



Tekoäly yhteiskunnassa – Eettinen ja sosiaalinen kestävyys

Jaana Leikas, Antero Karvonen, Nina Wessberg, Anton Sigfrids, Marinka Lanne

VTT Public

Raportin nimi Tekoäly yhteiskunnassa - Eettinen ja sosiaalinen kestävyys	
Projektin nimi Eettinen tekoälyn hyödyntäminen yhteiskunnan ohjauksessa	Projektin numero/lyhytnimi 124027/ETAIROS
Raportin laatija(t) Jaana Leikas, Antero Karvonen, Nina Wessberg, Anton Sigfrids, Marinka Lanne	Sivujen/liitesivujen lukumäärä 45/11
Avainsanat Tekoäly, etiikka, vastuullisuus, hallinto	Raportin numero VTT-R-00580-25
Tiivistelmä Tekoälyn hyödyntäminen tuo yhteiskuntaamme monia laaja-alaisia muutoksia. Julkis- ja valtionhallinnon on tärkeää kyetä vastaamaan yhteiskunnan haasteisiin ja ohjaamaan tekoälyn käyttöönotosta seuraavaa muutosta mahdollisimman eettiseen ja sosiaalisesti kestäväan suuntaan. ETAIROS – ”Eettinen tekoäly yhteiskunnan ohjauksessa” oli Strategisen tutkimuksen neuvoston rahoittama hanke ”Julkisen vallan muuttuva rooli ja yhteiskunnan ohjaamisen mahdollisuudet” - STEER-ohjelmassa (2019-2025). Hankkeen tavoitteena oli tutkia ja kehittää käytännön prosesseja ja viitekehyksiä tekoälyteknologioiden eettistä ja sosiaalisesti kestävää käyttöä varten. Monialaisen, usean tutkimusorganisaation ETAIROS-hankkeen VTT:n osahankkeessa tarkasteltiin tekoälykehityksen ennakoinnin ja tekoälyjärjestelmien yhteiskunnallisen hallinnan ja ohjauksen näkökulmia. ETAIROS toimi väylän avaajana ja keskeisenä vaikuttajana tekoälyjärjestelmien eettisessä kehitystyössä Suomessa. Hankkeessa tehty tutkimus- ja vuorovaikutustyö on nostanut tekoälyn eettisen ohjauksen kysymykset näkyvästi suomalaiseen yhteiskunnalliseen keskusteluun ja tarjonnut konkreettisia ratkaisuja ja työkaluja niin päätöksentekijöille kuin yrityksille.	
VTT Public	
Tampere-Espoo 30.11.2025	
Laatija Jaana Leikas Johtava tutkija	Tarkastaja Katri Valkokari Research Manager
<i>VTT:n nimen käyttäminen mainonnassa tai tämän raportin osittainen julkaiseminen on sallittu vain Teknologian tutkimuskeskus VTT Oy:ltä saadun kirjallisen luvan perusteella.</i>	

"Kuvittele, että teknologia..."

kohtelisi näkemyksiämme ja aikomuksiamme arvokkaina sen sijaan, että se ohjaa halujamme ja vie huomiomme,

olisi suunniteltu niiden ehdoilla, joihin se vaikuttaa, eikä niinkään niiden, jotka pyrkivät hyötymään siitä taloudellisesti,

auttaisi meitä tekemään viisaita valintoja sen sijaan, että se passivoi ja tekee meistä avuttomia ja hiljaisia uutisten seuraajia,

tukisi kestävien digitaalisten demokratioiden rakentamista sen sijaan, että se tyrkyttää valeuutisia, ääriajattelua ja hukuttaa ihmiset informaatiotulvaan,

edistäisi yhteisen ymmärryksen syntymistä ja mahdollistaisi syvällisen yhteistyön globaalien haasteiden edessä sen sijaan, että se vaikeuttaa yhteiseloja, polarisoi ihmisryhmiä ja synnyttää riitelyn kautta toimintakyvyttömyyttä." ¹

¹ Center for Humane Technology, Humane technology -kurssin materiaali, kirjoittajien käännös.

Alkusanat

ETAIROS² -tutkimushanke *"Eettinen tekoäly yhteiskunnan ohjauksessa"* (2019-2025) oli Strategisen tutkimuksen neuvoston rahoittama hanke (Julkisen vallan muuttuva rooli ja yhteiskunnan ohjaamisen mahdollisuudet - STEER-ohjelma), jonka tavoitteena oli tutkia ja kehittää käytännön prosesseja ja viitekehyksiä tekoälyteknologioiden käytön eettistä ja sosiaalisesti kestäväää käyttöä varten.

