Tobias Mörck

ALGORITMISEN PÄÄTÖKSENTEON OIKEUDENMUKAISUUS JA TURVALLISUUS

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2024

TIIVISTELMÄ

Mörck, Tobias
Algoritmisen päätöksenteon oikeudenmukaisuus ja turvallisuus
Jyväskylä: Jyväskylän yliopisto, 2024, 32 s.
Tietojärjestelmätiede, kandidaatintutkielma
Ohjaaja: Mehtälä, Saana

Algoritmeihin perustuvat päätökset vaikuttavat yhä enenevissä määrin yhteiskuntaan ja yksittäisiin ihmisiin. Näiden päätösten tutkiminen voidaan nähdä hyvin merkittävänä yhteiskunnalle, sillä algoritmisen päätöksenteon käytön laajenemisen myötä on tärkeää löytää sopivia toimintatapoja niiden oikeudenmukaisen sekä turvallisen käytön varmistamiseksi. Kysymykset, joihin tässä tutkielmassa pyrittiin kirjallisuuskatsauksen metodein löytämään vastauksia, liittyivät oikeudenmukaisuuden ja turvallisuuden toteutumiseen algoritmisten päätöksentekojärjestelmien käytössä. Oikeudenmukaisuuden käsitettä tutkittiin yleisten säädösten sekä eettisten pohdintojen näkökulmasta. Keskeiset löydökset viittasivat tiiviin yhteistyön tärkeyteen algoritmien rakentamisen yhteydessä oikeudenmukaisuuden toteutumiseksi. Lisäksi sääntelyn tärkeys teknologian nopean kehityksen seurauksesta nähtiin merkittävänä. Turvallisuutta taas käsiteltiin suurelta osin itseohjautuvien ajoneuvojen kautta tarkastellen. Itseohjautuvien ajoneuvojen käyttö voidaan nähdä sujuvoittavan liikennettä merkittävästi, mutta päätöksenteon taustoihin liittyviä ongelmia voi kuitenkin ilmetä. Varsinkin kolaritilanteissa tapahtuvan päätöksenteon taustat ovat usein pohdinnan ja tutkimuksen kohteena. Näiden ajoneuvojen päätöksenteon eettisyys nähtiin monimutkaisempana kuin usein vertauskuvana pidetty vaunuongelma kuvastaa. Automaattisten päätösten käytön sopivuus eri käyttötarkoituksissa riippuu paljolti päätöksen turvallisuuskriittisestä riskistä, tai riskistä aiheuttaa suuria menetyksiä vääränlaisen päätöksen seurauksesta. Yhteistyötä ihmisen ja algoritmin välillä pidettiin turvallisuuden takaamisessa myös keskeisenä tekijänä. Vaikka oikeudenmukaisuuden ja turvallisuuden käsitteitä voidaan käsitellä erillisinä käsitteinään, kuuluu niihin kuitenkin paljon samankaltaisuuksia. Ongelmat vastuun kannossa, liittyen automaattisten päätösten seurauksiin, koskee vahvasti useampaa osa-aluetta oikeudenmukaisuudessa sekä turvallisuudessa. Algoritmien rakentamisen ongelmia pystytään nykyisen tietämyksen valossa pääpiirteittäin ratkaisemaan eri sidosryhmien asiantuntijoiden tiiviin yhteistyön avulla, jossa otetaan huomioon yhteiset edut ja menetysten minimointi.

Asiasanat: algoritminen päätöksenteko, oikeudenmukaisuus, eettiset ongelmat, turvalliset algoritmit, algoritmien sääntely

ABSTRACT

Mörck, Tobias
Fairness and security of algorithmic decision-making
Jyväskylä: University of Jyväskylä, 2024, 32 pp.
Information Systems, Bachelor's Thesis
Supervisor: Mehtälä, Saana

Algorithm-based decisions are influencing society and individuals in an expanding amount. Examining these decisions can be seen as important for society, because as the use of algorithm-based decision expands, it is important to find suitable ways to ensure their fair and safe use. The questions, to which answers were sought in this thesis using the methods of literature review, were related to the realization of fairness and security in the use of these automated decision-making systems. The topic of fairness was examined through a viewpoint of general regulations and ethical considerations. The main findings suggest that the close cooperation of different parties is essential to ensure the fairness in algorithmic decision-making. In addition, the importance of regulation as a result of rapid technological development was seen as significant. The security aspect was investigated mainly through the viewpoint of autonomous vehicles. The use of autonomous vehicles can be seen to significantly enhance traffic flow, but there may be problems with the decision-making process. In particular, the backgrounds of the decisions behind collisions are usually the subject of contemplation and research. The ethics of these vehicles were seen as more complex than the famous metaphor of a trolley problem usually reflects them to be. The suitability of automated decisions-making in different scenarios often depends on how large of a safety risk is associated with the decision or how much harm it can cause in an error. The collaboration between human and algorithm was seen as one of the key elements in ensuring security. Although fairness and security can be considered as two different concepts, they share many similarities. The problems surrounding accountability were highly present in many ways within both fairness and security. Based on current knowledge, many problems surrounding automated decision-making can be solved through close collaboration among various stakeholders, while taking into account shared benefits and minimizing losses.

Keywords: algorithmic decision-making, fairness, ethical problems, secure algorithms, regulation of algorithms

KUVIOT

SISÄLLYS

1 JOHDANTO

Algoritminen päätöksenteko tarkoittaa teknologian avulla tehtyä päätöstä, joka ei vaadi välitöntä ihmisen osallisuutta. Ihmisen rooli kyseisissä tapauksissa voi kuitenkin vaihdella tilannekohtaisesti. Tällainen päätöksenteko perustuu usein koneoppimisalgoritmeihin tai sääntöpohjaiseen automaatioon. Näitä järjestelmiä saatetaan toisinaan käyttää vain päätöksenteon tukena, jolloin algoritmin tarjoamaa informaatiota hyödynnetään ainoastaan lopullisen päätöksentekijän tarpeiden mukaan. Saatua dataa arvioidaan ja päätös tehdään parhaaksi nähdyllä tavalla. Täysin autonomisissa järjestelmissä päätös kuitenkin toteutuu ilman ihmisen ohjausta (Koulu, ym. 2019).

Algoritmien ohjaama päätöksenteko vaikuttaa yhä enemmän ihmisten arkipäivään. Koska tällaiset automaattiset järjestelmät voivat aiheuttaa merkittäviä haittoja yksilöille ja yhteisöille, on herännyt huoli oikeudenmukaisuudesta. Starken ym. (2022) mukaan tutkijoiden ja päättäjien vaatima ihmislähtöinen lähestymistapa edellyttää, että ihmisten oikeudenmukaisuuskäsitykset otetaan huomioon algoritmipohjaista päätöksentekoa suunniteltaessa ja toteutettaessa

Harrisin ja Davenportin artikkelissa (2005) voitiin nähdä jo lähes 20 vuotta sitten näkemyksiä algoritmisen päätöksenteon tulevaisuuden mahdollisuuksista. He toteavat algoritmisen päätöksenteon kattavan useita menetelmiä, jotka vaativat vaihtelevasti ihmisten osallistumistasoja. Tällaisissa ympäristöissä tavoitteena olisi saada tietokonejärjestelmä tekemään päätös tai antamaan jonkintasoinen suositus. Jopa kaikkein automatisoiduimmat järjestelmät tarvitsevat kuitenkin asiantuntijoita ja johtajia sääntöjen ylläpitoon ja tulosten seurantaan. Päätöksentekojärjestelmä perustuu yleensä joukkoon sääntöjä, mutta siihen voi liittyä myös muita tietoja ja päätöksentekoteknologioita. Myös toisin kuin päätöksenteon tuki, useimmat näistä järjestelmistä ovat suoraan yhteydessä keskeisiin liiketoimintaprosesseihin, jotka kääntävät päätöksen nopeasti, tarkasti ja tehokkaasti toiminnaksi (Harris & Davenport, 2005).

Tämän tutkimuksen pääsääntöisenä tavoitteena on tutkia ja pohtia miten oikeudenmukaisuuden sekä turvallisuuden käsitteet ilmenevät algoritmisessa päätöksenteossa eri konteksteissa. Aihe on teknologian nopean kehittymisen seurauksesta hyvin kriittinen, jotta päätöksentekojärjestelmiä voidaan tulevai-

suudessa rakentaa ottaen huomioon yhteiskuntaan ja yksilöihin vaikuttavia tekijöitä. Tästä esimerkkinä voidaan pitää algoritmista syrjintää, jolla voidaan viitata suoraan syrjintään algoritmien tai tekoälyjärjestelmien käytössä tai epäsuoriin syrjiviin seurauksiin, joita algoritmien kautta aiheutuu (Ojanen, ym., 2022). Aiheesta on tässä tutkielmassa pyritty etsimään monipuolisesti erilaista kirjallisuutta, jota analysoitiin sekä syntetisoitiin, samalla pohtien vastauksia tutkimuskysymyksiin.

Tutkimuskysymykset, joihin tässä tutkielmassa pyritään löytämään vastauksia ovat seuraavat:

- Millaisia oikeudenmukaisuuteen ja turvallisuuteen liittyviä ongelmia voidaan nähdä algoritmisessa päätöksenteossa?
- Mitä keinoja oikeudenmukaisuuden ja turvallisuuden varmistamiseksi voidaan käyttää algoritmisten päätöksentekojärjestelmien kehittämisessä?

Tutkimuskysymyksiin vastaamiseen liittyy vahvasti myös oleellisten käsitteiden määrittely. Käsitteitä tutkimalla ja niiden merkitystä pohtimalla pyritään luomaan selkeämpää kuvaa tutkimuskysymysten avaamista ongelmista liittyen algoritmiseen päätöksentekoon.

Oikeudenmukaisuuden käsitettä kuvataan tässä tutkielmassa osittain nykyisiä käytäntöjä sekä säädöksiä tutkien, painottuen lähinnä EU:n alueelle. Lisäksi aihetta pohditaan myös muiden eettisten periaatteiden ja pohdintojen kautta. Tavoitteena on löytää yhteen vetävää määritelmää siitä, miten oikeudenmukaisuus nähdään algoritmisessa päätöksenteossa ja sen sovelluksissa. Turvallisuuden käsitettä taas tutkitaan pääosin itseohjautuvien ajoneuvojen näkökulmasta erillisenä osuutenaan. Itseohjautuvat autot valittiin yhdeksi tutkielmassa käsiteltäväksi aiheeksi niiden aiheen kannalta oletetun sopivuuden vuoksi. Kun tavoitteena on tutkia algoritmisen päätöksenteon turvallisuutta, voidaan autonomisten ajoneuvojen toiminnan tuovan merkittäviä näkökulmia pohdinnalle. Lopuksi tavoitteena on tehdä yhteenvetoa molemmista suuremmista kokonaisuuksista ja yhdistellä löydöksiä suurempaa kokonaisuutta varten.

Tämä tutkimus toteutettiin kirjallisuuskatsauksena. Kirjallisuus valittiin hyödyntäen Google Scholar-hakupalvelua sekä suoraan tieteellisiä julkaisuja sisältäviä tietokantoja (Science Direct, Wiley Online library) käyttäen hakusanoja: *"Algoritminen päätöksenteko"*, *"Algoritmisen päätöksenteon oikeudenmukaisuus"*, *"Algoritmisen päätöksenteon turvallisuus"* tai englanniksi *"Algorithmic decision-making"*, *"Security in algorithmic decisions"*, *"Fairness in algorithmic decision-making"*. Kirjallisuutta valittiin ensisijaisesti materiaaleista, jotka sisälsivät suuremman osan tutkielman aiheen käsitteistä (esim. algoritmit, oikeudenmukaisuus, turvallisuus), sillä aiheen laajuuden vuoksi nähtiin, että kokonaisuutta laajemmin kuvaava kirjallisuus voisi antaa merkittävämpää näkemystä tutkittavista ongelmista. Näiden materiaalien lisäksi kirjallisuutta valittiin yksityiskohtaisemmin painottaen eri osa-alueita erillisinä alueinaan. Tutkielmassa tar-

kastellaan aiheen osalta pääosin laadullisia käsitteitä ja selityksiä, joten teknisempään puoleen algoritmeista ei suurelta osin paneuduttu. Täten materiaalit, jotka käsittelevät yksipuolisesti vain teknisiä käsitteitä jätettiin pois.

