

# EETTISELLÄ TEKÖÄLYLLÄ LISÄARVOA SUOMELLE

Taustapaperi Tietopoliittisen selonteon etiikkatyöryhmälle 4.6.

Raija Koivisto, VTT

*Kehitys uusien teknologioiden alueella on ollut teknologiavetoista, mitä suuret toimijat ovat jopa hehkuttaneet julkisuudessa: “We work in an industry that does not respect tradition, only innovation.” (Satya Nadella, CEO of Microsoft, Hong Kong, 6.3.2018). Toisaalta, Facebookin ja Googlen rakentajanörtit kyllästyivät luomiensa jättialustojen epäeettiseen toimintaan ja perustivat yhdistyksen, jonka nimissä käyvät puhumassa mm. kouluissa asiasta Kaliforniassa. Euroopassa eettisyys on noussut agendalle niin komission kuin useiden jäsenvaltioidenkin toimesta. Keskustelu eettisyydestä on pääsääntöisesti leimannut eettisyyden suureksi uhaksi tekoälyn soveltamisessa, jonka vastapainoksi useat maat ovat julistaneet olevansa eettisen tekoälyn edelläkävijöitä. Tämä paperi esittää perusteluja visiolle, jonka mukaan eettisyys tekoälyn kehittämisessä ja soveltamisessa voisi tuoda lisäarvoa ja kilpailuetua Suomelle, sekä esittää vaiheita tekoälyn eettisyyden toteuttamiseksi ja todentamiseksi.*

## Eettisyyden tarkastelun kehitysvaiheet

Tekoälyn ennakoitaan muuttavan maailmaa jopa enemmän kuin höyrykone aikoinaan. Tekoälysovelluksia kehitetään kaikille elämän alueille ja siihen liittyvien liiketoimintavoittojen arvioidaan kasvavan kymmeniin miljardeihin lähitulevaisuudessa. Toistaiseksi kehitys on ollut teknologiavetoista eikä tekoälyn soveltamisen kaikkia vaikutuksia ole huomioitu ennen sen käyttöönottoa. Niinpä varoittavien esimerkkien myötä on herännyt tarve tarkastella tekoälyn kehittämistä ja soveltamista kokonaisvaltaisemmin.

Kuvaava esimerkki tekoälyn käytöstä epäilyttävään tarkoitukseen on Cambridge Analyticsin tarjoama ’vaalipalvelu’, joka hyödynsi Facebookin avulla kerättyä dataa - ilman, että ainakaan Facebookin käyttäjät olivat siitä tietoisia. Toinen hälyttävä esimerkki on ollut Kiinan kansalaisten profilointihanke. Itseajaville autoille sattuneet onnettomuudet ovat herättäneet keskustelua ja epäilyksiä, mutta suuren yleisön keskuudessa on vähemmän puhuttu lääketieteellisistä sovelluksista kuten CRISP, joka mahdollistaa dna:n muokkaamisen. Jokapäiväisessä elämässämme tekoälysovelluksia on käytössä jo monissa palveluissa ilman, että käyttäjä sitä edes ymmärtää.

Tekoälyn eettisyyden arvioimiseksi on siihen liittyvät ilmiöt ymmärrettävä. Tekoälyn eettisyyden suhteen tilannetta voidaan verrata teollisuuslaitoksen turvallisuusajattelun kehittymiseen. 40 vuotta sitten teollisuusautomaation yleistyttyä piti sen soveltamiseen liittyvät ilmiöt - ja niiden avulla riskitekijät - ymmärtää, jotta tilanne voitiin ottaa haltuun ja teollisuuslaitokset suunnitella turvallisiksi. Kehitettiin riskianalyysimenetelmiä (HAZOP, Vikapuu, FMEA, jne) vaarojen tunnistamiseksi, riskien arvioimiseksi ja hallitsemiseksi. Seuraavaksi siirryttiin laitosten ja suunnitelmien analysoinnista turvallisuuden huomioimiseen koko suunnittelun ajan (inherent safety, safety-conscious design). Lopulta, kun turvallisuuden huomiointi on pystytty tekemään läpinäkyväksi ja jopa sertifioitavaksi koko laitoksen elinkaaren ajalta, on turvallisuudesta tullut

yritystoiminnan lisäarvo. Vaikka eettisyyttä ei suoraan voikaan verrata turvallisuuteen, niin tekoälyyn liittyvän eettisyyden tarkastelussa voidaan ajatella olevan selviä analogioita turvallisuusajattelun kehittymisen kanssa.

