

VAASAN YLIOPISTO
TEKNIKAN JA INNOVAATIOJOHTAMISEN YKSIKKÖ
TIETOJÄRJESTELMÄTIEDE

Sinikka Sysky

FENOMENOGRAFINEN TUTKIMUS ALGORITMIEN ETIIKASTA

Tietojärjestelmätieteen
pro gradu -tutkielma

Teknisen viestinnän maasterikoulutusohjelma

VAASA 2019

SISÄLLYSLUETTELO

	sivu
1. JOHDANTO	6
1.1 Tutkimuksen tarkoitus ja tavoite	7
1.2 Tutkimusaineisto ja -menetelmä	7
2. ALGORITMIT	10
2.1 Mitä ovat algoritmit	10
2.2 Algoritmien eettiset ohjeet	12
3. INFORMAATIOTEKNOLOGIAN FILOSOFIA	14
3.1 Yleisiä IT-filosofian näkökulmia	14
3.2 Tietokone-etiikan historia	16
3.3 Yleistä tietokone-etiikasta	18
3.4 Tietokone-etiikkaan liitettyjä eettisiä näkökulmia	19
3.5 Floridin informaatioetiikka	20
4. ETIIKKA	22
4.1 Etiikan käsitteestä	22
4.2 Analyyttinen etiikka	23
4.3 Normatiivinen etiikka	25
4.4 Tutkimusetiikka soveltavan etiikan esimerkkinä	26
4.5 Deskriptiivinen etiikka	27
5. FENOMENOGRAFINEN TUTKIMUS	29
5.1 Fenomenografia tutkimusmenetelmänä	29
5.2 Fenomenografinen tutkimusprosessi	30
5.3 Aineiston hankinta ja analyysi	31
5.4 Fenomenografisen tutkimuksen reliabiliteetti ja validiteetti	33
6. ANALYYSI JA TULOKSET	34
6.1 Algoritminen ympäristö	34
6.1.1 Perusominaisuudet	34

6.1.2	Toiminta	35
6.1.3	Vaikutukset	39
6.1.4	Tunteminen	40
6.2	Algoritmit ja etiikka	42
6.2.1	Kontrolli	42
6.2.2	Sokaiseva pimeys	44
6.2.3	Digitaalinen identiteetti	46
6.2.4	Algoritminen toiminta	49
6.2.5	Eettinen toiminta	50
6.3	Synteesi	52
6.3.1	Yhdistävät tekijät ”Algoritminen ympäristö” -kategoriassa	52
6.3.2	Yhdistävät tekijät ”Algoritmit ja etiikka” -kategoriassa	54
6.3.3	Yhteenvedo	58
7.	DISKUSSIO	60
7.1	Merkittävimmät tutkimustulokset	60
7.2	Tutkimuksen tarkoitus ja tavoitteet	61
7.3	Tulosten merkitys ja jatkotutkimusaiheet	62
7.4	Tämän tutkimuksen reliabiliteetti ja validiteetti	63
	LÄHDELUETTELO	65

VAASAN YLIOPISTO**Tekniikan ja innovaatiojohtamisen yksikkö**

Tekijä:	Sinikka Sysky	
Tutkielman nimi:	Fenomenografinen tutkimus algoritmien etiikasta	
Ohjaajan nimi:	Tero Vartiainen	
Tutkinto:	Kauppätieteiden maisteri	
Ohjelma:	Teknisen viestinnän maisterikoulutusohjelma	
Pääaine:	Tietojärjestelmätiede	
Opintojen aloitusvuosi:	2017	
Tutkielman valmistumisvuosi:	2019	Sivumäärä: 71

TIIVISTELMÄ:

Elämme algoritmista aikaa. Algoritmien vaikutus ulottuu kaikkialle yhteiskunnassa. Algoritmit ohjaavat uutissyötteitä ja sosiaalisessa mediassa olevia henkilökohtaisesti kohdennettuja mainoksia ja postauksia, päättävät kuka pääsee työhaastatteluun tai saa lainaa, tarkastavat veroilmoituksia, ja paljon muuta. Keskustelu algoritmien vaikutuksesta on levinnyt laajalle parina viime vuonna. Tässä tutkimuksessa kartoitettiin neljästätoista tutkimusdokumentista tutkijoiden käsityksiä ja tutkimustuloksia algoritmien eettisyydestä vuosina 2016–2018. Tehdyllä aikarajauksella saatiin ajankohtainen kokonaiskuva.

Teoreettisessa viitekehyksessä kerrottiin mitä algoritmit ovat, kuinka informaatioteknologian filosofian kenttä näyttäytyi sen erityisenä painopisteenä tietokone-etiikka sekä miten etiikka yleisesti taustoitti eettisiä kysymyksiä. Fenomenografisessa analyysissä pureuduttiin tutkijoiden käsityksiin algoritmien etiikasta, etsittiin käsityksistä merkityssisältöjä ja tulkittiin ne. Näin muodostettua ensimmäistä kategoriaa luokiteltiin ja työstettiin edelleen, jotta aikaan saatiin toinen kategoria. Tutkimusaineiston kanssa tapahtuvan jatkuvan dialogin pohjalta saaduista tuloksista muodostettiin lopulta yhteinen tulosavaruus, joka tässä tapauksessa oli synteesi.

Tutkimuksen mukaan algoritmien merkittävimmät huolenaiheet liittyivät algoritmiseen päätöksentekoon ja sen aiheuttamiin vaihteleviin vaikutuksiin. Eettiset vaikutukset olivat tulleet esille nimenomaan tutkijakunnan taholta. Suurin ongelma oli yhteydessä siihen, että algoritmien arvolatausta ei tunnustettu. Tämän työn tutkimusaineistosta nousi vahva viesti kattavan algoritmien etiikan muodostamisesta. Erilaisia jatkotutkimusaiheita tutkimusmateriaalissa esitettiin runsaasti. Aineistossa esiintyi tarve monimuotoisemmalle ja poikkitieteelliselle tutkimukselle. Tutkimusaineistosta ilmeni myös, että erilaiset tutkimusryhmät olivat kehittäneet omia ratkaisuja joihinkin rajattuihin algoritmisiin ongelmiin. Kaikki tämä tietämys tulisi saada saman kokonaismallin alle. Algoritmien eettiset periaatteet saattavat ehkäistä ongelmatilanteita ja lisätä luottamusta digitaalisessa toimintaympäristössä.

AVAINSANAT: algoritmi, algoritmiikka, etiikka, fenomenografia

UNIVERSITY OF VAASA**School of Technology and Innovations**

Author: Sinikka Sysky
Topic of the Master's Thesis: Phenomenographic study of the ethics of algorithms
Instructor: Tero Vartiainen
Degree: Master of Science in Economics and Business Administration
Degree Programme: Degree Programme in Technical Communication
Major: Information Systems
Year of Entering the University: 2017
Year of Completing the Master's Thesis: 2019 **Pages:** 71

ABSTRACT:

We are living in an algorithmic era. The effect of algorithms extends throughout society. Algorithms guide news feeds and personally targeted ads and postings on social media, decide who can get a job interview or get a loan, check tax returns, and more. The discussion about the effect of algorithms has been widespread in the past couple of years. In this study, researchers' perceptions and research results on the ethics of algorithms in 2016–2018 are surveyed. The time constraint made gives the current overall picture.

The theoretical framework explained what algorithms are, how the field of information technology philosophy appears as its particular focus on computer ethics, and how ethics generally deal with ethical issues. Phenomenographic analysis focused on the scientists' perceptions of the ethics of algorithms, searching for and understanding conceptual content. The first category thus formed were categorized and further processed to provide second category. The results obtained from the ongoing dialogue with the research material eventually was formed a common space, which in this case was a synthesis.

According to the study, the most important concerns of algorithms were related to algorithmic decision-making and its varying effects. Ethical effects had come to light from the research community. The biggest problem was that the value-laden algorithm was not recognized. The research material of this work raised a strong message about the development of the ethics of algorithms. A variety of further research topics were presented in the research material. There was a need for more diverse and cross-disciplinary research. The research material also revealed that different research groups have developed their own solutions to some limited algorithmic problems. All this knowledge should be receivable under the same overall model. The ethical principles of algorithms might prevent problems and could increase confidence in the digital environment.

KEYWORDS: algorithm, algorithmics, ethics, phenomenography

1. JOHDANTO

Algoritmeja itsessään on tutkittu paljon. Niistä on tehty runsaasti pro graduja ja väitöskirjoja sekä muuta tieteellistä tutkimusta. Algoritmien etiikkaan kohdistuva tutkimus on jäänyt vähemmälle. Viimeisen muutaman vuoden aikana algoritmien etiikka on herättänyt kasvavaa kiinnostusta tutkijoiden parissa niin meillä kuin muualla. Mittelstadt, Allo, Taddeo, Wachter ja Floridi (2016) ovat kartoittaneet laajasti käytävää algoritmien etiikka-keskustelua ja esittävät yhteenvedona algoritmien eettisen kartan nykyisistä ongelma-alueista. Heidän mukaansa algoritmien eettiset ongelmat koskevat keskeneräisyyttä, läpinäkymättömyyttä, harhaanjohtavuutta, syrjiviä tuloksia, muuttuvia vaikutuksia ja jäljitämättömyyttä eli ns. mustia laatikoita. He esittävät useita lisätutkimusaiheita.

Toisenlaista näkökulmaa edustavat Kraemer, van Overveld ja Peterson (2011), jotka ovat tutkineet magneettikuvauksessa käytettävää lääketieteellistä kuvausalgoritmia ja sen antamia vääriä positiivisia ja vääriä negatiivisia tuloksia. Tutkijat kysyvät, kenen on vastuu, jos väärän negatiivisen tuloksen saanut potilas onkin vakavasti sairas ja jää hoidotta väärän tuloksen takia. Toisaalta väärän positiivisen tuloksen saanut voi pahimmassa tapauksessa menettää elämänhalunsa kokonaan, sen lisäksi, että häntä hoidetaan kalliisti ja turhaan. Tutkijat peräävät algoritmeihin läpinäkyvyyttä ja tunnistettavuutta sekä eettisten päätösten jättämistä käyttäjille, ei algoritmille.

Algoritmien eettiset näkökohdat näyttävät erilaisina eri alojen ammattilaisille. Ohjelmoija ei mieti syvällisiä moraalisia kysymyksiä keskittyessään ratkaisemaan algoritmista ongelmaa. Yhteiskuntatieteilijää sen sijaan kiinnostaa kovasti, mitkä ovat esimerkiksi algoritmien päätöksenteon vaikutukset yksilöihin ja laajemmin yhteiskuntaan. Jo edellä olevista tutkimuksista nähdään, että algoritmeihin liittyvät eettiset kysymykset ovat moniulotteisia, ja harvoin vain yhden ammattiryhmän ratkaistavissa. Laajemmassa mittakaavassa ollaan vasta heräilemässä informaatioteknologian ja lisääntyvän digitaalisuuden tuomiin eettisiin kysymyksiin.

1.1 Tutkimuksen tarkoitus ja tavoite

Tutkimuksen tarkoituksena on kartoittaa tutkijoiden käsityksiä ja tutkimustuloksia algoritmien eettisyydestä vuosina 2016–2018 ja näin muodostaa ajankohtainen kokonaiskuva. Mitään yhtenäistä, muotoon puristettua algoritmien etiikkaa ei ole olemassa, jota voitaisiin helposti kääntelemällä tarkastella eri puolilta. Kirjallisuudessa algoritmien etiikka esiintyy usein vastakohtanaan, puutteen tai ongelman kautta. Näistä puutteista ja ongelmakohdista nousevat esiin ne eettiset kysymykset, joita algoritmien parissa on.

Työn tavoitteena on tehdä näkyväksi algoritmeihin liitettyjä eettisiä ongelmia. Tästä syystä informaatiotutkimuksen filosofiaa selvitetään keskeisin osin ennen taustoittavaa etiikan osuutta. Merkittävän osan informaatioteknologian filosofian tarkastelusta muodostaa laaja tietokone-etiikan osuus. Tietokone-etiikka muodostaa tämän tutkimuksen selkärangan, jonka päälle muu rakentuu. Etiikan luku antaa laajempaa perspektiiviä ja suhteuttaa tarkastelua. Vuorovaikutteisen ja aineistolähtöisen tutkimusmenetelmän ansiosta voidaan tavoittaa uutta tietämystä tutkimusaiheesta. On oletettavaa, että algoritmien tutkimuskentästä löytyy useita jatkotutkimusaiheita.

1.2 Tutkimusaineisto ja -menetelmä

Tutkimuskysymys *mitä on tutkittu algoritmien etiikasta* syntyi ensimmäisessä tapaamisessa ohjaajan kanssa. Toisin sanoen tutkimuskysymys oli jo olemassa ennen teoreettista perehtymistä informaatioteknologian filosofiaan ja etiikkaan. Materiaali on kerätty syysmarraskuussa 2018. Joitakin yksittäisiä teorialäydennyksiä on haettu vielä joulukuussa 2018. Olen hyödyntänyt haun alkuvaiheessa Google Scholaria. Olen käyttänyt hyödykseni kirjallisuusluetteloita sekä tieteellisistä artikkeleista että kirjoista. Teoreettisen viitekehyksen aineisto koostuu tietokannoista haetuista tieteellisistä artikkeleista, joistakin internet-lähteistä ja e-kirjoista sekä painetuista kirjoista, jotka on lainattu kirjastoista. Tietokannoista on haettu tutkimusmateriaalia pääasiassa hakusanoilla ”ethics” ”algorithm” sekä ”algorithmic”. Materiaalia on haettu, kunnes tietty saturaatio on saavutettu, toisin sanoen samat artikkelit ovat alkaneet ilmestyä yhä uudestaan eikä relevanttia uutta

aineistoa ole ilmestynyt. Tutkimusaineiston mukaanottokriteereitä oli kolme: 1) *Aineisto on relevantti tutkimuskysymyksen kannalta* 2) *Aineistossa käsitelty eettinen kysymys kattaa jonkin algoritmin etiikan kysymyksen esim. läpinäkyvyys* 3) *Aineisto ajoittuu vuosille 2016–2018*. Viimeisen mukaanottokriteerin perustelu on se, että keskeisin aineisto kertyi näille vuosille, ja tekemällä rajauksen näihin vuosiin oli mahdollista saada ajankohtainen käsitys siitä, mitä algoritmien etiikasta ajatellaan.

Varsinainen tutkimusmateriaali koostuu tieteellisistä tietokannoista haetuista 9 tutkijan tai tutkijaryhmän artikkelista, kolmen tutkijan kirjoittamasta kahdesta *Datafied Society*-kirjan artikkelista, (Uricchio: *Data, Culture and the Ambivalence of Algorithms* sekä Paßmann & Boersma: *Unknowing Algorithms. On Transparency of Unopenable Black Boxes* 2017) ja kolmesta kirjasta (mediatutkija ja Kööpenhaminan yliopiston viestinnän ja informaatioteknologian apulaisprofessori Bucher: *If...Then. Algorithmic Power and Politics* 2018, Michiganin yliopiston Amerikan kulttuurin ja digitaalisten opintojen apulaisprofessori Cheney-Lippold: *We Are Data* 2017 sekä matemaatikko O’Neil: *Weapons of Math Destruction* 2016).

Taulukko1. Tietokannoista valitut 9 artikkelia.

Kirjoittaja	Vuosi	Artikkeli
Ananny, Mike	2016	Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness.
Crawford, Kate	2016	Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics.
Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S-M., Morison, J., Murphy, M. H., O’Brochain, N., Schafer, B. & Shankar, K.	2017	Algorithmic governance: Developing a research agenda through the power of collective intelligence.
Dourish, Paul	2016	Algorithms and their others: Algorithmic culture in context
Kitchin, Rob	2017	Thinking critically about and researching algorithms
Martin, Kirsten	2018	Ethical Implications and Accountability of Algorithms
Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L.	2016	The ethics of algorithms: Mapping the debate

Zarsky, Tal	2016	The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making
Ziewitz, Malte	2016	Governing Algorithms: Myth, Mess, and Methods

Tutkimusaineiston analyysissä pureudutaan tutkijoiden käsityksiin ja tutkimustuloksiin algoritmien etiikasta, etsitään käsityksistä merkityssisältöjä ja tulkitaan ne. Erilaisten käsitysten taustalla olevat filosofiset tai eettiset käsitykset muovaavat väistämättä niitä merkityssisältöjä, joita voidaan havaita. Näin muodostettua ensimmäistä kategoriaa luokitellaan ja työstetään edelleen, jotta aikaan saadaan toinen kategoria. Tutkimusaineiston kanssa tapahtuvan jatkuvan dialogin pohjalta saaduista tuloksista muodostetaan lopulta yhteinen tulosavaruus, joka tässä tapauksessa on synteesi.

2. ALGORITMIT

Seuraavassa selvitetään algoritmeista joitakin perusasioita menemättä kuitenkaan syvemmin ohjelmointiin. Tarkoitus on esitellä algoritmit yleisluonteisesti. Tämän jälkeen otetaan tarkasteltavaksi algoritmien yleiset eettiset ohjeet, jotka julkaistiin Yhdysvalloissa vuonna 2017. Ohjeistuksen tarkoituksena on luoda yhteiset toimintatavat algoritmien eettisten ongelmien vähentämiseksi.

2.1 Mitä ovat algoritmit

Algoritmi on saanut nimensä persialaiselta matemaatikolta Abu Abdallah Muhammad ibn Musa Al-Khwarizmilta, joka eli noin 1200 vuotta sitten (Rapaport 2016: 233). Jo antiikin babylonialaisten tiedetään käyttäneen algoritmeja (Rapaport 2016: 233). Halavan (2016: 1) mukaan algoritmien tutkimus on alkanut 1930-luvulla, jolloin kehitettiin ”laskettavuuden teoria, algoritmisesti ratkeamattomat probleemat” ja ”Turingin koneet”. Turingin kone on teoreettinen malli, jolla voidaan simuloida tietokonetta. 1950-luvulla tutkimus kehittyi algoritmien suunnittelun ja analysoinnin suhteen. Tällöin suunniteltiin monet tärkeät algoritmit. Tullessa 1970-luvulle kompleksisuusteoria oli päässyt tutkimuksen keskiöön. Muun muassa mahdottomat probleemat ja todennäköisyysalgoritmit ovat kompleksisuusteorian antia.

Algoritmi voidaan määritellä eri tavoin. Sen voidaan sanoa olevan matemaattinen konstruktio, toimintaohje tai resepti, jonka avulla saavutetaan tietty tavoite tiettyjen sääntöjen mukaan (Rapaport 2016: 234–236, 697; Hill 2016: 38, 47–49; Bucher 2018: 19). Algoritmi on siis menettelytapa, jolla jokin probleema ratkaistaan (Halava 2016: 4). Seuraavassa esimerkissä lasketaan kahden positiivisen kokonaisluvun tulo erilaisilla kertolaskualgoritmeilla. Viimeiseen kertolaskualgoritmiin on liitetty kuvaus.

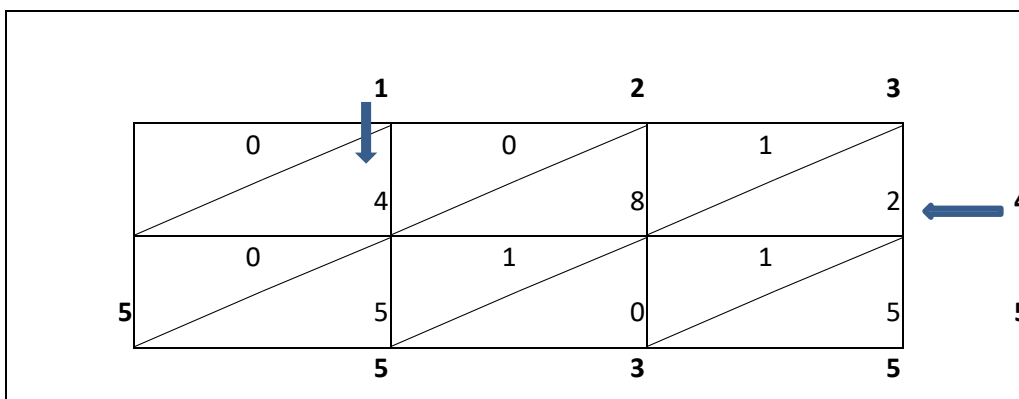
a) **klassinen kertolaskualgoritmi**

	1	2	3
		4	5
	6	1	5
4	9	2	
5	5	3	5

b) **arabialainen kertolaskualgoritmi:**

Sivujen pituudet = kerrottavien lukujen pituudet.

Kerrottavat luvut tulevat laatikoiden sivuille ja laatikkoon laitetaan tulot. Nuolet osoittavat lukusuunnan. Lukujen tulo tulee vasemmalle sivulle ja alasivulle, jonne lasketaan diagonaalien summat.

c) **venäläinen kertolaskualgoritmi:**

Toista kerrottavista luvuista puolitetään pyöristämällä alaspäin ja toista tuplataan, kunnes puolitettava luku on yksi. Laskemalla yhteen tuplatut luvut, joiden puolitettu luku on pariton, saadaan tulo.