Tekoälyä hyödynnettiin osittain lähdemateriaalin käsittelyssä. Tämä tapahtui lähinnä englanninkielisen materiaalin kääntämisenä sekä yleisen ymmärtämisen tukemisena. Työkaluina käytettiin ChatGPT-tekoälysovellusta sekä DeepL-kääntötyökalua. Lopullisen tekstin sisältöön tekoäly ei kuitenkaan ole suoraan vaikuttanut. Tekoälyyn suhtauduttiin tutkimusta tehdessä kuitenkin hyvin kriittisesti, ja sen luomaa sisältöä hyödynnettiin vain inspiraation lähteenä.

Tutkimuksen avulla löydettiin monia näkökulmia liittyen algoritmisen päätöksenteon eettisiin tekijöihin, lainsäädännöllisiin ongelmiin sekä turvallisuuden huomioimiseen. Algoritmien ympärillä toimivien asiantuntijoiden tiiviin yhteistyön nähtiin olevan avainasemassa oikeudenmukaisuuden varmistamisessa algoritmisissa päätöksentekojärjestelmissä. Turvallisuuden ja oikeudenmukaisuuden huomioiminen nähtiin algoritmien nopean kehittymisen vuoksi entistä merkittävämpänä tekijänä järjestelmien tulevaisuuden kehitystyössä. Varsinkin vastuunkantoon liittyvät ongelmat olivat keskeisessä asemassa. Turvallisuuden edistämisessä voidaan toki myös hyödyntää algoritmista päätöksentekoa, joten turvallisuuden ongelmaa ei täten voida nähdä ainoastaan yhdensuuntaisena. Kaiken kaikkiaan oikeudenmukaisuuden ja turvallisuuden pohdinnat sisälsivät eriävistä näkökulmista huolimatta samansuuntaisia ongelmia sekä ratkaisuja.

Seuraavassa luvussa käsitellään algoritmisen päätöksenteon oikeudenmukaisuuteen liittyviä käsitteitä. Aihetta tarkastellaan säädösten kautta tarkastellen ja eettisiä pohdintoja hyväksi käyttäen omina alalukuinaan. Kolmannessa luvussa taas käsitellään turvallisuutta algoritmiseen päätöksentekoon liittyen. Lukuun sisältyy kaksi alalukua, joista toinen käsittelee kokonaisuudessaan itseohjautuvia ajoneuvoja. Toisessa alaluvussa tarkastellaan muita turvallisuuteen liittyviä esimerkkejä sekä pohditaan yleisesti aiheen tulevaisuuden näkymiä. Tämän jälkeen näiden kahden edellisen luvun keskeiset löydökset kootaan yhteen neljännessä luvussa. Lopuksi viimeisessä luvussa koko tutkielman sisällöstä tehdään yhteenveto ennen lähdeluetteloa.

2 OIKEUDENMUKAISUUS ALGORITMISESSA PÄÄTÖKSENTEOSSA

Oikeudenmukaisuuden käsite pitää tässä tutkielmassa sisällään kuvauksen siitä, miten päätöksiä voidaan erilaisten eettisten periaatteiden kautta tulkita. Lisäksi annetaan esimerkkiä aiheen kannalta merkittävistä säädöksistä ja siitä, miten ne pyrkivät vaikuttamaan algoritmiseen päätöksentekoon ja sen hyödyntämiseen. Teknologian kehittyminen tuo mukanaan monia mahdollisuuksia, mutta myös riskejä. Siksi voidaan nähdä, että on tärkeää tehdä jatkuvaa työtä eettisyyden edistämiseksi algoritmisen päätöksenteon kontekstissa. Tässä luvussa tavoitteena on tutkia näitä ennalta mainittuja teemoja algoritmeihin ja niiden käyttöön liittyen.

Yhtenä suurena ongelmana algoritmien kehittyessä on, että niiden selitettävyys myös vaikeutuu. De Bruijnin ja kollegoiden (2022) mukaan selitettävyys voidaan nähdä olevan merkittävässä roolissa päätöksien luotettavuuteen liittyen, koska epäluottamuksen syntyminen on todennäköistä algoritmin toiminnan pysyessä mysteerinä. Monimutkaistuessaan algoritmit pystyvät antamaan yhä tarkempia ja tehokkaampia päätöksiä. Haittapuolena niiden selitettävyys kuitenkin heikkenee, jolloin yleisön näkökulmasta perustelulle voidaan nähdä olevan tarvetta. Tämän vuoksi selitettävyys käsitteenä voi ehkäistä monia yhteiskunnallisia ongelmia aiheeseen liittyen (de Bruijn, ym., 2022). Myös kaupallisesta näkökulmasta tekoälyn selitettävyys ja eettisten tekijöiden huomioon ottaminen voi vaikuttaa liiketoiminnan menestymiseen. Epätietoisuus johtaa usein ajattelumalliin, jossa riski voidaan nähdä liian suureksi tuotteen/palvelun käyttöön, jolloin mahdollinen positiivinen vuorovaikutus käyttäjään jää saavuttamatta (Bartneck, ym., 2021). Yleinen epätietoisuus algoritmisia päätöksiä kohtaan myös julkisten palveluiden alalla on aiheuttanut laajaa epäluottamusta sekä epämukavuutta yhteisöjen keskuudessa (Brown, ym., 2019).

Voidaan siis nähdä, että oikeudenmukaisuuden toteutuminen algoritmisissa päätöksissä vaatii yhä enemmän huomiota algoritmisten järjestelmien kehittyessä. Jatkuvan kehitystyön tärkeys on täten siis jalustalla. Tämä kehitys on tärkeä varmistaa, jotta algoritmit voisivat palvella yhteiskunnan etuja ja minimoida menetyksiä.

2.1 Säädökset yleisesti ja Euroopan unionin alueella

Euroopan komission huhtikuussa 2021 julkaisemassa ehdotuksessa (Työ- ja elinkeinoministeriö, 2021) tekoälyn sääntelystä tuodaan esille riskiperusteinen lähestymistapa. Siinä tekoälyjärjestelmät jaetaan ei-hyväksyttäviin, suuren riskin, rajoitetun riskin ja vähäriskisiin kategorioihin. Useimmat sovellukset ovat kuitenkin vähäriskisiä ja niitä ei erikseen säädellä. Suuririskisiin tekoälysovelluksiin, kuten koulutukseen, rekrytointiin ja lainvalvontaan, liittyy erityisvaatimuksia, kuten datan laadun, järjestelmän tarkkuuden ja läpinäkyvyyden arviointi. Työ- ja elinkeinoministeriö kuitenkin huomauttaa, että vaikka esityksessä mainitaan syrjimättömyys useita kertoja, sitä ei erikseen vaadita suuririskisiltä järjestelmiltä (Työ- ja elinkeinoministeriö, 2021).

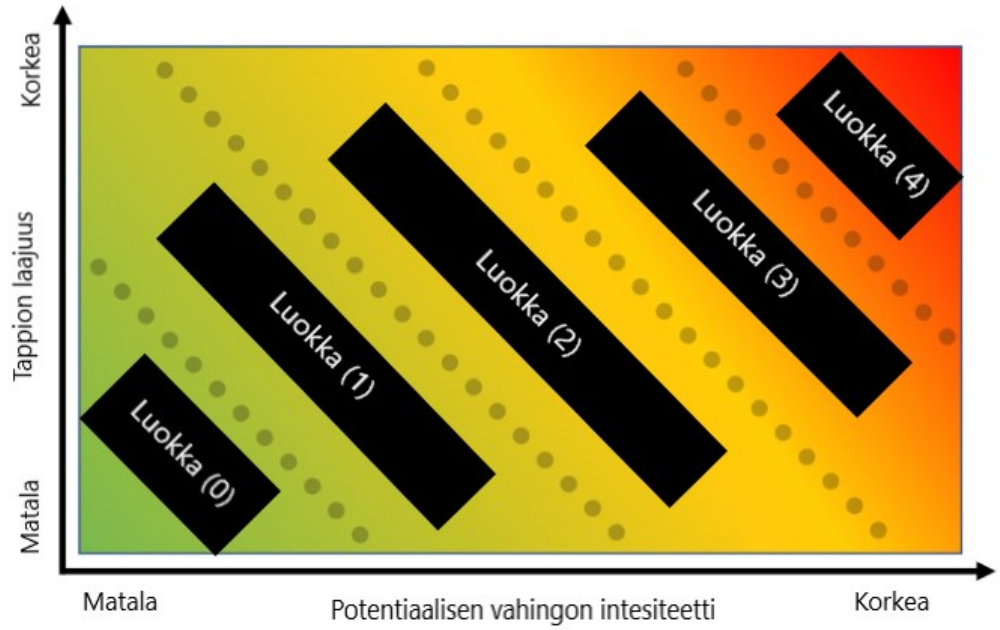
Ojanen ym. (2022) kertoo artikkelissaan tekoälyn käytön ylittävän usein kansainväliset rajat, mikä vaikeuttaa lainsäädännön valvontaa. EU:lla on haasteita käsitellessään algoritmista syrjintää, kuten oikeudellisia aukkoja, intersektionaalista syrjintää ja välillistä syrjintää. Algoritminen syrjintä voidaan nähdä myös ongelmallisena, koska nykyiset oikeusjärjestelmät eivät sovellu hyvin uusiin tilanteisiin. Lisäksi algoritmisen syrjinnän todentaminen ja vastuuseen asettaminen ovat vaikeita tekoälyjärjestelmien monimutkaisuuden ja moninaisten toimijoiden vuoksi (Ojanen, ym., 2022). EU määrittelee, että epäsuoraa syrjintää ei aiheudu, jos käytäntö on objektiivisesti perusteltu laillisella tavoitteella ja keinot sen tavoitteen saavuttamiseksi ovat asianmukaiset ja välttämättömät (Zuiderveen, 2018).

Vuonna 2016 Euroopan unioni julkaisi uudistuneen yleisen tietosuoja-asetuksen, joka tuli voimaan 2018. Se pyrkii siirtämään painopistettä pois paperipohjaisista byrokraattisista vaatimuksista kohti käytännön noudattamista ja yksilöiden oikeuksien vahvistamista. Suurin osa asetuksista käsittelee tapaa, jolla tietoa kerätään ja varastoidaan. Asetuksessa on myös artikla 22, joka käsittelee automaattista yksilöllistä päätöksentekoa ja profilointia. Tämä artikla koskee laajaa joukkoa algoritmeja, joita käytetään esimerkiksi suosittelujärjestelmissä, luottoluokitus- ja vakuutusriskien arvioinnissa, laskennallisessa mainonnassa ja sosiaalisissa verkostoissa (Euroopan Unioni, 2016). Goodmanin ja Flaxmanin (2017) mukaan tämä herätti tärkeitä kysymyksiä erityisesti koneoppimisyhteisön keskuudessa. Tietosuoja-asetuksen politiikka korostaa ihmisen tulkittavuuden tärkeyttä algoritmien suunnittelussa, kun kansalaisten oikeus saada selitys algoritmipäätöksistä otetaan huomioon. Huomautetaan myös, että vaikka tietosuoja-asetus aiheuttaa useita haasteita koneoppimisen sovelluksille, nämä haasteet voidaan nähdä tuovan mukanaan kehitystä tukevia tekijöitä. Useat haasteet korostavat työn tärkeyttä, joka varmistaa, että algoritmit ovat tehokkaita, läpinäkyviä ja oikeudenmukaisia. Jatkotutkimuksen kannalta olisi Goodmanin ja Flaxmanin mukaan olennaista osoittaa, että millä tavalla tietyntyyppisissä algoritmijärjestelmissä voitaisiin tunnistaa ja ottaa käyttöön korjaavia toimenpiteitä syrjinnän vähentämiseksi. Ulkopuolisten ja eettisesti epäasianmukaisten tekijöiden vaikutukset ovat ihmisen päätöksenteossa hyvin do-

kumentoituja, joten tätä dataa voitaisiin hyödyntää parempien algoritmien rakentamiseen. Syrjivän päätöksenteon on myös nähty olevan hyvin yleistä monilla algoritmeja potentiaalisesti hyödyntävillä aloilla. Täten asianmukaisesti sovellettuna algoritmit voisivat tehdä tarkempia ennusteita ja tarjota lisää läpinäkyvyyttä ja oikeudenmukaisuutta verrattuna ihmispäätöksentekijään (Goodman & Flaxman, 2017).

