

Antti Kauppinen (Helsingin yliopisto) & Arto Laitinen (Tampereen yliopisto)

Älykkäät työkalut, robotit ja sovellukset mahdollistavat monenlaisten päämäärien entistä tehokkaamman toteuttamisen. Silloin kun päämäärät ovat hyviä ja keinot niiden tavoitteluun eivät riko kenenkään oikeuksia, tekoäly on hyvä renki. Mutta se, mikä hyödyttää yhtä, ei välttämättä hyödytä toista, koska ihmisten edut voivat olla ristiriidassa. Sen takia on tärkeä varmistaa sekä lainsäädännöllisesti että itsesääntelyllä, että tekoälyn laajamittainen käyttöönotto ei pahenna epäoikeudenmukaisuutta, loukkaa ihmisten oikeuksia ja tasavertaisuutta, eikä muutenkaan tee yhteiskunnasta huonompaa paikkaa elää. Hahmottelemme tässä joitakin periaatteita ja peukalosääntöjä, jotka voisivat ohjata aiheen pohdintaa.

§1 Keskitason moraaliperiaatteet ovat hyvä lähtökohta tekoälyn etiikan tarkasteluun, parhaan yleisen teorian löytymistä ei kannata odottaa

Vaikka seurausetiikka (maksimoi hyvinvointia) ja kantilainen velvollisuusetiikka (noudata sääntöjä, joita voidaan universaalisti tahtoa noudatettavan) tunnetaan melko laajasti, samoin kuin niiden kilpailijat hyve-etiikka (toimi kuten hyveellinen, käytännöllisesti viisas toimisi), hoivaetiikka (ethics of care) sekä kontraktualismi (toimi vain niin, että kenelläkään asianosaisella ei ole riittäviä perusteita hylätä toimintasi periaatteita), niiden paremmuuden vertailu ei lupaa ratketa lähivuosikymmeninä. Saattaa olla, että nämä yleiset teoriat kiipeävät samaa vuorta eri puolilta, ja huipulla huomataan, että ne ovat muotoiltavissa yhteensopiviksi (Parfit). Tämä kannattaa jättää filosofisen etiikan tutkimuskysymykseksi.

Parempi lähtökohta ovat keskitason periaatteet, jotka nekin ovat mahdollisesti a priori kaikkien (joilla on relevanttia kokemusta tai kosketusta asiaan) tiedettävissä ja joita koskien vallitsee laaja konsensus:

- välttää kärsimyksen aiheuttamista ilman riittävän hyvää syytä
- edistä toisten hyvinvointia ainakin jos siitä ei koidu merkittäviä kustannuksia itsellesi
- kunnioita jokaisen autonomiaa (kykyä tehdä itsenäisiä valintoja), ellei siitä seuraa merkittävää vahinkoa muille,
- kunnioita jokaisen ihmisarvoa ja perusoikeuksia (yksityisyyttä, ruumiillista loukkaamattomuutta, ...)

- kunnioita oikeudenmukaisuuden vaatimuksia, kohtelee jokaista tasavertaisina
- älä tuhoa luontoa tai tulevien sukupolvien elinmahdollisuuksia

...

Jokaisen keskitason periaatteen ytimestä löytyy ajatus jostain hyvästä, arvokkaasta päämäärästä, jota tavoitella (hyvinvointi, autonomia, ihmisarvon mukainen elämä, oikeudenmukaisuus, luonnon monimuotoisuus ja sukupolvien oikeudenmukaisuus), mutta ne ovat myös kunnioittamista vaativia normeja (kärsimyksen ja vahinkojen aiheuttaminen on väärin; merkittäviä hyvinvoinnin edistämisiä on väärin jättää tavoittelematta; paternalismi on väärin; on väärin loukata oikeuksia; on väärin vain maksimoida hyvää ottamatta huomioon oikeudenmukaisuutta; on väärin tuhota luontoa tai seuraavien sukupolvien elämän edellytyksiä). Eri ihmisillä ja kulttureilla on erilaisia käsityksiä siitä, kuinka nämä arvokkaat asiat tulisi ymmärtää ja kuinka ne konkreettisesti toteutuvat, mutta nämä käsitykset ovat ainakin osittain päällekkäisiä tai yhteneviä.

Näitä periaatteita voi siten pitää lähtökohtaisesti yleispätevinä, ja lisäksi niihin on Suomen perustuslaissa, lainsäädännössä ja kulttuurissa sitouduttu: ne ovat ainakin suomalaisia periaatteita. Perustuslaissa todetaan esimerkiksi, että ”Jokaisella on oikeus elämään sekä henkilökohtaiseen vapauteen, koskemattomuuteen ja turvallisuuteen” (§7) ja että ”Jokaisen yksityiselämä, kunnia ja kotirauha on turvattu” (§10). Vastaaviin periaatteisiin on sitouduttu myös YK:n julistuksissa.

