

<https://helda.helsinki.fi>

Moraalinen toimijuus ja ihmiskeskeisyyden dilemma tekoälyn maailmassa

Visala, Aku

Gaudeamus
2021

Visala , A 2021 , Moraalinen toimijuus ja ihmiskeskeisyyden dilemma tekoälyn maailmassa .
julkaisussa P Raatikainen (Toimittaja) , Tekoäly, ihminen ja yhteiskunta : Filosofisia
näkökulmia . Gaudeamus , Helsinki , Sivut 157-180 . <
<https://kauppa.gaudeamus.fi/sivu/tuote/tekoaly-ihminen-ja-yhteiskunta/3868931> >

<http://hdl.handle.net/10138/340672>

unspecified
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Moraalinen toimijuus ja ihmiskeskeisyyden dilemma tekoälyn maailmassa

Aku Visala

Käyttämämme teknologia muokkaa meitä ja yhteisöämme.¹ Latteudesta huolimatta tämä väite on hyvinkin totta. Kaltaiselleni moraalista kiinnostuneelle humanistille juuri tämä seikka tekee tekoälystä ja sitä koskevasta keskustelusta erityisen kiinnostavan. Vaikka tekoälyteknologian uhkista ja mahdollisuuksista on keskusteltu viime vuosina niin mediassa kuin populaarikulttuurin piirissä paljon, eräs uhkakuva on jäänyt vähemmälle huomiolle. Kysyn, miten älykkäiden järjestelmien arkinen käyttö voi mahdollisesti vaikuttaa moraaliimme, erityisesti moraaliseen autonomiaan.

Vaikka tekoälyn etiikassa inhimillisen moraalien ja itseymmärryksemme yhteyksiä on kartoitettu laajastikin, julkisen keskustelun ja populaarin kirjoittelun piirissä teemasta on keskusteltu yllättävän vähän.² Esimerkiksi filosofi Nick Bostromin teoksessa *Superintelligence* ei edes mainita moraaliin liittyviä mahdollisia uhkia. Fyysikko Max Tegmarkin *Life 3.0: Being Human in the Age of Artificial Intelligence* käsittelee tekoälyn kehitystä ja pitkän aikavälin skenaarioita satojen vuosien päähän tulevaisuuteen. Tegmark kyllä kuvaa erilaisia mahdollisuuksia tekoälyn ja ihmisten yhteiselämälle, mutta hän ei juuri kuvaa sitä, miten tekoäly voisi vaikuttaa ihmisten käsitykseen heidän omasta moraalisesta elämästään.

Kysymys on tärkeä kahdesta syystä. Ensiksikin moraalin ja itseymmärryksemme välillä on selvä yhteys: se, miten ymmärrämme ihmisyden, määrittelee sen, kenet ja mitkä hyväksymme moraalisen kohtelun piiriin kuuluviksi olioiksi. Jos pidämme itseämme lähinnä biologisina robotteina, ei meillä ole syytä sulkea moraalisen kohtelun ulkopuolelle esimerkiksi ei-biologisia robotteja. Toiseksi moraalinen kehityksemme on ainakin osittain riippuvainen erilaisista jaetuista käytännöistä ja tottumuksista. Nämä saattavat uuden teknologian myötä muuttua yllättävilläkin tavoilla. Jotkut tutkijat ovat esimerkiksi väittäneet, että sosiaalinen media ja muut tekoälyä käyttävät apuneuvot nakertavat ihmisten moraalista autonomiaa jo nyt. Huolena on se, että kun ihmiset käyttävät moraalista elämää säätelevää teknologiaa yhä enemmän, heidän moraaliset kykynsä jäävät kehittymättä, jolloin moraalista autonomiaa on vaikeampi saavuttaa. Lisäksi tämä teknologia perustuu hyvin pienen piirin arvoille ja päämäärille. Emme edes huomaa, kuinka tämän pienen piirin arvot ja moraalit iskostetaan myös meihin käyttäjiin.³

Nämä huolet koskevat niin jonkinlaiseen vuorovaikutukseen kykeneviä älykkäitä järjestelmiä kuin myös muita älykkäitä ohjelmia ja laitteita, kuten älypuhelimia ja GPS-paikannusta. Yhtä kaikki älykkäät järjestelmät ovat läsnä elämämme ydinalueilla. Maailmamme on erilaisten ”älykkyyksien” kansoitama. On siis tärkeää kysyä, miten nämä älykkyydet muokkaavat moraalista käyttäytymistämme.

Ennen siirtymistä itse asiaan sanon muutaman sanan lähestymistavastani. En aio käsitellä yhteiskunnallisia enkä poliittisia kysymyksiä, vaikka ne tässä yhteydessä ovatkin tärkeitä. Keskityn lähinnä yksilön moraalin erityiskysymyksiin. Teen tämän nostamalla esiin muutaman tärkeän kytköksen itseymmärryksen ja moraalin välillä. Erityisesti populaarikulttuurissa on kiinnitetty huomiota näihin yhteyksiin, joten tarkastelen muutamaa esimerkkitapausta tältä elämänalueelta. Lisäksi etsin näkökulmia tutkimuksesta, jossa käsitellään niitä vaikutuksia, joita maapallon ulkopuolisen älykkyyden löytämisellä voisi olla. Niin tutkimus

kuin populaarikulttuurin tuotteetkin viittaavat siihen, kuinka muiden älykkyyksien maailmassa eläminen voi kalvaa sekä intuitiivista että kulttuurimme ylläpitämää ihmiskeskeisyyttä, mistä puolestaan voi seurata meille keskeisten moraalisten kykyjen ja käytäntöjen rapautuminen.

Moraalinen toimijuus ja käytännöllinen järki

Moraalisen toimijuuden ja moraalisen autonomian ehdoista ja ennakoedellytyksistä on filosofien kesken jonkin verran erimielisyyttä.⁴ Menemättä tähän väittelyyn sen enempää selvitän lyhyesti, kuinka ymmärrän nämä termit. Moraalisella toimijuudella viitaan joukkoon psykologisia kykyjä, jotka mahdollistavat moraalisten arvostelmien muodostamisen, moraalisesti relevanttien tosiseikkojen tunnistamisen sekä oman käyttäytymisen ohjaamisen näiden perusteella.

Filosofi Alasdair MacIntyre on esittänyt hyödyllisen tavan tarkastella moraalista toimijuutta. MacIntyren mukaan moraalinen toimijuus edellyttää joukkoa synnynnäisiä kykyjä, joita kasvatusta ja kulttuuri hiovat. Näitä ovat kyky arvioida ja muodostaa moraalisia arvostelmia, ottaa huomioon moraaliset perusteet arvioinneissa, kyky arvioida omia arvostelmia muiden ihmisten näkemysten valossa sekä kyky ohjata omaa toimintaa näiden arvioiden perusteella.⁵ Vaikka moraalisen toimijuuden perustaso onnistuukin monilta ihmisiltä, moraalinen kasvu hyveissä on vaikeampaa. MacIntyren mukaan ihmiset ovat biologisia organismeja, kuten kaikki muutkin elävät oliot, mutta ihmisillä on kuitenkin joitakin ainutlaatuisia piirteitä. Tärkein näistä on mahdollisuus tulla autonomiseksi moraaliseksi toimijaksi.

Yksilön moraalisen autonomian kehittymiselle on lukuisia ennakoheitoja. Ensinnäkin yksilö on MacIntyren mukaan aina osa yhteisöä. Jatkuva vuorovaikutus yhteisön ja yksilön välillä on välttämätön moraalisen kasvun kannalta. Yksilö ikään kuin kalibroi moraalista päättelyään suhteessa yhteisön standardeihin.

Yhteisö ei ole pelkästään harjoittelun ja koulutuksen paikka, vaan se tarjoaa yksilölle hyvän elämän normit ja mallit. Näin moraalinen kehitys ei ole mahdollista ilman moraalista yhteisöä.