45* 123	→	123
22 246		
11* 492	→	615
5* 984	→	1599
2 1968		
1* 3936	→	5535

venäläisen kertolaskualgoritmin kuvaus on seuraava:

SYÖTE: kokonaisluvut $a, b > 0$

TULOSTE: tulo ab

```

x ← 0
while a > 0 do
if “a on pariton” then x ← x+b
a ← [  $\frac{a}{2}$  ], b ← 2b
return x

```

Kuva 1. Esimerkki erilaisista matemaattisista kertolaskualgoritmeista Halavaa (2016: 3, 10–11) mukailten.

Sama algoritminen ongelma voidaan ratkaista monella tavalla. Oleellista on algoritmin vaatima tila (eli kuinka paljon algoritmi vie tietokoneen muistia) sekä aika (eli kuinka nopeasti algoritmi ratkaisee ongelman) (Laaksonen 2009). Erilaisiin tarpeisiin on olemassa erilaisia algoritmeja. Algoritmit ovat riippumattomia sekä ohjelmistokielestä että laitteistosta, mikä on niiden suuri etu. (Cormen, Leiserson, Rivest & Stein 2009.)

Koska algoritminen ongelma tulee voida ratkaista kaikenlaisilla syöteillä, ohessa listana yleisempiä toistorakenteita Halavan (2016: 11) mukaan.

```

for i = 1 to n do ...
repeat ... until
while ... do ...
if ... then (... else ...)
return
quit

```

2.2 Algoritmien eettiset ohjeet

Amerikkalainen Association for Computing Machinery US Public Policy Council loi vuonna 2017 algoritmien eettiset ohjeet. ACM:n (2017) mukaan algoritmien läpinäkyvyyden syyt ovat teknisiä, taloudellisia ja sosiaalisia. Eettisen ohjeistuksen tarkoituksena on luoda yhteiset pelisäännöt algoritmien läpinäkyvyyden ja vastuullisuuden lisäämiseksi. Eettiset periaatteet ovat lyhentäen ja hieman muunnellen seuraavat:

1. *Tietoisuus*: Tulee olla tietoinen suunnitteluun, toteutukseen ja käyttöön liittyvistä harhoista ja vahingoista.

2. *Käyttöoikeudet ja oikeussuojakeinot:* Viranomaisten on kannustettava sellaisten mekanismien käyttöönottoa, jotka mahdollistavat tehtyjen päätösten oikaisun.
3. *Vastuullisuus:* Instituutiot on pidettävä vastuullisina niiden käyttämien algoritmien tekemistä päätöksistä.
4. *Selitys:* Instituutioita, jotka käyttävät algoritmista päätöksentekoa, kannustetaan tuottamaan selvityksiä sekä algoritmin noudattamista menettelyistä että tehdyistä päätöksistä.
5. *Tietojen alkuperä:* Kehittäjien tulee dokumentoida kuinka aineisto kerätään sekä arvioida tiedonkeruuprosessin aiheuttamia vääristymiä. Tietojen julkinen tarkastelu tarjoaa mahdollisuuden korjata aineistosta johtuvia virheitä. Rajoitettu pääsy tietojen tarkasteluun voi olla perusteltua järjestelmän hyväksikäytön estämiseksi.
6. *Tarkastettavuus:* Kaikki aineisto on tallennettava siten, että ne voidaan tarkastaa, jos epäillään vahinkoa.
7. *Validointi ja testaus:* Instituutioiden tulee käyttää luotettavia menetelmiä mallien validoimiseksi, ja menetelmät ja tulokset tulee dokumentoida. Tämän lisäksi tulee suorittaa rutiininomaisesti testejä, jotta voidaan arvioida ja selvittää, aiheuttaako malli syrjivää haittaa. Testien tulosten tulee olla julkisia. (Association for Computing Machinery US Public Policy Council 2017.)

3. INFORMAATIOTEKNOLOGIAN FILOSOFIA

Informaatioteknologiaan liittyy useita filosofisia näkökulmia. Tästä syystä tässä tutkimuksessa tarkastellaan ensin informaatioteknologian filosofiaa ja sen osa-aluetta tietokone-etiikkaa, ja vasta sen jälkeen keskitytään taustoittamaan etiikkaa. Seuraavassa on ensin käsittely tärkeimpiä filosofisia suuntauksia sekä niiden tarkastelemia kysymyksiä. Sitten on katsaus tietokone-etiikan historiaan, joka pohjustaa varsinaista tietokone-etiikan käsittelyä. Tämän jälkeen kartoitetaan tietokone-etiikkaan liitettyjä normatiivisen etiikan näkökulmia. Lopuksi on lyhyt katsaus Floridin informaatioetiikkaan.

3.1 Yleisiä IT-filosofian näkökulmia

Rapaport (2016: 36) määrittelee filosofian ”henkilökohtaiseksi totuuden etsinnäksi millä tahansa kentällä järkevällä tavalla”. Filosofia tarkoittaa kirjaimellisesti ”rakastaa viisautta” (kr. *fileoo* rakastaa; *sophia* viisaus) (Perseus Digital Library 2018). Filosofian päähaaroiksi Rapaport (2016: 46–48) katsoo *metafyysiikan*, *epistemologian*, *logiikan*, *etiikan*, *estetiikan*, *filosofian historian*, *kielen filosofian*, *matematiikan filosofian*, *mielen filosofian*, *tieteen filosofian* ja *tuntemattoman filosofian haaran*, josta emme nyt tiedä, mutta joka varmasti vielä löytyy. Tietokone-etiikka on soveltavaa etiikkaa ja eräs filosofian alalaji.

Hassan, Mingers ja Stahl (2018: 265–269) pitävät informaatioteknologian filosofian tärkeimpinä filosofisina suuntauksina *metafyysiikkaa*, *epistemologiaa*, *rationalismia* ja *arvoteoriaa*. Heidän mukaansa esimerkiksi ontologia – oppi olevaisesta – ja mielen filosofia ovat metafyysiikkaan kuuluvia, samoin kuin tieteen filosofia ja logiikka rationalismiin ja estetiikka sekä etiikka arvoteoriaan kuuluvia. Informaatioteknologian kohdalla metafyyssisenä kysymyksenä voidaan tarkastella esimerkiksi teknologian sosiaalista ja teknologista suhdetta tai kysyä onko informaatiota olemassa ja mikä on sen luonne. Epistemologinen kysymys voi tutkia mikä on totuus tai miten hankimme tietämystä informaatioteknologiasta. Rationalistinen kysymys ottaa selvää mikä on hyvä teoria tai mikä on

järkevää. Arvoteoreettinen kysymys paneutuu esimerkiksi siihen, kuinka meidän tulee käyttäytyä. (Hassan ym. 2018: 265.)

Sana *metafysiikka* tarkoitti alun perin Aristoteleen kirjojen kokoelmaa, ja se tapahtui noin sata vuotta Aristoteleen kuoleman jälkeen (Sullivan ja van Inwagen 2014) (kr. *meta* jälkeen; *fysis* luonto) (Perseus Digital Library 2018). Metafysiikalla tarkoitettiin alussa eri asioita kuin myöhemmin. Alkuvaiheen pohdinnoissa esiintyivät kysymykset olemisesta sellaisenaan, asioiden perimmäinen syy ja ”se, mikä ei muutu” (Sullivan ja van Inwagen 2014). Ajan kuluessa metafysiikan kysymykset laajenivat mm. mielen ja ruumiin suhteen, aikaan, avaruuteen, Jumalan olemassaoloon, aineen ongelmaan, syy-seuraussuhteen ja kosmologiaan (Sullivan ja van Inwagen 2014). René Descartesia pidetään filosofina, joka on luonut pohjan luonnontieteiden kehittymiselle (Hassan, Mingers ja Stahl 2018: 265). Hän oli rationalisti, joka alkoi järjestelmällisesti kysyä mikä / mitä on – kysymyksiä, joita useat eri tieteenalojen pioneerit ovat sittemmin kysyneet (Hassan ym. 2018: 266).

Epistemologia tutkii filosofisin menetelmin tietoa ja sen ulottuvuuksia. Klassinen tiedon määritelmä on ”tieto on perusteltu tosi uskomus”. Epistemologia kysyy mitä tarkoitetaan tiedon oikeuttamisella ja ”mikä tekee uskonnuksesta perustellun”. Uskomuksen oikeuttaminen voi tapahtua kahdella tavalla: sisäisesti tai ulkoisesti. Sisäisesti oikeuttaminen tapahtuu mielen sisällä. Ulkoisesti uskomuksen oikeuttaminen tapahtuu luotettavien prosessien syntymisestä. Näin tapahtuu esimerkiksi eläimillä ja pikkulapsilla. Luotettavia uskomusten lähteitä ovat havainnot, introspektio, muisti, syy ja todistus. (Steup 2005.)

Rationalismin mukaan tietoa saadaan ennen kaikkea järjen avulla päättelöllä. Markie (2017) esittää kolme teesiä: *intuitio/päätely*, *luontainen tietämys* ja *luontainen mielle*. Intuitio/päätely -teesin mukaan voimme tunnistaa joitakin väitteitä intuition avulla, mutta parempiin tuloksiin päästään päättelöllä. Luontainen tietämys -teesin mukaan ennen kokemusta syntynyt tietämys on osa rationaalista luontoa. Luontainen mielle -teesi lähtee siitä, että kokemus laukaisee ymmärtämisprosessin.

Arvoteoria on moraalifilosofian alue, joka tutkii arvoja (Schröder 2016). Hassan, Mingers ja Stahl (2018: 269) käsittävät arvoteorian laajempänä käsitteenä, joka kattaa mm. etiikan ja estetiikan. Perinteisiä arvoteoreettisia kysymyksiä ovat hyvän ja hyvyyden ympärille rakentuneet kysymyksenasettelut: mitkä asiat ovat hyviä ja miten hyvyys näyttäytyy ja käsitetään (Schröder 2016). Arvotavoitteet voivat olla yksilöllisiä valintoja tai toisaalta objektiivisia tiloja, jolloin voidaan saavuttaa yksimielisyys arvoista. Arvoteoria tarkastelee luontaista arvoa ja välinearvoa. Esimerkiksi rahalla on välinearvo, koska sillä saadaan muita hyviä asioita kuten ruokaa kaappiin. (Schröder 2016.)

3.2 Tietokone-etiikan historia

Rakkaalla lapsella on monta nimeä. Näin on myös tietokone-etiikan laita. Kirjallisuudessa näkee puhuttavan dataetiikasta, informaatioteknologian etiikasta, tietokone-etiikasta ja tietojärjestelmien etiikasta. Kaikilla näillä yleensä tarkoitetaan tietokoneiden ja informaatioteknologian mukanaan tuomien moraalisten kysymysten etiikkaa. Kuitenkin Bynum (2015) näkee asian hieman toisin. Hänen mielestään tietokone-etiikasta puhuttaessa on viitattu normatiivisen etiikan suuntauksiin tietokoneiden yhteydessä, ammattietiikkaan ja internetin yhteydessä internetietiikkaan ja kybernetiikkaan. Myös ilmaisua informaatioetiikka tapaa, mutta tämä on niin ikään jalostuneempi muoto tietokone-etiikasta ja esiintyy Luciano Floridin tutkimuksissa (kts. 3.5 Floridin informaatioetiikka).

T.W. Bynum (2001: 109–110) valottaa tieteenalan synnyn alkuhämärää. Ensimmäiset askeleet otettiin 1940-luvulla, kun sotateollisuuden tutkija Norbert Wiener keksi uuden tieteenalan, jonka nimesi kybernetiikaksi (”*cybernetics*”). Hän yhdisti kybernetiikan idean tietokoneisiin ja näin ennusti tulevaa tietokone-etiikkaa. Seuraava merkittävä askel otettiin puolivälissä 1960-lukua, kun Donn Parker alkoi tutkia tietokoneiden epäeettistä ja laitonta käyttöä. 1960-luvun loppupuolella Joseph Weizenbaum loi tietokoneohjelman ELIZA, joka oli jäljitelmä psykoterapeutista. Ihmisten reaktiot järkyttivät Weizenbaumia, ja hän kirjoitti eettisistä huolistaan kirjan *Computer Power and Human Reason*. Termin *tietokone-etiikka* keksi yliopisto-opettaja Walter Maner, joka opettaessaan lääketieteen etiikkaa 1970-luvun puolivälissä huomasi, että tietokoneilla ja lääketieteen etiikalla oli

yhtymäkohtia. *Tietokone-etiikka* kuvasi niitä eettisiä ongelmia, joita tietokoneteknologia synnytti. Vuosikymmenen loppuun mennessä tietokone-etiikka oli terminä vakiinnuttanut paikkansa.

Bynumin (2015) mukaan vuosi 1985 oli tietokone-etiikalle keskeinen. Silloin ilmestyivät sekä Deborah Johnsonin ensimmäinen painos kirjasta *Computer Ethics* että *Metaphilosophy*-lehden numero, jossa oli James Mooren artikkeli. Vuonna 1995 Krystyna Górnjak-Kocikowska teki ennusteen, jonka mukaan tietokone-etiikka laajenisi globaaliksi etiikaksi. Ennustus tunnetaan nimellä ”Górniakin hypoteesi”.

Eettiset kysymykset ovat kulkeneet tietokoneiden rinnalla niin kauan kuin teknologiaa on ollut olemassa. Ensimmäiset keskustelut tekoälystä ja sen ihmisyyttä muuttavasta voimasta käytiin toisen maailman sodan jälkeen (Johnson 2001: viii). Pelättiin, että päätöksenteko siirtyy ihmisiltä tietokoneille (Johnson 2001: viii). Vielä vuonna 1979 James H. Moor oli sitä mieltä, että tekoäly ei tule saavuttamaan ihmisen älykkyyttä:

...there is enough evidence from artificial intelligence research to be suspicious of dogmatic claims that computers will never be able to accomplish certain feats of intelligence; yet, today there is not nearly enough evidence to support the conclusion that computers will someday match or exceed human intellectual ability in general or human decision-making ability in particular. (Moor, James H. 1979: 222.)

1970-luvun loppupuolella käytiin ensimmäisiä keskusteluja yksityisyyden suojasta. Yhdysvalloissa alettiin rakentaa suuria tietokantoja, joihin alettiin kerätä ihmisten henkilökohtaisia tietoja. Siihen asti tietokoneita oli käytetty lähinnä suurta laskentatehoa vaativiin projekteihin kuten asekehitykseen ja avaruusohjelmaan. (Johnson 2001: ix.)

1980-lukua leimasi henkilökohtaisten mikrotietokoneiden tulo. Eettinen keskustelu pyöri ohjelmistojen ympärillä. 80-luvulla tutustuttiin myös internettiin ja saatiin ensimmäiset hakkerit. Videopeliteollisuus sai alkunsa. (Johnson 2001: ix.)

1990-luku oli internetin. Internetin käyttö alkoi monipuolistua. Internet nähtiin demokration välineenä. Graafiset selaimet ja hakukoneiden tulo lisäsivät käyttäjien määrää.

Eettisessä kentässä yksityisyyden suoja, omistusoikeudet ja demokratian kysymykset olivat ajankohtaisia (Johnson 2001: ix–x).

3.3 Yleistä tietokone-etiikasta

Tietokoneiden ja informaatioteknologian nopea kehitys ovat avanneet valtavia mahdollisuuksia niin yksilön kuin yhteiskunnan tasolla. Toimintaympäristön muutokset ovat laajoja ja syvälle käyviä. Mooren (1985: 266) mukaan eettiset kysymykset nousevat toimintatavan tyhjiöstä, koska emme tiedä miten teknologiaa tulisi käyttää. Näyttää siltä, että toimintatavan tyhjiö on usein myös käsitteellinen tyhjiö (Moor 1985: 266). Tietokoneetiikan keskiössä on kysymys, miten määrittellä mitä meidän pitää tehdä ja mitä toimintatapojemme tulisi olla (Moor 1985: 266; Johnson 2001: 6). Bynumin (2015) mukaan Mooren vaikutuksesta tietokone-etiikkaan tulivat tärkeät ihmisen ydinarvojen käsitteet kuten yksityisyys ja turvallisuus. Mooren (1985: 275) mukaan tulisi päättää, missä asioissa voidaan luottaa tietotekniikkaan ja missä asioissa ei.

Johnson lähtee tietokone-etiikassa siitä, että kaikki uusi teknologia tulisi arvioida eettisesti jokaisessa sen kehitysvaiheessa (2001: 7). Tulee tietää mikä arvioinnin kohteena oleva teknologia on, toisin sanoen millaisia ominaisuuksia sillä on, koska muutoin ei voida määrittää yksityiskohtia tai sääntöjä sen käyttämiseen (Johnson 2001: 9). Tietotekniikkaa käytetään yhteydessä muihin ihmisiin eli sen luontainen konteksti on sosiaalinen (Johnson 2001: 12). Sosiaalinen viitekehys muokkaa tietotekniikan kehitystä ja tämä tapahtuu läpi leikaten kaikkialla (Johnson 2001: 12). Etiikka liittyy aina ihmiseen, sillä ihminen on rationaalinen olento, joka kykenee ajattelemaan tekonsa syitä ja seurauksia (Johnson 2001: 17). Johnsonin (2001: 17–21) mukaan tietotekniikka tuo tähän uuden ulottuvuuden, sillä nyt etiikka välineellistyy: Tekoja tehdään välineen avulla. Samalla etäisyys toisesta ihmisestä kasvaa ja tekojen seuraukset voivat jäädä näkymättömiin. Toiminta voi näyttäytyä arvoneutraalina, kunnes se kytkeytyy moraalikysymykseen. Johnsonin (2001: 205) mukaan mikään ihmistekoinen teknologia ei kuitenkaan voi olla arvoneutraalia. Teknologialla pyritään saavuttamaan aina jotakin: myös tehokkuus ja käyttäjäystävällisyys ovat arvoja.

Johnson (2001: 128) toteaa, että ”yksityisyyden hinta on noussut erittäin korkeaksi”. Yksityisyyttä uhkaavat tiedonkeruu eri tietokantoihin, kerättävän tiedon laadun muuttuminen, tiedon jakelun ja vaihdon laajuus, virheellisen tiedon skaalautuvuus ja kerätyn tiedon pitkäaikaisvaikutukset (Johnson 2001: 117). Meistä halutaan tietoja, koska uskotaan, että mitä enemmän tietoja, sitä helpompaa on päätöksenteko ja palvelu parempaa (Johnson 2001: 118). Masonin (2017: 1) mukaan yksityisyyttä uhkaavat erityisesti tietoteknologian kasvu ja lisääntynyt tiedon arvo päätöksenteossa. Mitä enemmän menetämme tietojen hallintaa itseltämme jollekin muulle taholle, sitä vähemmän meillä on yksityisyyttä ja autonomiaa.

Yksityisyyden ohella tietokone-etiikassa askarruttaa näkymättömyys. Moor (1985: 272–274) pitää suurena eettisenä huolenaiheena tietokoneiden toimintojen näkymättömyyttä. Hän jakaa näkymättömyyden kolmeen aspektiin. Ensimmäinen on näkymättömien operaatioiden tarkoituksellinen epäeettinen käyttö, joka kohdistuu omaisuuteen, yksityisyyteen tai valvontaan. Toinen aspekti on tietokoneohjelmiin upotetut näkymättömät arvot. Tällöin ohjelma ei toimi oletetusti, vaan kyseessä voivat olla ominaisuudet tai virheet. Kolmas aspekti on monimutkainen laskenta. Tietokoneiden laskentateho kasvaa jatkuvasti ja supertietokoneissa saavutetaan niin monimutkainen ja valtava teho, ettei ihminen enää pysty monitoroimaan saati ymmärtämään sitä.

Tietokone-etiikka käsitetään joskus pelkäksi ammattietiikaksi. Tällöin etiikalla ymmärretään koostettuja eettisiä ammattiohjeita kuten ACM Code of Ethics and Professional Conduct (Association for Computing Machinery 2018) tai Software Engineering Code of Ethics (IEEE Computer Society 1999). Ammatilaisen eettiseen toimintaan vaikuttavat monet muutkin seikat kuten lainsäädäntö, sopimukset, hyvät käytännöt, yrityksen käytännöt, henkilökohtaiset arvot ja uskomukset sekä yhteisön arvot (Bynum 2004: 63–64).

3.4 Tietokone-etiikkaan liitettyjä eettisiä näkökulmia

Tietokone-etiikkaan yhdistetään yleensä normatiivisen etiikan suuntauksia. Normatiivisen etiikan tärkeimmät suuntauksukset ovat *seurausetiikka*, *velvollisuusetiikka*, ja *hyveetiikka* (kts. 4.3 Normatiivinen etiikka). *Utilitarismi* on seurausetiikan suuntaus, jonka tavoitteena on toiminta, joka tuottaa suurimman hyödyn mahdollisimman suurelle joukolle ihmisiä (Pietarinen 2015). Utilitarismi siis tarkastelee toiminnan seurausta hyöty vs. haitta – akselilla. Käytännön esimerkki voisi olla vaikka kalliin hoidon saatavuus: Antaanko kyseistä hoitoa kaikille tarvitseville potilaille hoitoontulojärjestyksessä vai ovatko ensisijalla yhteiskunnallisesti tuottavat yksilöt?

