

Artikkeli

Tunnetekoälyn etiikka toimijuuden valossa – kohteena pedagogisesti orientoitunut ryhmävuorovaikutustutkimus

Eevastiina Gjerstad¹¹ Vastaava kirjoittaja, Haaga-Helia ammattikorkeakoulu

Viittaa: Gjerstad, E. (2023). Tunnetekoälyn etiikka toimijuuden valossa – kohteena pedagogisesti orientoitunut ryhmävuorovaikutustutkimus. *eSignals Research*, 3(1). <http://urn.fi/URN:NBN:fi-fe2023031031088>

Julkaistu: 10.03.2023



Vertaisarvioijat: Virpi-Liisa Kykyri, Jyväskylän yliopisto & Mika Simonen, Helsingin yliopisto.



Copyright: © 2023 by the authors and Haaga-Helia University of Applied Sciences. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC SA) license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Tiivistelmä: Artikkelin ”Tunnetekoälyn etiikka toimijuuden valossa – kohteena pedagogisesti orientoitunut ryhmävuorovaikutustutkimus” valottaa tekoälyn ja aivan erityisesti tunnetekoälyn käytön eettisiä haasteita toimijuutta painottaen. Tarkasteltavana ovat vuosien 2019–2021 välillä tehdyt ryhmävuorovaikutustutkimukset, joihin tunnetekoälyn eettisiä haasteita peilataan. Eettiset haasteet liittyvät käytettävään teknologiaan ja menetelmiin, taustalla oleviin tunneteorioihin, osallistujien eettiseen kohteluun ja huomioimiseen sekä laajempaan ihmis- ja toimijuuskäsitykseen.

Erilaisten tunneteorioiden ja käytettyjen menetelmien paradigmoja ylittävänä ja yhdistävänä teoriana käytetään arviointitunneteorioita. Tekoälyä sovellettaessa on suojeltava ja vahvistettava ihmistoimijuutta. Tekoälyn hyödyntämisessä olisi eduksi löytää yhteisiä eettisiä periaatteita ja normistoja sekä varautua jo ennakolta eettisiin riskeihin. Riittävän laaja eettinen näkökulma mahdollistuu yhdistämällä strukturaalista etiikkaa, yksilöllistä toimijuuden vahvistamista ja yhteisöllistä ”me-tilaisuutta”.

Avainsanat: ryhmävuorovaikutus, etiikka, toimijuus, tekoäly, tunnetekoäly, paradigmat

Abstract: The article ”Ethics of emotional artificial intelligence in the light of the agency of the human being - case pedagogical social group interaction research” sheds light on the ethical challenges of artificial intelligence, especially emotional artificial intelligence focusing on the principle of the human agency. I use my social interaction of groups researches from the years 2019-2021 that applies emotional AI as case studies. Ethical challenges I refer here are related to the technology and methods used, underlying emotional theories, the ethical treatment of participants, and the broader views of human beings and their agency.

The appraisal theories of emotions have transcended different paradigms between the different research methods and theories of emotions. Protecting and strengthening the conditions of human agency in any AI environment is a necessary ethical condition of the ethics of AI. It would be an advantage of setting common ethical principles and norms and preparing ourselves for the ethical risks. Combining structural ethics, strengthening human agency and communal ”we-ness” we might reach the broad ethical view enough.

Keywords: social interaction of groups, ethics, human agency, artificial intelligence, emotional artificial agency, paradigms

1. Johdanto

Keskustelua tekoälystä leimaa usein dikotominen suhtautuminen, jossa tekoäly on nähty joko uhkana ja ihmisten mahdollisena hallitsijana tai kaiken ratkaisevana vastauksena ihmiskunnan kysymyksiin (Pietikäinen & Silvén, 2019, 11–12, 141). Tekoälyn eettiset kysymykset on tuotu esille niin Suomen, Euroopan kuin maailmankin tasolla (mm. Ailisto ja muut, 2018; 2019; A European approach to artificial intelligence, 2023).

Tämä artikkeli käsittelee tekoälyä ja erityisesti erästä tekoälyn osa-aluetta, tunnetekoälyä ja siihen liittyviä eettisiä kysymyksiä. Tunnetekoälyn sisällä artikkeli keskittyy vielä yhteen sen alueeseen eli kasvojentunnistusteknologiaan ja siinä tunnereaktioiden tunnistamiseen. Tunnetekoäly määritellään tässä teknologioina, jotka hyödyntävät ihmisen tunnetiloihin liittyvää affektiivista laskentaa sekä tekniikkoina, jotka havainnoivat, oppivat ja vuorovaikuttavat ihmisten tunteiden kanssa (McStay, 2020, 1). Yksittäinen tutkija kohtaa monia eettisiä haasteita tunnetekoälyä soveltaessaan. Peilaan näitä haasteita ja niihin liittyviä ratkaisuja esimerkinomaisesti itse tekemiäni ryhmävuorovaikutustutkimusten kautta.

Tein vuosina 2019–2021 kaksi tutkimusta Haaga-Helian ammatillisessa opettajakorkeakoulussa, jossa hyödynsin tunnetekoälyä käyttävää labia. Koska olen esitellyt näiden tutkimushankkeiden varsinaisia empiirisiä tutkimustuloksia muualla (Gjerstad, 2021; 2022; Aunimo ja muut, 2021; Pedaforum, 2022), en esittelen niitä uudestaan tässä, vaan keskityn tässä artikkelissa näitä tutkimuksia sivuaviin eettisiin lähtökohtiin ja kysymyksiin. Esittelen kuitenkin tutkimushankkeiden lähtökohdat ja tutkimusmenetelmät ohessa.

VALAA (vaikuttava ja laadukas ohjaus opettajan työssä) oli ohjausvuorovaikutusta ja tunteita vuorovaikutuslabissa tutkiva tutkimushanke, johon osallistui kolme opettajankouluttajaa ohjaajan roolissa sekä heidän pienryhmänsä (neljä opettajaopiskelijaa per ryhmä) (N=16). Pienryhmät kokoontuivat keskustelemaan etukäteen määritellyistä pedagogisista, opintoihin liittyvistä teemoista, jotka olivat opetusharjoittelu ja osaamisen arviointi. Tutkimuksessa kunkin ryhmän käymä ryhmäohjausvuorovaikutuskeskustelu videoitiin ja ryhmä katsoi sen jälkeenpäin niin sanotussa reflektiovideoinnissa (myös tämä videoitiin). Tällaisia kierroksia oli yhtä ryhmää lukuun ottamatta kaksi. Kokonaisuudessaan videoiden määrä oli 11. VALAA:n päätutkimuskysymykset olivat: Mitä ohjausvuorovaikutustilanteessa tapahtuu tunnereaktioiden, tunnekokemusten sekä dialogisuus- ja vuorovaikutuskokemusten tasoilla? sekä Miten vuorovaikutuksen hallinta näyttäytyy ryhmäohjaustilanteissa?, Kuka hallitsee vuorovaikutustilanteita ja miten? sekä Mitä merkitystä sillä on vuorovaikutustilanteiden kannalta?

SOEVE (sosio-emotionaaliset ulottuvuudet pedagogisessa vertaisryhmäoppimisessa) oli oppimista ja tunteita sosiaalisessa vertaisryhmävuorovaikutuksessa tutkiva hanke, johon osallistui yksi opettajaopiskelijoista koostuva vertaisryhmä ja osittain heidän opettajakouluttajansa ohjaajan roolissa (N=5). SOEVE-tutkimus perustui opiskelijoiden vertaisryhmän kolmen kokoontumisen videoinneille, joista yhdessä myös heidän ohjaajansa oli mukana. SOEVE:ssa etsin vastauksia siihen, miten vertaisryhmä muokkaa oppimiskokemusta sekä miten sosiaaliset ja emotionaaliset ulottuvuudet vertaisryhmässä vaikuttavat ryhmäprosesseihin ja oppimiseen.

Tutkimusmenetelmät molemmissa tutkimuksissa koostuivat labissa tapahtuvasta biometrisestä ja fysiologisesta tunnereaktioiden mittaamisesta (tunnereaktioita

mittaava kasvojenilmetunnistusanalyysi sekä osittain ihon sähkönjohtavuuden mittaaminen, josta data jäi kuitenkin liian vähäiseksi), osallistujien omien tunne-, dialogisuus-, ryhmäidentiteetti- oppimis-, ryhmävuorovaikutus- ja/tai reflektiokokemusten tutkimisesta kyselylomakkeiden, avovastausten, niin sanotun visuaalisen dialogisuusjatkumon sekä stimulated recall -tyyppisten reflektiovideointien keinoin sekä itse muotoilemani vuorovaikutuksen hallinta (niin sanotun dominanssiteoria-idean) kehikon ja Järvelän ja kollegoiden (2016) oppimisen säätelyn tasot -analyysien kautta. Mittaristot ja kyselyt rakensin pitkälti jo olemassa olevien mittaristojen ja tutkimustulosten pohjalta, mutta suomensin, muokkasin ja yhdistin mittaristoja myös uusiksi mittaristoiksi.¹

Biometriikka viittaa tieteeseen, jossa pyritään arvioimaan ihmisen identiteettiä fysiologisten, kemiallisten tai käyttäytymiseen liittyvien (yksilöllisten) piirteiden perusteella (Jain & Ross, 2008, 1). Tällaisia piirteitä ovat esimerkiksi kasvojen piirteet, sormenjäljet, käden piirteet, silmän iiris, DNA, sydämen syke tai ääni (Sutrop & Laas-Mikko, 2012, 21). Tätä niin sanottua biometriikkaa käytetään, jotta ymmärretään paremmin mielen, aivojen ja kehon välistä vuorovaikutusta (Human Behavior, 2017). Tunnetekoäly hyödyntää käytännössä usein biometrisiä tutkimusmenetelmiä, jonka vuoksi nämä molemmat käsitteet ovat keskiössä tunnetekoälystä puhuttaessa.

Artikkelini pyrkii vastaamaan seuraaviin kysymyksiin: Miten ihmisten toimijuus pitäisi huomioida tunnetekoälyä soveltavassa tutkimuksessa? Entä mitä eettisiä periaatteita meillä olisi oltava tunnetekoällyn käytön suhteen, jotta se vahvistaisi ihmisen toimijuutta?

Toisaalta on tärkeää huomata, että vaikka tällaisia eettisiä periaatteita voitaisiin formuloida, kaikki toimijat, kuten valtiot, eivät piittaa näistä periaatteista eivätkä siis myöskään vahvista sen kautta ihmisten toimijuutta. Esimerkiksi kasvojentunnistusteknologian suurimpia eettisiä haasteita on, miten voidaan välttää sen avulla tapahtuva kontrollointi, kriminalisointi tai syrjintä. Suostumuksesta ei ole useinkaan välitetty, eikä pelkkä suostumuskaan takaa teknologian eettistä käyttöä. Teknologiaa voidaan käyttää ihmisoikeuksien vastaisilla tavoilla. Ei pidä myöskään olettaa, että tutkijat noudattavat väistämättä ihmisoikeuksia, vaan vaikkapa Kiinassa osa tekoälytutkijoista on sortotoimien kannalla. (van Noorden, 2020.)

Maailman talousfoorumi onkin nostanut esille muun muassa tekoölyyn liittyvät mahdolliset seuraukset epätasa-arvon, puolueellisuuden ja ei-toivottujen sivuvaikutusten näkökulmasta. Yksi keskeinen kysymys koskee sitä, miten voidaan suojata eettisiä periaatteita noudattaen luodut tekoälyjärjestelmät pahantahtoislta toimijoilta. (Bossmann, 2016.)

Tämä on johdantoluontoinen artikkeli tunnetunnetekoällyn eettisiin haasteisiin erityisesti toimijuuden ja ryhmävuorovaikutustutkimuksen näkökulmista. Tarkastelen aihetta sosiaalieettisesti painottuneen toimijuusnäkökulman sekä kasvatustieteellisesti ja pedagogisesti² painottuneen ryhmävuorovaikutusnäkökulman kautta. Kehikkoa täydentävät empiiriset case-havaintoni omista tutkimuksistani.

2. Ihmisen toimijuus ja tekoäly

Euroopan komissio on kiinnittänyt huomiota tekoällyn käytössä yksityisyyden, datan hallinnan, läpinäkyvyyden, moninaisuuden ja sosiaalisen ja ekologisen hyvinvoinnin kysymyksiin (Ethics guidelines for trustworthy AI, 2019). Edellä mainitut kysymykset ovat tärkeitä, ja useita niitä yhdistää vielä perimmäisempi

kysymys ihmisen toimijuudesta ja siihen liittyvästä moraalisesta vastuullisuudesta.

Toimijuudella on erilaisia määritelmiä eri tieteenaloilla, mutta yleisesti sillä viitataan aktiivisuuteen, aloitteellisuuteen, osallisuuteen sekä jonkinlaiseen hallintaan ja vaikutusmahdollisuuksiin (Eteläpelto ja muut, 2017, 7). Toimija on Hallamaan (2017, 17) määritelmän mukaan ihminen, joka pystyy asettamaan päämääriä, ymmärtää ja osaa arvioida syys-seuraussuhteita ja kykenee toimintansa kautta tavoittelemaan asettamiaan päämääriä. Toimijuuteen voidaan katsoa liittyvän käsitykset kyvykkyydestä, intentionaalisuudesta ja rationaalisuudesta sekä vastuullisuudesta (Hallamaa & Kalliokoski, 2020).

Toimijuus on aina myös relationaalista eli suhteessa toisiin tapahtuvaa. Samalla myös ymmärrys omasta itsestä muokkautuu. Vuorovaikutuskin on tavallaan 'vuorotellen vaikuttamista' eli molemmat osapuolet vaikuttavat toisiinsa. Tähän liittyvät eettiset kysymykset oikeuksista, velvollisuuksista, vapaudesta ja vastuusta. (Mellenius, 2013, 80, 91.) Nämä kysymykset puolestaan ovat jo osa toimijuuden *eettistä* ja *eksistentiaalista* puolta, joihin kuuluvat myös kysymykset elämän tarkoituksesta ja merkityksestä, kuolemasta, osattomuudesta ja voimaantumisesta. (ks. Vanhalakka-Ruoho, 2014, 194–199.)

Moraalinen toimijuus viittaa ihmisiin ja heidän moraaliseen vastuullisuuteen teoistaan, heidän kyvykkyyteensä erottaa oikea väärästä ja hyödyntää moraalista harkintaansa toiminnassaan (Brey, 2014, 126). Moraalinen toimijuus perustuu ajatukselle, että toimija on vapaa toimimaan intentionaalisesti ja vapaaehtoisesti, jolloin häntä voidaan pitää vastuussa omasta toiminnastaan. Tämä ajatus on sisäenrakennettuna oikeastaan kaikissa moraaliteorioissa. (Johnson, 2006, 198.) Moraalisen toimijuuden myöntäminen on tärkeää, sillä se velvoittaa myös yhteisön jäsenet myöntämään toistensa moraalisen toimijuuden (Brožek & Janik, 2019, 103).

Moraalista toimijuutta on mahdotonta paeta, sillä elämme täysin normatiivisessa maailmassa, jossa ilmenee oikean ja väärän sekä hyvän ja pahan välisiä erotteluja ja niille annettuja merkityksiä (Pihlström, 2021, 211). Kasvatuksessa ihmisen sivistäminen vastuuntunnon ja arvotajunnan omaavaksi moraalisubjektiksi hahmotetaan kasvatuksen tärkeimmiksi tehtäviksi. Tästä seuraa, että kasvatusta pyrkii kehittämään kasvatettavan moraaliresursseja ja eettistä toimintakykyä. (Väri, 2018, 123.)