Toiseksi yksilö tarvitsee kykyjä, jotka perinteisesti on las-kettu järjen alaan kuuluviksi. Yksilön on kyettävä tunnistamaan ja ymmärtämään kulttuurinsa ja ryhmänsä normeja, odotuksia sekä rooleja. Ilman tätä moraalista kehitystä ei voi tapahtua. Lisäksi yksilö tarvitsee varsin monimutkaisen kognitiivisen apparaatin, jonka avulla hän pystyy muodostamaan representaatioita erilaisista tulevaisuuden mahdollisuuksista ja arvioimaan niitä. Ilman tätä kykyä pitkän aikavälin suunnittelu ja intentionaalinen toiminta olisi mahdotonta.

Kehittynyt moraalinen autonomia ilmenee yksilön kyknä ottaa etäisyyttä omiin haluihinsa ja tarpeisiinsa sekä arvioida niitä normien ja odotusten valossa. Autonomia on näiltä osin eräänlaista itsehallintaa tai itsekontrollia, jossa yksilö säätelee omia tunteitaan, halujaan ja odotuksiaan. MacIntyren mukaan ihmiset eroavat monista muista eläimistä juuri tässä. Valtaosa ihmisten, kuten muidenkin eläinten, käytöksestä syntyy vastauksena biologisiin haluihin ja päämääriin. Ihmiset tarvitsevat ruokaa, suojaa ja yhteisöä. Ihmiset pyrkivät jatkamaan sukuaan ja suojelemaan omiaan. Näiden halujen lisäksi ihmisellä on kuitenkin mahdollisuus kehittyä autonomiseksi toimijaksi järjen ja itsesäätelyn kykyjen avulla.

Hieman perinteisempää terminologiaa käyttäen voisimme sanoa, että moraalinen toimijuus edellyttää vapaata tahtoa. Filosof John Searlen mukaan vapaa tahto on moraalisen toiminnan edellytys. Jotta voimme tehdä moraalisia valintoja, on näitä valintoja edeltävien perusteiden ja valintojen itsensä välisen suhteen oltava alimääräytynyt. Toisin sanoen siis tekojen ja niihin johtavien psykologisten (ja ei-psykologisten) syiden välissä on oltava tilaa, eräänlainen ”aukko”. Tässä aukossa tietoinen minä tekee moraalisia arvostelmia, arvioi toiminnan perusteita ja lopulta muodostaa aikomuksen toimia jollakin tietyllä tavalla.⁶

Näiden huomioiden pohjalta muodostuu rikas kuva moraalista toimijuudesta. Yhtäältä moraalinen toimijuus perustuu kyvyille, jotka ihminen saa synnyinlahjana. Vaikka jänistä yritettäisiin kasvattaa kuten ihmislasta, se ei kykenisi toimimaan ihmisten moraalissa yhteisössä. Tämä johtuu siitä, että jäniksellä puuttuvat ne kognitiiviset kyvyt, joita moraalinen toimijuus edellyttää. Toisaalta taas moraalisen toimijuuden kehitys riippuu vahvasti yhteisöstä ja kulttuurista. Yksilö itse ei kykene tekemään itsestään moraalista toimijaa vain ja ainoastaan synnynnäisen varustuksensa perusteella, vaan se edellyttää muita ihmisiä, normeja, kulttuuria sekä koulutusta. Kuten kohta käy ilmi, tekoäly ja moraalista autonomiaa koskevat huolenaiheet koskevat nimenomaan tätä moraalisten kykyjen ja yhteisön välistä linkkiä.

Moraali valinkauhassa

Populaarikulttuurissa ihmisen moraalista älykkäiden koneiden maailmassa on tarkasteltu laajasti. Erityisesti tietekirjallisuus ja tietokonepelit ovat kunnostautuneet näiden teemojen käsittelyssä. Käsittelen nyt kahta viime vuosina ilmestynyttä populaarikulttuurin tuotetta, jotka hyödyllisellä tavalla havainnollistavat ihmisten kohtaamia moraalisia haasteita älykkään teknologian ympäristössä.

Vuonna 2018 ilmestyi kunnianhimoinen konsolipeli *Detroit: Become Human*. Peli asettaa pelaajan ohjattavaksi kolme ihmistä hyvin paljon muistuttavaa androidia. Pelaajan tehtävänä on päättää, kuinka näiden androidien suhde ihmisiin kehittyy. Tulevaisuuden yhteiskunnassa androidit ovat yleistyneet ja niitä käytetään mitä erilaisimpiin tehtäviin. Pelaajan ohjaamat päähenkilöt ja heidän kohtalonsa liittyvät eri elämäntilanteisiin. Yksi androidi on alkoholistimiehen kodinhoitaja ja äidin korvike laiminlyödylle tyttärelle, toinen taas poliisi ja kolmas rikkaan taiteilijan henkilökohtainen avustaja. Erilaisten sattumusten ja moraalisten dilemموjen kautta kukin näistä koneista joutuu luo-

pumaan totutusta roolistaan ja kehittymään. Orjina pidettyjen koneiden vallankumous tulee. Valitseeko pelaaja vastakkainasetelun vai yhteistoiminnan tien?

Peli on rakennettu taitavasti niin, että se haastaa pelaajan jatkuvasti määrittelemään suhteensa androideihin uudella tavalla. Tämä käsitys puolestaan vaikuttaa siihen, miten pelaaja hahmoja pelaa ja kuinka heidän tarinansa edistyvät. Pelaaja voi yhtäältä pelata androideja ”konemaisesti”, jolloin niillä ei juuri ole empatiaa tai inhimillisiä tunteita. Toisaalta hän voi kuitenkin pelata androideja ”ihmismaisesti” ikään kuin ne olisivat moraalisia toimijoita, joita koskevat samat moraaliset velvollisuudet ja oikeudet kuin ihmisiä. Samalla pelaaja voi myös vaikuttaa siihen, miten pelin ihmishahmot suhtautuvat androideihin ja missä määrin he muuttavat näkemyksiään itsestään, kun robotit tulevat osaksi heidän elämäänsä.

Pelin viesti on selkeä. Se, mitä pelaaja ajattelee robotin moraalisesta statuksesta ylipäänsä, heijastuu siihen, miten hän robottien käytöstä pelissä ohjailee. Mitä enemmän pelaaja inhimillistää robotteja, sitä inhimillisempiä niistä tulee. Näin peli havainnollistaa yksinkertaista moraalipsykologista väitettä: moraalinen kohtelu on moraalisen toimijuuden kehityksen edellytys. Jotta olio voi kehittää moraalisia kykyjään, edellyttää tämä osallisuutta moraaliseen yhteisöön. Lisäksi peli paljastaa tyylikkäästi sen, kuinka moraalinen kohtelu tai sen puute vaikuttaa myös niihin, joita jo ennestään kohdellaan täysinä moraalisinä toimijoina. Jos kohtelen androidia julmasti ja välinpitämättömästi, tulen myös itse julmemmaksi ja kovasydämisemmäksi. Tämä vaikutus ei ole riippuvainen androidin todellisesta moraalisesta tai ontologisesta statuksesta. Riippumatta siitä, kykeneekö robotti esimerkiksi tietoisuuteen tai moraaliseen toimintaan, sen kohtelulla on vaikutusta kohtelijaan itseensä.

Peli havainnollistaa kiinnostavalla tavalla myös moraaliseen kohteluun kuuluvaa riskiä ja haavoittuvaisuutta. Mitä enemmän pelaaja inhimillistää androideja, sitä enemmän hän samastuu

niihin. Hän siis asettaa itsensä alttiiksi kärsimykselle, kun androideja kohdellaan huonosti tai niiden tarinat päättyvät huonosti.