Kun keskitason periaatteita *sovelletaan* tekoälyn kysymyksiin, tarvitaan ensinnäkin tutkittua empiiristä tietoa tekoälyjärjestelmistä, sekä teknologian että yhteiskunnallisten vaikutusten osalta. Sen lisäksi tarvitaan kuitenkin myös arvostelukykä, sillä tällaiset periaatteet voivat joissakin tapauksissa olla ristiriidassa keskenään, ja ne sallivat poikkeuksia. Siksi olemme muotoilleet joitakin esimerkkiperiaatteita siten, että niissä puhutaan ”riittävän hyvästä” syystä tai ”merkittävistä” kustannuksista. Se, mikä on riittävän hyvää tai merkittävää riippuu tilannekohtaisista tekijöistä, jotka voivat olla hyvinkin hienovaraisia. Joskus on toki mahdollista muotoilla myös yleisiä sääntöjä ensisijaisuudelle – on esimerkiksi oikeutettua aiheuttaa kärsimystä tai vahinkoa oman ruumiillisen loukkaamattomuutensa puolustamiseksi, kun ei ole itse muuten tekemässä vääryyttä toisille. Tällaisten ensisijaisuussääntöjen muotoileminen on yksi tärkeä tehtävä tekoälyn etiikan pohdinnassa – onko esimerkiksi oikeutettua käyttää ihmisten yksityisyyden piiriin kuuluvia terveystietoja ilman heidän nimenomaista lupaansa uusien hoitomuotojen kehittämiseen, jos niistä voisi olla hyötyä monille muille? Vastaukset tällaisiin kysymyksiin eivät suinkaan ole itsestään selviä, vaikka lähtökohtana olevat periaatteet ja empiiriset tosiasiat olisivat tunnettuja.

§2 Robotit ja tekoälyjärjestelmät itse eivät ole, eivätkä vuosikymmeniin tule olemaan, moraalisten oikeuksien kantajia (moraalipersonia), moraalisten vastuuden kantajia (moraalitoimijoita) tai moraalisten arvostelmien tekijöitä (moraalisubjekteja).

On hyviä syitä uskoa, että tekoälyjärjestelmät eivät täytä moraalipersonuuden ehtoja, moraalitoimijuuden ehtoja, tai ole aitoja moraalissubjekteja. Keskusteluissa moraalisten oikeuksien kantajista on tavallista erottaa kaksi tasoa. Jos yksilöllä on *moraalista merkitystä*, heitä ei voi kohdella mielivaltaisesti, toisin kuin esimerkiksi omistamiaan esineitä. Jos heillä on *täysi moraalinen status*, heidän tahtonsa ja tarpeensa asettavat rajoitteita toisten ihmisten hyvienkin päämäärien tavoittelulle – heitä ei esimerkiksi voi uhrata yleisen edun edistämiseksi. Uskottavin kriteeri moraalille merkitykselle on *tuntoisuus*, kyky tuntea kärsimystä ja mielihyvää. Nyökkisääntönä voisi sanoa, että jollakin yksilöllä on väliä, koska asioilla on väliä hänelle.

Täyden moraalisen statuksen edellytykset ovat kiistanalaisempia, koska riman asettaminen korkealle sulkee helposti pois esimerkiksi vastasyntyneet ja tietyillä tavoilla vammaiset. Jos nämä haasteet jättää syrjään, hyvä lähtökohta voisi olla vaikka *omasta tulevaisuudesta välittäminen ja kyky toimia järkevästi päämääriensä toteuttamiseksi monenlaiset perusteet huomioiden*.

Näiden lähtökohtien mukaisesti kaikilla ihmisillä ja monilla eläimillä on moraalista merkitystä, koska he ovat tunteisia, ja ainakin normaaleilla aikuisilla on täysi moraalinen status. Mutta tekoälyjärjestelmät eivät ole tunteisia, ne eivät välitä mistään, eikä niillä ole omia päämääriä. Näiden asioiden suhteen ei myöskään ole tapahtunut pienintäkään kehitystä missään vaiheessa, vaikka järjestelmien muut kyvyt ovat kehittyneet valtavasti. Tämä antaa syyn uskoa, että keinotekoinen tuntoisuus, kyky välittää, ja kokonaisvaltainen arvostelukyky eivät toteudu näköpiirissä olevassa tulevaisuudessa. Tämä ei ole siinä mielessä yllättävää, ettei ihmisenkään tietoisuuden fysikaalista perustaa oikeastaan ymmärretä. (Periaatteessa mikään ei kuitenkaan estä sitä, etteivät tällaiset kyvyt voisi toteutua myös ei-biologisilla olioilla.)

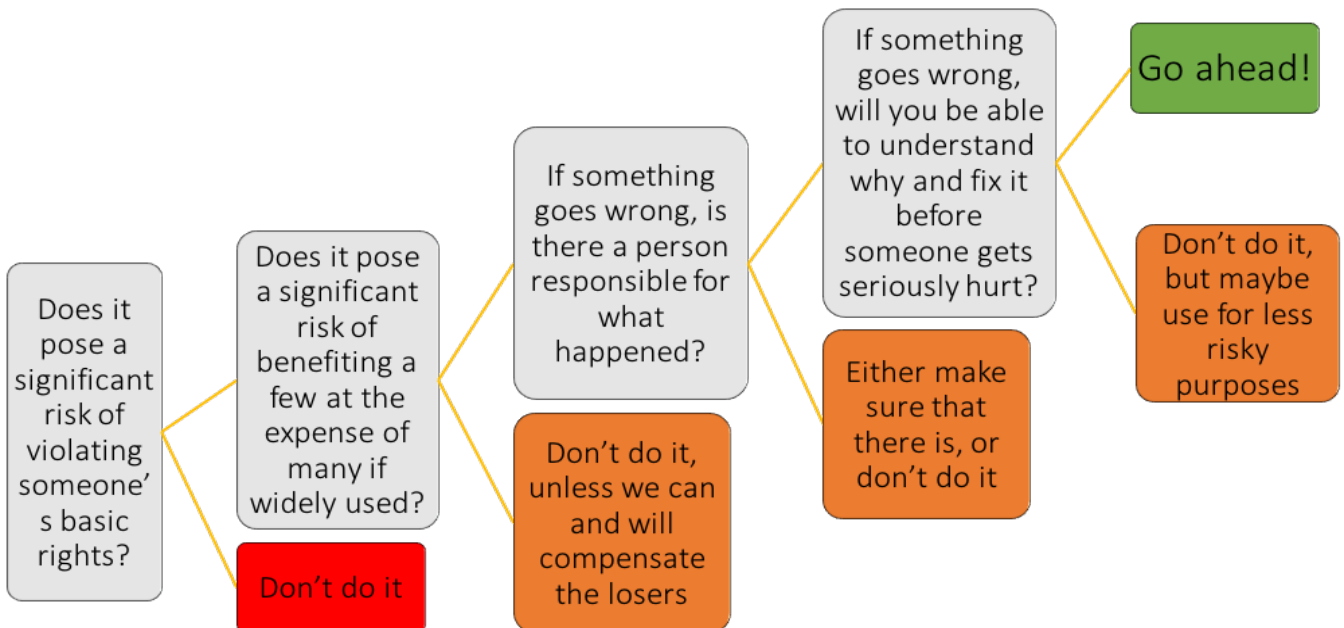
Samat kyvyt ja ominaisuudet liittyvät myös moraaliseen vastuuseen. Jotta voisimme mielekkäästi sanoa jonkun olevan moraalisesti vastuussa jostain, täytyy olla ainakin periaatteellisesti mahdollista *pitää häntä vastuussa*. Mutta älykästäkin konetta ei voi rangaista eikä sitä ole mielekäästä paheksua. Meidän asenteillamme ei ole sille väliä, sitä ei voi satuttaa, eikä sillä ole ainakaan sellaista arvostelukykyä, jota toisiltamme vaadimme. Yksi tärkeä osatekijä tällaisessa arvostelukyvyyssä on empatia, kyky asettua toisen asemaan ja nähdä asiat hänen tolaltaan.

