

TEKOÄLYTEKNOLOGIAT JA TURVALLISUUS

JAANA HALLAMAA, RAINE HAIKARAINEN JA TAINA KALLIOKOSKI

Tekoälyjärjestelmät ovat jo osa arkielämää. Ne muuttavat ihmistoimijan roolia, vaikka keskeiset fyysisen todellisuuden tekijät ja biologiset perustarpeet säilyvät entisellään. Digitaaliset järjestelmät parantavat toimintojen turvallisuutta, mutta synnyttävät uudenlaisia riskejä. Näihin kysymyksiin perehdytään Strategisen tutkimuksen neuvoston ETAIROS-hankkeessa.

”**A**lypelin kuningatar” otsikoi Helen-energiailaitoksen asiakaslehti kansisivulla pääjuttunsa vuoden 2021 ensimmäisessä numerossa. Vaikka kuvituksena on shakkipelin kuningatarnappula, yhdeksänsivuisessa kokonaisuudessa ei puhuta lautapelistä vaan tekoälystä.¹ Lukijalle selviää, ettei tekoälyssä ole kysymys älypelistä, eikä se ole minkään valtakunnan kuningatar. Vetävät metaforat houkuttelevat energia-yhtiön asiakkaat lukemaan kansantajuistettua katsausta tekoälysovelluksista ja niiden kehittämispyrkimyksistä, mutta ne jättävät pimentoon mutkikkaat tekoälyn järjestelmäturvallisuutta koskevat kysymykset.

Otsikoinnista huolimatta jutussa ei hahmotella transhumanistista tekoälyutopiaa vaan esitellään näkyviä, jotka vastaavat käsitystä erityisiin tai kapeisiin sovelluksiin kykenevästä tekoälystä (*Artificial Narrow Intelligence*). Vaikka suurelle yleisölle hahmotellaan mieluummin, kuinka robotit korvaavat tulevaisuudessa ihmistyön, skenaarioissa pohditaan vain harvoin, mitä ihmistä muistuttavan monenlaisista tehtävistä selviävän tekoälyn (*Artificial General Intelligence*, AGI) laajamittainen käyttöönotto tarkoittaisi puhumattakaan siitä, että arviotaisiin useimmista tehtävistä ihmistä paremmin selviävän superteškoälyn (*Artificial Super Intelligence*) vaikutuksia.² Seuraavassa pohdimme meneillään olevaa siirtymää tekoälyteknologioiden käyttöön ihmisen ja tekniikan välisen suhteen kannalta ja kysymme, kuinka se vaikuttaa käsitykseemme turvallisesta teknologiasta.

Digitaaliset sovellukset ovat jo osa arkipäivää

Energiayhtiö ei ole ainoa, jonka on selvitettävä palveluidensa käyttäjille, mihin tehtäviin ja kuinka digitaalisuuteen perustuvia älykkäitä laitteita ja ohjelmistoja hyödynnetään, vaan sama tarve koskee paitsi laajaa joukkoa kaupallisia toimijoita myös lähes kaikkia julkisten palveluiden edustajia. Yhteisenä toiveena on käyttää uutta teknologiaa parantamaan palveluita toiminnasta kerättävää dataa hyödyntämällä. Helen-yhtiö kertoo esimerkkinä data-analytiikkaan tukeutumisesta, kuinka Hel-

singin viemäreihin laskettuun jäteveteen sitoutunut lämpöenergia saadaan optimaalisesti talteen.³

Toisin kuin monissa tekoälyä kansantajuisesti käsittelevissä kirjoituksissa Helenin jutussa tekoälyä ei tarkastella kuvaamalla akrobaattisia tempuja tekeviä robotteja ja itseään ajavia ajoneuvoja eikä uumoilla, milloin valta siirtyy ihmiseltä tekoälylle, vaan esitellään digitaalisuuteen perustuvaa älykystä – toisin sanoen oman toimintansa tuottamaa dataa toimintansa säätelyyn ja optimoimiseen hyödyntävää teknologiaa. Tekoälyjärjestelmät ovat nykyään lähinnä mutkikkaita laskevia systeemejä ja suunnittelijoiden laatimaa algoritmia toteuttavia koneita. Osaa sovelluksista voidaan pitää oppivina järjestelminä, mutta siihen, että niistä tulisi täysin autonomisia toimijoita ihmisten rinnalle, on vielä matkaa.⁴

Asiakaslehden kirjoitus edustaa kansantajuistavaa journalismia parhaimmillaan. Tiiviissä juttukokonaisuudessa tarjotaan realistinen ja esimerkein ymmärrettäväksi tehty tietopaketti ihmisten arkeen jo nyt vaikuttavista digitaalisista sovelluksista. Lisäksi keksintöjä luonnehditaan suhteessa aiempiin edistysaskeliin. Samalla tuodaan esiin, että tekoälyjärjestelmät ovat – tosin termiä mainitsematta – sosio-tekniisiä systeemejä, joissa ihmisen ja teknisten sovellusten toiminta punoutuvat toisiinsa.⁵