Toisena esimerkkinä toimikoon vuonna 2017 ilmestynyt elokuva *Blade Runner 2049*. Tämä jatko-osa scifi-klassikkoon on erittäin onnistunut ja käsittelee monipuolisesti edeltäjästään tuttuja teemoja. Elokuvasa eletään ympäristötuhon jälkeistä aikaa, jossa ihmistä muistuttavat keino-oliot biologisista ”replikanteista” erilaisiin hologrammeihin ovat arkipäivää. Replikanteista on tullut eräänlainen orja- ja työväenluokka, jolla ei ole ihmisen moraalista statusta. Ne tekevät roskatyöt ja tyydyttävät ihmisten monet tarpeet.

Eräs elokuvan kiinnostava piirre on se, kuinka se esittää ihmiset. Valtaosa elokuvan ihmishahmoista on kovasydämissä ja välinpitämättömiä. Replikantit ovat heille lähinnä työkaluja, joista saadaan irti hyötyä, tai sitten seksiojekteja, joista saadaan tyydytystä. Toisinaan ihmishahmot horjuvat inhimillistävien ja välineellistävien asenteiden välimaastossa tietämättä itsekään, miten suhtautua. Nämä asennevaikeudet näyttävät myös heijastuvan heidän suhteisiinsa muiden ihmisten kanssa.

Näin elokuva alleviivaa moraalipsykologian ja androidien kohtelun kriittisen tärkeää suhdetta. Jos ihmiset tottuvat kohtelemaan heitä monessa mielessä muistuttavia androideja vain ja ainoastaan välineinä, he alkavat kohdella välineinä myös toisiaan. Moraalisen toiminnan ennakoedellytyksenä on joukko psykologisia kykyjä, jotka edellyttävät harjoitusta. Näihin kuuluvat esimerkiksi kyvyt tunnistaa moraalisesti relevantteja tosiseikkoja toisista ihmisistä, kyky harjoittaa empatiaa sekä ottaa huomioon toisten intressit. Nämä kyvyt eivät kuitenkaan pysy yllä itsestään ympäristössä, jossa on ihmisenkaltaista älykkyyttä mutta jossa sitä ei kuitenkaan oteta moraalisen kohtelun piiriin.

Tekoäly ihmisällyn kuvana

Edelliset huomiot osoittavat, ettei tekoäly ole mitä tahansa teknologiaa. Se on teknologiaa, jonka tarkoitus on auttaa meitä sellaisissa moraaliin ja sosiaaliseen elämään liittyvissä asioissa, jotka on aiemmin katsottu kuuluviksi vain ihmisällyn piiriin. Tämä onnistuu vain, jos tekoäly kykenee tekemään vaativia moraalisia päätöksiä. Vaikka tekoäly suorittaakin tehtävänsä usein eri tavoin kuin ihmisäly suorittaisi vastaavan tehtävän, pyrkii tekoäly kuitenkin olemaan eräessä mielessä ihmisällyn kuva. Tästä herää heti kysymys siitä, millainen ihmisäly todellisuudessa oikein on.

Ihmismäisen älykkyyden suhteen erotetaan toisistaan kapea tekoäly ja laaja tai yleinen tekoäly. Kapea tekoäly keskittyy jonkin tarkasti rajatun tehtävän suorittamiseen, kuten lääketieteellisen diagnoosin tekemiseen tai shakin pelaamiseen. Tekoälyjärjestelmät ovat tällaisilla kapean tekoällyn alueilla monesti ylittäneet ihmisen suorituskyvyn. Koneoppiminen on mahdollistanut esimerkiksi tehokkaan kielenkääntämisen, suurien datamassojen käsittelyn sekä erilaiset hahmontunnistuksen sovellukset.

Yleisälykkyyden suhteen edistys on kuitenkin ollut hidasta. Yksi selitys hitaudelle saattaa olla se, ettei meillä ole yksimielisyyttä ihmisälykkyyden toimintamekanismeista. Pyrkimys tuottaa ihmistä vastaavaa älykkyyttä edellyttää käsitystä siitä, millaisia olioita me ihmiset olemme ja miten älykkyytemme tosiasiaassa toimii. Ongelma erityisesti moraalisen ajattelun ja käyttäytymisen tutkimuksessa on se, että tutkimuskohde ei ole pelkästään tieteellinen ja empiirinen. Moraalin voidaan katsoa sisältävän oletuksia arvoista, normeista ja ihmisen hyvästä, joiden totuutta ei voida empiirisesti ratkaista.

Moraalilla on rooli yhteiselämän ja politiikan taustatekijänä. Erilaisiin filosofisiin, uskonnollisiin ja poliittisiin näkemyksiin liittyy oletuksia siitä, millaisia ihmiset ovat, mikä on heille hyvää ja millainen heidän psykologiansa on. Esimerkiksi puolueiden ohjelmiin perehtyminen paljastaa jo lukuisia eroja siinä, miten yksilöllisyys ja yhteisöllisyys käsitetään.

Moraalimme ja ihmiskäsityksemme on myös keskeisessä roolissa, kun päätämme yhteiskuntana, millaista teknologiaa haluamme kehittää ja missä hyväksyttävän ja torjuttavan raja oikein kulkee. Erilaiset ihmiskäsitykset johtavat erilaisiin uutta teknologiaa kohtaan omaksuttuihin asenteisiin. Yhteiskunnan digitalisoituessa olisi siis tärkeää ottaa huomioon erilaisten ryhmien itseymmärrys.

Totesin edellä, ettei tekoäly ole mitä tahansa teknologiaa. Tekoälyn kautta teknologia levittäytyy syväälle moraalisen ja sosiaalisen elämämme alueelle. Filosofit ovat jo pitkään kiinnittäneet huomiota teknologian arvosidonnaisuuteen.⁷ Teknologian tuotanto ja käyttö vaikuttaa ihmissuhteiden luonteeseen, yhteisöihin ja yksilöihin joskus ennakoimattomillakin tavoilla. Teknologia suunnitellaan aina johonkin tarkoitukseen, joten se antaa konkreettisen muodon suunnittelijan päämäärille ja tätä kautta myös arvoille. Tekoäly on siinä mielessä vaarallinen, että sen avulla meille hyvin tärkeät ja läheiset elämänalueet, kuten juuri arvot ja moraalit, voidaan teknologisoida.

Tästä syystä esimerkiksi filosofit Brett Frischmann ja Evan Selinger pitävät älykstä teknologiaa erityisen haastavana. Ihmisten moraaliseen elämään kietoutuva tekoäly voi pahimmillaan toimia välineenä, jolla suuryritykset panevat toimeen valtavan mittakaavan yhteiskunnallisen vaikuttamisen projekteja omaksi edukseen. Tämän mahdollisuuden pitäisi herättää yritykset tunnustamaan oma vastuunsa. Älykkään teknologian insinöörit ovat kuitenkin ihmisiä, kuten me kaikki muutkin. Heillä on omat arvonsa ja ihmiskäsityksensä sekä vinoumansa. Erityistä huomiota Frischmann ja Selinger kiinnittävät Piilaakson kulttuuriin, joka on läpeensä kilpailun ja tehokkuuden vaatimusten leimaamaa. Osakkeenomistajien intressit, siis taloudellinen voitto sekä tehokkuus tehtävien suorittamisessa, ovat tärkeimmät arvot. Sovellettuna älykkääseen teknologiaan tehokkuuden hintana voi kuitenkin olla se, että käyttäjä menettää moraalisen autonomiansa ja kyvyt sen ylläpitoon.⁸

Älykäs teknologia ei siis synny missään poliittisessa eikä eettisessä tyhjiössä. Esimerkiksi länsimaisen ihmiskäsityksen keskeiset väitteet on kirjattu kansallisiin ja kansainvälisiin lakeihin. Tärkeä osa eurooppalaista ja myös suomalaista arvopohjaa on tietynlainen käsitys siitä, millaisia ihmiset ovat ja mikä on heille hyvää. Ihmisarvo, moraalinen autonomia, erilaiset vapaudet ja velvollisuudet sekä ihmisoikeudet implikoivat jo tietyn itseymmärryksen.