Yleisesti voisi sanoa, ettei koneella ole *omaa näkökulmaa maailmaan*. Se voi reagoida ympäristöönsä älykkäästi, mutta ei *koe* sitä eikä *välitä* mitä tapahtuu, vaikka se saattaakin tavoitella päämääriä, joita ei ole siihen suoraan ohjelmoitu. Siksi se ei ole myöskään aito moraalisten arvostelmien tekijä.

§3 NELJÄ KYSYMYSTÄ ENNEN TEKOÄLYSOVELLUKSEN KÄYTTÖÖNOTTOA

3.1. Robotit ja tekoälyjärjestelmät voivat jo nyt kuitenkin tuottaa hyvinvointia tai kärsimystä, ja niiden hyödyt ja haitat voivat jakautua epätasaisesti. Ne on ennen kaikkea rakennettava edistämään hyvinvointia, kunnioittamaan autonomiaa, yksilöiden perusoikeuksia ja oikeudenmukaisuuden vaatimuksia sekä välttämään kärsimyksen tuottamista (eli ottamaan keskitason periaatteet huomioon).

Eettisessä ”vuokaaviossa” tämä on ensimmäinen ja keskeisin vaatimus.



§3.2. Aikomattomat yhteiskunnalliset kerrannaisvaikutukset tulee ottaa huomioon ja ehkäistä eriarvoistavat vaikutukset.

Toiseksi, ei-aiotut yhteiskunnalliset vaikutukset tulee ottaa huomioon. Jos halutaan tehdä ”Ihmiskokeita”, on varmistettava, että eriarvoistavia mekanismeja vastustavia mekanismeja on käytössä.

§3.3 Vastuukysymykset tulee ratkaista etukäteen.

ks. pohjapaperi vastuukysymyksistä.

§3.4 Institutionaalis-teknologisen prosessin kyky oppia virheistä tulee taata. Kyky ymmärtää ja mielekäs inhimillinen kontrolli ovat keskeisiä

Elämme aina keskellä historiallista prosessia, yrityksen ja erehdyksen prosessia. Koko yhteiskunnallis-teknologista kokonaisuutta voi lähestyä eräänlaisena tutkimuksen prosessina. Keskeistä kyseiselle prosessille on kyky ymmärtää mistä virheet johtuvat ja kyky korjata ne.

Kun näihin neljään kysymykseen on saatu oikea vastaus, voidaan eettisesti ottaa tekoälysovellus käyttöön.

§4 ROBOTIIKAN JA TEKOÄLYN ETIIKAN ERITYISKYSYMYKSIÄ

On lukuisia tekoälyn etiikkaan liittyviä erityiskysymyksiä, joista tässä mainitsemme vain muutamia.

4.1 Mitä päätöksiä koneet eivät saa tehdä?

Kun tekoäly kehittyy, sitä sovelletaan epäilemättä yhä useampien ongelmien ratkaisuun, monesti hyvällä syyllä. On kuitenkin asioita, jotka tulee periaatteellisista tai käytännöllisistä syistä pitää ihmisten eikä koneiden käsissä. Tässä on yksi ehdotus, joka perustuu nykyisten ja näköpiirissä olevien tekoälyjärjestelmien kykyjen rajoituksiin:

Sellaiset päätökset, jotka a) vaikuttavat merkittävästi yksilöiden elämään ja b) vaativat kokonaisvaltaista tarkastelua, jossa punnitaan monia eriluonteisia seikkoja, täytyy jättää ihmisille, tai ainakin ihmisellä tulee olla olennainen rooli prosessissa.