Tekoälyjärjestelmät muuttavat ihmistoimijan roolia

Tekniset keksinnöt laajentavat ihmisen toimintakeinoja. Ihminen hyödyntää luonnon mahdollisuuksia omiin tarpeisiinsa soveltamalla luonnontieteiden tuloksia aineellisten tuotteiden valmistukseen, tehtävien suorittamiseen ja erilaisten tarpeiden tyydyttämiseen. Tekniikalla on apuvälineen rooli, ja ihanteellinen apuväline palvelee tarkoitustaan täsmällisesti. Erilaiset tekniikat muodostavat kuitenkin nykyään laajoja järjestelmiä, teknologisia systeemejä, mikä vaikeuttaa niiden tarkastelemista yksinkertaisina apuvälineinä. Tekoälysovellukset korostavat vaikeutta, koska ne ovat myös käyttöjärjestelmiä, sovelluksia ja palveluita, jotka ihmisen on integroitava osaksi toimintaansa. Ne muuttavat inhimillistä todellisuutta vielä enemmän kuin pe-

1 Helen 1/2021, 1.

2 Kuusi & Heinonen 2020, 16, 24–25.

3 Kelola 2021, 11–19.

4 Kelola 2021, 12–15.

5 Kelola 2021, 12–13.

rinteiset mekaaniset laitteet.⁶

Tekniikka on syrjäyttänyt ihmistoimijan teollistumisesta alkaen yhä useammissa toistoa, nopeutta, kestävyyttä, suorituksen yhdenmukaisuutta ja voimaa edellyttävissä tehtävissä. Kehitystä on pidetty ihmisen kannalta sekä kielteisenä että myönteisenä muutoksena. Automatisoituminen voidaan esitellä kehityskulkuna, joka tekee ihmisistä koneiden apulaisia ja muuttaa työnteon yksitoikkoiseksi linjastopuurtamiseksi niin, että inhimilliselle panokselle ei ole sijaa. Raskaiden, samana toistuvien ja vaarallisten työtehtävien jättäminen koneille tukee kuitenkin myös mahdollisuutta siihen, että ihmiset vapautuvat raskaista ja vaarallisista töistä ja voivat suunnata aikansa ja energiansa luovuutta ja kekseliäisyyttä edellyttäviin, inhimillisesti mielekkäämpiin tehtäviin.

Tekoälyteknologiat muuttavat ihmistoimijoiden roolia myös vaativissa, luovuutta ja harkintaa edellyttävissä tehtävissä, joiden uskomisen tekniikan hoidettavaksi on tähän asti vaikuttanut mahdollisuuksien rajoittaminen. Kykyä itsenäiseen ja luovaan päätöksentekoon eli autonomisuutta on pidetty erityisen inhimillisenä kykenä. Autonomisuus liitetään yhä useammin myös tekoölyyn sen keskeisenä ominaisuutena siitä huolimatta, että monet sääntöpohjaisen koodaamisen taitajat, koneoppimisalgoritmien ja neuroverkko-pohjaisen tekoälyteknologian ohjelmoijat korostavat, ettei tekoöllylle voida muotoilla kattavaa määritelmää. Ylätason luonnehdinnan mukaan tekoälyksi kutsutaan tietoteknologisia ohjelmia, jotka toimivat annettujen parametrien ja tehtävien mukaan autonomisesti ja adaptiivisesti.⁸

Tyypillisesti tekoölyn toimintaan kuuluu datan kerääminen ympäristöä havainnoimalla jonkin ihmisaistia vastaavan sensorin avulla tai soveluksen käyttöön annetusta digitaalisesta poolista. Esimerkiksi lentokoneiden ohjauslaitteisto havainnoi koneen liikettä, sijaintia, ulkoisia sääolosuhteita ja säilyttää ohjaussuunnan, tasaa kulkua ja keventää miehistön työkuormaa näiden datavirtojen mukaan. Oppiva tekoälyjärjestelmä pystyy dataa hyödyntämällä muuttamaan toimintaansa joko etukäteen koodattujen sääntöjen ohjaamana tai omaa toiminta-algoritmiaan mukauttamalla.

Jo nyt koneohjelmat keräävät, järjestävät, luokittelevat ja vertailevat eri alojen dataa ja hoitavat lähes kaikki laskutoimitukset. Kieltä on pidetty ihmisen ainutlaatuisena ominaisuutena samoin taiteellista luovuutta. On ajateltu, että vain ihminen voi laatia tekstejä ja kääntää niitä kielestä toiselle, säveltää ja luoda kuvataideteoksia. Tekoölyn ohjelmointikehitys osoittaa kuitenkin, että koneet voivat oppia paitsi tuottamaan tekstejä myös tulkitsemaan ja kääntämään niitä. Suurten tekstiasojen läpikäymistä edellyttävissä tehtävissä konepeittoon ihmisen vaivatta. Toisin kuin on kuviteltu, ei ole yksinkertaista tapaa erottaa koneen säveltämää hittiä ihmisen luovuuden tuottamasta musiikkikappaleesta.⁹

Digitaalisuus ja kehittyvä tekoäly muuttavat vauhdilla ihmisen työ- ja elinympäristöä ja -oloja, minkä vuoksi enää ei voida kysyä, kohtaako menemään oleva muutos sitä tai tätä ryhmää vaan kuinka se heihin vaikuttaa. Monilla aloilla, kuten ilmailussa, tietokoneet ohjaavat valtaosaa toiminnasta, jota ei voitaisi enää lainkaan hallita ilman niitä.