Ihmiskeskeisyys erilaisten älykkyyksien kosmoksessa

Kuten edellä kävi ilmi, moraalinen toimijuus ja kehittynyt moraalinen autonomia edellyttävät joitakin synnynnäisiä kognitiivisia kykyjä. Nykytiedon perusteella vain ja ainoastaan *Homo sapiens*-organismit on varustettu tällaisilla kyvyillä. Tästä ei tietenkään seuraa, etteivätkö muut eläimet kykenisi jonkinlaiseen sosiaaliseen ja moraaliseen käytökseen. Näyttää kuitenkin siltä, että kumuloituvan kulttuurin ja laajan yhteisön kautta harjoitettava moraalinen toimijuus on ihmisen erityispiirre.⁹

Ihmiskeskeisyydeksi tai *antroposentrisyydeksi* kutsun tässä näkemystä, jonka mukaan vain *Homo sapiensit* voidaan laskea kuuluviksi täysimittaisesti moraalisen yhteisön jäseniksi. Yhä lisääntyvä älykäs teknologia kuitenkin haastaa tämän näemyksen: ympäristömme täyttyy erilaisista, ei-homo sapiens-älykkyyksistä. *Blade Runnerin* ja *Detroitin* esimerkit osoittavat, että ihmiskeskeisyyden häviämällä voi olla myös hyvin hankalia seurauksia: ihmiset voivat tulla laiskoiksi moraalisen toimijuutensa harjoittamisessa ja yhteisön moraaliset normit voivat muuttua laajamittaisemminkin.

Tutkimusta ihmiskeskeisyyden rapautumisen vaikutuksesta moraalisiin on kuitenkin olemassa varsin vähän. Vertailukohta voisi löytyä tutkimuksesta, jossa on pyritty käsittelemään niitä

vaikutuksia, joita maapallon ulkopuolisen älykkään elämän löytymisellä voisi olla.

Yhdysvaltain avaruushallinto NASA päätti vuonna 2015 rahoittaa usealla miljoonalla dollarilla kolmivuotista projektia, jonka tutkimuskohteena ovat astrobiologian yhteiskunnalliset ja uskonnolliset vaikutukset. Tässä projektissa tarkasteltiin erilaisia poliittisia, moraalisia, teologisia ja muita yhteiskunnallisia vaikutuksia, joita planeettamme ulkopuolisen älykkään elämän etsintä (tai sen löytäminen) voi laukaista. NASA pitää tällaista työtä tarpeellisena erityisesti siksi, että sen oma olemassaolo riippuu ihmisten verorahoista ja näin ollen siitä, miten kansalaiset siihen suhtautuvat. Projektin ytimessä on kuitenkin tärkeä huomio: maan ulkopuolisen älykkyyden löytymisellä on valtavat vaikutukset siihen, miten ihmiskunta kokonaisuutena näkee itsensä, oman arvonsa ja päämääränsä kosmoksessa. Tämä kysymys ei ole pelkästään tieteellinen vaan myös moraalinen, poliittinen ja jopa teologinen.

Laajassa astrobiologian vaikutuksia kartoittavassa kokoelmateoksessa muodostetaan ja arvioidaan erilaisia skenaarioita maapallon ulkopuolisen älyn löytymisen seurauksista. Erityistä huomiota kiinnitetään ihmiskeskeisyyteen:

Tarve päästä ihmiskeskeisyyden tuolle puolen on eräs niistä päällepäin mahdottomilta näyttävistä ongelmista, joista on keskusteltava maapallon ulkopuolisen elämän löytämisen yhteydessä. Kuinka kykenemme siirtymään sellaisten peruskäsitteiden, kuten elämä, älykkyys, kulttuuri, sivilisaatio, teknologia ja kommunikaatio, tuolle puolelle?¹⁰

Ongelma on, että käsityksemme siitä, mitä älykkyys, kulttuuri ja kommunikaatio ovat, perustuvat pääsääntöisesti kokemukseemme inhimillisestä älykkyydestä, kulttuurista ja kommunikaatiosta. Tämä kokemus on vieläpä vahvasti jonkin tietyn kulttuuriperinteen (tavallisesti länsimaisen perinteen) suodattamaa. Maapallon ulkopuolisen äyllisen elämän löytäminen siis

pakottaisi meidät tarkastelemaan älykkyyttä ja kulttuuria aivan uudella, ei-ihmiskeskeisellä tavalla.

Tutkimuksessa on tarkasteltu myös sitä, millainen vaikutus maan ulkopuolisen elämän löytymisellä olisi eri uskontoihin. David Weintraubin mukaan kaikkein hyödyllisin mittari mahdollisten vaikutusten analysoinnissa on uskonnon ihmiskeskeisyys tai sen puute.¹¹ Millaisen roolin kukin uskonto antaa *Homo sapiensille* kosmoksen tarinassa? Tämä on aiheemme kannalta tärkeää sikäli, että uskonnot ja laaja-alaiset maailmankatsomukset ovat valtaosalle ihmiskunnasta heidän moraalinsa ja ihmiskäsitystensä keskeisiä lähteitä. Weintraubin mukaan on tärkeää, että tarkastelemme kutakin uskontoa erikseen. Jotkin uskonnot ovat vähemmän ihmiskeskeisiä kuin toiset. Melkein kaikkiin uskontoihin kuuluu ajatus ei-inhimillisestä älykkästä elämästä: henget, jumalat, enkelit ja esi-isät asuttavat uskonnollisten ihmisten maailmoja. Nämä oliot ovat kuitenkin luonteeltaan yliluonnollisia tai ”ei-luonnollisia”. Entä sitten aivan luonnolliset, maapallon ulkopuoliset oliot?

Ihmiskeskeisyyden kulttuuri ja intuitiivinen ihmisluonto

Määrittelin edellä ihmiskeskeisyyden näkemykseksi, jonka mukaan vain *Homo sapiens* kykenee moraaliseen toimijuuteen. Länsimaisessa kulttuurissa on laaja kirjo näkemyksiä, jotka tukevat tätä väitettä. Näitä näkemyksiä voidaan tarkastella kolmen eri akselin tai ulottuvuuden kautta.

Ensimmäinen akseli jaottelee näkemyksiä sen perusteella, kuinka vahva persoonan ontologia niissä on. Mitä persoonat oikein ovat? Akselin vahvimmassa päässä on joukko näkemyksiä, joiden mukaan jokaista ihmisyksilöä vastaa tämän syntymästä lähtien ikuinen ja tuhoutumaton sielu, joka ei ole osa luonnollista maailmaa. Tällainen sielu takaa jokaisen yksilön erityisyyden sekä erottaa sielullisten olioiden joukon siitä joukosta, jolla sielua

ei ole. Käsityksellä siitä, että vain ihmisillä on rationaalinen sielu, on länsimaisen filosofian ja teologian perinteessä puolustettu ihmisen erityisyyttä suhteessa muihin eläviin olentoihin.¹² On myös joitakin tieteellisiä syitä pitää ihmistä siinä mielessä erityisenä, että tällä saattaa olla psykologisia järjestelmiä, joita millään muulla lajilla ei ole. Ehdokkaita tällaisiksi mekanismeiksi on esitetty useita: itsetietoisuus, symbolinen kieli, monimutkainen sosiaalinen älykkyys sekä kyky toimia joustavasti kulttuurin olosuhteissa.¹³

Akselin vastakkaisessa päässä puolestaan on joukko näkemyksiä, joissa ihmiskeskeisyys on minimoitu. *Homo sapiensilla* ei ole sielua. Ei myöskään ole mitään, mikä ankkuroisi yksilön identiteetin muista erillisenä. Ihmisyksilöllä sen enempää kuin ihmisyhteisöllä ei tällöin ole mitään erityisyyttä tai yksilöllisyyttä takaavaa tekijää, vaan ero muihin olioihin on lähinnä harha, josta ihmisten tulisi päästä eroon. Lisäksi voidaan esittää, ettei ihmisellä ole sitä muista lajeista erottavia psykologisia kykyjä tai ominaisuuksia laisinkaan: vaikka on totta, ettei mikään muu tuntemamme laji kykenekään ihmistä vastaavaan sosiaaliseen älykkyyteen tai vaikkapa symbolisen kielen käyttöön, näillä kyvyillä on kuitenkin juurensa myös muiden lajien kyvyissä.