Velvollisuusetiikka keskittyy toimintaan eikä ota huomioon tekojen seurauksia. Velvollisuusetiikka tunnustaa jokaisen ihmisarvon ja puolustaa jokaisen henkilön oikeuksia. Ihmisen tekoja ei kuitenkaan voida tarkastella erillisinä ilman teoista tulevia seurauksia, joten mikä viime kädessä määrää sen, mikä on moraalisesti oikein? Jos moraalilaki sanoo ”puhu aina totta” ja henkilö voisi valehtelemalla pelastaa useiden ihmisten hengen, niin velvollisuusetiikan mukaan henkilön tulisi puhua totta ja antaa niiden ihmisten kuolla, jotka olisi voinut valehtelemalla pelastaa. (Pietarinen 2015.)

Hyve-etiikka esittää kysymyksen moraalisesta luonteesta ja pyrkii saavuttamaan korkeimman hyvän, *eudaimonian*. Hyveet ovat ominaisuuksia, jotka rakentavat luonnetta ja auttavat saavuttamaan hyvän elämän. Esimerkkejä hyveistä ovat oikeudenmukaisuus, rohkeus, rehellisyys. Hyve-eettinen kysymys on esimerkiksi: Millaisia meidän tulisi IT-ammattilaisina olla, jotta toimisimme aina moraalisesti oikein erilaisissa tilanteissa? (Johnson 2001: 51; Pietarinen 2015.)

3.5 Floridin informaatioetiikka

Floridi (1999: 37–42) kritisoi sitä, että tietokone-etiikassa ei ole kunnollista filosofista perustaa eikä se ole kiinnostunut tekojen moraalisisista arvoista, vaan ensisijaisesti olemisesta. Floridin mukaan monelle tietokone-etiikka on lähes ammattietiikan synonyymi. Vastauksena tähän Floridi on luonut *informaatioetiikan*, jonka eettinen kysymys on ”mikä on hyvää informaatioyksikölle ja infosfäärille yleisesti”.

Floridi (2010: 85–87) esittelee yksinkertaistetun RPT-mallin, jossa informaatio näyttäytyy infosfäärissä sisällä olevana resurssina (*resource*), tuotteena (*product*) tai päämääränä (*target*). Resurssinäkökulmasta tarkasteltuna tietoa tarvitaan paljon eettiseen päätöksentekoon. Moraalialiehteita ovat mm. digitaalinen jakaminen, tietoähky, luotettavuus, todennukaisuus, saavutettavuus ja täsmällisyys. Tuotenäkökulmasta ihminen ei ole pelkästään kuluttaja vaan myös tiedon tuottaja. Keskiössä olevia aiheita ovat mm. vastuullisuus, luotettavuus, tahaton väärä tieto sekä viestinnän käytännölliset säännöt. Informaatio päämääränä, missä yksilön moraaliset arvioinnit ja toiminnot vaikuttavat informaatioympäristöön, moraalialiehteinä nousevat mm. yksityisyys, luottamuksellisuus, turvallisuus, hakkerointi, vandalismi, avoin lähdekoodi, ilmaisunvapaus ja sisällön suodattaminen ja kontrollointi.

Koska jokainen elollinen ja eloton *olemassa oleva tai ollut tai tulevaisuudessa oleva* (Floridi 2010: 90) kokonaisuus tulkitaan informaalisena objektina, jolla on vähintään minimaalinen moraalinen arvo, informaatioetiikka voi mennä perinteisten eettisten teorioiden ulkopuolelle, ja jokainen informaallinen objekti voidaan tulkita tekijäksi, jolla on edelleen vaikutusta muihin tekijöihin. Itsenäisissä objekteissa on tarkoituksenmukaiset tietorakenteet, jotka muodostavat objektin tilan, identiteetin ja ominaisuudet. Tämä muodostaa kokonaisen universumin, *infosfäärin*. (Bynum 2015.)

Infosfääri on jakaantunut neljään luokkaan, joihin kaikki tieto-ominaisuudet on jaettu. Luokat ovat modaalinen, humanistinen, illuministinen ja konstruktionistinen. Jokainen informaallinen objekti voi vahingoittua ja tätä vahingoittumista Floridi kuvaa *entropiaksi*. Informaatioetiikka sisältää neljä moraalisääntöä, jotka pyrkivät rajoittamaan tai estämään entropian tapahtumista sekä lisäämään, laajentamaan ja rikastamaan informaation määrää, laatua ja monimuotoisuutta. Metaeettisesti informaatioetiikka on liikkunut *antroposentrisestä* (ihmiskeskeisestä) näkökulmasta *biosentriseen* ja edelleen *ontosentriseen* (olemassaolokeskeiseen). Floridin mukaan objektorientoitunut ontosentrinen näkökulma sopii paremmin tietoyhteiskuntaan ja tarjoaa mm. robottikäyttäytymiseen oivallisen eettisen teorian. (Floridi 1999: 46–56.)

4. ETIIKKA

Etiikka voidaan jakaa analyttiseen, normatiiviseen, soveltavaan ja deskriptiiviseen etiikkaan. Suurelle yleisölle normatiivinen etiikka ja sen kysymykset (kuten kuinka meidän tulisi elää ja millainen on hyvä ihminen) lienevät tutuimpia soveltavan etiikan (esim. lääketieteen etiikka) ohella. Analyttinen ja normatiivinen etiikka ovat teoreettista etiikkaa. Deskriptiivistä etiikkaa esiintyy lähinnä empiirisissä tieteissä ja sen pyrkimyksenä on kuvata neutraalisti ihmisten arvovalintoja.

4.1 Etiikan käsitteestä

Arkipäiväisessä kielenkäytössä etiikka ja moraalit esiintyvät usein toistensa synonyymeina. Sanan *moraali* katsotaan yleensä juontuvan latinan sanasta *mos* (mon. *mores*), joka tarkoittaa tapaa, tottumusta, käytäntöä (Perseus Digital Library: 2018). Kuitenkin myös koineekreikka tuntee vastaavaa tarkoittavan sanan *mor* (mon. *mores*) (Mäkinen 2016: 11; Perseus Digital Library: 2018). Sana *etiikka* tulee koineekreikan sanasta *ethos*, jolla on tarkoitettu paitsi vakiintuneita tapoja ja luonnetta (Perseus Digital Library: 2018), myös omien tekojen seurausten pohtimista ja pohdinnan päätöstä, onko teko eettisesti oikein vai väärin (Mäkinen 2016: 11).

Kielitoimiston sanakirjan (2018) mukaan *etiikka* tarkoittaa ”hyvää ja paha sekä ihmisen moraalista toimintaa tutkiva tieteenhaara, siveysoppi; eettiset normit, moraalisäännöt, moraalit”. Sana *moraali* tarkoittaa ”1. yhteisössä vallitsevat eettiset käsitykset, arvostukset ja käyttäytymissäännöt, etiikka. 2. henkinen ryhti, henkinen kestävyys 3. eettinen sisältö, opetus, ohje.” (Kielitoimiston sanakirja 2018).

Tässä yhteydessä etiikalla tarkoitetaan tieteenhaaraa, jonka tutkimuskohteena on moraalit. Filosofiasa moraalit on ”empiirinen ilmiö, josta voidaan tehdä havaintoja” (Pietarinen ja Poutanen 1998: 12). Arkipäivän moraalikeskustelu on niin sanottua ”ensimmäisen tason moraalidiskurssia” (Pietarinen ja Poutanen 1998: 12). Etiikan tutkimus on ”toisen tason moraalidiskurssia”, mikä tarkoittaa ensimmäisen tason moraalidiskurssin tutkimusta

(Pietarinen ja Poutanen 1998: 13). Teoreettinen etiikka voidaan jakaa analyyttiseen ja normatiiviseen etiikkaan.

4.2 Analyyttinen etiikka

Analyyttinen etiikka tai *metaetiikka* tutkii yleisesti moraalin luonnetta ja käsityksiä. Olemukseltaan analyyttinen etiikka on käsiteanalyysia (Pietarinen ja Poutanen 1998: 13). Se ei ota kantaa moraalikeskusteluun eikä siltä saada neuvoja tai ohjeita. Fieser (2018) painottaa: ”Kuitenkin kaksi asiaa ovat merkittäviä: (1) metafysiset kysymykset, jotka koskevat sitä, onko moraalit riippumaton ihmisistä ja (2) psykologiset kysymykset, jotka liittyvät moraalisten tuomioiden ja käyttäytymisen taustalla olevaan henkiseen perustaan.” Analyyttisen etiikan tärkeimmät suuntaukset ovat *intuitionismi*, *emotivismi*, *preskriptivismi*, *eettinen relativismi*, *eettinen realismi* ja *eettinen konsensualismi* (Pietarinen ja Poutanen 1998).

Tunnettu *intuitionismin* edustaja on G.E. Moore ja hänen teoksensa *Principia Ethica*. Hänen kiinnostuksen kohteenaan oli moraalinen hyvä itsessään. Hän pohti, milloin joku ihminen tai teko on hyvä ja mistä se voidaan tietää, ja tuli siihen tulokseen, ettei teon hyvydestä voida tehdä havaintoja. Mooren mukaan hyvyys voidaan oppia ”tuntemaan moraalisen intuition avulla”. (Pietarinen ja Poutanen 1998: 15, 30.)

1920-luvulla Euroopassa syntyi filosofinen suuntaus nimeltä *looginen positivismi*. Suuntaukseen liittyi kiinteästi ajatus kielen kuvateoriasta. Toteamuksella oli sisältö vain silloin, kun se luonnehti vallitsevaa asiantilaa. Teoria joutui ongelmiin pohtiessaan moraalilmaisujen luonnetta, sillä ne eivät olleet totuuksia eikä niitä pystytty todistamaan. Moraalilmaisut eivät kuvanneet mitään. Tästä kehittyi edelleen *emotivismi*, joka keskittyi pohtimaan moraalilmausten merkityssisältöjä. (Pietarinen ja Poutanen 1998: 49–54.)

Preskriptivismin mukaan ”moraali ohjaa ihmisen toimintaa” (Pietarinen 2015). Tunnetuin suuntauksen edustaja on Richard M. Hare. Hänen mukaansa ”moraaliväitteen preskriptiivisyys... tarkoittaa sitä, että väitteen esittäjä joutuu itse toimimaan väitteen edellyttämällä

tavalla” (Pietarinen ja Poutanen 1998: 75). Moraalissa nähdään kaksi tasoa, intuitiivinen ja kriittinen taso. Intuitiivisella tasolla ratkaisut syntyvät liikoja pohtimatta kasvatuksen ohjaamina (Pietarinen ja Poutanen 1998: 77). Kriittisellä tasolla ratkaisuja pohditaan ja testataan (Pietarinen ja Poutanen 1998: 79).

Pietarisen ja Poutasen (1998: 100) mukaan *eettisessä relativismissa* ”ei ole ehdottomia moraalistandardeja”. Eettinen relativismi on kiinnostunut moraali-ilmaisujen pätevyyden perusteluista, ei väitteiden oikeellisuudesta. Millään yleispätevällä menetelmällä ei voida ratkaista moraalisesti oikeaa toimintaa. Eettisen relativismin teorioita ovat sosiaaliin tapoihin perustuva *konventionalismi* sekä perustaviin moraaliarvostelmiin keskittyvä *käsitteellinen relativismi* (Pietarinen ja Poutanen 1998: 106, 123).

Kielellisen merkityksen teoria, eli niin sanottu *totuusehtoinen semantiikka*, on *eettisen realismin* keskiössä (Pietarinen ja Poutanen 1998: 129). Eettisen realismin mukaan tietoa moraalista tosiasioista ja ominaisuuksista saadaan intuition ja havaintojen avulla (Pietarinen ja Poutanen 1998: 146–147). ”Moraaliset tosiasiat ovat moraaliväitteen totuusehdoista” (tosi/epätosi) (Pietarinen ja Poutanen 1998: 134–135). Moraaliset ominaisuudet käyvät ilmi henkilön käytöksestä (Pietarinen 2015). Kuitenkaan eettinen realismi ei selvitä moraalisten ominaisuuksien luonnetta tai alkuperää (Fisher 2011: 5). Eettisen realismin teorioita ovat *intuitionistinen* ja *naturalistinen realismi* (Pietarinen ja Poutanen 1998: 138).

Eettinen konsensualismi uskoo, että moraaliväitteiden pätevyyden kriteeri on yksimielisyys (Pietarinen ja Poutanen 1998: 153). Tunnetun konsensusteoreetikon Jürgen Habermasin mukaan viestintä on erilaisia puheakteja. Jotta ihmiset ymmärtäisivät toistensa puheakteja, tarvitaan kielellistä toimintaa. Keskustelu voi olla teoreettista tai käytännöllistä. Tosiasiat kuuluvat teoreettiseen diskurssiin. Moraalikysymykset ovat käytännöllisen puheaktin aluetta. Habermasin diskurssin malli on vapaa kaikista esteistä ja häiriöistä, ja keskustelun osapuolten tavoitteena on saavuttaa yksimielisyys moraaliväitteestä tai normista. (Pietarinen 2015.)

4.3 Normatiivinen etiikka

Normatiivinen etiikka pohtii millaista on hyvä elämä tai millainen ihmisen tulisi olla (Oksanen, Launis ja Sajama 2010: 8). *Kultainen sääntö* on klassinen esimerkki normatiivisesta etiikasta: Kohtele toista ihmistä niin kuin haluaisit itseäsi kohdeltavan (Fieser 2018). Normatiivisen etiikan tunnettuja suuntauksia ovat *velvollisuusetiikka*, *seurausetiikka* ja *hyve-etiikka* eli *virtuismi*.

Velvollisuusetiikassa toiminnan arvon määrää toiminnan sisäinen luonne (Pietarinen 2015). Velvollisuusetiikka keskittyy itse toimintaan eikä ota huomioon tekojen seurauksia. On olemassa neljä keskeistä velvollisuusetiikan suuntausta. Saksalainen filosofi Samuel Pufendorf jakoi velvollisuudet kolmeen luokkaan: velvollisuudet Jumalaa, itseä ja muita ihmisiä kohtaan (Fieser 2018). Toinen velvollisuusetiikan suuntaus on *oikeusteoria*, jota kutsutaan myös oikeuksien ja velvollisuuksien korrelaatioksi (Fieser 2018). Kolmas velvollisuusetiikan suuntaus on Immanuel Kantin kategorinen imperatiivi. Hänen mukaansa ihmisellä on autonominen tahto, joka luo ohjeita moraalialue varten (Pietarinen 2015). ”Moraalilaki esittää ehdottoman vaatimuksen”, jonka mukaan on toimittava (Pietarinen 2015). Neljäs velvollisuusetiikan teoria on W.D. Rossin yleiset *prima facie* -velvollisuudet, joita ovat luotettavuus, hyvitys, kiitollisuus, oikeudenmukaisuus, hyväntekeväisyys, itsensä parantaminen ja vahingoittamattomuus (Fieser 2018).

Seurausetiikka keskittyy ”hyötyjen ja haittojen määrittelemistä ja mittaamista koskeviin kysymyksiin” (Pietarinen 2015). Seurausetiikka voidaan määritellä siten, että ”toiminta on moraalisesti oikein, jos tämän toiminnan seurauksista on enemmän hyötyä kuin haittaa” (Fieser 2018). Tunnettu seurausetiikan suuntaus on *utilitarismi*. Utilitarismi pyrkii hyödyn maksimointiin. Periaate on, että ”jokaisen tulisi toimia niin, että saavutettaisiin mahdollisimman suuri hyöty mahdollisimman monelle” (Johnson 2001: 36). Hyöty voi olla yleistä hyvää, hyvinvointia, mielihyvää tai arvoja (Pietarinen 2015). Fieserin (2018) mukaan Jeremy Bentham on kehittänyt kokonaisen utilitarismin järjestelmän, josta erityisesti ”kaksi suuntausta ansaitsee tulla mainituksi”: *tekoutilitarismi* ja *hedonistinen utilitarismi*. Tekoutilitarismi tarkoittaa, että jokaisessa tilanteessa on erikseen arvioitava,

toimitaanko moraalisesti oikein vai väärin ja miten saavutetaan moraalisesti paras tulos. Hedonistinen utilitarismi taas tarkastelee tekoja nautinnon ja kivun perustan kautta.

Hyve-etiikka tarkastelee ”millaisia luonteenominaisuuksia ja kykyjä moraalisella toimijalla tulee olla, jotta hän osaisi menetellä eettisesti hyvin tai oikein” (Pietarinen ja Poutanen 1998: 164). Luonnetta kehitetään hyveitä harjoittamalla. Aristoteleen mukaan moraaliteoria kertoo ihmiselle, mikä hänen päämääränsä eli *teloksensa* on (Pietarinen ja Poutanen 1998: 164). Korkein hyvä, *eudaimonia*, voidaan saavuttaa hyveitä harjoittamalla ja suotuisilla ulkoisilla tekijöillä (Pietarinen 2015).

4.4 Tutkimusetiikka soveltavan etiikan esimerkkinä

Soveltava etiikka on etiikan haara, joka käsittelee erityisiä kysymyksiä. Käsiteltävän asian tulee täyttää kaksi ehtoa. Sen on oltava ristiriitainen, toisin sanoen mielipiteitä on oltava puolesta ja vastaan. Lisäksi asian on oltava selkeästi moraalinen asia. (Fieser 2018.)

Etiikan teorioita on sovellettu eri elämänaloilla (Pietarinen 2015). Näitä ovat esimerkiksi lääkintäetiikka, ympäristöetiikka, tutkimusetiikka ja tietokone-etiikka. Koska tietokone-etiikka muodostaa teoreettisen tutkimuskehityksen selkärangan ja on käsitelty omana kokonaisuutenaan, tässä katsaus tutkimusetiikan alaan.

Mäkisen (2006: 10) mukaan tutkimusetiikan ”määrittely riippuu siitä, mihin toiminnan osa-alueeseen keskitytään”. Vaihtoehtoja ovat toimijat (tutkijat, tutkittavat), yhteisöt (yliopisto, tiedeyhteisö), toimet (esim. raportointi) tai taustayhteisöt kuten liike-elämä ja viranomaiset. Tutkimusetiikassa käsitellään tutkimuksen päämäärään liittyvää moraalialaa ja keinoja tuon päämäärän saavuttamiseksi sekä ”pohditaan miten tämä moraaliala joko voidaan ylläpitää tai miten se pitäisi ylläpitää” (Mäkinen 2006: 10). Mäkinen (2006: 15) toteaa, että tutkimusetiikkaa käsitellään joko tutkijoiden ammattietiikkana tai moraaliteorianä. Lääketiede on stimuloinut tutkimusetiikan kehittymistä. Tutkimusetiikan

näkökulmasta tärkein dokumentti on Nürnbergin koodi vuodelta 1947, jossa on kansainvälisesti määritelty tutkimusetiikan periaatteet (Mäkinen 2006: 18).

Itä-Suomen avoin yliopisto (2018) esittää sivuillaan tieteelle seuraavia tuntomerkkejä: perusteltavuus, julkisuus ja intersubjektiivisuus, kriittisyys, itsensä korjaavuus, autonomia sekä edistyvyys. Mäkinen (2006: 29–33) avaa keskeisiä tieteen ominaisuuksia seuraavasti. Objektiivisuus merkitsee tutkijan neutraalisuutta ja epäpersoonallista toimintaa. Kriittisyys on tutkijan asenne, kaikkea tulee epäillä, kunnes toisin todistetaan. Autonomia on perinteisesti liitetty sekä tieteeseen että tiedeyhteisöihin. Ulkopuoliset tahot uhkaavat tieteen autonomiaa ja tutkimusta. Ne voivat vaikuttaa siihen, mitä ja miten tutkitaan. Edistyvyys tarkoittaa, että tiede korjaa omia aikaisempia virheitään ja jalostuu paremmaksi. Tieteen tulee olla myös julkista ja avointa kaikille sekä kirjoitettu ymmärrettävällä kielellä (Itä-Suomen avoin yliopisto 2018).

Tutkimusetiikassa risteävät sekä yhteiskunnan että tieteen intressit. Mäkinen (2006: 153) toteaa: ”Tiede ja yhteiskunta käyvät...tasapainoleikkiä: tieteen autonomisuus pitää turvata, mutta tieteen tulokset vaikuttavat yhteiskunnan ja tavallisten ihmisten elämään, ja poliitikot edustavat heitä”.

4.5 Deskriptiivinen etiikka

Deskriptiivinen eli kuvaileva etiikka pyrkii selittämään ja kuvaamaan moraalin ilmiöitä. Deskriptiivinen etiikka ei tee moraalisuosituksia, vaan havainnoi puolueettomasti. Se hyväksyy jopa toisilleen vastakkaiset näkemykset. Empiiriset tieteet, kuten psykologia ja yhteiskuntatieteet, suosivat deskriptiivistä etiikkaa. (Oksanen, Launis ja Sajama 2010: 7–11).