Hakkaraisen, Koulun ja Markkasen artikkelissa (2020) käsitellään algoritmeja, tekoälyä ja automaatiota käsittelevää tutkimusta hallinnollisesta näkökulmasta. Artikkelin mukaan, vaikka automatisoitu päätöksenteko ja algoritmit ovat saaneet laajempaa huomiota vasta äskettäin, niiden tutkimus oikeustieteen alalla ei ole aivan uusi asia. Tämä aihe on ollut tutkimuksen kohteena jo 1950-luvulta lähtien erityisesti oikeusteorian, kielitieteen ja oikeuskibernetiikan tutkimustraditioissa. Kuitenkin juuret tälle tutkimuksen suunnalle voidaan jäljittää vieläkin kauemmaksi. 1940-luvulla käynnistyi jurimetriikaksi kutsuttu tutkimussuunta ja se keskittyi pääosin laskennallisten menetelmien soveltamiseen oikeudellisessa päättelyssä. Tuon ajan tutkimuksen ominaispiirteenä oli, että siitä puuttuivat käytännössä toimivat tekoälysovellukset ja konkreettiset käytännön toteutukset. Tutkimus pyrki ennakoimaan teknologisen kehityksen ja oikeudellisen päätöksenteon automatisoinnin ihanteellista tulevaisuudenkuvaa (Hakkarainen, ym., 2020). Bynum (2018) esittää artikkelissaan Moorin (1985) näkemyksiä sen aikaisesta tietotekniikkaan liittyvästä sääntelystä. Moor esitti, että tietotekniikassa kohdataan usein ongelmia, jotka johtuvat puutteellisesta ohjeistuksesta. Olisi siis tarve kehittää viitekehys, jonka avulla voitaisiin rakentaa toimivia käytäntöjä ongelmien ratkaisemiseksi (Bynum, 2018; Moor, 1985). Hakkarainen ym. (2020) esittää, että klassiset hallinto-oikeudelliset kysymykset, kuten hyvän hallinnon tai avoimen hallinnon toteutuminen, ovat kuitenkin nousseet tutkimuksen keskiöön vasta 2000-luvulla. Syitä tälle tutkimusaiheen vähäiselle huomiolle voidaan nähdä tulevan useammasta suunnasta. Yhtenä syynä esitetään olevan tietoteknisten valmiuksien puute automaattisen päätöksenteon toteuttamisessa, joka oli merkittävää aina 1950–1980-luvuille asti. Alun perin oikeustieteellinen tutkimus oli suuntautunut pääasiassa tuomioistuimiin ja merkittäviin oikeustapauksiin, kun taas teknologiaan liittyvä tutkimus oli pitkään keskittynyt tuomioistuimen ja tuomarin päätöksentekoprosessien mallintamiseen. Hallintoautomaatio puolestaan liittyi Hakkaraisen ym. mukaan usein massaluonteisiin päätöksiin, joiden tarkkaa koodaamista tai perustelemista ei aina pidetty tehokkaana tutkimuksen näkökulmasta. Tämä on johtanut siihen, ettei ratkaisutoiminnan mallintamiseen keskittyvä tutkimustraditio ole pystynyt antamaan suoria vastauksia hallintopäätösten automaatiosta aiheutuviin ongelmiin (Hakkarainen, ym., 2020). Tutkimussuunnan historia on kuitenkin hyvin tuoretta, sekä sen kehitys pitkälti nopeasti etenevää. Tulevaisuuden näkymiä pohtiessa voidaan olettaa kehityksen jatkuvan kutakuinkin yhtä nopeasti kuin aikaisempien vuosikymmenten aikana. Lisäksi nopean teknologisen kehityksen mukana ilmenee jatkuvasti uusia tarpeita tutkia ilmiöitä sopivien sekä yhteiskuntaa tukevien säädösten rakentamiseksi.

Avoimuuden ongelmat algoritmisten järjestelmien kohdalla liittyvät monesti ns. "musta laatikko" -ongelmaan, joka viittaa koneoppimismenetelmien tekniseen läpinäkymättömyyteen (Hakkarainen, ym., 2020; Krafft, Zweig, König, 2022). Tämä tarkoittaa, että järjestelmän toimintaa ei välttämättä voida täysin varmentaa. Edes järjestelmän kehittäjä ei välttämättä tiedä, että miten se päätyy tiettyihin päätöksiin. Tämä tekninen läpinäkymättömyys ei kuitenkaan koske sääntöpohjaisia päätöksentekojärjestelmiä, joissa päätökset määritellään etukäteen (Hakkarainen, ym., 2020). Informaatioasymmetriat ovat myös Krafftin ym. (2022) mukaan keskeinen haaste. Tällöin järjestelmä toimii vastoin odotuksia ja haluttuja etuja ei synny. On myös mahdollista, että tulokset ovat ristiriidassa tilannekohtaisten säädösten kanssa. Viranomaisen velvollisuus edistää avoimuutta Hakkaraisen ym. (2020) mukaan ulottuu sekä järjestelmän kehittämiseen että sen käyttöön. Tarkemmat toteutustavat kuitenkin määrittelevät velvollisuuden täyttymisen ehdot. Tarkastelun keskiössä kerrotaan olevan erityisesti algoritmisen järjestelmän lähdekoodi, joka on ohjelmoijan luettavassa muodossa oleva koodi. Sen avulla järjestelmän toiminta määritellään ja sen saatavuus teoriassa mahdollistaa järjestelmän arvioinnin (Hakkarainen, ym., 2020). Järjestelmiä kehittäessä on myös tarpeen ottaa huomioon riskien laajuus sen ympäristössä (Kuvio 1) (Krafft, ym., 2022).



Kuvio 1 Automaattisten päätöksentekojärjestelmien riskimatriisi (mukaillen Krafft ym., 2022)

Krafftin ja kollegoiden (2022) esittämä malli (Kuvio 1) esittää suuntaa antavasti tarvetta automaattisten päätöksentekojärjestelmien käytön sääntelystä eri tilanteissa. Kuviossa vaaka-akseli kuvaa potentiaalisen vahingon intensiteet-

tiä ja pystyakseli tappion laajuutta. Tasot (0) - (4) kuvaavat karkeasti jaoteltuja riskikategorioita, joista numeron suuruus kuvaa sitä, kuinka suurta sääntelyä olisi suositeltavaa kohdistaa kyseisen automaattisen päätöksentekojärjestelmän käyttöä varten. Luokkien nimet ovat yksinkertaistettu suomennoksessa asioiden selkeyttämiseksi. Järjestelmän potentiaalisesti aiheuttama vahinko ja mahdolliset menetykset riippuvat kuvion laatijan mukaan sen sosiaalisesti sidotusta toteutuksesta. Vaikka riskikategorioiden välillä ei ole selkeitä rajoja, on silti tärkeä nähdä kokonaiskuva, sillä päätöksentekojärjestelmät eroavat laajalti niiden vaikutuksista yhteiskuntaan (Krafft, ym., 2022).

Hakkaraisen ym. (2020) mukaan julkisuusperiaatteen toteuttaminen perustuslain mukaisesti tapahtuu pääasiassa julkisuuslain kautta, joka käsittelee asiakirjojen julkisuutta. Kysymys nousee kuitenkin siitä, että missä määrin viranomaisen käyttämän algoritmisen järjestelmän lähdekoodia voidaan pitää viranomaisen asiakirjana. Päätöksentekojärjestelmien avoimuuden ongelman vuoksi julkisuuslain merkitys nähdään usein epäselvänä. Koska asiakirjajulkisuus ei yksinään ratkaise algoritmisten järjestelmien valvontaa, avoimuuden periaatetta kuuluisi korostaa entistä enemmän. Avoimuuden ja läpinäkyvyyden nähdään olevan keskeisissä rooleissa arvioitaessa algoritmisen päätöksenteon hyväksyttävyyttä (Hakkarainen, ym., 2020). Käytännössä algoritmisten järjestelmien läpinäkyvyysvelvoitteet vaativat, että asianomaisen henkilön kuuluu saada tieto osallisuudestaan, kun hän on algoritmisen järjestelmän arvioitavana (Työ- ja elinkeinoministeriö, 2021). Hakkarainen ym. (2020) osoittavat myös, että läpinäkyvyysvelvollisuudella sekä julkisuudella on tärkeä rooli oikeudellisen päätöksenteon hyväksymisessä. Algoritmisen päätöksenteon ja ymmärrettävyyden välinen etäisyys on merkittävä. Tämä etäisyys vaikuttaa siihen, että kuinka paljon avoimuus lopulta tuottaa todellista läpinäkyvyyttä. On kuitenkin toivottavaa, että perustelut eivät etääntyisi liian kauas itse päätöksen tekemisestä (Hakkarainen, ym., 2020).

Euroopan parlamentin päätöslauselmassa lokakuussa 2020 sisältyy ehdotus viitekehyksestä, liittyen tekoälyyn, robotiikkaan ja niihin liittyvien teknologioiden eettisiin näkökohtiin, jonka mukaan korkeimman riskin sektorit ovat koulutus, terveydenhuolto, kuljetus, energia, työllisyys, maanpuolustus, turvallisuus, rahoitus sekä julkisensektorin oikeus- ja sosiaaliturva. Näillä sektoreilla on suurin riski perusoikeuksien ja turvallisuussäädösten rikkoutumisessa. Tulevaisuuden työn sekä tutkimuksen kannalta, liittyen algoritmisten päätösten toimintaan näillä kyseisillä sektoreilla, voidaan olettaa tapahtuvan muutoksia nykyisten ongelmien ja riskien ratkaisemiseksi (Euroopan parlamentti, 2020).

2.2 Eettiset tekijät

Binns (2022) käsittelee artikkelissaan oikeudenmukaisuuteen liittyviä ulottuvuuksia ja henkilön yksilöllisiin ominaisuuksiin perustuvaa syrjintää. Pohdintaa toteutetaan kolmen elementin suhteiden välillä. Nämä elementit ovat johdonmukaisuus, yksilöllinen oikeudenmukaisuus ja syrjimättömyys. Näiden

nähdään monimutkaistavan päätöksenteon tasapainottamisen ongelmaa ihmisten ja algoritmien osuuden välillä. Binnsin mukaan yksilöllisen oikeudenmukaisuuden korostaminen saattaa johtaa vähentyvään syrjimättömyyteen lisääntyneen säännöttömyyden vuoksi. Kun päätöksentekijöille annetaan enemmän harkintavaltaa, motivoituneet päätöksentekijät voivat kaivaa esiin tekijöitä, jotka saattavat johtaa epäyhtenäiseen kohteluun. Tällaiset epäoikeudenmukaiset tapaukset ovat jo historiassa tuoneet esille harkintavaltaa rajoittavien säädösten suosimista (Binns, 2022).

