualustan laatu- luokitus, tutkimuksissa käytettyjen lähteiden laatu sekä niihin suoritettujen viittausten lukumäärä. Lisäksi tarkoituksena on ollut valita hyödynnettäväksi mahdollisimman monia, ja eri näkökulmista tutkimusaihetta käsitteleviä lähteitä, jotta niitä on voitu vertailla ja arvioida kriittisesti keskenään. Artikkelien valinnassa on otettu huomioon yllämainittujen rajausten lisäksi myös julkaisuvuosi. Aihe on ollut viime aikoina paljon esillä, minkä ansiosta siitä on

tarjolla suhteellisen paljon tuoretta tutkimusmateriaalia. Gartnerin hypekäyrä kuvaa uusien ja kehittyvien teknologioiden leviämistä markkinoille. Sen uusimman raportin mukaan muun muassa tekoälyn turvallisuus ja läpinäkyvyys tulevat olemaan strategisesti ratkaisevassa asemassa vuonna 2020 (Cearley ym., 2019).

Tutkielman tavoitteena on luoda selkeä kokonaiskuva tekoälystä julkisella sektorilla sekä sen eettisistä vaatimuksista. Tämän tavoitteen saavuttamiseksi tutkielma pyrkii vastaamaan kysymyksiin :

- *Millä tavalla julkisella sektorilla hyödynnetään tekoälyä?*
- *Millaiset ovat tekoälyn eettisyyden vaatimukset (julkisella sektorilla)?*

Tässä tutkielmassa tekoälyn hyödyntämistä ja sen lähtökohtia on pyritty tarkastelemaan erityisesti Suomen näkökulmasta. Tutkielma on toteutettu systemaattisena kirjallisuuskatsauksena ja se on koostettu Jyväskylän yliopiston IT-tiedekunnan raportointiohjeen mukaisesti. Lähdeaineistoa on kerätty pääosin erilaisista verkossa sijaitsevista tietokannoista, kuten Google Scholarista, Scopussesta, IEEE Xplore Digital Librarysta ja JYKDOK-hakupalvelusta.

Tutkielma koostuu johdannon lisäksi kolmesta sisältöluvusta, joiden jälkeen koostetaan yhteenveto katsauksen tärkeimmästä sisällöstä ja syntyneistä havainnoista. Johdannon jälkeisessä, toisessa luvussa käydään läpi tekoälyn ominaispiirteitä ja historiaa sekä etiikan keskeisiä teorioita. Kolmannen luvun alussa tarkastellaan julkisen sektorin ja hyvinvointivaltion määritelmiä, joita myöhemmissä alaluvuissa hyödynnetään tekoälyn soveltamismahdollisuuksien ja haasteiden käsittelyssä. Viimeinen sisältöluke pitää sisällään tekoälyn etiikan jäsentelyn, sekä aiemmissa tutkimuksissa ilmenneitä eettisyyden vaatimuksia ja esiteltyjä periaatteita erityisesti läpinäkyvyyden, vastuullisuuden sekä lainsäädännön näkökulmista.

2 TEKÖÄLY JA ETIIKKA

Tässä luvussa käsitellään tekoälyä, sen kehitystä sekä etiikkaa ja sen eri osa-alueita. Tarkoituksena on luoda riittävä, mutta tiivis selvitys tutkielman kannalta keskeisistä avainsanoista, jotka toimivat pohjana tarkasteltavan aiheen perusteellisemmalle tarkastelulle. Tekoälyn ja etiikan yhdistelmästä muodostuvan kone-etiikan kenttää jäsennetään myöhemmin omassa alaluvussa 4.1.

2.1 Tekoäly

Tekoäly eli keinoäly on tietokone tai tietokoneohjelma, joka kykenee suorittamaan älykkäinä pidettäviä toimintoja. Tekoälyn avulla koneet, laitteet, ohjelmat, järjestelmät ja palvelut voivat toimia tehtävän ja tilanteen mukaisesti järkevällä tavalla (Russell & Norvig, 2016). Tekoälylle ei ole toistaiseksi olemassa, formaalia, yksiselitteistä määritelmää, koska älykkyyttä ja oppimista on vaikea määrittellä. On tärkeä huomata, ettei älykkyys itsessään ole yksi ulottuvuus tai asteikko, kuten vaikkapa lämpötila.

Nilsson (2009) kuvailee kirjassaan *The Quest for Artificial Intelligence* keino-älyä toimintana, jonka on tarkoitus tehdä koneista älykkäitä, ja älykkyyttä ominaisuutena, joka antaa tälle kokonaisuudelle mahdollisuuden toimia asianmukaisesti ja ennakoivasti ympäristössään. Oman ehdotuksensa tekoälyn määrittelyyn on laatinut myös Gordon (2011) kirjassaan *Artificial Intelligence: Approaches, Tools and Applications*: tekoäly voidaan esittää kokoelmana analyttisiä työkaluja, jotka kollektiivisesti yrittävät jäljitellä elämää, ja jotka helpottavat sellaisten ongelmien ratkaisemista, joita oli aiemmin vaikea tai mahdoton ratkaista.

Tekoäly ei siis ole ainoastaan yksi teknologia, vaan nimikkeen alle kuuluu joukko erilaisia menetelmiä, teknologioita, sovelluksia ja tutkimussuuntia. Tekoäly ja sen menetelmät, teknologiat ja sovellukset ovat osa digitalisaation laajempaa ilmiötä ja kehitystä (Ailisto, 2018).

Tekoälyn määrittelyssä hyödynnetään yleisesti kahta tai joissain tapauksissa kolmea ominaisuutta, jotka yhdistävät kaikkea tekoälyä:

- Autonomisuus
 - Kyky suorittaa tehtäviä monimutkaisessa ympäristössä ilman jatkuvaa käyttäjän ohjausta.
- Adaptoituminen
 - Kyky kehittää toimintakykyä oppimisen kautta.
- Suorituskyvyn laaja-alaisuus
 - Kyky soveltua ja laajentua eri alueille.

Kaikkia kolmea ominaisuutta yhdistävät peruskysymykset vastuusta, sekä moraalisesta että juridisesta läpinäkyvyydestä. Tekoälyn tulisi osata tunnistaa erilaisia tilanteita, toimia ilman ihmisen ohjausta ja suoriutua tehtävistä perustellulla tavalla. Autonomisuuden, oppivuuden ja suorituskyvyn lisäksi tekoälyteknologioita voidaan arvioida avoimuuden, riskien ja pätevyyden suhteen (Ailisto, Heikkilä, Helaakoski, Neuvonen & Seppälä, 2018).

Nykyinen tekoäly koostuu lähinnä erityisiin tiedonkäsittelytehtäviin tarkoitettuista ohjelmistoista. Tällä hetkellä olemassa olevat kehittyneimmätkin tekoälysovellukset edustavat niin sanottua heikkoa tekoälyä, sillä ne toimivat ainoastaan siihen ohjelmoidun logiikan perusteella ja vain siinä tietyssä tehtävässä, johon ne on kehitetty. Heikolla tekoälyllä ole siis sisällöllistä ymmärrystä käsittelemästään datasta, eikä se pysty asettamaan itselleen tiedollisia tai moraalisia päämääriä (Davidson, 2019).

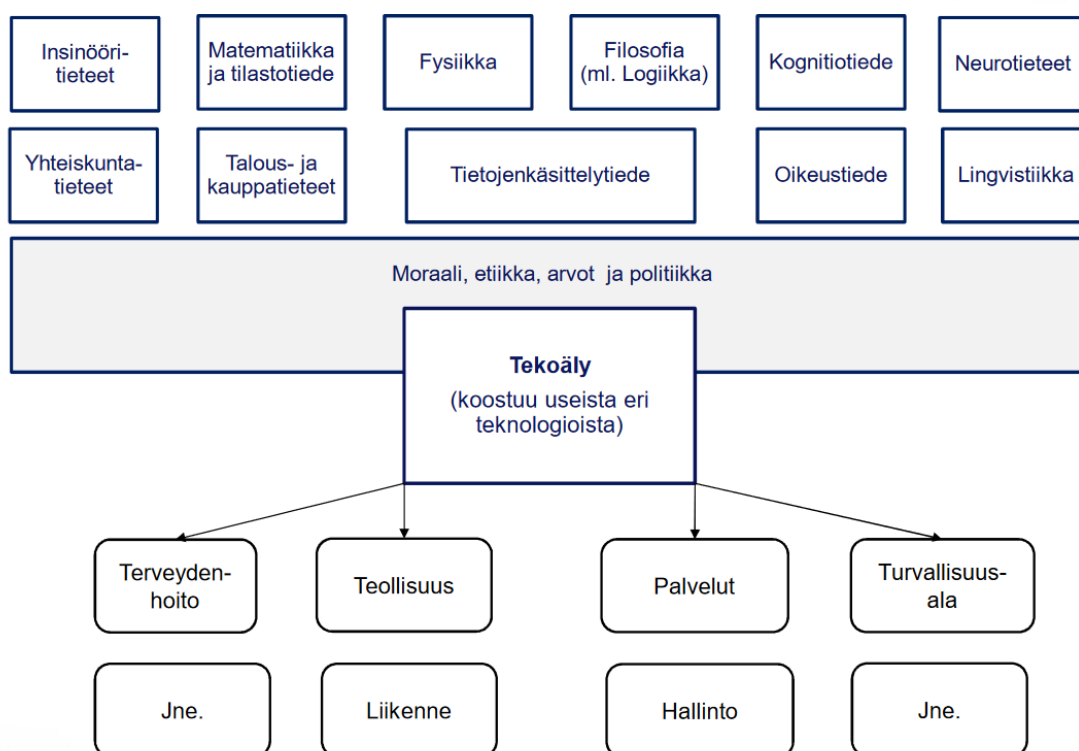
Seuraavana edistysaskeleena pidetään vahvaa tekoälyä, jota ei vielä tähän päivään mennessä ole onnistuttu kehittämään. Vahva (ts. yleinen) tekoäly kykenee ymmärtämään asioita ja niiden välisiä yhteyksiä selvästi laajemmin kuin nykyinen heikko tekoäly. Vahvan tekoälyn ominaispiirteisiin kuuluu maalaisjärki (englanniksi *common sense*) ja jonkin asteinen ymmärrys ympäröivästä maailmasta (Ailisto, 2018). Uskaliaimmissa näkemyksissä puhutaan vahvaa tekoälyä seuraavasta supertekoälystä, singulariteetista, joka tulee kyseeseen siinä vaiheessa, kun tekoäly ohittaa ihmisen kyvykkyyden.