Eettisyyden tekemisessä lisäarvoksi on ainakin seuraavia vaiheita:

- ilmiön ymmärtäminen: 1) mitä eettisyys on ja 2) miten konteksti (tekoälysovellus ja sen käyttö) määritellään ja kuvataan niin, että sitä voidaan tarkastella eettisyysnäkökulmasta
- eettisyyden arviointi: systematiikka, jonka avulla kohteen eettisyys voidaan arvioida
- eettisyyden sisällyttäminen tekoälyn ja sen sovellusten suunnittelemisen jokaiseen vaiheeseen - eettisyystietoinen suunnittelu ja toteutus
- hyväksytyt eettisyyden todentamismenetelmät, joiden avulla eettisyyden toteutuminen tekoälyn elinkaaren eri vaiheissa ja sovelluksissa voidaan osoittaa.

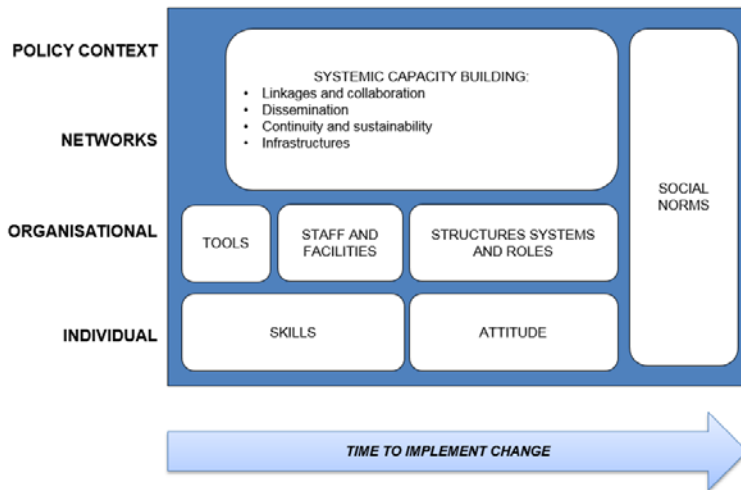
Edellytys sille, että eettisyys tulee huomioiduksi on yleinen eettisyystietoisuuden lisääntyminen. Kun on kyse eettisyydestä, ei ole valmiiksi tarjottavia ratkaisuja, vaan eettisyys määräytyy kunkin arvojen kautta. Sen vuoksi eettisyydestä tulee keskustella: keskustelua voidaan kuitenkin tukea esittämällä tärkeitä näkökohtia (vrt. Antti Kauppinen, HY & Arto Laitinen, TY), jotka eettisyyskeskustelussa tulee huomioida. Suomalaisessa ja Pohjoismaalaisissa kulttuureissa yleensä on jo perinteisesti monia eettisyyttä tukevia piirteitä, kuten rehellisyys, tasa-arvo, luottamus, alhainen korruptio, sosiaalinen vastuu, jne., jotka toivottavasti eivät ole jäämässä vallitsevan länsimaisen itsekeskeisyyden ja näyttäytymiskulttuurin jalkoihin.