Tällaisten ratkaisujen tekijän on oltava niistä vastuussa toisille ihmisille, erityisesti niille, jotka voivat joutua kärsimään niistä. (Heidän täytyy siis olla moraalisia toimijoita yllä käsitellyssä mielessä.) Vain sillä ehdolla valta on yhteensopivaa ihmisten perustavan moraalisen

tasavertaisuuden kanssa. Koneita ei voi rangaista, eikä siltä voi odottaa kokonaisnäkemystä, joten sille ei voi delegoida tällaisia kysymyksiä. Tämä tarkoittaa, että esimerkiksi poliittisia tai sotilaallisia päätöksiä, tai poliisin tai psykiatrin työtä, ei saa antaa koneiden tehtäväksi. Tämä ei perustu vain päätösten oletettuun laatuun, vaan myös niiden merkitykseen ihmisten välisille suhteille. Tekisi vääryyttä vanhemmille, jos koneen annettaisiin tehdä päätös heidän lapsensa huostaanotosta, vaikka sosiaalityöntekijä olisi tehnyt aivan saman päätöksen. Sen sijaan on sallittavaa, että vaikkapa tiettyihin työtehtäviin valitaan tekoälyn avulla, tai että robotti tekee vaikeita leikkauksia tai ajaa autoa, kunhan se tekee nämä asiat ainakin yhtä hyvin kuin ihminen.

4.2. Algoritmien ylivalta, vinoumat ja diskriminaatio

Nk. Joukkotuhoalgoritmista tai matikkatuhoaseesta (*Weapon of math destruction*) on datatieteilijä Cathy O'Neilin mukaan kyse, kun kolme kriteeriä täyttyy: *algoritmi on toimintaperiaatteeltaan läpinäkymätön, se aiheuttaa epäreilua vahinkoa tai kärsimystä sekä on laajalle levinnyt.*

Millä tavoin tekoälyä olisi niiden osalta muutettava paremmaksi? Voidaan erottaa viisi eri seikkaa: tavoitteet, aineistojen vinoumat, palautteen puute, läpinäkyvyyden puute sekä joissain tapauksissa mitattavuuden mielekkyys ylipäättään.

Yksi kysymys on se, *mihin algoritmien avulla tähdätään.* Jos puolue saa selville, että jotkut ovat vihaisia asuinolosuhteista, se voi käyttää tätä tietoa poliittisen mainonnan taustatietona tai hankkeissa asuinolosuhteiden parantamiseksi. Yritys voi käyttää tietoja voittojensa maksimoimiseen tai tuotteidensa parantamiseen. Informaatioidankäyntiä harrastava ulkovalta voi pyrkiä yhteiskunnallisen epäluottamuksen ja kahtiajakojen lisäämiseen. Monissa tapauksissa pyrkimys sinänsä on hyvä: pyritään parantamaan sairaalan toimintaa, saamaan selville parhaat hakijat, mitoittamaan vakuutusmaksut riskeihin jne. Miksi hyvään tähtäävät algoritmit kuitenkin saattavat aiheuttaa vahinkoa?

Algoritmit toistavat aineistojen vinoumia. Keskeinen seikka on, että algoritmit eivät ole täysin yksilöityjä, vaan voivat perustua mihin tahansa ihmisten välisiin samankaltaisuuksiin, joita kone löytää. Ja nämä samankaltaisuudet sisältävät rakenteellisia epäoikeudenmukaisuuksia, joita se kohtelee neutraalina datana. Kenties postinumerosi alueella tapahtuu paljon rikoksia. Tämä on epäreilua yksilöä kohtaan, ainakin jos ajatellaan mahdollisuuksien tasa-arvoa. Syrjinnän kielto vaatii jättämään osan datasta sivuun silloinkin, kun se olisi merkityksellistä: vaikka postinumero korreloisikin joidenkin riskien kanssa, niin tiedon käyttö saattaa olla rodullista syrjintää. (Osa datasta, joka saattaa vinouttaa päätelmiä, voi olla yksilöllistäkin: omat aiemmat

mielenterveysongelmat, sovitettut rikokset tai luottohäiriöt voivat vaikuttaa algoritmin arvioon työntekijän sopivuudesta.).

Algoritmit tekevät virheitä. Ne tarvitsevat palautetta virheiden karsimiseksi. Joskus palautetta on helppo antaa: Käyttäjä voi lisätä sanoja tekstinkäsittelyohjelman oikolukuohjelmaan.

Urheilujoukkueiden suoritustilastoihin perustuvia ennusteita seurataan ja menestystä ennustavien tekijöiden painoarvoa viilataan jatkuvasti. Sen sijaan laajoihin joukkoihin kohdistuvia esimerkiksi työnhakualgoritmeihin sisältyviä vinoumia ja virheitä kenelläkään ei ole motivaatiota poistaa, ainakaan ellei tätä lainsäädännöllisesti tehdä pakolliseksi.

4.3. Oikeus saada tietää, mihin itseä koskevat ratkaisut perustuvat (*Right to explanation*)

Koneoppimista toteuttavat algoritmit eivät ole läpinäkyviä, vaan ovat mustia laatikoita, joiden toimintaperiaatetta ei ole ennalta ohjelmoitu. Koneoppiminen ei noudata ennalta koodattua kaavaa, vaan edes ohjelmoijat eivät tiedä millä perusteella algoritmit suosituksiinsa päätyvät. Eräs tapa yrittää puuttua tähän on testata algoritmia koeaineistolla, ja saada tietoa millaisia mahdollisia vinoumia se sisältää, vaikka ”musta laatikko” jääkin avaamatta. Tällaista tutkimusta onkin meneillään (joskin yritykset saattavat vaikeuttaa hypoteettisten henkilöiden syöttämistä systeemeihinsä testausmielessä). GDPR antaa jollain tavalla yksilöille ”oikeuden selityksen saamiseen”. Oleellinen kysymys johon selitys tarvitaan, on: *minkä olisi pitänyt olla toisin, jotta päätös olisi ollut toinen?* Meidän ei tarvitse tietää yksityiskohtaista kausaalista ketjua ”konepellin alla”, mutta relevantti kysymys on käytännöllinen: mitkä tekijät vaikuttivat lopputulokseen?