Esimerkkinä deskriptiivisen etiikan edustajasta voidaan mainita sosiologi Émile Durkheim, jonka kiinnostuksen kohteena olivat yhteiskunnalliset ilmiöt. Hän pyrki tutkimaan niitä selittävästi. Moraali oli Durkheimille ennen kaikkea kokoelma sääntöjä, joita voitiin luonnehtia ja ryhmitellä. Deskriptiiviseen etiikkaan kuuluvana pidetään myös

evolutiivista etiikkaa ja sen edustajaa Charles Darwinia. Hänen mukaansa ihmisen erottaa eläimestä moraalista tai omatunto. (Oksanen, Launis ja Sajama 2010: 10.)

5. FENOMENOGRAFINEN TUTKIMUS

Fenomenografinen tutkimus on laadullista tutkimusta. Kettunen ja Tynjälä (2018: 1) toteavat, että laadullisella tutkimuksella on mahdollista saavuttaa laajempi horisontti tutkitavaan asiaan:

Methodological literature suggests that qualitative research may also reveal broader perspectives on the phenomenon under investigation.

Laadullinen tutkimus pyrkii ymmärtämään elämän ilmiöitä monipuolisesti tarkastellen. Silloin, kun halutaan saada ilmiöstä hyvä kuvaus tai tavoittaa syvällistä näkemystä, laadullinen tutkimus on paikallaan. Kun ilmiö on ymmärretty, voidaan tämän perusteella kehittää teorioita ja oletuksia ilmiön toiminnasta. (Kananen 2015: 70–71.)

5.1 Fenomenografia tutkimusmenetelmänä

Fenomenografialla sanana tarkoitetaan yleensä ilmiön kuvaamista (kr. *fainomenon* ilmiö, verbi *fainoo*, inf. *fainesthai* loistaa, ilmestyä, tulla näkyiin; verbi *grafoo* raapustaa, kirjoittaa, kaivertaa, piirtää) (Perseus Digital Library 2018). Fenomenografia on empiirinen laadullinen tutkimusmenetelmä, ja se tutkii niitä käsityksiä, joita tutkimuskohteella on tutkittavasta ilmiöstä (Ahonen 1995: 116; Huusko ja Paloniemi 2006: 163). Fenomenografian mukaan rationaaliselle ihmiselle ”maailma on sekä todellinen että koettu” (Huusko ja Paloniemi 2006: 164). Ahosen (1995: 116–117) mukaan fenomenografialle tärkeitä termejä ovat ilmiö, käsite ja käsitys. Ilmiö ja käsitys ovat saman kolikon kaksi puolta. Ilmiöllä tarkoitetaan ihmisen kokemuksia, joista muodostetaan käsityksiä. Käsitys on kuva ilmiöstä, joka on rakennettu kokemuksista. Käsite on dynaaminen konstruktio, jossa uusi tieto jäsentyy vanhan päälle. Todellisuutta jäsennetään yksilön oman tulkinnan kautta (Häkkinen 1996: 24).

Fenomenografiassa pyritään tekemään käsityksiä ymmärrettäviksi merkitysten tulkinnan ja luokitellun kautta. Luokittelu perustuu ennalta omaksuttuun teoriaan. Oleellista on

löytää laadullisesti erilaisia merkityksiä aineistosta sekä luoda kuvaavia ja relevantteja kategorioita (Ahonen 1995: 127). Varsinainen teorianmuodostus tapahtuu vuorovaikutuksessa tutkimusaineiston pohjalta (Ahonen: 1995: 123).

Fenomenografinen tutkimus syntyi 1970-luvulla Göteborgissa, kun Ference Marton tutki opiskelijoiden käsityksiä oppimisesta. Kuitenkin vasta 1980-luvun alussa Marton piti tutkimustaan fenomenografiana. Ihmistieteet, ennen kaikkea kasvatustiede ja oppimisen tutkiminen, ovat suosineet fenomenografista tutkimusta. Viime vuosina on jonkin verran muilla tieteenaloilla käytetty tutkimusmetodina fenomenografiaa, erityisesti tutkittaessa tiedonmuodostusta. (Ahonen 1995: 115; Huusko ja Paloniemi 2006: 163.)

Tutkimustavassa on kaksi tiedon tasoa (Häkkinen 1996: 30). Ensimmäinen taso on niin sanottu ensimmäisen asteen näkökulma, joka tarkastelee asioita niin kuin ne ovat. Toisen asteen näkökulma on kiinnostunut tarkastelemaan sitä, miten edellinen käsitetään. Esimerkiksi jos ensimmäisen asteen näkökulman havainto on, että algoritmeihin liittyy eettisiä ongelmia, niin toisen asteen näkökulma kysyy, millaisena havaitut eettiset ongelmat käsitetään. Fenomenografinen analyysi tapahtuu toisen asteen näkökulmasta (Häkkinen 1996: 32, 39).

5.2 Fenomenografinen tutkimusprosessi

Tutkimusprosessin ensimmäisessä vaiheessa tutkija kartuttaa teoreettista tietämystään aiheesta, jota aikoo tutkia. Tämä on tarpeellista, jotta kykenee esittämään oikeita kysymyksiä ja tunnistamaan eri merkityksiä aineistosta. Teoreettisten lähtökohtien perusteella tehdään tutkimuksen kysymysten asettelu. Tämän jälkeen hankitaan aineisto. Fenomenografiassa tavallisin aineistonhankintamentelmä on haastattelu. Kuitenkin aineiston voi hankkia millä tahansa menetelmällä. (Ahonen 1995: 132–134.)

Tämän jälkeen on vuorossa aineiston analyysi. Analyysin vaiheiden lukumäärä vaihtelee hieman tutkimuksesta riippuen, mutta yleensä niitä on neljä. Nämä ovat merkityskokonaisuuksien etsiminen käsityksistä, ensimmäisten kategorioiden luominen ja vertailu,

toisten, ylempien kategorioiden muodostaminen ja viimeisenä vaiheena kokoavan kuvauskategoriajärjestelmän laatiminen (Huusko ja Paloniemi 2006: 167; Kettunen ja Tynjälä 2018: 5–6). Kun aineisto on tulkittu ja luokiteltu, tuloksena on kokoava teoria ja kuvausavaruus. Viimeisessä tutkimisprosessin vaiheessa raportoidaan tutkimus.

5.3 Aineiston hankinta ja analyysi

Fenomenografiassa tavallisin aineistonhankintamentelmä on haastattelu. Silti aineiston voi hankkia millä tahansa menetelmällä. Aineisto voi olla esimerkiksi havainnointia, piirroksia tai dokumentteja (Huusko ja Paloniemi 2006: 164). Fenomenografisen tutkimuksen aineiston koko on yleensä melko pieni, noin 10–30 haastateltavaa tai dokumenttia (Kettunen ja Tynjälä 2018: 3). Laatu on tärkeämpää kuin määrä. Tutkijan tulee tuntea aineisto hyvin, sillä hän on vuorovaikutuksessa sen kanssa kaikissa vaiheissa (Ahonen 1995: 125).

Aineiston analysointi alkaa perehtymisellä tutkimusmateriaaliin. Kun käsitykset ovat selvillä, aletaan etsiä merkityssisältöjä. Tässä vaiheessa selvitetään myös, millaisina algoritmien eettiset ongelmat käsitetään ja miten ne näyttäytyvät. Merkityssisällöt järjestellään ja tulkitaan. Merkityssuhteen tulkinnassa huomioidaan ympäröivä konteksti. Näistä muodostuvat ensimmäisen vaiheen kategoriat. Ahosen (2005: 143) mukaan:

Vain ilmaisusta, joka on ajatuksellinen kokonaisuus, voi perustellusti tulkita jonkin merkityksen. ...tutkija määrittelee lukemalla ilmaisut ja tarkkailemalla, miten laajalle niiden ajatusyhteydet tekstissä ulottuvat...Tulkintayksikköjä ei siis voi määrittellä ulkoisin perustein etukäteen.

Seuraavassa vaiheessa tulkintoja luokitellaan etsien samankaltaisuuksia ja poikkeavuuksia. Tämän jälkeen muodostetaan toiset kategoriat. Lopuksi rakennetaan kokoava tulosavaruus.

Tämän työn tutkimusmateriaali on kerätty syys-marraskuussa 2018. Joitakin täydentäviä lähteitä on haettu vielä joulukuussa 2018. Tiedonhaun alkuvaiheessa hyödynsin Google

Scholaria. Olen käyttänyt hyödykseni kirjallisuusluetteloita sekä tieteellisistä artikkeleista että kirjoista. Teoreettisen viitekehyksen aineisto koostuu tietokannoista haetuista tieteellisistä artikkeleista, joistakin internet-lähteistä ja e-kirjoista sekä painetuista kirjoista, jotka on lainattu kirjastoista. Tietokannoista on haettu tutkimusmateriaalia pääasiassa hakusanoilla ”ethics” ”algorithm” sekä ”algorithmic”. Materiaalia on haettu, kunnes tietty saturaatio on saavutettu, toisin sanoen samat artikkelit ovat alkaneet ilmestyä yhä uudestaan eikä relevanttia uutta aineistoa ole ilmestynyt.

Perehdyin aluksi etiikkaan, informaatioteknologian filosofiaan ja erityisesti sen osa-alueeseen tietokone-etiikkaan sekä algoritmeja käsittelevään tutkimuskirjallisuuteen. Tämän jälkeen aloin lukea tutkimusaineistoa läpi ensin saadakseni käsityksen aineistosta. Tämän vaiheen jälkeen orientoiduin tutkimusaineistooni uudestaan nyt tehden miellekarttoja ja etsien käsityksiä. Jatkoin edelleen työstämällä teemoja omaan miellekarttaansa, joka toimi yleisenä analyysikehikkona. Keräsin tähän kehikkoon käsityksiä ja tutkimustuloksia. Samaan aikaan tein miellekartat myös filosofisista ja eettisistä suuntauksista sekä vaikuttajista, jotka nousivat kustakin dokumentista sekä keräsin omaan miellekarttaansa esiin nousseita jatkotutkimusaiheita ja kysymyksiä. Varsinaisessa analyysin kirjoitusvaiheessa kävin jatkuvaa dialogia tutkimusmateriaalin kanssa.

Ensimmäinen kategoria sijoittuu *algoritmiseen ympäristöön*, jossa ontologia, epistemologia ja normatiivisen etiikan kysymykset käyvät dialogia algoritmisen kokonaisuuden kanssa. Käytin apuna tutkimusaineistosta kerättyä tutkimuskysymyslistaa, jonka tarkastelulinssi on ollut eettinen. Käytännössä tämä on tarkoittanut sitä, että on arvioitu jokaisen tutkimuskysymyksen filosofinen tai eettinen luokka, jota tämä kysymys on edustanut. Tarkastelun perusteella loin kategorisen algoritmisen ympäristön. Algoritmiseen ympäristöön kerättiin mahdollisimman kattavia käsityksiä ja tuloksia, joista työstettiin algoritmiset merkityssisällöt. *Algoritmit ja etiikka* on algoritmisen ympäristön ylempi kategoria. Merkityssisällöt työstettiin uudelleen yhdistellen ja muodostaen viisi uutta luokkaa. Tarkastelulinssi pysyi eettisenä. Luokkien nimet kuvaavat keskeisintä sisältöä. Seuraavassa vaiheessa molemmat kategoriat yhdistettiin samaan kokoavaan kuvaukseen, johon liitettiin kaikkien luokkien läpi taustalla olevat *yhdistävät tekijät*. Nämä ovat *algoritmisen*

päätöksenteko, vastuullisuuden vaatimus ja algoritmeja on tutkittava enemmän. Näin saatiin muodostettua lopullinen tulosavaruus.

5.4 Fenomenografisen tutkimuksen reliabiliteetti ja validiteetti

Ahosen (2005: 129) mukaan aineiston validiteetti toteutuu, kun materiaali on aitoa (täsmää tutkijan oletukseen samasta aiheesta) ja tarkoituksenmukaista eikä ylitulkintaa merkitysisältöjen tulkinnassa ole tapahtunut. Samoin luotujen kategorioiden tulee olla relevantteja. Aineiston ja kategorioiden validiteetti on tiivistetty seuraavaan taulukkoon.

Taulukko 2. Tutkimuksen luotettavuuskriteerit Ahosen mukaan (2005: 130).

	AINEISTO	KATEGORIAT
AITOUS	koskeeko aineisto tutkijan ja tutkittavien kannalta samaa asiaa	vastaavatko kategoriat tutkittavien tarkoittamia merkityksiä
RELEVANSSI	onko aineisto relevanttia tutkimuksen teorian kannalta	ovatko kategoriat relevantteja tutkimuksen teorian kannalta

Häkkinen (1996: 44–45) arvioi tutkimuksen luotettavuutta Larssonin (1993) viiden kriteerin mukaan. Nämä ovat *pragmaattinen kriteeri*, *diskurssikriteeri*, *heuristinen arvo*, *konsistenssi* ja *empiirinen ankkurointi*. Pragmaattinen kriteeri saavutetaan tutkimustulosten käytännön merkityksen kautta. Diskurssikriteerin mukaan ”tutkimus on laadullisesti hyvä, jos sen todetaan tuottaneen tuloksia, joista kukaan ei ole löytänyt heikkouksia” (Häkkinen 1996: 44). Heuristinen arvo kertoo kuinka hyvin vakuutetaan vastaanottaja näkemään jokin aspekti toisin. Konsistenssi tarkoittaa kategorioiden sisäisen logiikan pätevyyttä. Empiirinen ankkurointi tarkoittaa tutkimusprosessin kuvauksen avaamista ja aineistoesimerkkejä lukijalle.

6. ANALYYSI JA TULOKSET

Alkaessani työstää tutkimusmateriaalia, aloin tutkia siinä ilmeneviä käsityksiä. Tein jokaisesta dokumentista oman miellekartan, johon kokosin esiin tulevia asioita, niin käsityksiä kuin tutkimustuloksiakin. Jatkoin edelleen työstämällä teemoja omaan miellekarttaansa, joka toimi yleisenä analyysikehikkona. Keräsin tähän kehikkoon käsityksiä ja tutkimustuloksia. Samaan aikaan tein miellekartat myös filosofisista ja eettisistä suuntauksista sekä vaikuttajista, jotka nousivat kustakin dokumentista sekä keräsin omaan miellekarttaansa esiin nousseita jatkotutkimusaiheita ja kysymyksiä. Varsinaisessa analyysin kirjoitusvaiheessa kävin jatkuvaa dialogia tutkimusmateriaalin kanssa.

6.1 Algoritminen ympäristö

Ensimmäinen kategoria sijoittuu algoritmiseen ympäristöön, jossa ontologia, epistemologia ja normatiivisen etiikan kysymykset käyvät dialogia algoritmisen kokonaisuuden kanssa. Pohtiessani ensimmäistä kategoriaa, käytin apuna tutkimusaineistosta kerättyä tutkimuskysymyslistaa, jonka tarkastelulinssi on ollut eettinen. Käytännössä tämä on tarkoittanut sitä, että on arvioitu jokaisen tutkimuskysymyksen filosofinen tai eettinen luokka, jota tämä kysymys on edustanut. Tarkastelun perusteella loin kategorisen algoritmisen ympäristön. Tähän algoritmiseen ympäristöön on kerätty mahdollisimman kattavia käsityksiä ja tuloksia, joista on edelleen työstetty eettisen linssin läpi algoritmiset merkityssällöt. Luokkien nimet ilmentävät niitä kiteytyneitä kokonaisuuksia, joita tutkimusaineistosta on löydetty. Jokaisen kappaleen alussa on esitetty oleellinen käsitys kursivoina lauseena, jonka jälkeen muu kappale avaa käsitystä. Kappaleen viimeinen lause kertoo algoritmisen merkityssällön.

6.1.1 Perusominaisuudet

Ontologia on oppi olevaisesta eli se tarkastelee olemassaolon kysymyksiä. Algoritmisessa ympäristössä tutkimusmateriaalista nousevia ontologisia kysymyksiä ovat muun muassa

mikä on algoritmi sekä millaisia ominaisuuksia algoritmilla on. Näiden kysymysten avulla kartoitetaan algoritmin perusolemusta ja algoritmiin itseensä liittyviä ominaisuuksia.

Algoritmi on monimuotoinen. Algoritmi on sekä ohjelmointikieli- että alustariippumaton (Bucher 2018: 22). Algoritmi voidaan määritellä toimintaohjeeksi, reseptiksi tai menetelmäksi (Rapaport 2016: 234–236, 697; Hill 2016: 38, 47–49; Bucher 2018: 19), mutta sitä voidaan tarkastella useista näkökulmista. Tarkastelulinssi voi olla tekninen, matemaattinen, poliittinen, kulttuurinen, taloudellinen, kontekstuaalinen, filosofinen jne. (Kitchin 2017: 16; Bucher 2018: 19–40). Kun lisäksi erilaisia algoritmeja voidaan käyttää hyvin monenlaisiin tarkoituksiin (Cormen, Leiserson, Rivest & Stein 2009), voidaan algoritmin sanoa olevan monipuolinen eli merkityssisällöksi havaitaan *monipuolisuus*.

Algoritmit ovat sosioteknisiä eikä niitä voida irrottaa kontekstista, jossa niitä kehitetään ja otetaan käyttöön (Ananny 2016: 98; Crawford 2016: 81, 89; Dourish 2016: 2–3; Ziewitz 2016: 10; Kitchin 2017: 18, 20; Uricchio 2017: 127; Bucher 2018: 126, 152). Sosioteknisyys tarkoittaa, että algoritmit ovat vuorovaikutuksessa ihmisten ja tekniikan kanssa eikä näitä puolia voida erottaa vaan asiaa on tarkasteltava kokonaisuutena. Algoritmit ovat siis kontekstiriippuvaisia. Algoritmiset merkityssisällöt ovat *sosioteknisyys* ja *kontekstiriippuvuus*.

Algoritmit ovat arvoladattuja (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 1; Bucher 2018: 23; Martin 2018: 1). Arvolataus tarkoittaa, että algoritmi kytkeytyy moraliin eli sillä on kosketus ihmistoimintaan (Johnson 2001: 17–21). Mikään teknologia ei voi olla arvoneutraalia, sillä teknologialla pyritään aina johonkin, ja nämä asetetut tavoitteet ovat itsessään arvoja, kuten esimerkiksi tehokkuus (Johnson 2001: 205). Koska algoritmit ovat sekä sosioteknisiä että arvoladattuja, tästä väistämättä seuraa, että niillä on eettisiä vaikutuksia (Dourish 2016: 7; Mittelstadt 2016; O’Neil 2016; Bucher 2018: 120; Martin 2018: 1). Merkityssisältö on siis *eettinen vaikutuksellisuus*.

6.1.2 Toiminta

Algoritmisessa ympäristössä toiminnan osalta esiintyy ontologinen kysymys millä tavalla algoritmi on olemassa (Tedre 2011: 65). Algoritmi näyttäytyy aktiivisena toimijana. Toimintaan liittyvät filosofiset ja eettiset kysymykset nivoutuvat suurelta osin epistemologisiin ja normatiivisen etiikan kysymyksen asetteluihin. Miten ja kuinka algoritmit toimivat? Millaisia eettisiä ongelmia esiintyy?

Algoritminen voima perustuu pääsääntöisesti neljään toimintoon. Nämä ovat priorisointi, luokittelu, yhdistäminen ja suodatus (Ananny 2016: 97, 103; O’Neil 2016: 70; Kitchin 2017: 18; Uricchio 2017: 130; Bucher 2018: 34). Data on muokattava algoritmille luettavaksi ennen kuin sitä voi prosessoida (Bucher 2018: 5). Pelkkä luettavaksi saatettu data ei ole hyödyllistä, vaan sitä on käyttötarkoitusten mukaan priorisoitava, luokiteltava, yhdistettävä, suodatettava ja niin edelleen. Algoritmisina merkityssisältöinä voidaan havaita *tiedonkeruu* ja *tiedon hallinta*.

Algoritmit eivät aina käyttydy ennustettavasti (Crawford 2016: 80; Kitchin 2017: 19, 21, 25). Algoritmit ovat osa suurempaa kokonaisuutta. Algoritmit voivat toimia odottamattomasti, niillä voi olla sivuvaikutuksia tai toiminnalla voi olla arvaamattomia seurauksia (Kitchin 2017: 19). Merkityssisältönä voidaan tunnistaa *ennustamattomuus*.

Algoritmit tekevät puolueellisia päätöksiä (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 7, Kitchin 2016: 19; O’Neil 2016; Zarsky 2016: 123, 128; Martin 2018: 4). Mittelstadt ym. (2016: 7) mukaan puolueellisuus voi syntyä sosiaalisista asenteista, teknisistä rajoitteista, virheistä tai suunnittelupäätöksistä. Myös käyttökontekstissa olevat ulottuvuudet, kuten esimerkiksi järjestelmämuutos tai käyttäjä, saattavat aiheuttaa puolueellisuutta. Algoritminen merkityssisältö on *puolueellisuus*.