Parhaimmillaan kasvatusta ja kulttuuri välittävät ihmiselle sellaisia hyvän elämän malleja ja normeja, jotka edistävät hänen moraalisen toimijuuden kykyjään (Visala, 2021, 159–160) ja mahdollistavat sekä yhteisen hyvän että yksilön persoonan kehittymisen (Gjerstad, 2011, 84–89). Jos ihmisyys määritellään toimijuuden ja moraalisen toimijuuden kautta näin laajasti ja syvällisesti, se asettaa tekoälyn tuottamalle lisäarvoa suurina haasteita.

Tekoälyyn perustuva ympäristö saattaa itsessään jo muokata moraalisuuden peruseriaatteita kuten ihmisen arvokkuutta ja moraalista vastuullisuutta (Hallamaa & Kalliokoski, 2020, 54–55). Algoritmien ja niiden käyttötapojen on katsottu sisältävän kätkeytyä valtaa, jonka avulla ne voivat vähitellen muuttaa ihmisenä olemisen sääntöjä (Fry, 2018, 15). Biometrisiä menetelmiä on kritisoitu "algoritmista hallinnasta", jonka avulla luodaan ja ylläpidetään valvontaa (Crampton, 2019).

Laajasti tunnetun kantilaisen periaatteen mukaan ihmistä ei saa koskaan kohdella välineenä vaan aina päämääränä sinänsä, mutta teknologiat on jo lähtökohtaisesti suunniteltu toimimaan välineinä johonkin (Hallamaa & Kalliokoski, 2020, 54–55). Ironisesti kukaan ihminen ei kuitenkaan kykene täysin toimimaan kantilaisen moraalisen periaatteen mukaisesti, vaan se voi olla koneelle joissakin tapauksissa jopa helpompaa kuin ihmiselle. Tämä perustuu siihen, että ihmisen kognitio ja motivaatio eivät riitä kantilaisen moraalisen periaatteen johdonmukaiseen toteuttamiseen. Kone kykenee huomioimaan nopeasti valtavan määrän moraalisia sääntöjä ja niihin liittyviä monimutkaisia systeemejä sekä suorittamaan niihin liittyviä tehtäviä. (Brožek & Janik, 2019, 102.)

Lisäksi kone voidaan ohjelmoimaan siten, että se seuraa vain tiettyjä sääntöjä, kun puolestaan ihmisen on vaikeampi ulkoistaa tiettyjä asioita päätöksenteon ulkopuolelle. Toisaalta kone ei välttämättä kykene havaitsemaan ja keräämään kaikkea relevanttia tietoa ympäristöstä, jota moraalisiin päätöksiin tarvitaan. (Brožek & Janik, 2019, 102.) Ihmisen moraaliseen päätöksentekoon voivat vaikuttaa esimerkiksi yksittäisen ihmisen tarpeet ja kärsimykset, jolloin halukkuus poikkeuksen tekemiseen formaaliin sääntöön tietyn yksilön auttamiseksi voi kasvaa (Galinsky ja muut, 2011, 21–24). Tämä voi olla tilanteesta riippuen hyöty tai haitta.

Lähtökohtaisesti siis näyttäisi siltä, että tekoäly on tietyissä asioissa ylivertainen ihmiseen verrattuna, mutta toimijuus ja moraalinen toimijuus sisältävät ulottuvuuksia, joita tekoälyyn ei sisälly. Kuitenkin sen käyttö voi lisätä ihmisen toimijuutta – vai voiko?

3. Tunnetekoälyn toiminta ja siihen liittyviä eettisiä huomioita

Edellisessä alaluvussa tarkastelin toimijuutta laajemmasta tekoälyn näkökulmasta. Tässä alaluvussa siirryn tarkastelemaan tarkemmin tunnetekoälyn toimintaperiaatteita ja siihen liittyviä eettisiä kysymyksiä. Koska olen hyödyntänyt omissa tutkimuksissani tunnetekoälystä nimenomaan kasvojentunnistus- / tunnereaktioiden tunnistamisteknologiaa, keskityn lähinnä sen tarkasteluun osana tunnetekoälyä. Kun sovelletaan tunnetekoälyä tutkimuksessa, on huomioitava sekä tutkimuksen tekoon että tutkimuksen kohteena olevien kohteluun liittyvät eettiset seikat. Tunnetekoäly, kuten yleensäkin teknologiat, saatetaan ottaa ”annettuna”. Tunnetekoäly sovelluksineen pohjautuu kuitenkin aina johonkin tunneteoriantai sen sovellukseen, jotka itsessään ovat ristiriitaisia tai ainakin monimerkityksisiä.

Käytännössä kaikki kasvonilmeisiin liittyvien tunteiden tunnistusjärjestelmät pohjautuvat tavalla tai toisella Ekmanin ja Friesenin kehittämään koodausjärjestelmään (FACS eli Facial Action Coding System) (ks. esim. Ekman, 1992; Stöckli ja muut, 2018, 1447–1448³). FACS on ollut laajimpia kasvonilmeiden kehikkoja jo yli 30 vuoden ajan. Kasvonilmeiden koodauksen automatisointi on nopeuttanut ja monipuolistanut siihen perustuvaa tutkimusta. (Martinez ja muut, 2019.) Näin ollen aikaisemmin ihmisten tekemät koodaukset kasvonilmeistä ovat korvautuneet koneiden opettamisella tekemään saman moninkertaisella nopeudella.

Esimerkiksi iMotionsin AFFDEX-algoritmi ottaa ihmiskasvoista 25 kuvaa sekunnissa ja tekee analyysin salamannopeasti. On itsestään selvää, että ihmisen kognitiivinen kapasiteetti ei riitä samaan. Toisaalta ihminen on opettanut koneelle periaatteet, joiden pohjalta analyysi tehdään. Algoritmit tunnistavat tunnereaktioita aina tietyllä todennäköisyydellä. Käyttämäni iMotion-ohjelmiston Affectivan Affdex-algoritmin lisäksi kasvonilmeisiin perustuvia tunnejärjestelmiä

ovat muun muassa FaceVideo, Facet, Cognitive Services: Face, EmotionalTracking, EmotionRecognition, FaceReader ja Face Analysis (ks. Dupré ja muut, 2020).

Myös omista tutkimuksistani käyttämäni vuorovaikutuslabin tunnetekoäly eli iMotions -ohjelmiston AFFDEX-algoritmi pohjautuu FACS:iin. AFFDEX-algoritmi kykenee tunnistamaan ihmisten oletettuja tunnereaktioita kasvojenilmeiden analyysin (facial expression analysis) avulla (Poria ja muut, 2017, 100). Se hyödyntää kasvojen piirteitä ja soveltaa psykologisia teorioita ja tilastotieteellisiä menetelmiä, joiden kautta kone opetetaan tunnistamaan tunteita. (Stöckli ja muut, 2018, 1447.)

Konkreettisesti analyysi eteni siten, että tein kasvojenilmetunnistusanalyysin yhdestä osallistujasta suoraan jo videointitilanteessa ja jälkikäteen muista osallistujista. Analysoin biometrisen datan ensin koneen omilla analyysiohjelmilla, jonka jälkeen vertasin biometristä tietoa vuorovaikutuksen sisältöön ja tapahtumiin ja suhteessa sekä verbaaliseen että non-verbaaliseen kommunikaatioon. Kasvojentunnistusanalyysistä tehdyssä yhteenvedossa huomioin pääosin vain ne kohdat, joissa Hz on 25 tai enemmän (eli kasvoista on otettu ainakin 25 kuvaa sekunnissa) (tämä on oletettu taso AFFDEX-algoritmissa, jotta tuloksia voidaan pitää pätevinä).

Kynnysarvon asettamisen avulla voidaan määrittää koska tietty tunne otetaan huomioon ja erotetaan toisista tunteista sekä neutraaleista kasvoista (Calvo ja muut, 2018, 2). Tein analyysit todennäköisyystasolla 50 tai 70, mikä vastaa kohtalaista tai korkeaa evidenssiä (ks. esim. Haidt ja Keltner, 1999, 238). Koska kasvojentunnistusvideota on yhteensä 20 ihmisestä ja jokaisesta monen tunnin ajalta, en ole ottanut analyysieihin mukaan kaikkea, vaan validoinut keskeisiä huomioita vuorovaikutuksesta. Tilastotieteellistä analyysia on vain pienestä osasta aineistoa. Sen sijaan olen tarkastellut erityisesti niin sanottuja tunnepiikkejä eli vuorovaikutuksen kohtia, joissa tunnereaktioita ilmenee erityisen runsaasti. Koska olen käyttänyt biometriikkaa tutkimuksistani muiden tutkimusmenetelmien rinnalla, datan käsittely on erilaista kuin puhtaasti luonnontieteellisissä biometrisissä tutkimuksissa, joissa koko tutkimusasetelma perustuu niihin.

FACS taas perustuu *erillisten tunteiden teoriaan*, jota kutsutaan myös *perustunteiden teoriaksi* tai diskreetiksi *tunneteoriaksi*. Ajatuksena tässä teoriassa on, että perustunteet ovat diskreettejä eli toisistaan riippumattomia, lajinkehityksessä syntyneitä, yleismaailmallisia tunteita, joihin liittyvät kasvonilmeet tunnistetaan lähes kaikkialla samalla tavalla. Tämän mukaisesti uskotaan, että jopa tekoäly kykenee erottelemaan eri tunteet toisistaan, koska eri tunteiden sisältämä tieto on niin erilaista. Toki kulttuurien ”murteet” sekä tunteisiin liittyvät esiintymissäännöt vaikuttavat siihen, miten näitä ilmeitä esitetään ja tulkitaan. (Nummenmaa, 2010, 33–41, 89–90.)

FACS-luokittelujärjestelmää on kuitenkin myös kritisoitu. Se ei kykene huomioimaan kontekstiin liittyvää tietoa tunteiden tunnistamisessa (ks. Stöckli ja muut, 2018, 1447–1448). Kasvonilmeiden avulla voidaan kommunikoida eri asioita eri konteksteissa, eivätkä läheskään kaikki ihmisten tunnetilat mahdu muutamaankin perustunteeseen (Krywko, 2017). Samallakin kasvonilmeellä ja siitä tulkitulla tunnereaktiolla voi olla eri funktioita (ks. Darwiche ja muut, 2008), joita täytyisi tutkimuksessa selvittää. Esimerkiksi Ruisrockissa kasvojentunnistuksessa käytetty ohjelmisto ei pystynyt erottelemaan ilosta seuraavaa itkua surullisesta itkusta (Rissanen, 2017). Nämä seikat liittyvät usein nimenomaan ihmisen toimijuuteen, jota tunnetekoäly ei tässä kykene havainnoimaan tai ainakaan tulkitsemaan.

Lisäksi kasvojentunnistusohjelmistojen ja etenkin tunteidentunnistusohjelmistojen validaatiossa on vielä parantamisen varaa, joten niiden käyttö yksinään tai ehdottomien johtopäätösten teko niiden varassa ei ole eettisesti kestävä. Tunnetekoälyn algoritmit pystyvät yleensä arvioimaan ilon tai onnellisuuden hyvin, mutta negatiivisten tunteiden kuten vihan tai pelon tai monitulkintaisten tunteiden tunnistaminen on heikompaa tai toisinaan jopa virheellistä (Calvo ja muut, 2018; Cojocea ja muut, 2019; Stöckli ja muut, 2018; Tavares ja muut, 2016). Validaatiotutkimus on saanut jalansijaa ihan EU-tasoisesti (ks. Lassalle ja muut, 2019), ja validaatiotutkimukset ovat laajentuneet kattamaan kasvoniilmeiden lisäksi myös ääntä emootioiden välittäjänä (Bänziger ja muut, 2009; Samadiani ja muut, 2019). Tämä onkin välttämätöntä, jotta tunnetekoälyn suhteen saavutetaan tutkimuksessa riittävä tarkkuus.

Tunnereaktioiden tunnistamista voivat vaikeuttaa myös arkisemmat, esimerkiksi valaistukseen, pään ja kasvojen väliseen suhteeseen tai hiuksiin ja koruihin liittyvät tekijät (Calvo ja muut, 2018, 2).

Etenkin laboratorio-olosuhteiden ulkopuolella haasteina ovat muun muassa valaistus ja pään asento (Samadiani ja muut, 2019; Corneanu ja muut, 2016, 1560; Kumar ja muut, 2019, 946; Martinez ja muut, 2019, 340). Kasvot ja pää eivät ole todellisessa elämässä kovin suorassa (Krumhuber ja muut, 2020).

Kasvoniilmeisiin perustuvien analysointimenetelmien vaikeus on, että ne vaativat kunnolla toimiakseen edestäpäin otetut videot, joilla pään asento ei saa juurikaan vaihdella eikä olla peitossa. Niin sanotussa luonnollisessa ympäristössä tai puheen aikana kasvoniilmeiden tunnistaminen on haastavaa. (Pietikäinen & Silvén, 2019, 199.) Luonnollisessa vuorovaikutuksessa varsinkin ryhmässä ihmisen pää liikkuu kuitenkin jatkuvasti, jolloin ryhmävuorovaikutustutkimuksessa on tasapainoiltava jatkuvasti tunnereaktiodatan määrän ja vuorovaikutuksen luonnollisuuden välillä. Jompikumpi kärsii väistämättä.

Vuorovaikutuksen dynamiikan osalta varsinkin kasvojentunnistusanalyysin kohdistuminen VALAA:ssa nauhoituksen hetkellä yhteen osallistajaan vaikutti siten, että hän ei voinut kääntää päätään juuri ollenkaan. Tämä teki ryhmävuorovaikutuksesta hänen osaltaan epäluonnollisen. Erityisesti non-verbaalinen vuorovaikutus, jossa osallistajat kääntyivät välillä toisiinsa päin, hymyilivät, nyökkäilivät ja ilmehtivät, jäi suorassa kasvojentunnistusanalyysissä olleen osallistujan kohdalla lähes kokonaan puutteelliseksi.

Tästä osallistajat puhuivat myös reflektiovideoinnissa keskustellessaan kasvojentunnistusanalyysin vaikutuksesta toimintaansa todeten, että katsekontakti toimii merkinä, että haluaa sanoa jotakin. Jos osallistuja ei pysty ottamaan katsekontaktia, se vaikeuttaa myös oman puheenvuoron ottamista, koska hän ei pysty seuraamaan toisten sanatonta viestintää.

Kasvojentunnistuksessa ollut opiskelija kuvasi asiaa näin: "miettimään sitä kameraa, että kuinka paljon mä voin liikkua ja sitten ahah, ne on jo toisessa aiheessa". Tämän huomaaminen sai minut muuttamaan tutkimusasetelmaa hiukan siten, että annoin luvan vahvempaan kääntymiseen seuraavilla videotinkerroilla. Kääntyily johti kuitenkin kasvojentunnistusanalyysin kattavuuden heikkenemiseen. Näin ollen tilanteessa ilmeni tasapainoilu vuorovaikutuksen luonnollisuuden ja tutkimusdatan kattavuuden välillä. Molempien saavuttaminen maksimaalisesti yhtä aikaa ei ollut mahdollista.

Koronapandemian vuoksi joitakin videointeja jouduttiin toteuttamaan etäyhteydellä Zoomin välityksellä. Tällöin kaikki kohdistivat katseen kameraan,

jolloin tunnereaktiot saatiin määrällisesti hyvin talteen, mutta luonnollinen vuorovaikutus osallistujien väliltä jäi vajaaksi. Ruudulta etäyhteyden välityksellä ei ollut mahdollista esimerkiksi huomata, koska tietty osallistuja katsoi tiettyä toista osallistujaa.