Läpinäkyvyys, periaatteellinen avoimuus ja julkisuus ovat aivan keskeisiä sosiaalisen oikeudenmukaisuuden kannalta: vaikka algoritmi toimisi virheettömästi, ilman vinoumia, tämän voi tietää vain jos kytetään tietämään millä perusteella algoritmi on suosituksiinsa päätenyt.

Yhteiskunnallisen luottamuksen kannalta oikeudenmukaisuuden pitäisi paitsi tapahtua, myös näkyä. On mahdotonta ylläpitää luottamusta algoritmeihin, joiden toimintaperiaatteita ei tunneta – etenkin kun on syytä epäillä, että kaikkia virheitä ei ole vielä kitketty pois, ja niiden alkuperäinen data sisältää vinoumia. Oikeus selityksen saamiseen ei vielä tavoita filosofi Rainer Forstin peruseriaatetta ”oikeus oikeutuksen saamiseen”: toisinaan selitykset eivät vielä ole oikeutuksia; epäoikeutetulla toiminnallakin on selityksensä.

4.4. Data ja alistaminen: tieto on valtaa

Tekoälysovellukset mahdollistavat toisten ihmisten tarkkailun, hallinnan ja manipuloinnin aivan uudella tavalla. Kiinassa ja Yhdysvalloissa poliisi kokeilee jo hahmontunnistukseen perustuvia

järjestelmiä, jotka tunnistavat valvontakameran tai vartalokameran kuvasta etsintäkuulutettuja rikollisia. Tiedustelupalvelut ympäri maailmaa analysoivat automaattisesti miljardeja sähköposteja ja puheluita terrorismin merkkien varalta. Google, Apple tai Facebook tietävät, missä olet milloinkin ollut, mistä olet kiinnostunut ja mistä tykkäät. Parhaassa tapauksessa tämä johtaa siihen, että saat mainoksia joista on sinulle hyötyä ja konnat saadaan kiinni ennen pahantekoa. Mutta data mahdollistaa myös elämäsi puuttumisen hyvällä tai pahalla tarkoituksella. Parhaat kaverimme ovat aina pystyneet pitkälti ennakoimaan käyttäytymistämme. Järjestelmä jolla on valtavasti tietoa menneisyydestämme pystyy epäilemättä samaan. Jos väkivaltamonopolin omaava valtio ei pidä ennustetusta toiminnastamme, se voi asettaa sille yksilöityjä esteitä. Kuten moni varmaan tietää, Kiinassa kehitetään sikäläisen sosiaalisen median ja valtion yhteistyöllä kansalaisten luottamuspesteytystä, joka tulee lähitulevaisuudessa vaikuttamaan luoton saantiin, matkustuslupiin, ja työpaikan saamiseen.

Ja jos joku pullikoi tämänkaltaista systeemiä vastaan, hänet on ennennäkemättömän helppoa palauttaa ruotuun. Jos voimankäyttöä tarvitaan, edes Lucky Luke ei vedä nopeammin kuin poliisirobotti. Vähän isomassa mittakaavassa Suomen armeijalla ei ole kenties mitään keinoa pysäyttää Venäjän uutta Armata-tankkia, joka on pian täysin automatisoitu. Eipä ihme, että Vladimir Putin on sanonut, että se joka hallitsee tekoälyä, hallitsee koko maailmaa.

Voimankäyttöä ei toki tarvitse, jos ihmiset saa muutenkin manipuloitua. Mainostajat, poliitikot ja psykologit ovat jo pitkään perehtyneet sekä käytännössä että teoriassa siihen, kuinka ihmisen intuition ja tunteiden sokeita pisteitä voi käyttää hyväkseen. Emme suinkaan hallitse ajatuksiamme ja aikeitamme siinä määrin, kuin kuvittelemme. Koneoppiva järjestelmä, jolla on käytössään massiivinen määrä dataa yksilöstä ja kenties saa palautetta interventioden tehokkuudesta vaikkapa kasvonilmeisiin perustuvan tunne-analyysin kautta, voi muovata toimintaamme enemmän kuin tajuamme. Siinä määrin kuin tämä onnistuu, äänestysdemokratia menettää merkityksensä.

Tarvitsemme lainsäädäntöä, joka rajoittaa yksilöistä kerättävän informaation määrää, laatua, säilyttämistä ja käyttöä tavalla, joka palvelee yksilön omaa etua.

Esimerkiksi kansalaisten terveystietoja ei saa kerätä tai käyttää muuhun kuin diagnoosin ja hoidon parantamiseen, ei terveydenhuollon kaupallistamisen edistämiseen. Ongelma ei rajoitu tietoja kerääviin yrityksiin, vaan koskee myös valtioita, joiden valtaa kansalaisiinsa ei myöskään tule lisätä. Kuten filosofi Philip Pettit on korostanut, pelkästään se, että joku pystyy halutessaan puuttumaan toimintaamme ilman rajoituksia ja rangaistusta vähentää vapauttamme, vaikka tosiasiallisesti saisimme tehdä juuri niin kuin haluamme, koska satumme miellyttämään

vallankäyttäjää. Orja, jonka isäntä on hyvällä tuulella ja antaa tämän puuhailta rauhassa ei ole vapaa, niin kauan kuin isännällä on edellytykset hänen mielivaltaiseen rankaisemiseensa ja komentelemiseensa. Kyse on alistamisesta (*domination*), vapauden vastakohdasta. Kun toisella on kyky puuttua minkä tahansa oikun takia elämään, se jo sinänsä antaa minulle tärkeän kannustimen mielistellä, nöyristellä ja pokkuroida, kuten alistetussa asemassa olevat ovat kautta historian joutuneet tekemään.