Strategisen rahoituksen avulla kanavoidaan tutkimusrahoitusta valtioneuvoston tunnistamiin yhteiskunnan merkittävien ja suurten haasteiden tutkimukseen: tutkimukseen, joka palvelee elinkeinoelämän uudistamista ja kilpailukykyä, työelämän kehittämistä ja julkisen sektorin kehittämistä.

Hankkeen toteuttivat yhdessä VTT:n kanssa Tampereen, Helsingin, Jyväskylän ja Turun yliopistot. Tampereen yliopisto vastasi hankkeen koordinoinnista (prof. Jyrki Nummenmaa) ja koneoppimisen tutkimuksesta (prof. Jaakko Peltonen), Helsingin yliopisto etiikan tutkimuksesta (prof. Jaana Hallamaa), Jyväskylän yliopisto suunnitteluajattelun tutkimuksesta (prof. Pertti Saariluoma) ja Turun yliopisto tekoälyn regulaatiotutkimuksesta (prof. Mika Viljanen). VTT:n tutkimus keskittyi erilaisten tekoälytulevaisuuksien ennakkoinnin sekä tekoälyn käytön yhteiskunnallisen ohjauksen tutkimukseen (dos. Jaana Leikas). Hankkeen vuorovaikutuksesta ja viestinnästä vastasi ensimmäisellä rahoituskaudella 4FRONT Ltd. (Kalle Piirainen), ja toisella rahoituskaudella VTT (Antero Karvonen).

Monialainen ETAIROS-hanke edisti tekoälyn soveltamista eettisesti eri liike-elämän aloilla ja julkisessa hallinnossa. Tekoälyn eettisen hyödyntämisen tarpeita ja ratkaisuja tarkasteltiin erilaisten teemojen kautta: yhteiskunnan ohjauksessa ja sääntelyssä, sosiaali- ja terveyspalveluissa, ikääntymisessä, kansallisessa turvallisuudessa, yhteiskunnallisessa vakaudessa, autonomisissa järjestelmissä ja tulevaisuuden ennakkoinnissa.

Tässä raportissa tarkastellaan ETAIROS-hankkeen toteutusta ja tuloksia, sekä joitakin hankkeessa toteutettuja case-esimerkkejä VTT:n tutkimusalueen - *tekoälykehityksen ennakkoinnin ja tekoälyjärjestelmien yhteiskunnallisen hallinnan ja ohjauksen* - näkökulmasta.

Tampere-Espoo 30.11.2025

Jaana Leikas (VTT:n ETAIROS-osahankkeen johtaja), Antero Karvonen (ETAIROS-hankkeen vuorovaikutusvastaava), Nina Wessberg, Anton Sigfrids ja Marinka Lanne

Kiitokset:

Nadezhda Gotcheva, Mika Nieminen, Nina Rilla, Tuisku Salonen, Santtu Lehtinen, Emmi Heinisuo

² kreik. έταίρος - kumppani, yhteistyökumppani, ystävä; kuinka hyvää voidaan tehdä ja edistää ja pahaa välttää ja ehkäistä silloin, kun ihmisten ja tekoälyjärjestelmien toiminta punoutuvat yhteen.