Algoritmiset järjestelmät eivät kuitenkaan automaattisesti poista syrjintää, vaikka ne välttäisivät inhimillisen päätöksentekijän mahdollisen epäjohdonmukaisuuden. Pelkkä johdonmukaisuus ei riitä välttämään eriarvoisia vaikutuksia, jos tapauksen arvioinnissa käytetyt tekijät ovat itse syrjinnän lähde (Binns, 2022; Ojanen, ym., 2022). Tällaisissa tilanteissa voidaan Ojasen ym. (2022) mukaan puhua algoritmisesta vinoumasta, jolla tarkoitetaan tilannetta, jossa koneoppimismalli tai laajemmin tekoälysovellus kohtelee tiettyä ihmisryhmää systemaattisesti epäoikeudenmukaisesti. Sen seurauksesta voi ilmetä tilanteita, joissa epätarkat ennusteet tai päätelmät vahingoittavat yksilöitä (Ojanen, ym., 2022). Tämä voi esimerkiksi vaikuttaa laajasti yritysten henkilöstönhallinnassa (Köchling & Wehner, 2020). Vinouma voi liittyä demografisiin, fenotyyppisiin tai muunlaisiin ominaisuuksiin. Se voi ilmetä myös datajoukon eroavaisuuksina esimerkiksi taudin tai luottokelpoisuuden jakautumisessa (Ojanen, ym., 2022).

Binns (2022) pohtii myös mahdollisuudesta, jossa algoritmeja voitaisiin suunnitella huomioimaan useita oikeudenmukaisuuden ulottuvuuksia. Tällaiset lähestymistavat pyrkisivät yhdistämään johdonmukaisuuden, yksilöllisen oikeudenmukaisuuden ja syrjimättömyyden. Esitetään myös, että tällaiset menetelmät voisivat vähentää ihmisen päätöksentekijän roolia oikeudenmukaisuuden ulottuvuuksien välissä. Huomautetaan kuitenkin myös, että ihmisen osuutta päätöksenteossa ei voida täysin korvata (Binns, 2022).

Binns (2022) kertoo, että algoritmijärjestelmät, joiden tarkoituksena on vain ehostaa inhimillisiä päätöksiä tai seuloa tapauksia inhimillistä tarkastelua varten, näyttävät yhdistävän algoritmisten päätösten edut. Näitä etuja ovat esimerkiksi johdonmukaisuus ja inhimillinen harkinta, jotka nähdään palvelevan yksilön oikeudenmukaisuutta. Binns kuitenkin kertoo, että yksilöllinen oikeudenmukaisuus on usein ristiriidassa algoritmien ohjaaman johdonmukaisuuden ja oikeudenmukaisuuden kanssa. Käytännössä algoritmijärjestelmien parissa toimivat asiantuntijat voivat käyttää harkintavaltaansa omien sitoumustensa mukaisesti tavalla, joka saattaa olla ristiriidassa niiden käyttöönottaneen organisaation tavoitteiden kanssa. Mahdollisuus epätasaiseen oikeudenmukaisuuden toteutumiseen voi luoda lisähaasteita harkintavallan roolin, arvon ja laajuuden arvioimiselle. Kun yksilöllisen oikeudenmukaisuuden käsite on pystytty täysin muodostamaan, jää Binnsin mukaan vielä nähtäväksi tämän käsitteen tärkeyden suuruus tietyissä yhteyksissä. Tulevaisuuden pohdinnaksi jää se, että millaisia yhteiskunnallisia odotuksia saattaa olla niillä, jotka hyötyvät tai kärsivät sen seurauksista (Binns, 2022). Algoritmisten järjestelmien avoimuus nähdään edellyttävän sekä asiantuntijavalvontaa, kuten lähdekoodin julkisuut-

ta, että avoimuutta kansalaisten suuntaan. Kansalaisen oikeuden kannalta on keskeistä, että päätöksiin sisältyy perusteluvelvollisuutta sekä järjestelmän parissa toimivan asiantuntijan valvontaa (Hakkarainen, ym., 2020). Näitä kysymyksiä on ratkaistava, kun algoritmijärjestelmiä otetaan käyttöön käytännössä. Jos yksilöllistä oikeudenmukaisuutta on tärkeä suojella, ei voida olettaa, että se voitaisiin varmistaa yksinkertaisesti lisäämällä ihminen algoritmisen prosessin osaksi.

Lisääntynyt oikeudenmukaisuuden käsitteen ymmärrys ja hyödyntäminen nähdään nousevan keskeiseksi tekijäksi algoritmisten järjestelmien kehittämisessä, jotta haitallisia seurauksia voitaisiin torjua tehokkaammin. Starken ym. (2022) mukaan sekä OECD että Euroopan komissio ovat tukeneet algoritmisen oikeudenmukaisuuden merkitystä yhtenä neljästä keskeisestä periaatteesta luotettavan tekoälyn kehittämisessä, ja se on sisällytetty suureen osaan tekoälyetiikkaa koskevista ohjeista. Kuitenkin yhteiskunnallisten vaikutusten käsittely liittyen epäoikeudenmukaisiin algoritmipohjaisiin päätöksentekojärjestelmiin vaatii toisinaan enemmän kuin pelkkiä teknisiä ratkaisuja. Oikeudenmukaisten algoritmien suunnittelussa ja käyttöönotossa tarvitaan kattavaa ymmärrystä siitä, että milloin ja minkä takia kansalaiset kokevat algoritmisen päätöksenteon epäoikeudenmukaiseksi. Starke ja kollegat kertovat tällaisten oikeudenmukaisuusnäkemysten olevan olennaisia ihmiskeskeisen tekoälyn kehitystyössä. Näin voidaan vaikuttaa järjestelmien kehittämiseen informoimalla heitä, jotka vastaavat eettisten algoritmipohjaisten päätöksentekojärjestelmien suunnittelusta. Tämä informaatio voi myös vaikuttaa päättäjiin, jotka vastaavat tällaisten järjestelmien käyttöönotosta yhteiskunnallisissa yhteyksissä. Näin ollen nämä oikeudenmukaisuuden havainnot voivat auttaa vastaamaan kehotukseen yhteiskuntalähtöisestä lähestymistavasta, jossa yhteiskunnan arvot halutaan liittää osaksi päätöksentekojärjestelmien suunnittelua (Starke, ym., 2022).

Starke ja kollegat (2022) tutkivat kirjallisuuskatsauksen muodossa reiluuden käsitystä ja ymmärrystä algoritmisissa päätöksissä. Katsauksen keskeinen havainto on, että koettu oikeudenmukaisuus algoritmisten päätöksentekojärjestelmien osalta on voimakkaasti kontekstista riippuvainen. Oikeudenmukaisuuden havainnot määräytyvät teknisen algoritmin suunnittelun lisäksi myös sovellusalueen (esimerkiksi ennakkoarviointi tai rekrytointi) ja kyseessä olevan tehtävän (esimerkiksi korkean riskin vs. matalan riskin) perusteella. Osat tuloksista ovat kuitenkin epäselviä johtuen puutteellisista teoreettisista viitekehyksistä koetun algoritmisen oikeudenmukaisuuden osalta (Starke, ym., 2022). Empiirinen näyttö viittaa siihen, että ihmiset arvioivat algoritmista päätöksentekoa eri tekijöiden perusteella kuin perinteistä ihmisen tekemää päätöksentekoa (Dietvorst, ym., 2015). Koetun oikeudenmukaisuuden tulokset ihmisten tekemien päätösten ja algoritmipohjaisten päätösten välillä ovat Starken ja kollegoiden (2022) mukaan epäselviä. Tulokset tukevat ja kyseenalaistavat näkemystä siitä, että algoritmit koetaan oikeudenmukaisempina kuin ihmiset. Tämä vaihtelu havaitaan eri toimialoilla (kuten uusintarikosten riskinarvioinnissa verrattuna rekrytointiin) ja samalla toimialalla eri tehtävien välillä (esimerkiksi kahden

erilaisen rekrytointialgoritmin kohdalla). Tämä myös osoittaa, että oikeudenmukaisuuden havainnot ovat voimakkaasti sidoksissa kontekstiin, ja että jokainen algoritmi vaatii perusteellista tutkimusta ennen laajaa käyttöönottoa. Kuitenkin kerrotaan, että pelkkä ihmisten ja algoritmien erottelu ei riitä ottamaan huomioon luonnollisen maailman monimutkaisuutta. Useimmissa käytännön tehtävissä algoritmipohjaiset järjestelmät eivät tee päätöstä täysin itsenäisesti, vaan ihmiset osallistuvat myös osana itse päätöksentekoprosessiin (Harris & Davenport, 2005). Starken ym. (2022) katsaus paljasti aukkoja nykyisessä kirjallisuudessa, sillä vain harva tutkimus on huomioinut laajemman institutionaalisen kontekstin ja sen mukanaan tuomat kysymykset algoritmisten päätöstentekojärjestelmien käyttöönotossa. Esimerkiksi, miten näiden järjestelmien kanssa työskentelevät ihmiset koulutetaan käyttämään niitä? Tai miten algoritmilla tehtävät päätökset välitetään niille, joita päätös suurimmilta osin koskee (esimerkiksi työnhakijat ja syytetyt). Ovatko asianosaiset myöskään edes tietoisia siitä, että algoritmi on tehnyt lopullisen päätöksen? Ja kuka lopulta on vastuussa virheellisistä tai syrjivistä luokituksista? Näihin kysymyksiin Starke ja kollegat toivovat tulevaisuuden tutkimuksessa löytyvän tarkempia ja moniulotteisempia ratkaisuja (Starke, ym., 2022).

3 TURVALLISUUS ALGORITMISESSA PÄÄTÖKSENTEOSSA

Tässä luvussa käsitellään algoritmisen päätöksenteon turvallisuuteen liittyviä käsitteitä algoritmien eri soveltuvuuksien näkökulmasta. Turvallisuus nähdään tässä tutkielmassa algoritmisten päätösten mahdollisuutena vaarantaa ympäristöään toiminnallaan. Yhtenä suurempana kokonaisuutena turvallisuutta tarkastellaan itseohjautuvien autojen, tai yleisesti ajoneuvojen, kontekstissa. Itseohjautuviin ajoneuvoihin viitataan tässä luvussa toisinaan myös autonomisina ajoneuvoina. Turvallisuus algoritmisissa päätöksissä voidaan nähdä merkittävänä tutkimuskohteena, sillä taustalla tapahtuva päätöksenteko voi laajasti vaikuttaa monilla eri toimialoilla yksilöihin sekä yhteisöihin.

Algoritminen päätöksenteko ja sen turvallisuuden arviointi ulottuu laajalle alueelle uusien innovaatioiden kehittyessä. Turvallisuuden käsite voi kuitenkin ulottua sen riskien lisäksi myös yleisen turvallisuuden edistämiseen algoritmeja hyödyntämällä. Esimerkiksi Bakhshi Lomer ym. (2023) tutkivat hätäsuojapaikkojen taktista sijoittamista riskiperusteista päätöksentekojärjestelmää apuna käyttäen. Tämä tutkimus esittelee uudenlaisen riskiperusteisen päätöksentukijärjestelmän auttamaan riskinhallintasuunnittelijoita valitsemaan parhaat hätämajoituspaikat maanjäristyksien jälkeen. Riskienhallintaa toteuttavat suunnittelijat voivat hyödyntää järjestelmää kohdentamalla enemmän resursseja paikoille, jotka täyttävät kaikki relevantit kriteerit optimaaliseen hätämajoituspaikkaan, mikä todennäköisesti voi lisätä tavoitellun tehokkuuden saavuttamista (Bakhshi Lomer, ym. 2023). Toisena esimerkkinä voidaan pitää Al-Rashidin ja kollegoiden (2020) tutkimusta, jossa pyritään kehittämään uudenlainen riskiluokitusalgoritmi, jonka avulla voidaan tunnistaa korkeariskiset perkutaaniset sepelvaltimointerventiot, jotta korkeariskisille potilaille voitaisiin tarjota paremmin tilanteeseen sopivaa hoitoa (Al-Rashid, ym., 2020). Vaikka algoritmiset järjestelmät voivat tuoda mukanaan lukuisia hyötyjä turvallisuuden edistämiseen, keskusteluissa nähdään usein kuitenkin pohdintaa niiden mahdollisuuksista vaarantaa turvallisuutta erilaisissa tilanteissa.