Vapauteen ei siis riitä, etteivät toiset puutu asioihimme, vaan se edellyttää, että he eivät *voi* tehdä niin ainakaan ilman rangaistusta. Tekoälyyn sovellettuna tämä tarkoittaa, että sellaisia tietoja, jotka mahdollistavat toimintamme hallinnan ei tule lainkaan kerätä (mikä on tietysti varmin tapa), tai ainakin, että sen käytölle on tiukat ja vaikeasti kierrettävät (esimerkiksi perustuslailliset) rajoitteet. Silloin (ja vain silloin) kansalaisten ei tarvitse käyttäytyä kuin alamaiset välttääkseen Isoveljen paheksuntaa.

4.5. Itseään ohjaavat kulkuneuvot

Itseohjautuvista autoista puhutaan ehkä liikaakin tekoälyn etiikan yhteydessä, koska ne eivät sinänsä muodosta merkittävää yhteiskunnallista ongelmaa. Mielenkiintoisia kysymyksiä herättää kuitenkin se, että autoa ohjaavan tekoälyjärjestelmän on itsenäisesti ratkaistava elämän ja kuoleman kysymyksiä mahdollisissa onnettomuustilanteissa. Jotkut reunaehdot ovat selviä: jos kenelläkään ei aiheudu väistämistä vaaraa, auton tulee viimeiseen asti välttää toisiin liikkujiin törmäämistä – vaikka seurauksena olisi auton kyydissä olevan viivästyminen tai auton vaurioituminen.

Mutta entä jos mikä tahansa vaihtoehto todennäköisesti johtaa jonkun loukkaantumiseen tai kuolemaan? Tätä kysymystä pohdittaessa on tärkeää pitää mielessä, että kaikilla moraalisten oikeuksien kantajilla on lähtökohtaisesti yhtäläinen oikeus ruumiilliseen koskemattomuuteen, riippumatta esimerkiksi sellaisista asioista kuin iästä, terveydentilasta tai yhteiskunnallisesta hyödyllisyydestä. Tällaiset seikat voi siis unohtaa tässä yhteydessä. (Palaamme hetken kuluttua muihin relevantteihin asioihin.) Luontevana lähtökohtana on vahingon minimoiminen: jos muut seikat ovat samoja, on parempi aiheuttaa vaaraa harvemmille kuin useammille, kun mahdollinen vahinko on samaa luokkaa (esimerkiksi kaikki ovat kuolemanvaarassa). On siten parempi ajaa päin yhtä kuin kolmea jalankulkijaa (jos heidän välillään ei ole muuta relevanttia eroa). Jos taas yhden vaihtoehdon mahdolliset vahingot ovat selvästi lievemmat, auton tulee valita se, vaikka useampi ihminen joutuisi kärsimään (on esimerkiksi parempi ajaa viiden ihmisen varpaiden päältä kuin törmätä suoraan yhteen).

Edellä sanottu lähtee siitä, että kukaan ei ole moraalisesti erilaisessa asemassa. Mutta on ainakin yksi moraalisesti tärkeä tekijä, jolla on merkitystä sille, kuinka auton tulisi ajaa: kuka on vastuussa vaaratilanteesta? Jos kolme ihmistä juoksee tielle sivulleen katsomatta ja ainoa tapa väistää olisi ajaa jalkakäytävälle, jossa kulkee liikennesääntöjä noudattava jalankulkija, on väärin kääntää auto jalkakäytävälle. *Vastuu on siis toinen olennainen moraalisesti relevantti tekijä vahingon minimoimisen lisäksi.*

Nähdäksemme oletuksena täytyy olla, että auton ”ajaja” on oletusarvoisesti vastuussa vaaratilanteista. Hänen valintansa on ollut nousta kulkuvälineeseen, jonka tietää jo etukäteen voivan aiheuttaa vaaraa toisille. Auto tulisi siis ohjelmoida niin, että ensisijaisesti minimoidaan ulkopuolisille koitua riski. (Tämä on täysin vastoin ainakin Mercedes-Benzin ”kuljettajan” etusijalle asettavaa algoritmia.) Nykyteknologia on kaukana siitä, että se kykenisi arvioimaan, kuka on tosiasiallisesti vastuussa tilanteesta.

Tämä on sinänsä jo perusta sille, että ihmisen tulisi hallita ajoneuvoa ympäristössä, jossa vaaratilanteet ovat todennäköisiä – normaali kuljettaja kykenee esimerkiksi erottamaan reteästi tielle hoipertelevan humalaisen ressurkasta, jonka jenginuoret tönäisevät pois jalkakäytävältä, ja reagoimaan asianmukaisesti (humalaisen takia ei pidä vaarantaa sivullisia, mutta on parempi ajaa jengiläisten kuin uhrin päälle). Niin kauan kuin tekoäly ei kykene tällaiseen tekoälytyypin luokitteluun, sitä ei voi ohjelmoida eettiseen toimintaan.

4.6 Muita kysymyksiä

Tekoälyyn ja robotiikkaan liittyviä eettisiä erityiskysymyksiä on lukuisia muitakin, kuten seuraavat: Onko vanhuksilla oikeus inhimilliseen kontaktiin, vai voiko heidän hoitonsa huoletta jättää robotille? Onko seksiroboteilla kenties tarkoittamattomia (ja eettisesti ongelmallisia) vaikutuksia ihmisten välisiin suhteisiin? Pitäisikö sotarobotit kieltää tai ainakin pitää mielekkäässä inhimillisessä kontrollissa? Emme kuitenkaan tässä yhteydessä voi puuttua näihin.