Kun vuorovaikutuksen tutkiminen siirretään laboratorioon, haasteena on muutenkin säilyttää luonnollisen vuorovaikutuksen dynamiikka (Stevanovic, 2016, 405–406). Erilaiset mittauslaitteet saavat ihmiset toimimaan epäluonnollisesti noin tunnin ajan (ks. Pentland, 2012, 5; Peräkylä ja muut, 2016, 260). Parhaimmillaan vuorovaikutustutkimus on yhtä aikaa sekä kontekstiherkkää että yleistettävää (Voutilainen, 2016, 387–388). Kontekstiherkän koodausjärjestelmän luominen on kuitenkin vaikeaa, koska jokaisella tilanteella on oma dynamiikkansa (Haakana, 2002, 227). Kontekstiherkkyys tarkoittaa myös sitä, että videointien analysointia ei voida koskaan kokonaan automatisoida, vaan tutkijan tulkinnoilla ja luokittelulla on niissä keskeinen rooli.

Luotettavuutta lisää joka tapauksessa videoiden käyttö valokuvien sijaan: Kasvojen ilmeiden tärkein piirre on sen dynaaminen luonne, ja dynaaminen (liikkuva) informaatio parantaa tunteen tunnistuksen koherenssia ja virheettömyyttä. Lisänä on kuitenkin hyvä käyttää erilaisia itsearviointeja. (Krumhuber ja muut, 2013.) Tekoäly ei nimittäin kykene kovin tulkitsevaan työhön (Saranpää, 2019). Rikkain näkemys ihmisen käyttäytymisestä saadaan usein nimenomaan monimenetelmäisesti, sillä silloin menetelmät tuottavat keskenään erilaista tietoa (Murray, 2019).

Vaikka tällä hetkellä perustunteista vallitsee melko laaja yksimielisyys etenkin viiden tunteen eli vihan, pelon, inhon, surun ja ilon / onnellisuuden osalta (Ekman, 2016), myös muut näkemykset, kuten tunteet konstruktiona näkevät teoriat, epäkognitiiviset emootioteoriat, enaktivismi sekä arviointitunneteoriat (ks. mm. Barrett ja muut, 2019, 4; Salmela, 2017, 33) ovat voimissaan. Toisaalta tunteiden universaalisuus täydennettynä tunteiden ”murteiden” ja tunteiden esittämissääntöjen teorioilla on tieteellisesti vahvasti toteennäytetty (ks. esim. Nummenmaa, 2019). Dimensionaalinen (eli näkemys tunteista eri ulottuvuuksia sisältävänä) ja diskreettinen (eli näkemys tunteista erillisinä perustunteina) näkemys tunteista on ehkä mahdollista osittain yhdistää, jos tunteet käsitteellistetään sekä fysiologisina, kokemuksellisinä että behavioraalisina vastauksina henkilökohtaisesti merkityksellisiin ärsykkeisiin (Mauss ja Robinson, 2009).

Tunnereaktioiden ja tunnekokemusten ja niiden välisen paradigmaeron ylittäminen ja yhdistäminen on mahdollista, jos nojautumme *arviointitunneteorioihin* (*appraisal theories of emotion*). Niissä tunteet nähdään sopeutuvina reaktioina ja havaintoina ympäristön haasteisiin siten, että niihin sisältyy myös neuraalisia ja fysiologisia muutoksia (Ellsworth, 2013). Näitä fysiologisia muutoksia voidaan tutkia biometriikan kautta ja reaktioita ja havaintoja ihmisten kokemusten kautta. Tilanne ja aikaisemmat kokemukset vaikuttavat siihen, miten mentaalisten prosessien eli arviointien kautta tunteet syntyvät (Moors, 2014).

Näin ollen kasvojenilmeet nähdään dynaamisina ilmiöinä, joihin ihmisten omat arviot vaikuttavat (Scherer ja muut, 2019). Tämä näkemys auttaa tieteenfilosofisesti yhdistämään kokemuksia painottavat menetelmät biometrinen menetelmien kanssa. Arviointitunneteoriolla on yhtäläisyyksiä Maussin ja Robinsonin (2009, 210) esittämän mallin kanssa, jossa tietty tapahtuma aiheuttaa

arvion sen merkityksestä, joka puolestaan aiheuttaa emotionaalisia reaktioita niin kokemusten, toiminnan kuin autonomisen hermoston suhteen.

Tämä näkemys on mahdollista yhdistää myös esimerkiksi kasvatustieteissä usein ilmenevään ihmiskäsitykseen, jossa keskiöön halutaan nostaa yksilön omat kokemukset niin tunteista, oppimisesta, dialogisuudesta kuin vuorovaikutuksestakin. Vain nämä huomioimalla katsotaan voitavan kunnioittaa ja edistää ihmisen toimijuutta. Toisaalta arviointiteorian primaarireaktio – eli ensin seuraava, yleensä nopea reaktio ilman siihen liittyvää arviointia ja tulkintaa⁴ – jota olen itse tutkinut ja tulkinnut diskreetin tunneteorian mukaisella kasvojenilmeiden ilmenemisellä, täydentää nähdäkseni erinomaisesti ihmistieteisiin sisältyvää tunnekokemuksiin ja niiden tulkintaan perustuvaa traditiota. Sen mahdollistama luonnontieteellisempi paradigma tuo ihmisen toimintaan liittyvät, osittain tiedostamattomatkin, reaktiot fokukseseen.

Näin ollen vuorovaikutuksesta saadaan monipuolisempi ja vivahteikkaampi kokonaiskuva. Lisävalaistusta tuo myös ryhmäprosesseihin liittyvä viitekehys, joka laajentaa kehystä sosiaalisempaan suuntaan. Liittyväthän tunteisiin ja niiden tutkimukseen niin tunteisiin liittyvät subjektiiviset kokemukset, niihin linkittyvät psykofyysiset prosessit kuin niiden sosiaalinen funktiokin (Nummenmaa, 2010, 53; Peräkylä, 2009, 261; van Kleef & Fischer, 2016, 4). Ylipäänsä sisäiset kognitiiviset ja emotionaaliset sekä ulkoisemmat käyttäytymiseen ja toisten kanssa vuorovaikutukseen liittyvät prosessit ovat yhteydessä toisiinsa (Janusz & Peräkylä, 2017, 427), joten on hyödyllistä, jos tutkimus pystyy valottamaan useampaa niistä.

Paradigmojen yhdistämisen konkreettisena seurauksena on usein monimenetelmäisyys tai multimodaalisuus. Monimenetelmäisyys viittaa monien erilaisten tutkimusmenetelmien käyttöön samassa tutkimuksessa. Esimerkiksi tunteiden tutkimisessa on suositeltavaa käyttää montaa eri menetelmää tai komponenttia kuten osallistujien kokemuksia, käyttäytymistä tai toimintaa luotaavia menetelmiä sekä psykofysiologisia menetelmiä (Eteläpelto ja muut, 2018, 8). Multimodaalisen datan käyttö yhdistää niin sanottuja subjektiivisia ja objektiivisia tutkimusmenetelmiä yhdistäessään fysiologisten reaktioiden, subjektiivisten kokemusten ja tunteiden sosiaalisten funktioiden tutkimisen vuorovaikutuksessa. Parhaimmillaan multimodaalinen data pystyy ylittämään ontologisia rajoja ihmiskehon (neurologiset prosessit), ympäristön (toiminta) ja mielen välillä. (Järvelä ja muut, 2019.)

VALAA:ssa ja SOEVE:ssa monimenetelmäisyys oli jälkikäteen ajatellen viisas valinta, sillä silloin tutkittavista ilmiöistä kuten tunteista saatiin monipuolisempi ja laajempi kuva. Tunnetekoälyä hyödyntävän algoritmin hyödyntäminen ryhmävuorovaikutustilanteissa ei tuottanut luonnontieteellisesti katsottuna riittävän tarkkaa koeasetelmaa. Toisaalta ryhmävuorovaikutustutkimuksessa tutkimusasetelmaa ei voikaan rakentaa yhtä tarkaksi koeasetelmaksi kuin vaikkapa muistin tutkimista, sillä siinä myös toiminnan sosiaalinen puoli on sisällytettävä tutkimusasetelmaan.

Luonnollisesti monimenetelmäisyys tai -modaalisuus asettaa haasteita luotettavuuden arvioinnille (Eteläpelto ja muut, 2018, 8). Sen lisäksi, että laboratorio-olosuhteissa haasteena on säilyttää luonnollisen vuorovaikutuksen dynamiikka, luotettavuuden kannalta vertailukelpoista materiaalia tulee myös olla riittävästi (Stevanovic, 2019, 405–406). Jo ilmiö itsessään, kuten ihmisen emootiotilan luotettava tunnistaminen ja seuranta, voi on hyvin haastavaa ja monimutkaista (Pietikäinen & Silvén, 2019, 106). Tieteenfilosofisesti erilaisista

lähtökohdista lähteviä menetelmiä voi olla vaikea yhdistää, tai ainakin niiden vertailussa on vaikea löytää yhteismitallisia kriteereitä.

4. Tunnetekoälytutkimukseen osallistuvien ihmisten toimijuuden huomiointi

Tunnetutkimusta tehtäessä on tarpeen erottaa biologisesti määräytyneet ja automaattisesti toimivat tunnereaktiot niihin liittyvistä tietoisista kokemuksista eli tunnekokemuksista (Nummenmaa, 2010, 16). Vuorovaikutukseen liittyvässä tunteiden tutkimisessa on keskitytty usein puhtaasti tunnekokemuksiin, vaikka ne edustavat ”jäävuoren huippua” tunneprosessissa. Toisaalta tunteiden ohjaavuus sosiaalisissa vuorovaikutustilanteissa on ilmeistä ja osittain jopa automaattista, joten myös tunnekokemusten ymmärtäminen on keskeistä. (Nummenmaa, 2017, 37; 2010, 52–53.)

Pelkkien tunnereaktioiden tutkiminen jättää ihmisen kokemuksellisenä ja sosiaalisena olentona tutkimuksen ulkopuolelle, mutta toisaalta tunnetekoälyn mahdollistama tunnereaktioiden löytäminen tuottaa tietoa, jota pelkkä tunnekokemuksiin keskittyminen ei paljasta. Ihmiset eivät ole aina tietoisia tunnereaktioistaan, sillä ne ovat usein hyvin automaattisia ja nopeasti esiintyviä. Näin ollen tunnereaktioiden tutkiminen voi tuottaa tietoa, jota tunnekokemuksia tutkimalla ei saada esille. Esimerkiksi on huomattu (Reed ja muut, 2012), että aidot tai sosiaaliset hymyt saattavat ennustaa todennäköisyyttä yhteistyöhön molempien osapuolten taholta, kun puolestaan halveksunnan ilmeneminen kasvoilla saa todennäköisemmin peräytymään yhteistyöstä. Pesosen (2018) tutkimuksessa taas kyselytutkimukset antoivat positiivisemmän kuvan koehenkilöiden suhtautumisesta materiaaliin kuin heidän tunnereaktionsa biometriikan perusteella.

Tunnereaktiot voivat paljastaa vuorovaikutuksesta asioita, joita ei muuten ”näe”. Yksi tällainen havainto koski VALAA:ssa toisiin liittyviä asenteita. Yksittäisistä tunnereaktioista ei voi vielä päätellä mitään, mutta kun esimerkiksi Hannan tunnereaktiot ovat lähes aina positiivisia, kun Jaana puhuu ja negatiivisia, kun Pekka puhuu, tunnereaktiot todennäköisesti paljastavat Hannan asenteita ryhmäläisiä kohtaan. Asenteet eivät toisaalta ole täysin johdonmukaisia, vaan vaihtelevat vuorovaikutustilanteen aikana. Vaikka Otolla on välillä ylenkatsetta Viiviä ja Niinaa kohtaan ja toisinpäin, humoristinen Niinan lohkaisu aiheuttaa Otossa iloa ja kiinnostusta.

Konflikteihin ajautuvassa vertaisryhmässä tällaista ”asenteellista kertymää” oli erityisen paljon. Tässä voi esiintyä luontaisia suhtautumiseroja ja toisista pitämisen eroja ihmisten välillä, mutta myös vuorovaikutustilanne itsessään voi joko lisätä tai vähentää näitä eroja. Esimerkiksi ryhmän juuttuminen konflikteja ruokkivaan vuorovaikutuksen kehään näytti lisäävän tällaista asenteellista kertymää. Tärkeä rooli on sillä, että vuorovaikutuksessa pyritään välttämään ongelmia toistavia tai syventäviä vuorovaikutusastelmia (Leiman, 2015). Tällaisia asetelmia ovat muun muassa kapeuttaminen, jossa sovittelevaa puhesävyä polarisoidaan tai kärjistetään tai jopa palautetaan uudelleen riitaisampi ääni mukaan keskusteluun (Kankaanpää & Savolainen, 2012, 14, 31).

Vaikka tällaista asenteellista kertymää ei varmaankaan voi kokonaan poistaa, sitä voi pyrkiä vähentämään ryhmäytymisen ja dialogisen vuorovaikutuksen keinoin. Vuorovaikutus ryhmän kesken voi myös pysyä positiivisena tällaisista asenteellisista tunnereaktioista riippumatta, joten pelkästään niiden perusteella ei pidä tehdä liiallisia tulkintoja.

Lohdullista on, että ryhmävuorovaikutustaitoja voidaan parantaa merkittävästi. Ryhmän saama palaute auttaa tasapainottamaan vuorovaikutusta muun muassa siten, että puheaika jakautuu tasaisemmin ja puheenvuorot vaihtuvat tiheämmin, hallitsevaa käyttäytymistä on vähemmän ja yhteistyötä enemmän sekä ryhmän suoritusaste, luottamus ja sitoutuminen kasvavat. Erityisesti visualisoitu palaute lisää ryhmän jäsenten tietoisuutta ryhmädynamiikasta. (Lederman ja muut, 2018, 32; Kim ja muut, 2012; Pentland 2012.)

Ylipäänsä vuorovaikutustilanteiden aiheuttamien negatiivisten tunnereaktioiden tulkinnassa on oltava suorastaan ylivarovainen. Esimerkiksi on epäeettistä tulkita osallistujan tunnereaktio absoluuttisesti halveksunnaksi, koska tunne on niin kielteinen, jopa leimaava ihmissuhteissa ja vuorovaikutuksessa. Nykyinen tunnetekoaalyteknologia soveltuukin parhaiten sovelluksiin, joissa kriittisenä tekijänä ei ole emootioiden tunnustustarkkuus (Pietikäinen & Silvén, 2019, 199).

Olisikin hyödyllistä, jos useammassa tutkimuksessa tutkittaisiin yhtä aikaa tunnereaktioita ja tunnekokemuksia. Pelkkä tunnereaktioihin keskittyminen voi tosin vääristää tuloksia, kun ihmisen toimijuutta ja siihen liittyviä omia kokemuksia ei huomioida. Jos ihmisen fyysiset piirteet tai käyttäytyminen tulkitaan ihmisen käyttäytymistä ennustavaksi siten, että häntä itseään ei kuulla eikä hänen tahtoa tai toimintaa määrittäviä motiivejaan kartoiteta, riskit diskriminaatioon ja stigmatisointiin väärin tulkintojen muodossa kasvavat (Sutrop & Laas-Mikko, 2012, 22–23, 34). Tunnekokemusten huomointi vähintään tunnereaktioiden rinnalla huomioi ihmisen kokonaisvaltaisemmin.

Omissa vuorovaikutustutkimuksissani biometrinen tieto osallistujien tunnereaktioista tuotti tietoa heidän oletetuista tunnetiloistaan vuorovaikutuksen aikana. Tämä tieto oli osittain osallistujien oman havaitsemisen ja reflektoinnin ulottumattomissa, ja toisaalta jälkikäteen arvioiden he eivät kyenneet kaikkia yksityiskohtia tunnereaktioistaan muistamaan. Osallistujien tunnekokemuksia kartoittavat kyselyt sen sijaan toivat tietoa osallistujien havainnoimista ja reflektioista tunnekokemuksista. Tunnekokemusmittaristossa sisälsi kymmeniä tunteita, joten tunteiden skaala oli laajempi kuin tunnereaktioiden tarkastelussa.

