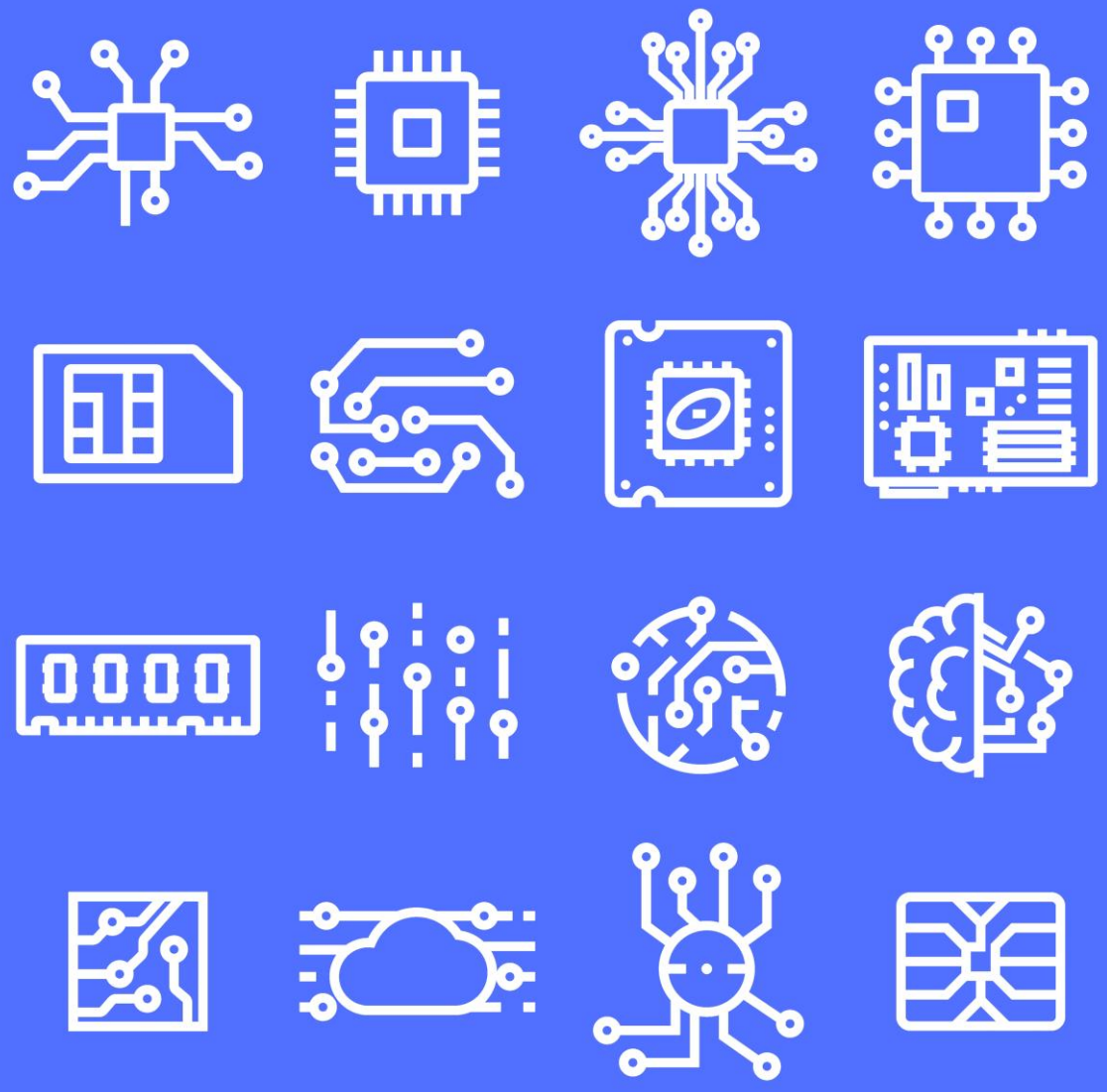
Turku AI Society

Opas tekoälyn etiikkaan

2019



Ojanen – Oljakka – Sahlgren – Tuikka – Vaiste

Esipuhe

Tekoälyn ja robotiikan kehityskulku on herättänyt laajaa julkista keskustelua etiikasta. Kansainvälisesti erilaisia suosituksia on ilmestynyt runsaasti, ja Suomessakin on tehty lukuisia selvityksiä ja ministeriöiden selontekoja. Tämä opas pyrkii toimimaan yleistajuisena johdantona näihin keskusteluihin. Tärkeimpiä lähteitä on mainittu oppaan lopussa.

Tekoälyn etiikan keskustelua voi lähestyä kysymällä retorisesti: mitkä moraaliset periaatteet eivät ole relevantteja tässä kontekstissa? Vastaus on tietystikin, että kaikki periaatteet ovat yhä relevantteja. Etiikkaa ei tarvitse keksiä uudelleen tyhjästä. Uusi teknologia voi kuitenkin vahvistaa ennestään olemassa olleita eettisiä ongelmia sekä synnyttää uusia. Jotta näitä voi ymmärtää, täytyy tuntea uuden teknologisen kontekstin erityispiirteitä. Vaikka tekoälyn etiikassa ollaan päätymässä kannattamaan samoja periaatteita kuin vaikkapa lääketieteen etiikassa (erityisesti autonomian ja yksityisyyden kunnioittamista, vahingoittamisen kieltoa, hyvinvoinnin turvaamista, oikeudenmukaisuuden vaatimusta), uusi teknologia voi synnyttää myös uudenlaisia kysymyksiä. Esimerkin tarjoaa syväoppimista hyödyntävien algoritmien niin kutsuttujen ”mustien laatikoiden” ongelma, jossa algoritmin toimintalogiikka ylittää käyttäjien käsityskyvyn, eikä algoritmi tarjoa inhimillisesti ymmärrettäviä perusteita suosituksilleen.

Tämän oppaan tarkoitus on tarjota yleiskatsaus näihin eettisiin haasteisiin, uusiin ja vanhoihin, sellaisina kuin ne ilmenevät autonomisen ja älykkään teknologian yhteydessä. Toivottavasti se osaltaan kannustaa eettiseen pohdintaan tekoälyn suunnittelun, käytön ja yhteiskunnallisen hyödyntämisen eri vaiheissa, ja osaltaan poistaa esteitä eettisiin ja yhteiskunnallisiin pulmakohtiin tarttumiselta.

Arto Laitinen
Filosofian professori, Tampereen yliopisto

Kirjoittajilta

Turku AI Society on tekoälyn yhteiskunnallisia ja taloudellisia vaikutuksia tutkiva poikkitieteellinen tutkimusyhteisö. Pyrkimyksenämme on edistää eettistä teknologiasuunnittelua, edesauttaa sidosryhmien keskinäistä yhteistyötä teknologian etiikan alalla sekä tarjota käytännön työkaluja, joiden avulla tekoälyn etiikan haasteisiin voidaan vastata. Opas tekoälyn etiikkaan palvelee näitä pyrkimyksiä luomalla yleisen ja ymmärrettävän katsauksen tekoälyn etiikan kenttään, sen keskeisiin eettisiin ongelmiin ja haasteisiin. Tämän lisäksi esitämme yleisen tason toimenpidesuosituksia näihin haasteisiin vastaamiseksi. Ehdotettujen toimenpiteiden painopiste on pitkäjänteisissä investoinneissa laaja-alaiseen tutkimukseen, lainsäädännön ja sääntelyn tarpeellisuudessa sekä kansainvälisessä yhteistyössä, kuluttajien asemassa ja oikeudessa osallistua keskusteluun teknologisen kehityksen suuntaviivoista.

Tahdomme kiittää Arto Laitista (Tampereen yliopisto, filosofian professori) sekä Kai Kimppaa (Turun yliopisto, dosentti, yliopistotutkija) arvokkaista kommenteista ja ohjauksesta. Kiitämme Suomen Akatemian strategisen tutkimuksen konsortiota Robotit ja hyvinvointipalvelujen tulevaisuus (ROSE) sekä Turun yliopiston Future Ethics -tutkimusryhmää antoisasta yhteistyöstä. Lisäksi haluamme kiittää Turku Science Park Oy:tä oppaan painatuksen sponsoroinnista.

Kirjoittajat

Atte Ojanen (VTK, Turun yliopisto) on käytännöllisen filosofian maisteriopiskelija, kiinnostuksenaan teknologisen kehityksen luomat uudet eettiset haasteet.

Nea Oljakka (OTM, Turun yliopisto) on oikeustieteen tohtorikoulutettava ja valmistelee tällä hetkellä väitöskirjaa uusien teknologioiden mukanaan tuomista rikosvastuun toteutumisen haasteista. Oljakka on erityisen kiinnostunut toimijuuden muutoksen vaikutuksista vastuun kohdentamiseen lähitulevaisuudessa, kun tekoälysovellukset kehittyvät yhä autonomisemmiksi.

Otto Sahlgren (HuK, Tampereen yliopisto) on tekoälyn ja robotiikan filosofisiin ja eettisiin kysymyksiin erikoistunut humanististen tieteiden kandidaatti.

Anne-Marie Tuikka (FM, Turun yliopisto), tekee väitöskirjatutkimusta sosiaalipalveluiden digitalisoitumisesta kansalaisten näkökulmasta. Hän osallistuu Turun yliopiston Future Ethics –tutkimusryhmän toimintaan ja on jäsenenä IFIP Joint Working Group 9.2:ssa.

Juho Vaiste (KTM, Turun yliopisto) tekoälyn etiikan väitöskirjatutkija ja asiantuntija. Erikoistumisalueina teknologiakehittäjien eettinen toimijuus sekä tekoälyn etiikka liiketoimintavastuun näkökulmasta.

Sisällys

Johdanto

Mitä on tekoäly?

Tekoäly ja älykkäät järjestelmät ovat olleet kiistatta yksi viimeisen vuosikymmenen eniten keskustelua herättäneistä aiheista. Taloudellisen ja poliittisen keskustelun ytimeen ovat nousseet kysymykset esimerkiksi tulevaisuuden työn luonteesta, datan käytöstä, yksityisyydestä sekä tulonjaosta. Nämä kysymykset ovat erottamattomasti yhteydessä tekoälyteknologioiden nykytilanteeseen ja kehitykseen. Mutta mitä itse asiassa on tekoäly?

Tekoälyä on pyritty määrittelemään usealla eri tavalla useista eri näkökulmista lähtien. Kognitiotieteen ja filosofian piirissä yksi keskeinen jaottelu on ollut heikon ja vahvan tekoälyn välillä: jaottelu on ollut heikon ja vahvan tekoälyn välillä: edellinen viittaa ohjelmaan tai malliin, joka pyrkii simuloimaan kognitiivisia prosesseja tai yksinkertaisesti vaikuttamaan älykkäältä, kun taas jälkimmäinen termi viittaa aidosti älykkääseen, tietoiseen ohjelmaan tai esineeseen.[1] Nykykeskustelussa tekoäly samaistetaan kuitenkin usein koneoppiviin algoritmeihin (ja niiden pohjalta toimiviin koneisiin), jotka kykenevät mukautumaan ja oppimaan, muuttamaan omia parametrejään ja toimintaperiaatteitaan uuden tiedon pohjalta.[2]

Tässä oppaassa viittaamme termillä “tekoäly” yhtäältä joukkoon moderneja teknologioita, joka käsittää esimerkiksi erilaiset oppivat algoritmit ja erilaiset runsasta datamäärää hyödyntävät sovellukset. Tarkastelumme ei kuitenkaan rajoitu ainoastaan koneoppimiseen vaan käsittelemme myös esimerkiksi autonomisia kulkuneuvoja, esineiden internetiä (internet of things, IoT), robotiikkaa ja ubiikkia teknologiaa (tai nk. ambienttia älykkyyttä). [3]

Toisaalta tekoälystä puhuttaessa on mielekästä tarkastella sitä myös yhteiskunnallisena ilmiönä, jolla on potentiaalia muuttaa yksityiselämäämme, ympäristöämme ja laajemmin yhteiskuntaamme perustavanlaatuisella tavalla. Edellä mainitut teknologiat ohjaavat käyttäytymistämme hienovaraisin tavoin jo tälläkin hetkellä, ja näiden teknologioiden käyttöönoton vaikutukset voivat näkyä radikaalistikin yhteiskunnallisella tasolla. Tekoälyteknologioiden kehittyminen tuo mukanaan kysymyksiä työstä, taloudesta, (poliittisesta) päätöksenteosta, yksityisyydestä ja turvallisuudesta.

Tämän määritelmän pohjalta huomaamme, että tekoälyyn liittyvä eettinen problematiikka voi koskettaa meitä monella tasolla. Seuraavassa taulukossa esittelemme keskeistä tekoälyyn ja kehittyneeseen teknologiaan liittyvää sanastoa. Myöhemmin oppaassa käsittelemme näihin teknologioihin liittyviä eettisiä ongelmia.

[1] Searle 1980.

[2] Vaikka tekoälystä puhutaankin usein nykykeskustelussa synonyymisesti koneoppimisen kanssa, ei kaikki tekoälyn käsitteen piiriin luettavissa ole teknologia suinkaan ole koneoppimisalgoritmeihin pohjautuvaa. Perinteinen koodaaminen elää ja voi hyvin yhä vieläkin, osittain varmastikin siksi, että laskentatehon kasvaminen on mahdollistanut älykkäiden sovellusten luomisen ilman oppimismekanismejakin.

[3] ITS Finlandin sivuilta löytyvä sanasto: http://www.its-finland.fi/index.php/fi/mita-on-its/its-sanasto.html.

Taulukko 1. Tekoälyteknologioiden sanastoa

Algoritmi **(engl. algorithm)**	Tarkka ja yksiselitteinen kuvaus ongelman (tai ongelmien luokan) ratkaisemisesta; eräänlainen lista askelista, jotka täytyy ottaa ongelman tai tehtävän ratkaisemiseksi. Algoritmeja käytetään esimerkiksi haku-, valinta-, optimointi- ja lajittelutehtävien suorittamiseen.
Tekoälyagentti	Tekoälyä hyödyntävä järjestelmä, joka kykenee suhteellisen itsenäiseen toimintaan ja suorittaa (laskennallisia) toimintoja, kuten päätöksentekoa, suunnittelua sekä optimointia, itsenäisesti ilman ihmisen valvontaa.
Koneoppiminen (machine learning)	Koneoppimisalgoritmit luovat malleja valitun datajoukon pohjalta. Näitä malleja voidaan käyttää analyysissa, joka voi paljastaa (ilmeisten yhteyksien lisäksi) uudenlaisia piirteitä ja suhteita datasta, kuten yllättäviä korrelaatioita erilaisten muuttujien välillä.
Ohjattu (kone)oppiminen (supervised learning)	Koneoppimisen muoto, jossa mallin oppimisen prosessia valvotaan esimerkiksi tarjoamalla tiettyä datasyötettä (input) vastaava maalivaste (target output). Tällöin koneen tehtäväksi jää muodostaa malli yhteyksistä syötteiden ja vasteiden välillä. Käytetään erityisesti luomaan ennusteita datan pohjalta.
Ohjaamaton (kone)oppiminen (unsupervised learning)	Koneoppimisen muoto, jossa mallin oppimisen prosessia ei valvota. Malli muodostetaan löytämällä (osa)joukko datainstanssien välisiä yhteyksiä. Käytetään erityisesti löytämään yllättäviä ja informatiivisia yhteyksiä datasta.
Vahvistusoppiminen (reinforcement learning)	Koneoppimistekniikka, jossa oppimista ohjaavat positiiviset ja negatiiviset palautteet. Vahvistusoppiminen on hyödyllinen tekniikka erityisesti tilanteissa, joihin liittyy pitkän tähtäimen päätöksentekoa ja laaja mahdollisten tilojen kirjo.
Yleistekoäly (artificial general intelligence, AGI)	Tekoäly, joka kykenee yleiseen älykkääseen toimintaan, suorittamaan yhtä paljon tehtäviä yhtä tehokkaasti tai tehokkaammin kuin ihminen (vrt. vahva tekoäly).
Supertekoäly (superintelligence)	(Hypoteettinen) tekoälyn muoto, joka ylittää inhimillisen älykkyyden.
Esineiden internet (internet of things, IoT)	Esineiden välinen verkko, joka mahdollistaa koneiden keskinäisen kommunikoimisen. Verkottuneet laitteet voivat mitata ja monitoroida toimintaansa esim. sensoreiden avulla, tuottaen samalla dataa, joka mahdollistaa prosessien reaaliaikaisen valvomisen. Tällöin on mahdollista optimoida tuotantoprosesseja sekä ennustaa mahdollisia poikkeamia (esim. teollisten laitteiden toimintakykyä).
Ubiikit teknologiat/ambientti älykkyys (ubiqutious or pervasive technologies, ambient intelligence)	Teknologiat tai älykkäät järjestelmät, jotka sulautuvat erilaisiin ympäristöihin, toimien saumattomasti ja huomaamattomasti. Tällaiset teknologiat saattavat myös kommunikoida keskenään, kerätä dataa ympäristöstä ja muokata sen ominaisuuksia tämän datan pohjalta (esim. Lämpötilaa, valaistusta, musiikkia tai ilman kosteutta muuttamalla).