Tekoälyn viimeaikainen eksponentiaalinen kasvu ja näkyvyys perustuvat syvien neuroverkkojen ja suurten datamassojen käyttöön, suorittimien laskentatehon ja oppivan tekoälyn harjoittamiseen käytettävän datan määrään kasvuun sekä algoritmien parantuneeseen saatavuuteen (Ailisto, 2018). Neuroverkoilla tarkoitetaan malleja, jotka perustuvat yhdistävään laskentaan. Monikerroksisten neuroverkkojen avulla tekoälyn keskeinen oppimisjärjestelmä, eli syväoppiminen, pyrkii ratkaisemaan sille asetetut ongelmat (Patterson & Gibson, 2017).

Tekoälyn olemus herättää keskustelua erilaisista filosofisista kysymyksistä. Voidaan muun muassa pohtia sitä, edellyttääkö älykäs toiminta mielen tai tietoisuuden olemassaoloa ja sitä, missä määrin tietoisuutta on mahdollista tuottaa

laskennallisesti. Toisaalta, McCarthy (2006) uskoo, ettei tekoälyn filosofialla ole sen enempää vaikutusta tekoälyn harjoittamiseen käytännössä, kuin tieteen filosofialla on tieteen harjoittamiseen.

Tekoälyn nähdään kirjallisuudessa liittyvän laajaan kirjoon eri tieteenaloja, kuten alla esiteltävästä kuvioista ilmenee (kuvio 1). Tämän tutkielman kannalta olennaisia tieteenaloja ovat erityisesti yhteiskuntatieteet, tietojenkäsittelytiede sekä oikeustiede. Teknologisen kehityksen lisäksi tekoäly vaikuttaa koko yhteiskuntaan ja ihmisiin, jolloin tarkastelun näkökulmia ovat moraalitieteet, etiikka, arvot ja politiikka sekä yhteiskunta-, oikeus-, talous- sekä kauppatieteet. Kaavio on tarkoituksenaan havainnollistaa tekoälyn sijoittumista eri tieteenalojen ja sitä hyödyntävien sektoreiden kentälle (Ailisto ym., 2018).



KUVIO 1 Tekoälyn taustatieteet ja sovellusalueet (Ailisto, 2018)

2.2 Tekoälyn historia ja kehitys

Tekoälyä alettiin kehittää ensimmäisen kerran 1950-luvulla eri tutkijoiden toimesta. Yleisesti tekoälyn isänä pidetään matemaatikko Alan Turingia, jonka kehittämä Turingin testi oli ensimmäinen menetelmä koneiden älykkyyden mittaamiseen. Turingin testissä kone asetetaan keskustelemaan ihmisen kanssa ja se läpäisee testin siinä vaiheessa, kun sen vastauksia ei pystytä erottamaan ihmisen vastauksista (Turing, 1950). Samanlaisia mielipiteitä ajattelevista koneista esitettiin myös Dartmouthin konferenssissa vuonna 1956, joka on yleisesti tunnustettu tekoälyn alan ja termin viralliseksi syntymäpäiväksi (Crevier & Lepage,

1993). Dartmouthissa Allen Newell, Herbert A. Simon ja Cliff Shaw esittelivät Logic Theorist -tietokoneohjelmansa, jota kutsutaan ensimmäiseksi tekoälyohjelmaksi (Russell ym., 2016). Se oli samalla myös ensimmäinen ohjelma, joka suunniteltiin tarkoituksellisesti jäljittelemään ihmiselle tyypillisiä ongelmanratkaisumenetelmiä ja se kykeni todistamaan matemaattisia lausekkeita lahjakkaan matemaatikon tasolla.

Tekoälyn kehitystä voidaan havainnollistaa jaottelemalla se teknologia-aaltoihin. Launchbury (2017) ehdottaa jakoa kolmeen aaltoon:

1. Käsin rakennetut toteutukset
2. Tilastollinen oppiminen
3. Tilanteeseen mukautuva oppiva tekoäly

Ensimmäistä aaltoa nimitetään myös klassiseksi tai symboliseksi tekoälyksi, koska se oli tekoälytutkimuksen keskiössä aina 1960-luvulta 1980-luvulle asti. Nykyinen teknologia on tilastollista oppimista, jota edustavat koneoppiminen ja syvät neuroverkot. Kolmannen aallon, tilanteeseen mukautuvan oppivan tekoälyn odotetaan nousevan 2020-luvulla. (Työ- ja elinkeinoministeriö, 2019).

Vaikka tekoälyllä on pitkäikäinen historia ja sen käyttö lisääntyy jatkuvasti eri toimialoilla, siihen liittyvät eettiset kysymykset ovat jääneet vähemmälle huomiolle. Tällä hetkellä tekoälyn tutkimus keskittyy pohtimaan esimerkiksi sitä, voiko tekoäly oppia toimimaan tavoilla, jotka eivät kestä eettistä tarkastelua. (Bostrom & Yudkowsky, 2011). Tämän vuoksi on mietitty mahdollisuutta opettaa tekoälylle etiikkaa, mutta toistaiseksi ongelmaksi on muodostunut se, ettei ihminen kykene täysin ymmärtämään algoritmeja, joiden avulla kehittynein tekoäly oppii. Tekoälyn kehityksen yksi seuraavista askeleista on opettaa tekoälylle selkeää viestintää ihmisen kanssa.

2.3 Etiikka

Etiikka on filosofinen oppiala, joka tutkii ja argumentoi moraalialia sekä siihen keskeisesti kuuluvia kysymyksiä, kuten esimerkiksi eettisen toiminnan periaatteita, oikeaa ja väärää, hyvää ja pahaa, sekä arvojen ja eettisten väittämien luonnetta. Etiikka on siis moraalin teoreettista pohdiskelua, kun taas moraalii on käytännön hyvän elämän tavoittelua: arvoja, käytäntöjä ja toimintatapoja (Ollila, 2019). Etiikka näin ymmärrettynä on moraaliililosofiaa, joka tutkii filosofisesti kiinnostavia asioita moraalisisissa ilmiöissä (Pietarinen, 2015). Etiikka liittyy yleensä keskeisesti ajatusrakenteisiin ja vastaukset eettisiin kysymyksiin riippuvat usein vastauksista todellisuuskäsitystä koskeviin kysymyksiin (Crisp, 1998). Länsimainen etiikka jakaantuu kolmeen suurempaan kokonaisuuteen, jotka ovat metaetiikka, normatiivinen etiikka ja soveltava etiikka. Näiden osalueiden lisäksi filosofisen etiikan alaan voidaan laskea mukaan myös deskrip-

tiivinen etiikka, joka on kiinnostunut ihmisten eettisen toiminnan ja arvojen soveltamisesta käytännössä (Stahl, Timmermans & Mittelstadt, 2016).

Meta-etiikka tutkii moraalisten käsitteiden luonnetta ja niiden pätevyyden ehtoja sekä eettisten käsitteiden alkuperää ja merkitystä. Metaetiikka tutkii moraalifilosofian kysymyksiä ja mahdollisuuksia ennen kaikkea ontologian, epistemologian sekä filosofisen ja empiirisen moraalipsykologian näkökulmista (DeLapp, 2007). Sen tarkoituksena ei varsinaisesti ole selvittää sisällöllisiä moraaliperiaatteita, vaan päämääränä on selvittää, minkälaista moraalinen ajattelu on. Metaetiikan tärkeimpiä suuntauksia ovat intuitionismi, emotivismi, relativismi, realismi, preskriptivismi ja konsensualismi (Tieteen termipankki, 2020).

Normatiivinen etiikka tutkii eettisiä normeja ja sitä, miten yksilön tai yhteisön tulisi toimia. Se pyrkii vastaamaan kysymyksiin liittyen siihen, mitkä ovat oikeat eettiset säännöt, joita yksilön tulisi noudattaa. Normatiivisen etiikan teorit luovat, osoittavat ja puolustavat aina joitakin tiettyjä moraalisia sääntöjä ja arvoja. Normatiiviset teorit jaetaan tavallisesti kolmeen luokkaan: velvollisuusetiikka, seurausetiikka ja hyve-etiikka (Fieser, 2000).

Soveltava etiikka pyrkii soveltamaan teoreettista etiikkaa ja sen eri teorioita, kuten utilitarismia, yhteiskuntasopimusteorioita tai velvollisuusetiikkaa tosielämän ongelmiin. Soveltavaan etiikkaan kuuluvat muun muassa lääketieteen etiikka, ympäristöetiikka, sodan etiikka, tietokone-etiikka, liike-elämän etiikka ja yritysten yhteiskuntavastuu (Marturano, 2002). Tässä tutkielmassa painoarvoa kiinnitetään erityisesti tietokone-etiikkaan, jota voidaan osaltaan soveltaa myös tekoälyyn.

Deskriptiivinen etiikka, eli kuvaileva etiikka käsittelee moraalisia instituutioita ja moraaliteorioita sekä arvioi niiden sisäistä perusteltavuutta. Sen tarkoituksena on kuvata ja selittää moraalisia käsityksiä. Deskriptiivinen etiikka ei pyri tarjoamaan eettisiä ratkaisuja kuten normatiivinen etiikka, vaan välineitä näiden eettisten ratkaisujen löytämiseksi. Deskriptiivisessä etiikassa tavoitellaan neutraaliutta (Fieser, 2000).