Kun turvalliseksi uskottu muuttuu vaaralliseksi

Ilmailussa mutta yhä laajemmin myös muussa liikenteessä ja sen ohjaamisessa käytetään älylaitteita, joilla päätöksentekoa siirretään algoritmisille järjestelmille nopeita päätöksiä vaativissa tilanteissa, joissa ratkaisujen kannalta relevantteja tekijöitä on suuri määrä. Ihmisen havainto- ja tietokyvyt eivät yksinkertaisesti riitä ottamaan vastaan, käymään läpi ja suhteuttamaan keskenään kaikkea nykyaikaisen lentokoneen kulloinkin vallitsevista olosuhteista välittämää dataa. Modernien koneiden aerodynaaminen tehokkuus perustuu sekin tietokoneohjattuun hallintaan. Onnettomuuksilta ei kuitenkaan ole vältytty.¹⁰

Syksyllä 2018 ja keväällä 2019 tapahtui kaksi liikenneilmailun historiaan jäävää lento-onnettomuutta, kun kaksi uudenuutukaista Boeing 737 MAX -matkustajakonetta tuhoutui lennolla. Molemmat onnettomuudet olivat kaikkien mukana olleiden kannalta tuhoisia – yhteensä 346 ihmistä kuoli. Onnettomuustutkinta paljasti, että tur-

6 Hallamaa ja Kalliokoski 2020.

7 Grace ym. 2018; Marr 2019.

8 Tekoölyn perusteet -kurssi. <https://course.elementsofai.com/fi/1/1>.

9 Needham 2017.

10 Ilmailun turvallisuusajattelun kehittämisestä, ks. Haikarainen 2021.

mien syynä oli anturivika, joka sekoitti konetyypin tietokoneisiin perustuvan ohjausjärjestelmän toiminnan.

Ohjaamomiehistö ei kyennyt hallitsemaan yllättäen ilmaantunutta vikaa, koska he eivät olleet selvillä järjestelmään konetyypin uudistamisen yhteydessä tehdyistä muutoksista. Koska he eivät kyenneet ymmärtämään, mistä ohjausongelmat johtuivat, heidän mahdollisuutensa reagoida niihin olivat olemattomat. Onnettomuuksien tapahtuttua lähes 400 aivan uutta Boeing 737 MAX-konetta määrättiin yli 20 kuukautta jatkuneeseen lentokielttoon, mikä oli ilmailuhistoriassa ainutlaatuinen toimi.

Onnettomuustutkinta, josta Yhdysvaltain kongressin edustajainhuoneen liikennekomitea julkaisi syyskuussa 2020 perusteellisen loppuraportin, paljasti, että yhden sähköisen komponentin viasta aiheutunut säätöjärjestelmän sekoaminen oli vain viimeinen tekijä onnettomuuksiin johtaneissa tapahtumaketjuissa.¹¹ Selvityksen mukaan koko Boeing-yhtiöön mutta myös viranomaisvalvontaan oli päässyt leviämään turvallisuuden kannalta vahingollisia toimintakulttuurin vinoutumia.

Raportoiijat kiinnittivät huomiota niin Boeingin insinöörien tekniikkaa koskeviin virheellisiin oletuksiin ja yhtiön johdon haluttomuuteen toimia avoimesti kuin viranomaisvalvonnan puutteisiin. Kiivas taistelu kansainvälisen liikenneilmailun markkinaosuuksista oli luonut tehtaalle paineen karsia kustannuksia ja nopeuttaa koneiden teknisiä tarkastusprosesseja. Koska lentäjiltä ei haluttu edellyttää kalliita lisäkoulutuskursseja, yhtiö ei esitellyt koneeseen tehtyjä järjestelmämuutoksia. Näin voitiin sanoa, että konetyyppi oli pysynyt entisenlaisena.¹² Onnettomuuksia edeltäneinä vuosina liikenneilmailu oli ollut poikkeuksellisen turvallista. Esimerkiksi vuonna 2017 suurten kansainvälisten reittilentoyhtiöiden koneet eivät joutuneet yhteenkään onnettomuuteen, vaikka matkustajalentokertoja kertyi peräti 4,1 miljardia.¹³

B737 MAX -onnettomuudet tekivät näkyväksi, että tuhoisat onnettomuudet ovat edelleen mahdollisia, ja palauttivat perinteiset teemat turvalli-

suuskeskusteluun. Selvitys korosti, että ultraturvallisena pidetty järjestelmä oli punonut teknisiä ja sosiaalisia tekijöitä toisiinsa tavalla, joka vahvisti onnettomuusriskiä. Tapahtunutta analysoineiden Phillip Johnstonin ja Rozi Harrisin mukaan onnettomuudet olivat traaginen esimerkki yritysmallista, jossa yrityksen uuden, markkinamenestystä korostaneen johdon visiot sivuuttivat suunnitelluinsinöörien näkemykset niin, että perustehtävään – toiminnaltaan luotettavien lentokoneiden suunnittelemiseen ja tuottamiseen – keskittyminen unohtui.¹⁴ Onnettomuus muistutti kovalla tavalla siitä, että turvallisuustekijät sivuuttavat ansaintamallit tuottavat tappiota ja että turvallisuus on sosio-teknistä järjestelmää kokonaisvaltaisesti koskeva suure, eikä sitä voida supistaa teknisten järjestelmien kehittämiseksi.