Mitä on tekoälyn etiikka – mitä se on ja miksi sitä tarvitaan?

Yksittäiset tekoälyteknologiat vaikuttavat sekä yksilöihin käyttäjäkokemustasolla, yksilöllisessä vuorovaikutuksessa teknologian kanssa, että koko yhteiskuntaan. Tekoälyn laajamittainen käyttöönotto voi mahdollisesti muuttaa yhteiskunnallisia rakenteita, käytänteitä ja toiminnan tapoja merkittävästi. Näistä syistä tarvitsemme tekoälyn etiikkaa.

Etiikka tai moraalifilosofia on filosofian osa-alue, jossa tutkitaan kysymyksiä hyvästä ja pahasta, oikeasta ja väärästä, oikeudenmukaisuudesta ja epäoikeudenmukaisuudesta. Etiikka voidaan jakaa edelleen osiin sen tutkimuskohteiden ja -menetelmien perusteella.[4] Näistä osa-alueista on tässä yhteydessä tarpeen mainita vain kaksi: metaetiikka ja normatiivinen etiikka. Normatiivisen etiikan teoriat pyrkivät esittämään tarkasti jäsenneltyjä ja johdonmukaisia näkemyksiä siitä, mitä ja minkälaista toimintaa voidaan pitää oikeudenmukaisena, hyvänä, pahana, oikeana tai vääränä. Kun tällaisia näkemyksiä sovelletaan tietyllä osa-alueella, kuten lääketieteen tai yritystoiminnan eettisiä kysymyksiä tarkasteltaessa, voidaan puhua myös soveltavasta etiikasta, joka sijoittuu normatiivisen etiikan alle. Metaetiikka puolestaan keskittyy tarkastelemaan, minkälaisiin perustavampiin näkemyksiin erilaiset normatiivisen etiikan teoriat sitoutuvat ja miten etiikassa operoitavat keskeiset käsitteet, kuten hyvä ja paha, tulisi ymmärtää.[5]

Tekoälyn etiikkaa voidaan pitää soveltavana osana normatiivista etiikkaa, joka keskittyy erityisesti tekoälyteknologioihin liittyvään problematiikkaan sekä suunnitteluteknisellä että yhteiskunnallisella, rakenteellisella tasolla. Tässä vaiheessa on hyvä huomauttaa, että metaeettiset ja normatiiviset käsitykset ovat usein hyvin yhteen kietoutuneita.

Yhtäältä, kun pyrimme antamaan vastauksia eettisiin ongelmiin käytännön tasolla, tulemme aina tehneeksi oletuksia perustavammalla tasolla - oletuksia siitä, mikä on yleisesti hyvää tai tavoittelemisen arvoista (esimerkiksi erilaiset arvot kuten yksityisyys, turvallisuus tai tehokkuus). Toisaalta tämä tarkoittaa myös, että ratkaisumme käytännön tasolla voivat potentiaalisesti tarjota meille välineitä ratkaista myös teoreettisen tason ongelmia. Tämän huomautuksen tarkoituksena on tuoda ilmi yhtäältä tekoälyn eettis-teoreettisen tarkastelun tärkeys ja välttämättömyys, ja toisaalta se, että toimivat ratkaisut tekoälyn eettisiin ongelmiin saattavat tarjota meille mahdollisuuden oppia jotakin etiikasta yleisesti.

Tekoälyn etiikka sijoittuu myös laajemmin teknologian etiikan perinteeseen, jossa on tarkasteltu etiikan keinoin teknologiaa, joka ei ole älykästä tai oppivaa. Teknologian filosofiassa on historiallisesti tarkasteltu ihmisen ja teknologian välisiä suhteita ja näiden suhteiden seurauksia ja vaikutuksia inhimillisen elämän eri tasoilla. Tästä syystä tekoälyn etiikasta kiinnostuneet voivat löytää hedelmällistä tarttumapintaa myös teknologian filosofiasta ja
etiikasta yleisemmin.

[4] Ks. esim. Fieser, James. “Ethics”, Internet of Encyclopedia of Philosophy. URL: https://www.iep.utm.edu/ethics/. Viitattu 13.10.2018.

[5] Sayre-McCord (2014)

Tekoälyn etiikkaa käsittelevässä tutkimuskirjallisuudessa on laajasti otettu kantaa vastuullisen ja eettisen tekoälykehityksen puolesta. Tekoäly tuo meille uudenlaisia uhkia, haasteita ja eettisiä kysymyksiä, joiden selvittämiseen ja keskustelemiseen on varattava aikaa ja resursseja.[6] Tekoälyn etiikan työtä on tehtävä monella eri tasolla: yliopistoissa ja tutkimusinstituutioissa, kansalaiskeskustelun foorumeilla, kehittäjäyhteisöissä sekä yrityksissä. Kansainvälisesti olemme vielä vaiheessa, jossa keskiössä ovat tietoisuuden ja tutkimuksen lisääminen. Monet yritykset kuitenkin ottavat jo ensiaskeliaan tekoälyn etiikkaan: tekoälyä hyödynnetään jo runsaasti, ja teknologia-alan yksittäiset toimijat osaltaan määrittelevät tekoälyn kehityksen ja soveltamisen eettisiä periaatteita. Julkisessa ja kansalaiskeskustelussa sekä mediassa puolestaan suuren painoarvon ovat saaneet - osittain varmastikin vetovoimaisuutensa takia - erilaiset dystopia-skenaariot. Näistä etäisistä skenaarioista ja spekulaatiosta tulisi kuitenkin siirtyä aktiiviseen keskusteluun, johon ideaalitilanteessa osallistuisivat niin tutkijayhteisö ja asiantuntijat kuin poliittiset päättäjätkin, kansalaisia unohtamatta.

Tässä oppaassa tarkastelemme tekoälyn eettisiä ongelmia ja tarjoamme suuntaviivoja sille, kuinka näitä ongelmia voidaan hedelmällisesti lähestyä. Näitä erityisiä ongelmia tarkastellaan myöhemmissä luvuissa, mutta aluksi on syytä nostaa esille muutama huomio näiden kysymysten erityisluonteesta. Moni tekoälyetiikan kysymys kietoutuu yhteen teknologia- ja erityisesti informaatioteknologian eettisen problematiikan kanssa laajemmin. Ongelmat voivat liittyä jo olemassa olevien teknologioiden käyttöön, mutta tekoälyn ja yleisen teknologisen kehityksen myötä voi syntyä myös uusia eettisiä haasteita. Kahdenlaiset prosessit voivat täten synnyttää eettisiä haasteita:

1) Aikaisemmat eettiset ongelmat kumuloituvat ja saavat suuremmat mittasuhteet, kun tekoälyteknologiaa implementoidaan laajemmalla skaalalla ilman tarvittavaa ymmärrystä kyseisten teknologioiden mahdollisista epätoivottavista ominaisuuksista ja vaikutuksista.

2) Tekoälyn teknologinen erityisluonne (autonomisuus, mukautuvuus, näkymättömyys) antaa olemassa oleville eettisille kysymyksille uudenlaisia piirteitä, muuttujia, ulottuvuuksia tai vaikutusalueita.

Tekoälyteknologian suunnittelun ja käyttöönoton eettistä ulottuvuutta arvioitaessa tulee ottaa huomioon nämä kaksi prosessia. Tämän arvioinnin onnistuminen edellyttää riittävää ymmärrystä teknologioille ominaisista ongelmista (sekä yksittäisten teknologioiden osalta että yleisemmällä tasolla), mutta myös vallitsevasta kulttuurista ja yhteiskunnallisesta kontekstista, joissa lainsäädäntö ja arvot määrittävät toimintamme mahdollisia suuntaviivoja.

Yhtä lailla meidän tulee olla tietoisia siitä, että teknologia voi tuottaa yllättäviäkin seurauksia, jotka voivat olla eettisesti kestämättömiä ja hankalia ennakoida. Näitä vaikutuksia tulee pyrkiä kartoittamaan, ja yhteistyö tutkimusyhteisön sekä muiden asiantuntijoiden kanssa voi vähentää riskiä niiden ilmenemiselle.

[6] Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. AI Magazine, 36(4): 105-114.

Eettinen tekoäly tulevaisuuden vastuullisen toiminnan peruspilarina

Jos tekoälyn etiikka on niin tärkeää kuin olemme esittäneet, miksi siihen herätään toden teolla vasta nyt? Vastaus tähän kysymykseen piilee osittain suhteellisen tuoreessa mahdollisuudessa hyödyntää dataa koneoppimisen ja datatieteen keinoin. On esimerkiksi arvioitu, että vuoteen 2003 mennessä noin 5 eksatavua (n. 1015 gigatavua) dataa oli tuotettu ihmisten toimesta. Vuodesta 2013 eteenpäin ihmiskunta on tuottanut ja varastoinut vastaavan määrän dataa päivittäin.[7] Nykyinen koneoppimiseen painottuva tekoälyn aalto, joka on mahdollistanut tämän datavirran hyödyntämisen, on siis vielä suhteellisen nuori. 2010-luvun alussa otettiin tärkeitä askelia, mutta uusien teknologioiden laajempi käyttöönotto pienemmissä ja keskikokoisissa yrityksissä on ollut käynnissä vasta muutaman vuoden. Sama pätee julkiseen sektoriin, jossa datatiedettä ja koneoppimista on alettu viime vuosina hyödyntää esimerkiksi älykkään infrastruktuurin ja älykkäiden ympäristöjen luomisessa.

Vaikka tekoälyn aaltoliike ei ole ehtinyt tuottaa vielä julkisessa keskustelussa ajoittain maalailtavia laaja-alaisia ja radikaaleja muutoksia, on sitä jo sovellettu laiminlyöden eettisiä standardeja. Keskeisiä tapauksia ovat esimerkiksi profilointi- ja diskriminaatiotapaukset. Esimerkiksi Googlen online-mainontajärjestelmän on todettu näyttävän korkeapalkkaiseen työhön liittyvää mainosta herkemmin henkilölle, joka identifioitui mieheksi Google-profiilissaan.[8] Tällaisia tapauksia ei ole todettu vain yksityisten sektorin toimijoiden piirissä. Yhdysvalloissa ja Yhdistyneessä kuningaskunnassa käytettyä datatyökalu PredPolia, joka ennustaa rikollisen toiminnan ilmenemistä paikallisesti ja ajallisesti, on kritisoitu sen mahdollisista vinoumista, jotka johtavat syrjintään esimerkiksi etnisyyden tai sosio-ekonomisen aseman perusteella.[9] Tulevissa luvuissa tarkastelemme tällaisia sekä lukuisia muita tekoälyteknologioihin liittyviä eettisiä ongelmia yksityiskohtaisemmin.

Tekoälyn riskien arvioiminen, ongelmakohtien osoittaminen ja eettisten suunnitteluperiaatteiden muotoileminen tulee ottaa vakavasti, mutta tämän ei tarvitse johtaa yleiseen teknologiapessimismiin. Tekoälyn eettinen kehitystyö voi parhaimmillaan olla innovatiivista toimintaa, joka luo sekä taloudellista hyötyä että hyvinvointia yhteiskunnallisella tasolla. Erilaiset sidosryhmät huomioon ottava, kauaskantoinen ja kestävä eettinen design voidaan nähdä rajoitteen sijasta mahdollisuutena luoda skaalautuvia eettisiä teknologioita, jotka palvelevat kaikkien osapuolien intressejä.

Huolenaiheet, kuten yksityisyys ja turvallisuus, ovat nousseet keskeisiksi julkisessa tekoälykeskustelussa, joten on syytä olettaa, että tulevaisuuden tekoälyteknologioiden vetovoimaisuus kuluttajamarkkinoilla on vahvasti riippuvainen näiden asioiden huomioimisesta suunnittelu- ja kehitystyössä. Samoin yleinen tietoisuus datan keräämiseen, omistamiseen ja käyttöön liittyvistä mahdollisista eettisistä ongelmista on lisääntynyt.[10] Tämän voidaan olettaa vaikuttavan kuluttajien odotuksiin siitä, minkälaisia datan omistamis- ja hallinnointiratkaisuja yritykset ja muut palveluntarjoajat tekevät. Eettisesti sensitiivinen

[7] Kelleher, J. D., & Tierney, B. (2018, 9). Data Science. The MIT Press.

[8] Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. Proceedings on Privacy Enhancing Technologies, 2015(1), 92-112.

[9] Baldridge, J. (2015). Machine Learning and Human Bias: An Uneasy Pair. TechCrunch, 2. elokuuta. URL: http://social.techcrunch.com/2015/08/02/machine-learning-and-human-bias-an-uneasy-pair.

[10] Esimerkiksi Suomessakin vaikuttava MyData (https://mydata.org/) painottaa yksilöiden oikeutta oman datansa hallitsemiseen.

lähestymistapa tekoälyteknologioiden suunnittelu- ja kehitystyöhön sekä implementointiin ja käyttöön voidaan nähdä mahdollisuutena tehdä hyvää - tarkoitusperä, joka ei itsessään vaatine perusteluja - mutta myös mahdollisuutena profiloitua eettisesti vastuullisena toimijana aikana, jolloin teknologia tuottaa epävarmuutta sekä yksilöllisellä että yhteiskunnallisella tasolla.

Oppaassa luomme katsauksen tekoälyn etiikan kysymyksiin ja tarjoamme tilannekuvan akateemisesta kentästä, yritysmaailmasta ja yhteiskunnasta yleisesti. Opasta päivitetään tasaisin väliajoin vastaamaan tekoälyteknologioiden kehityksen suuntaa ja ajankohtaista tilannetta. Olemme edelleen vaiheessa, jossa tietoisuuden lisääminen tekoälyn etiikkaa kohtaan on tärkeää. Oppaan päätarkoituksena on tarkastella ongelmia, jotka vaativat ensisijaista huomiotamme tällä hetkellä, toimia tekoälyn etiikkaa käsittelevän keskustelun avaajana ja tarjota sidosryhmille yleisiä suuntaviivoja kohti eettistä teknologiasuunnittelua. Toivomme, että opas kannustaa eri sidosryhmiä avoimeen keskusteluun tekoälyn eettisestä kehityksestä ja käytöstä.