3 TEKOÄLY JULKISELLA SEKTORILLA

Tässä kappaleessa käydään läpi esimerkkejä julkisella sektorilla käytettävästä tekoälystä sekä sen mukanaan tuomista mahdollisuuksista ja mahdollisista ongelmista. Ensimmäisessä alaluvussa käsitellään julkisen sektorin erityispiirteitä, roolia sekä siihen kohdistuvia odotuksia yhteiskunnassa. Sen tarkoituksena on luoda mahdollisimman selkeä kokonaiskuva julkisen sektorin toiminnasta, jotta voimme lopulta tarkastella tekoälyn hyödyntämistä julkisen sektorin näkökulmasta.

3.1 Julkinen sektori

Julkista sektoria koskevaa tutkimusta läpikäydessä on huomattavissa, että sekä julkisen sektorin määritelmää että julkisen sektorin toimialueita pidetään niin itsestään selvinä, ettei niitä ole useinkaan erikseen määritelty. Yleisesti julkisella sektorilla tarkoitetaan sitä kansantalouden osaa, joka on kuntien tai valtion omistuksessa. Valtion ja kuntien perinteisenä pidetyn toiminnan lisäksi myös muun muassa julkiset sosiaaliturvarahastot, kuntayhtymät, liikelaitokset ja valtionyhtiöt lasketaan kuuluvan julkisen sektorin piiriin. Julkinen hallinto edustaa eri hallinnon tasoilla toimivia julkisista toimenkuvista vastaavia organisaatioita (Harisalo, Aarrevaara, Stenvall & Virtanen, 2007).

Käsite julkinen voidaan määritellä myös demokratiana tai kansanvaltana. Julkinen hallinto toimii demokraattisessa ohjauksessa ja valvonnassa, mikä erottaa sen esimerkiksi yksityisistä yrityksistä ja vapaaehtoisesta kansalaistoiminnasta. Keskeistä julkisen sektorin toiminnassa on, että se perustuu lakeihin, sääntöihin ja normeihin. Hallinnollisen sääntelyn päämääränä on ohjata muun muassa päätöksentekoa ja sen tarkoituksena on luoda julkisen sektorin toiminnasta avointa, oikeudenmukaista ja ennakoivaa (Harisalo ym., 2007).

Yleisesti ottaen julkisen sektorin organisaatioiden ajatellaan toimivan niin, että enemmistö kansalaisista hyötyy niiden toiminnasta ja niiden tarjoamat palvelut ovat kaikkien saatavilla (Sydänmaanlakka, 2015). Julkinen sektori rahoite-

taan suurelta osin veroilla. Lähes kaikkialla maailmassa julkinen sektori nähdään parhaana tapana toteuttaa järjestystä ja turvallisuutta, esimerkiksi oikeuslaitosten, poliisin ja armeijan kautta. Yleensä julkiseen sektoriin kuuluu useita muitakin toimia, kuten koulutusta, terveydenhuoltoa ja ympäristönsuojelua (Honkanen, 2015).

Paul Spickerin (2000) hyvinvointivaltion yleisessä teoriassa puhutaan hyvinvoinnin edistämisestä, turvaamisesta, tarjoamisesta, tuottamisesta ja uudelleenjakamisesta. Vaikka näiden käsitteiden sisällöt ovat hieman epätäsmällisiä, voidaan niiden katsoa määrittävän julkisen sektorin hyvinvointitehtävää.

Julkinen hyvinvointivastuu nähdään usein pitkälle oikeudellistuneen ja vastuullisen yhteiskunnan käsitteenä. Sen onnistumista arvioidaan sen perusteella, toteutuuko se lakien tai jonkin eettisen koodiston määrittelemällä tavalla. Oikeudellisessa mielessä julkisella hyvinvointivastuulla tarkoitetaan esimerkiksi perustuslailla julkiselle vallalle säädettyä velvollisuutta turvata ja edistää yksilöiden sosiaalisia oikeuksia, toimeentuloa ja palveluja. (Julkunen, 2006).

Työ- ja elinkeinoministeriön (2019) tekoälyohjelman loppuraportin mukaan tekoälyaikana hyvinvointiyhteiskunnan kehittäminen edellyttää rohkeita valintoja ja riittäviä investointeja. Hyvin toteutetut julkiset palvelut mahdollistavat ihmiskeskeisen ja ennakointikykyisen yhteiskunnan. Tällaisessa yhteiskunnassa ihminen oppii ymmärtämään omaa hyvinvointiaan, ja palvelut haakeutuvat ihmisen luokse oikea-aikaisesti, tietoturvallisesti ja eettisesti kestäväällä tavalla. Sekä julkisen sektorin että lainsäätäjien tehtävänä on poistaa esteitä ja luoda mahdollisuuksia kasvulle.

3.2 Käyttökohteet ja -mahdollisuudet

Täysin automatisoidun päätöksenteon sijaan tekoälyn käyttämisen uskotaan ensivaiheessa voivan hyödyttää viranomaistoimintaa mahdollistamalla entistä laajemman algoritmeihin ja laskentaan perustuvan tietopohjan päätöksenteon tueksi. Samalla voidaan välttyä yksinomaan tekoälyyn nojautuvan päätöksenteon opetusmateriaalin ja mallien virheiden synnyttämiltä tai näihin kohdistuvien tahallisten väärinkäytösten vaikutuksilta päätöksentekoon (Kääriäinen ym., 2018).

Tekoälyn hyödyntämisen selkeistä mahdollisuuksista huolimatta se ei ole ratkaisu hallinnon systemaattisiin ongelmiin ja se voi mahdollisesti jopa pahentaa palveluiden toimittamiseen, yksityisyyteen tai etiikkaan liittyviä ongelmia, ellei sitä toteuteta harkitusti ja strategisesti. Tekoälyn toteuttamisesta kiinnostuneet virastot voivat oppia paljon hallitusten aiemmista muutospyrkimyksistä sekä tekoälyä hyödyntäviltä yksityisen sektorin toimijoilta (Mehr, 2017). Hänellä on tarjota myös oman kuusijakoisen strategiansa hallitusvirastoille, jonka avulla tekoälyä voidaan harkita sovellettavan julkisella kentällä. Tämän strategian keskeiset tekijät ovat tekoälyn rakentaminen osaksi kansalaiseskeistä ohjelmaa, kansalaisten panoksen ja olemassa olevien resurssien hyödyntäminen, tietojen käsittelystä ja yksityisyyden suojasta huolehtiminen, eettisten riskien ja

tekoälyn suorittaman päätöksenteon vähentäminen sekä työntekijöiden lisääminen.

Stirling (2017) esittää oman viisitasoisen viitekehjoksensä tekoälyn hyödyntämisestä julkisella sektorilla, jossa tasolta seuraavalle siirryttäessä tekoäly suorittaa tehtäviä yhä itsenäisemmin (taulukko 1). Tämän taulukon tarkoituksena on auttaa ymmärtämään tekoälyn monipuolista soveltamista julkisella sektorilla.

TAULUKKO 1 Tekoälyn tasot julkisella sektorilla (Stirling, 2017)

Taso 5	Täysin automatisoitu	Tekoälyjärjestelmä toimii täysin itsenäisesti, eikä vaadi ihmisen väliintuloa missään tilanteessa.
Taso 4	Automaattinen	Tekoäly toimii automaattisesti ja itsenäisesti niin kauan, kunnes vastaan tulee äärimmäinen tilanne, joka vaatii ihmisen tulkintaa.
Taso 3	Puoliautomaattinen	Tekoäly suorittaa rutiinomaisia tehtäviä ilman valvontaa, mutta hälyttää apua, kun ihmisen työpanosta tarvitaan.
Taso 2	Tiukka valvonta	Ihminen valvoo toimintaa, on valmis puuttumaan tehtävään ja suorittaa vaikeimmat tilanteet tekoälyn suorittaessa rutiinomaisia työtehtäviä.
Taso 1	Yksinkertaisen tuen lisäys	Ihminen tekee suurimman osan työstä, mutta tekoälyä hyödynnetään esimerkiksi tiedon syötössä, prosessoinnissa ja identifioinnissa.
Taso 0	Ei automaatiota	Tekoälyä ei hyödynnetä, vaan julkisen sektorin palvelut tuotetaan kokonaan ihmisvoimin.

Tasojä tulkittaessa nähdään, että tekoälyjärjestelmien vastuualueet kasvavat siirryttäessä toimintaa tukevasta tekoälystä autonomisempaan päätöksentekoon. Tästä syystä tarvitaan sääntelyä ja lainsäädäntöä sekä myös eettisiä ja moraalisia sääntöjä järjestelmien kehittämiseen, operointiin ja ylläpitoon (Kääriäinen ym., 2018).