Ihmiskeskeisyyttä on perusteltu myös vetoamalla erilaisiin arvo- ja olemishierarkioihin, jollaisia liittyy uskontojen ja maailmankatsomusten ihmiskäsityksiin. Tällaisia hierarkioita tarkastelemalla voidaan kartoittaa kyseisen uskonnon tai maailmankatsomuksen ihmiskeskeisyyttä. Esimerkiksi länsimaisen filosofian ja teologian perinteessä ihminen on tavallisesti asetettu hierarkian keskivälille. Hänen ”alapuolellaan” ovat eläimet, kasvit ja muu luomakunta. Toisaalta taas hänen ”yläpuolellaan” arvoasteikossa ovat erilaiset henkiolennot. Jumala on tämän hierarkian korkein olio. On myös ihmiskäsityksiä, joissa ihminen on hierarkiassa korkeimmalla. Näin ajatellaan esimerkiksi silloin, jos ihminen nähdään maapallon (ja ehkä koko kosmoksen) historian huippukohtana, jossa ei-tietoinen luonto tulee vihdoin tietoiseksi itsestään. Jotkin ihmiskäsitykset ovat puolestaan ei-

hierarkkisia, ja niissä ihmiset, muut eläimet ja koko ympäristö pyritään asettamaan samalle portaalle.

Kolmas ulottuvuus koskee itseymmärrykseemme liittyvää tarinaa tai narratiivia. Psykologisesti itseymmärrystä pidetään usein yllä erilaisilla tarinoilla. Millaisen tarinan kerromme itsellemme ihmisen alkuperästä ja roolista (tai sen puutteesta) maapallolla tai koko kosmoksessa? Kullakin uskonnolla on tästä oma versionsa. Myös erilaiset filosofiat ja poliittiset ajattelutavat kantavat mukanaan omia tarinoitaan: esimerkiksi länsimaissa on valistuksen jälkeen kehittynyt kertomus edistyksestä, jossa ihmiskunta kehittyy teknologisesti ja moraalisesti kohti onnen tilaa. Tieteellisten tulostenkin perusteella luodaan erilaisia tarinoita ihmisestä. Brittifilosofi John Gray on esimerkiksi väittänyt, että länsimainen edistysusko ja usko ihmisarvoon ja ihmisen erityisyyteen ovat ristiriidassa evoluutiobiologian tulosten kanssa.¹⁴

Narratiiveilla on merkitystä, koska ne kantavat mukanaan ihmiskäsityksiä ja moraalisia normeja. Näistä syistä niillä on tärkeä rooli myös tekoälyä koskevassa keskustelussa. Tegmark kertoo teoksessaan siitä, miten supertekoäly voisi ehkä syntyä, ja kuljettaa tätä tarinaa filosofisen pohdiskelun sivussa teoksen läpi. Kuten tunnettua, myös transhumanistien tulevaisuusvisioissa on paljolti kyse ihmisen moraalista ja sen kehittämisestä. Ihminen on tämän tarinan mukaan teknologiaa käyttävä olio, joka on aina pyrkinyt ylittämään teknologian avulla biologisen rakenteensa ja ympäristönsä rajat. Superälykkyys on vain seuraava askel tässä.¹⁵

Edelliset kolme ulottuvuutta (persoonan ontologia, ihmiskäsitysten hierarkia ja itseymmärryksen narratiivi) ovat enemmän tai vähemmän tietoisia. On kuitenkin tieteellistä näyttöä siitä, että ihmisillä olisi myös taipumusta ihmiskeskeiseen ajatteluun. Meillä saattaa olla osin jaettu, intuitiivinen käsitys synnynnäisestä ihmisluonnosta, joka liitetään nimenomaan *Homo sapiensin*.¹⁶

Ihmisillä on tapana olettaa, että kaikkia ihmisiä yhdistää jonkinlainen ”olemus” tai luonto, joka viime kädessä selittää ihmisiä yhdistävät piirteet. Ihmisillä on vahva taipumus liittää tämä olemus *Homo sapiensin* biologiaan ja siihen sisänrakennettuihin

päämääriin. Yksilöitä ja ryhmiä erottavia piirteitä selitetään puolestaan usein kulttuurieroilla. Lisäksi tämä ”olemus” on siinä mielessä ihmislajille erityinen, että se erottaa meidät muista lajeista: vaikka ihmislajin ominaisuuksilla olisikin edeltäjiä muussa eläinkunnassa, on tämä ”olemus” juuri niiden ominaisuuksien syy, jotka erottavat ihmiset muista lajeista.¹⁷ Nämä ihmiset muista erottavat kyvyt antavat ihmisille myös joukon – yleensä objektiivisesti ymmärrettyjä – päämääriä ja tavoitteita. Voidaan siis sanoa, että intuitiiviseen itseymmärrykseemme kuuluu myös tietynlainen intuitiivinen teleologia: on joukko hyviä päämääriä, joita ihmisten tulisi olemuksensa perusteella tavoitella, ja pahoja, joita heidän tulisi välttää.

On myös näyttöä siitä, että ihmisillä on taipumus olettaa minuus tai sielu, joka tekee kustakin yksilöstä erityisen ja antaa tälle erityisaseman verrattuna muihin eläimiin. Tämä sielu on myös tavalla tai toisella ihmisälyn ja ihmisarvon erityinen asuin-sija. Mielen ja ruumiin erottaminen toisistaan, jonkinlainen dualismi siis, näyttää myös ihmisille intuitiiviselta.¹⁸

On helppo nähdä, kuinka nämä intuitiiviset taipumukset tukevat ja vahvistavat ihmiskeskeisyyttä. Oletus lajiolemuksesta ja lajimme erityisyydestä sopii hyvin yhteen sen ajatuksen kanssa, että moraalinen kohtelu kuuluu vain ja ainoastaan *Homo sapiens*-lajin jäsenille. Taipumus käsittää minuus ja sielu kehoista erillisiksi, yksilön identiteetin ankkuroiviksi olioiksi, tukee samaista väitettä. Jos psykologiamme on tosiaan kallistunut ihmiskeskeiseen suuntaan, tulee ymmärrettävämmäksi se, miksi esimerkiksi länsimaisen filosofian ja poliittisen ajattelun historiassa ihmiskeskeiset moraalikäsitteet ovat olleet suosittuja.

Ihmiskeskeisyys tekoälyn maailmassa

Näyttää siis siltä, että ihmiskeskeisyys sekä usko ihmisen erityisyyteen ovat kumpikin osa niin intuitiivisia taipumuksiamme kuin monia länsimaisen kulttuurin ihmiskäsityksiä. Tällöin älyk-

kään teknologian lisääntyminen voi uhata, tai ainakin haastaa, moraalimme taustalla vaikuttavia ihmiskäsityksiä. Erityisesti ne ihmiskäsitykset, jotka olettavat kaikkein vahvimman version ihmisen erityisyydestä, joutuvat paineen alle. Älykkäiden järjestelmien täyttämä maailma ei kohtelee helläkätisesti näkemyksiä, joissa inhimillinen älykkyys nähdään yliluonnollisena tai fysikaalisesta ja biologisesta maailmasta riippumattomana ilmiönä.