Eettiset kysymykset ja haasteet

Tämä opas keskittyy yleisimmin esillä olleisiin tekoälyetiikan riskeihin ja kysymyksenasetteluihin, jotka toistuvat tieteellisessä kirjallisuudessa, julkisessa keskustelussa ja mediassa. Tarkastelemme myöhemmin myös eri näkökulmia ja lähtökohtia, jotka ohjaavat kysymystenasettelujen arviointia ja eettistä analyysia. Tekoälyn etiikassa eri teemat limittyvät ja täydentävät toisiaan, joten oppaan kategorisointia ei tule arvioida kiveen hakattuna järjestyksenä, vaan mahdollisimman selkeänä yleisesityksenä.

Tällä hetkellä konkreettiset ja ajankohtaiset riskit ovat pääsääntöisesti tunnettuja aikaisemmasta informaatioteknologian etiikan perinteestä, mutta tekoälyteknologiaa vahvasti määrittelevä autonomisuus luo niihin uusia tasoja. Polttaviksi eettisiksi kysymyksiksi ovat nousseet esimerkiksi koneoppimismalleihin käytettävän dataaineiston vinoumat tai vääristymät, jotka voivat päätyä syrjiviin lopputuloksiin, tekoälyteknologian luomiin uudenlaisiin turvallisuusuhkiin liittyvät kysymykset sekä yksityisyyteen ja datan keskittymiseen liittyvät kysymykset.

Keskustelemme oppaassa myös kahdesta pidemmän aikavälin yhteiskunnallisesta muutoksesta ja haasteesta: työn muutoksesta ja murroksesta sekä teknologian vaikutuksesta siihen, miten ymmärrämme ihmisyyden ja inhimillisen elämän. Käsittelemme lyhyesti myös kehittyneisiin tekoälyteknologioihin liittyviä teemoja. On epävarmaa, milloin esimerkiksi skenaario ihmistasoisesta tekoälystä voi toteutua, mutta akateemisessa kirjallisuudessa teoreettisempiakin kysymyksiä on käsitelty runsaasti. Keskeisiä aiheita ovat myös muun muassa supertekoälyn uhkakuvat, ihmisen parantelu tekoälyn avulla, sekä robottien ja tekoälytoimijoiden oikeudet.

Datan keskittyminen ja yksityisyys

Tekoälyn myötä yksityisyydensuojasta käytävän keskustelun keskeinen asetelma on muuttunut. Tämä on seurausta kahdesta asiasta: tekoälyn seurauksena yksityistietojamme on yhä helpompi kerätä ja hyödyntää laajassa mittakaavassa, ja alustatalouskehityksen myötä tämä tieto keskittyy yksittäisten toimijoiden haltuun. Teknologisen kehityksen ansiosta lähes kaikista ihmiselämän osa-alueista voidaan kerätä tietoja, mikä mahdollistaa entistä tarkemman yksilöiden toiminnan seuraamisen ja ennustamisen. Yksityisyyden kannalta ongelmia voi aiheuttaa niin tiedon kerääminen, käsittely kuin jakaminenkin.

Ihmisellä on perustava oikeus omaan yksityisyyteensä, oli kyse sitten käytöksestämme, sijainnistamme tai tunteistamme. Oikeus yksityisyyteen on Suomessa perustuslain turvaama, jokaiselle kuuluva oikeus. Siitä on säädetty myös kansainvälisissä sopimuksissa, esimerkiksi YK:n ihmisoikeuksien yleismaailmallisen julistuksen 12 artiklassa. Yksityisyys voidaan nähdä perustana niin sananvapaudelle, yhdistymisvapaudelle kuin autonomiselle toiminnallekin. Tekoälysovellukset tulee suunnitella siten, että ne eivät loukkaa ihmisten oikeutta yksityisyyteen. Tekoälyn kohdalla yksityisyydensuojan toteutuminen edellyttää yksilön oikeutta muun muassa tarkistaa ja hallita henkilötietojensa käyttöä.

Tekoälysovellusten kehittäminen edellyttää valtavia datamääriä. Henkilötietojen kerääminen, prosessointi ja hyödyntäminen mahdollistavat myös perinteisesti vahvasti suojellun yksityisyyden piiriin tunkeutumisen ja näiden tietojen hyödyntämisen yksityisyydensuojan vaarantavalla tavalla. Esimerkiksi mahdollisuudet valvontaan, profilointiin ja käytöksen manipulointiin ovat lisääntyneet huomattavasti. Riskejä luovat tekoälysovellukset sosiaalisen median algoritmeista kehittyneempiin teknologioihin kuten kasvojen-tunnistukseen. Kansalaiselle ja kuluttajalle voi tuottaa hankaluuksia hahmottaa, kuinka systemaattisesti erilaiset älylaitteemme keräävät ja jakavat tietoa käyttäytymisestämme, usein ilman selkeää suostumustamme. Tekoälyn mahdollistama automaattinen profilointi ja päätöksenteko herättävätkin kysymyksiä datan hyödyntämisen perusteista, päätösten taustalla piilevästä logiikasta, sekä näiden oikeutettavuudesta ja mahdollisista syrjinnän riskeistä. Tämän lisäksi lisääntyvällä datankeruulla on huomattavia negatiivisia vaikutuksia yhteiskunnalliseen luottamukseen ja avoimuuteen.

Tieto on valtaa, data on tietoa. Datan keskittyminen yksittäisille yrityksille, kuten Googlelle ja Facebookille, antaa näille toimijoille ennennäkemättömän pääsyn ja vaikutusvallan ihmisten yksityiselämään. Tästä yksilödatasta saattavat olla tunnistettavissa myös ystävät ja perheenjäsenet, vaikka he eivät olisivatkaan antaneet suostumustaan näille palvelujentarjoajille. Tämä kehitys ei siis ainoastaan uhkaa omaa yksityisyyttämme, vaan lisäksi vahvistaa massavalvonnan mahdollisuutta.

Käyttäjillä on harvoin tietoa siitä, mihin kaikkeen heidän tietojaan käytetään, vaikka yrityksillä, jotka henkilötietoja käsittelevät, on velvollisuus tarjota riittävästi informaatiota. Tekoälyn mahdollistama valvonta ja yksityisyydensuojan rikkomukset tarkoittavat, että sitä ei voida hyödyntää ilman kuluttajien selkeää ja täsmällistä suostumusta.[11] Tämä luo yrityksille velvollisuuden kertoa, minkälaista tietoa tekoäly kerää, kenellä on pääsy siihen, mihin tarkoitukseen sitä käytetään, kuinka pitkään sitä säilytetään ja mihin sitä mahdollisesti luovutetaan. Kun tekoäly hyödyntää arkaluonteista tietoa, tulee tämä tieto anonymisoida ja tiedon tunnistamattomuus varmistaa. Käyttäjien tulee pystyä luottamaan, että heihin yhdistettävää tieto minimoidaan ja suojellaan asianmukaisesti. Mikäli tekoälysovellukset eivät huolehdi käyttäjiensä yksityisyydensuojasta, voi tällä olla tuhoisat vaikutukset yksilöiden omantunnon-, sanan- ja ilmaisunvapauden toteutumiselle.[12]

Yksityisyydensuojan haasteet vaativat entistä parempaa valistusta yksityistietojen ongelmista sekä uusia ratkaisuja datan luotettavaan hallinnointiin EU:n tietosuoja-asetuksen, GDPR:n[13] mukaisesti. Tietosuoja-asetus on korostanut yk-

[11] AI Now 2018, 4.

[12] AccessNow 2018, 29.

[13] Säädösteksti: https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex%3A32016R0679

silöiden itsemääräämisoikeuden kunnioittamista datan käytössä – ihmisillä tulee olla perimmäinen valta päättää mihin tarkoituksiin ja millä tavoin heidän tietojaan hyödynnetään. Lainsäädäntö dataturvallisuuden varmistamiseksi sekä datan perustelemattoman keräämisen rajoittamiseksi on myönteistä kehitystä. Luotettava tekoäly vaatii eettisesti, käyttäjien suostumuksella kerättyä dataa – keskinäinen luottamus on sekä käyttäjien että yritysten eduksi.

Turvallisuus

Tekoälyyn liittyy uudenlaisia, tekoälyteknologian autonomiseen toimintaan liittyviä kysymyksiä ja riskejä. Vaikka turvallisuus on - tai ainakin sen pitäisi olla - vahvasti huomioituna suunnittelu- ja ohjelmointiprosessissa, tekoäly on lähtökohtaisesti luonteeltaan itseoppivaa ja siten myös vaikeasti tai jopa mahdottomasti ennakoitavaa. Ennakoinnin vaikeus lisää tarvetta miettiä tekoälyn toimintaa ja sen seurauksia myös eettiseltä kannalta. Usein keskustelussa nostetaan esille kysymys siitä, missä määrin ihmisen on oltava varmistamassa, että tekoälyn tekemät päätökset tai toimet ovat oikeudenmukaisia, perusteltuja ja hyväksyttäviä (ns. "human-in-the-loop" problematiikka).

Autonomisiin kulkuneuvoihin liittyviä dilemmoja on esitetty mielenkiintoisina eettisinä ajatusharjoituksina, mutta vasta viimeaikaiset onnettomuustapaukset ovat tuoneet konkretiaa riskeihin ja eettisiin kysymyksenasetteluihin. Kuinka luotettavia autonomisten autojen on oltava niiden käyttöönottamiseksi? Onko yksittäisten yritysten oikeutettua testata itseohjautuvia autoja liikenteessä, ilman ihmisten suostumusta?

Pahansuopien toimijoiden, kuten rikollisten käyttöön tekoäly antaa valitettavasti uusia mahdollisuuksia. Haavoittuvuus kavalluksille, hakkeroinnille ja identiteettivarkaudelle kasvaa. Näiden uhkien ydinkysymyksenä on nyt hajauttaminen, joka tuo mukanaan uusia huolenaiheita. Esimerkiksi huume- ja asekauppa sekä muu laiton toiminta on pysynyt elinvoimaisena Torverkon kaltaisissa hajautetuissa ja salatuissa palveluissa. Bitcoin sekä muut kryptovaluutat ovat tarjonneet mahdollisuuksia muun muassa pimeän rahan siirtämiseen.

Eettisiä ongelmia liittyy myös keinoihin, joilla näitä ilmiöitä pyritään estämään. Suora tekniikan regulointi voi vaikuttaa houkuttelevalta, mutta nopeasti kehittyvällä alalla tällainen sääntely on auttamatta ajastaan jäljessä. Erilaisten teknologioiden rajoittamiseen tähtäävä lainsäädäntö voi kyllä jossain määrin vaikeuttaa laitonta toimintaa, mutta samalla se kaventaa ihmisten oikeuksia ja toimintamahdollisuuksia. Yleisemmän tason sääntely, jossa pyritään vaikuttamaan teknologioiden suunnittelu- ja tuotantoprosessiin ja luomaan teknologioiden vaikutuksille oikeudenmukaiset rajat lienee lopulta kestävämpi ratkaisu. Tavoitteena olisi varmistaa, että jo suunnitteluvaiheessa kartoitettaisiin riittävällä tarkkuudella riskit ja mahdollisuudet sekä pyrittäisiin varmistamaan tekoälysovellusten turvallinen ja oikeudenmukainen toiminta. Yksi keskeinen seikka on, että yritysten ja tekoälyn kehittäjien suunnittelemat tekoälysovellukset rakennetaan mahdollisimman vastustuskykyisiksi ulkopuolisia hyökkäyksiä vastaan.[14]

Hajautetut, pääosin internet-verkkoon tukeutuvat ja tekoälyteknologiaa hyödyntävät rikolliset nostavat esiin myös kysymyksen teknologisesta tasa-arvosta. Jo tähän asti vähemmän kokeneet käyttäjät, kuten vanhukset, ovat olleet haavoittuvaisempia esimerkiksi sähköpostihuijauksille.[15]

[14] Tekoälyn sääntelystä ks. esim. Turner 2019.

[15] Esim. Cook, D. M., Szewczyk, P., Sansurooah, K. (2011). Securing the Elderly: A Developmental Approach to Hyper-

Tekoälyteknologioiden mukanaan tuomat profilointi, psykologinen vaikuttaminen ja kehittyneet kommunikaatiomahdollisuudet voivat suurentaa tätä teknologisen osaamisen aiheuttamaa kuilua entisestään.

Hajautetut uhat koskevat tietenkin myös valtiollista maanpuolustusta ja sodankäyntiä. Hajautetut ja autonomiset taistelurobotit, informaatiosodankäynti, palvelunestohyökkäykset sekä muut tekoälyn mahdollistamat sodankäynnin muodot muuttavat kansainvälisten kriisien kenttää. Lisääntyvä informaatiosodankäynti muuttaa myös kuvaamme yhteiskunnan kokonaisturvallisuudesta. Eettisesti huomionarvoista on tarkastella myös, kuinka perustavat arvomme, kuten turvallisuus ja yksityisyys voivat ajoittain olla myös ristiriidassa keskenään.

Algoritmisen päätöksenteon ongelmat

Algoritmit ohjaavat päätöksentekoa erilaisissa tilanteissa. Niiden on ensinnäkin oletettu analysoivan erilaisia tekijöitä objektiivisemmin ja kattavammin kuin ihmiset, johtaen näin parempiin päätöksiin. Algoritmien toisena etuna on niiden tehokkuus – laaja datamäärä voidaan käydä läpi huomattavasti ihmistä nopeammin. Algoritmeja voidaan käyttää myös itsenäisinä päätöksentekijöinä, ilman ihmisten ohjausta ja valvontaa. Tällaiset algoritmit, kuten suosittelujärjestelmät, vaikuttavat esimerkiksi siihen, mitä meille mainostetaan, millä hinnalla ja miten.

Algoritmeja voidaan hyödyntää myös määrittämällä riskiluokituksia esimerkiksi rikokseen syyllistymiselle tai antamalla suosituksia rangaistuksien kovuudelle. Tekoälyä käytetään tällaisiin tarkoituksiin muun muassa Yhdysvalloissa.[16] Myös Iso-Britanniassa poliisi suunnittelee hankkivansa tekoälyyn pohjautuvan sovelluksen, jonka avulla voidaan arvioida henkilöiden todennäköisyyttä syyllistyä väkivaltarikoksiin.[17] Suomalaisetkin instituutit ovat kiinnostuneita soveltamaan tekoälyä erilaisissa tilanteissa, mistä kertoo esimerkiksi se, että Espoossa on selvitetty tekoälyn kykyä ennakoida lastensuojelun tarvetta.[18] Lisäksi Maahanmuuttovirasto on kehittänyt valmiuttaan automaattiseen päätöksentekoon ja hallitus on valmistellut lakiesityksen henkilötietojen käsittelystä maahanmuuttohallinnossa siten, että se mahdollistaisi automatisoidut yksittäis-päätökset.[19] Esityksessä ehdotetaan, että automatisoitua päätöksentekoa sovellettaisiin ensisijaisesti vain niihin tilanteisiin, joissa oleskelulupa päätettäisiin myöntää.

Tekoälyn tekemät päätökset eivät kuitenkaan välttämättä ole luotettavampia kuin ihmisten tekemät päätökset, vaikka päätöksenteon taustalla olevat algoritmit olisivat avoimesti tarkasteltavissa ja vaikuttaisivat virheettömiltä. Ongelmaksi voi muodostua algoritminen diskriminaatio. Tällöin syynä on useimmiten tekoälyn kehittämiseen hyödynnetty harjoitusdata, joka on sisältänyt vinoumia ja vääristymiä. Tekoälyn toiminnan taustalla olevat algoritmit suunnitellaan suhteessa dataan, joka ilmentää syrjivää historiaa ja nykyisyyttä, minkä takia lopputuloksena on usein syrjivä tekoäly.[20] Tämän johdosta tekoälyn hyödyntäminen päätöksenteossa, joka vaikuttaa suoraan ihmisen saamaan palveluun, seurantaan tai rangaistukseen, voi toistaa ja vahvistaa syrjiviä mekanismeja yhteiskunnassamme.

media Based Online Information Security for Senior Novice Computer Users.

[16] Edellisestä ks. esim. Berk ym. (2017); jälkimmäisestä ks. esim. Kugler (2018).