Tekoäly turvallisuuskriittisillä aloilla

Ilmailu on vain yksi turvallisuuskriittiseksi kutsutuista aloista, jotka hyödyntävät tekoälyteknologiaa keskeisissä toiminnoissaan. Turvallisuuskriittinen on suomennos termistä *safety critical*. Sillä tarkoitetaan toimintaa, jossa turvallisuudesta huolehtiminen on keskeistä toiminnan onnistumisen kannalta. Turvallisuuskriittisten organisaatioiden tavanomainen toiminta sisältää vaaroja ja uhkia, jotka voivat huonosti hallittuina aiheuttaa vahinkoa ihmisille tai ympäristölle.¹⁵ Turvallisuuskriittisiä aloja ovat muun muassa yhteiskuntien energiantuotanto, sairaanhoito ja meri- ja maaliikenne. B737 MAX koneiden kohtaamista onnettomuuksista on keskusteltu lähinnä osana ilmailun turvallisuutta. Niiden tarkasteleminen osana tekoälyn etiikasta käytyä keskustelua voisi kuitenkin olla hyödyllistä ja laajentaa näkökulmaa.

Viimeksi kuluneen vuosikymmenen kuluessa on luotu erilaisia vaatimuslistoja, joiden valossa on pyritty arvioimaan tekoälyohjelmien ja -järjestelmien eettisyyttä ja ohjaamaan järjestelmien kehittämistä. Käytännössä ne koostuvat ohjelmistojen kehittäjille, tilaajille ja käyttäjille suunnatuista eettisistä periaatteista ja toimintaohjeista, joiden noudattamisen uskotaan tuottavan eettisesti kestäviä tekoälyratkaisuja. Tekoälyn eettisistä ohjeista ja tutki-

11 The Final Report 2020.

12 Final Committee Report B737 MAX, 6.

13 https://www.icao.int/safety/Documents/ICAO_SR_2018_30082018.pdf

14 Johnston ja Harris 2019.

15 Reiman ja Oedewald 2008, 17–18.

muksesta nousee muutama kiinnostava näkökulma.

Turvallisuudesta puhutaan ihmisyyhteisöjen ole-
massaolaa uhkaavana eksistentiaalisena riskinä Fu-
ture of Life -instituutin kokoamissa tekoälyn Asi-
lomar-periaatteissa. Niissä kiinnitetään huomiota
siihen, että kilpailu voi saada tekoälyn kehittäjät tin-
kimään turvallisuusstandardeista. Erityiseksi tur-
vallisuusuhaksi se mainitaan autonomisten aseiden
kehittämiseen tähtäävässä kilpavarustelussa.¹⁶

Muutoin ohjeissa puhutaan yllättävän vähän
ja hahmottomasti turvallisuudesta tekoälyteknolo-
gioiden eettisenä vaatimuksena. OECD:n te-
koälyteknologioita koskevissa eettisissä ohjeissa
esitetään, etteivät teknologiat saisi aiheuttaa koh-
tuutonta turvallisuusriskiä. Vaatimus jää kuiten-
kin epäselväksi, koska ohjeissa ei tarkenneta, mi-
hin asioihin kohdistuvia riskejä olisi tarkattava,
eikä eritellä, mitä turvallisuus tai kohtuuton tur-
vallisuusriski tarkoittavat.¹⁷

Kohtuuttoman vaaran määrittelemisen onkin
vaikeaa. Autonomisten kulkuneuvojen kohdalla
kohtuuttoman vaarallisena pidetään sitä, että te-
koälyn kontrolloima ajajärjestelmä on vähemmän
turvallinen kuin ihmisten kuljettama ajoneuvo tai
edellinen onnettomuuden aiheuttanut autonomi-
nen ajajärjestelmä.¹⁸ Näyttää siltä, että tunnettu
vaarallisuus on kohtuullista, kun taas kohtuuton
vaarallisuus tarkoittaa haittaa tai onnettomuutta,
jollaista ei vielä tunneta.

Turvallisuus esitetään tekoälyn eettisissä oh-
jeissa myös ominaisuutena, jota suojellaan ja pide-
tään yllä tekoälysystemien avulla. Kansainvälinen
insinöörijärjestö IEEE vetoaa eettisissä ohjeissaan
siihen, että tekoälyohjelmistojen on toimittava tur-
vallisesti koko niiden käyttöiän.¹⁹ Samoin Euroopan
komission korkean erityistason asiantuntijaryhmän
laatimassa *Trustworthy AI* -kirjasessa turvallisuus
mainitaan samassa yhteydessä, jossa muotoillaan
vaatimus suunnitella teknisesti vakaita järjestelmiä.
Turvallisuus rinnastetaan järjestelmän luotettavaan
toimintaan.²⁰ Sitä, mitä tekoälyohjelman turvallinen
toiminta käytännössä tarkoittaa, ei kuitenkaan näis-
sä ohjeissa tarkenneta.