Algoritmit toimivat portinvartijoina (Uricchio 2017: 131; Bucher 2018: 7). Portinvartijana toiminen tarkoittaa, että algoritmi tekee valintoja ja päätöksiä ilman, että käyttäjä voi millään lailla osallistua kyseisten valintojen ja päätösten tekemiseen. Näin tapahtuu esimerkiksi silloin, kun algoritmi suodattaa ja valitsee käyttäjän nähtäväksi tietyt uutissyötteen tai päättää keiden käyttäjien viestit näkyvät. Prosessi on tavallisesti käyttäjälle täysin läpinäkymätön eli ei ole mitään mahdollisuutta saada selville millä perusteilla päätöksiä

tehdään (Mittelstadt ym. 2016: 3; Uricchio 2017: 131). Merkityssisällöksi voidaan tunnistaa *läpinäkymättömyys*.

Koneoppivat automaattiset tai puoliautomaattiset ”algoritmit eivät ainoastaan muutu tapahtumassa, vaan niillä on kyky muuttaa tapahtumaa” (Bucher 2018: 28). ”Algoritmit muuttuvat ja kehittyvät käyttöönnotossa” (Dourish 2016: 8). Bucherin (2018: 24) mukaan koneoppivia algoritmeja on kolmenlaisia: ohjattuja, ohjaamattomia ja puoliohjattuja. Ohjatussa algoritmista harjoitusdatasta tiedetään haluttu tulos. Ohjaamaton algoritmi ei sisällä dataa halutusta tuloksesta. Puoliohjattu algoritmi oppii vuorovaikutuksessa sisään syötetyn mallin ja ympäristönsä kanssa. Algoritmisen merkityssisältö on *algoritmisen päätöksenteko*.

Mustat laatikot ovat haasteellisia (Kitchin 2016: 20; Paßmann & Boersma 2017: 139–142; Bucher 2018: 41–65). Musta laatikko tarkoittaa prosessia tai objektia, jonka toimintaa ei tunneta (Bucher 2018: 43). Tämä saattaa aiheuttaa monia ongelmia. Ongelmat voivat olla niin teknisiä kuin sosiaalisiaakin. Mustien laatikoiden ongelma liittyy usein algoritmiseen päätöksentekoon (Dourish 2016: 7; Danaher ym. 2017: 3). Yritykset suojaavat liikesalaisuuksiaan luomalla läpinäkymättömiä prosesseja. Usein yritykset eivät tiedä itsekään miten algoritmi toimii prosessissa. Joskus tietojärjestelmät ovat niin monimutkaisia, että niihin muodostuu mustia laatikoita. Algoritmisen merkityssisältönä voidaan tunnistaa *tuntemattomuus*.

Kaikesta datasta, jota tuotamme internetissä, muodostuu henkilökohtainen digitaalinen identiteetti (Cheney-Lippold 2017: 26–31, 58–60), *jolla ei ole mitään tekemistä sen kanssa, keitä oikeasti olemme*. Digitaalinen identiteetti on eräänlainen palapeli. Sen yhtenä osana ovat ne tiedot, jotka vapaaehtoisesti annamme esimerkiksi sosiaalisessa mediassa ja muilla alustoilla. Yleinen nettikäyttäytymisemme muodostaa toisen palan. Se muodostuu siitä tiedosta, joka syntyy kun vieraillemme eri sivustoilla. Kolmas pala on metadata, data datasta. Esimerkiksi yhdestä puhelinsoitosta tiedetään soittajan ja vastaanottajan puhelinnumerot, kellonaika ja päivä, puhelun kesto sekä soittajan ja vastaanottajan maantieteellinen sijainti (Cheney-Lippold 2017: 188). Algoritmisen merkityssisältö on *digitaalinen identiteetti*.

*Kerätyn datan perusteella muodostetaan datamalleja (O’Neil 2016: 25–27; Cheney-Lippold 2017: 47, 51, 55–56, 121–123, 155–156) ja erilaisia käyttäytymiskuvioita (O’Neil 2016: 171–172; Cheney-Lippold 2017: 43–44), joita hyödynnetään monin eri tavoin mm. Googlen markkinoinnissa ja rikosten ennaltaehkäisyssä (Ananny 2016: 100–101; O’Neil 2016: 84–87; Cheney-Lippold 2017: 22; Martin 2018: 4). Hyvän esimerkin tästä antaa Edward Snowden Brian Williamsin haastattelussa 28.5.2014 (Williams 2014). Jääkiekko-ottelun tulosten haun perusteella Googlen hakukoneella voidaan päätellä henkilön tottumuksista ja käyttäjäprofiilista seuraavanlainen yleinen elämän malli: milloin henkilö käy kyseisillä sivuilla, onko henkilö kotona vai matkoilla, milloin hän herää ja menee nukkumaan, mitä muita datalaitteita on ympärillä ja keiden, onko henkilö mukana laittomassa toiminnassa, kenen seurassa henkilö on ja tekeekö henkilö jotain, mitä ei pitäisi tehdä. O’Neil (2016: 120–121) havainnollistaa mallia toisella esimerkillä. Kyseessä on yritys, joka profiloii työnhakijakandidaatteja asiakkailleen. Yritys väittää mm. pystyvänsä ennustamaan, milloin yrityksen tähtityöntekijä on aikeissa vaihtaa työpaikkaa. Yrityksen malli perustuu työntekijän sosiaalisen pääoman laadun määrittelyyn. Ohjelmoija, joka viettää paljon työajan jälkeistä aikaa ratkoen ohjelmointipulmia toisten ohjelmoijien kanssa, arvioidaan intohimon, sitoutumisen, taitojen ja sosiaalisen pääoman perusteella laadukkaammaksi työntekijäksi. Jos lisäksi ohjelmoijan kontaktit ovat erityisen merkityksellisiä, hänen sosiaalinen pisteytyksensä ”ampuu katosta läpi.” Algoritmiseksi merkityksellöiksi voidaan tunnistaa *datamallit* ja *datan tulkinta*.*

Yksityisyydestä tulossa harvojen etuoikeus (O’Neil 2016: 170; Cheney-Lippold 2017: 207–208). Facebookin perustaja Mark Zuckerberg kertoi The Guardian -lehdessä vuonna 2010, että ”yksityisyys ei ole enää sosiaalinen normi” (Johnson 2010). Edellisenä vuonna 2009 Googlen johtaja Eric Schmidt totesi CNBC:lle antamassa haastattelussaan:

If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place (Esguerra 2009).

Cheney-Lippold (2017: 208) toteaa, että somejättien johtajien mielestä ”yksityisyys on kuollut”, koska se häiritsee liiketoimintaa. Sittemmin erilaisten kohujen jälkeen Facebookin perustaja on joutunut hieman pyörtämään puheitaan ja parantamaan Facebookin

yksityisyyskäytäntöjä. Myös Google on joutunut parantamaan omia toimintatapojaan. Edellä kerrotun perusteella algoritminen merkityssisältö on *yksityisyys*.

6.1.3 Vaikutukset

Algoritmeilla on kiistatta eettisiä vaikutuksia. Nämä kysymyksenasettelut kuuluvat sekä epistemologian että normatiivisen etiikan alaan. Miten algoritmien vaikutuksista voidaan saada tietoa? Millaisia seurauksia algoritmien toiminnalla ja algoritmisella päätöksenteolla on? Vastuukysymykset ovat normatiivisen etiikan kysymyksiä. Kuka vastaa eettisistä vaikutuksista? Mitä pitäisi tehdä? Millainen algoritmin tulisi olla?

Algoritminen valta muokkaa todellisuutta (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 5; O’Neil 2016: 28–31; Kitchin 2017: 18; Bucher 2018: 73, 114, 153–156). Algoritmit muovaavat käsitystämme todellisuudesta. Esimerkiksi somekupla on tilanne, jossa algoritmi syöttää jatkuvasti saman aihepiirin sisältöä käyttäjälle. Koska vastakkaista informaatiota ei tule, sama sisältö jää ikään kuin kaikumaan (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 9; Uricchio 2017: 131) ja täyttää koko tilan. Tästä on seurauksena näkemyksen puolueellisuus. Bucher (2018: 53) toteaa Facebookin käyttäjän roolista ja algoritmin yhteydestä:

Furthermore, it helps to know how central a role users have in shaping algorithmic outcomes...Users matter because it is their data, their clicking behaviour, preferences, network relations and communicative actions that provide the data for algorithms to act on.

Edellä sanotun perusteella algoritminen merkityssisältö on *puolueellisuus*.

Automaattinen päätöksenteko luo eettisiä ongelmia ihmisten ja algoritmien välillä (Mittelstadt ym. 2016: 11; Zarsky 2016; O’Neil 2016: 153; Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar 2017: 3–4). Mittelstadtin ym. (2016: 11) mukaan inhimillisille päätöksentekijöille on haastavaa määritellä toiminnan eettiset periaatteet. Jos inhimillinen ja algoritminen päätöksentekijä toimivat yhdessä, on ehdottomasti kuvattava milloin ja kuinka inhimillinen väliintulo on välttämätöntä. Danaherin ym. (2017: 4) mukaan algoritmiseen

päätöksentekoon liittyy epätarkkuutta, tehottomuutta ja arvaamattomia seurauksia. Myös Zarsky (2016: 121) toteaa, että epätarkkuudet voivat haitata prosessia ja että analyysi voi tehdä virheitä ennustaessaan ihmisen käyttäytymistä. Läpinäkymättömyys pahentaa epäoikeudenmukaisuutta (Zarsky 2016: 125). Algoritmisenä merkityssisältönä voidaan havaita *ennustamattomuus*.

Algoritmien eettisistä vaikutuksista ei olla tietoisia (Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar 2017: 13; Kitchin 2017: 21). Sekä Danaher ym. (2017: 13) että Martin (2018: 2) esittävät, että yrityksillä ja järjestelmäsuunnittelijoilla ei ole riittävää ymmärrystä algoritmien eettisistä vaikutuksista. Kuvitelma neutraalista algoritmista elää sitkeästi. Toisin sanoen uskotaan, että algoritmillä ei ole eettisiä vaikutuksia, johon pitäisi reagoida. Danaher ym. (2017: 13) kannattavat aiheesta lisätutkimusta kaikille sidosryhmille. Algoritmisen merkityssisältö on *tietämättömyys*.

Yritykset ovat vastuussa kehittämistään algoritmeista, sillä suunnittelupäätöksillä on suuri merkitys (Bucher 2018: 154; Martin 2018: 10). Martinin (2018: 2, 9–10) mukaan yritykset välttelevät vastuutaan eivätkä myönnä, että tekevät moraalisia valintoja kehittäessään algoritmeja ja myydessään niitä eteenpäin. Martin (2018: 10) huomauttaa, että mikäli yritys ei halua kantaa omaa vastuutaan, sen ei pitäisi olla liiketoiminnassa mukana ollenkaan. Sekä Bucher (2018: 154) että Martin (2018: 9) esittävät, että suunnittelijat ja päätöksentekijät ovat paljon haltijoita: heillä on valta päättää, mitä algoritmi tekee, mutta myös kyky ja mahdollisuus korjata virheet ja epäoikeudenmukaisuudet. Algoritmisen merkityssisältö on *vastuullisuus*.

6.1.4 Tunteminen

Algoritmin tuntemiseen liittyvät kysymykset ovat epistemologisia, tietoon liittyviä. Mitä algoritmista voidaan tietää? Miten algoritmeja voidaan parhaiten tutkia?

Algoritmin olemusta ei ymmärretä riittävästi (Ziewitz 2016). Kuten aiemmin on todettu, algoritmi voidaan määritellä lukuisilla tavoilla, mutta tyhjentävää esitystä siitä, mikä algoritmi lopulta on, ei voida antaa. Ziewitzin (2016: 11) mukaan:

Rather, the goal should be to keep our inquiries generative enough to invite us to revisit some of our own assumptions and beliefs of what an algorithm actually is. (Kursivointi Ziewitzin.)

Edellä sanotun perusteella merkityssisältö on *ymmärtämättömyys*.

Algoritmeja tulee tutkia enemmän (Ananny 2016: 109; Dourish 2016: 8–9; Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 11–13; Danaher ym. 2017: 15; Kitchin 2017: 27; Uricchio 2017: 137; Bucher 2018: 147, 155; Martin 2018: 13). Mittelstadt ym. (2016: 12) näkevät algoritmin vaikutusten moniulotteisuuden tärkeimpänä syynä siihen, että ratkaisujen on tultava laajasta tietämyskentästä. Uricchio (2017: 137) painottaa algoritmin käyttöönoton ja toiminnan laajempaa tuntemista. Eräs huomiota vaativa huolenaihe on humanin ja ei-humanin toimijoiden roolit algoritmisessa päätöksenteossa (Mittelstadt ym. 2016: 11; Martin 2018: 13). Algoritminen merkityssisältö on *tutkimattomuus*.

Algoritmeja tulisi tutkia poikkitieteellisesti (Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar 2017: 14; Kitchin 2017: 26–27). Kitchin (2017: 27) näkee arvokkaana erilaiset tutkimusmenetelmät ja niiden tuottaman tietämyksen. Danaher ym. (2017: 14) toteavat, että on olemassa selkeä tarve rakentaa poikkitieteellistä yhteistyötä. Eri tieteiden väliset kommunikaatio-ongelmat rajoittavat syvällisen ymmärryksen kokoamista algoritmeista. Tutkijat ehdottavat monitieteellisten verkostojen luomista, jotta tietoa ja tutkimustuloksia voidaan jouhevasti jakaa. Algoritminen merkityssisältö on *poikkitieteellisyys*.

ALGORITMINEN YMPÄRISTÖ			
PERUSOMINAISUUDET	TOIMINTA	VAIKUTUKSET	TUNTEMINEN
MONIPUOLISUUS	TIEDONKERUU	PUOLUEELLISUUS	YMMÄRTÄMÄTTÖMYYS
SOSIOTEKNISYYS	TIEDON HALLINTA	ENNUSTAMATTOMUUS	TUTKIMATTOMUUS
KONTEKSTIRIIPPUUUS	ENNUSTAMATTOMUUS	LÄPINÄKYMÄTTÖMYYS	TUNTEMATTOMUUS
EETTINEN VAIKUTUKSELLISUUS	PUOLUEELLISUUS	TIETÄMÄTTÖMYYS	POIKKITIETEELLISYYS
	ALGORITMINEN PÄÄTÖKSENTEKO	VASTUULLISUUS	
	LÄPINÄKYMÄTTÖMYYS		
	TUNTEMATTOMUUS		
	DIGITAALINEN IDENTITEETTI		
	DATAMALLIT		
	DATAN TULKINTA		
	YKSITYISYYS		
	TIETOSUOJA		

Kuva 2. Algoritminen ympäristö eettisen linssin läpi tarkasteltuna.

6.2 Algoritmit ja etiikka

Algoritmit ja etiikka on algoritmisen ympäristön ylempi kategoria. Algoritmisen ympäristön merkitysisällöt on työstetty uudelleen yhdistellen ja muodostaen viisi uutta luokkaa. Tarkastelulinssi on yhä eettinen. Luokkien nimet kuvaavat keskeisintä sisältöä. Kursivoituja algoritmisia merkitysisältöjä avataan kussakin kappaleessa. Joissakin tapauksissa on ollut tarpeellista selvittää merkitysisältöä muutaman kappaleen verran. Kappaleen alussa oleva kursivoitu merkitysisältö kertoo milloin siirrytään uuden aiheen käsittelyyn.

6.2.1 Kontrolli

Algoritminen hallinta, tiedon hallinta ja algoritminen päätöksenteko ovat kaikki sidoksissa algoritmien toimintaan. Kaikissa kolmessa kulkee algoritmisen sääntelyn punainen

lanka, jonka vuoksi luokka on saanut nimen ”Kontrolli”. Mitä tämä kontrolli on ja miten se ilmenee, sitä avataan seuraavassa. Algoritmien toimintaa käsitellään luokassa ”Algoritmien toiminta”, joten tässä ei puututa algoritmisen päätöksenteon algoritmisiin vaikutuksiin kuin yhden esimerkin verran. Sen tarkoitus on havainnoida päätöksenteon haasteita.

Cheney-Lippoldin (2017: 99) mukaan algoritmien säätely edustaa ”kontrollia, jota on vaikea havaita tai kontrollia, joka on etäistä ja harvoin koettua”. Tämä valta ilmenee avoimen ympäristön kautta hyvin hienovaraisella mukauttamisella. Valtaan liittyy *datafikatio*, joka tarkoittaa arvon luomista datalle (Cheney-Lippold 2017: 107). Toisin sanoen, ihmisen toiminta muuttuu tai muutetaan laskettavaksi dataksi, jolla on *kaupallista* arvoa. (Cheney-Lippold 2017: 101.)

Googlen entisen operatiivisen johtajan Douglas Merrillin kuolemattomiin lausuntoihin kuuluu: ”All data is credit data” (O’Neil 2016: 158). Näin näyttää olevan. Eräs tuttavani valitti hiljattain, että kun puhuu kaverinsa kanssa puhelimesta vaikka säästä, niin on kohta säämainoksia Facebook täynnä. En ole Facebookissa, mutta aloitin eräässä WhatsApp-ryhmässä (WhatsApp on Facebookin omistama) keskustelun nivelrikosta ja niin alkoi älypuhelimien ilmestyä nivelrikolläkemainoksia. Facebookin omistama algoritmi siis kerää, yhdistää ja käyttää tietoa myös sieltä, minne käyttäjä ei ole antanut suostumusta.

Tiedon hallinta algoritmisessa mielessä ymmärrettynä on sidoksissa algoritmien toimintaan. Kun algoritmi lajittelee tietoa, priorisoi, luokittelee, suodattaa, yhdistää – kaikessa on kyse siitä, että algoritmi hallitsee ja käyttää tietoa ilman loppukäyttäjän suostumusta ja väliintuloa. Prosessi on tavallisesti läpinäkymätön eli prosessista ei saa tietoa. Tiedon hallinta on osa normaalia algoritmista säätelyä ja algoritmista päätöksentekoa.

Algoritmisella hallinnalla on tänä päivänä paremmat toimintamahdollisuudet kuin koskaan aiemmin. Teknologia on läsnä kaikkialla, nopeus ja skaalautuvuus ovat kasvaneet, ja prosessit ja datan käsittely ovat tehokkaampia kuin ennen. Tiedon kerääminen, tiedon käsittely, käyttäminen, palautteenanto ja oppiminen ovat ne toiminnot, jotka luotaavat

algoritmista hallintaa. (Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar 2017: 2–3.)

Päätöksenteossa algoritmien toiminta voidaan järjestää kahdella tavalla, joko *top-down* tai *bottom-up*. Top-down -mallissa ohjelmoija tai ohjelmointitiimi määrittelee algoritmille säännöt. Bottom-up -mallissa koneoppivalle algoritmille luodaan oppimissääntö ja se oppii harjoitusdatan avulla (Dourish 2016: 7; Danaher ym. 2017: 3). Bucherin (2018: 24) mukaan koneoppivia algoritmeja on kolmenlaisia: ohjattuja, ohjaamattomia ja puoliohjattuja. Ohjatussa algoritmista harjoitusdatasta tiedetään haluttu tulos. Ohjaamaton algoritmi ei sisällä dataa halutusta tuloksesta vaan oppii itsenäisesti. Puoliohjattu algoritmi oppii vuorovaikutuksessa sisään syötetyn mallin ja ympäristönsä kanssa. Koneoppivat algoritmit ovat itseohjautuvia autokraatteja (Crawford 2016: 82; Uricchio 2017: 127).

Danaherin ym. (2017: 4) mukaan algoritmiseen päätöksentekoon liittyy epätarkkuutta, tehottomuutta ja arvaamattomuutta. Algoritmisen päätöksenteon ja sen aiheuttaman vaikutuksen välillä voi olla aukko, jolla on vakavat seuraukset (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 2, 11). Äärimmäisessä tapauksessa kyseessä voi olla ihmishengen menetys (Crawford 2016: 83–85). Cheney-Lippold (2017: 203–204) kertoo tapauksesta, jossa amerikkalainen mies soitti hätäkeskukseen ja algoritmi esti operaattoria lähettämästä ambulanssia. Hätätapauksiin operaattorin avuksi oli laadittu kysymyslista. Ambulanssin lähettäminen oli sidoksissa kyseiseen listaan ja ”väärät vastaukset” olivat este ambulanssin lähettämiselle. Kyseisessä tapauksessa mies löytyi kotinsa lattialta maakaamasta kahden päivän kuluttua ja kuoli pian sairaalaan tuonnin jälkeen. Algoritmi toimi oikein eli teki sitä, mitä se oli suunniteltukin tekemään, mutta esti samalla inhimillisen puuttumisen päätöksentekoon. Mittelstadt ym. (2016: 11) painottavat, että eräs problematiikka algoritmisessa päätöksenteossa on humanin ja ei-humanin päätöksentekijän roolit. Tilanteet, joissa toimii sekä humaanin että ei-humaani päätöksentekijä, ja joissa ehdottomasti tarvitaan humanin päätöksentekijän puuttuminen, tulee kuvata riittävän huolellisesti etukäteen.

6.2.2 Sokaiseva pimeys

Sokaiseva pimeys käsittää niitä algoritmisia ulottuvuuksia, joiden läheisyydessä pimeys tuntuu vain sakenevan. Algoritmiin itseensä liittyy asioita, joiden vuoksi sitä on vaikea selittää tai ymmärtää. Algoritmiseen päätöksentekoon liittyy kiinteästi läpinäkymättömyys. Mustat laatikot ovat suuri tuntematon ja usein tutkimaton alue. Algoritmien tutkimattomuuteen liittyy omia erityiskysymyksiään.