[17] Baraniuk 2018.

[18] Vuolteenaho 2018.

[19] HE 224/2018, Hallituksen esitys eduskunnalle laiksi henkilötietojen käsittelystä maahanmuuttohallinnossa ja eräiksi siihen liittyviksi laeiksi, https://www.finlex.fi/fi/esitykset/he/2018/20180224

[20] Chander 2016.

Algoritmista päätöksentekoa voisi hyödyntää myös yhteiskunnallisten syrjivien rakenteiden havaitsemisessa. Algoritmien tekemistä päätöksistä voitaisiin seurata, millaisiin lopputuloksiin saatavilla oleva aineisto ja kriteeristö johtaisi. Algoritmeja ei käytettäisi lopullisten päätösten tekemiseen, vaan algoritmin antamia ehdotuksia verrattaisiin ihmisten tekemiin päätöksiin. Jos algoritmin tekemät päätökset näyttäisivät osoittavan syrjinnän merkkejä, tuloksien perusteella voitaisiin tarkastella uudestaan jo tehtyjä päätöksiä. Tavoitteena olisi löytää syitä algoritmin ehdottamalle syrjinnälle. Tämä voisi paljastaa olemassa olevia syrjiviä rakenteita, mikä kenties mahdollistaisi tällaisten rakenteiden vähittäisen purkamisen.

Vastuu

Oikeusfilosofisesta näkökulmasta toimijuuteen kohdistuvat muutokset vaativat pian vastauksia, jotta muun muassa oikeudelliset seuraukset saadaan kohdennettua oikeudenmukaisesti. Kun tekoäly levittäytyy älykkäiden ympäristöjen, erilaisten robottiassistenttien ja älylaitteiden mukana laajasti yhteiskuntaan ja osallistuu toimintaan, milloin suosituksin, milloin toimien omistajansa puolesta[21], milloin ympäristöönsä reagoiden, on koko ajan vaikeampi nähdä, kuka on lopulta aktiivinen toimija. Tekoäly luo uuden kerroksen toiminnan ja ihmisen väliin, jolloin sen toiminnan ymmärrettävyys ja läpinäkyvyys käytön yhteydessä muodostuu yhä merkityksellisimmäksi.

Vastuunalaisia tahoja on vaikeampi todentaa tekoälyn toimiessa entistä autonomisemmin, ilman ihmisen valvontaa. Onko esimerkiksi itseohjautuvan auton ajama kolari tekoälyn suunnittelijoiden, automerkin, viranomaisten vai peräti itse tekoälyn vastuulla? Entä miltä tämän vastuun jakautuminen näyttää, jos tekoälyn taustalla ei ole enää yksittäinen luonnollinen henkilö, vaan esimerkiksi kollektiivinen toimija, kuten yritys? Vaikka nämä vastuutyhjiöt eivät ole täysin vältettävissä, on ne pyrittävä ratkaisemaan mahdollisimman kattavasti ennen tällaisten teknologioiden laajamittaista käyttöä.

Kuten sanottu, tekoälyä on mahdollista soveltaa laajaan kirjoon tehtäviä, kuten luokitteluun, päätöksentekoon ja toiminnan ohjaamiseen, hahmontunnistukseen ja lukuisiin muihin käyttötarkoituksiin. Esimerkiksi automatisoitu päätöksenteko mahdollistaa erinäisten prosessien tehostamisen ja ihmistyön tukemisen eri tavoin, mutta voi johtaa myös eettisesti kestämättömiin lopputulemiin. Mikäli sidosryhmillä ei ole mahdollisuutta saada tietoonsa (1) päätöksentekoprosessin kulkua sekä (2) tehtyyn päätökseen vaikuttaneita tekijöitä, muodostuu läpinäkyvyyden puute ongelmaksi vastuun näkökulmasta. Yksilön näkökulmasta on olennaista, että prosessin lopputulokseen vaikuttaneet muuttujat voidaan saada tietää, koska tämä tieto on olennaista, jos tekoälyjärjestelmän tekemä päätös tahdotaan esimerkiksi kiistauttaa.[22]

Ymmärrys tekoälyn toimintaperiaatteista ja operoiduista muuttujista on olennaista vastuun attribuoimisen kannalta. Koneoppivat järjestelmät muodostavatkin mallinsa syötetyn datan perusteella, jolloin vastuukysymystä tarkasteltaessa on syytä ottaa huomioon datan laatu ja määrä sekä datankeruusta vastuussa olevien luonnollisten henkilöiden osallisuus. Se, missä määrin ihmisen osallisuus astuu kuvaan, on yleisestikin kysymys, jota tulee tarkastella vastuukysymyksen kohdalla monella tapaa.

Vastuun jakautumista tarkasteltaessa tulee siis

[21] Muun muassa Googlen Duplex, joka pystyy varaamaan vaikkapa kampaajan tai ravintolan, ks. https://www.digitaltrends.com/home/what-is-google-duplex/

[22] Ks. esim. Wachter ym. (2017).

ottaa huomioon ainakin seuraavat seikat:

1) Kuka on (suorasti/epäsuorasti) vastuussa datan keruusta? Datan keräämisessä on otettava huomioon mm. seuraavat seikat:
- Harjoitusdatan edustavuus, datan määrä, datankeruumenetelmät, datan julkisuus ja avoimuus, käytetyn datan rajoitteet ja ajankohtaisuus.

2) Miten järjestelmä toimii/mitkä ovat mallin toimintaperiaatteet?
- Mallissa käytetyt muuttujat, mallin operoimien käsitteiden/muuttujien konstruointi, mallin tarkkuus ja luotettavuus.

3) Missä määrin ihminen on osallisena prosessissa kokonaisuutena?
- Algoritmin käytöstä suorasti ja epäsuorasti vastuussa olevat henkilöt, käyttäjän toiminta, käyttöliittymän läpinäkyvyys (nk. algoritminen presenssi; tietääkö käyttäjä, että tekoälyä käytetään).

Vaikutukset ihmiselämään ja inhimillisyyteen

Tekoälyn kehitys haastaa meidät miettimään pohjimmaisia kysymyksiä arvoistamme ja elämästä. Teknologian on usein nähty rikastuttavan ihmisten elämää ja tukevan heidän kukoistustaan. Tekniikan edistyksellä voi kuitenkin olla myös päinvastaisia vaikutuksia. Esimerkiksi sosiaalisen median aktiivisen käytön on joissain tutkimuksissa todettu tekevän ihmisistä onnettomampia.[23][24] Tekoälyllä onkin merkittävä vaikutus elämäämme ja hyvinvointiimme, ja teknologiset edistysaskeleet vaativatkin mukanaan laajempaa kansalaiskeskustelua siitä, mitä todella arvostamme ja millaista elämää haluamme elää. Ihmisen itsemääräämisoikeutta tulee kunnioittaa kaikissa tilanteissa teknologiasta riippumatta.[25] Nämä arvokysymykset eivät ole vain yksittäisten yritysten päätettävissä vaan vaativat myös yksilöiden ja poliittisten instituutioiden osallistumista dialogiin.

Esimerkin omaisesti ihmiselämän muutokseen liittyvät kysymykset ovat erityisen polttavia terveyden- ja vanhustenhuollon kohdalla. Tekoäly voi merkittävästi tehostaa terveydenhuoltoa, mutta saattaa toisaalta johtaa myös ihmiskontaktin vähenemiseen erilaisten digitaalisten hoitopalvelujen sekä hoitorobottien myötä. Tämä saattaa heikentää potilaiden vapauden-, autonomian- ja omanarvontunnetta. Kysymykset potilaiden profiloinnista, potilastiedoista ja yksityisyydestä nousevat koko ajan tärkeämmäksi, eikä alati digitalisoituvan elämän ja valvontayhteiskunnan psykologisia vaikutuksia tule myöskään jättää huomiotta. Yleisesti suhteemme tekoälytoimijoihin luo meille uusia haasteita. Voimmeko muodostaa merkityksellisiä vuorovaikutussuhteita robottien kanssa ja onko tämä ylipäätään toivottavaa?

Eräs tekoälyn tuoma uhka on lisääntyvä eristäytyminen omiin kupliimme ja vieraantuminen laajemmasta todellisuudesta. Poliittinen polarisaatio on jo nyt merkittävässä nousussa ympäri maailman.[26] Algoritmit mahdollistavat entistä kehittyneemmät tavat levittää harhaanjohtavaa tietoa. Disinformaatio-kampanjat ja niiden aiheuttama epäluottamus instituutioita kohtaan uhkaavat yhteiskunnallista tasapainoa ja horjuttavat uskoamme yhteiskunnalliseen päätöksentekoon. Tällaisen kehityksen sijaan tekoäly tulisi

[23] Royal Society for Public Health: #StatusOfMind - Social media and young people's mental health and wellbeing https://www.rsph.org.uk/uploads/assets/uploaded/62be270a-a55f-4719-ad668c2ec7a74c2a.pdf

[24] Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No More FOMO: Limiting Social Media Decreases Loneliness and Depression. Journal of Social and Clinical Psychology, 37(10), 751-768.

[25] Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404): 751-752.

[26] Mccoy, J. & Rahman, T. & Somer, M. (2018).

valjastaa tukemaan aktiivisempaa osallistumista kansalaisyhteiskuntaan, julkiseen keskusteluun ja demokratiaan. Tekoäly myös pakottaa meidät miettimään suhtautumistamme demokratiaan. Mitkä asiat olemme valmiita luovuttamaan koneiden ja algoritmien päätettäväksi, varsinkin, kun ottaa huomioon, että useimmat näistä ovat yksityisten yritysten hallinnassa?

Moraalimme ja arvojärjestelmämme muuttuvat hitaasti, ja kehittyvä teknologia suuntaa tätä muutosta osaltaan. Haluammeko siirtyä maailmaan, jossa isovanhemmistamme huolehditaan robotein, kommunikaatiomme on digitalisoitunut, algoritmit päättävät meille näytettävät sisällöt sekä mikrosirut ovat tehokkaan työntekijän mitta? Kenellä on oikeus päättää tästä kehityssuunnasta? Tekoälysovellukset muuttavat radikaalisti toimijuuttamme: sitä miten toimimme ja olemme olemassa, ja minkälaisiin identiteetteihin ja kulttuurillisiin instituutioihin tukeudumme. Toimijuuden muutoksen ymmärtäminen vie meidät ihmiskäsityksemme ytimeen. Mikä on ihmisyytemme perusta - eroammeko todella koneista luovuutemme, empatiamme tai rationaalisuutemme perusteella? Teknologinen kehitys haastaa käsityksemme ihmisyyden olemuksesta.

Työllisyys ja taloudelliset vaikutukset

Tekoäly tulee todennäköisesti edistämään innovaatioita ja taloudellista kasvua merkittävästi. Tämä on pitkälti seurausta automaation mahdollistamasta tuottavuuden kasvusta. Tekoäly tuo kuitenkin mukanaan myös sosioekonomisia uhkia.[27] Nämä uhkat eivät sinällään liity teknologian mahdollistamiin taloudellisiin hyötyihin itseensä, vaan näiden hyötyjen jakaantumiseen yhteiskunnassa. Tarkastelussa on tärkeä erottaa toisistaan lyhyen aikavälin työn murros ja sen aiheuttama epävakaus pidemmän aikajänteen talouskasvun oikeudenmukaisuudesta. Lisääntyvän automatisaation ja robotisaation johdosta taloudellinen epätasa-arvo saattaa lisääntyä merkittävästi, ainakin hetkellisesti. Tuottavuushyödyt siis kätkevät allensa potentiaalisesti lisääntyvät tulo- ja varallisuuserot. Uhkakuvia ovat työpaikkojen katoaminen ja palkkatason lasku, erityisesti matalan palkkatason tehtävissä. Rakenteelliset muutokset taloudessa tarkoittavat myös, että työntekijöiden taidot eivät välttämättä enää vastaa markkinoiden tarpeita. Työmarkkinat jakautuvat entistä vahvemmin häviäjiin ja voittajiin.

Vaikka tekoäly mullistaa toimialoja läpi elinkeinoelämän, ei ole selvää, että se johtaa välttämättä työpaikkojen vähenemiseen. Näin ei ole tapahtunut ainakaan historiallisesti.[28] Teknologisesta kehityksestä huolimatta työhön käytetty aika on vähentynyt vuosikymmenten saatossa vain vähän - olemme olleet taitavia keksimään uusia tarpeita ja niihin liittyviä työtehtäviä. Vanhat elinkeinot ovat korvautuneet uusilla. Tekoälyratkaisut selviytyvät kuitenkin jo monista työtehtävistä ihmistä paremmin, näin tehden tästä teollisesta vallankumouksesta edellisiä haastavamman. Lisäksi automatisaatio voi johtaa etenevissä määrin varallisuuden keskittymiseen harvojen käsiin. Työpaikkojen katoaminen ja varallisuuserot ovat erityisen huolestuttavia, sillä ne iskevät pahiten jo valmiiksi yhteiskunnan heikko-osaisimpiin. Tutkimusten mukaan automaatioille alttiimmat työtehtävät liittyvät järjestäen alhaisen palkkatason, osaamisen ja koulutuksen työpaikkoihin.[29] Tekoälyn nousun myötä erot pääomassa ja omistuksissa muodostuvat talouden tarkastelussa tärkeämmiksi kuin

[27] Koski, O., & Husso, K. (2018). Tekoälyajan työ: neljä näkökulmaa talouteen, työllisyyteen, osaamiseen ja etiikkaan. Työ- ja elinkeinoministeriön julkaisuja 2018(19).

[28] Royal Societyn ja British Academyn raportti tekoälyn työllisyysvaikutuksista: https://www.britac.ac.uk/projects/ai-and-work

[29] Furman & Seamans 2018, 15.

perinteiset tuloerot.

On tietenkin totta, että useimmat nykyisistä tekoälysovelluksista on suunniteltu tiettyyn tehtävään, jossa ne avustavat ihmisiä heidän korvaamisensa sijaan. Vaikka harva ammatti on vielä täysin automatisoitavissa, merkittäviä osia nykyisistä työtehtävistä voidaan kuitenkin automatisoida. On esimerkiksi arvioitu, että nykyisistä työtehtävistä 50% on automatisoitavissa nykyteknologian avulla.[30] Tämä johtaa vähintäänkin merkittäviin muutoksiin tehtävien työnkuvassa. Oletettavasti sosiaalisuutta, abstraktia ajattelua ja luovuutta vaativat tehtävät eivät ole yhtä haavoittuvaisia automatisaatiolle. Samaten on esitetty, että kaoottiset ja nopeaa reagointikykyä vaativat tehtävät ovat tekoälyltä parhaiten suojassa.[31] Mekaanisista työtehtävistä vapautuva aika voidaan käyttää hyväksi siirtämällä työntekijöitä enemmän tällaisiin koneille haastaviin tehtäviin.

Teknologian kehityksen muuttamassa yhteiskunnassa yhä useampi voi tarvita tukea. Tekoälyn kehityksen tuomiin sosioekonomisiin haasteisiin, kuten eriarvoisuuden kasvuun on esitetty vastaukseksi sosiaaliturvan uudistuksia, kuten perustuloa. Kaikille kansalaisille maksettu vastikkeeton sosiaaliturva saattaisi myös tehdä työmarkkinoistamme toimivammat teknologisoituneeseen toimintaympäristöön. Nämä ovat kuitenkin kalliita ratkaisuja, jotka vaatisivat lisää verotuloja. Jotkut tahot ovatkin esittäneet robottiveron käyttöönottoa automatisaation mahdollisesti vähentäessä ansiotuloverojen kertymää.