Miten teknologioiden käyttäminen vaikuttaa
laajemmin ihmisyyhteisöjen ja organisaatioiden
jouhevaan yhteistoimintaan turvallisuuskriitti-
sillä aloilla? Tulisiko tekoälyn kehittäjien ja käyt-
töönottajien huolehtia vain esimerkiksi systeemin
toiminnan sujuvuudesta sekä datankeruun ja -käy-
tön toteutumisesta ilman, että dataa päätyy asian-
kuulumattomien ihmisten tai ohjelmien haltuun?
Vai olisiko keskityttävä huolehtimaan siitä, ettei-
vät muihin kuin systeemin kehittäjiin ja käyttäjiin
kohdistuvat, yhä laajenevat turvallisuusuhat to-
teudu? Epätasällinen turvallisuuspuhe vaikut-
taa eettisten ohjeiden pyrkimykseltä lisätä luotta-
musta uutta teknologiaa kohtaan.

Esimerkiksi Bryant W. Smith, joka kirjoittaa
Oxford Handbook of Ethics of AI -teoksessa automa-
tisoidun liikenteen eettisistä kysymyksistä, pää-
tyy esittämään, että automatisoitujen kulkuneu-
vojen turvallisuus merkitsee niiden valmistajaan
kohdistuvaa luottamusta. Smith argumentoi, että
vaikka automatisoidut ajoneuvot voivat aiheuttaa
vaaratilanteita ja joutua onnettomuuksiin, syste-
min epäonnistumiset ovat jäljitettävissä ja korjat-
tavissa. Toisin on ihmisten ohjaamien ajoneuvojen
kohdalla. Yhdysvaltalaisen lakinormien mukaan
automatoitu ajoneuvo on kohtuuttoman vaaral-
linen vain, jos se on vähemmän turvallinen kuin
ihmisten ohjaama ajoneuvo tai edellinen onnetto-
muuden aiheuttanut autonominen ajajärjestelmä.
Koska ainoastaan kohtuuttoman vaarallisia auto-
matisoituja ajajärjestelmiä käytäviä kulkuneuvo-
ja ei tule tuoda markkinoille ja koska kohtuuton
vaarallisuus on laissa määritelty, turvallisuuden ta-
kaaminen kiteytyy valmistajan luotettavuuteen.²¹

Uusia teknologioita arvioidaan sen mukaan,
kuinka toimivia, luotettavia ja riskittömiä ne ovat.
Teknologisen sovelluksen turvallisuus voidaan il-
maista näiden arviointiperusteiden avulla. Samal-
la ne ilmaisevat vaatimuksen, joka on sovelluksen
suunnittelua ja toimintavarmuutta ohjaava keskei-
nen eettinen periaate.

Tekoälyteknologioiden kehittäjille suunnat-
tujen eettisten vaatimuslistojen epäselvä turval-
lisuuspuhe tekee uuden sosio-teknisiä systeemejä
muodostavan teknologian turvallisuuden arvioi-
misesta vaikeaa. Tekoälyteknologian turvallisuus-

16 Future of Life Institute 2017.

17 OECD 2019.

18 Smith 2020, 676.

19 IEEE 2017.

20 HLEG 2019.

21 Smith 2020, 675–677.

den samaistaminen käyttäjien valmistajaa kohtaa tuntemaan luottamukseen jättää pimentoon todelliset – ei vain laitteiston toiminnassa vaan sen yhteistoiminnassa ihmistoimijoiden kanssa piilevät – turvallisuusriskit, joiden hallinta vaatii tarkkaa-vaista tunnistamista, korjaamista, laitteiston testaamista ja vaaratilanteiden simuloimista.

Ilmailussa automaatiotekniikkaa on hyödynnetty vasta perusteellisten testauksien jälkeen. Rasmussen (1983) erottaa toisistaan taito-, määräys- ja taitopohjaisen toiminnan lajit, joilla voi kuvata hyvin ilmailun ja autonomisten järjestelmien suhdetta.²² Taitopohjainen toiminta perustuu rutiininomaiseen ja usein sensomotoriseen suorittamiseen. Perinteinen lentokoneen ohjaaminen edustaa sensomotorisuutta aidoimmillaan. Koska tällaista toimintaa on melko helppo mallintaa, lentämistä on voitu automatisoida. Myös lennonjohdon määräyksiin pohjautuvien raja-arvojen – lentokorkeus ja -nopeus sekä ilmatilarajoitukset – noudattaminen samoin kuin moottoreiden tehonsäätö on yhä useammin annettu tietokoneiden hoidettavaksi.