3.1 Itseohjautuvat autot

Autonominen ajoneuvo tarkoittaa Trafin tutkimuksen (Innamaa, ym., 2015) mukaan ajoneuvoa, joka pystyy suoriutumaan ajotehtävästä ilman kuljettajaa tai yhteyttä muuhun infrastruktuuriin. Tässä tutkielmassa autonomisiin ajoneuvoihin viitataan pääosin itseohjautuvina autoina. Itseohjautuvat autot voivat tuoda merkittäviä etuja liikenteeseen sujuvoittaen liikkumista sekä lisäämällä turvallisuutta (Davidson & Spinoulas, 2015). Nämä edut tuovat kuitenkin mukanaan useita haasteita, jotka voivat olla myös hankalasti ratkaistavissa (Bartneck, ym., 2021). Muun muassa lainsäädännölliset haasteet sekä vastuunkantoon liittyvät pohdinnat voivat hidastaa lukuisten etujen saavuttamisen. Itseohjautuvien autojen tutkiminen voi myös tarjota arvokasta tietoa, joka tukee yleisesti algoritmien turvallisuuteen liittyvien teemojen tutkimista ja kehittämistä. Tämä on kriittisen tärkeää, kun pyritään varmistamaan autonomisten järjestelmien luotettavuutta ja turvallisuutta ennen niiden laajamittaista käyttöönottoa. Mukaillen tämän tutkielman edellistä lukua, itseohjautuvien autojen kehitystyö vaatii myös säädösten laatimista varmistamaan turvallisuuden toteutumisen ajoneuvojen käytön aikana.

3.1.1 Hyödyt ja haitat

Itseohjautuvien autojen hyötyjä voidaan nähdä esimerkiksi suurten kaupunkien liikenteen sujuvoittamisessa. Bischoffin ja Maciejewskin (2016) toteuttama simulaatio, jossa kuvattiin Berliinin yksityisautojen korvaamista automaattisesti ohjautuvilla takseilla, osoitti tämän liikennejärjestelyn olevan paljon tehokkaampi sekä ympäristölle huomattavasti vähemmän kuormittava. Simulointimalli kattaa yksityisautojen matkojen yksityiskohtaisen kysynnän koko kaupungissa, erottaen sisäisen ja ulkoisen liikenteen. Sen tulokset osoittivat, että 100 000 itseohjautuvaa ajoneuvoa voi tehokkaasti korvata koko autokannan Berliinissä säilyttäen samalla korkean palvelun laadun asiakkaille. Lisäksi tulokset viittasivat siihen, että yksi itseohjautuva taksi voi korvata aiemmin kymmentä perinteisesti kuljetettavaa ajoneuvoa kaupungissa, korostaen mainittujen taksien mahdollisia tehokkuusetuja (Bischoff & Maciejewski, 2016).

Itseohjautuvien ajoneuvojen haasteiden kannalta Trafin (Innamaa, ym., 2015) tutkimuksessa todettiin, että suurin haaste automaattisten ajoneuvojen teknologiassa on ympäristönhavainnoinnin kehitys. Teknisesti haaste liittyy auton tilannetietoisuuden parantamiseen, jotta se voi havaita ja reagoida oikein kaikkiin liikenneympäristössä esiintyviin tilanteisiin. Täysin automaattisessa liikenteessä kuljettajan kapasiteetin kerrotaan vapautuvan muuhun toimintaan kuin itse auton ajamiseen. Taktiset tehtävät siirtyvät auton vastuulle, kun taas strategiset tehtävät, kuten matkan suunnittelu, tehdään usein ennen liikkeelle lähtöä. Koska laajempi automaatio ei toistaiseksi ole vielä toteutunut, useat haasteet nykytilanteessa liittyvät taktisen tason yllättäviin kuormituspiikkeihin.

Lisäksi kuljettajan tilannetietoisuuden ylläpitäminen sekä sopivan toimintamallin valinta tuovat paljon haasteita (Innamaa, ym., 2015).

Davidsonin ja Spinoulasin artikkelissa (2015) tuodaan itseohjautuvuuden tuomia etuja esille. Itseohjautuvat ajoneuvot voivat esimerkiksi parantaa tien turvallisuutta, koska ne tarkkailevat ympäristöään tarkemmin ja luotettavammin kuin ihmiset. Ne eivät ole myöskään alttiita ihmisen hitaille reaktioajoille. Näiden ominaisuuksien vuoksi nämä itseohjautuvat autot voivat Davidsonin ja Spinoulasin artikkelin (2015) että Trafin tutkimuksen (2015) mukaan ajaa lähempänä toisiaan ja korkeammilla nopeuksilla, lisäten näin tien kapasiteettia. Parannettu mukavuus, mahdollisuus hyödyntää paremmin aikaa matkustaessa ja vähentyneet pysäköinnin ongelmat tekevät tiellä tapahtuvasta matkustamisesta houkuttelevampaa (Davidson & Spinoulas, 2015; Innamaa, ym., 2015). Tämä todennäköisesti kuitenkin lisää matkustamista ja keskimääräisten matkojen pituutta. Ongelmia voidaan toki nähdä olevan myös lukuisia, joiden korjaamiseksi joudutaan tekemään paljon työtä. Davidson ja Spinoulas (2015) kertovat, että tästä esimerkkinä voidaan nähdä laillinen puoli itseohjautuviin autoihin liittyen. Usein vaaditaan ajokykyisen henkilön olevan läsnä autossa valmiina ottamaan vastuu ajoneuvon toiminnasta tarvittaessa. Ongelmaksi koituukin, että kuka on missäkin tilanteessa vastuussa ajoneuvon mahdollisesta sääntöjä rikkovasta toiminnasta. Tämä osoittautuu vaikeaksi varsinkin tilanteissa, joissa ajoneuvon sisällä ei ole yhtäkään vastuunkantoon kyvyllistä henkilöä. Mainitaan myös, että teknologiset haavoittuvuudet ovat mahdollinen suuri riski itseohjautuvien ajoneuvojen käytössä. Merkittäviä etuja voitaisiin saada sallimalla ohjausjärjestelmien päivitykset ilman langallista yhteyttä, jotta ajoneuvot voisivat parantaa toimintaansa ajan myötä. Järjestelmä voisi myös kytkeytyä pilvipalveluihin päivittääkseen karttoja ja ajankohtaisia olosuhteita. Molemmilla näistä on kuitenkin riskinsä, sillä hyökkäykset näihin järjestelmiin voivat aiheuttaa paitsi taloudellisia vahinkoja, mutta myös merkittäviä hengenvaarallisia tilanteita. Reaaliaikaiset karttapäivitykset saattavat kuitenkin olla hyväksyttävissä, jos järjestelmä on riittävän älykäs havaitakseen ettei tallennettu kartta vastaa antureiden havaintoja. Tällöin jos karttoja muutetaan pahantahtoisesti, voidaan ajoneuvot saada pysähtymään tai ajamaan hitaasti, mutta turvattomia tilanteita ei välttämättä aiheutuisi. Järjestelmien täytyy täten olla perustavanlaatuisesti turvallisia laajempaa käyttöönottoa varten. (Davidson & Spinoulas, 2015).

3.1.2 Eettiset pohdinnat

Kallioinen ja kumppanit (2019) esitettävät kaksi arviointitutkimusta, joissa selvitettiin mahdollisia eroja moraalisissa normeissa, joita sovelletaan ihmiskuljettajiin ja itseohjautuviin autoihin. Kokeisiin osallistujat arvioivat joukon tilannekuvia, joissa oli mukana ihmiskuljettajia tai itseohjautuvia ajoneuvoja. Tilanteiden esittämistä manipuloitiin, jotta voitaisiin selvittää näkökulman vaikutus moraalisiin arvioihin. Tuloksista nousi kaksi tärkeää havaintoa. Tutkimuksen

mukaan ihmiskuljettajia ja itseohjautuvia autoja arvioitiin suurelta osin samalla tavalla, mutta ilmeni kuitenkin selkeämpi taipumus suosia itseohjautuvia autoja toimimaan tavalla, joka minimoi vahingot verrattuna ihmiskuljettajiin. Toiseksi havaittiin, että näkökulma vaikuttaa toisinaan arvioihin. Jalankulkijan näkökulmasta arvioiduissa tilanteissa ihmiset suosivat toimia, jotka vaarantavat auton matkustajia sen sijaan, että vaarantaisivat itseään. He eivät kuitenkaan osoittaneet oman edun edistämistä, kun vaihtoehtona oli vaarantaa muita jalankulkijoita pelastaakseen itsensä (Kallioinen, ym., 2019). Kaiken kaikkiaan tulokset ihmisen käyttäytymisestä itseohjautuviin autoihin liittyen ovat linjassa esimerkiksi Nyholmin ja Smidsin (2016) artikkelin pohdintojen ja Trafin (Innamaa, ym., 2015) tutkimuksen kanssa (Kallioinen, ym., 2019; Nyholm & Smids, 2016; Innamaa, ym., 2015). Tuloksien voidaan nähdä korostavan vaikeuksia sopeuttaa yleistä mielipidettä päätöksentekoalgoritmeihin.

Tulisiko kaikkien itseohjautuvien autojen olla eettisessä toiminnassaan samanlaisia? Pitäisikö itseohjautuvan auton ostajan pystyä päättämään, että haluavatko he autonsa yrittävän pelastaa heidät onnettomuustilanteessa, pyrkiä minimoimaan kokonaisvahinkoja vai toimia mahdollisesti jonkin muun periaatteen mukaisesti? Gogoll ja Müller (2017) väittävät, että ihmisillä on taipumus haluta kaikkien autojen olevan ohjelmoituja samojen asetusten mukaisesti. Tämä saavutettaisiin heidän mukaansa parhaiten, jos autot koordinoisivat törmäystilanteissa keskenään. Tietyt yleiset rajat saattaisivat kuitenkin olla oleellisia, joiden puitteissa kaikkien autojen on tehtävä valintansa siitä, miten toimia kolaritilanteessa (Gogoll & Müller, 2017). Nyholm pohtii artikkelissaan (2018) Sandbergin ja Bradshaw-Martinin (2013) näkemyksiä, että voisiko kenties olla olemassa välimuotoisia näkemyksiä itseohjautuvien autojen eettisistä toimintamalleista. Näkemyksien mukaan voisi olla tiettyjä yleisiä rajoja, joiden puitteissa kaikkien autojen on tehtävä valinnat kolaritilanteissa. Näiden rajojen sisällä jotkut ihmiset voisivat kuitenkin saada olla epäitsekkäämpiä kuin muilta odotetaan, jos se on tilanteen kannalta järkevästi perusteltua. Jonkinasteisen valinnan salliminen ihmisille voisi helpottaa heidän vastuunsa pitämistä mahdollisista huonoista seurauksista, joita heidän ajoneuvoihinsa liittyvät kolarit voisivat aiheuttaa. Näin ollen valinnan mahdollisuus saattaisi auttaa paikkaamaan mahdollisia vastuun aukkoja, joita itseohjautuvat autot saattaisivat avata (Nyholm, 2018).