Erityisen tärkeänä voidaan pitää lisäpanostuksia uudelleenkoulutukseen ja elinikäiseen oppimiseen, jotta työelämän murrokseen pystytään vastaamaan. Valtion ohella myös yrityksillä on tärkeä rooli työntekijöiden ammattitaidon kehittämisessä pitkäjänteisesti. Tekoälyn myötä työmarkkinoiden joustavuus on entistä tärkeämpää. Kaiken tämän ohella on huomattava myös työelämän henkinen puoli: työ on tärkeä osa identiteettiämme ja omanarvontunnettamme. Mikäli tekoälyn kehitys ei luo uusia työpaikkoja vanhaan tapaan koneiden suoriutuessa ihmistä paremmin, saattaa tarpeellisuuden tunteemme kärsiä.

Kehittyneet AI-teknologiat ja supertekoäly

AI-teknologioiden kehittyminen yhä älykkäämpään ja autonomisempaan muotoon tuo pohdittavaksi filosofisesti, oikeudellisesti ja yhteiskunnallisesti kiinnostavia kysymyksiä. Populaarikulttuurissa esiin nousevat ihmisen kaltaiset robotit ovat vielä yleisen käsityksen mukaan kaukana, mutta jo lähitulevaisuudessa voi nousta esiin monimutkaisia, vaikutuksiltaan kenties ennakoimattomia teknologioita. Kehittyneet teknologiat eivät välttämättä ota kuitenkaan ihmistä muistuttavaa muotoa, vaan saattavat olla yhteydessä meihin ja ympäristöömme huomaamattomilla tavoilla, sulautuen ympäristöömme ja muovaten sitä alin omaan. Kehittynyt teknologia viittaa tässä yhteydessä joukkoon teknologioita, jotka hyödyntävät tekoälyä eri tavoin ja siirtävät päätöksentekoa ja toimintaa etäämmälle inhimillisestä toimijasta. Tällainen teknologia luo uuden kerroksen ihmisen ja toiminnan väliin, ja mahdollisesti tulevaisuudessa jopa täysin ihmisestä erillisen toimijuuden muodon.

Joissakin tällaisissa teknologioissa korostuvat ohjelmiston (software) lisäksi myös laitteiston (hardware) eettiset ulottuvuudet sekä vuorovaikutusta ohjaavat tekijät. Esimerkiksi sosiaalista robotiikkaa suunniteltaessa eettiset kysymykset

[30] McKinsey 2017, 2.

[31] Osoba O. & Welser IV W. (2017). The Risks of Artificial Intelligence to Security and the Future of Work. Santa Monica, CA: Rand Corporation. 12.

eivät koske vain robotin kuoren alla piilevää ohjelmointia vaan myös robotin käyttöliittymä sekä sen materiaalista suunnittelua. Tekijöiden, kuten robotin äänen, kehollisten piirteiden (tai kehon puuttumisen)[32], keinotekoisten emootioiden ilmaisun[33], yleisen esteettisen muodon sekä antropomorfismin asteen[34], on huomattu vaikuttavan ihmisten käsityksiin ja vuorovaikutukseen robottien kanssa. Robotin materiaalisten piirteiden lisäksi sen toimintaympäristö ja konteksti vaikuttavat tapoihin, joilla ihmiset suhtautuvat robottiin. Teknologiaa ei voida koskaan erottaa materiaalisesta, sosio-kulttuurisesta ja yhteiskunnallisesta käyttökontekstistaan. Eettinen suunnittelu ei siis koske vain ja ainoastaan robotin toiminnallisuutta ja käytettävyyttä, vaan sen istumista laajempaan ekologiseen ympäristöön, osaksi ihmisten toimintaa ja yleisesti hyväksyttyjä yhteiskunnallisia käytäntöjä.

Toiset kehittyneet teknologiat, kuten vielä hypoteettiset yleis- ja supertekoäly, nostavat keskustelunaiheeksi tekoälyagenttien oikeudet ja velvollisuudet sekä globaalin mittakaavan muutokset, joita tällaisen teknologian kehittäminen voivat mahdollisesti aiheuttaa.[35] Dystooppiset tulevaisuudenkuvat, kuten ihmislajin tuhoava supertekoäly, voivat hypoteettisuudestaan huolimatta valottaa tekoälyyn liittyviä eettisiä ongelmia. Esimerkiksi tekoälyn tavoitteiden määrittely on aina tärkeää, jotta vältytään tilanteilta, joissa tekoäly tavoittelee päämääriään ennakoimattomalla, vahingollisella tavalla. Vähintäänkin uhkakuvat osoittavat, että teknologisen kehityksen suuntaviivojen ja päämäärien eettinen arviointi on välttämätöntä, sillä teknologisilla luomuksillamme voi olla potentiaalisesti kyky aiheuttaa kriisejä globaalilla mittakaavalla.

Nano- ja bioteknologia sekä esimerkiksi kehittynyt proteesiteknologia voivat tulevaisuudessa tarjota mahdollisuuksia ihmisen fyysisen, kognitiivisen ja emotionaalisen kapasiteetin kehittämiseen (augmenting tai enhancing) sekä eliniän pidentämiseen ja yleisen terveydentilan parantamiseen. Hieman eri luokkaa edustavat esimerkiksi ympäristöön upotettu "jokapaikan teknologia" (ubiquitous technology) ja ambientti älykkyys, jotka vaikuttavat ihmisen toiminnan taustalla, osana ympäristöä, säätelemällä esimerkiksi huoneen lämpötilaa tai täydentämällä jääkaapin sisältöä preferenssiemme pohjalta. Niin kutsuttu esineiden internet (Internet of things, IoT) yhdistää erilaiset teknologiat toisiinsa internetyhteydellä toimivaksi verkostoksi. Se mahdollistaa vuorovaikutuksen ja tiedonsiirron eri laitteiden - esimerkiksi kodin valvontajärjestelmän, älykellon tai -puhelimen, valaistuksen tai minkä tahansa muun verkkoon kytketyn laitteen - välillä, ja siten eri teknologioiden toiminnan toistensa toimintaa täydentävinä ja yhdistävänä.

Kuten augmentaatio- ja paranteluteknologiat, ympäristöön upotetut teknologiat ovat eräänlainen jatke ihmiselle ja hänen toiminnalleen. Kaikki edellä mainitut teknologiat tähtäävätkin ihmisen toimintakyvyn tukemiseen ja parantamiseen, mutta muodostavat myös omanlaisiaan potentiaalisia ongelmia ja kysymyksiä. Raportissa Converging Technologies - Shaping the Future of European Societies[36] esitetään, että joka-

[32] Crowelly, C. R., Villanoy, M., Scheutzz, M., & Schermerhornz, P. (2009, October). Gendered voice and robot entities: perceptions and reactions of male and female subjects. Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on (s. 3735-3741). IEEE.

[33] Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of emotional body language displayed by a humanoid robot: A case study with children. International Journal of Social Robotics, *5*(3), 325-334.

[34] Duffy, B. R. (2003). Anthropomorphism and the social robot. Robotics and autonomous systems, 42(3-4), 177-190.

[35] Max Tegmark on esitellyt 12 mahdollista skenaariota, jotka voivat seurata supertekoälyn kehityksestä. Tiivistelmä skenaarioista: https://futureoflife.org/ai-aftermath-scenarios/?cn-reloaded=1.

[36] Altmann, D., Andler, K., Bruland, K., Nakicenovic, N., & Nordmann, A. (2004, 31-33). Converging Technologies - Shaping the Future of European Societies.

paikan teknologioiden tuottamat mahdollisuudet sisältävät myös ainakin neljänlaisia riskejä:

1) Teknologiat, jotka muokkaavat ihmisyyden rakennetta ja ihmisluontoa, voivat tuottaa riskejä, jotka koskevat ihmisten ja ihmisryhmien itseymmärryksen ja identiteetin muutoksia.

2) Edistys augmentointi- ja paranteluteknologioiden alalla voi saada meidät (2a) aliarvioimaan riippuvaisuuttamme luonnosta ja täten (2b) luoda illuusion siitä, että (erityisesti ekologiset) globaalit ongelmat ovat hallinnassamme.

3) (3a) Teknologisen kehityksen myötä ihminen delegoi yhä useampia toimintoja ja tehtäviä teknologialle, mikä voi (3b) johtaa hitaaseen autonomian katoamiseen ja (moraalisen) vastuutyhjiön syntymiseen.

4) Teknologian näkymättömyys tekee (4a) kyseistä teknologiaa käyttävien ja sen käytöstä hyötyvien ihmisten sekä (4b) sen läsnäolon tai poissaolon erottamisesta hankalaa. Lisäksi, (4c) tällaiset teknologiat sisältävät riskin vallan epätasaisesta jakautumisesta, mikä voi horjuttaa yhteiskunnallista tasapainoa.

Kehittyvien tekoälyteknologioiden mukanaan tuomat turvallisuusriskit on otettava vakavasti. Niiden potentiaalinen vaikutus yhteiskunnan toimintaan eri tasoilla voi luoda tilanteen, jossa tekoälyn satunnaiset viat tai tahallinen hakkerointi muodostavat täysin uudenlaisia, kompleksisesti rakentuneita ja vaikutuksiltaan laajaalaisia riskejä, joihin on syytä varautua. Tällä hetkellä riskinäkökulma on tieteellisessä keskustelussa vahvasti esillä, ja teknologioiden turvallisuus on suuressa roolissa tuotteita kehitettäessä. Riskien tunnistamisen lisäksi vaaditaan myös yhteiskunnallisella tasolla eettistä tarkastelua. Tämän tarkastelun tulee keskittyä kysymyksiin, jotka koskevat vallan jakautumista, teknologisen kehityksen suuntaa sekä sen mahdollisia (globaaleja) ekologisia ja taloudellisia vaikutuksia. Keskustelu filosofisista kysymyksistä koskien juridista ja moraalista vastuuta sekä ihmisyyden perustavaa luontoa on yhtä lailla tärkeää, sillä se voi auttaa meitä asettamaan päämääriä ja suuntaviivoja, mutta myös rajoja teknologisen kehityksen suunnalle.

Tekoälyn eettisen arvioinnin lähtökohdat

Läpinäkyvyys

Läpinäkyvyys on tärkeä periaate tekoälyteknologioiden kehittämisessä. Yksinkertaisimmillaan se tarkoittaa, että varmistetaan eri sidosryhmien - käyttäjien, kansalaisten, kehittäjien, omistajien - riittävä ymmärrys sovellettavan koneoppimisratkaisun tai tekoälysovelluksen toimintamekanismeista. Tapauskohtaisesti on tärkeää keskustella, mikä on riittävä ymmärryksen tai läpinäkyvyyden taso: luonnollisesti toimintamekanismin logiikan ymmärtäminen voi monimutkaisissa sovelluksissa vaatia pitkällistä perehtymistä käytettävään teknologiaan. Tätä toimintalogiikkaa tulisi kuitenkin avata myös yleisemmällä tasolla, niin että käytettäviä teknologioita ymmärtämättömät käyttäjät hahmottavat kokonaisuutta.

Läpinäkyvyyden tarve korostuu sitä vahvemmin, mitä tärkeämpiä päätöksiä tekoälysovelluksella tehdään. Usein julkisen sektorin päätöksenteko on korostuneessa asemassa: jos yhteiskunnassa siirrytään tekemään tai avustamaan tekoälyllä päätöksiä esimerkiksi tulonsiirroista, huoltajuusasioista tai kaavoitushakemuksista, on läpinäkyvyyden tarve suuri. Vaikka Euroopan unionin GDPR-lainsäädäntö antaa kansalaisille oikeuden kieltäytyä koneellisesta tai automaattisesta päätöksenteosta, tulee läpinäkyvyyden ihanne toteutua mahdollisimman hyvin niille, ketkä eivät tätä kieltävää oikeuttaan halua käyttää. Algoritmista päätöksentekoa käyttävillä toimijoilla on velvollisuus tarjota päätöksenteon logiikasta ja päätöksen kriteereistä selitys ymmärrettävässä muodossa.[37]

Toinen merkittävä läpinäkyvyyden vaatimuksen alue on laajassa käytössä olevat verkkopalvelut ja -alustat. On epäselvää, miten voimme tarkastella esimerkiksi Googlen, Facebookin ja muiden suurten teknologiayhtiöiden toimintaa läpinäkyvyyden näkökulmasta. Tällä hetkellä parhaana keinona on kuluttaja-aktiivisuus sekä lainsäädännön määrittelemiin vaateisiin liittyvä yhteistyö.[38]

Läpinäkyvyyden toteutumista voi käytännön

[37] GDPR 15 artiklan 1 kohdan h alakohta

[38] Brundage, M., & Bryson, J. (2016). Smart Policies for Artificial Intelligence. arXiv preprint arXiv:1608.08196.

kehitystyössä varmistaa eri suunnittelu- ja muotoilunäkökulmien hyödyntämisellä. Suunnittelutyötä pitäisi suorittaa käyttäjän näkökulmasta ja järjestelmällisesti kartoittaa sitä, miten käyttäjä ymmärtää toteutettavan tekoälysovelluksen toimintaa. On tärkeää ottaa huomioon myös erilaiset väestöryhmät, ikäluokat ja käyttötottumukset.

Läpinäkyvyyttä lisää myös käyttäjäkunnalle, kehitysyhteisölle, tutkijaryhmille tai julkiselle yleisölle annettavat tarkastusmahdollisuudet.[39] Esimerkiksi julkisen sektorin käyttämiä algoritmeja voisi päästä tarkastelemaan asiaan perehtynyt tutkijoiden joukko, joka varmistaisi, että algoritmin tekemät päätökset soveltavat tasavertaisuuden, oikeellisuuden ja teknisen oikeaoppisuuden periaatteita. Algoritmisen diskriminaation huomaamisessa ja korjaamisessa kontrolloidusti myönnetty pääsy tarkastelemaan päätöksentekoalgoritmin toimintaa voi olla eduksi.

Vastuullisuus

Tässä yhteydessä vastuullisuudella tarkoitetaan yrityksen ja julkisten toimijoiden velvollisuutta selittää ja perustella algoritmien tekemiä päätöksiä, raportoida niistä, sekä vähentää niiden mahdollisia haittavaikutuksia yhteiskuntaan ja yksilöihin.

Vastuullisuuden perustana ovat selkeät toimintaohjeet, joita seurata, jotta mahdolliset ongelmat voitaisiin ennaltaehkäistä, sekä varmistaa, että mikäli tuote aiheuttaa vahinkoa käyttäjälleen tai ympäristölleen, yrityksessä tiedetään miten toimia vahingon korjaamiseksi. Kysymykset siitä, kuka on vastuussa vahinkotilanteessa, kuinka vahinko korvataan, miten vahingoista raportoidaan ja miten algoritmin toimintaa säädellään, jotta vastaava ei toistuisi, ovat tärkeitä selvitettäviä jo ennen tuotteen julkaisemista.

Vastuullisuuden kysymykset voidaan karkeasti jakaa osakysymyksiin vastuusta, selitettävyydestä, virheettömyydestä, tarkastettavuudesta ja oikeudenmukaisuudesta.

Taulukko 2. Vastuullisuuden osakysymykset

Vastuu	• Kuka on vastuussa, mikäli tuote aiheuttaa haittaa käyttäjälle? • Minkälainen raportointiprosessi on, miten haitta korvataan? • Kenellä on mahdollisuus päättää tarvittavista muutoksista algoritmisen toimijan suunnittelun aikana, julkaisua edeltävänä ja julkaisun jälkeisenä aikana?
Selitettävyys	• Keitä ovat loppukäyttäjät ja sidosryhmät? • Kuinka paljon käyttäjille ja sidosryhmille on mahdollista kertoa tuotteesta/algoritmista? • Kuinka suuri osa datalähteistä on julkista materiaalia/mahdollista julkaista?
Virheettömyys	• Mitä virheen mahdollisuuksia on ja miten niiden vaikutus minimoidaan/hallitaan? • Kuinka luotettavia algoritmin tekemät päätökset ovat? • Mitä ovat realistiset "worst case scenariot", jos miettii vaikutuksia yhteiskuntaan, yksilöihin ja sidosryhmiin? • Ovatko datalähteet todenmukaisia, mitä muita vaihtoehtoisia lähteitä olisi olemassa?
Tarkastettavuus	• Onko julkinen auditointi mahdollinen, vai käyttääkö algoritmi arkaluontoisia lähteitä, joiden takia auditoinnin voisi suorittaa ainoastaan määrätty taho? • Kuinka varmistaa, että mahdollinen auditointi ei avaa algoritmia manipuloinnille?
Oikeudenmukaisuus	• Onko olemassa ryhmiä, jotka hyötyvät tai häviävät algoritmin toiminnan seurauksena? • Minkälaisia mahdollisia vahingollisia vaikutuksia, epävarmuuksia tai virheitä algoritmi voi tuottaa eri ryhmille?