Automaation yleistyessä on jouduttu kysymään sen soveltamisen rajoja. Ihminen häviää koneelle standardiratkaisuihin perustuvissa tehtävissä. Tilanteissa, joissa ongelmanratkaisu edellyttää tietopohjaisen²³ valinnan tekemistä seurauksiltaan tuntemattomien vaihtoehtojen välillä tai jonkin ennen kokeilemattoman ratkaisun muotoilemista – aina-kaan nykyiset – tekoälypohjaiset järjestelmät eivät korvaa ihmistä. Tietopohjainen toiminta on kulttuurisesti ja sosiaalisesti määrittyvää päämääräsuuntautuneisuutta, jossa tehty ratkaisu kuvastaa päätöksentekijän yhteisön edustamia arvoja. Ihminen poikkeaa koneesta siinä, että ihminen ei toimi algoritmin mukaan deterministisesti, vaan et-

sii aktiivisesti eri vaihtoehtoja. Ihmistoiminta on teleologista.²⁴

Ilmailussa koettujen poikkeamien, vaaratilanteiden ja onnettomuuksien tarkasteleminen on osoittanut, että ihmisellä on edelleen rooli etenkin tilanteissa, jotka edellyttävät tietopohjaista toimintaa. Yhä autonomisemman teknologian keskellä miehistön rooli on hallita poikkeustilanteita viimeisenä vikasietoisuuden varmistajana. Tällaisissa tilanteissa tulevat esiin myös ihmisen toiminnalle ominainen luovuus ja pyrkimys toimia moraalisesti kestäväällä tavalla.

Turvalliset tekoälysovellukset edellyttävät emergenttiä todellisuuskäsitystä

Turvallisuuskriittisillä aloilla tekoälysovellusten turvallisuudesta tulee yhä syvemmin järjestelmien emergentti ominaisuus: se ei ole toteutettujen toimenpiteiden suoraviivainen seuraus eikä niiden summa vaan järjestelmän ominaisuus, joka tulee ilmi tarkoituksenmukaisena sopeutumisenä ja itsekorjaavina muutoksina suhteessa toimintaympäristön haasteisiin. Nisula esittää tutkimuksessaan, että emergenttisyys ilmenevä ajattelu on yhteydessä laajaan todellisuuskuvan muutokseen. Perinteinen Descartesin ja Newtonin teorioihin pohjautunut näkemys maailmasta monimutkaisena (*complicated*), mutta selvitettävissä oleviin syysuhteisiin palautuvana mallina on korvautunut toisenlaisella maailmankuvalla. Sille on ominaista monivuorovaikutteisuus (*complexity*), ei-lineaarisuus ja dynaamisuus.²⁵

Myös turvallisuusmalleja on uudistettava vastaamaan toimintatodellisuuden vaatimuksia sopeutua yhä uudelleen ympärillä tapahtuviin muutoksiin ja pitää yllä valmiutta kohdata ennakkoimattomia tilanteita ja tekijöitä. Turvallisuuden ylläpitäminen on edellyttänyt monivuorovaikutteisuuden ja sopeutumiseen tähtäävien toimintatapojen hyödyntämistä (*Complex Adaptive Systems, CAS*) myös ilmailussa.²⁶

Turvallisuuden huomioon ottaminen on keskeinen tavoite myös, kun suunnitellaan ja hyödynnetään tekoälysovelluksia. Kehitystyöhön liitetään uusille aloille tyypillistä innokkuutta ja

22 Rasmussen 1983.

23 Reason (1990, 56) erottaa toisistaan kolme ihmiselle tyypillistä toimintatapaa. Rutiinoinainen toiminta on taitopohjaista (*skill-based*), eikä edellytä toimijalta harkintaa, vaan on usein automatisoitunutta. Ongelmanratkaisutilanteessa on turvauduttava joko sääntöpohjaiseen (*rule-based*) tai tietopohjaiseen (*knowledge-based*) toimintaan. Sääntöpohjaiselle toimijalle on tyypillistä, että toimija etsii ratkaisua käsillä olevaan pulmaan soveltamalla aiemmin omaksumiaan toimintaa ohjaavia sääntöjä. Tietopohjainen lähestymistapa on hidas mutta (useimmiten ainoa) tehokas tapa ratkaista ongelmia, joiden syyt toimija ei ymmärrä ja joiden ratkaisemiseen entiset toimintatavat eivät siksi sellaisenaan sovellu.

24 Rasmussen 1983, 257.

25 Nisula 2018, 42–44. Ks. myös Reiman ym. 2015, 81–84.

26 Nisula 2015b.

yskinomaista tulevaisuuteen suuntautumista ja laajemman yleisön vakuuttamista teknologioiden luotettavuudesta. Samalla voidaan unohtaa, että tekoälyteknologiat eivät muodosta erillistä utopiamaailmaa, vaan tulevat osaksi monimutkaisia yhteiskunnallisia, sosiaalisia ja teknisiä prosesseja. Monia tekoälynsovellusten kannalta keskeisiä kysymyksiä on pohdittu ennenkin. Aiemmin koetusta, ymmärretystä ja koetellusta kannattaa oppia. Tulevaisuusutopioiden rinnalla on harjoitettava asioiden aktiivista mieleen palauttamista ja reflektointia, *anamneesia*.