Ymmärtämättömyyden ongelma liittyy algoritmin kohdalla sen perusolemukseen. Se on ontologinen ja epistemologinen kysymys. Algoritmia ei ymmärretä, koska ei ole olemassa kaikille tutkijoille yhteistä kehystä. Määritelmiä on yhtä monta kuin tutkijoitakin. Algoritmin toimintaan liittyy myös asioita, joita ei ymmärretä eikä osata selittää. O’Neil (2016: 173) antaa esimerkiksi käyttäytymisdataa operoivat älykkäät järjestelmät, jotka automaattisesti määrittelevät kuinka data tulkitaan.

Läpinäkymättömyys ilmenee useimmiten algoritmisessa päätöksenteossa. Läpinäkymättömyys tarkoittaa siis tilannetta, jossa ei tiedetä millä perusteella päätökset tehdään tai miten algoritmi toimii. Martin (2018: 13) huomauttaa, että läpinäkymättömyys on erilaista eri konteksteissa. Kitchinin (2017: 20) mukaan järjestelmään upotetut algoritmit ovat läpinäkymättömiä. Ne esiintyvät usein monimutkaisissa kokoonpanoissa. Mittelstadtin, Allon, Taddeon, Wachterin & Floridin (2016: 6) mukaan ”läpinäkyvyyden peruskomponentit ovat saavutettavuus ja informaation ymmärrettävyys.” Päätöksenteko on huonosti läpinäkyvää, koska yritykset käyttävät patentoituja algoritmeja liiketaloudellisista syistä (Dourish 2016: 6; Mittelstadt ym. 2016: 6; Kitchin 2017: 20; Bucher 2018: 41). Läpinäkyvyyttä halutaan, koska algoritmeja on vaikea valvoa, seurata tai korjata (Mittelstadt ym. 2016: 6). Zarskyn (2016: 122) mukaan läpinäkyvät prosessit ovat tarkempia ja tehokkaampia, mutta eivät ongelmattomia. Läpinäkyvyyden kustannukset voivat kohota korkeiksi eivätkä läpinäkyvät prosessit vaikuta vahvistavan sosiaalista hyvinvointia (Mittelstadt ym. 2016: 7; Zarsky 2016: 122).

Tuntemattomuus liittyy ennen kaikkea niin sanottuihin mustiin laatikoihin. Musta laatikko tarkoittaa prosessia tai objektia, jonka ”sisäistä toimintaa ei tunneta” (Bucher 2018: 43). Tavallisesti mustat laatikot aiheuttavat ongelmia (Bucher 2018: 45). Tutkijoiden

suhtautuminen mustiin laatikoihin on kahtalaista. Toiset pyrkivät avaamaan niitä, toiset etsivät muita ratkaisuja pyrkiessään tuntemaan tuntemattoman.

Eräs ratkaisu mustien laatikoiden tuntemiseen on käänteinen suunnittelu. Jokaisessa mustassa laatikossa on kaksi kohtaa, joista voidaan saada tietoa: syöttö ja ulostulo (Kitchin 2017: 24; Bucher 2018: 41–65). Havainnoimalla toimintaa ja tapahtumia voidaan tehdä päätelmiä toimintalogiikasta. Tässä on kyseessä tarkastelukulman muutos. Ei yritetä käydä käsiksi suoraan tuntemattomaan, vaan tarkkaillaan sitä, mistä voidaan saada tietoa. Mahdollista on myös tarkkailla esimerkiksi online foorumilla, miten käyttäjät ymmärtävät ja kokevat algoritmin muuttuneen. Sosiaalisen median algoritmit muuttuvat koko ajan (Ananny 2016: 107). (Kitchin 2017: 23–24; Paßmann & Boersma 2017: 142–145; Bucher 2018: 59–61.)

Toinen mahdollisuus tarkastella mustia laatikoita on koodin tutkiminen. Tämä tapahtuu joko tutkimalla pseudokoodia tai lähdekoodia. Pseudokoodi tarkoittaa jäljelle jätettyä algoritmin perusrakennetta, josta on muu koodi puhdistettu pois. Koodin tutkimisessa koodi pitää ensin purkaa. Käytännössä koodin tutkiminen on vaikeaa ja edellyttää sekä ohjelmoijan että tutkijan taitoja. (Kitchin 2017: 22.)

Tutkimattomuus liittyy paitsi mustiin laatikoihin, niin algoritmiin. Edellä tarkasteltiin kahta tapaa tutkia mustia laatikoita, joten tässä keskitytään algoritmeihin. Algoritmien tutkimus on ollut hyvin toimintakeskeistä. Kitchinin (2017: 16) mukaan algoritmien tutkimus on ollut lähinnä kolmenlaista. Ensimmäisessä joukossa ovat yksityiskohtaiset tapautkimukset tai tutkimukset algoritmien luokista, joissa on tutkittu algoritmien luonnetta yleisesti. Toisena joukkona ovat jonkin tietyn osa-alueen käyttöön liittyvät yksityiskohtaiset tarkastelut, kuten esimerkiksi journalismi. Kolmannen joukon muodostavat yleisemmät, arvioivat tarkastelut algoritmin luonteesta ja toimivuudesta. Huomattavasti laajemmalle ja monipuolisemmalle algoritmitutkimukselle näyttää olevan tilausta.

6.2.3 Digitaalinen identiteetti

Digitaalinen identiteetti on tietokoneistettu malli ja tulkinta siitä kootusta datasta, jota meistä on saatu kerättyä sekä vapaaehtoisesti luovuttamalla että luvatta. Näin saatuja datamalleja ja käyttäytymiskuvioita hyödynnetään monin tavoin. Seuraavassa avataan datamallien ja tulkinnan muodostumista sekä paneudutaan yksityisyyden ja tietosuojan merkitykseen digitaalisen identiteetin yhteydessä.

Tiedonkeruu on kaiken internetissä kerättävän hyödylliseksi muokatun datan perusta. Koska etukäteen ei voi tietää, mikä data osoittautuu hyödylliseksi, vallalla on ideologia ”mitä enemmän sen parempi” (Johnson 2001: 118; O’Neil 2016: 89). Tällainen varmuuden vuoksi toimiminen aiheuttaa tarpeettomien tietojen kirjaamista. Tiedonkeruussa ongelmana on myös se, ettei ole varmuutta kuka tietoja tosiasiallisesti käyttää ja mihin tarkoitukseen.

Datamallit perustuvat kerättyyn dataan. Malleja käytetään yleensä osoittamaan käyttäjän identiteetti (Cheney-Lippold 2017: 47). Uutta dataa verrataan olemassa olevaan dataan. Tällä tavoin muodostuu algoritminen konstruktio, jota voidaan hyödyntää kulloinkin tarkoituksenmukaisella tavalla. Cheney-Lippold (2017: 66) huomauttaa, että ”datamallit eivät ole totuus, vaan konstruktioita, algoritmisesti tuotettuja totuuksia.”

Cheney-Lippold (2017: 43–44) esittää erään käytökseen perustavan datamallin ja sen tulkinnan. Bill ja Ted asuvat samassa osoitteessa. Bill vuokraa kuorma-auton ja Ted ostaa ammoniumnitraattia, jota käytetään paitsi perunoiden kasvatuksessa myös kotitekoisten pommien tekemiseen. Jostakin syystä kaksikko vierailee tarkkailtavana olevassa kohteessa. Käyttäytymiskuvion perusteella (sama osoite, kuorma-auton vuokraus, ammoniumnitraatin osto ja vierailu ”kohteessa”) voidaan Bill ja Ted luokitella silmälläpidettäviksi terroristeiksi.

Datan tulkintaa voidaan siis tehdä esimerkiksi siten, että kun tietyt tunnusmerkit henkilön käyttäytymisessä täyttyvät, hänet kategorisoidaan johonkin ryhmään kuuluvaksi. Edellä oleva ”Bill ja Ted” -esimerkki voisi todellisuudessa selittyä siten, että he ovat veljeksiä, jotka ovat päättäneet ryhtyä perunankasvatukseen, jota varten tarvitaan kuorma-auto. Vierailu ”kohteessa” voisi olla pelkkä paikallisten perunankasvattajien yhdistyksen

järjestämä yhdistystapahtuma historiallisesti merkittävälle paikalle. Perunankasvattajien yhdistykseen Bill ja Ted ovat vasta liittyneet, joten tietoja ei ole saatavilla.

Toinen esimerkki on yhdysvaltalainen COMPAS uusintarikollisuuden riskipisteytysjärjestelmä. COMPAS todettiin hyvin kalibroiduksi järjestelmäksi, koska sen antama prosenttiluku vastasi todellista tilannetta. Kuitenkin tuloksissa ilmeni ongelmia. Järjestelmä antoi kaksi kertaa enemmän vääriä positiivisia tuloksia mustaihoisille rikollisille ja kaksi kertaa enemmän vääriä negatiivisia tuloksia valkoihoisille rikollisille. Järjestelmä toimi oikein, mutta antoi eriarvoistavia, vääriä tuloksia. COMPAS kykeni ennustamaan 20 % uusintarikoksista eli suurin osa tapauksista jäi järjestelmän ulottumattomiin (Martin 2018: 4). (O’Neil 2016: 224; Martin 2018: 4.)

Yksityisyys ja tietosuojat kulkevat käsi kädessä, toista ei ole ilman toista. Digitaalinen kehitys haastaa yksityisyyden. Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brochain, Schafer & Shankar (2017: 13) ovat huolissaan riittämättömästä yksityisyyden suojasta, sillä se on murenemassa. Cheney-Lippold (2017: 219) kertoo esimerkin huomiota herättävästä yksityisyysloukkauksesta. Kyseessä on Facebookissa operoinut Gaydar -algoritmi. Algoritmi päätteli henkilöiden ystävyysuhteiden perusteella ketkä ovat todennäköisesti homoseksuaaleja. Lähtöoletuksena oli tavallista suurempi yhteys samaa sukupuolta oleviin henkilöihin. Gaydar käytti hyväkseen metadatan, jonka avulla se ryhmitteli henkilöt homoseksuaaleihin ja heteroseksuaaleihin.

Konepohjainen analytiikka seuloa aina vain suurempia massoja dataa. Entistä tärkeämpään osaan on nousemassa oikeus omaan dataan. Käyttäjä ei voi tälläkään hetkellä olla täysin varma siitä, mitä hänen lataamilleen omille tiedoilleen tapahtuu internetissä. Oma data voi kuitenkin jollakin tasolla hallita erilaisilla alustoilla. Horisontissa siintävät silti mustat pilvet. Internetin vapauden idea on jo kääntynyt markkinajättien leikkikentäksi, jossa vain voitot ratkaisevat. Yksittäisen ihmisen mahdollisuus hallita omia tietoja käy vaikeammaksi. Tämä siitä huolimatta, että Euroopassa on jo voimassa uusi tietosuojalaki. Globaalissa mittakaavassa on paljon työtä jäljellä. Maailmassa on useita valtioita, joilla

on suuri intressi kontrolloida kansalaisiaan. Tulevaisuudessa koettanee päivä, jolloin joudumme maksamaan omasta datastamme (Cheney-Lippold 2017: 256).

6.2.4 Algoritminen toiminta

Algoritmin monimutkaisuuden taustalla ovat algoritmin perusominaisuudet (kts. 6.1.1 Perusominaisuudet). Monimutkaisen algoritmista tekee kuitenkin sen muuntuvuus eri kokoonpanoissa ja ympäristöissä. Kun tähän lisätään eri algoritmien erilaiset tavoitteet, jotka voivat olla ristiriitaisia keskenään (Crawford 2016; Bucher 2018) ja se seikka, että samaa algoritmia voidaan käyttää eri tavoin eri tilanteissa (Dourish 2016: 4; Kitchin 2017: 21), näyttäytyy algoritmi erittäin muuntautumiskykyisenä ja laajana objektina. Pitää myös muistaa, että kaikella teknologialla on taustamotiivinsa. Algoritmeihin liitetyt oletukset tai odotukset (Bucher 2018: 23; Martin 2018: 6) eivät välttämättä pidä paikkaansa. Tästä aiheutuu virheitä tai vaikutuksia, joita on vaikeaa tai mahdotonta ennakoita.

Ennustamattomuus on dynaamisten algoritmien toiminnan luonnollinen seuraus. Teoriassa suunniteltu käyttö ja algoritmin käytännön toiminta voivat suuresti poiketa toisistaan (Kitchin 2017: 25). Ennustamattomuutta aiheuttavat paitsi virheet ja sivuvaikutukset, myös käytetyillä teknologioilla voi olla yllättäviä seurauksia (Ananny 2016: 108; Kitchin 2017: 25). Zarsky (2016: 120) toteaa, että organisaatioiden automaattiset järjestelmät olettavat ihmisen käyttäytymisen olevan sekä johdonmukaista että ennustettavissa. Koska algoritmi ei pysty ennustamaan ihmisen käyttäytymisen ennustamattomuutta, sen on perustettava arvionsa olemassa olevaan dataan. Tästä syystä esimerkiksi Netflixin suosittelija-algoritmi suosittelee lisää samanlaista sisältöä (Bucher 2018: 2).

Puolueellisuus liittyy algoritmiseen toimintaan. ”Algoritmit tekevät väistämättä puolueellisia päätöksiä” (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 7). Mittelstadt ym. (2016: 7–8) mukaan puolueellisuus voi syntyä monella tapaa. Se voi olla tahaton seuraus tai tarkoituksella suunniteltu. Se voi olla peräisin sosiaalisista käytännöistä tai asenteesta, teknisistä rajoitteista tai virheistä, suunnittelusta tai käyttäjästä. Eriarvoisuutta

ja suoranaista syrjintää voidaan estää mm. reiluun tietoon perustavalla tiedonlouhinnalla, päätösten kohtuullistamisella ja algoritmeille asetettavilla kriteereillä.

Zarsky (2016: 127) toteaa, että ”algoritmien sisältämä läpinäkymättömyys luo selviä ongelmia, jotka liittyvät oikeudenmukaisuuteen”. Toisiinsa yhteydessä olevat tietojärjestelmät lisäävät syrjintää samojen ihmisten kohdalla (Zarsky 2016: 128). Tämä tapahtuu tietojen skaalautumisen vuoksi. O’Neil tunnistaa saman ongelman (2016: 28–31).

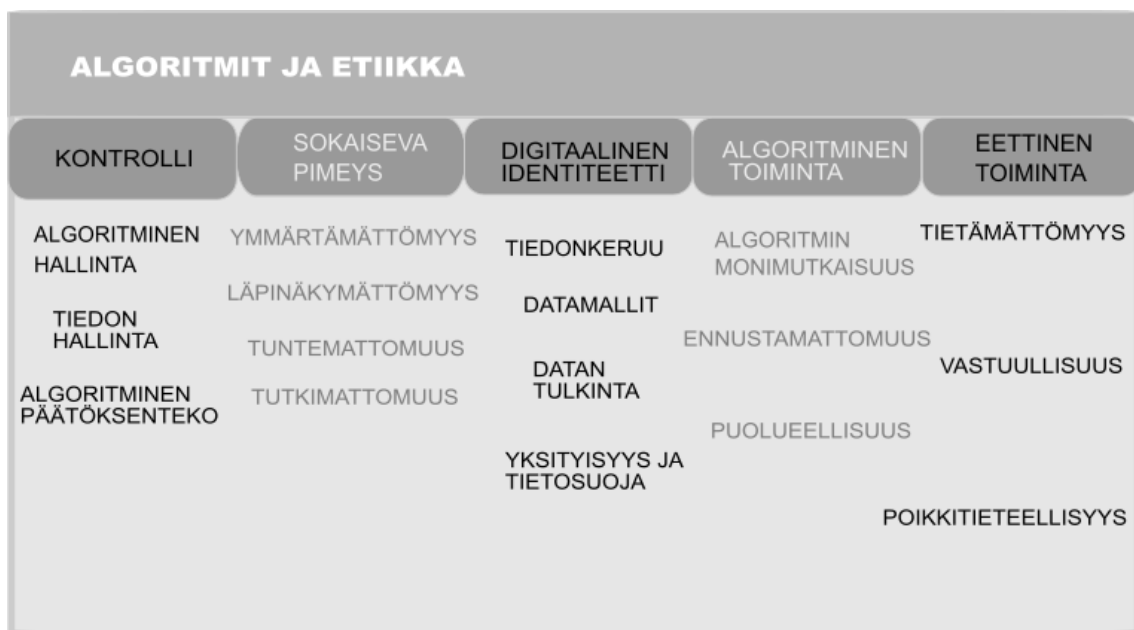
6.2.5 Eettinen toiminta

Tietämättömyys algoritmien aiheuttamista eettisistä vaikutuksista on laaja-alaista. Martin (2018: 3) toteaa, että yrityksissä halutaan uskoa ja toivoa algoritmien olevan arvoneutraaleja, joilla ei ole eettisiä vaikutuksia tai on niitä vain hyvin vähän. Tällöin ei asiaan tarvitse ottaa kantaa eikä eettisiä vastuukysymyksiä ole. Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brochain, Schafer & Shankar (2017: 13) laajentaisivat keskustelua eettisten vaikutusten tietämättömyydestä kaikkiin sidosryhmiin, niin yleisöön kuin päättäjiin. Heidän mukaansa asiaa tulisi tutkia monipuolisesti erilaisin tutkimusmenetelmin (Danaher ym. 2017: 18).

Vastuullisuus on entistä tärkeämpää, sillä algoritmit vaikuttavat ihmisten elämään monella tavalla, josta syntyy eettisiä vaikutuksia. Algoritmit hallitsevat osallistumista alustoilla, jakavat informaatiovirtaa, vahvistavat sosiaalisia ennakkoluuloja ja algoritmit saavat ihmiset käyttäytymään erityisillä tavoilla (Bucher 2018: 120). Tutkijat (O’Neil 2016: 223; Bucher 2018: 154; Martin 2018: 2) peräävät vastuullisuutta yrityksiltä, jotka suunnittelevat algoritmeja. Suunnittelijat tekevät moraalisia valintoja päättäessään millaisia algoritmeja kehittävät ja mihin tarkoituksiin. Suunnittelupäätöksillä on merkitystä (Bucher 2018: 154; Martin 2018: 9). Martinin (2018: 4–5) mukaan arvoladatut algoritmit vaikuttavat myös eettisten periaatteiden ilmenemiseen sekä ihmisoikeuksien ja ihmisarvon toteutumiseen. Yrityksen kehittäessä tarkoituksellisesti eettisiä vaikutuksia omavia algoritmeja yrityksestä itsestään tulee päätöksenteon osapuoli (Martin 2018: 10).

Algoritmien eettiselle vastuunjaolle ei ole olemassa valmista karttaa. Mittelstadt, Allo, Taddeo, Wachter & Floridi (2016: 10–12) esittävät, että vastuuta tulisi pyrkiä jakamaan. He näkevät käsikirjoitetun, ei-oppivan algoritmin vastuunkantajana sen suunnittelijan, koska tämä voi kontrolloida algoritmin toimintaa. Mutta koneoppivissa järjestelmissä tulisi olla heidän mukaansa toisenlainen vastuukehys. Autonomisella, interaktiivisella kone-toimijalla tulee olla syyperustainen vastuu, mutta ei voida olettaa humanin toimijan moraalista vastuuta. Tutkijoiden mukaan myös yritysten ja suunnittelijoiden on kannettava oma vastuunsa.

Poikkitieteellisyydelle on algoritmitutkimuksessa selkeä tarve (Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar 2017: 14). Syvempää ja laajempaa näkemystä tarvitaan, jotta eettisen toiminnan kysymyksiin kyettäisiin paremmin vastaamaan. ”Algoritmien eettiset huolenaiheet ovat moniulotteisia, joten vaaditaan moniulotteisia ratkaisuja” (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 12). Bucher (2018: 147) huomauttaa, että algoritmien ymmärtäminen edellyttää niiden tapojen tuntemista, joilla algoritmit on tehty merkitseviksi.



Kuva 3. Algoritmit ja etiikka eettisen linssin läpi.

6.3 Synteesi

Synteesi muodostuu, kun molemmat kategoriat yhdistetään samaan kattaukseen ja mukaan pujotetaan kaikkien luokkien läpi taustalla olevat *yhdistävät tekijät*. Nämä ovat *algoritmisen päätöksenteko*, *vastuullisuuden vaatimus* ja *algoritmeja on tutkittava enemmän*. Näiden aspektien esiintymistä avataan seuraavassa luokkakohtaisesti.