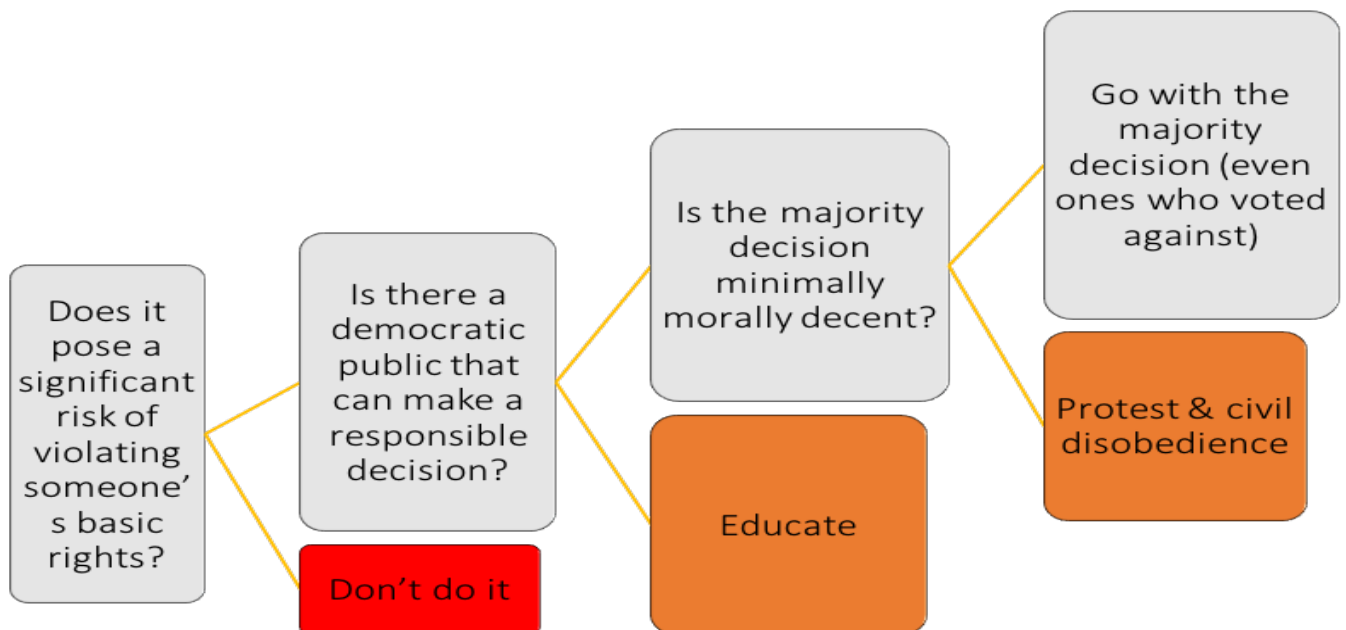
§5 HAASTEITA YHTEISKUNTAETIIKALLE

Kukin yksilö ja yhteisö voi periaatteessa muodostaa oman moraalisen kantansa jokaiseen kysymykseen. Vaikka keskitason periaatteita koskee laaja konsensus, voi syntyä enemmän

erimielisyyttä niiden soveltamisesta käytäntöön. Kun erimielisyyksiä syntyy (moraalisissa ja muissa käytännöllisissä kysymyksissä), tulisiko jokaisen noudattaa ehdottomasti omaa moraalista kantaansa, omaatuntoaan?

Esitämme tässä lopuksi demokratian auktoriteettiin liittyvän näkökohdan:

Sen sijaan, että jokainen noudattaisi kaikissa kysymyksissä ainoastaan omaa näkemystään, tulee sopia yhteisistä normeista, jotka ovat julkisesti tiedettyjä. Yhteiset normit ja päätökset tulee laatia demokraattisesti, ja niitä tulee noudattaa myös niiden, jotka ovat hävinneet äänestyksessä. (Muutoinhan jokainen vain noudattaisi omaa näkemystään, eikä yhteisiä julkisia normeja olisi). Velvollisuudelle noudattaa yhteisiä ja julkisia normeja on kuitenkin kaksi reunaehto: jos ne loukkaavat yksilöiden oikeuksia, niitä ei tule noudattaa. Ja jos ne muutoin ovat moraalisesti muilla tavoin riittävän huonoja, niitä ei tule noudattaa. Näissä tapauksissa kansalaisilla on pikemminkin velvollisuus kansalaistottelemattomuuteen ja julkiseen protestointiin.



5.1 Yksilöiden oikeudet ja demokraattinen päätöksenteko pysyvät keskeisinä lähtökohtina

Jotkut perusoikeudetkin saattavat olla uhattuna uudessa tilanteessa, ne on kyettävä takaamaan. Kollektiivinen vastuu siitä, millaista teknologiaa ja yhteiskuntaa halutaan, tulee lopulta kantaa demokraattisesti: kansalaisilla on lopullinen päätösvalta (joskin ekspertejä tarvitaan asioiden

muotoilemisessa sulatettavaan muotoon). Myös maailmankatsomusten moneuden ja näkemyserimielisyyksien takia oikeudet (mielipiteenvapaus) ja demokratia (harkitsevan keskustelun jälkeen enemmistö ratkaisee) pysyvät parhaina menettelytapoina.