Julkishallinnon tekoälysovellukset kohdistuvat toistaiseksi suurelta osin kyselyihin vastaamiseen, asiakirjojen täyttämiseen ja etsimiseen, pyyntöjen ohjaukseen, käännöstyöhön ja dokumenttiluonnosten laatimiseen. Potentiaalisim-

pina tekoälyn sovelluskohteina nähdään ne, joissa tekoälyllä voidaan vähentää byrokratiaa, ratkaista resursointiongelmia tai suorittaa erityisen monisyisiä tehtäviä avustavasti. On huomattu, että julkisen sektorin tekoälymahdollisuudet voivat vähentää hallinnollista taakkaa ja vapauttaa viranomaisia vaativampiin tai ihmisläheisempiin tehtäviin. Kun julkinen sektori pystyy tarjoamaan kansalaisille heidän tarvitsemiaan digitaalisia palveluja, voi tekoälystä tulla yksi tapa kansalaisten osallistamiseen ja julkisten palveluiden kehittämiseen (Mehr, 2017)

Tekoälyn soveltamisen osalta Suomessa on edetty pääosin pilottien avulla. Erilaisia hankkeita on meneillään sekä kuntasektorilla että ministeriöissä (Klemettinen & Peltola, 2018). Esimerkkinä hankkeista ovat kokeilut ja käyttö Kansaneläkelaitoksella, Maahanmuuttovirastolla ja Verohallinnolla. Terveystieteiden puolella Helsingin ja Uudenmaan sairaanhoitopiirillä on meneillään kymmeniä tekoälyhankkeita (Vierula, 2018). Myös Oulun kaupungilla on havaittu, että tekoälyä voidaan hyödyntää asiakaspalvelussa ja avustavassa roolissa erilaisissa lupaprosesseissa sekä päätöksenteossa (Kääriäinen ym., 2018). Kuntaliiton vuonna 2018 teettämä kuntien tietotekniikkakartoituskysely osoitti, että tekoälyratkaisut tulevat olemaan ajankohtaisia asioita vasta kolmen vuoden päästä, eli vuonna 2021. Kysely lähetettiin kuntien ja kuntayhtymien tietohallintoon ja sen osatarkoituksena oli selvittää osallistuneiden näkemyksiä uusien teknologioiden mahdollisuuksista kuntasektorilla osana palvelutuotantoa (Hyvärinen & Parviainen, 2018).

Suomi on lähtenyt Euroopan kärkijoukoissa tekemään kansallista tekoälytyötä elinkeinoministeri Mika Lintilä käynnistettyä Tekoälyohjelman vuonna 2017. Tekoälyohjelman tavoitteena on viedä Suomi tekoälyn soveltamisen kärkimaiden joukkoon ja osoittaa suuntaa Suomen kansalliselle kehitykselle sekä osallistua kansainväliseen keskusteluun (Työ- ja elinkeinoministeriö, 2019). Osana Tekoälyohjelmaa Valtiovarainministeriö on käynnistänyt kansallisen tekoälyohjelma AuroraAI:n, joka pohjautuu nykyisen pääministeri Sanna Marinin hallitusohjelman strategiseen tavoitteeseen elinvoimaisesta Suomesta. AuroraAI:n päämääränä on vauhdittaa julkisen hallinnon siirtymistä tekoälyaikaan turvallisesti ja eettisesti sekä mahdollistaa kansalaisille erilaisten palveluntuottajien palveluista koostuvat sujuvat kokonaisuudet erilaisiin elämäntilanteisiin. AuroraAI -ohjelma pyrkii luomaan tekoälyn hyödyntämiseen pohjautuvat toimintamallit mahdollistamaan Suomeen maailman parhaan julkisen hallinnon. Tämä edellyttää, että julkiset organisaatiot sekä muiden sektorien palvelut kytketään yhteen tekoälyä hyödyntävän AuroraAI -verkon avulla (Valtiovarainministeriö, 2019).

Klemettisen ja Peltolan (2018) mukaan tekoälyn nykyistä laajempi käyttöönotto julkisella sektorilla edellyttää ainakin kansalaiskeskustelun käynnistämistä, mittavaa ja avointa hallinnonalojen ja kuntien yhteistyötä sekä kustannustehokkuuden huomioimista. Lisäksi esiin nousee perus- ja ihmisoikeuksien huomioiminen, jonka mukaan oppivat ja kehittyvät algoritmit eivät saa heikentää marginaaliryhmien tilannetta.

3.3 Tekoälyn hyödyntämisen haasteet

Keinotekoisien älykkyyden hyödyntämistä julkisella sektorilla arvioidaan ristiriitaisilla tavoilla. Vaikka keskustelu sen tuomista vaaroista ja eduista on lisääntynyt, empiiristä tutkimusta kummankaan näkökulman tukemiseksi on hyvin vähän. Avainkysymyksenä tekoälyn hyödyntämisessä lienee se, millä ehdoilla päätöksenteko voidaan siirtää ihmiseltä järjestelmälle tai jollekin tekniselle ratkaisulle, kuten tekoälylle. Automatisoitavan, tekoälylle siirrettävän tehtävän luonne on tällöin tarkastelun keskipisteenä (Rusthollkarhu, 2019). Tietyn toiminnan eettinen arvioiminen vaatii syventymistä tarkastelun kohteena olevaan tapaukseen, sillä moraalisesti väärät toimintamallit ovat yleensä helposti tunnistettavissa. Eettinen ongelma syntyy silloin, kun kysymykset ovat monisyisiä ja päätös on tehtävä kahden oikealta tuntuvan vaihtoehdon välillä (Koivisto ym., 2019).

Tänä päivänä tekoälyn käyttöönottoa viranomaistoiminnassa rajoittavat vielä esimerkiksi julkishallinnon rajalliset resurssit sekä tarve säilyttää ihmisten luottamus hallintoon (Mehr, 2017). Kansalaisten suhtautuminen tekoälyn käyttöön muuttuu sitä kielteisemmäksi, mitä tuntemattomampi sovelluskohde on kyseessä tai mitä itsenäisempi päätöksentekoasema sille annetaan. Kansalaisten näkökulmasta tekoälyn hyväksyntää lisääviä tekijöitä ovat selkeän sääntelyn laatiminen, päätöksenteon avoimuus ja läpinäkyvyys sekä riittävän ihmiskontrollin takaaminen (Koivisto ym., 2019). Suomessa hidasteena ja ongelmana nähdään jokseenkin myöhäinen herääminen digitalisaation muutostarpeisiin, pienen talouden rajalliset investointimahdollisuudet sekä tekoälyn ja tietotalouden osaamisvajae. Myös harvaan asutun Suomen infrastruktuuri voi osaltaan vaikeuttaa esimerkiksi palveluiden ja liikenteen automaation kehitystä. (Valtiovarainministeriö, 2018).

Tekoälyyn liittyy kahdenlaisia tiedollisia riskejä, helppoja ja vaikeita. Helppona tiedollisena riskinä voidaan nähdä datan vinoutuminen. Järjestelmä ei itse tunnista vinoutumista, mutta sitä käyttävä ihminen voi havaita sen helposti ja korjata tilanteen. Vaikeita tiedollisia ongelmia ovat monimutkaiset riskitilanteet, joissa järjestelmässä itsessään on ominaisuus, jota on ulkoapäin vaikea havaita. Nämä tilanteet johtuvat siitä, että tekoälyjärjestelmässä käytetyt neuroverkot sisältävät niin paljon piilokerroksia, ettei tiedetä, miksi järjestelmä toimii niin kuin se toimii. Tällöin on kyse niiden läpinäkymättömyydestä. Tämän takia tekoälyn tutkijatkaan eivät ole valmiita ulkoistamaan kokonaisvaltaista ihmistä koskevaa päätöksentekoa koneelle, vaan tekoälyn halutaan toimivan vain päätöksenteon tukena. (Rusanen & Koskinen, 2018). Neuroverkkojen heikkouksiksi on tunnistettu myös niiden suuri energian ja laskentatehon tarve sekä herkkyyys syöttötiedon muutoksille ja virheille. Nämä ongelmat ovat oikeustieteen näkökulmasta kriittisiä ja voidaan nähdä tekoälyn hyödyntämistä hankaloittavina tekijöinä viranomais- ja tuomiovallan toiminnassa (Ailisto, 2018).

Keskeisimmiksi tekoälyä hyödyntävien päätöksentekojärjestelmien haasteiksi on tunnistettu muun muassa viranomaisten puutteellinen keskinäinen

yhteistyö informaationvälityksessä sekä tietovarantojen vähäisyys. Ongelmallisenä nähdään myös algoritmisen päätöksentekoprosessin pirstaloituminen eri verkostotoimijoille, koska tekoälyjärjestelmien kohdalla ei aina voida osoittaa, kenen tulisi vastata järjestelmän mahdollisesta virheestä. Tämän kaltaisten riskitilanteiden ehkäisemiseksi vastuutahot tulisi eritellä tarkasti ja kansalaisille olisi taattava asianmukaiset oikeussuojamahdollisuudet (Koulu, Mäihäniemi, Kyyrönen, Hakkarainen & Markkanen, 2019).

Vakavimpia tekoälystä aiheutuvia uhkakuvia kartoitettaessa on kuitenkin otettava huomioon, että toistaiseksi julkisen sektorin tekoälysovellukset ovat hyvin epäitsenäisiä ja toimivat suurelta osin ihmisen valvonnan alaisina. Lähes mitä tahansa teknologiaa voidaan käyttää väärin ja sen mahdollisuus täytyy huomioida myös tekoälyn kanssa. Väärinkäytön todennäköisyys nähdäänkin usein paljon suurempana kuin sen, että ihminen menettäisi tekoälyn hallinnan (Torresen, 2018). Tärkeää on ymmärtää, etteivät kaikki tekoälyn virheet välttämättä aiheuta vakavia seurauksia, ja että myös ihmisen päätöksenteossa voi olla taipumusta puolueellisuuteen. Puutteellisesta tekoälylainsäädännöstä huolimatta ei tekoälyn suhteen olla oikeudellisessa tai lainsäädännöllisessä tyhjiössä, vaan esimerkiksi Euroopan Unionin lainsäädännössä on vakiintuneet periaatteet ja lainsäädännölliset vaatimukset tietojen oikeudenmukaiselle ja vastuulliselle käytölle, kansalaisoikeuksille sekä julkisen sektorin toimille (Copeland, 2019).

4 TEKOÄLYN EETTISYYDEN VAATIMUKSET

Tässä luvussa käsitellään tekoälyn etiikkaa, tekoälyn eettisyyden vaatimuksia ja periaatteita sen läpinäkyvyyden, vastuiden sekä lainsäädännöllisten asetusten näkökulmasta. Kyse on sekä eettisestä että juridisesta perusasiasta. Luotettavan tekoälyn olisi paitsi oltava kaikkien sovellettavien lakien ja asetusten mukainen, myös täytettävä sille asetettavat vaatimukset.