Tunnekokemukset ovat luultavasti vaihtelevampia ja monipuolisempia kuin tunnereaktiot (Nummenmaa, 2010, 37), käyttäväthän suomalaiset käyttävät yleisesti noin 50 sanaa tunteidensa kuvaamiseen (Tuovila, 2005). Tämän vuoksi tunnekokemusmittarissa oli mukana laajempi skaala tunteita kuin tunnereaktioita mittaavassa iMotionsin AFFDEX-algoritmissa. Tunnekokemukset lienevät monin osin valideja kokemuksia osallistujien koko videoinnista, mutta mitattuina vasta vuorovaikutusvideoinnin lopussa ne eivät kykene mittaamaan yksittäisen hetken tunnekokemusta. Ylipäänsä on nähty ongelmalliseksi, jos tunteita tutkitaan haastattelemalla (tai kyselemällä) vasta oppimisen jälkeen (Eteläpelto ja muut, 2018, 8).

Lisäksi tunnekokemusmittarista puuttui kokonaan kysymys, mihin tunnekokemus kohdistui. Reflektioilanteissa osallistujat toivat kuitenkin esille useita hetkellisiä tunnekokemuksiaan, jotka eivät ilmenneet tunnekokemusten kokonaisarviointinsa. Hetkellisyys näkyi vahvasti myös tunnereaktioissa. Tunnekokemukset ovatkin VALAA:sa osittain mittaustavasta johtuen olleet kokonaisvaltaisempia kuin tunnereaktiot, joita ilmenee kasvojenilmetunnistuksen perusteella lähes koko vuorovaikutuksen aikana. Täsmällistä vertailtavuutta ei niiden välillä voida tehdä. Tunnereaktiot ovat spesifimpiä ja paljastavat tiettyyn vuorovaikutuksen hetkeen sisältyviä tunnereaktioita.

Sen sijaan SOEVE:ssa osallistujat täyttivät tunnekokemusmittariston myös kesken vuorovaikutustilannetta. Lisäksi kysyin tunnekokemusten kohdistumisesta. Tämä tuotti välittömästi tarkempaa tietoa: Paljastui, että esimerkiksi suhteessa vuorovaikutukseen osallistujat kokivat hyväksyntää, iloa, yhteenkuuluvuutta ja ylpeyttä ja suhteessa keskustelun sisältöön kiinnostusta ja oivallusta. Omaan oppimiseen liittyen koettiin eniten tyytyväisyyttä ja luottamusta. Näin ollen tunnekokemuksistakin saatiin aikaisempaa tarkempaa tietoa.

Yksi tapa lisätä tutkimukseen osallistuvien mahdollisuutta omaan toimijuuteen ja omiin tulkintoihin vuorovaikutuksestaan on reflektio. Reflektiossa tapahtuva jälkikäteinen tilanteen tarkastelu antaa osallistujalle mahdollisuuden omaan tulkintaan ja oman toiminnan ja ajattelun sanallistamiseen. Tällöin sekä osallistujien omat tunnekokemukset että heidän omat tulkintansa heidän vuorovaikutuksestaan ovat keskeisessä osassa.

Osana tätä tulkintaa ja kokemusten jakamista voidaan hyödyntää niin sanottuja reflektiovideointeja tai jälkihaastatteluja, joissa osallistujat katsovat tai kuuntelevat vuorovaikutustallenteensa tai osia siitä. Tällaista menetelmää kutsutaan *Stimulated Recall* (STRI) -menetelmäksi (tai lähes rinnakkaisena terminä käytettynä *interpersonal process recall* (IPR) -menetelmäksi) (alun perin mm. Kagan ja muut, 1963; ks. myös Ryan & Gass, 2012). Tätä menetelmää on käytetty ammatillisen kehittymisen välineenä muun muassa opettajilla (ks. Hennissen ja muut, 2010; Gazdag ja muut, 2019), kliinisessä työssä ja psykoterapiassa (ks. Janusz & Peräkylä, 2017, 429) sekä osana ammatillisen kommunikaation parantamista eri ammattilaisilla (esim. Hussin, 2013). Tämä menetelmä tarjoaa uusia näkökulmia sekä ihmissuhteisiin että ryhmäprosesseihin (Kagan ja muut, 1963). Menetelmä soveltuu sekä yksilöiden, dyadien että myös ryhmien tutkimiseen.

Tämän menetelmän käyttäminen vuorovaikutustutkimuksissa mahdollistaa osallistujille aiempaan tilanteeseen liittyvien ajatusten ja kokemusten muistelun (Juutilainen, 2003, 69). Tällöin keskustelun pohjana on usein havainnointimateriaalia, kuten juuri videoituja tilanteita, joiden avulla tilanne palautuu osallistujien mieleen (Atjonen, 1996, 229–231). Reflektiovideointien käyttöä puoltaa myös se seikka, että osallistujien on vaikeaa muistaa asioista jälkikäteen pelkästään muistinvaraisesti. Videointi ”hidastaa aikaa” ja antaa mahdollisuuden sekä oman toiminnan reflektiiviseen tarkkailuun, että kokemuksesta oppimiseen. (ks. Aaltonen & Eskola-Kronqvist, 2011, 49.)

Menetelmä voi olla pehmeän joustava tai tiukemmin strukturoitu riippuen muuan muassa siitä, kuka valitsee katsottavat yksityiskohdat, miten kysymykset muodostetaan ja mikä on tutkijan ja osallistujan rooli (Gazdag ja muut, 2019, 66). Jo menetelmän alkuaajoista saakka fysiologisia reaktioita on tutkittu myös suhteessa tähän menetelmään ja toimijuuteen (Kagan & Schauble, 1969, 312), joten omat tutkimukseni sijoittuvat tälle jatkumolle.

Stimulated recall -menetelmän tavoitteena on yleensä ihmisen toimijuuden lisäämiseen jollakin tavalla. Esimerkiksi psykoterapiassa menetelmän avulla voidaan tarkastella, miten terapeutin toiminta voisi lisätä asiakkaan reflektiota, itsetietoisuutta ja toimijuutta (Rennie, 2001). Menetelmän avulla ammattilainen, kuten psykoterapeutti, voi oppia oman toiminnan reflektointia ja sisäisen puheensa tavoittamista ”kokevan minän” ja ”ammatillisen minän” sisäisenä dialogina (esim. Rober, 2005).

Menetelmän tavoitteena on päästä käsiksi osallistujien sisäisiin kokemuksiin tietyissä hetkissä. Ihmisten välisessä vuorovaikutuksessa ilmeneviä kognitiivisia ja

emotionaalisia prosesseja voi olla vaikea tai mahdoton havainnoida varsinaisessa vuorovaikutustilanteessa. Niinpä menetelmän sisältämän palautuksen tai uudelleen muistelun kautta voidaan pyytää osallistujia palauttamaan mieliin heidän sen hetkisiä ajatuksia, tunteita ja muistoja. (Janusz & Peräkylä, 2017, 429-430.) Osallistujat voivat kokea videon katsomisen hetkellä tai sen jälkeisessä keskustelussa erityyppisiä oivalluksia ja muutoksia. Nämä voivat liittyä esimerkiksi itsensä ymmärtämiseen, tunteiden kokemiseen ja ilmaisemiseen, toisen ihmisen havainnointiin sekä omaan vuorovaikutukseen ja sen ristiriitoihin. (Huhtamäki & Lehtinen, 2016, 21–30.)

Vuorovaikutteisessa reflektiossa ihminen suuntautuu myös muita ihmisiä ja heidän ajatuksiaan ja tarpeitaan kohti (Laine, 2004, 24). Tällainen reflektio sopii hyvin pedagogisesti tai kasvatustieteellisesti orientoituneeseen kontekstiin, jossa painotetaan yhteistoiminnallisuuden kautta syntyviä vuorovaikutuskvalifikaatioita ja oman persoonan kehittymistä toimimalla suhteessa muihin (Pasanen, 2014, 15). Kun osallistujat katsovat heistä otettua videota ja keskustelevat siitä, pohdinta ja ymmärrys syvenee jokaisella kierroksella. Katsomistilanteesta voi tulla jopa merkityksellisempi ihmisten välisen vuorovaikutuksen kannalta kuin varsinaisesta videoinnista. Yllättävin ja arvokkain hetki ei ole etukäteen tiedossa. (Mäkelä ja muut, 2013, 25–26.)

Stimulated recall -menetelmä voi myös auttaa ymmärtämään mitkä asiat ovat osallistujille tärkeitä ja mitä he yrittävät toisilleen välittää. Sen avulla voidaan yrittää löytää vastauksia siihen, miksi vuorovaikutus on joskus sujuvaa ja joskus taas tahmeaa. (Dempsey, 2010.) Reflektioon rohkaisevana ja ei-arvostelevana menetelmänä se voi auttaa pienentämään ihmisen intentioiden ja toiminnan välistä kuilua. Parhaimmillaan tuloksena ovat odottamattomat ”löydökset”, joita muut menetelmät eivät välttämättä tavoita. (Wass & Moskal, 2017.)

Reflektiovideot olivat tärkeä osa osallistujien kokemusten ja ryhmävuorovaikutuksen kehittymisen väline. VALAA:ssa osallistujat täyttivät reflektiovideota katsoessaan visuaalista dialogisuusjatkumoa, joka oli keskeinen osa tutkimuksen toteuttamista. Dialogisuusjatkumo on itse kehittämäni visuaalinen työkalu, joka oli erittäin tärkeässä osassa reflektiota. Olin pilkkonut vuorovaikutusvideon temaattisesti sopiviin pätkiin (6–12 pätkää / video). Jokainen osallistuja arvioi itsenäisesti jokaisen teemapätkän dialogisuuden. Näistä otin tarkasteluun dialogisimmiksi ja vähiten dialogisiksi rankatut teemajaksot (ryhmän antamien arvioiden keskiarvojen mukaan) ja tarkastelin niitä dominanssien eli vuorovaikutuksen hallinnan kautta, jolloin sain tietoa siitä, millaiset piirteet ovat tyypillisiä dialogisiksi ja ei-dialogisiksi koetulle vuorovaikutukselle. (ks. tästä tarkemmin Gjerstad, 2022.)

Jokaisen pätkän jälkeen ryhmä keskusteli ajatuksistaan ja tunteistaan niin kuin myös siitä, miksi kyseinen vuorovaikutuskohta tuntui tai ei tuntunut dialogiselta. Osallistujat tekivät huomioita sekä omasta vuorovaikutuksestaan (esimerkiksi puheajan käytöstään) että ryhmänsä vuorovaikutuksesta (esimerkiksi siitä, milloin puhe näytti soljuvan parhaiten tai milloin kaikki näyttivät osallistuvan parhaiten vuorovaikutukseen).

Reflektiovideoinnissa on myös mahdollista palata sisäisiin jännitteisiin ja jopa väärinymmärryksiin, joita varsinaisen videoidun vuorovaikutustilanteen aikana on tapahtunut. Reflektiovideointi soveltuu täten hyvin myös ryhmän oppimisen reflektion välineeksi.

Näin kävi VALAA:ssa, jossa eräs ryhmä keskusteli kauan epämiellyttävistä tunnekokemuksistaan sekä erilaisista konflikteja aiheuttaneista väärinymmärryksistä reflektiovideoinnin aikana. Pelkkä tunnereaktioiden tai tunnekokemustenkaan tutkiminen ei mahdollista tällaista osallistujaryhmän omien kokemusten jakamista ja uudelleentulkintaa.

Koska havainnoinnin kohteena oleminen niin kuin myös itsensä katsominen videolta esimerkiksi reflektiotilanteessa, on aina ihmiselle stressitilanne, myönteistä palautetta täytyy käyttää kompensoimaan niistä seuraavaa stressiä. Psykiatri Jukka Mäkelän mielestä videointi siten, että ihminen ”ei saa kuulla mitään palautetta, on haitan luomista ihmisen elämään”, eikä meillä ole oikeutta tällaisen haitan luomiseen. Videoinnin tarkoitus ja työskentelylogiikka täytyy myös esitellä erityisen selvästi ja ”läheisesti” paikalla olevan ihmisen kertomana. Tähän kuuluu se, että osapuolet ymmärtävät, miksi vuorovaikutusta kuvataan. (Mäkelä ja muut, 2013, 21–30.)

Tämän vuoksi osallistuin intensiivisesti koko tutkimusprosessiin, ja viestin mieluummin liikaa kuin liian vähän. Mäkelän ja muiden (2013, 23–24) periaate antaa saman tien videoinnista osallistujille ainakin yksi myönteinen palaute analysoimatta kuitenkaan sen syvemmin tilannetta on linjassa sen periaatteen kanssa, että tutkittaville ei tehdä haittaa. Noudatin samaa periaatetta VALAA:ssa ja SOEVE:ssa: Sanoin jotakin positiivista ryhmän vuorovaikutuksesta ja toiminnasta heti videoinnin jälkeen, vaikka tarkemmat tulokset tulivatkin vasta myöhemmin.

On myös hyvä myöntää, että vuorovaikutustilanteet eivät ole neutraaleja, vaan niihin sisältyy kysymys, vaaditaanko osapuolia jatkuvasti muuttumaan (Mäkelä ja muut, 2013, 29, 31). Toimijuuden kannalta on väliä myös sillä, painotetaanko ryhmävuorovaikutuksessa sosiaalipsykologisesti mukautumisen tai ryhmäpaineen mekanismeja (ks. esim. Logren, 2019, x) vai kuten kasvatustieteellisesti orientoituneessa vuorovaikutustutkimuksessa yleensä ihmisen autonomisuuden, kunnioituksen, dialogisuuden ja hyvän elämän ideaaleja.

Mukautuvaisuus, myöntöväisyys tai samaistuminen eivät nimittäin useinkaan johda arvotason muutoksiin, jotka taas ovat edellytyksiä ihmisen sisäiselle motivaatiolle ja ihmisen identiteetin kehittymiselle (esim. Kelman, 2006, 3–4). Mukautumisen korostamisessa ei myöskään huomioida välttämättä kunkin yksilön hyvää elämää. Arvoihin perustuva toimijuus on yleensä pysyvämpää ja pitkäkestoisempää kuin mukautumiseen perustuva toimijuus. Se on myös enemmän linjassa toimijuuden ideaan sisältyvien ihmisen oman aktiivisuuden, aloitteellisuuden ja päämäärätietoisuuden kanssa. Kasvatuksessa pyritään tällaiseen toimijuuteen, sillä se mahdollistaa sivistyneen, autonomiseen ajatteluun kykenevän persoonan kehittymiseen.

Paradoksaalisesti yksilön kunnioituksen kanssa samanaikaisesti vuorovaikutukseen sisältyy sisäänrakennettuna (oikeutettujakin) vaikuttamisen ja oppimisen vaatimuksia. Pedagogiset vuorovaikutustutkimuksetkaan eivät ole immuuneja näille eettisille haasteille. Kuitenkin pedagogiikkaan sisältyvä vahva, lähes mahdottomalta tuntuva hyvän elämän ideaali edesauttaa eettisyyden huomaamisessa ja toteuttamisessa. Sen mukaisesti eettisyyden mittana on, että toisen parhaita kykyjä ja ominaisuuksia edistetään ja tarpeita ja oikeuksia kunnioitetaan nimenomaan nykyhetkessä eikä vasta tulevaisuudessa (Värri, 2000, 23).