6.3.1 Yhdistävät tekijät ”Algoritmisen ympäristö” -kategoriassa

Perusominaisuudet -luokka kuvaa algoritmin monimuotoisuutta eri puolilta. Erilleen otettu, yksittäinen algoritmi ei ole eettisesti merkityksellinen. Kuva muuttuu heti, kun algoritmi on vuorovaikutuksessa muihin toimijoihin (Kitchin 2017: 20). Toimijat voivat olla teknisiä komponentteja tai ihmisiä. Vuorovaikutuksessa ympäristönsä kanssa algoritmi on laaja-alainen ja moniroolinen objekti (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016; Kitchin 2017; Bucher 2018). Algoritmeja käytetään usein päätöksentekotilanteissa. Näitä ovat esimerkiksi erilaiset nettikauppojen ohjelmistoagentit ja terveydenhuollon päätöksentekojärjestelmät (Mittelstadt ym. 2016: 3). Algoritmisen päätöksenteko on siis vuorovaikutteinen sosiotekninen kokonaisuus ja sillä on eettistä merkitystä, kuten on aiemmin todettu.

Algoritmeja tulee tutkia enemmän. Dourish (2016: 8) kiinnittää huomiota algoritmien erilaisiin ilmentymiin ja perää huomiota algoritmin ymmärtämiseen erilaisissa ympäristöissä. Samaa tutkimuksen vaatimusta tukee Bucher (2018: 147) todetessaan, että tulee ymmärtää kaikkia niitä ympäristöjä, joissa algoritmi on tehty merkitseväksi ja miten tämä on tapahtunut.

Toiminta -luokka esittelee algoritmien toimintaan liittyviä piirteitä ja ulottuvuuksia. Algoritmisen voima perustuu algoritmien kykyyn toimia erilaisissa kokoonpanoissa ja suorittaa erilaisia toimintoja. Tämä tapahtuu algoritmisessa päätöksenteossa. Algoritmisen päätöksenteko on prosessi, jonka heikentävänä tekijänä pidetään automaatiota ja läpinäkymättömyyttä (Zarsky 2016: 125, 129–130). Voi esiintyä virheitä, epätarkkuutta,

ihmisten henkilökohtaisia tietoja käytetään luvatta eikä sen paremmin algoritmien kuin ihmisten käyttäytymistä kyetä luotettavasti osoittamaan. Läpinäkymättömyys estää virheiden ja arvaamattomien tulosten syyn selvittämistä. Virheet saattavat myös korvata toisiaan, korjaantua itsestään tai niillä saattaa olla vain rajallinen vaikutus lopputulokseen. (Zarsky 2016.)

Toisaalta algoritmisessa päätöksenteossa piilevät suuret eettiset mahdollisuudet. Ananyn (2016: 104) mukaan ”algoritmien eettinen voima luodaan hallitulla verkostolla ihmisistä ja koneista, jotka uudelleenluovat todennäköisyyksiä... Tehokas algoritmien kokoonpano käyttää todennäköisyyksiä ehdottaakseen mitä pitäisi tehdä.”

Sekä algoritmien toimintaa että ihmisten käsityksiä tulee tutkia enemmän. Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar (2017: 7) esittävät, että tulisi tutkia yleisön ja asiantuntijoiden näkemyksiä siitä, kuinka algoritmiset järjestelmät toimivat. Tämä tulisi heidän mukaansa tehdä eri tutkimusmenetelmiä yhdistelmällä kattavan näkemyksen aikaansaamiseksi. Dourish (2016) ja Kitchin (2017: 26) tarjoavat jatkotutkimusaiheeksi algoritmien ja muiden tietorakenteiden keskinäisten yhteyksien tutkimista. Uricchio (2017: 137) ehdottaa lisää tutkimusta nimenomaan algoritmien käyttöönottoon liittyen sekä niiden kykyjen, rajojen, vaikutusten ja mahdollisuuksien selvittämistä.

Vaikutukset -luokka tarkastelee algoritmien toiminnan aiheuttamia eettisiä vaikutuksia. Algoritmien päätöksenteko ilmenee maailmankuvan muokkaajana. Tämä päätöksenteon linssi näyttää meille digitaalisen järjestelmän kontrollin, alati paisuvat tietovarastot, demokraattisen osallisuuden vähenemisen sekä yksityisyyden ja saavutettavuuden ongelmat (Dourish 2016; O’Neil 2016). Uudet teknologiat monipuolistavat entisestään algoritmien potentiaalia (Dourish 2016: 6).

Päätöksentekoprosessilla on eettisiä vaikutuksia, vaikka niiden olemassaoloa ei aina tunnusteta. Algoritmi on päätöksentekoprosessissa arvoladattu. Algoritmisen päätöksentekoketjun molemmissa päässä on jossain kohdin ihminen. Alkuperäessä se on algoritmin

suunnittelija. Jos algoritminen päätöksentekoprosessi ei suoraan vaikuta ihmiseen, on josain prosessin tulokohdassa inhimillinen vastaanottaja, joka tekee tulkintaa tuloksista.

Algoritmin aiheuttamia vaikutuksia tulee tutkia enemmän. Kitchin (2017: 26) näkee erityisesti sosiaalisten ja humanististen tieteiden tutkimuksen algoritmeista ja algoritmisen hallinnan muodoista tarpeelliseksi. Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar (2017: 13) kokevat hyödylliseksi tutkimuksen eettisen tietämyksen ja valvutuneisuuden tasosta niin yleisölle, päättäjille kuin järjestelmänrakentajille. He ehdottavat myös tutkimusta teknologisen kehityksen vaikutuksista niin historialliselta kuin eettiseltä kannalta (Danaher ym. 2017: 18).

Tunteminen -luokka keskittyy algoritmin tuntemiseen liittyviin kysymyksiin. Vaikka algoritminen päätöksenteko on luettavissa suoraan vain päätöksenteon toimijoiden rooleista, ovat sekä ymmärtämättömyys että tutkimuksen vaatimus relevantteja myös algoritmiselle päätöksenteolle. Algoritminen päätöksenteko näyttäytyy siten tutkimattomana rakenteena, jonka aiheuttamia vaikutuksia ei ole läpivalaistu. Tähän viittaavat mm. edellä esitetyt tutkimusaiheet usean tutkijan suulla. Tutkiminen asettaa haasteita. Pääsyä järjestelmään voi olla vaikea saada, useiden algoritmien kokonaisuudet voivat olla upotettuina laajoihin ja kompleksisiin rakennelmiin eikä kokonaisuus avaudu kuin omassa kontekstissaan (Kitchin 2017: 20–21).

Algoritminen päätöksenteko aiheuttaa sekä eriasteista haittaa että syrjintää. Mittelstadt, Allo, Taddeo, Wachter & Floridi (2016: 12–13) ehdottavat tutkimusaiheena kuinka voidaan vähentää algoritmiseen päätöksentekoon liittyviä haittoja. Eräs ratkaisu algoritmisen päätöksenteon problematiikkaan voi olla vuonna 2017 Association for Computing Machinery US Public Policy Councilin luomat algoritmien eettiset ohjeet, joilla pyritään vähentämään läpinäkymättömyyden tuomia ongelmia sekä lisäämään vastuullisuutta (kts. 2.2. Algoritmien eettiset ohjeet).

6.3.2 Yhdistävät tekijät ”Algoritmit ja etiikka” -kategoriassa

Kontrolli -luokka käsittelee algoritmien toimintaa nimenomaan säätelyn ja hallinnan näkökulmista. Algoritmisen päätöksentekojärjestelmä havaitaan keskeisenä kontrollirakenteena. Päätöksenteossa vastuukysymys on erityisen korostunut. Kuka on vastuullinen, kun algoritmi tekee päätökset? Erityisen kiinnostavaksi vastuunjakamisen rajanveto käy, kun se on tehtävä humanin ja ei-humanin toimijan kesken (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 11; Martin 2018: 8, 13).

Tärkeänä tutkimusaiheena Mittelstadt ym. (2016: 11) näkevät kysymyksen humanin ja ei-humanin päätöksenteon vastuista. Martin (2018: 13) ehdottaa lisätutkimusta algoritmisten päätösten etiikasta sekä ihmisen vaikutuksesta algoritmisessa päätöksenteossa ei-humanin toimijan kanssa. Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar (2017: 7) esittävät lisätutkimusaiheina kuinka koneoppivat järjestelmät kehittävät sääntöjä, joita käytetään päätöksenteossa ja missä määrin etiikka sisältyy koneoppimiseen sekä tutkimusta lohkoketjuteknologiasta vaihtoehtoisena hallintamekanismina.

Sokaiseva pimeys -luokka käsittelee niitä algoritmien ulottuvuuksia, jotka aiheuttavat hämmennystä: ymmärtämättömyyttä, läpinäkymättömyyttä, tuntemattomuutta ja tutkimattomuutta. Samat aspektit tuovat algoritmiseen päätöksentekoon mustan aukon kaltaisen massatihentymän. Toisin sanoen algoritmisen päätöksenteko ilmenee rakenteena, jonka kaikkia osia ei tunneta. Kuten aiemmin on tuotu esille, useat tutkijoiden esittämät lisätutkimusaiheet antavat viitteitä siihen, että algoritmista päätöksentekoa ei ymmärretä riittävästi. Läpinäkymättömyys on algoritmisen päätöksenteon perusongelma. Suljettu algoritmisen päätöksentekoprosessi on saavuttamaton eikä sen tuottamaa informaatiota voida täysin ymmärtää (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016: 6), koska ei tiedetä päätöksenteon perusteita (Uricchio 2017: 131). Näin suljettu järjestelmä voi tuottaa eettisesti mielivaltaisia tuloksia, vaikka järjestelmä itsessään toimii oikein (O’Neil 2016; Zarsky 2016: 129; Cheney-Lippold 2017).

Tuntemattomuus on mustien laatikoiden ongelma myös algoritmisessa päätöksenteossa. Päätöksentekojärjestelmät voivat olla niin monimutkaisia, että mustia laatikoita on mahdotonta selvittää. Tutkimusryhmä Danaher, Hogan, Noone, Kennedy, Behan, De Paor,

Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar (2017: 16–18) tuo esille useita tutkimusaiheita, jotka keskittyvät algoritmisen päätöksenteon parempaan tuntemiseen. Erilaisia tapoja tutkia algoritmeja ovat esittäneet niin Kitchin (2017: 22–26) kuin Bucher (2018: 41–65). Dourish (2016: 8) toteaa, että algoritmeja tulisi ymmärtää nykyistä paljon paremmin ja kiinnittää huomiota algoritmien erilaisiin ilmentymiin erilaisissa ympäristöissä. Tätä ajatusta tukevat myös Ziewitz (2016), Uricchio (2017) ja Bucher (2018). Paßmann ja Boersma (2017: 145) nimeävät jatkotutkimusaiheeksi mustien laatikoiden ”tunnettujen tuntemattomien” tutkimisen.

Digitaalinen identiteetti -luokka tarkastelee tiedonkeruun, datamallien ja datan tulkinnan kautta tietojen käsittelyä päätöksentekojärjestelmissä. Algoritmisen päätöksenteko ilmenee tietojen käsittelyä rakenteena. Oletuksena olevat mallit eivät välttämättä vastaa lainkaan todellisuutta, koska ”malli on aina yksinkertaistus todellisuudesta” (O’Neil 2016: 20). Väärä turvallisuuden tunne algoritmisen päätöksenteon virheettömyydestä saa yritykset luottamaan järjestelmän toimivuuteen (O’Neil 2016: 48). Kuitenkin algoritmisen päätöksenteko tuottaa virheellisiä päätöksiä (Mittelstadt, Allo, Taddeo, Wachter & Floridi 2016; Zarsky 2016), luo epäoikeudenmukaisuutta (O’Neil 2016; Martin 2018) ja loukkaa yksityisyyttä (Cheney-Lippold 2017). Tiedonkeruun tulisi tapahtua aina todelliseen tarpeeseen. Sen lisäksi kerättävän tiedon tulisi olla virheetöntä ja oikea-aikaista. Yksilön tulisi voida luottaa, että hänen tietojensa käytetään eettisesti. Keskustelua reilun datatalouden tarpeesta käydään jo, mm. Sitralla on menossa IHAN-hanke, joka tähtää reilun datan ratkaisujen löytymiseen (SITRA 2019).

Tiedonkeräämiseen, datamalleihin, datan tulkintaan ja yksityisyyteen kohdistuvaa tutkimusta tarvitaan lisää. Mittelstadt ym. (2016: 12) esittävät yksityisyyteen liittyviä lisätutkimusaiheita: Kuinka yksityisyys toimii ryhmätasolla, yksityisyyden mahdollistavat mekanismit analytiikassa sekä yksityisyyden säilyttävät tiedonlouhintatekniikat. Tutkijaryhmä huomauttaa, että ”algoritmit muuttavat tapaa, jolla identiteetti rakentuu, hallitaan ja suojataan”. Tulisiko siis henkilökohtainen data määritellä uudelleen? Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar (2017: 13) mainitsevat tutkimusaiheeksi yleisön

valveutuneisuuden selvittämisen kuinka dataa käytetään algoritmisen päätöksenteonjärjestelmissä.

Algoritminen toiminta -luokka painottaa algoritmien monimutkaisuutta, algoritmisen päätöksenteon ennustamattomuutta ja puolueellisuutta. Algoritminen päätöksenteko ilmenee rakenteena, jossa teoria ja käytäntö voivat poiketa huomattavasti toisistaan (Kitchin 2017: 25; Martin 2018: 6–7). Tämä voi tuoda harhaanjohtavia tuloksia. Mittelstadt, Allo, Taddeo, Wachter & Floridi (2016: 4–5) huomauttavat, että ”algoritmeja otetaan käyttöön tilanteissa, missä luotettavimmat ratkaisut eivät ole saatavilla tai ovat liian kalliita.” He muistuttavat, että ”päätelmät voivat olla vain niin luotettavia kuin on data, johon ne perustuvat.” Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar (2017: 13) esittävät tutkimusaiheeksi sen selvittämistä, miten epätasa-arvo ja puolueellisuus sisältyvät algoritmiseen päätöksentekoon.

Eettinen toiminta -luokka paneutuu tietämättömyyteen, vastuullisuuteen ja poikkitieteellisyyteen. Algoritminen päätöksenteko nähdään vastuullisuutta vaativana konstruktiona. Tässä työssä esitetyn tutkimusaineiston perusteella yritykset ovat pikemminkin halukkaita pesemään kätensä vastuukysymyksistä kuin panostamaan eettiseen tutkimukseen. Yritykset kehittävät päätöksentekojärjestelmiä ja myyvät niitä. Martinin (2018: 10) mukaan yritykset ovat merkittävä päätöksenteon osapuoli. Tähän asti yritys-elämä on saanut omillaan ehdoillaan suunnitella algoritmisia päätöksentekojärjestelmiä. Tutkijat ovat pidelleet sitä valonheitintä, joka on osoittanut algoritmisen päätöksenteon eettisiin vaikutuksiin.

Mittelstadt, Allo, Taddeo, Wachter & Floridi (2016: 13) toteavat, että on tarve saada ”algoritmeille realistiset säätelyjärjestelmät ja standardit”. Tutkimusryhmä on luonut oman eettisen mallinsa algoritmien yleisimmistä huolenaiheista. Ananny (2016: 109) on osallistunut tieteelliseen keskusteluun omalla algoritmien normatiivisen etiikan kehityksen mallillaan. Hän puhuu algoritmien etiikan laajan mallin puolesta. Danaher, Hogan, Noone, Kennedy, Behan, De Paor, Felzmann, Haklay, Khoo, Morison, Murphy, O’Brolchain, Schafer & Shankar (2017: 16–18) näyttävät toteen algoritmisen

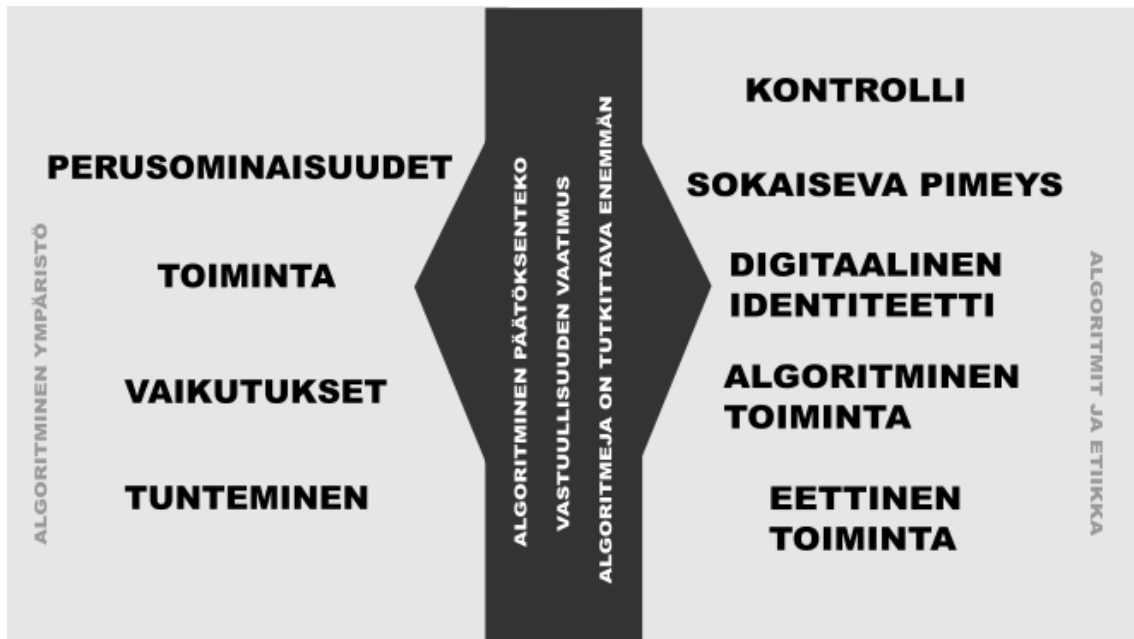
tutkimuksen tarpeen. Tutkimusryhmä on kartoittanut erittäin laajan algoritmisen hallinnon tutkimusohjelman, josta tässä työssä on esitelty vain murto-osa tutkimusaiheista.



Kuva 4. Algoritmisen päätöksenteon ominaisuuksien ilmeneminen synteesissä.

6.3.3 Yhteenveto

Edellä on etsitty fenomenografisen tutkimusmenetelmän kautta vastausta kysymykseen mitä on tutkittu algoritmien etiikasta. Algoritmien haasteet vaikuttavat keskittyvän nimenomaan algoritmiseen päätöksentekoon niissä tapauksissa, joissa käytetään joko puoliautomaattisia tai täysautomaattisia koneoppivia algoritmeja. Autokraattisiin järjestelmiin liittyy hyvin monimuotoisia ja kompleksisia haasteita. Eettiseltä kannalta katsottuna suurin ongelma on yhteydessä siihen, että algoritmien arvolatausta ei tunnusteta. Tämä vaikeuttaa eettisten ratkaisujen kehittämistä. Tämän työn tutkimusaineistosta nousee vahva, kiireellinen viesti kattavan algoritmien etiikan muodostamisesta. Erilaisia jatko-tutkimusaiheita tutkimusmateriaalissa esitettiin runsaasti. Jotta kattavaa kuvausta algoritmien etiikkaan voidaan tulevaisuudessa tehdä, on tarve monimuotoisemmalle ja poikki-tieteelliselle tutkimukselle. Tutkimusaineistosta ilmeni myös, että erilaiset tutkimuskoonpanot ovat kehittäneet omia ratkaisuja joihinkin rajattuihin algoritmisiin ongelmiin. Kaikki tämä tietämys tulisi saada saman kokonaismallin alle.



Kuva 5. Synteesi. Yhdistävät tekijät läpileikkaavat algoritmisen tulosavaruuden.

7. DISKUSSIO

Algoritmit ovat kaikkialla. Eettisiin näkökohtiin tulee kiinnittää huomiota, sillä tulevaisuudessa algoritmien käyttö lisääntyy ja monipuolistuu uusien teknologioiden myötä. Tutkijat ovat pystyneet aukottomasti osoittamaan, että algoritmisella päätöksenteolla on muuttuvia vaikutuksia. Osa näistä on eettisiä. Jos mitään ei tehdä, tulevaisuudessa odotavat entistä monimutkaisemmat ja vaikeasti ratkaistavat eettiset ongelmat. Ei voi olla edes globaaleiden yhtiöiden etu, että nykyiset haitat saavat jatkua, jatkuessaan ne kumuloiduvat.

7.1 Merkittävimmät tutkimustulokset

Tutkimuksen mukaan algoritmien merkittävimmät huolenaiheet liittyvät algoritmiseen päätöksentekoon ja sen aiheuttamiin vaihteleviin vaikutuksiin. Eettiset vaikutukset ovat tulleet esille nimenomaan tutkijakunnan taholta. Yritysmaailman suhtautuminen eettisiin huolenaiheisiin on ollut vähättelevää. Tarve algoritmin kokonaisvaltaiselle eettiselle mallille on olemassa. Kyseisen mallin tulisi ulottua kaikille tasoille. Yksittäisiä ratkaisuja on olemassa (mm. Association for Computing Machinery US Public Policy Councilin luomat algoritmien eettiset säännöt), mutta yksinään nämä tarjoavat vain osaratkaisuja suurempaan problematiikkaan. Tarvitaan globaali eettinen malli, johon myös yritykset sitoutuvat. Tämä globaalin mallin täytyy sisältää myös se, että algoritmisesta päätöksenteosta saadaan eettisesti kestävää.