4.1 Tekoälyn etiikka

Vaikka roboteilla ja tekoälyllä ei vielä toistaiseksi ole sellaisia kognitiivisia kykyjä, että niitä voitaisiin pitää itsenäisinä tai moraalisisina toimijoina, on kirjallisuudessa huomattavissa konsensus siitä, että niiden tekemien valintojen tulisi vastata eettisiä arvojamme. Robottien moraalialia on pohdittu pitkään ja jopa ennen kuin ensimmäistäkään robottia oli edes valmistettu. Tieteiskirjailija Isaac Asimov kehitti jo vuonna 1942 robotiikan eettisen säännösten kirjojen maailmaan, jossa lain tärkeimpänä tarkoituksena oli estää robotteja vahingoittamasta ihmisiä missään tilanteessa (Asimov, 1942).

Tekoälyn etiikan (ts. kone-etiikan) määrittely nähdään usein haasteellisena tutkimukselle. Tekninen tekoäly ja filosofinen etiikka yhdessä muodostavat luonteeltaan niin monitieteisen ja metodologisesti rikkaan tutkimuskentän, ettei alalle ole vielä vakiintunut yksiselitteistä käsitystä siitä, mistä tekoälyn etiikka tarkalleen ottaen koostuu tai mistä sen tulisi koostua (Vakkuri & Abrahamsson, 2018).

Tekoälyn etiikka eroaa merkittävästi perinteisestä teknologian etiikasta, jossa moraalisten valintojen tekijänä nähdään yleensä teknologiaa suunnitteleva ihminen (Polonski, 2017) ja tutkimuksen kohteena ovat teknologian suorat tai välilliset vaikutukset ihmisiin ja yhteiskuntaan. Tekoälyn etiikkaa tarkasteltaessa täytyy ottaa huomioon myös sen kyky kehittyä, oppia sekä soveltua erilaisiin muuttuviin tekijöihin. Oppimiskyky voi johtaa tekoälyn sellaisiin tilanteisiin, jotka ovat mahdollisesti vain kaukaisesti yhteydessä alkuperäiseen ja suunnitel-

tuun käyttötarkoitukseen. Yleisesti nähdäänkin, että koneen itsenäisesti tekemät päätökset ovat tekoälyn etiikan keskiössä, koska ne haastavat ihmisen ja koneen välisen suhteen perustavanlaatuisella tavalla. Keskeinen kysymys on, kuinka etiikkaa tulisi opettaa itsenäisiä päätöksiä tekevälle tekoälylle. Koneetiikka puolestaan tarkastelee asiaa koneen näkökulmasta (Anderson & Anderson), jolloin pohdittavana on, kuinka tekoälyä opetetaan tekemään eettisiä valintoja. Oikeustieteellisestä näkökulmasta kysymys on puolestaan siitä, miten saada kone tekemään juridista harkintaa, kuten vaikkapa ottamaan huomioon suhteellisuusperiaatteen tehdessään valintoja (Ailisto ym., 2018).

Dignum (2018) mukaan tekoälyn etiikassa tulisi ottaa huomioon yhteiskunnalliset arvot sekä punnita niitä eri sidosryhmien monikulttuurillisista näkökulmista. Eettinen tekoäly liittyy ihmisten vastuuseen kehittää älykkäitä järjestelmiä inhimillisten perusperiaatteiden ja arvojen mukaisesti. Hän näkeekin, että nimenomaan vastuullisuus on autonomian kannalta perustavanlaatuista ja sen tulisi olla yksi keskeisistä lähtökohdista tekoälyn etiikan tutkimuksen taustalla. Näiden periaatteiden pohjalta hän on luonut oman mallinsa tekoälyn etiikalle, joka koostuu kolmesta kentästä (taulukko 2).

TAULUKKO 2 Tekoälyn etiikan kolme tasoa (Dignum, 2018)

Eettisten tekijöiden suunnittelu	Eettisen päättelykyvyn integroiminen autonomisiin tekoälyjärjestelmiin teknisten ja algoritmisten ratkaisujen avulla.
Suunnitteluetiikka	Säätely- ja suunnittelumenetelmät, jotka tukevat tekoälyjärjestelmien eettisten vaikutusten analysointia ja arviointia niiden su-lautuessa sosiaalsiin ja yhteiskunnallisiin rakenteisiin.
Ammattietiikka	Käytännösäännöt, standardit ja sertifiointiprosessit, jotka tukevat kehittäjiä ja käyttäjiä heidän tutkiessaan, suunnitellessaan, valmistessaan, käyttäessään tai hallitessaan tekoälyjärjestelmiä.

Etiikan huomioon ottamiseksi osana tekoälyn tai tekoälysovellusten kehittämistä on olemassa kaksi ensisijaista lähestymistapaa: metodinen ja periaatteellinen. Metodiset tarkastelumallit perustuvat suunnittelua ohjaaviin lähestymistapoihin, joissa etiikkaa käsitellään suunnitteluajattelun ja vaiheittain etenevien menetelmien kautta. Periaatteellisissa lähestymistavoissa etiikkaa lähestytään ohjaavien periaatteiden ja periaatekokoelmien kautta. Periaatteelliset tarkastelumallit pyrkivät tarkastelemaan tekoälyn eettisiä periaatteita (Koivisto ym., 2019).

4.2 Eettiset periaatteet

Tekoälyn eettisyyttä pohdittaessa ihminen nähdään moraalisten ratkaisujen tekijänä sekä vastuun kantajana. Toimiva ja demokraattinen digitaalinen yhteiskunta perustuu ennen kaikkea luottamukselle. Jotta luottamus yhteiskuntaa ja sen instituutioita kohtaan syntyy, tarvitaan kokemus osallisuudesta, sananvapaudesta ja yhteiskunnallisen muutoksen mahdollisuudesta. Yhteiskunnan eettiseen toimintaan sisältyy sen jatkuvuuden turvaaminen ja kehittäminen esimerkiksi merkittävän tiedon säilyttämisen sekä yhteiskunnalle kriittisen tiedon yhdenmukaisen käytön avulla (Valtiovarainministeriö, 2018).

IEEE on laatinut A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems -julkaisussaan (taulukko 3) ehdotuksen autonomisten ja älykkäiden järjestelmien eettisessä ja arvopohjaisessa suunnittelussa, kehittämisessä ja toteuttamisessa noudatettavista yleisistä periaatteista. Näistä periaatteista tutkielman kannalta keskeisiä ovat erityisesti vastuullisuus ja läpinäkyvyys, joita käsitellään tarkemmin seuraavassa alaluvussa 4.3.

TAULUKKO 3 Eettisen tekoälyn periaatteet (The IEEE, 2019)

Ihmisoikeudet	Tekoäly on luotava ja sitä on käytettävä kunnioittamaan, edistämään ja suojelemaan kansainvälisesti tunnustettuja ihmisoikeuksia.
Hyvinvointi	Tekoälyä kehittävien valmistajien on otettava ihmisten lisääntynyt hyvinvointi ensisijaiseksi menestyskriteeriksi kehitystyössä.
Tiedonvälitys	Valmistajien on annettava yksilöille mahdollisuus jakaa turvallisesti tietoaan ylläpitääkseen ihmisten kykyä hallita omaa henkilöllisyyttään.
Tehokkuus	Tekoälyvalmistajien ja muiden toimijoiden on esitettävä todisteita tekoälynsä tehokkuudesta ja tarkoituksenmukaisuudesta.
Läpinäkyvyys	Tietynlaisen tekoälyratkaisun perustan tulisi aina olla löydettävissä.
Vastuullisuus	Tekoäly on luotava ja sitä on käytettävä tarjoamaan yksiselitteinen peruste kaikille tehdyille päätöksille.
Tietoisuuden väärinkäytön kitkeminen	Tekijöiden on suojauduttava kaikilta mahdollisilta tekoälyyn liittyviltä väärinkäytöiltä ja riskeiltä toiminnallaan.

(jatkuu)

Taulukko 3 (jatkuu)	
Pätevyys	Valmistajien on määriteltävä turvallisen ja tehokkaan toiminnan edellytykset ja käyttäjillä on oltava tarvittava tietotaito niiden noudattamiseen.

Yllä listattujen periaatteiden (taulukko 3) lisäksi tekoälykeskustelussa on esitetty myös muita tekoälyn tai tekoälyjärjestelmien kehitystä ohjaavia periaatteita. Omia eettisiä koodistoja tarjoavien tahojen selvityksissä toistuvia periaatteita ovat ainakin vakaus ja turvallisuus, ihmiskeskeinen ja ihmisarvoja kunnioittava tarkoitus, jäljitettävyyden sekä reiluus ja syrjimättömyys.

Vakaus ja turvallisuus edustavat algoritmien riittävää luotettavuutta virheellisyyksien tai epäjohtamismukaisuuksien varalta jokaisessa tekoälyjärjestelmän elinkaaren vaiheessa (European Commission, 2019). Ihmisarvoja kunnioittava tarkoitus koskee muun muassa yksityisyyden suojaa, datan hallintaa sekä tiedonvälitystä. Kansalaisilla tulee säilyttää mahdollisuus valvoa omia tietojaan, eikä tällaisia tietoja saa käyttää heitä vahingoittavalla tavalla. Reiluus ja syrjimättömyys liittyvät ihmisoikeuksiin ja siihen, ettei tekoälyjärjestelmien päätöksenteko saa lisätä yhteiskunnassa vallitsevia rakenteellisia epätasa-arvoisuuksia, vaan se toimii yhdenvertaisesti niin yksilöitä kuin ihmisryhmiäkin kohtaan (Alasoini, Koivula & Kurki, 2018). Tekoälyä hyödynnettäessä tulisi pystyä säilyttämään kyky selittää päätökset järjestelmän käyttäjille ja ymmärtää järjestelmän päätöksiä sekä niiden lähtökohtia. Jäljitettävyyden puolestaan liittyy keskeisesti läpinäkyvyyteen. Tekoälyjärjestelmien toiminnan ja algoritmien tulisi olla läpinäkyviä ja ihmisten ja järjestelmien täytyisi kyetä vastaamaan päätöksistä ja tunnistaa virhetilanteet (Valtiovarainministeriö, 2018).