Toimijuuden osana voitaneen pitää myös mahdollisuutta anonymiteettiin keskeisenä tunnetekoälyn vaatimuksena. Biometrinen menetelmien tavoitteena on yleensä joko tunnistamisen avulla erottaa persoonat toisistaan tai ennustaa jonkun käyttäytymistä tai intentioita (ks. Sutrop & Laas-Mikko, 2012, 22), mikä on anonymiteetin ja toimijuuden kannalta ongelmallista. Toimijuuteen sisältyy sisäänrakennettuna ajatus oman itsensä määrittelystä ainakin osittain, joten ulkopuolelta rakennettu tulkinta ihmisen intentioista on erityisen ongelmallinen. Ratkaisuna lienee mahdollista harkita niin sanottua pehmeää biometriikkaa, jossa ihmistä ei voida tunnistaa prosessin kuluessa (McStay, 2020, 1). Silloin ideana voisi olla tietynlaisen tiedon kerääminen yleisellä tasolla yksilöimättä sitä tiettyyn ihmiseen.

Esimerkiksi parhaillaan menossa olevassa, useamman tutkijan toteuttamassa tiimien vuorovaikutusta tutkivassa TIITU-hankkessa keräämme biometristä tietoa, mutta muutamme sen saman tien tiimitasoiseksi tiedoksi. Emme ole kiinnostuneita yksilöiden biometrisistä ominaisuuksista, vaan siitä, miten tiimeissä työskentelyn aikana kasvojentunnistuksessa ilmenevät tunnereaktiot ovat yhteydessä tiimien vuorovaikutukseen ja ihmisten kokemuksiin siitä.

Tutkija on aina viime kädessä vastuussa tutkimuksensa eettisyydestä. Tutkimusta suunniteltaessa tutkijan on arvioitava tutkimukseen osallistumisesta ”mahdollisesti koituvia haittoja ja vahinkoja suhteessa tutkimuksella tavoiteltavaan tietoarvoon” niin osallistujien, heidän läheistensä kuin tutkijankin osalta. Tutkimuksen avulla oletettavan uuden tiedon merkitys on suhteutettava tutkimuksen mahdollisiin haittoihin ja riskeihin. Tutkimustilanteisiin saa kuulua saman tyyppistä ”henkistä rasitusta ja tunteiden kokemista” kuin arkeenkin. (Ihmiseen kohdistuvan, 2019, 8, 14–15.)

Ihmisen toimijuuden keskeisyys näyttäytyy suostumuksen vaatimuksena. Koska tutkittavan ei voida edellyttää hankkivan hänelle kuuluvista oikeuksistaan itse, jo tutkimusta suunniteltaessa täytyy ottaa kantaa tutkimusaineiston mahdolliseen avaamiseen, ja informoitava tutkittavia tästä (Ihmiseen kohdistuvan, 2019, 12–13). Jos aineistoa näytetään muille kuin tutkimukseen osallistuville ihmisille, siitä pitäisi siis eksplisiittisesti ilmoittaa jo tutkimuksen suostumisvaiheessa.

Suostumuksen käsite on tutkimuseettisesti selviö, mutta kasvojentunnistusteknologian käytössä usein rikottu periaate. Kasvojentunnistusteknologian käyttö on johtanut useisiin yksityisyyden loukkauksiin (Chochia & Nässi, 2021). Laajoja kuvatietokantoja on hyödynnetty kasvojentunnistukseen ilman siinä esiintyvien ihmisen lupaa. Esimerkiksi Yhdysvalloissa julkisen sektorin työntekijät ovat käyttäneet kasvojentunnistusteknologiaa ilman ihmisten suostumusta (Mac ja muut, 2021). Tiedelehti *Naturen* ja alan tutkijoiden enemmistön mukaan biometriisiin tutkimuksiin osallistuvilta ihmisiltä on saatava selkeä ja informoitu suostumus. (van Noorden, 2020.)

Suostumus yksinään ei kuitenkaan välttämättä riitä, sillä haavoittuvassa asemassa olevien ryhmien suostumus ei perustu aina heidän vapaaseen tahtoonsa. Aineistoa on myös mahdollista käyttää ihmisoikeuksien vastaisilla tavoilla eivätkä kaikki tutkijat noudata välttämättä yleisten ihmisoikeuksien mukaisia toimintatapoja. Kasvojentunnistusteknologian voidaankin lähtökohtaisesti ajatella olevan vaarallista, jos se voi tunnistaa tai luokitella ihmisiä ilman heidän suostumustaan. (van Noorden, 2020.)

Päätin helpottaa opiskelijoiden kynnystä lähteä mukaan VALAA- ja SOEVE-hankkeisiin lupaamalla aineiston käytölle ja hävitykselle tiukat rajat: Tunnistettava aineisto tuhotaan 1–1,5 vuoden päästä tutkimuksesta, ja aineistoa voivat katsoa vain tutkimushankkeiden työntekijät sekä heitä avustavat ihmiset. Litteroin videonauhat kokonaan itse. Tutkimusten ohjausryhmät päätyivät yksimielisesti rajaamaan aineiston käytön siten, että vain anonymisoituja, yleisiä tutkimustuloksia voidaan käyttää laajemmin. Esimerkiksi videonauhojen opetus- tai konferenssikäyttö on eettisistä syistä suljettu pois.

Omissa vuorovaikutustutkimuksissani – niin kuin yleisemminkin organisaatiossamme - vapaasta tahdosta annettu suostumus on itseäänselvyys, josta ei poiketa. Haaga-Helian vuorovaikutuslabin eettinen linjaus on tässä kohdassa tiukka: kasvojentunnistusta hyödyntävät tahot eivät saa edes kaupallisissa projekteissa vuorovaikutuslabista katsottavakseen videoita eivätkä niistä kerättyä dataa, vaan suoraan datasta tehdyt analyysit. Sama linjaus oli käytössä Ruisrockissa, jossa kerättiin vuonna 2017 kasvojentunnistusanalyysin avulla tietoa festivaalikävijöiden tunnetiloista (kävijöitä oli informoitu asiasta etukäteen). Festivaalijärjestäjät eivät saaneet nähdä videota tai kerättyä dataa, vaan he saivat suoraan siitä tehdyt analyysit. Videomateriaalia ei myöskään tallennettu minnekään. (Rissanen, 2017.) Tutkimuksellisesti tämä johtaa tosin toiseen ongelmaan eli siihen, että tutkimustulosten validaatiota biometrisen aineiston osalta ei voi kukaan enää jälkikäteen todentaa.

Suostumuksen tärkeyttä lisää se fakta, että biometriset tiedot luokitellaan erityisiksi henkilötietoryhmiksi. Ne viittaavat tietosuojalainsäädännössä henkilötietoihin, joista ilmenee ”rotu tai etninen alkuperä, poliittisia mielipiteitä, uskonnollinen tai filosofinen vakaumus tai ammattiliiton jäsenyys, geneettiset tiedot tai henkilön tunnistamista varten käsitellyt biometriset tiedot” (Ihmiseen kohdistuvan, 2019, 11). Vaikka biometristä tietoa ei kerättäisikään tai biometrisiä menetelmiä ei käytettäisikään henkilön tunnistamista varten, biometrisen tiedon mukana olo missä tahansa muodossa on tietyllä tavalla arkaluontoista, jonka vuoksi on tärkeää noudattaa tavallista tiukempia rajoituksia tutkimusaineiston kanssa.

5. Tekoälyn ja tunnetekoälyn eettiset periaatteet ihmisen toimijuuden valossa

Tekoälyn etiikasta puhuttaessa meidän olisi ymmärrettävä samanaikaisesti sekä moderneja teknologioita että eettisiä näkökulmia (Chochia & Nässi, 2021). Teknologia ei synny irrallaan poliittisista, yhteiskunnallisista ja eettisistä tekijöistä (Visala 2021, 66; Brey 2018). Tekoälyssä ilmenee usean tason haasteita: teknologiaan liittyviä, ihmiseen liittyviä, prosesseihin liittyviä ja liiketaloudellisen vaikuttavuuden haasteita (Pietikäinen & Silvén, 2019, 90). Jos halutaan kehittää ihmiskeskeistä tekoälyä, on huomioitava muun muassa yksityisyyden, datan omistajuuden, vastuullisuuden, tilivelvollisuuden, läpinäkyvyyden sekä oikeudenmukaisuuden ja reilun asettamat haasteet (Lepri ja muut, 2021).

Esimerkiksi Helsingin kaupungin linjaukset tekoälyn ja algoritmin hyödyntämisestä rekistereitä koskien linjaavat, että vastuullisuuden kriteerien pitää olla samat kuin ihmisten toiminnassakin (Haataja ja muut, 2020, 4). Eettisesti korkeatasoiset periaatteet voivat johtaa kuitenkin toisenlaisiin eettisiin haasteisiin: jos esimerkiksi läpinäkyvyys tulkitaan ymmärrettävyytenä, se voi johtaakin uudenlaisiin selittämisen, välitteisyyden ja tulkinnan muotoihin. Näin ollen läpinäkyvyyttä voidaankin käyttää jopa vallan oikeuttamiseen demokraattisen valvonnan sijaan. (Koivisto, 2023.)

Tunnetekoälyä hyödyntävän vuorovaikutustutkimuksen eettiset haasteet puolestaan liittyvät sekä käytettyyn teknologiaan, menetelmiin, osallistujiin, taustateorioihin ja paradigmoihin sekä niihin liittyviin oletuksiin ihmisestä ja yhteisöstä. Tästä syystä eettiset kysymykset ovat tällaisissa tutkimuksissa laajat ja monimutkaiset. Vaikka eettiset periaatteet ja jopa normistot voivat selkeyttää tunnetekoälytutkimusten ja –sovellusten käyttöä, säännöt tai normit eivät kuitenkaan koskaan takaa tunnetekoälyn eettisyyttä. Eettisten periaatteiden soveltamisessa on aina kyse myös sen taustalla olevasta ihmiskäsityksestä.

Teknologian hyödyntämisen mahdollisuudet ovat ainutlaatuiset ja niihin liittyvät prosessit äärimmäisen nopeita. Eettiset vaikutukset mietittyvät samalla, kun teknologia on jo osa elämäämme. (Chochia & Nässi, 2021.) Teknologia kehittyi nopeammin kuin hidas lainsäädäntö, moraali ja etiikka. Paradoksaalisesti tällainen toimintaympäristön muutos on tehnyt valvonnan lisäämisestä yhteiskunnassa helpompaa. Biometriseen tunnistamiseen liittyvät oikeudelliset kysymykset liittyvät siihen, miten biometristä tunnistamista käytetään lainmukaisesti. Eettiset kysymykset puolestaan koskevat sitä, pitäisikö biometristä tunnistamista ylipäänsä käyttää yhteiskunnassa. Periaatteena pitäisi olla, että biometristä tietoa ei kerätä vain sen helppouden tai nopeuden takia, vaan että sille on syvällisempi tarve. (Korja, 2016, 10, 162, 417, 421–422, 454.)

Utilitaristinen vääristymä viittaa siihen, että suostumme hyväksymään uuden teknonologian käytön, jos se vaikuttaa turvalliselta ja hyvinvointia yhteiskunnassa lisääväältä, mutta samaan aikaan emme pysty kuvittelemaan itseämme kyseisen teknologian käytön tai soveltamisen uhreina (Laakasuo & Palomäki, 2018, 50). Teknologia, joka on kehitetty tiettyyn tarkoitukseen, laajenee kuitenkin helposti sen alkuperäisestä käytöstä myös muuhun käyttöön. Tätä ilmiötä kutsutaan nimellä *function creep*. (Dekkers, 2019; Korja, 2016, 160–161, 454.) Esimerkiksi viattomalta tuntuvasta kasvojen tunnistuksesta liu'utaan helposti tunteiden tunnistukseen ja siitä päätelmien tekemiseen ihmisen tunteista tai luonteesta (Barrett, 2012, 414; Crampton, 2019, 59).

Esimerkiksi koronapandemian aikana oli usein epäselvää, miten paljon ihmisistä kerättyä tietoa käytettiin vain koronan leviämisen estämiseen ja minkä verran muihin tarkoituksiin (ks. Chochia & Nässi, 2021). Suomessa poliisi ja tulli saivat kesäkuussa 2019 oikeuden automaattiseen kasvojen tunnistamiseen: viranomaisilla on oikeus verrata automaattisen kasvojen tunnistusjärjestelmän ottamia valvontakameroiden kuvia henkilöstörekistereihin tallennettuihin kasvokuviin silloin, kun yritetään estää, paljastaa tai selvittää rikoksia tai tavoittaa etsintäkuulutettuja. Tämä on herättänyt huolta yksityisyyden suojasta, mutta viranomaiset ovat sanoneet kansalaisten oikeusturvan jopa paranevan, kun toiminta kohdentuu tarkemmin suoraan oikeisiin ihmisiin. (Hjelt, 2019.)

Rikosten parempi selvittäminen ei olekaan huolta aiheuttava asia, vaan pikemminkin parannus. Huolta aiheuttaa sen sijaan se, että valvontaa ulotetaan melko varmasti ennemmin tai myöhemmin myös alueille, joihin sitä ei teknologian kehittämisen aikaan ole suunniteltu tai tarkoitettu. Voi olla, että ihmisen toimijuus on huomioitu ja sitä on pyritty kunnioittamaan tietyn tunnetekoälyteknologian käytössä aluksi, mutta kun käyttötarkoitus muuttuu tai laajenee, toimijuuden suojelukin voi unohtua tai se voidaan nähdä vähemmän tärkeäksi. Jos toimijuudelle määritellään tietyt vähimmäisehdot, joihin sitoudutaan jo alussa, teknologian käytön laajenemisen vahingollisia seurauksia voitaisiin sillä osittain välttää.

Riskeistään huolimatta hyväksymme usein profilointialgoritmien käytön, sillä niiden käyttö vapauttaa meidät vaikeista päätöksistä ja sen myötä myös vastuusta. Esimerkiksi vuonna 2017 amerikkalaisen United Airlines -lentoyhtiön lennolta poistettiin väkivalloin matkustaja, jonka tietokonealgoritmi oli arponut poistettavaksi ylivaratusta koneesta. Henkilökunta ei kyseenalaistanut järjestelmän päätöstä eikä siten myöskään joutanut tilanteessa. Suuresta kohusta ja rahallisista tappioista ja korvauksista huolimatta työntekijät eivät joutuneet myrskyn silmään, sillä päätöksenhän olivat tehneet algoritmit. (Laakasuo & Palomäki 2018, 51–52.) Sen sijaan yrityksen johdon pätevyys kyseenalaistettiin.

Ettinen sokeus johtaa meidät toimimaan sokeasti sääntöjen mukaisesti sekä asettamaan ihmisille hyvin erilaisia moraalisia kriteereitä kuin koneille: esimerkiksi ihmishoitajan maine vaikuttaa hänen toimintansa arviointiin, mutta tietokoneen ei. (Laakasuo & Palomäki 2018, 51–52.) Tämä on kuitenkin myös perusteltua, sillä riippumattoman moraalisen toimijuuden olettaminen tekoälylle on vaarallista. Todellisuudessa nimenomaan ihminen luo ja kehittää tekoälyä. (Johnson, 2006, 204.)

6. Tekoälyn soveltamiseen liittyviä eettisiä suosituksia

Jotta tietyn tekoälysovelluksen relevanttiutta voidaan arvioida tutkimuksen tai kehittämistyön taustalla, on eettisesti työhönsä suhtautuvan tutkijan seurattava (1) sen ohjelmistoon liittyvää validaatiotutkimusta, (2) sitä käyttäen tehtyjä tieteellisiä julkaisuja ja käytännön raportteja, (3) sen taustalla olevan metodin tai teknologian (tässä kasvonilmeiden tunnistukseen perustuvan tunnereaktioanalyysin) yleisempää kehitystä ja siihen liittyvää tutkimusta sekä (4) taustalla olevasta teoriasta käytyä keskustelua.