## Ilmiön ymmärtäminen ja eettisyystietoisuuden lisääminen - SATORI roadmap

EU SATORI -projektissa (<http://satoriproject.eu/>) kehitettiin menettely tutkimuksen eettisyyden arviointiin. Menettely julkaistiin kaksiosaisena CEN esistandardina:

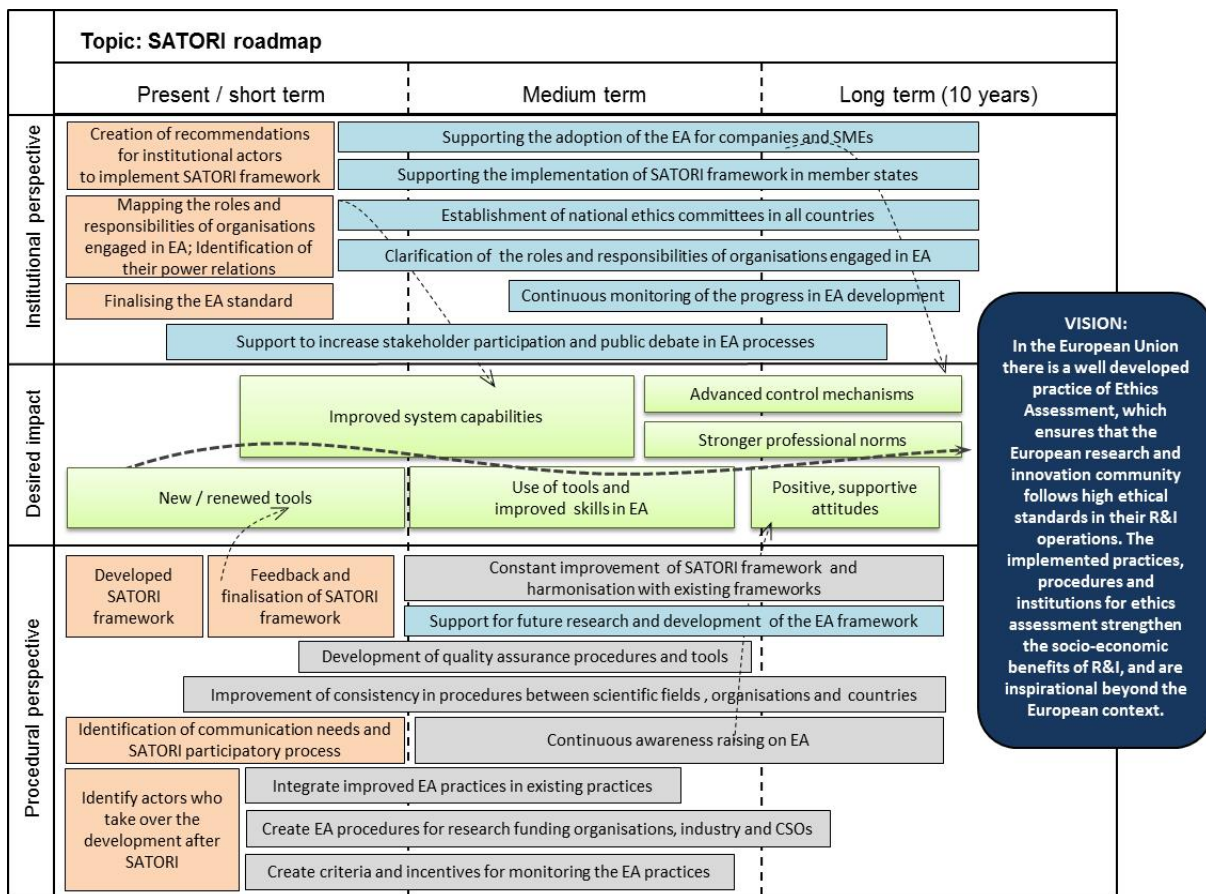
1. CWA 17145-1, Ethics assessment for research and innovation - Part 1: Ethics committee
2. CWA 17145-2, Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework.

SATORIssa tehtiin myös tiekartta siitä, miten eettisyystietoisuutta voitaisiin lisätä Euroopassa ja siten auttaa menetelmien käyttöönotossa. Yhtenä sisäänrakennettuna tavoitteena oli tutkimuseettisten käytäntöjen yhtenäistäminen ainakin Euroopassa. Tiekartan perustaksi muotoiltiin eri käyttäytymiseen ja yhteiskunnalliseen muutokseen liittyvistä teorioista yksinkertainen kehikko, jossa haluttua muutosta tarkastellaan eri tasoilla: politiikkatoimet, verkostot, organisaatiot ja yksilöt (kuva 1). Tarkasteluajanjaksoksi valittiin 10 vuotta.