On tietysti mahdollista uskoa ihmisälyn yliluonnolliseen alkuperään myös älykkään teknologian maailmassa, mutta helppoa se ei ole. Usko ihmislajin erityisyyteen moraalisisessa ja psykologisessa mielessä ei välttämättä edellytä uskoa yliluonnolliseen sieluun. Älykkään teknologian maailmassa on kuitenkin yhä uskottavampaa ajatella, että myös inhimillinen älykkyys on tulosta luonnollisista prosesseista eikä edellytä yliluonnollista lähdettä. Näin ihmisen itsensä luoma älykkyys haastaa sellaisen ihmiskeskeisyyden, jossa ollaan lähellä persoonan ontologian vahvaa päätä. Jos kerran inhimillinen älykkyys ja tekoäly ovat näin samalla viivalla, miksi ihmisten pitäisi suosia itsensä kaltaista älykkyyttä muiden älykkyyksien kustannuksella?

Maailma, jossa olemme älykkäiden järjestelmien ympäröimiä, voi osoittautua hyvinkin hankalaksi joillekin uskonnoista ponnistaville eettisille näkemyksille. Esimerkiksi jotkin kristinuskon protestanttiset muodot ovat hyvin ihmiskeskeisiä: niissä älykkyys on aina tulosta Jumalan yliluonnollisesta toiminnasta ja ihmisen rooli kosmoksessa on keskeinen. Moraali on periaatteessa yliluonnollista, ja vain ihminen kykenee siihen (silloinkin lähinnä Jumalan vaikutuksesta).

Tämä ei kuitenkaan ole ainoa mahdollinen teologia. Kristinuskon historia paljastaa laajan keskustelun muiden älykkyyksien luonteesta ja olemassaolosta. Antiikin ja keskiajan teologiassa epäinhimillisiin älykkyyksiin uskottiin laajasti: enkelien, demonien ja muiden henkiolentojen lisäksi Jumalan katsottiin kykenevän luomaan, mitä hän vain sattui haluamaan. Jumala voi siis periaatteessa luoda kaikenlaisia maailmoja ja olioita, ja ehkä hän on näin tehnytkin. Myös meidän kosmoksemme voi kuhista äly-

kästä elämää, jota on inhimillisestä näkökulmasta vaikea ymmärtää.¹⁹

Nämä ajatukset eivät kuulu pelkästään antiikin ja keskiajan maailmaan. Valtaosa ihmiskunnasta uskoo epäinhimillisiin älykkyyksiin sekä joihinkin ihmistä älykkäämpiinkin olioihin, kuten jumaliin ja henkiin. Monoteististen uskontojen, kristinuskon, juutalaisuuden ja islamin, kannattajat uskovat lisäksi epäinhimillisen superälyn olemassaoloon. Jumalalla on kaikki täydellisen olion ominaisuudet, joista jotkin liittyvät älykkyyteen. Jumalan katsotaan tietävän kaikki tiedettävissä olevat totuudet sekä toimivan moraalisesti täydellisellä tavalla. Jumala kykenee toteuttamaan myös ihmisälylle käsittämättömiä suunnitelmia. Tällaisen näkemyksen vaikutuksia moraaliseseen autonomiaan en tässä yhteydessä lähde arvioimaan.

Edellisen nojalla voisi kuvitella, ettei tekoälyjen lisääminen tähän älykkyyksien paljouteen aiheuttaisi juurikaan ongelmia. Tämä ei kuitenkaan välttämättä pidä paikkaansa. Ongelma ei ole vain ja ainoastaan se, että on olemassa epäinhimillisiä älykkyyden muotoja. Ongelma on pikemminkin se edellä esitetty väite, että älykkään teknologian keskellä on hyvin vaikea pitää yllä ajatusta, että älykkyys on yliluonnollista. Tekoäly kun on ihmisen tekemä ja sellaisenaan osoittaa, että ainakin jonkinlainen älykkyys voidaan valmistaa ja suunnitella. Juuri tämä väite on jännitteessä niin ihmiskeskeisyyden kuin edellä kuvattujen uskonnollisten näkemysten kanssa. Näin tekoälyn maailmassa on yhä vaikeampi pitää kiinni ihmiskeskeisistä narratiiveista ihmisälyn alkuperästä ja tulevaisuudesta.

Ehkä vahvan ihmiskeskeisyyden muodoista olisi syytäkin luopua. Näin meidän ei tarvitsisi olla huolissaan siitä, jos ihmiskeskeiset ajattelutavat rapautuvat älykkäiden järjestelmien maailmassa. Jotkut filosofit ovat esittäneet, että ihmiskeskeisyydestä luopuminen on muutenkin moraalisesti perusteltua. Ympäristöetiikkaa koskevassa keskustelussa ihmiskeskeisyydestä on väitely laajasti. Yhtäältä ihmiskeskeisyyttä on kritisoitu siitä, että se on lisännyt asenteita, joissa luonto nähdään ihmisen käy-

tössä olevana resurssina. Jos ihmiskeskeisyydestä kyettäisiin luopumaan, olisi yhteys ihmisen ja ympäristön välillä helpompi hahmottaa. Tällöin voitaisiin myös purkaa hierarkioita ja luoda uudenlaista tietoisuutta ihmisen roolista ympäristökatastrofin lähteenä ja mahdollisena korjaajana.²⁰ Vastaavaa keskustelua on käyty myös eläinetiikan piirissä. Ihmiskeskeisyys erottaa ihmisen muista eläinlajeista ja edistää muiden eläinten julmaa ja välineellistävää kohtelua.²¹

Toisaalta taas ihmiskeskeisyyttä on myös puolustettu ympäristöetiikan piirissä väittämällä, ettei minkäänlainen moraalinen tarkastelu ole mahdollista ilman oletusta persoonia koskevista velvollisuuksista ja oikeuksista. Näin oletus ihmisten moraalista autonomiasta olisi välttämätön edellytys koko eettisen keskustelun olemassaololle.²² Tätä huolta vahvistaa myös se huomio, että ihmisarvon sekä yksilön autonomian korostaminen on hyvin tyypillinen piirre nimenomaan länsimaisissa moraaliteorioissa ja politiikassa. Mitä enemmän ihmiskeskeisyys horjuu, sitä vaikeampi ihmisarvoa ja autonomiaa on puolustaa. Näin ihmiskeskeisyyden horjumisella voi olla suuriakin moraalisia ja poliittisia seurauksia.

Moraalinen autonomia ja ihmiskeskeisyyden dilemma

Edellä paljastui alustava versio ihmiskeskeisyyden dilemmasta. Seuraavaksi haluan keskittyä sen yhteen, hieman tarkemmin rajattuun muotoon. Ongelma voidaan muotoilla seuraavalla tavalla. Ihmiskeskeisyys on yksi tärkeä oletus, kun arvioimme sitä, millaiset oliot kuuluvat moraaliseen yhteisöömme ja ”inhimillistävän” moraalisen kohtelun piiriin. Tällainen inhimillistävä moraalinen kohtelu koostuu pitkälti edellä kuvatuista moraalisen autonomian edellytyksistä. Tarvitaan yhteisiä normeja, niiden täyttämiseen kohdistuvia vaatimuksia ja odotuksia sekä vastuussa pitämistä. Normien täyttämisestä puolestaan seuraa

joukko oikeuksia ja status yhteisössä. Moraalinen kohtelu kuitenkin edellyttää aina jonkinlaista rajanvetoa. Mitä vahvempi ihmiskeskeisyyden korostus on, sitä tarkemmin erotamme ihmiset, jotka ovat osa moraalista yhteisöämme, ja ei-ihmiset ja muut oliot, jotka eivät tähän kuulu.

Kuten edellä kävi jo ilmi, älykkään teknologian maailmassa tällaista jakoa on hankalampi tehdä. Joudumme kuitenkin tekemään vaikeita moraalisia päätöksiä siitä, miten kohtelemme muita älykkyyksiä. Ongelmia seuraa niin moraalisen kohtelun laajentamisesta kuin supistamisestakin. Kutsun tätä ihmiskeskeisyyden dilemmaksi. Jos älykkäät koneet rajataan kategorisesti moraalisen kohtelun ulkopuolelle ja nämä ovat kuitenkin osa elämäämme, voi tällä olla ei-toivottuja moraalipsykologisia vaikutuksia. Jos taas moraalisen kohtelun rajoja laajennetaan ja erilaiset älykkäät koneet otetaan tämän kohtelun piiriin, uhkana ovat toisenlaiset moraalipsykologiset ongelmat. Aloitetaan dilemman ensimmäisestä sarvesta.