[39] Social Impact Statement for Algorithms, www.fatml.org

Oikeudet ja oikeusvaltio

Perustavanlaatuisten oikeuksien ja oikeudenmukaisuuden ymmärtäminen on väistämätön toiminnan eettisen analyysin perusta. Oikeuksien kunnioittaminen ja eettinen ajattelu rakentuvat näiden varaan. Perustavan-laatuiset oikeudet luovat pohjan lainsäädännölle ja kansalaisten toiminnalle. Kaikilla valtion kansalaisilla ja viranomaisilla on velvollisuus noudattaa oikeussääntöjä. Ylimpänä oikeusohjeena Suomessa on perustuslaki, joka takaa jokaiselle tietyt vähimmäisoikeudet, jotka perustuvat kansainvälisiin ihmisoikeussopimuksiin, jotka nekin sitovat valtioita myös sellaisenaan.[40]

Tekoälysovellukset eivät saa rikkoa perustuslain takaamia perusoikeuksia vastaan. Varsinkin yksityisyydensuoja ja syrjimättömyys ovat seikkoja, joihin on kiinnitettävä huomiota ensimetreiltä lähtien. Toisaalta myös muualta lainsäädännöstä löytyviä toimialaspesifimpiä oikeusohjeita on seurattava. Tekoälysovelluksen kehittäjän - ja myös hyödyntäjän - on siis tiedettävä, mitkä lainsäädännön vaatimukset koskevat häntä ja toimittava niitä kunnioittaen. Jotta tarvittavia oikeusohjeita on mahdollista noudattaa, yrityksellä on oltava riittävästi tietoa toiminta-alueensa kulttuurisesta ja oikeudellisesta normistosta.

Vaikka tekoäly on teknologiana uusi, siihen liittyviin sovelluksiin pätevät monet yleisemmät säännökset. EU-alueella markkinoille tuotavien tuotteiden on täytettävä tuotevastuulainsäädännön ehdot, ja markkinoitaessa tekoälytuotteita on huomioitava kuluttajansuojan vaatimukset.[41]

EU-alueella toimivan on tunnettava EU:n tietosuoja-asetus (GDPR). GDPR pyrkii varmistamaan, että jokaisella EU-kansalaisella säilyy oikeus omiin henkilötietoihinsa, ja että henkilötietoja hyödyntävät tahot toimivat oikeudenmukaisesti ja eettisesti. Käytännössä tämä vahvistaa kansalaisten tietosuojaa ja oikeutta yksityisyyteen säätämällä, että henkilötietojen keräämiseen on oltava laillinen peruste, ja antamalla kansalaisille oikeuden tarkistaa heistä kerätyt tiedot ja mahdollisuuden oikaista virheelliset tiedot tai jopa poistaa tietonsa rekisteristä. GDPR asettaa samat lailliset velvoitteet myös EU-markkinoilla toimiville EU:n ulkopuolisille toimijoille, kun he käsittelevät EU-kansalaisten henkilötietoja.

Vaikka olemassa oleva lainsäädäntö kattaakin suuren osan tekoälysovellusten esiin nostamista kysymyksistä, on kuitenkin paljon alueita, joilla lainsäädäntö ei anna selkeää vastausta. Näitä pyritään selvittämään tällä hetkellä sekä kansallisesti että EU:n tasolla. Avoimet kysymykset liittyvät etenkin kyberturvallisuuteen ja todistustaakan määrittelemiseen sekä vastuun muodostumiseen: onko mahdollista soveltaa isännänvastuuta, mitä katsotaan korvattavaksi ja kuinka paljon vastuun määräytymiseen vaikuttaa se, olisiko vahinko ollut mahdollisesti estettävissä vai ei.

[40] Ks. YK:n ihmisoikeussopimuksista https://www.yk.fi/node/257 ja Euroopan ihmisoikeussopimuksista https://eur-lex.europa.eu/summary/glossary/eu_human_rights_convention.html?locale=fi

[41] Tämänhetkisestä EU-tasoisesta lainsäädännöstä voi katsoa lisää englanninkielisestä dokumentista: Commission Staff Working Document - Liability for emerging digital technologies https://ec.europa.eu/digital-single-market/en/news/european-commission-staff-working-document-liability-emerging-digital-technologies

Moniarvoisuus

On äärimmäisen tärkeää varmistaa tekoälykehityksen tapahtuvan moniarvoisessa ympäristössä. Monet algoritmisen syrjinnän muodot ovat juurikin seurausta siitä, että tekoälyn kehityksessä ei ole otettu huomioon näkökulmien moninaisuutta. Esimerkiksi kasvojentunnistustekoälyn kouluttaminen vain tietyn alueen kantaväestön kuvilla johtaa luultavasti siihen, että tekoäly ei tunnista etnisiä vähemmistöjä ihmisiksi lainkaan tai tekee huomattavia virheitä näiden kohdalla. Datan monipuolisuus on siis erityisen tärkeää, jotta vääristymiltä vältytään. Haittojen minimoimiseksi nämä syrjivät vinoumat tulisi huomata ja kitkeä pois jo kehittelyvaiheessa. Tämä vaatii moninaista taustojen ja arvojen kirjoa tekoälyn kehitykseen, sillä vain yhdestä perspektiivistä syrjiviä käytäntöjä on vaikea tunnistaa. Näin on eritysesti siksi, että nämä vinoumat liittyvät syvälle juurtuneisiin yhteiskunnallisiin asenteisiin ja stereotypioihin, joille olemme usein sokeita.

Puhutaankin niin sanotusta tekoälyn monimuotoisuuskriisistä: tekoälyn kehittäjät edustavat vain tiettyä osaa moninaisesta yhteiskunnastamme. Aliedustus nostaa riskin kokonaisten ihmisryhmien huomaamattomasta syrjinnästä, joka perustuu heidän etnisyyteensä, sukupuoleensa, ikäänsä, kulttuurinsa tai arvoihinsa. Kyse ei tällöin ole tarkoituksellisesta toiminnasta, vaan implisiittisistä ennakkoasenteistamme, jotka vaarantavat yhteiskunnallisen tasa-arvon. Monimuotoisuuden varmistaminen on erityisen tärkeää yhteiskunnallisesti merkittävien alueiden, kuten oikeuslaitoksen, terveydenhuollon ja rekrytoinnin piirissä, jotta vallitseva epätasa-arvo ei entisestään kärjisty. Ongelman selättäminen vaatii aktiivista inklusiivisuuden edistämistä tekoälyn kehityksessä. Algoritmit ovat käytännössä yhtä tasapuolisia kuin heidän kehittäjänsä. Yritysten, valtioiden ja järjestöjen on varmistettava, että monimuotoisuus toteutuu heidän työympäristössään, jotta tekoälyä ei kehitetä tai arvioida vääristynein perustein. Erityisesti yritysten tulee varmistaa, että tekoälyn kehittäjät edustavat moninaisia taustoja ja arvoja, ja että käytettävä data ei ole vääristynyttä. Inklusiivista tekoälyä voidaan myöhemmin käyttää myös vääristymien huomaamiseksi ja estämiseksi, kuten esitimme luvussa 2.3.

Moniarvoisuus on tärkeää myös talous- ja tuottavuusnäkökulmasta. Entistä moninaisempi työympäristö lisää luovuutta ja innovatiivisuutta, jotka ovat erityisen tärkeitä tehokkaassa tekoälyn kehityksessä. Useiden eri lähtökohtien yhdistyessä ongelmanratkaisu tehostuu. Diversiteetin on todettu lisäävän työn tehokkuutta, tuottavuutta ja luovuutta[42] ja se myös auttaa tekemään tekoälysovellukset sopiviksi laajemmalle joukolle potentiaalisia käyttäjiä, mikä voi laajentaa sovelluksen markkinoita. Koko maailman markkinoiden hyödyntäminen vaatii monimuotoisuuden tekemistä keskeiseksi yrityksen toimintaperiaatteeksi.

Yritysvastuu ja liiketoimintaetiikka

Tekoälyn kehitys korostaa yritysten yhteiskuntavastuuta, sillä yritysten sosiaalinen vastuu kasvaa samassa suhteessa heidän liiketoimintansa yhteiskunnallisten vaikutusten kanssa. Tekoäly on lisäksi kehittymässä pitkälti markkinaehtoisesti, ilman valtiovallan kontrollia. Yritysten itsesääntely on erityisen tärkeää nopeasti kehittyvällä alalla, jonka perässä lainsäädännön on vaikea pysyä. Tämä pakottaa yritykset laajentamaan näkemystään yhteiskuntavastuustaan. Yritysten ainoaksi tehtäväksi ei voida enää mieltää voiton tahkoamista osakkeenomistajille. Koska tekoälyllä on jopa ennennäkemättömän laajat vaikutukset yhteiskuntaan, ovat yhtiöt myös

[42] McKinsey (2018) Delivering through diversity.

vastuussa laajemmin yhteiskunnallisille sidosryhmilleen: työntekijöille, asiakkaille, rahoittajille ja yhteisöille. Lyhytnäköinen oman edun tavoittelu ei ole kannattavaa näin tärkeän teknologian kohdalla. Parhaat edut saavutetaankin sitomalla tekoälykehitys yhteiskunnallisten ongelmien ratkaisemiseen ja yhteisen arvon tuottamiseen.

Jokainen yritys on omalta osaltaan vastuussa turvallisesta tekoälykehityksestä, joka takaa ihmisten yksityisyyden, syrjimättömyyden ja hyvinvoinnin. Vastuullisuus alkaa huipulta, eettisestä johtamisesta, luoden avoimempaa ja turvallisuuskeskeisempää yrityskulttuuria. Erityisesti riskienhallinta, laadunvalvonta ja varovaisuusperiaatteen noudattaminen korostuvat tekoälyn kohdalla. Vastuullinen yritystoiminta on myös kannattavaa. Näin on erityisesti tekoälyn aikakaudella, jolloin henkilötietoihin ja dataan liittyvien väärinkäytösten hinta on korkea yritysten maineelle. Lisäksi kilpailu asiantuntijoista vaatii läpinäkyvää yrityskulttuuria ja mainetta luotettavana alan toimijana. Vastuullisen liiketoiminnan onkin huomattu edistävän tuottavuuden kasvua, kustannussäästöjä ja asiakaskollisuutta[43], eikä vastuullinen tekoälykehitys liene tässä poikkeus. Innovatiivinen ja kestävä liiketoiminta on sekä yrityksen että yhteiskunnan parhaaksi ja vähentää sääntelyn tarvetta.

Käytännössä tietoturvallisuuden ja yksityisyydensuojan kysymykset nousevat tekoälyn myötä yritysvastuun keskiöön. Samoin algoritmisen syrjinnän välttämiseksi tekoälysovellusten pitää olla ennustettavia ja ymmärrettäviä. Yritysten on panostettava tekoälysovellusten turvallisuuteen ja pyrkiä estämään niiden tahallinen väärinkäyttö. Ymmärtääkseen paremmin kuluttajia ja toimintaympäristöään yritysten on kannattavaa osallistua keskusteluun oikeudenmukaisesta tekoälystä ja sen tavoitteista kansalaisyhteiskunnan kanssa. Automatisaation mahdollisesti aiheuttama työn murros vaatii yrityksiltä myös aiempaa kattavampaa uudelleenkoulutusta ja työtehtävien uudelleenjärjestelyä. Edellä mainitut toimet ovat tärkeitä, jotta yritykset saavat luotua itsestään kuvan vastuullisina ja luotettavina toimijoina. Tekoälyn aikakaudella luotettavuus onkin yritysten suurin yksittäinen markkinavaltti. On tärkeää nähdä yritysvastuu mahdollisuutena: asettamalla yhteiskunnallisten ongelmien ratkaisun liiketoimintansa perustaksi yritykset avaavat uusia mahdollisuuksia ja saavuttavat kilpailuetua. Yritykset eivät kuitenkaan ole yksin vastuussa tekoälyn riskeistä, sillä yhtä lailla vaaditaan relevanttien yhteiskunnallisten toimijoiden, kuten valtion ja tutkimuslaitosten yhteistyötä.

Tekoälyn eettinen kehitys pohjaa tekoälykehittäjien, yritysten, tutkijoiden sekä päätöksentekijöiden rakentavaan kanssakäymiseen. Yhteistyö on tarpeellista informaatioteknologian etiikan perinteen sekä tekoälyteknologioihin liittyvien erityiskysymysten saumattomassa yhdistämisessä. Yhteiskunnallisten instituutioiden ja oikeusjärjestelmän on pysyttävä mukana tekoälyn kehityksessä riskien hallitsemiseksi lainsäädännön keinoin. Yritysvastuulinjaukset tukevat ja täydentävät lainsäädäntöä. Erityistä sääntelyä saattavat vaatia yhteiskunnallisesti keskeiset alat, kuten terveydenhuolto, oikeuslaitos, politiikka ja kansallisen turvallisuus. Tekoälyn sääntelyelinten perustamisen on tapahduttava yhteistyössä elinkeinoelämän ja tutkimuslaitosten kanssa. Erityisen tärkeää tämä on alan kansainvälisen hallinnan ja sääntelyn näkökulmasta. Suomen on osallistuttava aktiivisesti kansainväliseen yhteistyöhön sääntelyn yhdenmukaistamiseksi, jotta vältetään tekoälyn kilpavarustelu ja "race to the bottom" -ilmiö.

[43] Asemah, E.S., Okpanachi, R.A. & Edegoh, L.E.O. 2013.

Tekoälyn etiikan käytännön työkalut

Eettiset ohjeistukset ja oppaat

Tekoälyn etiikka on yhteiskunnallisen tutkimuksen alueena suhteellisen tuore. Suurin osa tieteellisestä kirjallisuudesta on 2010-luvulta ja erityisesti viime vuosilta. Tieteellisen kirjallisuuden lisäksi omaa tietämystään tekoälyn etiikasta voi syventää myös eri organisaatioiden julkaisemien raporttien ja teosten avulla.

Viime vuosien tuotoksia ovat tekoälyn ja teknologian eettiset ohjeistukset. Kansainvälisistä järjestöistä IEEE ja ACM ovat kunnostautuneet alueella: IEEE:n Global Initiative on Ethics of Autonomous and Intelligent Systems julkaisi toisen version ohjeistuksestaan vuonna 2018, ja ACM päivitti yleistä teknologiaetiikan ohjeistustaan samana vuonna.

Eettiset ohjeistukset voi nähdä ensiaskeleena kohti tarkempia eettisiä standardeja tai muodostettavaa lainsäädäntöä.[44] Oman huomionsa ansaitsevat yleiset etiikan tai ammattietiikan ohjeistukset – ne muodostavat vahvan perustan eettisen tekoälyn ja informaatioteknologian etiikan saralla.

Suomalaisista vastaavista organisaatioista mainittavia ovat erityisesti TEK - Tekniikan Akateemiset sekä Tivia. TEK on tuottanut erinomaisen oppaan tekniikan etiikasta ja se on saatavilla avoimesti verkosta.[45] Tieto- ja viestintätekniikan ammattilaiset Tivia taas ylläpitää Etiikan ohjeita -säännöstöä, joka tarjoaa perustan eettiselle teknologian kehitykselle.