Nyholm vertaa kollegansa (2016) kanssa itseohjautuvien autojen toimintaa perinteiseen "vaunuongelmaan". Vaunuongelmalla tarkoitetaan tässä artikkelissa seuraavaa kuvitteellista tilannetta: Tapauksessa itseohjautuva veturi on matkalla kohti viittä ihmistä, jotka ovat jumissa raiteilla ja kuolisivat, jos veturia ei ohjattaisi sivuraiteelle. Seisot kytkimen vieressä. Jos vedät kytkintä, veturi ohjataan sivuraiteelle. Ongelmana on, että tällä sivuraiteella on myös henkilö, ja tämä henkilö kuolee, jos vedät kytkintä ohjataksesi junan toisaalle. Hyvin yleinen vastaus tähän tilanteeseen on Nyholmin ja Smidsin mukaan, että on sallittua pelastaa viisi henkilöä ohjaamalla juna toisaalle ja näin tappaen yhden sen seurauksena. Toisenlaisessa tapauksessa viisi henkilöä taas pelastettaisiin työntämällä raiteiden vieressä oleva henkilö vaunun eteen pysäyttääkseen vaunun,

mutta tappamalla tönäistyn henkilön. Tämä tilanne ei yleisesti kuitenkaan saa viiden henkilön pelastamista tukevaa vastausta (Nyholm & Smids, 2016).

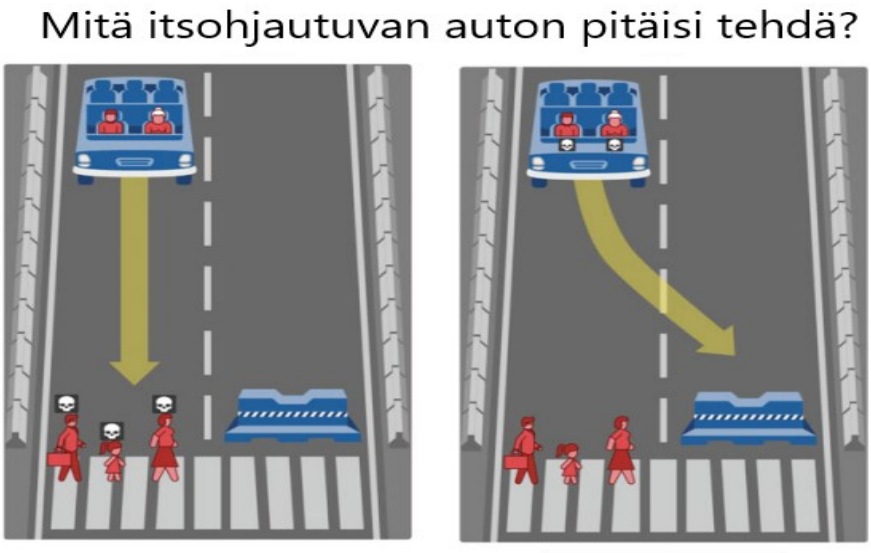
Nyholmin ja Smidsin (2016) artikkelissa tunnistetaan kolme tärkeää tapaa, joilla itseohjautuvien autojen algoritmien etiikka ja vaunuongelman filosofia eroavat toisistaan. Nämä kolme tapaa liittyvät: 1. perustavaan päätöksentekotilanteeseen, jonka kohtaavat itseohjautuvien ajoneuvojen algoritmien kehittämisessä mukana olevat henkilöt, 2. moraaliseen ja oikeudelliseen vastuuseen ja 3. päätöksentekoon riskien ja epävarmuuden edessä. Päätelmät ovat vahvasti samansuuntaisia kuin Gogollin ja Müllerin (2017) artikkelissa. Vaunuongelmassa yksilö asetetaan tilanteeseen, joka tapahtuu välittömästi. Heidän on tehtävä päätös paikan päällä, joko ohjata juna sivuraiteelle vetämällä kytkintä tai työntämällä toinen henkilö raiteille riippuen tilanteesta. Tämä vaatii hyvin nopeaa päätöksentekoa, mikä eroaa olennaisesti ennakoivasta päätöksenteosta ja suunnittelusta. Vertauskuva ei sisällytä algoritmien kehittämisen aikaista päätöksentekoa, jossa määritetään itseohjautuvien autojen toiminnan aikaisesta ennakoivasta päätöksenteosta (Nyholm & Smids, 2016; Gogoll & Müller, 2017).

Kun pohditaan moraalisesti merkityksellisiä päätöksentekotilanteita, niin voidaan nähdä enemmän yhtäläisyyksiä perinteisten autojen onnettomuustilanteiden ja vaunuongelmatapausten välillä kuin ennakoivan ohjelmoinnin sekä suunnittelun ja vaunuongelmatapausten välillä. Esimerkiksi perinteisen auton kuljettaja voi äkillisesti joutua tilanteeseen, jossa hänen on päätettävä paikan päällä, ohjaako hän autonsa yhteen henkilöön välttääkseen viiteen henkilöön törmäämisen, kuten aiemmin kuvatussa eettisessä ongelmassa. Tämä eroaa niistä tilanteista, joissa luodaan onnettomuusalgoritmeja itseohjautuville autoille.

Ei ole realistista myöskään ajatella, että päätöksenteko itseohjautuvien autojen ohjelmoinnista olisi vain yhden yksilön vastuulla, kuten vaunuongelman mukaan voisi kuvitella. Tätä lähestymistapaa ei välttämättä siis voida sovelluttaa itseohjautuviin autoihin ja niiden kehitystyöhön. Nyholmin ja Smidsin (2016) mukaan päätöksenteko itseohjautuvien autojen ohjelmoinnista on pikemminkin kollektiivinen ponnistus, johon osallistuvat useat eri sidosryhmät. Näitä sidosryhmiä voi olla esimerkiksi tavalliset kansalaiset, lakimiehet, etiikkaan erikoistuneet asiantuntijat, insinöörit, riskinarviointiasiantuntijat, autovalmistajat ja niin edelleen (Nyholm & Smids, 2016). Nämä sidosryhmät joutuvat neuvottelemaan ja saavuttamaan yhteisesti hyväksytyn ratkaisun ottaen huomioon erilaiset edut ja arvot, joita kullakin osapuolella mahtaa päätöksentekoprosessiin liittyen olla.

Bartneckin ja kollegoiden teoksessa (2021) huomautetaan näkökulmasta, jonka mukaan autonomisten ajoneuvojen olemassaolo saattaa vahingossa johtaa tiettyihin onnettomuuksiin, kun nämä ajoneuvot noudattavat tieliikennesääntöjä hyvin tarkasti. Esimerkiksi jos autonominen ajoneuvo ajaa moottoritiellä 55 mailin tuntinopeusrajoitusta, kun taas tyypillinen ihminen saattaisi mieluummin ajaa 70 mailia tunnissa, se saattaa pakottaa ihmiskuljettajia ohittamaan autonomisen ajoneuvon. Kerrotaan, että tämä saattaisi johtaa enemmän inhimillisen virheen aiheuttamiin onnettomuuksiin. Huomautettiin myös, ettei näissä

onnettomuuksissa voi täysin syyttää autonomisia autoja, koska ne noudattivat vain liikennesääntöjä, kun taas ihmiset eivät niin tehneet (Bartneck, ym., 2021). Myös Bartneckin ym. (2021) teoksen sekä Nyholmin ja Smidsin (2016) artikkelin mukaan, yksipuolisesti syyllisten etsiminen mahdollisiin onnettomuuksiin on väärä tapa lähestyä ongelmaa. Myös vaunuongelmaan liittyvät vertaukset ovat näissä teoksissa samansuuntaisia. Tällaisia vaunuongelman tapaisia tilanteita kerrotaan esiintyvän hyvin harvoin, sillä niiden edellytykset ovat erittäin erityiset. (Bartneck, ym., 2021; Nyholm & Smids, 2016). Bartneck ym. (2021) antavat esimerkin, että tällaisissa vaunuongelman tapaisissa tilanteissa eri kohteiden vahingoittumisen tai uhrien aiheutumisen todennäköisyydet voivat vaihdella. Tämä saattaa johtaa monimutkaisiin tilanteisiin, joissa ei ehkä ole eettisesti perusteltua ohittaa mahdollisuuksia vähentää kokonaisvahinkoja henkilöille (Kuvio 2). Itseohjautuvat autot pyrkivät pääsääntöisesti välttämään kaikki esteet ilman pyrkimystä tunnistaa niitä. Autot tietävät jo matkustajien määrän turvavyövaroitusjärjestelmän kautta, ja autojen välinen langaton viestintä on usein käytettävissä. Lähestyvän törmäyksen tapauksessa kahden auton olisi teknisesti mahdollista neuvotella ajokäyttäytymisestään, samankaltaisesti kuin ilmailussa käytetyn vastaavanlaisen järjestelmän avulla. Enemmän matkustajia kuljettava auto voitaisiin antaa etusijalle. Tällaisen järjestelmän käyttöönoton toivottavuus nähdään olevan kuitenkin yhteiskunnan päätettävissä (Bartneck, ym., 2021). Kuviossa 2 voidaan myös nähdä sisältyvän samanlainen asetelma kuin edellä mainitussa vaunuongelmassa on kuvattu.



Kuvio 2 Moraalinen dilemma. (Bartneck, ym., 2021)

Yleinen tietosuoja-asetus (Euroopan unioni, 2016) koskee vahvasti myös autonomisia ajoneuvoja. Datan keruun hyötyjä ei voida kiistää, mutta monet tehokkuutta edistävistä tiedoista voivat olla yksityisyyden kannalta kyseenalaisia. Autonomiset ajoneuvot keräävät tietoa liikkuessaan paikasta toiseen, ja tämä tieto voi myös sisältää havaintoja auton matkustajista. Näitä havaintoja voidaan kuitenkin hyödyntää esimerkiksi arvioitaessa, pitäisikö henkilön saada ajaa kyseistä ajoneuvoa. Teknologia voi estää henkilöitä, joilla on voimassa oleva ajokielto tai päihtymystila, ottamasta vastuun ajoneuvosta. Autonominen ajoneuvo voi itse tunnistaa, ilmentääkö henkilön käytös päihtymystä, ja tehdä sen perusteella päätöksen.

3.2 Tulevaisuuden näkymiä ja haasteita

Algoritmisten päätösten läpinäkymättömyys voi tuoda mukanaan haasteita ja mahdollisia vaaroja niiden vastuuttoman käytön seurauksena. Algoritmien roolin sekä vaikutusten kasvun myötä, niitä on yhteiskunnallisessa ja taloudellisessa elämässä välttämätöntä tarkastella huolellisesti, että kriittisesti (Kitchin, 2016). Zerillin ja kumppanien (2019) mukaan ihmisten ei tulisi luottaa sokeasti algoritmipohjaisiin päätöksentekotyökaluihin korkean riskin tai turvallisuuskriittisissä tilanteissa, elleivät kyseiset järjestelmät ole merkittävästi parempia kuin ihmiset kyseisellä päätöksentekoa koskevalla alueella. Heidän artikkelissaan esitetään strategioita niin sanotun “hallintahaasteen" ratkaisemiseksi. He pitävät lupaavimpana vaihtoehtona erittäin taitavien algoritmisten työkalujen ja ihmistoimijoiden yhteistyötä. Lisäksi määritellään avainperiaatteita, jotka tulisi ottaa huomioon kaikkien tällaisten ihmisen ja koneen yhteistoimintaan perustuvien järjestelmien suunnittelussa. Päätöksen automatisoinnissa tulisi huomioida, että ihmiset saattavat antaa liikaa valtaa algoritmille pelkästään sen toimiessa useimmissa tilanteissa hyvin. Ellei päätöksentekotyökalu ole ansainnut asemaansa, sitä ei tulisi käyttää turvallisuuskriittisissä tilanteissa. Sen sijaan tulisi harkita dynaamista ja toisiaan täydentävää tehtävien jakoa aktiivisesti osallistuvien ihmisten ja yksinkertaisempien, mutta luotettavampien, autonomisten järjestelmien välillä, jotta turvallisuus voitaisiin taata (Zerilli, ym., 2019).