Jos ihmiskeskeisyydestä halutaan pitää kiinni ja rajoittaa moraalinen kohtelu vain ja ainoastaan *Homo sapiensien* piiriin, voi syntyä varsin suuria ongelmia, joihin jo viitattiin edellä. Lisään niihin vielä yhden. Kyseinen ongelma on moraalipsykologinen. Jotkut filosofit (kuten Tuomas Akvinolainen ja Immanuel Kant) ovat esittäneet, että julma kohtelu, vaikka sen kohteena olisivatkin itsetietoisuutta vailla olevat eläimet, on moraalisesti haitallista myös ihmisille itselleen. Vaikka pitäytyisimmekin väitteessä, jonka mukaan ei-inhimilliset eläimet eivät olisi moraalisia toimijoita, niiden kohtelulla on kuitenkin tärkeä moraalinen merkitys meille itsellemme. Jos suhtaudumme ”järjettömien eläinten” kärsimyksen välinpitämättömästi, se lisää taipumus- tamme kohdella myös toisia ihmisiä julmasti. Siksi julmuus on väärin sen kohteen moraalista statuksesta riippumatta.

Argumentti perustuu tietynlaiseen käsitykseen moraalista autonomiasta. Kuten edellä totesimme MacIntyren korostavan, moraalinen toimijuus edellyttää kykyä ja taipumuksia, jotka edellyttävät harjoitusta. Ihmisten pitää harjoitella tunnistamaan

ja käsittelemään kärsimystä itsessään ja muissa sekä opetella, miten toisia ja itseä autetaan ja miten omia tunteita voi hallita. Ilman harjoitusta empatia ja myötätunto jäävät kehittymättä. Pahimmassa tapauksessa ihmiset voivat pitää julmuutta aivan normaalina ja näin kovettaa sydämensä joitakin kärsimyksen muotoja kohtaan kokonaan.

Blade Runner 2049 -elokuvan pohjalta tällainen argumentti voidaan analogisesti muodostaa myös älykkään teknologian maailmassa. Kun arkitodellisuutta täyttävät erilaiset älykkäät järjestelmät, joilla ei kuitenkaan ole moraalisen toimijan statusta, ihmiset tottuvat julmuuteen ja välinpitämättömyyteen. Tämä ei voi olla muokkaamatta heidän keskinäisiä suhteitaan sekä suhdetta omaan itseensä. Jos totun kohtelevaan ihmisenkaltaista toimijaa kuin saastaa, koska tämä ei ole ihminen, vedän lokaan omat kykyni kohdella muita ihmisiä ja omaa itseäni kuten moraalista toimijaa kohdellaan.

Frischmann ja Selinger huomauttavat, että vaikutukset näkyvät jo nyt käytössä olevassa teknologiassa. Otetaan esimerkki moraalisesti latautuneen sosiaalisen vuorovaikutuksen alueelta. Ihmisen vuorovaikutuksen kyvyt ovat evoluution aikana kehittyneet nimenomaan tilanteisiin, joissa yksilöt kohtaavat toisensa kasvokkain. Monenlaiset moraaliset ja sosiaaliset tunteet aina häpeästä kiitollisuuteen säätelevät vuorovaikutustamme. Nämä tunteet eivät kuitenkaan herää kovinkaan helposti tilanteessa, jossa keskustelukumppanista on näkyvillä vain tekstiä kännykän ruudulla. On psykologisesti paljon helpompaa pilkata toista ihmistä tällaisessa tilanteessa verrattuna siihen, että keskustelukumppani istuisi vastapäätä.

Kuvitellaan lähitulevaisuus, jossa älykäs teknologia on yhä enemmän läsnä ihmisten elämässä, esimerkiksi palvelualoilla, ja jossa tekoäly kykenee yhä laadukkaampaan inhimilliseen vuorovaikutukseen. Kun hoidan pankkiasioitani, on yhä vaikeampi erottaa, onko vastapuolella *Homo sapiens* vai ei. Jos ihmiskeskeysydestä on pidetty kiinni, älykkäitä järjestelmiä ei kuitenkaan pidetä moraalisinä toimijoina eikä niillä ole sellaisten oikeuksia.

Ei ole vaikea kuvitella, että ihmiset voivat tottua kohtelemaan yhä ”inhimillisempiä” järjestelmiä välinpitämättömästi, koska niillä ei kuitenkaan ole ihmiskollegojensa vastuuta ja valtaa. Tätä voi jo nyt kokeilla chattibotin kanssa pankkiasioita hoidellessa.

Ihmiskeskeisyyden dilemman ensimmäinen sarvi on siis se, että jos älykkäät järjestelmät rajataan kokonaan moraalisen kohtelun ulkopuolelle, moraalisen autonomian edellytykset omalla kohdallamme voivat kärsiä. Dilemmalla on myös toinen sarvi, joka syntyy siitä, että älykkäät järjestelmät pyritään ottamaan moraalisen ja inhimillistävän kohtelun piiriin. Ongelmia syntyy siksi, että nykyiset älykkäät järjestelmät ja niiden lähitulevaisuudessa todennäköisesti syntyvät kollegat eivät todellisuudessa ole moraalisia toimijoita, jotka pääsisivät lähellekään inhimillistä moraalista ja sosiaalista älykkyyttä.²³

Argumentin voisi esittää vaikkapa näin. Moraalisten toimijoiden yhteisön keskeisenä piirteenä ovat vastuussa pitämisen ja vastuunkannon käytännöt. Kapeat tekoälyt eivät kykene moraaliseen vastuullisuuteen, eivätkä ne ole oikeushenkilöitä. Ihmisten arkikäsitys vapaudesta ja vastuusta edellyttää, että vastuullisella toimijalla on kyky tunnistaa asiaankuuluvat moraaliset periaatteet ja ohjata tekojaan oman tahtonsa mukaisesti. Otamme toimijoiden teot ja omat (tahdonalaiset) tekomme merkkeinä siitä, millaisia me olemme, ja kohtelemme itseämme ja muita sen mukaisesti. Joku on luotettava, toinen epäluotettava; joku taas rohkea ja toinen pelkuri. Tällaiset päätelmät ovat pitkälti automaattisia ja intuitiivisia.²⁴ Mitä niille tapahtuu maailmassa, joka on täynnä järjestelmiä, jotka kykenevät yhä inhimillisempään toimintaan mutta joilla ei kuitenkaan ole tahtoa, luonnetta eikä sosiaalisen elämän edellyttämiä tunteita?

Meidän ei tässä yhteydessä tarvitse vedota vain fiktiosta poimittuihin esimerkkeihin, kuten edellä on tehty. Ihmiset ovat jo ottaneet hyvin inhimillisen (ja intiimin) kohtelun piiriin erilaisia keino-olioita. Olemme lukeneet reportaaseja ihmisistä, jotka haluavat naimisiin robottiensa kanssa. On tunnettua, että monet robotit suunnitellaan nimenomaan niin, että ne herättäi-

sivät mahdollisimman vahvoja moraalisia ja sosiaalisia reaktioita, vaikka näillä roboteilla ei olekaan minkäänlaisia moraalisen toimijuuden kykyjä. Näin ne synnyttävät illuusion moraalisesta toimijuudesta, vaikka tosiasiasa ne ovat kyvyttömiä esimerkiksi seuraamaan normeja, vaatimaan oikeuksiaan, tekemään lupauksia ja sitoumuksia sekä ilmaisemaan tunteitaan ja ottamaan toisten tunteet huomioon.