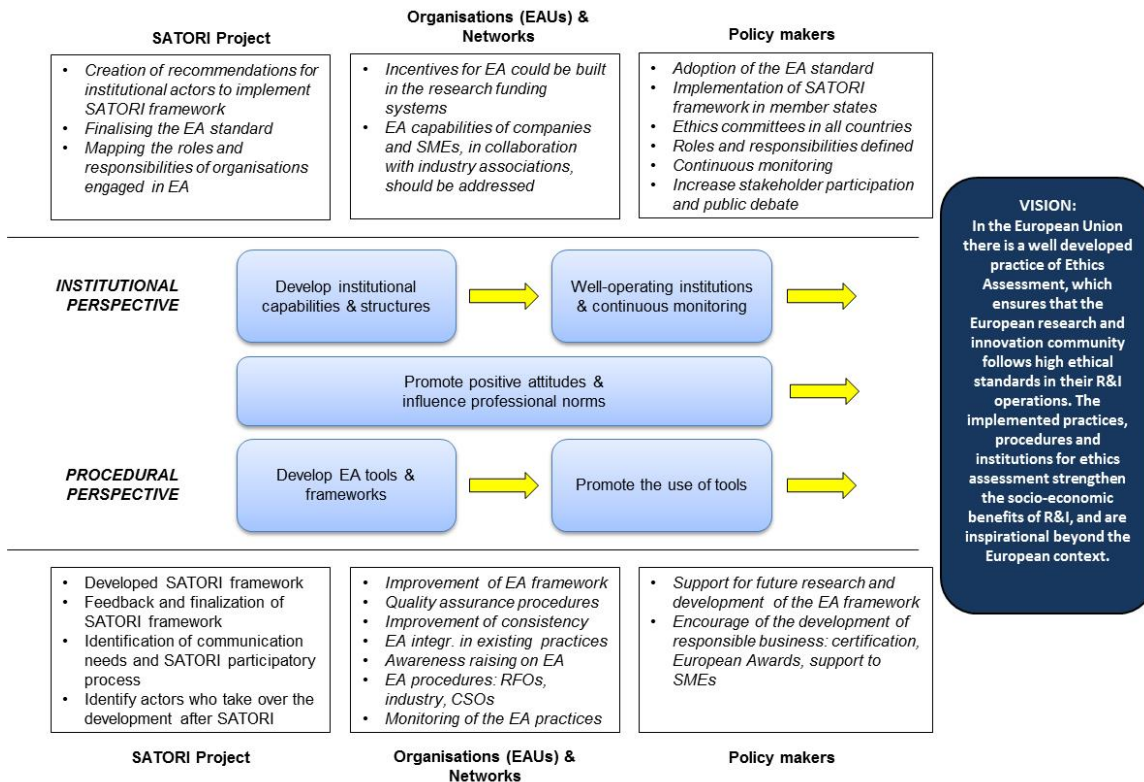
Kuva 1. Haluttuun muutokseen vaikuttavat tekijät (Leinonen et al. 2017).

Kun muutostekijöitä tarkasteltiin edelleen, jaoteltiin tarvittavat toimenpiteet menettelytapoihin ja institutionaalsiin toimenpiteisiin (kuva 2). Eettisyyden luonteesta johtuen ulkoa tuodut toimenpiteet eivät täysin takaa muutoksen toteutumista, vaan sen lisäksi tarvitaan vielä positiivista asennetta eettisyyden huomioimiseen.



Kuva 2. SATORI -tiekartta (Leinonen et al. 2017).

Kuvaan 3 tarvittavat toimenpiteet on ryhmitelty toimijoiden mukaan.