Tekoälyn käytössä hyvät tavoitteet tai aiheet eivät riitä, sillä ne voivat muuttua. Olisi pikemminkin kehitettävä hyviä hallintotapoja. (Tegmark, 2018, 201.) Avuksi voisi tulla myös yleisempi normisto. Osa ammattikunnistahan on onnistunut luomaan selkeitä eettisiä normeja. Toisin kuin vaikkapa lääketieteen etiikassa, tekoälyn etiikassa ei ole riittävän yleisiä tavoitteita, velvoitteita eikä menetelmiä, joiden avulla periaatteet siirretään käytäntöön. Laillisia ja ammatillisia vastuullisuusmekanismeja ja normeja ei myöskään vielä ole. Tämä johtaa muun muassa kilpailutilanteisiin yleisten ja yksityisten intressien kesken. Euroopan komission tekoälyryhmä suosittelee neljää eettistä periaatetta: ihmisen autonomian kunnioittamista, harmin välttämistä, reiluutta ja selitettävyyttä. (Mittelstadt, 2019.) Nämä ovat hyvä lähtökohta, mutta lisäksi tarvitaan konkreettisempia eettisiä periaatteita ja käytäntöjä, jotka muodostuisivat samalla tavalla itsestäänselvyyksiksi kuin monien ammattikuntienkin etiikka.

Tällä hetkellä tekoälyn käytölle ei ole riittävää ja selkeää regulaatiota eli sääntelyä, eikä tekoälyteknologian käytön varmistusmekanismejakaan ole riittävästi. EU on yrittänyt ottaa tässä portinvartijan roolia: muun muassa ”Ethics Guidelines for Trustworthy Artificial Intelligence” (2019) on hyvä suunta. Silti teknologia kehittyy niin nopeasti, että tällaiset pyrkimykset uhkaavat olla liian hitaita. (Chochia & Nässi, 2021.) Joka tapauksessa laajempien instituutioiden vaatima ja asettama sääntely on välttämätön osa kokonaisuutta, vaikka se ei yksinään kykenekään haasteita ratkaisemaan.

Toisaalta on huomattava, että myös eettiset periaatteet voivat olla ristiriidassa keskenään eivätkä ne kykene vastaamaan kaikkiin kysymyksiin. Eettisiä periaatteita pitää aina tulkita ja painottaa. (Beauchamp & Childress, 2009; Siipi, 2017.) Näin ollen eettinen arviointi on aina laajempi kysymys kuin tietyt eettiset periaatteet (Siipi, 2017). Osa filosofisista ja moraalisisista ongelmista on tosi syviä ja

vaikeasti ratkaistavia (Harari, 2022), joten helppoja tai yksinkertaisia ratkaisuja ei useinkaan ole.

Uutena, täydentävänä eettisenä kehyksenä tunnetekoälyn ja ihmisen toimijuuden arvioimiseen voidaan ehdottaa Brey'n mainitsemaa *struktuurista eli rakenteellista etiikkaa*. Siinä keskitytään laajempiin rakenteisiin ja verkostoihin, sosiaalisten ja materiaalien asetelmien moraalisiin aspekteihin ja siihen millaisia vaikutuksia niillä on ihmisten toimijuuteen. Etuna on, että tällainen etiikka täydentää individualistista etiikkaa. Strukturaalinen etiikka 1) analysoi olemassa olevien asioiden ja asetelmien moraalisia seurauksia; 2) arvioi niiden moraalista hyvyttä tai sopivuutta ja 3) kuvaa normatiivisesti moraalisesti toivottavia asetelmia tai sitä, miten nykyisiä asetelmia pitäisi muuttaa. (Brey, 2014, 135.)

Tällainen etiikka on perusteltua, sillä vaikka eettisesti tarkastellaan yleensä yksilötoimijaa, ihmisten välinen toiminta on pääsääntöisesti yhteisöllistä. Valtaosa ihmisen kohtaamista tilanteita on ryhmätilanteita ja niissä ilmenevä toiminta yhteistoimintaa. (Hallamaa, 2017, 97.) Varsinkin tekoälyn kohdalla yksilöllisen etiikan painotus ei riitä, sillä laajemmat järjestelmät ja rakenteet määrittävät vahvasti tekoälyn periaatteita, soveltamista ja niiden seurauksia yksilöille, ryhmille ja niiden toimijuudelle.

Strukturaalisessa etiikassa ei tarvitse pitää teknologisia artefakteja moraalina toimijoina sinänsä, mutta niillä voidaan nähdä olevan tärkeitä moraalisia rooleja. Niiden moraaliset roolit eroavat kuitenkin ihmisen moraalista rooleista. Vaikka artefakti voi olla syy moraalisesti hyviin tai huonoihin seurauksiin, tällaiset seuraukset ovat silti aina riippuvaisia laajemmasta ihmisten ja asioiden verkostosta. (Brey, 2014, 126, 136.)

Toimintamme tekoälyn kanssa on jo muuttanut käsitystämme toimijuudestamme. Ratkaisuna on siis tuskin vain tekoölyyn liittyvien eettisten periaatteiden listaus, sillä ne ovat abstrakteja ideaaleja, joita on vaikea soveltaa todellisessa käytössä (ks. myös Brožek & Janik, 2019, 104). Sen sijaan tulisi pyrkiä syvempään ymmärrykseen siitä miten teknologiset prosessit muuttavat maailmaa. Tekoöly keskittyy instrumentaalisiin hyötyihin, kun sen sijaan moraalinen vastuullisuus on meille ihmisille tyypillinen piirre. Meidän tehtäväksi jää siten kehittää tekoöly, joka palvelee moraalisesti hyviä päämääriä. Jos tämä ei onnistu, pahimmillaan tekoälyn instrumentaaliset arvot voivat syrjäyttää ihmisten sisäiset moraaliset arvot. Tärkeä osa tätä tehtävää on suojella ja vahvistaa ihmistoimijuuden ehtoja kaikissa tekoöly-ympäristöissä. (Hallamaa & Kalliokoski, 2020, 55, 63.) Tässä prosessissa on huomioitava sekä tekoälyn suunnittelija, itse tekoöly, että myös tekoälyn käyttäjä (Johnson, 2006).

Tässä artikkelissa on oletettu ihmisen toimijuus ja moraalinen toimijuus lähtökohtaisesti hyvänä ja kannatettavana asiana, joita tunnetekoälyn tulisi lisätä. Tunnetekoölyä hyödyntävässä tutkimuksessa monimenetelmäisyys näyttäytyy lähes väistämättömältä valinnalta, jotta ihmisen toimijuus-näkökulma säilyy mukana. Toimijuus on kuitenkin tosiasiallisesti mutkikkaampi eettinen käsite kuin mitä tässä kuvataan, sillä tässä artikkelissa esitetyissä toimijuuden määritelmässä ei määritellä sitä mihin (moraaliseen hyvään tai pahaan) toimijuus kohdistuu. Näin ollen rikollisillakin voi olla vahva toimijuus, mutta harvoin ajattelemme sen olevan vahvistamisen arvoinen tai yhteistä hyvää tuottava seikka.

Dilemmana on, että älykkäät koneet pitää havainnoida moraalisesti relevantteina siitäkin huolimatta, ettei niillä ole varsinaista moraalitajua (Laakasuo & Palomäki, 2018). Toisaalta tekoölyä ei voida pitää moraalisenä agenttina tai vastuullisena

toimijana, jolloin ihmisen ja tekoälyn yhteistoimintakaan ei ole varsinaista kollektiivista toimintaa (Hallamaa & Kalliokoski, 2020, 61). Näin ollen tekoäly ei voi saavuttaa eettisesti korkeatasoista "me-tilaisuutta", jota Hallamaa (2017, 115, 218) pitää tärkeimpänä ja tehokkaimpana yhteistoiminnallisuuden muotona.

Me-tilassa yhteistoimintaan osallistuvilla on aidosti yhteinen tavoite ja halu toimia yhdessä. Yhteistoimintaan osallistuvat muokkaavat tavoitteita siten, että yhteisön tavoitteesta tulee kaikkien tavoite. Tällöin yhdessä toimiminen on jo osa toiminnan tavoitetta. Tässä on kuitenkin varottava me-tilan muuttumista vahingolliseksi symbioottiseksi tilaksi, jossa yksilöllisten päämäärien asettaminen vaikeutuu. (Hallamaa, 2017, 115, 221.) Näin ollen lähes kaikki eettinen arviointi kohdistuu jollakin tavalla yksilön ja yhteisön väliseen tasapainoon. Ryhmävuorovaikutustilanteissa osallistujat kokevatkin dialogisuuden samanaikaisesti yksilön kunnioituksena ja ryhmän onnistuneena ilmapiirinä ja yhteistoimintana (Gjerstad, 2022).

Me-tila on yhteistoimijuudelle välttämätön tila, jota tekoäly ei kykene saavuttamaan. Tämän vuoksi tekoälyn soveltamisessa on oltava varovainen. Paradoksaalisesti tunnetekoäly – esimerkiksi tunnereaktioiden analyysi ryhmävuorovaikutuksesta - voi kuitenkin auttaa ryhmää lisäämään "me-tilaisuutta" esimerkiksi paljastamalla ihmisten tai ihmisryhmien toiminnasta jotakin sellaista, jonka löytämisessä tai tulkitsemisessä ihmisarvio voisi olla vaikeaa, mahdotonta tai epäeettistä kuten tiettyihin puheenaiheisiin liittyviä erityisen vahvoja tunnereaktioita.

Tunnetekoäly voi siis paljastaa ihmisyksilöiden tai -ryhmien toimijuudesta sellaisia säännönmukaisuuksia, jotka voivat hyödyttää yksilöä, ryhmää tai laajempaa yhteisöä tai yhteiskuntaa. Tämä on eettisesti perusteltua kuitenkin vain silloin, kun yksilön tai ryhmän toimijuuden ja yksityisyyden suojelusta onnistutaan huolehtimaan tutkimuksen tai ylipäättään tunnetekoälyn käytön aikana. Tämä ei kuitenkaan itsestään selvästi lisää yksilön tai ryhmän toimijuutta, vaan sitä on tutkittava erikseen.

Ryhmävuorovaikutustutkimuksissani tällaiset hyödyt konkretisoituvat stimulated recall -tyyppisissä reflektiotilanteissa, joissa osallistujat saavat katsoa omia, alkuperäisiä vuorovaikutusvideoitaan ja nähdä myös tutkijan tekemiä analyyseja. Hyöty voi olla vaikkapa siinä, että tutkija on koonnut ryhmän tunnetekoälyn poimimista tunnereaktioista vahvimmat piikit ja verrannut niitä vuorovaikutuksen sisältöön ja muotoon, ja ryhmä voi keskustellen löytää näiden pohjalta vahvuuksia ja kehittämisen kohteita vuorovaikutuksestaan.

Toisaalta tällaisella tunnetekoälyllä voi olla vaikutuksia myös yksilöiden välisiin (valta)suhteisiin, kun yksilö ei enää kykene kontrolloimaan itse sitä mitä hänestä tulkitaan (Korja, 2016, 160–161). Voidaan myös perustellusti kysyä, tarvitsemmeko me ihmiset kaikkea tätä uutta tutkimusta ja ennen kaikkea sen tarkkuustasoa? Esimerkiksi Kiinassa kasvonilmetunnistusta käytetään jo oppilaiden vireystason tarkkailuun (Crampton, 2019, 61). Kaupalliset sovellukset diktatuuriin kallellaan olevien valtioiden sovelluksista puhumattakaan eivät välttämättä piittaa sovellusten ja ihmisten oikeuksien välisestä ristiriidasta. Näin ollen dilemma eettisesti käytettyjen ja ihmisten toimijuutta lisäävien tekoälysovellusten ja niiden epäeettisen, jopa sortavan käytön välillä jää vaille ratkaisua. Sovellus yksinään ei tosin sanoen ratkaise tekoälysovelluksen eettisyyttä, vaan viime kädessä kyse on ihmisen valinnasta.

Kannatan samanaikaisesti tunnetekoölyn ammattilaisten eettisen normiston juurruttamista, hyviä hallintotapoja sekä eettisten riskien ennakoivaa tiedostamista ja varautumista. Strukturaalinen etiikka yhdistettynä yksilöllisen toimijuuden ja yhteisöjen eettisen me-tilaisuuden painotukseen mahdollistaa riittävän laajan eettisen näkökulman. Tämä tarkoittaa sitä, että eettinen arviointi on monimutkaista.

Lähteet

A European approach to artificial intelligence. European Commission. 6.3.2023. <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

Aaltonen, K. & Eskola-Kronqvist, A. (2011). Menetelmäopista pedagogiseen ajatteluun. Teoksessa K. Aaltonen & A. Eskola-Kronqvist (toim.), *Meidän Helmet. Aikuispedagogiikan ja -koulutuksen hyviä käytänteitä, uusia toimintamalleja ja välineitä* (ss. 40-51). Hämeen ammattikorkeakoulu, Ammatillinen opettajakorkeakoulu.

Ailisto, H. (toim.), Heikkilä, E., Helaakoski, H., Neuvonen, A. & Seppälä, T. (2018). *Tekoälyn kokonaiskuva ja osaamiskartoitus*. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 46/2018. <http://urn.fi/URN:ISBN:978-952-287-549-5>

Ailisto, H. (toim.), Neuvonen, A., Nyman, H., Halén, M. & Seppälä, T. (2019). *Tekoälyn kokonaiskuva ja kansallinen osaamiskartoitus – loppuraportti*. Valtioneuvoston kanslia 15.1.2019. <https://www.etla.fi/julkaisut/tekoalyn-kokonaiskuva-ja-kansallinen-osaamiskartoitus-loppuraportti/>

Atjonen, P. (1996). "Pitäs olla opella viis silimää". Video- ja observointimateriaali opettajaksi opiskelevan didaktisen ajattelun kehittämisessä. *Kasvatus*, 27(3), 228–237.

Aunimo, L., Gjerstad, E. & Raulinaitis, V. (2021). *How to foster dialogicality in group interaction?* ICERI-konferenssin artikkeli, 8431–8437. <https://doi.org/10.21125/iceri.2021.1944>

Bales, R. F. (1950). *Interaction Process Analysis. A Method for the Study of Small Groups*. Addison-Wesley Press.

Bales, R.F. & Cohen, P. (1979). *SYMLOG: A system for the multiple level observation of groups*. Free Press.

Barrett, L.F., Adolphs, R., Marsella, S., Martinez, A. M. & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, 20 (1), 1–68. <https://doi.org/10.1177/1529100619832>

Beauchamp T. L, Childress J, F. (2009). *Principles of biomedical ethics*. 6th ed. Oxford University Press.

Bossmann, J. (2016). Top 9 ethical issues in artificial intelligence. World Economic Forum. 21.10.2016. <https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/>

Brey, P. (2014). From Moral Agents to Moral Factors: The Structural Ethics Approach. In P. Kroes & P-P. Verbeek (Eds.) *The Moral Status of Technical Artefacts*. Philosophy and Technology 17 (pp. 125–142.) Springer.

Brey, P. (2018). The strategic role of technology in a good society. *Technology in Society*, 52, 39-45. <https://doi.org/10.1016/j.techsoc.2017.02.002>

Brožek, B. & Janik, B. (2019). Can artificial intelligences be moral agents? *New Ideas in Psychology*, 54, 101-106. <https://doi.org/10.1016/j.newideapsych.2018.12.002>

Bänziger, T., Grandjean, D. & Scherer, K. (2009). Emotion Recognition From Expressions in Face, Voice, and Body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, 9 (5), 691-704. <https://doi.org/10.1037/a0017088>

Calvo, M.G., Fernández-Martin, A., Recio, G. & Lundqvist, D. (2018). Human Observers and Automated Assessment of Dynamic Emotional Facial Expressions: KDEF-dyn Database Validation. *Frontiers in Psychology*, 26 (9), 1–12. <https://doi.org/10.3389/fpsyg.2018.02052>

Cameron, J. E. (2004). A Three-Factor Model of Social Identity. *Self and Identity*, 3, 239–262. <https://doi.org/10.1080/13576500444000047>

Chochia, A. & Nässi, T. (2021). Ethics and emerging technologies – facial recognition. *IDP. Internet, Law and Politics E-Journal*, 34. <http://dx.doi.org/10.7238/idp.v0i34.387466>

Cojocea, M.-E., Berbaru, R.-C., Boiangiu, C.-A., Bran, M., Aprodu, C.A., Hurezeanu, F.C. & Sicaru, I.A. (2019). Smile-O-Meter: A Pilot Project to Monitor a Patient's Emotional Changes Through an On-Line Therapy Session. *Journal of Information Systems & Operations Management*, 13 (1), 113–124.