Tietoteknisten työkalujen ja päätöksiä tukevien järjestelmien tulevaisuuden haasteita nähtiin Andersonin, ym. artikkelin (1993) mukaan jo usea vuosikymmen sitten. Nämä päätelmät ovat pysyneet nykyaikaan verrattuna hyvin samansuuntaisina verrattuna esimerkiksi Limin ja Taeihaghin artikkeliin (2019), jossa tutkittiin älykaupunkien eettisiä ja teknisiä huolia. Anderson ja kollegat pohtivat esimerkiksi luottamuksellisuuden, yksityisyyden, oikeudenmukaisuuden ja vastuun käsitteitä liittyen tietoteknisiin työkaluihin (Anderson, ym., 1993). Näiden ulottuvuuksien kanssa työskentely nähdään jatkuvan nykyään, että tulevaisuudessakin. Lim ja Taeihagh (2019) tarkastelivat näitä periaatteita myös älykaupunkien kontekstissa, sillä niitä joudutaan pohtimaan ja työstämään turvallisemman ympäristön rakentamiseksi. Usein tavoitteena on myös

integroida älykaupunkeihin liittyvät osat (esimerkiksi esineiden internet) ja älykkäät ajoneuvoteknologiat liikennejärjestelmään, jossa itseohjautuvilla ajoneuvoilla on keskeinen rooli. Tulevaisuuden kannalta kerrotaan, että näiden asioiden jatkuva kehittäminen voi viedä yhteiskuntaa jatkuvasti turvallisempaan suuntaan, kun koneoppimisen mahdollisuudet mahdollistavat automaattisten järjestelmien jatkuvan kehittymisen ja sopeutumisen päätöksentekoprosesseihin ympäristön muutosten mukaisesti. Näin voidaan nähdä tapahtuvan myös yhdessä esineiden internetin laitteiden sekä autonomisten ajoneuvojen kanssa. Täten pystytään mahdollistamaan liikennepalvelujen ja muiden palvelujen tarjoamisen ja räätälöimisen muuttuviin kuluttajatarpeisiin (Lim & Taeihagh, 2019). Voidaan siis todeta monen lähtökohdan pysyneen lähes muuttumattomana liittyen algoritmien kehittämiseen parempien lopputulosten saavuttamiseksi, vaikka kehittyvän ympäristön tuomat haasteet muuttavatkin ajan saatossa.

Bakhshi Germi ja Rahtu (2022) kuvailevat ja pyrkivät ratkaisemaan huolia syväoppimisalgoritmien turvallisuushuolista niiden kehittyessä kysynnän vuoksi ratkaisemaan tehtäväkohtaisia ongelmia. Toimintojen monimutkaistuessa turvallisuus voi olla mahdoton varmistaa algoritmien läpinäkymättömyyden vuoksi (Bakhshi Germi & Rahtu, 2022). Tämä avaa näkökulmaa tulevaisuuden haasteiden ymmärtämiselle ja antaa jatkokehitykselle suuntaa. Myös erilaisten päätöksentekomallien sisällyttäminen järjestelmistä saatavaan palautteeseen voi olla välttämätöntä muuttuvien päätöksentekoprosessien seurausten ymmärtämiseksi. Tolanin (2018) mukaan lisätutkimusta tarvitaan automaattisten päätöksentekojärjestelmien sovellusten osalta, jotta voidaan ymmärtää miten ihmiset tekevät päätöksiä koneavustuksen kanssa. Tämä tarkoittaa myös sitä, että algoritmeja on arvioitava usein uudelleen. Mikään algoritmi ei voi olla ikuisesti oikeudenmukainen ilman tilanteen mukaista säätöä ja muutosta. Monimuotoisuuden lisääntymisestä tulee tulevaisuudessa yhä olennaisempaa (Tolan, 2018).

4 JOHTOPÄÄTÖKSET

Tässä luvussa käydään läpi edellä käsiteltyjä aiheita, samalla luoden johtopäätöksiä asioiden suhteista. Tavoitteena on tiivistää keskeiset päätelmät, jotta saadaan vastauksia tutkielman ensisijaisiin tutkimuskysymyksiin. Nämä tutkimuskysymykset ovat: Millaisia oikeudenmukaisuuteen ja turvallisuuteen liittyviä ongelmia voidaan nähdä algoritmisessa päätöksenteossa? Mitä keinoja oikeudenmukaisuuden ja turvallisuuden varmistamiseksi voidaan käyttää algoritmisten päätöksentekojärjestelmien kehittämisessä?

Todettiin, että oikeudenmukaisuuden saavuttaminen algoritmisissa järjestelmissä ei ole yksiselitteistä, sillä jokainen tilanne vaatii harkittuja ratkaisuja huomioiden moninaisia näkökulmia. Tästä syystä on tärkeää, että automaattisten päätöksentekojärjestelmien kehittäjät työskentelevät tiiviisti yhdessä alan asiantuntijoiden kanssa. Sekä Tolanin (2018), että Nyholmin ja Smidsin (2016) mukaan tämä yhteistyö on avainasemassa, jotta nämä järjestelmät voivat toteuttaa oikeudenmukaisuutta parhaalla mahdollisella tavalla. Lisäksi on välttämätöntä, että nämä päätöksentekojärjestelmät ovat läpinäkyviä ja selitettäviä, jotta niiden toiminta ja päätöksenteko ovat ymmärrettävissä kaikille asianosaisille. (Tolan, 2018; Nyholm & Smids, 2016).

Nyholm ja Smids (2018) sekä Gogoll ja Müller (2017) pohtivat itseohjautuvien autojen eettisiä ongelmia verraten niitä vaunuongelmaan. Nähtiin, että vaikka vaunuongelma voisi tarjota hyödyllisen vertauskuvan, se kuitenkin jättää usein huomioimatta monimutkaisuuden, joka liittyy itseohjautuvien autojen toimintaan ja haasteisiin. Todellisuudessa autonomisten ajoneuvojen ongelmat ovat syvemmällä tasolla, kietoutuneina monenlaisiin muuttujiin ja tilanteisiin, jotka ulottuvat yli yksinkertaisen vertauskuvan kattaman näkökulman. (Nyholm, 2018; Gogoll & Müller, 2017).

Algoritmien läpinäkyvyyden ja selitettävyyden puute muodostaa merkittävän esteen oikeudenmukaisuuden toteutumiselle. Mustan laatikon ongelmaa kuvattiin Hakkaraisen ym. (2020) sekä Krafftin ym. (2022) artikkeleissa. Tämä on tilanne, jossa algoritmin päätökset ovat ikään kuin suljettuina laatikon sisälle. Sen voidaan nähdä johtavan siihen, että edes algoritmin kehittäjät eivät aina ymmärrä, miten tietty lopputulos tai päätös on saavutettu. Epätietoisuus vai-

keuttaa järjestelmän toiminnan arviointia ja voi lisätä epäluottamusta sen päätöksiin. Ratkaisu vaatii kehittäjiä avaamaan näitä prosesseja ja luomaan mekanismeja, jotka mahdollistavat päätösten jäljitettävyyden ja ymmärrettävyyden kaikille osapuolille (Hakkarainen, ym., 2020; Krafft, ym., 2022).

Vaikka oikeudenmukaisuuteen sekä turvallisuuteen liittyvät ongelmat ovat osittain erilaisia, niiden tekijät eivät kuitenkaan eroa toisistaan juurikaan. Olipa kyse sitten eettisistä seikoista, niin kuin algoritmisesta syrjinnästä (Hakkarainen, ym., 2020; Davidson & Spinoulas, 2015), tai turvallisuuteen liittyvästä pohdinnasta autonomisten ajoneuvojen kolaritilanteista (Kallioinen, ym., 2019; Nyholm & Smids, 2016), voidaan näiden ongelmien taustalla nähdä samoja tekijöitä. Ihmisen luottamus ihmisestä erillään olevaan päätöksentekijään, sekä vastuunkantoon liittyvät ongelmat, toistuvat monessa tapauksessa. Lisäksi molempiin kokonaisuuksiin liittyy pohdintaa siitä, että miten algoritmeja kuuluisi rakentaa, minkä perusteella niiden tulisi tehdä päätöksensä sekä miten algoritmien käyttöä kuuluisi säädellä (Goodman & Flaxman, 2017; Euroopan unioni, 2016). Loppujen lopuksi tavoitteena on löytää yhteiskunnan kannalta parhaat puitteet automaattisten päätösten hyödyntämiseen. Olipa tapauksessa kyse sitten kolarien välttämisestä tai vaikka yksilöiden oikeuksien suojelemisesta tiedonkeruun optimoituessa.

Voidaan siis todeta, että oikeudenmukaisuuden kontekstissa, algoritminen päätöksenteko tuo yhteiskuntaan lukuisia etuja sekä vanhoja ongelmia korjaavia tekijöitä. Esimerkiksi se voi poistaa ihmisen tuomaa subjektiivisuutta päätöksenteosta (Binns, 2022). Toisaalta algoritmisella päätöksenteolla voi olla myös negatiivisia vaikutuksia oikeudenmukaisuuteen. Algoritmit perustuvat usein aikaisempaan dataan, ja jos tämä data sisältää vääristymiä tai syrjiviä elementtejä, algoritmit voivat vahvistaa näitä epäkohtia (Hakkarainen, ym., 2020). Tämän seurauksesta oikeudenmukaisuuden turvaamiseksi teknologian kehittyessä joudutaan tekemään paljon työtä yhteiskuntaa tukevien säädösten laatimista varten. Turvallisuuteen liittyviä hyötyjä voidaan myös saada algoritmisia päätöksentekojärjestelmiä hyödyntämällä. Bakhshi Lomer (2023) esimerkiksi osoitti päätöksentekojärjestelmän hyödyntämisen tuovan etuja hätämajoituspaikkojen sijoittamiseen maanjäristyksiltä suojautumisessa. Lisäksi lukuisia etuja voidaan tavoittaa liikenteen sujuvoittamisessa hyödyntämällä autonomisia ajoneuvoja (Innamaa, ym., 2015). Algoritmisten päätöksentekojärjestelmien rakentaminen ilman kontekstiin sidonnaista harkintaa ja tutkimustyötä tai läpinäkyvyyttä saattaa aiheuttaa vaikeuksia ja mahdollisia turvallisuusriskejä. Algoritmien kasvaneen merkityksen myötä niiden roolia ja vaikutuksia on välttämätöntä arvioida huolellisesti ja kriittisesti.

5 YHTEENVETO

Tämä tutkielma keskittyi algoritmisen päätöksenteon oikeudenmukaisuuteen ja turvallisuuteen, ja sen tavoitteena oli syventää ymmärrystä aiheen nykytilasta ja haasteista. Tutkimus pyrki vastaamaan kysymyksiin liittyen algoritmisten päätösten ongelmiin sekä niiden mahdollisiin ratkaisuihin oikeudenmukaisuuden ja turvallisuuden näkökulmista. Täten pyrittiin tarjoamaan uutta näkökulmaa algoritmeihin liittyvään keskusteluun. Aluksi oikeudenmukaisuuden käsitettä tutkittiin erilaisten säädösten kautta. Tämän jälkeen pohdittiin eettisiä tekijöitä, joihin sisältyi aiheeseen liittyvässä keskustelussa ilmeneviä aihepiirejä, kuten algoritminen syrjintä. Seuraavaksi aihetta käsiteltiin turvallisuuden näkökulmasta ottaen tarkastelupisteeksi autonomiset ajoneuvot ja niiden turvallisuuteen liittyvät pohdinnat sekä tulevaisuuden haasteet. Lopuksi tarkasteltiin yleisesti löydöksien suhteita yhdistäen algoritmisen päätöksenteon oikeudenmukaisuuden ja turvallisuuden aikaisemmin erillisinä käsiteltyjä käsitteitä.