Valtioneuvoston kanslian julkaisemassa Tekoäly viranomaistoiminnassa -eettiset kysymykset ja yhteiskunnallinen hyväksyttävyyden -julkaisusarjassa julkisen sektorin toimijoiden tekoälyn käyttöönoton suunnitteluun ja toteutukseen esitetään seitsemän portaalan malli eettisten kysymysten tarkasteluun (taulukko 4). Tämän eettisen toimintamallin perustana on yleinen lähtökohta, jossa eettisyyttä tarkastellaan ihmisen hyvän elämän näkökulmasta (Koivisto ym., 2019).

TAULUKKO 4 Eettinen toimintamalli julkisella sektorilla (Koivisto ym., 2019)

1. Eettisen näkökulman valinta	Eettinen näkökulma, joka kohtelee kaikkia ihmisiä oikeudenmukaisesti ja kunnioittaa yksilöiden tarpeita sekä oikeuksia.
2. Arvioitavan tapauksen ymmärtäminen	Selkeä ja perinpohjainen kuvaus käsiteltävästä sovelluksesta, siihen liittyvistä toimijoista, heidän rooleistaan ja yhteyksistään muihin toimijoihin. (jatkuu)

Taulukko 4 (jatkuu)	
3. Sovellukseen liittyvien eettisten kysymysten löytäminen	Sovellukseen liittyvien eettisten avainkysymysten tunnistaminen ja perinteisten eettisten ohjesääntöjen hyödyntäminen niiden käsittelyssä.
4. Muihin vastaaviin tapauksiin vertaaminen	Vastaavanlaisten ennakkotapausten tai vastakohtaisten tapausten löytäminen ja niihin vertaaminen.
5. Systemaattisten analyysitekniikoiden hyödyntäminen	Eettisten ratkaisujen ja ohjesääntöjen valinnassa voidaan hyödyntää esimerkiksi eri ammattiryhmien eettisiä koodistoja, roolien ja vastuiden arviointia sekä tekoälyyn liittyvien eettisten periaatteiden huomiointia.
6. Eettisten johtopäätösten tekeminen	Pohdintaa tapauksen avainkysymyksistä ja toimenpiteiden eettisestä oikeuttamisesta.
7. Tapauksen pohdinta tulevaisuuden kannalta	Pohdintaa mahdollisen epäeettisyyden ehkäisykeinoista, ja siitä millaisia uusia toimintatapoja olisi mahdollista suositella.

Latvasen (2019) mukaan tekoälypalveluiden eettiseen vaikutusarviointiin kannattaisi soveltaa joustavaa mallia, jonka mukaan osa periaatteista tunnistettaisiin kategorisiksi ja loput aktivoituisivat tapauskohtaisesti käyttöyhteyden mukaan. Kategoriset periaatteet liittyisivät muun muassa turvallisuuteen ja yksityisyydensuojaan, kun taas tapauskohtaiset periaatteet esimerkiksi hallintoon ja talouteen. Tekoälyn eettisesti kestävä käyttö edellyttää yhteiskunnalta omien eettisten rakenteiden ymmärtämistä, jonka jälkeen ne on mahdollista aktivoida sääntelyyn ja järjestelmien suunnitteluun.

4.3 Läpinäkyvyys ja vastuut

Tekoälyn kehittämisessä ja arvioinnissa kiinnitetään enenevässä määrin huomiota teknologian avoimuuteen, riskeihin ja validointiin. Avoimuuden eli läpinäkyvyyden lisäämisellä pyritään parantamaan tekoälyn käytön turvallisuutta, hyväksyttävyyttä ja sitä kautta myös luotettavuutta. Läpinäkyvien viranomaisprosessien tarve liittyy sekä tekoälyjärjestelmän toimintaan, että järjestelmän prosessoimaan dataan (Koivisto ym., 2019). Läpinäkyvyys voi koskea esimerkiksi tiedon käyttöä, avointa lähdekoodia, tutkimustiedon avoimuutta, turvallisuusteknologian avoimuutta sekä organisaatioiden tavoitteisiin ja toimintakulttuuriin liittyvää avoimuutta (Valtiovarainministeriö, 2018). Termi lä-

pinäkyvyys tekoälyn kontekstissa nähdään käsittävän myös jäljitettävyyden, selitettävyyden ja tulkittavuuden käsitteet (The IEEE, 2019). Läpinäkyvyys viittaa avoimuuteen siitä, millaista aineistoa kerätään ja missä tarkoituksessa sekä mihin päätöksiä tukevien ja tekevien algoritmien avulla pyritään. Tämän pitäisi olla työntekijöiden tiedossa tekoälyä päätöksenteossa hyödyntävissä organisaatioissa (Alasoini ym., 2018). Kansalaisten kannalta ehdottomia vaatimuksia läpinäkyvyydelle ovat omistaja- ja kontaktiosapuolien esiin tuominen, palvelun toimintalogiikan yleistajuinen selittäminen, palveluntuottajaverkossa olevien toimijoiden tarjoaminen sekä täsmällinen ohje palvelun käyttöön ja käyttäjän oikeuksiin mahdollisissa virhetilanteissa. (Haataja & Latvanen, 2019).

Läpinäkyvyyden puute lisää vahingon riskiä ja suuruutta tilanteissa, joissa käyttäjät eivät ymmärrä käyttämiään järjestelmiä tai virheitä ei ole pystytty korjaamaan vahinkojen jälkeen. Samalla se lisää myös vaikeuksia vastuuvollisuuden varmistamisessa (The IEEE, 2019).

Moraalinen vastuu nähdään ihmisillä liittyvän velvollisuuksiin ja se koskee sekä tekoja että niiden tekemättä jättämistä. Moraalinen vastuu edellyttää teon tarkoituksellisuutta ja mahdollisuutta ennakoida sen aiheuttamia seurauksia (Kokkonen, 2016). Tekoälyn vastuullisuus liittyy keskeisesti siihen, ettei tekoälyyn pohjautuva päätöksenteko ole vaaraksi kenenkään fyysiselle tai psyykkiselle terveydelle tai uhkaa turvallisuutta, kuten yksilön yksityisyydensuojaa tai tietoturvaa (Alasoini ym., 2018). Tekoälyn vastuullisuus ei ole mahdollista ilman riittävää läpinäkyvyyttä. Tekoälyjärjestelmän ohjelmointi, tulostedata tai perimmäinen tarkoitus eivät usein ole yleisesti havaittavissa. Tämän vuoksi ihmiset ja instituutiot tarvitsevat selkeyttä näiden järjestelmien valmistuksessa ja käyttöönotossa vastuiden todentamiseksi sekä mahdollisten haittojen välttämiseksi. Haittatilanteissa syyllisyyden pitäisi olla tarpeen mukaan jaettavissa vastuullisten tekijöiden ja toimijoiden kesken, jotta vältetään sekaannukset sekä pelot suuremman yleisön keskuudessa. Lainsäädännön olisi selvitettävä tekoälyyn liittyvät vastuuvollisuuskysymykset mahdollisuuksien mukaan jo ennen tekoälyjärjestelmien kehittämistä ja käyttöönottoa, jotta valmistajat ja käyttäjät voivat ymmärtää heidän oikeutensa ja velvollisuutensa (The IEEE, 2019).

4.4 Lainsäädäntö ja valvonta

Tekoälyn hyödyntämisen osalta kysymyksiä ja keskustelua on herättänyt nyky-lainsäädännön riittävyys ja hyvin monella eri juridiikan osa-alueella joudutaan miettimään uusia lainsäädäntöratkaisuja, jotta ne vastaavat tarpeita myös tekoälyn aikakaudella. Monissa kansallisissa tekoälystrategioissa sääntelyn perustana käytetään jo olemassa olevaa lainsäädäntöä, koska varsinaisia tekoälyyn liittyviä lakeja ei vielä ole olemassa (Työ- ja elinkeinoministeriö, 2019).

Tehokas oikeusjärjestelmä edistää ihmisten hyvinvointia ja lait ovat olennainen osa sosiaalista järjestystä. Oikeusjärjestelmän luonne kertoo perustavanlaatuisella tavalla yhteiskunnan olemuksesta, kyvystä edistää väestönsä hyvinvointia sekä mahdollisuuksista talouskasvuun ja teknologisiin innovaatioihin.