- Corneanu, C. A., Simón, M. O., Cohn, J. F. & Guerrero, S. E. (2016). Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 38 (8), 1548-1568. <https://doi.org/10.1109/TPAMI.2016.2515606>
- Crampton, J.W. (2019). Platform Biometrics. *Surveillance & Society*, 17 (1/2), 54–62. <https://doi.org/10.24908/ss.v17i1/2.13111>
- Darwiche, J., de Roten, Y., Stern, D.J., von Roten, F.C., Corboz-Warnery, A. & Fivaz-Depeursinge, E. (2008). Mutual Smiling Episodes and Therapeutic Alliance in a Therapist-Couple Discussion Task. *Swiss Journal of Psychology*, 67 (4), 231–239. <https://doi.org/10.1024/1421-0185.67.4.231>
- Dekkers, T. (2019). Technology driven crimmigration? Function creep and mission creep in Dutch migration control. *Journal of Ethnic and Migration Studies*, 46 (9), 1849-1864. <https://doi.org/10.1080/1369183X.2019.1674134>
- Dempsey, N.P. (2010). Stimulated Recall Interviews in Ethnography. *Qualitative Sociology*, 33, 349–367. <https://doi.org/10.1007/s11133-010-9157-x>
- Dupré, D., Krumhuber, E.G., Küster, D. & McKeown, G.J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLOS ONE*, 15(4), e0231968. <https://doi.org/10.1371/journal.pone.0231968>
- Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, 6 (3-4), 169–200. <https://doi.org/10.1080/02699939208411068>
- Ekman, P. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science*, 11 (1), 31-34. <https://doi.org/10.1177/17456916155969>
- Ellsworth, P.C. (2013). Appraisal Theory: Old and New Questions. *Emotion Review*, 5 (2), 125–131. <https://doi.org/10.1177/1754073912463617>
- Eteläpelto, A., Vähäsantanen, K., Hökkä, P. & Paloniemi, S. (2017). Tutkimus- ja kehittämishankkeen tausta ja lähtökohdat. Teoksessa Vähäsantanen, K., Paloniemi, S., Hökkä, P. & Eteläpelto, A. *Ammatillinen toimijuus. Rakenne, mittari ja tuki* (ss. 5–13). Jyväskylän yliopisto.
- Eteläpelto, A., Kykyri, V.-L., Penttonen, M., Hökkä, P., Paloniemi, S., Vähäsantanen, K., Eteläpelto, T., & Lappalainen, V. (2018). A multi-componential methodology for exploring emotions in learning : using self-reports, behaviour registration, and physiological indicators as complementary data. *Frontline Learning Research*, 6 (3), 6-36. <https://doi.org/10.14786/flr.v6i3.379>
- Ethics guidelines for trustworthy AI*. (2019). European Commission. 18.4.2019. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Fry, H. (2018) Hello world. *Kuinka selviytyä algoritmien aikakaudella*. Suom. J. Iso-Markku. Bazar.
- Galinsky, A. D., Rus, D. & Lammers, J. (2011). Power: A Central Force Governing Psychological, Social, and Organizational Life. In De Cremer, D., van Dick, R. & Murnighan, J. K. (Eds.) *Social Psychology and Organizations* (pp. 17-38). Routledge.
- Gazdag, E., Nagy, K., & Szivák, J. (2019). “I Spy with My Little Eyes...” The use of video stimulated recall methodology in teacher training – The exploration of aims, goals and methodological characteristics of VSR methodology through systematic literature review. *International Journal of Educational Research*, 95, 60-75. <https://doi.org/10.1016/j.ijer.2019.02.015>
- Gjerstad, E. (2011). Arvokasvatuksen mahdollisuudet ja rajat. Teoksessa Purjo, T. & Gjerstad, E. (toim.) *Arvojen ja tarkoitusten pedagogiikka* (ss. 75-94). Non Fighting Generation ry.
- Gjerstad, E. (2021). Tutkittua tietoa vuorovaikutus- ja ryhmäprosesseista. *eSignals* 7.5.2021. Haaga-Helia. <https://esignals.fi/teemat/tutkittu/tutkittua-tietoa-vuorovaikutus-ja-ryhmaprosesseista/#682255a3>
- Gjerstad, E. (2022). *Tunteet pelissä – ryhmänohjausvuorovaikutustutkimus biometrisiä menetelmiä hyödyntäen* (VALAA – Vaikuttava ja laadukas ohjaus opettajan työssä -projektin tulokset). eSignals Pro toukokuu 2022. <https://julkaisut.haaga-helia.fi/valaa-vaikuttava-ja-laadukas-ohjaus-opettajan-tyossa/>

- Haakana, M. (2002). Laughter in medical interaction: From quantification to analysis, and back. *Journal of Sociolinguistics*, 6 (2), 207–235. <https://doi.org/10.1111/1467-9481.00185>
- Haataja, M., van de Fliert, L. & Rautio, P. (2020). Public AI Registers. *Realising AI transparency and civic participation in government use of AI*. Version 1.0. City of Helsinki. https://uploads-ssl.webflow.com/5c8abedb10ed656ecfb65fd9/5f6f334b49d5444079726a79_AI%20Registers%20-%20White%20paper%201.0.pdf
- Haidt, J. & Keltner, D. (1999). Culture and Facial Expression: Open-ended Methods Find More Expressions and a Gradient of Recognition. *Cognition & Emotion*, 13 (3), 225–266. <https://doi.org/10.1080/026999399379267>
- Hallamaa, J. (2017). *Yhdessä toimimisen etiikka*. Gaudeamus.
- Hallamaa, J. & Kalliokoski, T. (2020). How AI Systems Challenge the Conditions of Moral Agency? In M. Rauteberg (Ed.), *Culture and Computing* (pp. 54-64). Springer.
- Harari, Y. N. (2022). *The Most Important Skills for the Future of Work*. Nordic Business Forum -esitelmä Helsingissä 21.9.2022.
- Hennissen, P., Crasborn, F., Brouwer, N., Korthagen, F., & Bergen, T. (2010). Uncovering contents of mentor teachers' interactive cognitions during mentoring dialogues. *Teaching and Teacher Education*, 26 (2), 207–214. <https://doi.org/10.1016/j.tate.2009.02.022>
- Hjelt, Y. (2019). Poliisi ja Tulli saivat oikeuden automaattiseen kasvojen tunnistamiseen ihmisvirrasta – lupa on, mutta laitteet puuttuvat. Viranomaiset uskovat, että rikostutkinta nopeutuu automaattisen kasvojentunnistuksen myötä. Yle 5.6.2019. <https://yle.fi/uutiset/3-10815487>
- Huhtamäki, H. & Lehtinen, R. (2016). *Oivaltamisen hetket pariterapia-asiakkaiden jälkihaastatteluissa*. [Pro gradu –tutkielma, Jyväskylän yliopisto].
- Human Behavior*. Pocket Guide (2017). Biometric Research Platform. *iMotions*.
- Hussin, V. (2013). Student and teacher reflections on indirectness as a pragmatic feature of pharmacist–patient simulations. *English for Specific Purposes*, 32, 110–121. <https://doi.org/10.1016/j.esp.2013.01.001>
- Ihmiseen kohdistuvan tutkimuksen eettiset periaatteet ja ihmistieteiden eettinen ennakoarviointi Suomessa* (2019). Tutkimuseettisen neuvottelukunnan ohje. Tutkimuseettisen neuvottelukunnan julkaisuja 3/2019. TENK. https://tenk.fi/sites/default/files/2021-01/Ihmistieteiden_eettisen_ennakoarvioinnin_ohje_2020.pdf
- Jain, A.K. & Ross, A. (2008). Introduction to Biometrics. In K. Anil, P.F. Jain & A.A. Ross (Eds.) *Handbook of Biometrics* (pp. 1-22). Springer.
- Janusz, B. & Peräkylä, A. (2021). Quality in conversation analysis and interpersonal process recall. *Qualitative Research in Psychology*, 18 (3), 426–449. <https://doi.org/10.1080/14780887.2020.1780356>
- Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology*, 8, 195–204. <https://doi.org/10.1007/s10676-006-9111-5>
- Juutilainen, P-K. (2003). *Elämään vai sukupuoleen ohjausta?* Tutkimus opinto-ohjauskeskustelun rakentumisesta prosessina. Joensuun yliopiston kasvatustieteellisiä julkaisuja No. 92. [Väitöskirja, Joensuun yliopisto]. <http://urn.fi/URN:ISBN:952-458-330-5>
- Järvelä, S., Malmberg, J. & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction*, 42, 1–11. <https://doi.org/10.1016/j.learninstruc.2015.10.006>
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M. & Kirscher, P.A. (2019). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*. <https://doi.org/10.1016/j.learninstruc.2019.04.004>
- Kagan, N., Krathwohl, D.R. & Miller, R. (1963). Stimulated Recall in Therapy Using Video Tape-A Case Study. *Journal of Counseling Psychology*, 10 (3), 237–243. <https://doi.org/10.1037/h0045497>

- Kagan, N., & Schauble, P. G. (1969). Affect simulation in interpersonal process recall. *Journal of Counseling Psychology*, 16 (4), 309–313. <https://doi.org/10.1037/h0027718>
- Kankaanpää, A. & Savolainen, R. (2012). *Vuorovaikutuksen dominanssi ja dialogisuus masennuksen pariterapiassa. "Te voitte ihan valita mistä te puhutte"*. [Pro gradu –tutkielma, Jyväskylän yliopisto]. <http://urn.fi/URN:NBN:fi:juyu-201301271125>
- Kim, T., McFee, E. K., Olguin, D. O., Waber, B. & Pentland, A. (2012). Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior*, 33 (3), 412–427. <https://doi.org/10.1002/job.1776>
- Kelman, H.C. (2006). Interests, Relationships, Identities: Three Central Issues for Individuals and Groups in Negotiating Their Social Environment. *Annual Review of Psychology*, 57, 1–26. <https://doi.org/10.1146/annurev.psych.57.102904.190156>
- Koivisto, I. (2023). Transparency goes digital in the EU –What is it good for? *Open Government in the EU*. 3.2.2023. <https://www.eu-opengovernment.eu/?p=3627>
- Korja, J. (2016). *Biometrinen tunnistaminen ja henkilötietojen suoja*. Tutkimus biometrinen tunnistamisen lainsäädännöllisestä asemasta. Acta Universitatis Lapponiensis 325. [Väitöskirja, Lapin yliopisto]. <https://urn.fi/URN:ISBN:978-952-484-900-5>
- Krumhuber, E.G., Kappas, A. & Manstead, A.S.R. (2013). Effects of Dynamic Aspects of Facial Expressions: A Review. *Emotion Review*, 5 (1), 41–46. <https://doi.org/10.1177/1754073912451349>
- Krumhuber, E.G., Küster, D., Namba, S. & Skora, L. (2020). Human and machine validation of 14 databases of dynamic facial expressions. *Behavior Research Methods*, 53, 686–701. <https://doi.org/10.3758/s13428-020-01443-y>
- Krywko, J. (2017). The premature quest for AI-powered facial recognition to simplify screening. "This technology at the airport... is premature. It's not the right way to go." *Ars Technica* 6.2.2017. <https://arstechnica.com/information-technology/2017/06/security-obsessed-wait-but-can-ai-learn-to-spot-the-face-of-a-liar/>
- Kumar, A., Kaur, A. & Kumar, M. (2019). Face detection techniques: a review. *Artificial Intelligence Review*, 52, 927–948. <https://doi.org/10.1007/s10462-018-9650-2>
- Laakasuo, M. & Palomäki, J. (2018). Robotiikan moraalipsykologian näkökulmia yhteiskuntaan ja työelämään. *Työn tuuli*, 1, 44–54.
- Laine, T. (2004). *Huomisen opettajat. Luokanopettajakoulutus ammatillisen identiteetin rakentajana*. [Väitöskirja, Tampereen yliopisto]. <https://urn.fi/urn:isbn:951-44-5996-2>
- Lassalle, A., Pigat, D., O'Reilly, H., Berggen, S., Fridenson-Hayo, S. & Tal, S., Elfström, S. Råde, A., Golan, O., Bölte, S., Baron-Cohen, S. & Lundqvist, D. (2019). *The EU-Emotion Voice Database*. *Behavior Research Methods*, 51, 493–506. <https://doi.org/10.3758/s13428-018-1048-1>
- Lederman, O. Mohan, A., Calacci, D. & Pentland, A. (2018). Rhythm: A Unified Measurement Platform for Human Organizations. *IEEE MultiMedia*, 25 (1), 26–38. <https://doi.org/10.1109/MMUL.2018.112135958>
- Leiman, M. (2015). Dialoginen ohjaus. Teoksessa P. A. Kauppila, J. Silvonen & M. Vanhalakka-Ruoho (toim.) *Toimijuus, ohjaus ja elämäntulkku* (ss. 57–68). Publications of the University of Eastern Finland. Reports and Studies in Education, Humanities, and Theology. Itä-Suomen yliopisto.
- Lepri, B., Nuria, O. & Pentland, A. (2021). Ethical machines: The human-centric use of artificial intelligence. *iScience*, 24 (3), 102249. <https://doi.org/10.1016/j.isci.2021.102249>
- Logren, A. (2019). *The Management of Experience as a Platform for Social Influence in Health Promotion Groups*. Tampere University Dissertations 112. [Väitöskirja, Tampereen yliopisto]. <https://urn.fi/URN:ISBN:978-952-03-1208-4>
- Mac, R., Haskins, C., Sacks, B. & McDonald, L. (2021). Surveillance Nation. BuzzFeed.News. 9.4.2021. <https://www.buzzfeednews.com/article/ryanmac/clearview-ai-local-police-facial-recognition?bfsource=relatedmanual>
- Martinez, B., Valstar, M. F., Jiang, B. & Pantic, M. (2019). Automatic Analysis of Facial Actions: A Survey. *IEE Transactions on Affective Computing*, 10 (3), 325–347. <https://doi.org/10.1109/TAFFC.2017.2731763>