Vaikka monet meneillään olevat ja tulevat muutokset mullistavat arkeakin, keskeiset fyysisen todellisuuden tekijät säilyvät entisellään – ihmiset ovat luonnosta ja toisistaan riippuvaisia, monella tavalla haavoittuvia biologisia olentoja. Ihmisen osaa eksistentiaalisesti määrittävä kuolema kulkee ihmisen rinnalla taivaita kohti kurrottavan keskeliäisyyden ja rajojen ylittämisen varjona. Tällaisten kysymysten ratkaisemiseen tekoälysovellukset eivät riitä, vaan tarvitaan ihmisyyhteisöjen muotoilemia kulttuurisia käsitteitä ja perinteiden synnyttämiä symbolisia järjestelmiä tuomaan merkitystä ja arvioimaan myös kriittisesti meneillään olevaa kehitystä. Tekoäly ei pidä vieraannuttaa ihmistojimijasta pitämällä sitä älypelin kuningattarena vaan muistaa, että se on ihmisen ihmistä palvelemaan luoma apuväline.

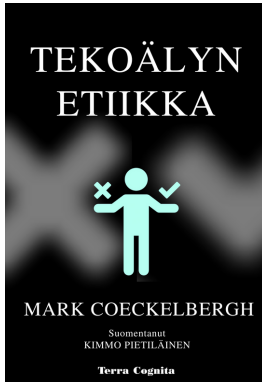
Kirjallisuutta

- Dubber, Markus D., Pasquale, Frank ja Das, Sunit, toim. (2020). *The Oxford Handbook of Ethics of AI*. New York: Oxford University Press.
- Eteläpelto, A., Heiskanen, T. ja Collin, K. (2011). ”Vallan ja toimijuuden monisäikeisyys”. TamPub. <http://urn.fi/URN:NBN:fi:uta-201612162853>
- Final Committee Report (2020). The design, development and certification of the Boeing 737 MAX. September 2020. The House Committee on Transportation & Infrastructure.
- Future of Life Institute (2017). Asilomar AI Principles. <https://futureoflife.org/ai-principles/>. 27.4.2021.
- Grace, K., Salvatier, J., Dafoe, A., Zhang, B. ja Evans, O. 2018. Viewpoint: When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research* 62: 729–754.
- Haikarainen, Raine. 2021. *Taivaalla opittua. Ilmailu ja turvallisuusajattelun kehittyminen* (julkaisematon käsikirjoitus).
- Hallamaa, J. ja Kalliokoski, T. (2020). ”How AI Systems Challenge the Conditions of Moral Agency?” Rauterberg M. (toim.): *Human Computer Interaction (HCI) 2020*, Springer Nature. LNCS 12215, 54–64. https://doi.org/10.1007/978-3-030-50267-6_5
- Helen. Suomen energisin asiakaslehti 1/2021. Helsinki: Helen Oy.
- IEEE (2017). The Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent

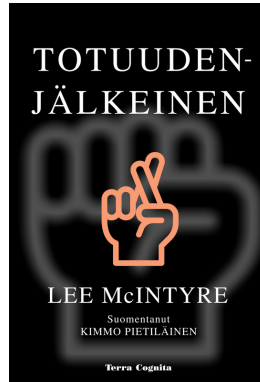
- Systems, Version 2.
- HLEG (2019). Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. *Ethics Guidelines for Trustworthy AI*.
- Johnston, Phillip ja Harris, Rozi (2019), The Boeing 737 MAX Saga: Lessons for Software Organizations. *Software Quality Professional*. Volume 21, Issue 3, 4–12.
- Kelola, K. (2021). ”Tekoäly. Kaikki mitä olet halunnut tietää tekoälystä - ja vähän enemmänkin”. *Helen* 1/2021, 11–19.
- Kuusi, O. ja Heinonen, S. (2020). Tulevaisuuspolkuja kapeasta tekoälystä vahvaan tekoälyyn. *Tieteessä Tapahtuu*, 38(3). Noudettu osoitteesta <https://journal.fi/tt/article/view/9556120.4.2021>
- Marr, B. (2019) *Artificial intelligence in practice: How 50 successful companies used AI and machine learning to solve problems*. Chichester, West Sussex, UK: Wiley.
- Needham, J. (2017) We are the robots: is the future of music artificial? *FACT Magazine*, <https://www.factmag.com/2017/02/19/we-are-the-robots-could-the-future-of-music-be-artificial>, 27.4.2021.
- Nisula, Jari M. (2015a). Moderni turvallisuusajattelu. *Finnanest* 48(4), 328–331.
- Nisula, Jari M. (2015b). *Riskiperusteinen vaikuttaminen. Tiedosta toimenpiteisiin -projekti*. Trafin tutkimuksia 11/2015.
- Nisula, Jari M. (2018). *A risk management framework for a complex adaptive transport system*. *Multiagent Systems [cs.MA]*. Toulouse III: Université Paul Sabatier.
- OECD (2019). Recommendation of the Council on Artificial Intelligence. <https://oecd.ai/ai-principles>.
- Packer, M. J. ja Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist* 35 (4), 227–241.
- Rasmussen, Jens (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*. SMC-13, No 3, 257–266.
- Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.
- Reiman, T. ja Oedewald, P. (2008). *Turvallisuuskriittiset organisaatiot. Onnettomuudet, kulttuuri ja johtaminen*. Helsinki: Edita.
- Smith, B. W. (2020), ”Ethics of Artificial Intelligence in Transport”, teoksessa *The Oxford Handbook of Ethics of AI* (toim. M. D. Dubber & F. Pasquale & S. Das). New York: Oxford University Press 669–683.
- Tekoälyn perusteet -Mooc-kurssi. Helsingin yliopisto ja Reaktor. <https://course.elementsofai.com/fi/>.