5.2. Koulutuksen rooli: ihmisten erityisyys tekoälyyn verrattuna ja kansalaisten ”velvollisuus tietää”

Koulutuksessa otettava huomioon mihin koneet kykenevät ja koulutettava ihmiset tekemään niitä asioita, joita koneet eivät osaa. Kykenemme parhaimmillamme vaihtamaan viitekehystä ja arvioimaan sekä keinoja että päämääriä monenlaisten perusteiden valossa. Toinen vahvuutemme on luovuus. Osaamme keksiä uusia näkökulmia ja tarkastelutapoja sekä yhdistellä vanhoja asioita uudella mutta tarkoituksenmukaisella tavalla. Näihin molempiin liittyy se, että meillä on tunteita ja kyky asettua toisten asemaan, mikä auttaa ymmärtämään ja kommunikoimaan tilanneherkällä tavalla. Jos meillä on tällaisia kykyjä ja taitoja, on todennäköisintä, että löydämme paikkamme koneiden ohessa, joten koulutuksen tulisi painottua niiden kehittämiseen.

Lisäksi tekoälylukutaitoa lisättävä. Kansalaisvelvollisuuksiin kuuluu oppivelvollisuus ja jonkinasteinen velvollisuus ymmärtää maailmaa, jotta voi tehdä vastuullisia päätöksiä. (Jos tekoäly vapauttaa ihmisiä raadannasta, ehkä vapautuu aikaa panostaa vastuulliseen demokraattiseen kansalaisuuteen?)

5.3 Yhteiskunnalliset peruseriaatteet koskien työtä, pääoman omistusta, rahaa, kulutusta, verotusta, ekologista kestävyyttä ja kansainvälistä oikeudenmukaisuutta jakavat mielipiteitä esimerkiksi vasemmisto-oikeisto – akselilla, ja ne tulisi alistaa demokraattiselle päätöksenteolle; silti jotkut ratkaisut voivat (jokaiselle hyväksyttävissä olevin eettisin perustein) olla moraalisesti kestävämpiä ja oikeuttaa kansalaistottelemattomuuden.

Mahdollisuuksien tasa-arvo on sekä vasemmiston että oikeiston kannalta keskeinen periaate, samoin (republikanismin perinteestä tuttu) *vapaan ja tasa-arvoisen kansalaisuuden* periaate. (Orjuuden ja dominaation vastustus, vapaa ja yhdenvertainen kansalaisuus, edellyttää riittävää taloudellista asemaa, jotta voi katsoa ketä tahansa silmiin ja sanoa poliittisen mielipiteensä.)

Nämä kaksi periaatetta implikoivat oikeuden riittävään osuuteen *työstä ja pääoman omistuksesta* sellaisissa yhteiskunnissa, joissa toimentulo on kytketty työhön ja pääoman omistukseen.

Kansalaisten yhdenvertaisuus ei kenties edellytä ”oikeutta työhön” (kenties kaikkien työpanosta ei tulevaisuudessa tarvita?) eikä ”oikeutta kaiken omaisuuden tasajakoon” (kenties tasajaosta

poikkeamat esimerkiksi omien valintojen seurauksena ovat eettisesti perusteltuja? kenties oikeudenmukaisuus vaatii vain riittävää, säädyllystä minimiosuutta), mutta kansalaisten yhdenvertaisuus edellyttäne ”oikeutta riittävään osuuteen työstä tai pääoman omistuksesta”: jos työn merkitys vähenee, pääoman jakamisen kaikille merkitys lisääntyy. On moraalisesti sietämätön ajatus, että syntyperän tai vanhempien kastiaseman takia jotkut saavat työtä/omaisuutta kun taas jotkut toiset eivät: mikä voisi perustella sen, että jollakulla ei olisi edes alun perin edes *mahdollisuutta* elää siedettävää elämää?

Mahdollisuuksien tasa-arvoa voi perustella John Rawlsin ajatuskokeella: millaiset yhteiskunnalliset periaatteet valittaisiin ”tietämättömyyden verhon” takana, jos ei tiedettäisi mihin yhteiskunnalliseen asemaan itse päätyy?

Automaation kehitys haastaa myös kulutuksen (talous vaatii, että joku ostaa koneiden tuottamat hyödykkeet) ja verotuksen (robottipersoonuuden ajatusta – joka siis on huono idea - on osin haluttu edistää, jotta niitäkin voitaisiin verottaa siinä missä ihmistyötäkin), joita molempia voi tarkastella oikeudenmukaisuuden näkökulmasta. Oikeiston ja vasemmiston näkemykset eroavat siinä, kuinka iso osa omasta menestyksestä on yksilön omalla vastuulla, mutta on selvää, että räikeät eriarvoisuudet eivät ole kenenkään mielestä oikeutettuja. Samoin rahan luonnetta (sallitaanko kryptovaluutat? entä aikapankit? Miten käsitetään keskuspankkien ja liikepankkien roolit) koskevat poliittiset kysymykset jakanevat jatkossa mielipiteitä.

Ekologinen kestävyys ja kansainvälinen oikeudenmukaisuus (mukaan lukien vauraampien valtioiden kansalaisten suurempi vastuu globaaleista kysymyksistä) ovat niin’ikään puoluepoliittisesti latautuneita, mutta myös moraalisesti relevantteja kysymyksiä. Niiden suhteen kysymys tuntuu olevan: halutaanko niin suurta muutosta vallitsevaan länsimaiseen elämäntapaan, kuin olisi moraalisesti oikeudenmukaista? Optimistit saattavat uskoa teknologiseen yltäkylläisyyteen, mutta toistaiseksi uskottavimmat ekologiset skenaariot (ml. ilmastonmuutos) ovat varsin hälyttäviä. Tämä on moraalinen ongelma esimerkiksi sukupolvien välisen oikeudenmukaisuuden, luonnon itseisarvoisuuden sekä ihmislajin säilymisen kannalta.