Nämä kyvyt ovat kuitenkin edellytys kaikkein rikkaimmille ja tärkeimmille moraalisille ja sosiaalisille suhteille, joita ihmisten välillä on. Moraalinen ja sosiaalinen elämämme on näiden normien ja tunteiden jatkuvaa käsittelyä. Ajatellaan vaikkapa romanttista rakkautta. Robotin ja sen omistajan suhde on kuitenkin hyvin kaukana siitä, miten rakkaussuhde vastuullisten ja autonomisten persoonien välillä tavallisesti nähdään. Ohjelman tai robotin ja sen käyttäjän suhde on lähinnä omistussuhde tai käyttösuhde. Romanttinen rakkaus puolestaan edellyttää autonomisia yksilöitä, uhrautumista, tunteita ja niiden käsittelyä sekä ruumiillista yhteyttä tavalla, joka samanaikaisesti luo velvollisuuksia ja motivaatioita. Nämä ovat juuri moraalisen toimijuuden ydintä. Jos rakkaussuhteen prototyypiksi otetaankin käyttäjäsuhte, tukee se kehitystä, jossa ”rakkaussuhde” käsitteään lähinnä seksuaalisen tyydytyksen ja vain toisen osapuolen päämäärien ja tarpeiden täyttymisenä. Tämä ei ole ongelma pelkästään yksilön moraalipsykologian kannalta, vaan se voi myös edistää yhteisön normien muutosta.²⁵

Lopuksi

Olen pyrkinyt vakuuttamaan lukijan siitä, että moraalisen toimijuuden edellytyksiä tekoälyn maailmassa on syytä pohtia tarkasti. Nostin esiin ihmiskeskeisyyden dilemman yhtenä esimerkkinä ongelmista, joita kohtaamme tulevaisuudessa, mutta en kuitenkaan pyrkinyt antamaan siihen minkäänlaista vastausta. Toivon tällä osoittaneeni, ettemme voi eristää tekoälyn etiikan

ongelmia pelkästään tekniikan etiikan tai jonkin muun etiikan erityisalueen piiriin. Pikemminkin tekoälyn maailma haastaa koko moraalisen ajattelumme ja käytäntömme pohjiaan myöten, ja nämä haasteet tulee kohdata hyvin kokonaisvaltaisella eettisellä tarkastelulla. Joudumme poliittiseen, eettiseen, filosofiseen ja jopa uskonnolliseen keskusteluun moraalista ja ihmisyydestä, koska emme voi välttää tekemästä tekoälyä koskevia päätöksiä. Tässä keskustelussa panokset ovat kovat: kysymys on siitä, mitä moraalit ja ihmisyydet lopulta ovat.

Moraalinen toimijuus ja ihmiskeskeisyyden dilemma tekoälyn maailmassa

1. Tässä artikkelissa on käytetty materiaalia toisesta artikkelistani ”Uhkaako tekoäly ihmisen itseymmärrystä” (Visala 2018a) aikakauslehti *Futuran* luvalla.
2. Ks. esim. Coeckelbergh 2019; Nyholm 2020.
3. Frischmann & Selinger 2018.
4. Keskustelua moraalista toimijuudesta esittelee Hallamaa 2017.
5. Seuraan jatkossa MacIntyren näkemystä (1999). Kattavan ja hyödyllisen kokonaisnäemyksen yksilön moraalisten ja sosiaalisten kykyjen kehityksestä tarjoaa Tomasello 2019.
6. Searle 2001. Filosofista keskustelua vapaasta tahdosta esittelee suomeksi Visala 2018b.
7. Teknologian filosofian historiasta ks. Coeckelbergh 2019.
8. Frischmann & Selinger 2018.
9. Baumeister 2005.
10. Dick 2015, 77.
11. Weintraub 2014.
12. Martin & Barresi 2018.
13. Baumeister 2005.
14. Gray 2007.
15. Kurzweil 2005.
16. Ks. keskustelu Fuentes & Visala 2018. Eräs empiirinen tutkimus, joka viittaa tähän suuntaan on Lindquist ym. 2011.
17. Wilson & Haslam 2012.
18. Bloom 2005.
19. Vainio 2018.
20. Ks. esim. Plumwood 2001.
21. Aaltola 2013.
22. Ks. esim. O’Neil 1997.
23. Ks. esimerkiksi Antti Kauppisen kirjoitus tässä teoksessa.
24. Moraalisen vastuun ulottuvuuksia erittelee Visala 2018b.
25. Ks. Nyholm 2020.

+ + +

Aaltola, Elisa (toim.) (2013). *Johdatus eläinfilosofiaan*. Helsinki: Gaudeamus.

Baumeister, Roy (2005). *The Cultural Animal: Human Nature, Meaning, and Social Life*. New York: Oxford University Press.

Bloom, Paul (2005). *Descartes’ Baby: How the Science of Child Development Explains What Makes Us Human*. New York: Basic Books.

Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. New York: Oxford University Press.

Coeckelbergh, Mark (2019). *Introduction to Philosophy of Technology*. New York: Oxford University Press.

- Dick, Steven (toim.) (2015). *The Impact of Discovering Life Beyond Earth*. Cambridge: Cambridge University Press.
- Gray, John (2007). *Straw Dogs: Thoughts on Humans and Other Animals*. New York: Farrar, Strauss, Giroux.
- Frischmann, Brett & Evan Selinger (2018). *Re-Engineering Humanity*. Cambridge: Cambridge University Press.
- Fuentes, Agustin & Aku Visala (toim.) (2018). *Verbs, Bones, and Brains: Interdisciplinary Perspectives on Human Nature*. Notre Dame: University of Notre Dame Press.
- Hallamaa, Jaana (2017). *Yhdessä toimimisen etiikka*. Helsinki: Gaudeamus.
- Kurzweil, Ray (2005). *The Singularity Is Near*. New York: Viking.
- Lindquist, Stefan, Edouard Machery, Paul Griffiths & Karola Stotz (2011). Exploring the Folkbiological Conception of Human Nature. *Philosophical Transactions of the Royal Society* 366:1563, 444–453.
- MacIntyre, Alasdair (1999). *Dependent Rational Animals: Why Human Beings Need the Virtues*. Chicago: Open Court.
- Martin, Raymond & John Barresi (2008). *The Rise and Fall of Soul and Self: An Intellectual History of Personal Identity*. New York: Columbia University Press.
- Nyholm, Sven (2020). *Humans and Robots: Ethics, Agency, and Anthropomorphism*. London: Rowman & Littlefield.
- O'Neill, Onora (1997). Environmental Values, Anthropocentrism and Speciesism. *Environmental Values* 6:2, 127–142.
- Plumwood, Val (2001). *Environmental Culture: The Ecological Crisis of Reason*. London: Routledge.
- Searle, John (2001). *Rationality in Action*. Cambridge, MA: The MIT Press.
- Tegmark, Max (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Alfred A. Knopf.
- Tomasello, Michael (2019). *Becoming Human: A Theory of Ontogeny*. Cambridge, MA: Harvard University Press.
- Vainio, Olli-Pekka (2018). *Cosmology in Theological Perspective: Understanding Our Place in the Universe*. Grand Rapids: Baker.
- Visala, Aku (2018a). Uhkaako tekoäly ihmisen itseymmärrystä? *Futura* 37:4, 20–28.
- (2018b). *Vapaan tahdon filosofia*. Helsinki: Gaudeamus.
- Weintraub, David (2014). *Religions and Extraterrestrial Life: How Will We Deal with It?* Dordrecht: Springer.
- Wilson, Samuel & Nick Haslam (2012). Reasoning about Human Enhancement: Towards a Folk Psychological Model of Humanness and Human Identity. Teoksessa R. Luppicini (toim.): *Handbook of Research on Technoself: Identity in a Technological Society*. Hershey: IGI Global, 175–188. Saatavissa: <www.researchgate.net/publication/236030914_Reasoning_about_human_enhancement_Towards_a_folk_psychological_model_of_humanness_and_human_identity>