Jaana Hallamaa on Helsingin yliopiston sosiaalietiikan professori. Taina Kalliokoski on tutkijatohtori Helsingin yliopistossa. Raine Haikarainen on vapaa tutkija Helsingin yliopistossa. Kaikki kolme työskentelevät tekoälyn eettistä suunnittelua ja käyttöä julkisen hallinnon ohjauksessa tutkivassa Strategisen tutkimuksen neuvoston rahoittamassa ETAIROS-hankkeessa. www.etairos.fi

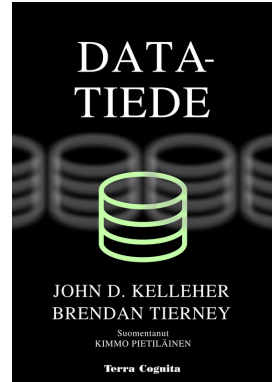
PARASTA SUOMENKIELISTÄ TIETOKIRJALLISUUTTA.



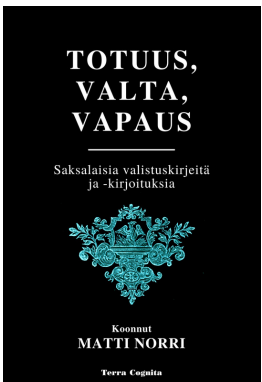
Mark Coeckelbergh:
Tekoälyn etiikka.
Ovh. 30 €.



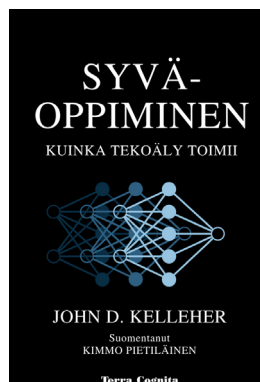
Lee McIntyre:
Totuudenjälkeinen.
Ovh. 30 €.



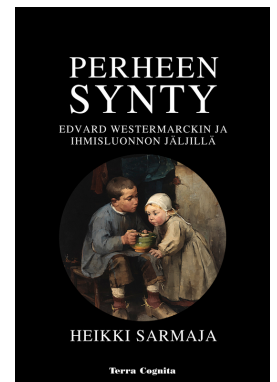
Kelleher & Tierney:
Datatieide.
Ovh. 30 €.



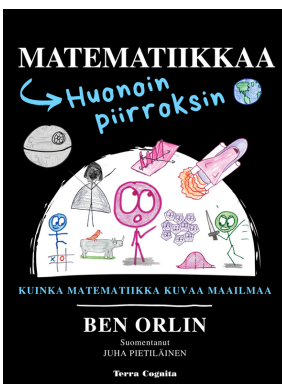
Matti Norri, toim: *Totuus, valta, vapaus - Saksalaisia valistuskirjeitä ja -kirjoituksia..*
Ovh. 40 €.



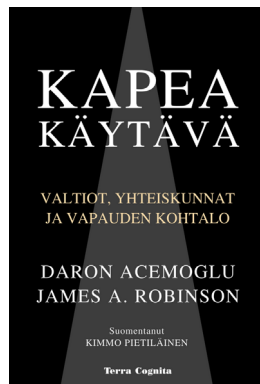
John D. Kelleher: *Syvä-oppiminen, Kuinka tekoäly toimii.*
Ovh. 30 €.



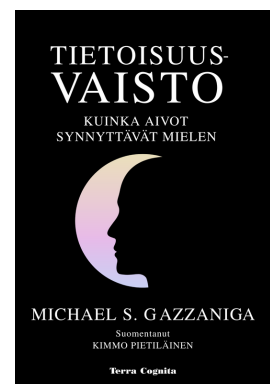
Heikki Sarmaja: *Perheen synty, Edvard Westermarckin ja ihmisluonnon jäljillä.*
Ovh. 50 €.



Ben Orlin: *Matematiikkaa huonoin piirroksin, Kuinka matematiikka kuvaa maailmaa.*
Ovh. 50 €.



Daron Acemoglu ja James A. Robinson: *Kapea käytävä, Valtiot, yhteiskunnat ja vapauden kohtalo.*
Ovh. 60€.



Michael S. Gazzaniga:
Tietoisuusvaisto, Kuinka aivot synnyttävät mielen.
Ovh. 40 €.

Terra Cognita

Hyvin varustetuista kirjakaupoista tai suoraan kustantajalta. www.terracognita.fi