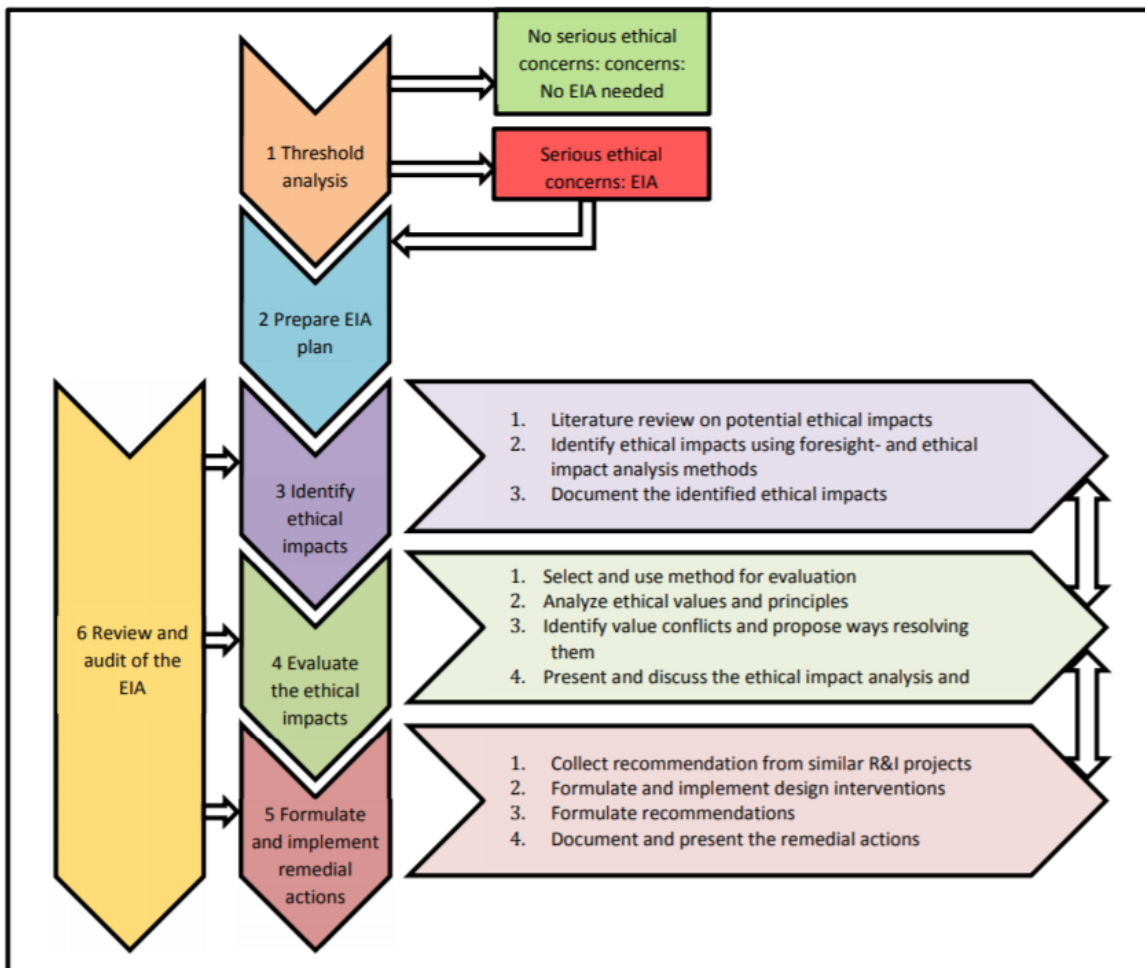


Kuv 3. Eettisyysvision toteutumiseen tarvittavia toimenpiteitä ryhmiteltyinä toimijoittain: poliittikatoimet, organisaatiot ja verkostot sekä SATORI -projekti (joka päättyi 2017 syksyllä, Leinonen et al. 2017).

## Eettisyyden arviointi

Eettisyyden arvioinnin ovat perinteisesti tehneet siihen varta vasten nimetyt asiantuntijaryhmät (lautakunnat, toimikunnat, jne.). Eri maissa on erilaisia käytäntöjä, mutta esim. lääketieteessä ja biotekniikassa käytännöt ovat vakiintuneet. Tekniikan kehityksessä taas, jos kyse ei ole lääketieteen sovelluksesta, on tähän mennessä edetty teknologia edellä ja usein eettisyys on tullut esille vasta, kun teknologian soveltamisessa on havaittu eettisiä ongelmia.