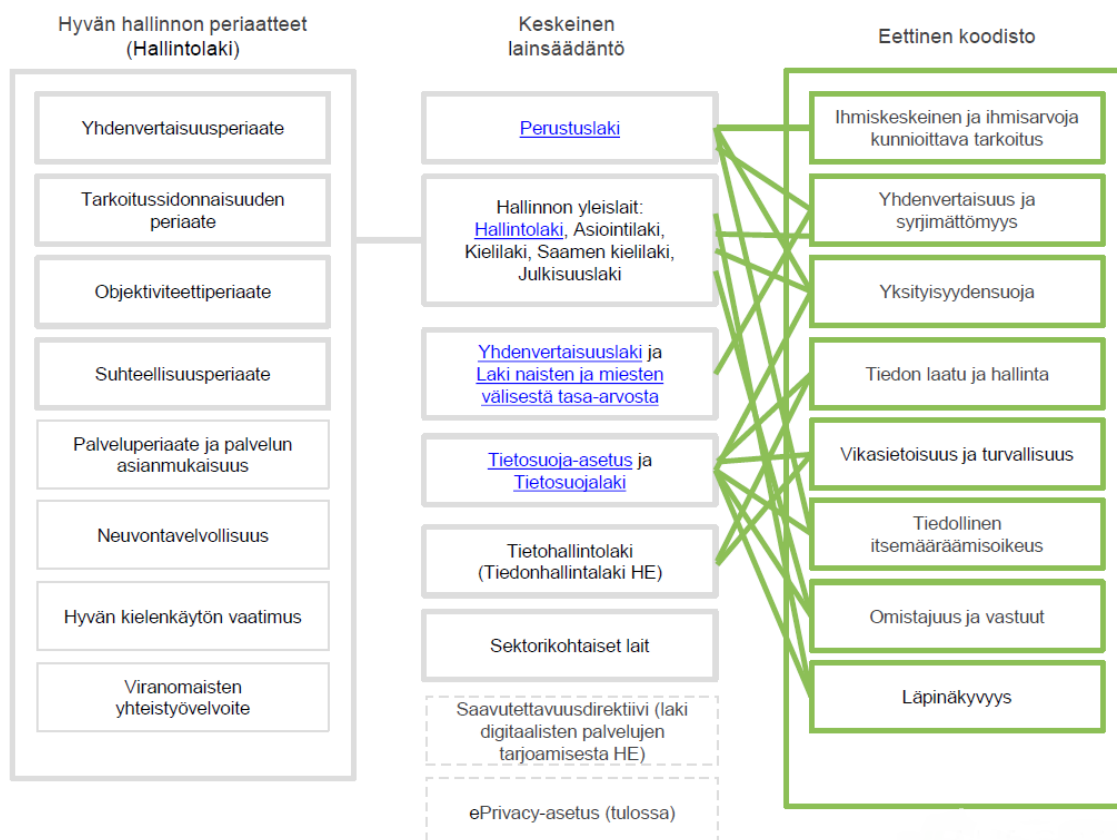
Tämän perusteella oikein toimivan tekoälyn lainsäädännön ominaisuuksia tulisi olla nopeus, reiluus, puolueettomuus, johdonmukaisuus, läpinäkyvyys, saatavuus, tehokkuus, tarkkuus ja mukautuvuus (The IEEE, 2019).

Valtiovarainministeriön Eettistä tietopolitiikkaa tekoälyn aikakaudella -selonteossa (2018) peräänkuulutetaan tekoälyn opettamiseen tarvittavia eettisiä standardeja, joiden avulla voidaan varmistaa kansalaisen tietosuojaa, datan laatu sekä luotettavuus. Selonteon mukaan tiedon luotettavan käsittelyn ja jakamisen sekä läpinäkyvien käytäntöjen lisäksi tarvitaan linjauksia, joiden avulla algoritmeja ja arkkitehtuuria voidaan eettisesti kehittää.

Teknologisen sääntelyn, kuten erilaisten standardien lisäksi sääntely voidaan jakaa kolmeen alajoukkoon, jotka ovat taloudellinen, sosiaalinen ja hallinnollinen sääntely. Tutkielman kannalta erityisen kiinnostava on hallinnollinen sääntely, jonka tarkoituksena on ohjata julkisen ja yksityisen sektorin toimia esimerkiksi verotusta, yritystoimintaa, jakelujärjestelmiä ja terveydenhuoltoa koskevilla säännöksillä (OECD, 2011).

Selvitys- ja tutkimustyössä on ilmennyt kysyntä ja tarve johdonmukaisen sekä kokonaisvaltaisen sääntelystrategian kehittämiseksi, jotta tekoälypohjaista päätöksentekoa voidaan lisätä julkisella sektorilla vaarantamatta kansalaisten oikeusturvaa (Koulu ym., 2019). Nykyinen lainsäädäntö Suomessa on sektori-kohtaista, jonka seurauksena myös eettiset käytänteet tulisi luoda toimialakohtaisesti. Tällä tavoin horisontaaliset ohjenuorat muuttuvat käytännöiksi, jotka pystyvät asianmukaisesti ohjaamaan kullekin toimialalle ominaisten sovellusten eettistä arviointia ja riskienhallintaa (Haataja, 2019).

AuroraAI -hankkeen teettämä etiikkaselvitys osoittaa, että tekoälyn eettisiä vaatimuksia säädellään jo monin eri tavoin esimerkiksi perustuslailla ja tietosuojalailla. Eettisillä ohjenuorilla pyritään määrittelemään hyväksytyjä toimintaperiaatteita ja vahvistamaan lain yhdenmukaista tulkintaa tekoälyn kontekstissa (Haataja ym., 2019). Alla esitetään Haatajan ja Latvasen luoma malli tekoälyn lainsäädännöllisestä perustasta, joka pohjautuu AuroraAI -hankkeen eettiseen koodistoon (kuviot 2).



KUVIO 2 Tekoälyn lainsäädännöllinen perusta (Haataja & Latvanen, 2019)

Oikeusjärjestelmän ja tekoälyn suhdetta tarkasteltaessa on otettava huomioon kaksi erillistä kysymystä: Voiko oikeusjärjestelmä tarjota odotettua oikeusturvaa, kun tekoälyjärjestelmä tuottaa epäreilun tuloksen? Ja mihin oikeusjärjestelmään tekoäly tulisi sisällyttää tai millä ehdoilla? (Campolo, Sanfilippo, Whittaker & Crawford, 2017). Hashmin (2019) kolme pääpoliittista suositusta tehokkaan ja maailmanlaajuisen etiikan käytännesäännösten kehittämiseksi ovat tekoälyn sidosryhmäyhteisön tiivis yhteistyö, olemassa olevien hallintotapojen hyödyntäminen ja tekoälytietoisuuden levittäminen institutionaalisella tasolla.

Julkisessa keskustelussa tekoälyjärjestelmistä ilmaantuneet eettiset huolet koskevat etenkin sodankäyntiä, yksityisyyden suojaa ja valvontaa, vastuuta, tekijänoikeuksia ja tasa-arvoa. Tästä syystä on tarpeen selvittää sellaiset mahdolliset tiedon käsittelyn ja tekoälyn soveltamisen muodot, joita on väärin kehittää ja joiden kehittämisen kieltäminen olisi perusteltua esimerkiksi kansainvälisten sopimusten avulla. Lisäksi tärkeää on linjata, millaista tekoälykehitystä priorisoidaan julkisella rahoituksella toteutettavaksi, sekä keskustella siitä, kenellä on oikeus päättää tekoälyn hyväksyttävät sovelluskohteet (Valtiovarainministeriö, 2018).

5 POHDINTA JA YHTEENVETO

Tässä tutkielmassa selvitettiin tekoälyn hyödyntämistä julkisella sektorilla sekä sen eettisyyden vaatimuksia. Tutkielman ensimmäisen tutkimuskysymyksen avulla pyrittiin selvittämään millä tavalla julkisella sektorilla hyödynnetään tekoälyä ja mitkä asiat rajoittavat toistaiseksi laajempaa käyttöönottoa. Mehrin (2017) mukaan tekoälyn soveltaminen julkiselle sektorille täytyy tehdä huolellisesti, jotta vältytään ongelmilta esimerkiksi palvelun tuottamisessa, yksityisyydensuojaan tai eettisiin kysymyksiin liittyvissä kysymyksissä. Keskeisimpinä ratkaisuinä tekoälyn eettiselle hyödyntämiselle nähdään riskienhallinta tilanteissa, joissa tekoäly toimii päätösten tekijänä. Lisäksi tärkeänä nähdään avoin kuntien välinen yhteistyö (Klemettinen ym., 2018), olemassa olevien resurssien hyödyntäminen sekä tekoälyn nostaminen mukaan tavoiteohjattuun strategiaohjelmaan (Mehr, 2017).

Tekoäly ja koneoppiminen kykenee helpottamaan julkishallinnon toimia muun muassa rutiiniprosesseissa ja tilanteissa, joissa hallittavaa ja analysoitavaa tietoa on paljon. Näissä tapauksissa tekoälyn on mahdollista automatisoida erilaisia toimenpiteitä ja vähentää viranomaisten hallinnollista kuormaa kehittämällä samalla toimintaansa oppimisen kautta (Deloitte, 2017). Suomen vahvuuksina tekoälyn hyödyntämisessä nähdään korkea koulutustaso, kyky soveltaa uusia teknologioita nopeasti sekä valmiudet yhteistyöhön yritysten ja tutkimuksen välillä. Lisäksi Suomessa on avoin asenneilmapiiri uusille teknologioille, ja lainsäädäntö sekä viranomaiset suhtautuvat tekoälypohjaisen automaation käyttöön myönteisesti (Ailisto, Helaakoski, Dufva & Tuikka, 2017).

Tutkielman toinen tutkimuskysymys pyrkii vastaamaan kysymyksiin liittyen tekoälyn eettisiin vaatimuksiin, joita tutkielmassa käsitellään muun muassa vastuuden, läpinäkyvyyden sekä sääntelyn kautta. Koska ihmiset suunnittelevat ja kehittävät algoritmeja sekä niitä ohjaavaa tietoa, myös algoritmin tekemät tai sen tietoon perustuvat päätökset nähdään olevan viime kädessä ihmisen vastuulla. Tekoäly, kuten muukaan teknologia ei tunne kansallisia rajoja, minkä vuoksi nykyisen lainsäädännön puutteet synnyttävät epävarmuutta julkisen sektorin digitalisoinnin kehitykselle sekä kansalaisten luottamukselle perustuvalla päätösten lainmukaisuudelle. Tekoälyn lisääntyvän hyödyntämisen

myötä myös muita haasteita saattaa ilmaantua, ellei niihin kiinnitetä riittävää huomiota jo uusien ratkaisujen kehitysvaiheessa (Kääriäinen ym., 2018). Erityisen paljon huolta on herättänyt tekoälyn luotettavuus, minkä vuoksi etenkin käytettävän datan valintaan tulisi kiinnittää huomiota, jotta ihmisen ajattelun vääristymät eivät kopioitu niitä uudelleen toteuttaviin algoritmeihin (Valtiovarainministeriö, 2018).