Taulukko 3. Tekoälyn etiikkaan liittyviä ohjeistuksia ja raportteja

IEEE	Ethically Aligned Design, Version 2	2018
ACM	Code of Ethics	2018
AI Now	AI Now 2018 Report	2018
AI Now	Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability	2018
Future of Life Institute	Asilomar principles for beneficial AI	2017
Internet Society	Artificial Intelligence and Machine Learning: Policy Paper	2017
Montreal Declaration	Montreal Declaration for a Responsible Development of Artificial Intelligence	2017
Tivia	Etiikan ohjeet	2002
World Economic Forum	How to Prevent Discriminatory Outcomes in Machine Learning	2018

[44] Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems.

[45] Heikkerö, T. (2009). Tekniikka ja etiikka: Johdatus teoriaan ja käytäntöön. Tekniikan akateemisten liitto.

Akateemisista instituutioista ja tutkimusprojekteista mainittavia ovat erityisesti AI Now Institute (New York University), Stanfordin yliopiston One Hundred Year Study on Artificial Intelligence, The Royal Society (UK) ja Future of Humanity Institute (Oxford). New Yorkissa toimiva AI Now Institute on julkaissut useita raportteja tekoälyn etiikasta yleisellä tasolla sekä tietyistä erityisalueista (algoritmien vaikutusarviointi, julkisorganisaatioiden algoritmit). Oxfordin FHI taas on tunnettu pitkälle kehittyneiden tekoälyn muotojen tutkimuksesta sekä tekoälyteknologioihin kohdistuvan kansainvälisen hallinnon kehittämisestä. Yksityisiä tutkimusorganisaatioita taas edustavat Future of Life Institute, OpenAI ja Machine Intelligence Research Institute.

Myös julkisen hallinnon organisaatiot ovat julkistaneet tekoälyä ja tekoälyn etiikkaa käsitteleviä raporttejaan viimeisen muutaman vuoden aikana. Huomattavimpina valtioina ja julkisyhteisöinä mainittakoon Yhdysvallat, Euroopan Unioni, Iso-Britannia ja Ranska. Suomessa tekoälyyn liittyvä kansallinen tekoälyohjelma Tekoälyaika on vielä käynnissä kevääseen 2019 asti. Euroopassa on julkisessa keskustelussa ja strategisena linjana korostettu erityisesti eettisen tekoälyn varmistamista. Euroopan Unionin tekoälyä koskeva eettinen ohjeistus julkaistaan huhtikuussa 2019.[46]

Käytännön etiikan arvioinnin malleja ja työkaluja

Kehittäjän näkökulmasta innovaation eettisten ongelmien ja etujen arviointi voi vaikuttaa hitaalta ja hankalalta asialta. Saatavilla on kuitenkin useita työkaluja tekoälyn etiikan tarkasteluun. Työkalut vaihtelevat algoritmien ja koneoppimissovellusten eettisyyden varmistamisen viitekehyksistä design-prosessien tueksi suunnitteluihin etiikkapaketteihin. Osa listatuista paketeista on seikkaperäisiä ohjeita siitä, millä tavoin tekoälyratkaisujen eettisyyttä tulisi tarkastella.
AI Now -instituutin tarjoama malli algoritminen vaikutusten arviointiin on mainio aloituspiste. Se tarjoaa rakennetta ja taustatietoa siihen, miten algoritmien vaikutusta tulisi eri tilanteissa arvioida. Sen lähtökohtana on julkishallinnon algoritmien tarkastelu ja se soveltuu myös muun algoritmisen päätöksenteon arviointiin.

Matemaatikko Cathy O'Neilin The Ethical Matrix auttaa myös algoritmien päätösten vaikutusten arvioinnissa. Se ottaa huomioon eri sidosryhmät ja niille aiheutuvat erilaiset vaikutukset: matriisi auttaa kokonaisvaikutuksen ja ongelmakohtien havainnoinnissa sekä korjaamisessa. Käytännön tekoälykehitystyöhön löytyy erilaisia suunnittelutyökaluja- ja kehyksiä kokonaisuuden hahmottamisen vahvistamiseksi. Ethics Kit tarjoaa valmiiksi mietittyjä kysymyksenasetteluja ja suunnittelukehyksen niiden huomioonottamiseksi. Kitistä löytyy apua tekoälyprojektien eri vaiheisiin, niin alkupalaverista datanhallintaan ja monimuotoisen kehittäjätiimin muodostamiseen.

Ethics Canvasin on tarkoitettu erilaisten teknisten innovaatioiden eettisten vaikutusten arviointiin, joka muistuttaa liiketoimintamallin suunnitteluun tarkoitettua työkalua, jonka nimi on englanniksi "Business Model Canvas". Sen kehittämisessä on pyritty huomioimaan kehitystyön nopea sykli ja mahdollistamaan keskustelu eettisistä kysymyksistä ja hyvistä ratkaisuista kehitykseen osallistuvien henkilöiden välillä suunnittelutyön aikana. Sen avulla määritellään, keihin innovaatio vaikuttaa, millaisia vaikutuksia innovaatiolla voi olla ja miten eettisesti kyseenalaisiin vaikutuksiin voidaan vastata.

[46] Draft Ethics guidelines for trustworthy AI https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

Suurista teknologiayhtiöistä IBM on panostanut erityisesti algoritmisen päätöksenteon vinouma-ongelman ratkaisuun.[47] IBM AI Fairness 360 on avoimena lähdekoodina julkaistu kokoelma erilaisia metriikoita ja algoritmeja, joilla voi paikantaa, raportoida ja korjata syrjiviä tai tuloksia vääräntäviä vinoumia koneoppimismalleista.

Vinkkejä organisaatioille tekoälyn etiikan huomioon ottamiseen

Organisaatioissa tekoälyn eettisesti relevantit teemat tulee muistaa pitää kehitysprosesseissa mukana kaikilla eri organisaatiotasoilla: strategisen ohjauksen ja johdon, suunnittelun ja designin sekä käytännön toteutus- ja ylläpitotyön tasolla. Vaikka keskustelu tekoälyn etiikan ympärillä käy kiivaana, on tärkeimmistä suuntaviivoista löytymässä yhteistä ymmärrystä ja niitä voi käyttää käytännön toiminnan ohjaamisessa.

Strategisen ohjauksen tasolla vastuullisen tekoälyn kehittämisessä tärkeä lähtöpiste on määritellä organisaatiokohtaisesti eettiset periaatteet ja säännöt tekoälyn kehittämiselle ja soveltamiselle. Periaatteita luodessa on hyvä käydä keskustelua organisaation eri tasojen ja henkilöstöryhmien kanssa ja hyödyntää erilaisia osaamisia. Periaatteista löytyy esimerkkejä muun muassa Työ- ja elinkeinoministeriön Etiikkahaasteen sivuilta.[48]

Tekoälyn eettiset periaatteet on kuitenkin hyvä nähdä vain ensiaskeleena: eettisten periaatteiden toteutumista tulee seurata, ja niissä pysymiseen ja niiden vaatiman toiminnan varmistamiseen tulee varata tarpeeksi resursseja. Mahdollisena jatkotoimenpiteenä eettisen kehitys- ja soveltamistyön seurantaa ja valvontaa voi delegoida organisaation eri henkilöille. Johdon, kehittäjien ja suunnittelijoiden välinen dialogi aiheesta on tärkeää.

Tekoälyn teemat on olennaista pitää mukana kaikissa tekoälyteknologioita hyödyntävissä kehitysprojekteissa, niin suunnittelun kuin toteutuksen osalta. Suunnittelussa voi tukeutua eettisen designin menetelmiin ja apuvälineisiin, jotka auttavat eettisten haasteiden huomaamisessa ja ratkaisemissa. Eettinen suunnittelu voi suoraviivaistaa myös toteutustyötä: jos mahdollisia eettisiä kysymyksiä on työstetty suunnitelmallisesti ja ratkaisuhakuisesti jo suunnittelupöydällä, on toteutustyössä helpompi seurata sovittuja eettisiä linjauksia. Tämä antaa toteuttajille myös selkeyttä tilanteisiin, joissa he havaitsevat esimerkiksi kehitysprojektin tulosten olevan eettisesti arveluttavia – kun etiikasta on keskusteltu ja siitä on olemassa linjauksia, on kehittäjän helpompi nostaa varoituslippu ylös eettisten ongelmien ilmentyessä.

Keskusteleva lähestymistapa on etiikassa keskeistä, ja näin tulisi olla myös organisaatioiden sisäisissä etiikan haasteiden ratkaisuissa. Eri näkökulmat voivat täydentää toisiaan, ja teknisesti monimutkaisten tekoälyratkaisujen läpikäynti laajemmassa joukossa lisää itsessään ratkaisujen läpinäkyvyyttä ja ymmärrettävyyttä. Etiikan osalta avointa asennetta kannattaa suosia myös kaikkien sidosryhmien suuntaan

[47] IBM launches tool aimed at detecting AI bias. Zoe Kleinman. BBC News, 19.9.2018. https://www.bbc.com/news/technology-45561955

[48] Työ- ja elinkeinoministeriön Tekoälyaika-hankkeen Etiikkahaaste. https://www.tekoalyaika.fi/mista-on-kyse/etiikka/

Tilannekuva tekoälyn etiikasta käytännössä

Tilannekuva yhteiskunnassa ja yrityksissä

Tekoälystä on käyty runsaasti julkista keskustelua eri foorumeilla ja mediassa viime vuosien aikana. Keskustelu on ymmärrettävästi hakenut suuntaansa: on yritetty ymmärtää mitä tekoäly on, mitä hyötyä siitä voisi olla, sekä minkälaisia riskejä se sisältää. Ymmärryksen lisääntyessä keskustelu on syventynyt ja pureutunut oikeisiin ongelmiin.

Tekoälyn etiikan osalta keskustelua ovat leimanneet liiaksi dystooppiset skenaariot. Nämä teoreettiset rakennelmat käsittelevät kieltämättä filosofisesti ja yhteiskunnallisesti tärkeitä asioita, mutta suurin osa kysymyksistä siintää vasta kaukana tulevaisuudessa. Suurin osa tekoälyn etiikan haasteista koskee kuitenkin nykyhetkeä, ja ovat itsessään jo sen verran hankalia, että ansaitsevat jakamattoman huomiomme.

On hyvä ymmärtää, että tekoälyn soveltaminen on käytännössä vasta alkamassa. Samalla hahmottuu tekoälyn ja tekoälyn etiikan kiinnittyminen ajallisesti menneeseen ja tulevaan: Ensinnäkin tekoälyn kehitys linkittyy vahvasti pidempiaikaisiin muutostrendeihin, kuten digitalisaation ja automaatioon. Toisekseen nojautuminen jo ennakolta tulevaisuuteen on tarpeellista: monet tekoälyn etiikan kysymykset tarvitsevat keskustelua ja tutkimusta jo nyt, vaikka teknologian soveltaminen seuraisikin perässä vasta viiden tai kymmenen vuoden päästä.

Monesta tekoälyn etiikan ongelmasta löytyy jo käytännön esimerkkejä. Rasistisesta Tay-botista on muutama vuosi aikaa, mutta samat vinouma-ongelmat ovat vaivanneet niin mobiililaitteiden tunnistuskameroita[49] kuin suurten yhtiöiden rekrytointijärjestelmiä[50]. Turvallisuus- ja vastuuhaasteisiin on herätty autonomisten ajoneuvojen onnettomuuksien myötä. Yksityisyydensuojasta on huolestuttu erityisesti lisääntyvän valvonnan näkökulmasta: Kiinan suorasanaiset

[49] Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency (pp. 77-91). http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

[50] Reuters: Amazon scraps secret AI recruiting tool that showed bias against women https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

suunnitelmat kansalaistensa seuraamisesta ja pisteyttämisestä tekoälyjärjestelmien avulla on tuonut dystopiatarinoita lähemmäs todellisuutta, mutta samankaltainen kehityskulku näyttää hivuttautuvan kaikkialle maailmaan.[51]

Tekoälyn etiikan keskustelussa aikajänteen valinta on hankalaa. Suuri osa teknologiaratkaisuista voi näyttäytyä neutraaleilta suunnittelupöydällä, ja tulevaisuuden skenaarioista on helppo keskustella kahvihuoneessa. Trendinomaiset, keskipitkän tähtäimen muutokset yhteiskunnassamme ovat monimutkaisia ja alttiita joko liialliselle teknologiaoptimismille tai ”ennen kaikki oli paremmin” – ajattelulle. Keskustelua ei helpota se, että monet tärkeät tekoälyläpimurrot tapahtuvat maailmalla huippututkimus-keskuksissa tai -yrityksissä: vaikka suomalaisyrityksissä tekoälyn tuleminen on lapsenkengissä, se ei tarkoita, etteikö tekoälyn etiikassa tulisi ottaa huomioon myös esimerkiksi DeepMindin ylivoimaa tieteellisessä tutkimuksessa, Boston Roboticsin robottien edelläkävijyyttä taikka Facebookin ja Googlen valtavaa vaikutusvaltaa.

Tekoälyyn liittyvät kysymykset nousevat yrityksissä ja julkisorganisaatioissa keskusteluun askel askeleelta. Suomessa työ- ja elinkeinoministeriön tekoälyn etiikan alatyöryhmä on tehnyt tärkeää työtä ja kannustanut yrityksiä tekoälyn eettisten periaatteiden määrittelyyn. TEM:n työryhmä on haastanut suomalaisia yrityksiä sitoutumaan eettisen tekoälyn kehittämiseen, ja haaste jatkuu keväällä 2019.[52] Kyseessä on ensimmäinen askel sitoutumisessa eettisen tekoälyn kehitykseen: samaa tematiikkaa tulisi myös konkreettisesti nostaa esille koulutuksen, toimintatapojen muutoksen ja kuluttajien mukaan sitouttamisen kautta. Muutosta voi olla luvassa: tekoälyn etiikka on mainittu Gartnerin vuoden 2019 trendilistauksessa[53], ja tekoälyeetikot yhtenä tärkeimpänä rekrytointikohteena tekoälyyrityksille konsulttiyhtiö KPMG:n toimesta.[54]

Tekoälyn etiikan teemojen ja eteneminen tarvitsee panoksensa myös kuluttajilta: monet eettiset kysymykset ja aiheet ovat vaarassa hautautua juhlapuheisiin ja komiteamietintöihin, mikäli kuluttajat ja käyttäjät eivät vaadi käyttämiltään palveluilta, yrityksiltä ja julkiselta sektorilta huolehtimista ja tarkkaa otetta eettisten kysymysten osalta. Esimerkiksi tietoturvan, datankäytön ja algoritmisen päätöksenteon eettisyyden vaatimusten soisi voimistuvan. Eettisestä käytännöstä on olemassa paljon ohjeistuksia, mutta kuluttajien paine ohjaa palveluntarjoajia käytännön sitoutumiseen.

Yhteiskunnan tasolla varautumista suuriin muutoksiin ei voi korostaa liikaa. Tekoälyteknologiat tulevat vaikuttamaan työelämäämme ja koulutuksen tarpeeseemme - tämä on asia, josta vallitsee jo laaja yhteisymmärrys. Suomen perustulokokeilu on ideatasolla hyvä aloite ratkaisuvaihtoehtojen kartoittamisessa. Koulutus- ja tutkimuspolitiikan voi sen sijaan nähdä olevan liian poukkoilevaa, jos Akavan ja muiden järjestöjen kannanottoa on uskomista.[55] Toivoaksemme tämä opas nostaa esille tarpeen keskustelulle myös siitä, mihin suuntaan haluamme ihmiselämämme ja inhimillisyyden perustavimman luonteen muuttuvan.