- Mauss, I. B., Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, 23 (2), 209–237. <https://doi.org/10.1080/02699930802204677>
- McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, 7 (1), 1–12. <https://doi.org/10.1177/2053951720904386>
- Mellenius, N. (2013) Vuorovaikutuksen hoito on lapsuuteen vaikuttamista. Lapsi Videoavusteisessa vuorovaikutuksen ohjauksessa. Teoksessa N. Mellenius & N. Remu (toim.) *Vuorovaikutus kuvaassa. Videoavusteisen ohjauksen eettisyys ja käytäntö* (ss. 75-94). Mannerheimin Lastensuojeluliiton Lasten ja Nuorten Kuntoutussäätiö.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1 (11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Moors, A. (2014). Flavors of Appraisal Theories of Emotion. *Emotion Review*, 6 (4), 303–307. <https://doi.org/10.1177/1754073914534477>
- Murray, B. (2019). *System 1 and System 2: Facts and Fictions*. iMotions blog. <https://imotions.com/blog/system-1-and-system-2/>
- Mäkelä, J., Mellenius, N. & Remu, N. (2013). Videointerventioiden etiikan pohdintaa. Teoksessa N. Mellenius & N. Remu (toim.), *Vuorovaikutus kuvaassa. Videoavusteisen ohjauksen eettisyys ja käytäntö* (ss. 21-36). Mannerheimin Lastensuojeluliiton Lasten ja Nuorten Kuntoutussäätiö.
- Nummenmaa, L. (2010) *Tunteiden psykologia* (2. painos). Tammi.
- Nummenmaa, L. (2016). Tunteiden neurobiologia. *Suomen Lääkärilehti*, 71 (10), 725–731.
- Nummenmaa, L. (2017). Mistä puhumme kun puhumme tunteista? *Tieteessä tapahtuu*, 35 (2), 35-39.
- Nummenmaa, L. (2019). *Tunnekartasto. Kuinka tunteet tekevät meistä ihmisiä*. Tammi.
- Pasanen, H. (2014). Opinto-ohjauksen kehittämisen dilemmat – eli kuinka talous tunkee ohjaukseen. Teoksessa P.-K. Juutilainen, H. Pasanen & M. Alanko-Turunen (toim.) *Arvokas ohjaus* (11-17). Haaga-Helian julkaisut. Kehittämisasiaportit ja tutkimukset.
- Pedaforum 2022 – Korkeakoulupedagogiikan päivät. J. Laukia (toim.) Haaga-Helia julkaisut (2022). <https://julkaisut.haaga-helia.fi/pedaforum-2022-korkeakoulupedagogiikan-paivat/>
- Pekrun, R., Goetz, T., Titz, W. & Perry, R.P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, 37 (2), 91–106. https://doi.org/10.1207/S15326985EP3702_4
- Pekrun, R. (2017). Emotion and Achievement During Adolescence. *Child Development Perspectives*, 11 (3), 215–221. <https://doi.org/10.1111/cdep.12237>
- Pentland, A. (2012). The New Science of Building Great Teams. *Harvard Business Review*, April, 1 – 11.
- Peräkylä, A. (2009). Mieli sosiaalisessa vuorovaikutuksessa. *Sosiologia*, 46 (4), 251–268.
- Peräkylä, A., Voutilainen, L., Henttonen, P., Kahri, M., Stevanovic, M., Sams, M. & Ravaja, N. (2016). Tarinankerronnan psykofysiologiaa. *Sosiologia* 53 (3), 258–274.
- Pesonen, E. (2018). *Tekoäly ja biometriikka neuromarkkinoinnin tutkimusvälineenä*. [AMK-opinnäytetyö: Haaga-Helia ammattikorkeakoulu]. <https://urn.fi/URN:NBN:fi:amk-2018120219472>
- Pietikäinen, M. & Silvén, O. (2019). *Tekoälyn haasteet – Koneoppimisesta ja konenäöstä tunnetekoälyyn*. Konenäön ja signaalianalyysin keskus. Oulun yliopisto. <http://urn.fi/urn:isbn:9789526224824>
- Pihlström, S. (2021). Ihmisen maailma. Esseitä humanismista, totuudesta ja ajattelun tilasta. *Niin & näin*.
- Poria, S., Cambria, E., Bajpai, R. & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>

- Reed, L.I., Zeglen, K.N. & Schmidt, K.L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. *Evolution and Human Behavior*, 33 (3), 200–209. <https://doi.org/10.1016/j.evolhumbehav.2011.09.003>
- Rennie, D. L. (2001). The client as a self-aware agent in counselling and psychotherapy. *Counselling and Psychotherapy Research*, 1 (2), 82-89. <https://doi.org/10.1080/14733140112331385118>
- Rissanen, V. (2017). Mikä tekee festarikävijän onnelliseksi? Ruisrock selvitti tunteiden kirjon videoanalytiikalla. *Helsingin Sanomat* 3.10.2017. <https://www.hs.fi/teknologia/art-2000005392883.html>
- Rober, P. (2005). The Therapist's Self in Dialogical Family Therapy: Some Ideas About Not-Knowing and the Therapist's Inner Conversation. *Family Process*, 44 (4), 477–495. <https://doi.org/10.1111/j.1545-5300.2005.00073.x>
- Ryan, J. & Gass, S. (2012). Stimulated recall. In R. Barnard & A. Burns (Eds.) *Teacher cognition and practice: International case studies to illuminate methodological issues* (pp. 144-161). Multilingual Matters.
- Salmela, M. (2017). Affektiivinen käänne. Yksi vai monia? Käsitteellisiä ja metodologisia pohdintoja. *Tieteessä tapahtuu*, 35 (2), 32–35.
- Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C-H., Xiang, Y. & He, J. (2019). A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. *Sensors*, 19 (1863). <https://doi.org/10.3390/s19081863>
- Saranpää, M. (2019). Oppimisen analysointia keinoälyn valossa, jonkin järjen talossa. *eSignals* 29.4.2019. Haaga-Helia. <https://esignals.fi/kategoria/pedagogiikka/oppimisen-analysointia-keinoalyn-valossa-jonkin-jarjen-talossa>
- Scherer, K.R., Ellgring, H., Dieckmann, A., Unfried, M. & Mortillaro, M. (2019). Dynamic Facial Expression of Emotion and Observer Inference. *Frontiers in Psychology*, 10, article 508. <https://doi.org/10.3389/fpsyg.2019.00508>
- Siipi, H. (2017). Tutkimuksen ennakoarviointi filosofisen etiikan näkökulmasta. *Duodecim*, 133, 582–586.
- Stevanovic, M. (2016). Keskusteluanalyysi ja kokeellinen vuorovaikutustutkimus. Teoksessa M. Stevanovic & C. Lindholm (toim.) *Keskusteluanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta* (ss. 390-409). Vastapaino.
- Stöckli, S. & Schulte-Mecklenbeck, M. & Borer, S. & Samson, A.C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, 50 (4), 1446–1460. <https://doi.org/10.3758/s13428-017-0996-1>
- Sutrop, M. & Laas-Mikko, K. (2012). From Identity Verification to Behavior Prediction: Ethical Implications of Second Generation Biometrics. *Review of Policy Research*, 29 (1), 21–36. <https://doi.org/10.1111/j.1541-1338.2011.00536.x>
- Tavares, G., Mourão, A. & Magalhães, J. (2016). Crowdsourcing facial expressions for affective-interaction. *Computer Vision and Image Understanding*, 147, 102–113. <https://doi.org/10.1016/j.cviu.2016.02.001>
- Tegmark, M. (2018). *Elämä 3.0. Ihmisenä olemisen tekoälyn aikakaudella*. K. Pietiläinen (suom.) Terra Cognita.
- Tuovila, S. (2005). *Kun on tunteet. Suomen kielen tunnesanojen semantiikkaa*. [Väitöskirja, Oulun yliopisto]. <http://urn.fi/urn:isbn:9514278070>
- van Kleef, G.A. & Fischer, A.H. (2016). Emotional collectives: How groups shape emotions and emotions shape groups. *Cognition and Emotion*, 30 (1), 3–19. <https://doi.org/10.1080/02699931.2015.1081349>
- Vanhalakka-Ruoho, M. (2014). Toimijuus elämäkulussa – ohjaustyön perusta? *Aikuiskasvatus*, 34 (3), 192–201. <https://doi.org/10.33336/aik.94099>
- Visala, A. (2021). Moraalinen toimijuus ja ihmiskeskeisyyden dilemma tekoälyn maailmassa. Teoksessa P. Raatikainen (toim.) *Tekoäly, ihminen ja yhteiskunta: Filosofisia näkökulmia* (ss. 157-180). Gaudeamus.
- van Noorden, R. (2020). The Ethical Questions That Haunt Facial-Recognition Research. *Nature*, 587 (7834), 354–358. <https://doi.org/10.1038/d41586-020-03187-3>

Voutilainen, L. (2016). Vuorovaikutusilmiöiden kvantifiointi. Teoksessa M. Stevanovic & C. Lindholm (toim.) *Keskusteluanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta* (ss. 372-389). Vastapaino.

Värri, V-M. (2002). Hyvä kasvatusta – kasvatusta hyvään. *Dialogisen kasvatuksen filosofinen tarkastelu erityisesti vanhemmuuden näkökulmasta*. Tampere University Press. <https://urn.fi/URN:ISBN:951-44-5558-4>

Värri, V-M. (2018). *Kasvatusta ekokriisin aikakaudella*. Vastapaino.

Wass, R. & Moskal, A.C.M. (2017). What can Interpersonal Process Recall (IPR) offer academic development?. *International Journal for Academic Development*, 22 (4), 293-306. <https://doi.org/10.1080/1360144X.2017.1340295>

Watson, D. & Clark, L. A. (1999). The PANAS-X: Manual for the Positive and Negative Affect Schedule – Expanded Form. University of Iowa. Department of Psychological & Brain Sciences Publications.

Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, 25(1), 163–194.

¹ VALAA:ssa muodostin tunnekokemusmittarin universaaleista perustunteista, joita ovat: ilo, suru, inho, pelko, viha, ylenkatse ja hämmästys (ks. esim. Nummenmaa 2016, samat vuorovaikutuslabrin iMotions AFFDEX-algoritmissa); VALAA-ohjajien näkemykset tärkeistä ohjausvuorovaikutukseen sisältyvistä tunteista, suomalaisten tärkeimmiksi nimeämistä tunteista (Tuovila 2005), PANAS-X:n (Watson & Clark, 1999), SYMLOG:n (Bales & Cohen, 1979), IPA:n (Bales, 1950) ja keskusteluanalyttisissä tutkimuksissa keskeisimmiksi oletetuista tunteista. Tunnekokemusmittaristossa mukana olleet tunteet olivat: ilo, innostus, suru, hämmästys, onnellisuus, pelko, ystävällisyys, epäystävällisyys, tyytyväisyys, viha, inho, alakuloisuus, ilahtuneisuus, pettymys, ahdistus, kateus, ymmärrys, hyväksyntä, yhteenkuuluvuus, toivo, välinpitämättömyys, epätoivo, myötätunto, helpotus, empatia, epäily, huoli, epävarmuus, tylsyyttä, halveksunta, vihamielisyys, aktiivisuus, luottamus, valppaus, tarkkaavaisuus, määrätietoisuus, kiinnostus, hermostuneisuus, jännittyneisyys (myönteinen), jännittyneisyys (kielteinen), kyllästyneisyys, uupumus, itsevarmuus, ujous, energisyys, syyllisyys, tyyneys, rentous, rauhallisuus, vapautuneisuus, kielteisyys, rohkeus, keskittyneisyys, vastenmielisyys, yksinäisyys, tarmokkuus, ärtyneisyys, ylpeys, häpeä, suuttumus, harmitus, yllättyneisyys, verkkaisuus, levollisuus, levottomuus, torjuvuus, turhautuneisuus ja päättäväisyys. Tunnekokemusmittaristossa käytin Likertin asteikkoa 1= en ollenkaan, 2= vähän, 3= kohtalaisesti, 4=melko paljon, 5=paljon kysymyksen: ”Mitä tunteita tai mielentiloja tunsit äskeisessä vuorovaikutustilanteessa (ei videointiin, vaan nimenomaan vuorovaikutustilanteeseen liittyen)?”

Olin myös huomannut VALAA:ssa, että siinä käytettyjen mittaristojen tunteet eivät kohdistuneet erityisesti oppimiseen ja ryhmävuorovaikutukseen. SOEVE:ssa sisällytin tunnekokemusmittaristoon nimenomaan oppimisessa keskeisimmiksi näyttäneitä tunteita: Näin ollen SOEVE:n tunnekokemusmittari rakentui Pekrunin (esim. Pekrun ja muut, 2002; Pekrun, 2017) oppimisessa tunnistamille tunteille sekä VALAA:ssa ryhmävuorovaikutuksessa vahvimmin osallistujien kokemille nimenomaan ryhmässä oloon liittyville tunteille (esimerkiksi yhteenkuuluvuus, ystävällisyys, hyväksyntä). SOEVE:n tunnekokemusmittaristossa mukana olleet tunteet olivat: ilo, tylsyyttä, hämmästys, empatia, kiinnostus, ylpeys, suru, ihailu, yhteenkuuluvuus, turhautuneisuus, tyytyväisyys, rentous, innostus, viha, hämmennys, hyväksyntä, nautinnollisuus, inho, toivo, sympatia, uteliaisuus, kateus, ahdistuneisuus, kiitollisuus, oivallus, helpotus, häpeä, ylenkatse, rakkaus, pettymys, myötätunto, toivottomuus, ystävällisyys, luottamus, rauhallisuus ja ymmärrys. Tässä ei ollut käytössä Likertin asteikkoa, vaan ohjeistus oli seuraava: ”Merkitse (esimerkiksi alleviivaten) ne tunteet tai mielentilat, joita koit jonkin verran tai paljon äskeisen vuorovaikutuksen aikana. (Niitä tunteita tai mielentiloja, joita et kokenut ollenkaan, ei tarvitse merkitä mitenkään.) Merkitse lisäksi kirjaimella, kohdistuiko tunnetilasi: a) vuorovaikutukseen sinänsä, b) ryhmän tiettyyn ihmiseen tai ihmisiin, c) keskustelun sisältöön, d) oppimiseesi.”

VALAA:ssa heti videoinnin jälkeen osallistujat (sekä opiskelijat että ohjaavat opettajat) täyttivät joka kerta tunnekokemuksia ja dialogisuuskokemuksia kartoittavan kyselylomakkeen. Toisen vuorovaikutusvideoinnin jälkeen osallistujat (vain opiskelijat) täyttivät ryhmäidentiteettiä mittaavan kyselyn, joka oli suomentamani ja suomalaisen opettajakoulutukseen lievästi muokkaamani versio Cameronin (2004) ryhmäidentiteettilomakkeesta.

SOEVE:ssa kehitin uuden mittariston, jolla voidaan tutkia ryhmän ryhmäkokemuksia. Kyselylomake perustuu lähes 20 jo olemassa olevaan mittaristoon ja tutkimustulosten operationalisointiin. Tämän kyselylomakkeen osallistujat täyttivät ennen ensimmäistä videointia, ja sitä on käytetty myös myöhemmin TIITU-tutkimushankkeessa.

² Kasvatustieteellisellä näkökulmalla viitataan kasvatustieteeseen tieteenalana, pedagogiikalla käytännön kasvatustoimintaan, jota ohjaavat tietyt pedagogisesti viisaiksi havaitut periaatteet. Nämä käsitteet ovat luonnollisesti osittain päällekkäisiä.

³ Tutkija Simo Järvelän pitämässä työpajassa haaga-helialaisille tunnelabin käyttäjille 17.12.2020 hän tutustui tutkimuksiini, ja piti tutkimusasetelmaani huomattavan monimutkaisen ja monimenetelmäisenä. Järvelä pohti sitä, että tämänkaltaisessa

tutkimusasetelmassa on vaikea sanoa mikä tietyn tunnereaktion aiheuttaa. Tässä on mahdollista käyttää psykofysiologiaa tavallaan osana laadullista kokonaismittaria, jolloin sen voi kontekstualisoida osana laajempaa kokonaisuutta, jossa tutkijan kokonaistulkinta on etualalla. Tällöin on kuitenkin perusteltava selkeästi mihin mittareihin kokonaisnäkemys perustuu. Arviointitunneteoriat soveltuu tähän erityisen hyvin, jos sitä soveltaa sosiaalisen vuorovaikutuksen lyhyiden tilanteiden analyyseihin (niin kuin olen tehnyt). Tässä on mahdollista hyödyntää myös takaisinpäätelyä, ja pitää fysiologisia signaaleita pointtereina.

⁴ Esimerkiksi viha ja halu satuttaa voi olla jopa pasifistille primaarireaktio (Ellsworth, 2013, 127).