SATORI -projektissa kehitetty tutkimuksen eettisyyden arviointiperiaate on esitetty kuvassa 4.



Kuva 4. Ethical impact assessment framework (Reijers et al. 2016).

## Eettisyyden huomioon ottaminen tekoälyn suunnittelussa ja käytössä

Prosessiteollisuuden inherent safety -ajattelua vastaavasti eettisyys voidaan tuoda osaksi tekoälyn suunnittelu- ja kehitystoimintaa: analogisesti turvallisuustietoisesta suunnittelusta (Koivisto 1996) kanssa voimme toteuttaa eettisyystietoista suunnittelua tekoälyn kyseessä ollessa. Eettisyyden tarkastelussa on tärkeää huomioida ihmiset, jotka liittyvät suunniteltavaan järjestelmään tavalla tai toisella. Tällöin puhutaan käyttäjälähtöisestä suunnittelusta. Organisaatiotasolla analogioita voidaan hakea responsible research and industry (RRI) ja Corporate Social Responsibility (CSR) -ajatteluista. Mitä tahansa menetelmää tai menettelyä sovelletaankin, on eettisyyden tapauksessa aina huomattava, että se ei ole luonnontieteellisesti määriteltävissä, vaan riippuu asiaa tarkastelevan henkilön tai yhteisön arvoista ja on keskusteltava asia. Järkevien yhteisesti hyväksyttävien linjojen löytämiseksi ja hyväksymiseksi tarvitaan tutkimusta ja yhteiskunnallista keskustelua.

## Eettisyydestä lisäarvoa ja kilpailuetua Suomelle

Suomi - ja muut pohjoismaat - tunnetaan hyvinvointivaltioiden kärkimaana: luottamus, avoimuus, tasa-arvo ja korruptoitumattomuus ovat kansallisia hyveitämme. Vaikka Suomi pärjää myös teknologiaosaamisessa, on kilpailu maailmalla kovaa ja vaatii isoja rahoja. Tarvitsemme erityisen

kilpailutekijän, jonka avulla luoda lisäarvoa ja erottua. Eettisyys voisi olla se kilpailu- ja lisäarvotekijä, johon myös muut uskovat. Kun eettisyys yhdistetään teknologiaosaamiseemme, pystymme tuottamaan käyttäjille lisäarvoa, jota koko maailma näinä itsekkyyden aikoina tarvitsee.

Eettisyys liittyy päätöksentekoon, joka enenevässä määrin siirtyy tekoälyn tehtäväksi. Siksi erityisesti tekoälyn eettisyyttä tulisi kehittää. Turvallisuuden kohdalla tilanne oli analoginen 40 vuotta sitten, mutta nyt turvallisuudesta on tunnetusti tullut lisäarvo- ja kilpailutekijä. Sama kehitys on mahdollista eettisen tekoälyn tapauksessa, mutta siihen meillä ei ole käytettävissämme edes kymmentä vuotta. Onneksi työ on jo alkanut.

## Lähteitä

Koivisto R. 1996. Safety-conscious process design. VTT Publications: 264. Doctoral thesis.

Leinonen A., Koivisto R., Tuominen A., Douglas D., Gurzawska A., Jansen P., Kapeller A., Gauttier S. and Brey P. 2017. Roadmap towards adoption of a fully developed ethics assessment framework. EU SATORI Deliverable D4.3. Saatavissa:

[http://satoriproject.eu/media/D4.3\\_SATORIRoadmap.pdf](http://satoriproject.eu/media/D4.3_SATORIRoadmap.pdf)

Reijers W., Brey P., Jansen P., Rodrigues R., Koivisto R. and Tuominen A. 2016. A Common Framework for Ethical Impact Assessment, Annex 1: A reasoned proposal for a set of shared ethical values, principles and approaches for ethics assessment in the European context. EU SATORI Deliverable D4.1.