[51] The New York Times: Warning! Everything Is Going Deep: ‘The Age of Surveillance Capitalism’ https://www.nytimes.com/2019/01/29/opinion/artificial-intelligence-surveillance.html

[52] TEM Etiikkahaaste 2018

[53] Gartner Top 10 Strategic Technology Trends for 2019 https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019/

[54] KPMG: Top 5 AI hires companies need to succeed in 2019 https://info.kpmg.us/news-perspectives/technology-innovation/top-5-ai-hires-companies-need-to-succeed-in-2019.html

[55] Akava: Kannanotto: Suomen menestys riippuu osaamisesta – seuraavalla hallituksella käsissään kohtalon kortit https://www.akava.fi/uutishuone/teemajutut/suomen_menestys_riippuu_osaamisesta

Ehdotetut toimenpiteet

Tekoäly ansaitsee voimakasta panostusta tutkimukseen ja koulutukseen. Olemme eläneet vauhdikasta teknologisen kehityksen aikaa jo viimeiset 10 vuotta ja vaikuttaa siltä, ettei vauhti ole hidastumassa. Tutkimuksessa tulee ymmärtää jo olemassa olevan tutkimustraditiomme vahvuudet, mutta olla rohkeasti havainnoimassa ja analysoimassa uusia suuntia. Teknologian tutkimuksen lisäksi on välttämätöntä panostaa myös tekoälyn yhteiskunnallisten vaikutusten ymmärtämiseen.

Tekoälyteknologiat antavat uusia mahdollisuuksia myös tutkimusten toteuttamiselle. Teknologioiden avulla voimme tutkia monimutkaisempia kokonaisuuksia ja syy-seuraus-suhteita kuin tähän asti. Avaimena tämän mahdollisuuden hyödyntämiseen on monitieteellinen lähestymistapa. Tekoäly-teknologiat voivat antaa paljon uutta esimerkiksi yhteiskunnalliselle tutkimukselle, mutta tekoälyn ammattilaisilta tarvitaan tähän tukea. Toisaalta teknologia-alan tulee ymmärtää kasvanut vastuunsa ja tehdä avoimesti yhteistyötä muiden tieteilijöiden kanssa.

Tiedeyhteisö on perustellusti huolissaan pitkäjänteisen rahoituksen puutteesta. Moni tekoälyn etiikkaan liittyvä teema vaatii laajaa ja kärsivällistä tutkimusta onnistuakseen - ihan kuten tekninen tekoälytutkimus itsessäänkin. Oppaan kirjoittajien harras toive on, että pitkäjänteiseen työskentelyn mahdollistavia rahoitusinstrumentteja kehitetään ja tuetaan.

Valtion tulisikin investoida pitkäjänteiseen tekoälytutkimukseen. Perustutkimuksen ohella tärkeää on panostaa tekoälyn turvallisuuteen sekä eettisiin ja yhteiskunnallisiin vaikutuksiin liittyvään tutkimukseen, jota yksityinen sektori ei samalla tavoin rahoita. Suomen on tuettava vastuullista tekoälyn kehitystä ja panostettava innovatiivisten tekoälysovellusten käyttöönottoon yhteiskunnassa. Tämä vaatii myös varautumaan työmarkkinoiden mullistukseen työvoiman tarvittavalla uudelleenkoulutuksella.

Lainsäädännössä tekoälyn kehitykseen on kiinnitettävä huomiota. Selkeitä alueita ovat esimerkiksi autonomisten kulkuneuvojen juridiset kysymykset, vastuu tekoälyjärjestelmien tekemistä päätöksistä sekä datan käyttöön liittyvän suostumuksen validiointi. Kansallisesti on varmistettava syrjimätön ja osallistava tekoälykehitys, joka ei uhkaa demokratiaamme. Erityisen tärkeää sääntelykysymysten selvittäminen on yhteiskunnallisesti keskeisillä alueilla, kuten työ-, rahoitus-, koulutus- ja terveyspalveluissa, jotta vältytään syrjinnältä, yksityisyyden loukkauksilta ja turvallisuusuhilta. Sääntelyä ei tule ajatella rajoitteena liiketoiminnalle vaan tapana edistää tasapuolista ja kestävää kilpailua, samalla, kun se suojelee yhteiskuntaa epätoivotuilta tekoälyn kehityssuunnilta.

Lainsäädännöllinen huomio ei välttämättä tarkoita merkittäviä lisäyksiä lainsäädäntöön, mutta sääntelyn tarkoituksen tosiasiallisesta toteutumisesta ja ohjaavuudesta tulee huolehtia. “Järkevän sääntelyn” (smart policies)[56] näkökulma painottaa tätä ajatusta ja on kiinnostunut uuden lainsäädännön lisäämisen sijaan siitä, että nykyinen, olemassa oleva sääntely toimisi entistä tehokkaammin ja ennakoitavammin. Toisaalta myös uusi, alakohtainen sääntely voi tulla kyseeseen, jotta eettisten periaatteiden noudattaminen ja turvallinen, osallistava tekoälykehitys varmistetaan. Tärkeää on myös kansainvälinen yhteistyö sääntelyn yhtenäistämiseksi, jotta vältetään kilpavarustelukierre ja ylläpidetään reilu kilpailuasetelma. Kaikki nämä tavoitteet myös vaativat tekoälyasiantuntijoiden rekrytointia valtionhal-

[56] Brundage, M., & Bryson, J. (2016). Smart Policies for Artificial Intelligence. arXiv preprint arXiv:1608.08196.

lintoon.

Yrityksillä on ensisijaisina tekoälyn kehittäjinä vastuu toimia avoimesti ja läpinäkyvästi eettisten periaatteiden mukaisesti syrjimättömien, turvallisten ja yksityisyyttä kunnioittavien tekoälysovellusten luomiseksi. Yritysten on varmistettava, että tekoäly näyttäytyy kuluttajille ymmärrettävänä. Tämä voidaan ideaalitilanteessa toteuttaa auditoimalla algoritmista prosessia teknisellä tasolla (esimerkiksi eksplikoimalla, kuinka algoritmi on päätynyt tiettyyn ratkaisuun) tai kontrafaktuaalisten selitysten avulla[57], jotka ilmaisevat käyttäjälle, mitkä ovat ne tekijät, joiden olisi tullut olla toisin, jotta myös algoritmisen prosessin lopputulema olisi ollut toisenlaisen. Eettiset periaatteet toteutuvat ainoastaan, jos ne sisällytetään keskeiseksi osaksi tekoälykehitystä, ja yritykset näkevät itsensä vastuullisina kansalaisille ja valvovat periaatteiden toteutumista.[58] Tämä tarkoittaa sääntelyn noudattamista, mutta myös oma-aloitteista yritysvastuun omaksumista vastuullisesta tekoälykehityksestä. Yritysten on kehitettävä sisäisiä vastuumallejaan, panostettava sisäiseen valvontaan ja sovellusten eettisyyteen. Varsinkin monimuotoisuuden varmistaminen on keskeistä tekoälyn kehityksessä. Myös datatietiikka ja käyttäjien suostumus tietojensa käyttöön on varmistettava heti kehitystyön alusta. Toivottavaa on myös, että yritykset tekevät yhteistyötä valtiovallan kanssa esimerkiksi erilaisten tekoälyn luotettavuuden ja läpinäkyvyyden mittareiden kehittämisessä. Tämän kehitystyön perusteella luodut sertifikaatit lisäisivät myös kuluttajien tietoisuutta ja luottamusta tekoälyyn.[59]

Osaan teknisistä haasteista on vaikea vastata käytännössä. Esimerkiksi algoritmista diskriminaatiota voi olla vaikea välttää tilanteissa, joissa lähdeaineiston vinoumat johtuvat yhteiskunnassa esiintyvästä epätasa-arvosta. Tällöin voi olla perusteltua hyödyntää algoritmista päätöksentekoa vain yhteiskunnassa tai organisaatiossa ilmenevien syrjivien rakenteiden havaitsemisessa sen sijaan, että algoritmin annettaisiin ohjata lopullisten päätösten tekemistä. Algoritmeja voidaan kuitenkin parantaa esimerkiksi IBM:n tarjoaman työkalun avulla, jolla voidaan havaita algoritmin mahdolliset vinoumat.[60]

Kansalaisyhteiskunnalla on olennainen rooli ylläpitää keskustelua oikeudenmukaisesta tekoälystä, johon myös valtiovallan ja yritysten on osallistuttava dialogisesti. Keskiössä ovat kysymykset siitä, millaista elämää arvostamme ja mitä haluamme jättää algoritmien päätettäväksi. Järjestöjen tehtävänä on varmistaa ja vaikuttaa, että tekoälyä kehitetään edistäen yleistä, yhteiskunnallista hyvää. Organisaatioiden on vaikutettava eettisen tekoälyn puolesta sekä kansallisesti että kansainvälisesti, esimerkiksi yhteisöjen, kuten The European AI Alliancen, kautta. Kansalaisyhteiskunnan rooli on siis toimia vallan vahtikoirana, erityisesti epätoivottavia sovellutuksia, kuten massavalvontaa ja autonomisia aseita vastaan. Tärkeää on yleisen tietoisuuden lisääminen tekoälyn eettisistä, oikeudellisista ja yhteiskunnallisista vaikutuksista. Käytännössä tämä voi tarkoittaa esimerkiksi osallistavaa koulutusta, digitaalisten taitojen edistämistä ja valistusta liittyen tietoturvaan sekä yksityisyyteen. Parhaat tulokset kuitenkin saavutetaan, kun kaikki edellä nimetyt tahot toimivat yhteistyössä tässä valistustyössä yhteiskunnallisesti hyödyllisen tekoälyn puolesta.

[57] Wachter ym. 2017.

[58] AI Now 2018, 9.

[59] Floridi et al. 2018, 703.

[60] Kleinman 2018.

Lähteet

Kirjallisuuslähteet

Asemah, E.S., Okpanachi, R.A. & Edegoh, L.E.O. (2013). Business Advantages of Corporate Social Responsibility Practice: A Critical Review. *New Media and Mass Communication*, 18: 45-54.

Baldridge, J. (2015). Machine Learning and Human Bias: An Uneasy Pair. [Verkkoartikkeli] *TechCrunch*. <http://social.techcrunch.com/2015/08/02/machine-learning-and-human-bias-an-uneasy-pair.> Viitattu 25.3.2019.

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2017). Fairness in criminal justice risk assessments: the state of the art. [arXiv preprint] arXiv:1703.09207.

Brundage, M., & Bryson, J. (2016). Smart Policies for Artificial Intelligence. [arXiv preprint] arXiv:1608.08196.

Chander, A. (2016). The racist algorithm. *Mich. L. Rev.*, 115: 1023.

Cook, D. M., Szewczyk, P., Sansurooah, K. (2011). Securing the Elderly: A Developmental Approach to Hypermedia Based Online Information Security for Senior Novice Computer Users. P*roceedings of the 2nd International Cyber Resilience Conference, Edith Cowan University, Perth Western Australia, 1st - 2nd August 2011.*

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, *2015*(1): 92-112.

Fieser, J. Ethics. [Verkkoartikkeli] *Internet of Encyclopedia of Philosophy*. <https://www.iep.utm.edu/ethics/>. Viitattu 13.10.2018.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.

Furman, J. & Seamans, R. (2018). AI and the Economy. *Innovation Policy and the Economy, 19*(1): 161–191.

Heikkerö, T. (2009). *Tekniikka ja etiikka: Johdatus teoriaan ja käytäntöön*. Tekniikan akateemisten liitto.

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No More FOMO: Limiting Social Media Decreases Loneliness and Depression. *Journal of Social and Clinical Psychology*, 37(10): 751-768.

Kelleher, J. D., & Tierney, B. (2018, 9). *Data Science*. Cambridge: MIT Press.

Koski, O., & Husso, K. (2018). Tekoälyajan työ: neljä näkökulmaa talouteen, työllisyyteen, osaamiseen ja etiikkaan. *Työ- ja elinkeinoministeriön julkaisuja* 2018(19). <http://urn.fi/URN:ISBN:978-952-327-311-5>.

Kugler, L. (2018). AI Judges and Juries. *Communications of the ACM*, 61(12): 19-21.

Mccoy, J. & Rahman, T. & Somer, M. (2018). Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities. *American Behavioral Scientist*, 62: 16-42.

Osoba O. & Welser IV W. (2017). The Risks of Artificial Intelligence to Security and the Future of Work. Santa Monica, CA: Rand Corporation. <https://www.rand.org/pubs/perspectives/PE237.html> Viitattu 25.3.2019.

Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, 36(4): 105-114.

Sayre-McCord, G. (2014). Metaethics. *The Stanford Encyclopedia of Philosophy*. Toim. Edward Zalta. <https://plato.stanford.edu/archives/sum2014/entries/metaethics/>. Viitattu 25.3.2019.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(3): 417-424.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, *361*(6404): 751-752.

Turner, J. (2019). *Robot Rules: Regulating Artificial Intelligence*. Palgrave Macmillan.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 2018.

Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180085.

Muut lähteet

AccessNow (2018) *Mapping artificial intelligence strategies in Europe.*
https://www.accessnow.org/mapping-artificial-intelligence-strategies-in-europe/

AI Now Institute, New York University (2018) *AI Now 2018 Report.*
https://ainowinstitute.org/AI_Now_2018_Report.html

BBC News, Zoe Kleinman (2018) *IBM launches tool aimed at detecting AI bias.* https://www.bbc.com/news/technology-45561955

DigitalTrends.com (2019) *What is Google Duplex? The smartest chatbot ever, explained.* https://www.digitaltrends.com/home/what-is-google-duplex/.

Euroopan unioni (2018) *Draft Ethics guidelines for trustworthy AI.* https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

Euroopan unioni, Euroopan komissio (2018) *European Commission Staff Working Document: Liability for emerging digital technologies.* https://ec.europa.eu/digital-single-market/en/news/european-commission-staff-working-document-liability-emerging-digital-technologies

Euroopan unioni. Euroopan parlamentin ja neuvoston asetus. (GDPR). https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex%3A32016R0679

Euroopan unioni, EUR-Lex. *Euroopan ihmisoikeussopimus.*
https://eur-lex.europa.eu/summary/glossary/eu_human_rights_convention.html?locale=fi

Fatml.org. *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.* http://www.fatml.org/resources/principles-for-accountable-algorithms

Future of Life Institute. *Summary of 12 Aftermath Scenarios.* https://futureoflife.org/ai-aftermath-scenarios/

ITS Finland. *ITS sanasto.* http://www.its-finland.fi/index.php/fi/mita-on-its/its-sanasto.html

McKinsey Global Institute (2018). *Delivering through diversity.* https://www.mckinsey.com/business-functions/organization/our-insights/delivering-through-diversity

McKinsey Global Institute (2017). *Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages.* https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages

MyData Finland. Verkkosivusto. https://mydata.org/

The Royal Society & The British Academy (2018) *The impact of artificial intelligence on work. An evidence synthesis on implications for individuals, communities, and societies.* https://www.britac.ac.uk/projects/ai-and-work

The Royal Society for Public Health, The Young Health Movement (2017) *Royal Society for Public Health: #StatusOfMind - Social media and young people's mental health and wellbeing.* https://www.rsph.org.uk/uploads/assets/uploaded/62be270a-a55f-4719-ad668c2ec7a74c2a.pdf

TIVIA - Tieto- ja viestintätekniikan ammattilaiset TIVIA ry (2002) *Etiikan ohjeet.* http://www.tivia.fi/julkaisut/etiikan-ohjeet

Työ- ja elinkeinoministeriö, Tekoälyaika-hanke (2018) *Etiikkahaaste - Yritykset mukaan tekoälyn eettiseen hyödyntämiseen.* https://www.tekoalyaika.fi/etiikka

Yhdistyneet kansakunnat. *Ihmisoikeussopimukset.* https://www.yk.fi/node/257