Suurimmat haasteet ovat olleet ajatuksellisia. Tutkimusmetodi oli minulle uusi ja sen käytössä oli opettelemista. Suuri apu, ja luultavasti koko työn pelastus, on ollut heti alkuvaiheessa luotu teemakehikko ja jokaisesta dokumentista tehdyt miellekartat. Tärkeässä roolissa on myös ollut tutkimusmateriaalista kerätty tutkimuskysymysten lista. Vaikka metodin teoria oli melko hyvin hallussa, käytännön soveltamisessa ajatustyötä ja ratkaisujen etsimistä oli paljon. Oli ilo huomata, että fenomenografia sopi menetelmäksi hyvin omaan

tutkimusaiheeseen. Fenomenografinen analyysi tapahtuu toisen asteen näkökulmasta, jota tässä työssä edustivat tutkijoiden käsitykset ja tutkimustulokset. Merkityssisältöjen etsiminen ja ensimmäisen kategorian luominen veivät analyysistä eniten aikaa. Tälle tutkielmalle ei ole vertailtavaa aineistoa. Uusi tutkimustulos on se, että tämä tutkimusaihe on taipunut käsiteltäväksi fenomenografisella metodilla ja saatu kombinaatio on uusi.

Kaikissa tutkimuksissa on rajoitteensa, niin tässäkin. Koska tutkimus on ollut laadullinen, tutkija itse on instrumentti, joka on omilla valinnoillaan vaikuttanut tutkimuksen tulokseen, vaikka pyrkimys on ollut objektiivisuuteen. Tehtyjen valintojen lisäksi oman rajoitteensa muodostaa tutkijan tulkinnalliset edellytykset. Kolmannen rajoitteen muodostaa käytetty tutkimusmenetelmä, jota olen pyrkinyt tarkasti noudattamaan. Kaikesta avoimuudesta ja objektiivisuudesta huolimatta tutkimus voi sisältää aukkoja. Mahdolliset virheet ovat luonnollisesti omiani.

7.2 Tutkimuksen tarkoitus ja tavoitteet

Tutkimuksen tarkoituksena oli selvittää tutkijoiden käsityksiä ja tutkimustuloksia algoritmien eettisyydestä vuosina 2016–2018. Tutkijoiden ääni kuuluu ja näkyy tässä työssä. Tämä on tapahtunut toisaalta luodessa ohjaavaa analyysikehikkoa, joka perustuu täysin tutkimusmateriaalista nouseviin teemoihin sekä jokaisesta dokumentista tehtyyn miellekarttaan, johon on kasattu relevantein tietämys tämän tutkimuksen kannalta. Toisaalta tutkijoiden näkemykset ovat saaneet runsaasti kuulua ja näkyä analyysivaiheessa käsitysesimerkkeinä ja runsaina viitteinä.

Työn tavoitteena oli tuoda esille algoritmeihin liitettyjä eettisiä haasteita. Tietokone-etiikan keskeiset kysymykset heijastuvat algoritmien eettisissä haasteissa. Mooren (1985: 266) mukaan eettiset kysymykset nousevat toimintatavan tyhjiöstä, koska emme tiedä miten teknologiaa tulisi käyttää. Tämä käsitys on yhtäpitävä algoritmien toimintatavan kanssa. Meneillään olevassa kehityksessä yritykset ovat saaneet määritellä algoritmiset toimintatavat (Martin 2018). Johnsonin (2001) eettisen arvioinnin perusteet ovat myös yhdenmukaiset algoritmisen toiminnan kanssa. Hän toteaa, että on tiedettävä mikä

arvioinnin kohteena oleva teknologia on, koska muutoin ei voida määrittää yksityiskohtia tai sääntöjä sen käyttämiseen (Johnson 2001: 9). Tämähän on se tilanne, mikä algoritmien etiikan kohdalla vallitsee. Teknologiasta ja vaikutuksista ei tiedetä riittävästi. Myös Masonin (2017) näkemys on keskeinen algoritmien toiminnassa. Sen mukaan yksityisyyttä uhkaa teknologian kasvu ja lisääntynyt tiedon arvo päätöksenteossa (Mason 2017: 1). Kaikkia näitä näkökulmia on havainnollistettu tässä työssä.

7.3 Tulosten merkitys ja jatkotutkimusaiheet

Tämän tutkimuksen tieteellinen merkitys on kolmenlainen. Ensimmäiseksi se vahvistaa sitä käsitystä, että algoritmisella päätöksenteolla on haitallisia vaikutuksia eikä prosessi suinkaan aina ole virheetön. Toiseksi se osoittaa, että lisää tutkimusta tarvitaan algoritmisen eettisen kokonaismallin luomiseksi. Kolmanneksi tutkimus näyttää, että algoritmien etiikan tutkimuksessa voidaan käyttää monenlaisia metodeja.

Työn suurin käytännön merkitys on siinä, että se kiinnittää huomiota siihen yhtälöön, jossa teknologia plus ihminen ei ole arvoneutraali. Syvästä algoritmisen neutraaliuuden Ruususen unesta on herättävä ja huomattava millaisia vaikutuksia algoritmeilla tosiasiallisesti on. Digitaalisuuden halutaan läpäisevän koko yhteiskunta, ja niin se jo nyt suurelta osin tekee, mutta tulisi pitää huoli myös siitä, että kaikki ymmärtävät niitä lainalaisuuksia, joita digitaalisuuteen liittyy. Tarvitaan paitsi tietotekniikkataitoja, niin ymmärrystä mitä digitaalisuus tarkoittaa, mikä on sen käsitteellinen merkitys. Tämän asian rinnalle tulisi saada eettinen ymmärrys kaikille. Käytännön työssä tätä voi hyödyntää esimerkiksi opetuksessa.

Jatkotutkimusaiheina tulisi selvittää miten yritykset saadaan motivoitua eettiseen järjestelmäkehitykseen, mitkä ovat reilun datan eettiset vaatimukset ja miten algoritmisesta päätöksenteosta saadaan eettisesti kestävä. Koska eri tieteenalojen tutkijoiden käsitykset saattavat poiketa toisistaan, tulisi algoritmista tehdä laaja käsiteanalyysi yhteisen keskustelun pohjaksi. Nyt olisi oivallinen aika aloittaa algoritmien aiheuttamista vaikutuksista

pitkittäistutkimus. Olisi myös houkuttelevaa tutkia algoritmeja vasten Floridin luomaa tetralogista eettistä kehystä.

7.4 Tämän tutkimuksen reliabiliteetti ja validiteetti

Tätä työtä arvioidaan fenomenografian luotettavuuskriteerien (Ahonen 2005: 130) kautta sekä Larssonin (2005) heuristisen arvon, pragmaattisen kriteerin, konsistenssin ja empiirisen ankkuroinnin avulla (Häkkinen 1996: 44–45). Fenomenografian luotettavuuskriteerit Ahosen (2005: 130) mukaan ovat aineiston ja kategorioiden aitous ja relevanssi. Aineiston aitous toteutuu, jos aineisto koskee tutkijan ja tutkittavien kannalta samaa asiaa ja aineiston relevanssi toteutuu, jos aineisto on relevanttia tutkimuksen kannalta. Aineiston aitous ja relevanssi toteutuvat, koska kaksi tärkeintä mukaanottokriteeriä olivat 1) aineisto on relevantti tutkimuskysymyksen kannalta ja 2) aineistossa käsitelty eettinen kysymys kattaa jonkin algoritmin etiikan kysymyksen. Tutkijan ja tutkittavan materiaalin välinen ymmärrys koskee samaa asiaa. Kategorioiden aitous toteutuu, jos kategoriat vastaavat tutkittavien tarkoittamia merkityksiä ja kategorioiden aitous toteutuu, jos kategoriat ovat relevantteja tutkimuksen teorian kannalta. Kategoriat vastaavat tutkimusdokumenttien tarkoittamia merkityksiä. Kategoriat ovat myös relevantteja tutkimuksen teorian eli IT-etiikan kannalta, sillä tutkimuslinssi on eettinen. Tämä tutkimus läpäisee esitetyt fenomenografiset luotettavuuskriteerit.

Seuraavaksi arvioidaan tutkimusta Larssonin (2005) heuristisen arvon, pragmaattisen kriteerin, konsistenssin ja empiirisen ankkuroinnin avulla (Häkkinen 1996: 44–45). Heuristisen arvon toteutuminen tarvitsee työn tarkastelua tuoreesta perspektiivistä tai ilmiöiden uudenlaista käsitteellistämistä. Algoritmien etiikan tutkimusta on lähdetty rakentamaan sikäli uudesta näkökulmasta, että on luotu kokonainen algoritminen tulosavaruus. Jotta tähän on päästy, on rakennettu kuvaa pala palalta aina lähtien algoritmien perusominaisuuksista.

Pragmaattinen kriteeri toteutuu tutkimuksen käytännön merkityksen kautta. Tutkimuksen käytännön merkitys nojaa IT-etiikkaan ja siihen, että teknologialla on eettisiä vaikutuksia.

Tietotekniikkataitojen yhteydessä olisi hyvä opettaa myös syvempää digitaalisuuden merkitystä ja etiikkaa.

Konsistenssi tarkastelee kategorioiden sisäisen logiikan pätevyyttä (Häkkinen 1996: 45). Käytetyt ilmaukset sopivat kategorioihin. Luodut kategoriat pystyy liittämään teoreettiseen käsitteistöön ja niitä on myös kulloisessakin kohdassa avattu. Lukija pystyy seuraamaan miten luokittelu on tehty. Tätä taustaa vasten kategorioiden sisäinen logiikka on pätevä.

Empiirinen ankkurointi edustaa tutkimusprosessin kuvauksen avaamista ja aineisto-esimerkkejä lukijalle. Olen pyrkinyt mahdollisimman avoimeen analyysipolkuun. Kaikki fenomenografisen tutkimusprosessin vaiheet on dokumentoitu. Kummankin kategorian muodostuminen on selvitetty huolellisesti esimerkkejä ja perusteluita esittäen. Kuvat havainnollistavat asiaa. Synteesi on keskittynyt kuvaamaan kaikissa luokissa esiintyviä yhdistäviä tekijöitä. Tutkimuksen aikana tapahtui aiheen kyllästymistä eli saavutettiin saturaatio. On oletettavissa, että muista tietotekniikan etiikkaa käsittelevistä tutkimuksista löydetään yhdenmukaisuutta tämän tutkimuksen löydösten kanssa. Edellä sanotun perusteella tutkielma täyttää tutkimukselle asetetut kriteerit.

LÄHDELUETTELO

- Ahonen, Sirkka (1995). Fenomenografinen tutkimus. Teoksessa: *Laadullisen tutkimuksen työtapoja*, 114–158. Syrjälä, L., Ahonen, S., Syrjäläinen, E. & Saari, S. Helsinki: Kirjayhtymä. ISBN 951-26-3948-3.
- Ananny, Mike (2016). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology & Human Values* Vol. 4 (1), 93–117.
- Association for Computing Machinery (2018). *ACM Code of Ethics and Professional Conduct*. [online]. [21.12.2018]. Saatavissa: <https://www.acm.org/code-of-ethics>.
- Association for Computing Machinery US Public Policy Council (USACM) (2017). *Statement on Algorithmic Transparency and Accountability*. [pdf]. [11.11.2018]. Saatavissa: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.
- Bucher, Taina (2018). *If...Then. Algorithmic Power and Politics*. Oxford: Oxford University Press. ISBN 978-0-1904-9302-8.
- Bynum, Terrell W. (2001). Computer ethics: Its birth and its future. *Ethics and Information Technology* 3, 109–112.
- Bynum, Terrell W. (2004). Ethical Decision-Making and Case Analysis in Computer Ethics. Teoksessa: *Computer Ethics and Professional Responsibility*, 60–85. (Edit.) Bynum, T. W. & Rogerson S. Oxford: Blackwell Publishing. ISBN 1-8 5 5 54-844-5.
- Bynum, Terrell (2015). Computer and Information Ethics. *Stanford Encyclopedia of Philosophy Archive. Winter 2016 Edition*. [online]. [6.11.2018]. Saatavissa: <https://plato.stanford.edu/archives/win2016/entries/ethics-computer/>

- Cheney-Lippold, John (2017). *We are data. Algorithms and the making of our digital selves*. New York: New York University Press. ISBN 978-1-4798-5759-3.
- Cormen, T.H, Leiserson, C.E, Rivest, R.L & Stein, C. (2009). *Introduction to Algorithms*. [e-kirja]. MIT Press.
- Crawford, Kate (2016). Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics. *Science, Technology & Human Values* Vol. 41 (1), 77–92.
- Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S-M., Morison, J., Murphy, M. H., O’Brolchain, N., Schafer, B. & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society* July–December 1–21.
- Dourish, Paul (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society* July–December 1–11.
- Esguerra, Richard (2009). Google CEO Eric Schmidt Dismisses the Importance of Privacy. *Electronic Frontier Foundation*. 10.12.2009. [online]. [23.02.2019]. Saatavissa: <https://www.eff.org/deeplinks/2009/12/google-ceo-eric-schmidt-dismisses-privacy>.
- Fieser, James (2018). Ethics. *Internet Encyclopedia of Philosophy. A Peer-Reviewed Academic Resource*. [online]. [28.10.2018]. Saatavissa: <https://www.iep.utm.edu/ethics/#H3>.
- Fisher, Andrew (2011). *Metaethics. An introduction*. [e-kirja]. Acumen.
- Floridi, Luciano (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology* 1, 37–56.

- Floridi, Luciano (2010). *Information. A Very Short Introduction*. [e-kirja]. Oxford University Press.
- Halava, Vesa (2016). *Algoritminen matematiikka*. Luentomoniste. Turun yliopisto. Matematiikan laitos. [pdf]. [11.11.2018]. Saatavissa: <https://www.utu.fi/fi/yksikot/sci/yksikot/mattil/opiskelu/kurssit/Documents/AlgMat2017.pdf>.
- Hassan, N. R., Mingers, J. & Stahl, B. (2018). Philosophy and information systems: where are we and where should we go? *European Journal of Information Systems* Vol. 27 (3), 263–277.
- Hill, Robin K. (2016). What an Algorithm Is. *Philosophy & Technology* Vol.29 (1), 35–59.
- Huusko, M. & Paloniemi, S. (2006). Fenomenografia laadullisena tutkimussuuntauksena kasvatustieteissä. *Kasvatus* 2, 162–173.
- Häkkinen, Kirsti (1996). *Fenomenografisen tutkimuksen juuria etsimässä. Teoreettinen katsaus fenomenografisen tutkimuksen lähtökohtiin*. Opetuksen perusteita ja käytänteitä 21. Opettajankoulutuslaitos. Jyväskylä: Jyväskylän yliopisto. ISBN 951-34-0742-X.
- IEEE Computer Society (1999). *Software Engineering Code of Ethics*. [online]. [21.12.2018]. Saatavissa: <https://www.computer.org/web/education/code-of-ethics>.
- Itä-Suomen avoin yliopisto (2018). *Tiedon määrittelyä*. [online]. [28.10.2018]. Saatavissa: <https://www.uef.fi/fi/web/aducate/tiedon-maarittelya>.
- Johnson, Bobbie (2010). Privacy no longer a social norm, says Facebook founder. *The Guardian*. 11.01.2010. [online]. [23.02.2019]. Saatavissa: <https://www.theguardian.com/technology/2010/jan/11/facebook-privacy>.

- Johnson, Deborah G. (2001). *Computer Ethics*. International Edition. 3rd edition. Upper Saddle River: Pearson Prentice Hall. ISBN 0-13-124788-3.
- Kananen, Jorma (2015). *Opinnäytetyön kirjoittajan opas. Näin kirjoitan opinnäytetyön tai pro gradun alusta loppuun*. Jyväskylän ammattikorkeakoulun julkaisuja 202. Jyväskylä: JAMK. ISBN 978-951-830-387-2.
- Kettunen, J. & Tynjälä, P. (2018). Applying phenomenography in guidance and counseling research. *British Journal of Guidance & Counselling* Vol.46 (1), 1–11.
- Kitchin, Rob (2017). Thinking critically about and researching algorithms. *Information, Communication & Society* Vol. 20 (1), 14–29.
- Kotimaisten kielten keskus. (2018). *Kielitoimiston sanakirja*. [online]. [25.10.2018]. Saatavissa: <https://www.kielitoimistonsanakirja.fi/>.
- Kraemer, F., van Overveld, K. & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology* Vol. 13 (3), 251–260.
- Laaksonen, Antti (2009). Algoritmien aakkoset: Osa 1 Aikavaativuus ja tilavaativuus. *Ohjelmointiputka*. Oppaat. [online]. [28.12.2018]. Saatavissa: https://www.ohjelmointiputka.net/opaat/opas.php?tunnus=alg_1.
- Markie, Peter (2017). Rationalism vs. Empiricism. *Stanford Encyclopedia of Philosophy*. [online]. [6.11.2018]. Saatavissa: <https://plato.stanford.edu/entries/rationalism-empiricism/>
- Martin, Kirsten (2018). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics* June 1–16.

- Mason, Richard O. (1986). Four Ethical Issues of the Information Age. *MIS Quarterly: Management Information Systems* Vol.10 (1), 5–12
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society* Vol. 3 (2), 1–21.
- Moor, James H. (1979). Are There Decisions Computers Should Never Make? *Nature and System* 1, 217–229.
- Moor, James H. (1985). What Is Computer Ethics? *Metaphilosophy* Vol.16 (4), 266–275.
- Mäkinen, Olli (2006). *Tutkimusetiikan ABC*. Helsinki: Tammi. ISBN-13: 978-951-26-5595-3.
- O’Neill, Cathy (2016). *Weapons of math destruction. How big data increases inequality and threatens democracy*. New York: Broadway Books. ISBN 978-0-553-41883-5.
- Oksanen, M., Launis, V. & Sajama, S. (2010). *Etiikan lukemisto*. Helsinki: Gaudeamus. ISBN 978-952-495-156-2.
- Paßmann, J. & Boersma, A. (2017). Unknowing Algorithms. On Transparency of Unopenable Black Boxes. Teoksessa: *The Datafied Society. Studying Culture through Data* ,125-138. (Edit.) Schäfer, M.T. & van Es, K. Amsterdam: Amsterdam University. ISBN 978-94-6298-136- 2. Saatavilla myös e-kirjana: <http://www.oopen.org/search?identifier=624771>.
- Pietarinen, Juhani (2015). *Etiikka*. [online]. [25.10.2018]. Saatavissa: <http://filosofia.fi/node/6985#Normatiivisen%20etiikan%20suuntauksia>.
- Pietarinen, J. & Poutanen, S. (1998). *Etiikan teorioita*. 5. painos. Helsinki: Gaudeamus. ISBN 978-951-662-741-3.

- Perseus Digital Library. (2018). *Greek Dictionary*. [online]. [25.10.2018]. Saatavissa: <http://www.perseus.tufts.edu/hopper/resolveform>.
- Rapaport, William, J. (2016). *Philosophy of Computer Science*. [pdf]. [29.10.2018]. Saatavissa: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.851.9153&rep=rep1&type=pdf>.
- Schröder, Mark (2016). Value Theory. *Stanford Encyclopedia of Philosophy*. [online]. [6.11.2018]. Saatavissa: <https://plato.stanford.edu/entries/value-theory/>.
- Sitra. (2019). *IHAN -hanke*. [online]. [23.03.2019]. Saatavissa: <https://www.sitra.fi/hankkeet/ihan-tekninen-kehittajayhteiso/>.
- Steup, Matthias (2005). Epistemology. *Stanford Encyclopedia of Philosophy*. [online]. [6.11.2018]. Saatavissa: <https://plato.stanford.edu/entries/epistemology/>.
- Sullivan, M. & van Inwagen, P. (2014). Metaphysics. *Stanford Encyclopedia of Philosophy*. [online]. [6.11.2018]. Saatavissa: <https://plato.stanford.edu/entries/metaphysics/>
- Tedre, Matti (2011). Informaatioteknologian ja perinteisten filosofian alojen yhtymäkoh-
tia. Teoksessa: *Informaatioteknologian filosofia*, 65–80. (Toim.) Laakkonen, M.,
Lamminpää, S. & Malaprade, J. Rovaniemi: Lapin yliopistokustannus. ISBN 978-
952-484-384-3.
- Uricchio, William (2017). Data, Culture and the Ambivalence of Algorithms. Teoksessa:
The Datafied Society. Studying Culture through Data, 125-138. (Edit.) Schäfer,
M.T. & van Es, K. Amsterdam: Amsterdam University. ISBN 978-94-6298-136- 2.
Saatavilla myös e-kirjana: <http://www.oapen.org/search?identifier=624771>.

Williams, Brian (2014). Edward Snowden's Motive Revealed: He Can 'Sleep at Night'. *NBCNews* 28.5.2014. [online]. [24.2.2019]. Saatavilla: <https://www.nbcnews.com/feature/edward-snowden-interview/edward-snowdens-motive-revealed-he-can-sleep-night-n116851>.

Zarsky, Tal (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology, & Human Values* Vol. 41(1), 118–132.

Ziewitz, Malte (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology & Human Values* Vol. 41 (1), 3–16.