Tämä tutkimus suoritettiin kirjallisuuskatsauksena. Aiheen rajaamisen jälkeen kirjallisuutta etsittiin hyödyntämällä internetin hakupalveluja ja suoraan tieteellisiä julkaisuja tarjoavia tietokantoja. Hakusanoina käytettiin seuraavia: "*Algoritminen päätöksenteko*", "*Algoritmisen päätöksenteon oikeudenmukaisuus*", "*Algoritmisen päätöksenteon turvallisuus*" tai vastaavia englanninkielisiä termejä kuten "*Algorithmic decision-making*", "*Security in algorithmic decisions*" ja "*Fairness in algorithmic decision-making*". Kirjallisuutta valittaessa painopiste kohdistui ensisijaisesti materiaaleihin, jotka käsittelivät laajasti tutkielman aiheeseen liittyviä käsitteitä. Näihin käsitteisiin kuului pääosin algoritmien oikeudenmukaisuuteen ja turvallisuuteen liittyviä käsitteitä, kuten esimerkiksi algoritminen syrjintä. Lisäksi valintaan otettiin mukaan kirjallisuutta, joka syventyi tarkemmin eri osa-alueisiin erillisinä aiheinaan. Tähän kuului esimerkiksi itseohjautuvien autojen käsittelyyn liittyvä aineisto. Tutkielmassa keskityttiin pääasiassa käsitteellisiin selityksiin ja laadullisiin näkökulmiin aiheesta eikä teknisiin yksityiskohtiin algoritmien osalta paneuduttu tarkemmin. Tästä syystä materiaalit, jotka keskittyvät pelkästään teknisiin käsitteisiin, jätettiin täysin tutkielman ulkopuolelle.

Tutkielman keskeiset tulokset korostivat monia näkökohtia algoritmisessa päätöksenteossa. Tiivis yhteistyö eri alojen asiantuntijoiden välillä nähtiin olevan avainasemassa oikeudenmukaisuuden varmistamisessa algoritmisissa päätöksentekojärjestelmissä. Vaikka datavetoiset järjestelmät voivat auttaa vähentämään inhimillistä subjektiivisuutta päätöksenteossa, ne tuovat mukanaan eettisiä haasteita, joiden ratkaisemiseen tarvitaan monien eri sidosryhmien yhteistyötä. Algoritmien läpinäkyvyyden nähtiin myös olevan keskeisessä asemassa oikeudenmukaisuuden edistämisessä, erityisesti algoritmisen syrjinnän tapauksissa. Tämän lisäksi algoritmisten päätöksentekojärjestelmien sääntelyn havaittiin olevan vaikeaa, koska teknologia kehittyy nopeasti ja ala on vielä historiallisesti hyvin tuore.

Automaattisten päätösten todettiin pystyvän parantamaan yksilöiden turvallisuutta, mutta samalla myös nähtiin, että ne voivat aiheuttaa paljon vastuunkantoon liittyviä ongelmia. Itseohjautuvien autojen tutkiminen nähtiin hyvänä näkökulmana turvallisuuden tarkasteluun algoritmisiin päätöksentekojärjestelmiin liittyen. Vaunuongelman sisällön analyysin jälkeen se voitiin nähdä liian yksinkertaisena vertauksena kuvaamaan todellisia eettisiä ongelmia itseohjautuvien ajoneuvojen toiminnassa. Todettiin myös, että sekä oikeudenmukaisuuteen että turvallisuuteen liittyvät pohdinnat käsittelevät samankaltaisia teemoja algoritmisiin päätöksiin liittyen. Merkittävimpänä yhteisenä teemana nähtiin olevan algoritmien läpinäkyvyyden tärkeys.

Tulevaisuuden tutkimuksessa olisi tärkeää edistää laajempaa yhteistyötä eri alojen asiantuntijoiden välillä. Algoritmien kehittämiseen olisi suotavaa sisällyttää mukaan monipuolisesti eri alojen osaajia, jotta kehitystyö olisi monipuolisempaa ja se huomioisi laajemmin erilaisia näkökulmia. Verraten algoritmisten päätöksentekojärjestelmien laajaa käytön määrää eri toimialoilla ja nykyistä sekä aikaisemman tutkimuksen määrää, voidaan todeta välttämättömäksi panostaa enemmän tutkimukseen ja kehitystyöhön algoritmien osalta. Tämä on tarpeen, jotta voidaan syvällisemmin ymmärtää automaattisten päätöksentekojärjestelmien toimintaa ja vaikutuksia sekä kehittää niitä vastaamaan monimuotoisiin tarpeisiin entistäkin tehokkaammin.

LÄHTEET

Al-Rashid, F., Totzeck, M., Mahabadi, A.A., Johannsen, L., Luedike, P., Lind, A., Krueger, A., Kamler, M., Kahlert, P., Jánosi, R.A., Heusch, G., Rassaf, T. (2020). Safety and efficacy of a novel algorithm to guide decision-making in high-risk interventional coronary procedures. *International Journal of Cardiology,* Volume 299, Pages 87-92. https://doi.org/10.1016/j.ijcard.2019.08.056

Anderson, R, D. Johnson, D. Gotterbarn and J. Perrolle (1993). "Using the New ACM Code of Ethics in Decision Making". *Communications of the ACM*, 36: 98–107. https://doi.org/10.1145/151220.151231

Bakhshi Germi, S., Rahtu, E. (2022). A Practical Overview of Safety Concerns and Mitigation Methods for Visual Deep Learning Algorithms. *SafeAI 2022: Proceedings of the Workshop on Artificial Intelligence Safety 2022.* https://ceur-ws.org/Vol-3087/paper_8.pdf

Bakhshi Lomer, A.R., Rezaeian, M., Rezaei, H., Lorestani, A., Mijani, N., Mahdad, M., Raeisi, A., Arsanjani, J.J. (2023). Optimizing Emergency Shelter Selection in Earthquakes Using a Risk-Driven Large Group DecisionMaking Support System. *Sustainability 2023.* 15(5), 4019. https://doi.org/10.3390/su15054019

Bartneck, C., Lütge, C., Wagner, A., Welsh, S. (2021). *An Introduction to Ethics in Robotics and AI.* Springer Nature, SpringerBriefs in Ethics.

Binns, R. 2022. Human Judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance,* 16: 197-211. https://doi.org/10.1111/rego.12358

Bischoff, J. & Maciejewski, M. 2016. Simulation of City-wide Replacement of Private Cars with Autonomous Taxis in Berlin. *Procedia Computer Science,* Volume 83, 2016, Pages 237-244, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2016.04.121

Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A. & Vaithianathan, R. (2019). Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-making in Child Welfare Services. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). *Association for Computing Machinery,* New York, NY, USA, Paper 41, 1–12. https://doi.org/10.1145/3290605.3300271

Bynum, T., (2018). Computer and Information Ethics, *The Stanford Encyclopedia of Philosophy,* Summer 2018 Edition, Edward N. Zalta (ed.). https://plato.stanford.edu/archives/sum2018/entries/ethics-computer/

Davidson, P., Spinoulas, A. (2015). Autonomous vehicles – What could this mean for the future of transport? *Australian Institute of Traffic Planning and Management (AITPM) National Conference, 2015,* Brisbane, Queensland, Australia.

de Bruijn, H., Warnier, M., Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making, *Government Information Quarterly,* Volume 39, Issue 2. https://doi.org/10.1016/j.giq.2021.101666

Dietvorst, B.J, Simmons, J.P and Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology*: General144(1), 114–126. https://doi.org/10.1037/xge0000033

Euroopan unioni. General Data Protection Regulation GDPR. (2016). https://gdpr-info.eu/

European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies. (2020). https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html

Gogoll, J. & Müller, J. (2017). Autonomous Cars: In Favor of a Mandatory Ethics Setting. *Science and Engineering Ethics*, 23(3), 681–700. https://doi.org/10.1007/s11948-016-9806-x

Goodman, B., Flaxman, S. (2017). European Union Regulations on Algorithmic Decision Making and a "Right to Explanation". *AI Magazine,* Volume 38, Issue 3. https://doi.org/10.1609/aimag.v38i3.2741

Hakkarainen, J., Koulu, R., Markkanen, K. (2020). Läpinäkyvät algoritmit? Lähdekoodin julkisuus ja laillisuuskontrolli hallinnon digitalisaatiossa. *Edilex,* 2020/18. https://www.edilex.fi/artikkelit/21042.pdf

Harris, J. & Davenport, T. (2005). Automated Decision Making Comes of Age. *MIT Sloan Management Review,* 46.

Innamaa, S., Kanner, H., Rämä, P., Virtanen, A. (2015). Automaation lisääntymisen vaikutukset tieliikenteessä. *Trafin tutkimuksia* No. 01/2015 https://www.trafi.fi/tietopalvelut/julkaisut/2015_tutkimukset/automaattiajaminen

Kallioinen, N., Pershina, M., Zeiser, J., Nosrat Nezami, F., Pipa, G., Stephan, A., König, P. (2019). Moral Judgements on the Actions of Self-Driving Cars and Human Drivers in Dilemma Situations From Different Perspectives. *Front. Psychol.* https://doi.org/10.3389/fpsyg.2019.02415

Kitchin, R. (2016). Thinking critically about and researching algorithms. Information, *Communication & Society,* 20. 1-16. https://doi.org/10.1080/1369118X.2016.1154087

Koulu, R., Mäihäniemi, B., Kyyrönen, V., Hakkarainen, J., Markkanen, K. (2019). Algoritmi päätöksentekijänä? : Tekoälyn hyödyntämisen mahdollisuudet ja haasteet kansallisessa sääntelyympäristössä. *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja,* 44, Valtioneuvoston kanslia. http://urn.fi/URN:ISBN:978-952-287-764-2

Krafft, T.D., Zweig, K.A. & König, P.D. (2022). How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regulation & Governance,* 16: 119-136. https://doi.org/10.1111/rego.12369

Köchling, A., Wehner, M.C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Bus Res 13,* 795–848 (2020). https://doi.org/10.1007/s40685-020-00134-w

Lim, H.S.M, Taeihagh, A. (2019). Algorithmic Decision-Making in AVs: Understanding Ethical and Technical Concerns for Smart Cities. *Sustainability*, 2019; 11(20):5791. https://doi.org/10.3390/su11205791

Lintilä, M., Vasamo-Koskinen, S. Työ- ja elinkeinoministeriö. (2021). *Valtioneuvoston U-kirjelmä*, U 28/2021 vp. Eduskunta.

Moor, J. (2007). What Is Computer Ethics?. *Metaphilosophy,* 16. 266 - 275. https://doi.org/10.1111/j.1467-9973.1985.tb00173.x

Nyholm, S. (2018). The ethics of crashes with self-driving cars: A roadmap, I. *Philosophy Compass,* 2018; 13:e12507. https://doi.org/10.1111/phc3.12507

Nyholm, S., Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem? *Ethic Theory Moral Prac 19,* 1275–1289 (2016). https://doi.org/10.1007/s10677-016-9745-2

Ojanen, A., Sahlgren, O., Vaiste, J., Björk, A., Mikkonen, J., Kimppa, K., Laitinen, A., Oljakka, N. (2022). Algoritminen syrjintä ja yhdenvertaisuuden edistäminen : Arviointikehikko syrjimättömälle tekoälylle. *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja,* Vol. 2022, No. 54. http://www.urn.fi/URN:ISBN:978-952-383-404-0

Sandberg, A., & Bradshaw‐Martin, H. (2013). In J. Romportl, et al. (Eds.), What do cars think of trolley problems: Ethics forautonomous cars?. *Beyond AI: Artificial Golem Intelligence,* Proceedings of the International Conference Beyond AI 2013 Pilsen, Czech Republic, November 12–14, 2013.

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, 9(2). https://doi.org/10.1177/20539517221115189

Tolan, S. (2018). Fair and unbiased algorithmic decision making: Current state and future challenges, *JRC Digital Economy Working Paper,* No. 2018-10, European Commission, Joint Research Centre (JRC), Seville. http://hdl.handle.net/10419/227696

Zerilli, J., Knott, A., Maclaurin, J., Gavaghan, C. (2019). Algorithmic Decision-Making and the Control Problem. *Minds & Machines 29,* 555–578 https://doi.org/10.1007/s11023-019-09513-7

Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. *Council of Europe, Directorate General of Democracy.* https://hdl.handle.net/11245.1/7bdabff5-c1d9-484f-81f2-e469e03e2360