Luvuissa 4 ja 4.1 esitetään kootusti kirjallisuuskatsauksessa tunnistettuja yksittäisiä eettisyyteen ja hyväksyttävyyteen liittyviä kysymyksiä sekä eri tahojen esittämiä tekoälyn eettisiä koodistoja ja periaatteita. Luvussa 4.2 eritellään The IEEE:n esittelemät eettisen tekoälyn periaatteet (taulukko 3). Euroopan komission tekoälystrategiaan perustuva selvitys (European Commission, 2019) listaa oman edotuksensa luotettavan tekoälyn seitsemästä ehdosta, joiden täyttymistä tarkastellaan arviointilistojen avulla. Nämä seitsemän ehtoa ovat ihmisen suorittama valvonta, vakaus ja turvallisuus, yksityisyyden suoja, läpinäkyvyys, syrjimättömyys, yhteiskunnallinen hyvinvointi sekä vastuuvollisuus. Copeland (2019) korostaa, että on hyvin hankalaa suunnitella yhtä ainoaa, yleispätevää eettistä säännöstöä, joka voisi kattaa kaikki tekoälyn hyödyntämiseen liittyvät monimutkaiset tapaukset. Ihannetilanteessa globaalin konsortion tai instituutin tulisi kehittää maailmanlaajuiset standardit tekoälyn etiikalle (Hashmi, 2019).

LÄHTEET

- Ailisto, H., Helaakoski, H., Dufva, M. & Tuikka, T. (2017). Tuottoa ja tehokkuutta Suomeen tekoälyllä. *VTT, Policy Brief 1/2017*.
- Ailisto, H. (2018). Tekoälyn käsitekartta. VN-TEAS projekti. Tekoälyn kokonaisjäsenitys ja kansallinen osaamiskartoitus.
- Ailisto, H., Heikkilä, E., Helaakoski, H., Neuvonen, A. & Seppälä, T. (2018). Tekoälyn kokonaiskuva ja osaamiskartoitus.
- Alasoini, T., Koivula A. & Kurki L. (2018). Tekoälyteknologian hyvä soveltaminen ja etiikka. *Tekoälyaika*.
- Anderson, M. & Anderson, S. (2011). Machine ethics. *Cambridge University Press*.
- Asimov, I. (1942). Runaround. *Astounding Science Fiction, Street & Smith*.
- Bostrom, N. & Yudkowsky, E. (2011). The ethics of artificial intelligence. Draft for Cambridge Handbook of Artificial Intelligence. *Cambridge University Press*.
- Campolo, A., Sanfilippo, M., Whittaker, M. & Crawford, K. (2017). AI Now 2017 Report.
- Cearley, D., Burke, B., Smith, D., Jones, N., Chandrasekaran, A. & Lu, CK. (2019). Top 10 Strategic Technology Trends for 2020. A Gartner Special Report.
- Copeland, E. (2019). Does the public sector really need a code of AI ethics? *Nesta*.
- Crevier, D. & Lepage, R. (1997). Knowledge-based image understanding systems: A survey. *Computer vision and image understanding, 67(2), 161-185*.

- Crisp, R. (1998). Ethics. *Routledge Encyclopedia of Philosophy*, Taylor and Francis. DOI: 10.4324/9780415249126-L132-1
- Davidson, L. (2019). Narrow vs. General AI: What's Next for Artificial Intelligence? *Springboard*.
- DeLapp, K. (2007). Moral Perception and Moral Realism: An 'Intuitive' Account of Epistemic Justification, *Review Journal of Political Philosophy*.
- Deloitte (2017). The new machinery of government: Robotic Process Automation in the Public Sector. *The Creative Studio at Deloitte, London*.
- Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics Inf Technol* 20, 1-3.
- European Commission (2019). Ethics guidelines for trustworthy AI. *Brussels*.
- Fieser, J. (2000). Moral Philosophy Through The Ages. *Mayfield Publishing Company*.
- Gordon, B. (2011). Artificial Intelligence: Approaches, Tools and Applications. *Nova Science Publishers, Inc., USA*.
- Haataja, M. & Latvanen, M. (2019). AuroraAI -esiselvityshanke. Etiikka-työkokonaisuuden suositukset.
- Harisalo, R., Aarrevaara T., Stenvall J. & Virtanen, P. (2007). Julkinen toiminta - julkinen politiikka. Tampere: Tampereen Yliopistopaino Oy.
- Hashmi, A. (2019). AI Ethics: The Next Big Thing In Government. Anticipating the impact of AI Ethics within the Public Sector. *World Government Summit & Deloitte*.
- Honkanen, P. (2015). Julkinen sektori: muistiinpanoja yksityistämisestä ja säästöpolitiikasta.
- Hyvärinen, S. & Parviainen, J. (2018). Kuntien tietotekniikkakartoitus 2018. Kuntien tietotekniikan tunnusluvut, organisointi, toiminnan kehittäminen ja haasteet. *Suomen Kuntaliitto, Helsinki*.
- Julkunen, R. (2006). Kuka vastaa? Hyvinvointivaltion rajat ja julkinen vastuu. *Stakes, Helsinki*.
- Klemettinen, M. & Peltola, V. (2018). Tekoälyaika, esitysmateriaali 15.2.2018, *Tietoyhteiskunta-Akatemia*.
- Koivisto, R., Leikas, J., Auvinen, H., Vakkuri, V., Saariluoma, P., Hakkarainen, J. & Koulu, R. (2019). Tekoäly viranomaistoiminnassa - eettiset kysymykset

ja yhteiskunnallinen hyväksyttävyys. *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 14/2019.*

- Koulu, R., Mäihäniemi, B., Kyyrönen, V., Hakkarainen, J. & Markkanen, K. (2019). Algoritmi päätöksentekijänä? Tekoälyn hyödyntämisen mahdollisuudet ja haasteet kansallisessa sääntely-ympäristössä. *Valtioneuvoston kanslia 2019:44.*
- Kääriäinen, J., Aihkisalo, T., Halén, M., Holmström, H., Jurmu, P., Matinmikko, T., Seppälä, T., Tihinen, M. & Tirronen, J. (2018). Ohjelmistorobotiikka ja tekoäly – soveltamisen askelmerkkejä. Valtioneuvoston kanslia. Selvitys- ja tutkimustoiminnan julkaisusarja 65/2018.
- Kääriäinen, J., Aihkisalo, T., Halén, M., Jurmu, P., Matinmikko, T., Seppälä, T. & Tihinen, M. (2018). Ohjelmistorobotiikan ja tekoälyn kehitysvaateet julkiselle sektorille – alustavia havaintoja.
- Launchbury, J. (2017). A DARPA Perspective on Artificial Intelligence.
- Latvanen, M. (2019). Tekoäly ja tekoälyn etiikka-muutamia perusasioita. *Kuntamarkkinat 2019.*
- Marturano, A. (2002). The role of metaethics and the future of computer ethics. *Ethics and Information Technology, 4(1), 71-78.*
- McCarthy, J. (2006). What has AI in Common with Philosophy? *Stanford University, USA.*
- Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government. *Harvard Ash Center Technology & Democracy Fellow.*
- Rusanen, A.-M. & Koskinen, I. (2018). Tiede, tekoäly ja tiedolliset riskit.
- Rusthollkarhu, S. (2019). Ihminen vastaan kone - milloin tekoäly erehtyy? Pro Gradu- tutkielma, *Tampereen Yliopisto, johtamisen ja talouden tiedekunta.*
- The IEEE (2019). Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition.
- Turing, A. M. (1950). Computing Machinery And Intelligence. *Mind vol. 49.*
- Nilsson, NJ. (2009). The quest for artificial intelligence. *Cambridge University Press.*
- OECD (2011). Regulatory reform and innovation.
- Ollila, M.-R. (2019). Tekoälyn etiikkaa. *Otava.*

- Patterson, J. & Gibson, A. (2017). *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc.
- Pietarinen, J. (2015). *Etiikka*. *Filosofia.fi*. Eurooppalaisen filosofian seura ry.
- Polonski, V. (2017). Can we teach morality to machines? Three perspectives on ethics for artificial intelligence.
- Russell, S.J., Norvig, P. (2016). *Artificial intelligence: A modern approach*. Malaysia; Pearson Education Limited.
- Stahl, B. C., Timmermans, J. & Mittelstadt, B. D. (2016). The ethics of computing: A survey of the computing-oriented literature. *ACM Computing Surveys (CSUR)*, 48(4), 55.
- Spicker, P. (2000). *The Welfare State. A general theory*. Sage, London.
- Stirling, R. (2017). Five levels of AI in public service.
- Sydänmaalakka, P. (2015). Älykäs julkinen johtaminen. Miten rakentaa älykäs verkostoyhteiskunta? *Talentum*.
- Tieteen termipankki (2020). *Filosofia: metaetiikka*.
- Torresen, J. (2018). A Review of Future and Ethical Perspectives of Robotics and AI. *Frontiers in Robotics and AI*.
- Työ- ja elinkeinoministeriö (2019). Edelläkävijänä tekoälyaikaan. Tekoälyohjelman loppuraportti. *Työ- ja elinkeinoministeriön julkaisuja 2019:23*.
- Vakkuri, V. & Abrahamsson, P. (2018). The Key Concepts of Ethics of Artificial Intelligence, *IEEE International Conference on Engineering, Technology and Innovation, Stuttgart*.
- Valtiovarainministeriö (2018). Valtioneuvoston selonteko tietopolitiikasta ja tekoälystä. Valtioneuvoston selontekomenettely. VNS 7/2018 vp.
- Valtiovarainministeriö (2018). Eettistä tietopolitiikkaa tekoälyn aikakaudella - selonteko.
- Valtiovarainministeriö (2019). AuroraAI -kohti ihmiskeskeistä yhteiskuntaa. Kansallisen tekoälyohjelma Auroran esiselvityshankkeessa tuotettu kehittämis- ja toimeenpanosuunnitelma 2019–2023.