Sisällysluettelo

1.	Johdanto	6
2.	Tekoälyn etiikka	9
2.1	Etiikka kontekstuaalisena ja toiminnallisena käsitteenä	9
2.2	Ihmiskeskeisyys kehittämisen keskiössä	9
3.	Miten ohjata sosioteknistä muutosta kohti eettisesti kestäviä tuloksia?	11
3.1	Tekoälyn vastuullinen hyödyntäminen Suomen kilpailukyvyyn ja hyvinvoinnin edistämiseksi	11
3.1.1	CASE OSKI-malli - Julkishallinto tukemassa eettisen tekoäly-yhteiskunnan rakentamista	13
3.2	Sosioteknisen muutoksen ennakointi	15
3.2.1	CASE JULKISHALLINTO -Tekoälytulevaisuuksien eettisten vaikutusten ennakkoiva arviointi	17
3.3	Osallistava eettinen visioiden arviointi	19
3.3.1	CASE CITYVERSE – Tampereen kaupunki	21
3.4	Kansalliset ja kansainväliset sääntelykehikot	22
3.5	Julkishallinto tekoälyn hyödyntäjänä.....	23
3.5.1	CASE VNK - Tekoälyn käyttöönoton johtaminen ja organisatoriset jännitteet.....	25
3.6	Luottamus tekoäly-yhteiskuntaan edellyttää osallisuuden kokemusta	29
3.6.1	CASE ETPK - Osallisuus ja osallistuminen	31
4.	ETAIROS eettisenä kompassina	33
4.1	Vaikuttavuus syntyy käytännön näkökulman kautta.....	33
4.2	Vuorovaikutus sidosryhmien kanssa: innovaatiofoorumit ja työpajat.....	34
4.3	Vaikuttavuus syntyy yhteistyöstä ja yhteiskehittämisestä	35
4.4	Vaikuttavuus syntyy julkisten toimijoiden muuttuvan roolin tukemisesta	36
4.5	Eettisen suorituskyvyn mittaristo	36
5.	Julkaisut ja media.....	38
6.1	VTT:n vertaisarvioidut artikkelit	38
6.2	ETAIROS-podcast 10 pulmaa tekoälystä	42
6.3	VTT:n toisen rahoituskauden (2022-2025) mielipidekirjoitukset, blogit ja populaariartikkelit	43
	Lähteet.....	44
	LIITE 1 ETAIROS-julkaisut, kaikki partnerit.....	46

Keskeiset käsitteet

Ennakointi	Prosessi, jossa tuotetaan näkemyksiä tulevaisuuteen varautumisesta ja tulevaisuuden luomisesta.
Teknologian etiikka	Yleisesti ottaen etiikka tutkii moraalisia kysymyksiä: mikä on oikein ja väärin, hyvää ja pahaa, sekä miten meidän tulisi toimia. Se tarkastelee toiminnan perusteita, arvoja ja velvollisuuksia sekä pyrkii muodostamaan perusteltuja näkemyksiä siitä, millainen elämä on tavoittelemisen arvoista.
Tekoälyn etiikka	Tarkastelee älykkäiden teknologioiden (kuten itseohjautuvien ajoneuvojen, kasvojentunnistuksen, robottien tai konekääntäjien) kehitystä ohjaavia arvoja ja sen suunnitteluun, kehitykseen ja käyttöön liittyviä eettisiä dilemmoja.
Tekoälykehitys	Tässä raportissa: tekoälyn käytön kehittyminen ja siitä johtuva sosiotekninen muutos yhteiskunnassa. Oleellisia kysymyksiä ovat mm. tekoälypohjaisten ratkaisujen yhteiskunnallinen haluttavuus ja hyväksyttävyys sekä ympäristön kestävyys.
Tekoälyn kehittäminen	Tässä raportissa: tekoälytekniikan tutkimus, kehitys ja innovaatiotoiminta. Oleellisia kysymyksiä ovat mm. tekoälyn luotettavuus, vastuuvollisuus, läpinäkyvyys, selitettävyys ja tietosuojat.
Ohjaus	Tässä raportissa: erilaiset toimintaa jäsentävät suunnittelun, päätöksenteon, ohjauksen ja organisoinnin muodot. Katso kansainvälisessä kirjallisuudessa käytetty käsite ”governance”.
Tekoälyn ohjaus	Tässä raportissa: tekoälykehitykseen vastaavat ja tekoälyn kehittämistä jäsentävät suunnittelun, päätöksenteon, ohjauksen ja organisoinnin muodot.
Vastuullinen tutkimus- ja innovaatiotoiminta	Responsible Research and Innovation (RRI); Tutkimus- ja kehitystoimintaa ja tuloksia kehitetään ja arvioidaan vastaamaan yhteiskunnallisia arvoja, tarpeita ja odotuksia yhteistyössä tutkijoiden ja sidosryhmien kanssa. RRI-periaatteita ovat ennakointi, osallistaminen, itserefleksio ja valmius muuttua.

1. Johdanto

Valtiovarainministeriön tietopoliittinen selonteko³ (2018) nosti hyvän tekoäly-yhteiskunnan arvoiksi *läpinäkyvyyden, vastuullisuuden ja laajan yhteiskunnallisen hyödyn*. Sen mukaan tulevaisuuden tekoälyyn pohjautuvat teknologiat tulee suunnata tukemaan elämänlaadun parantamista, helpottamaan palveluiden saatavuutta ja käyttöä sekä tehostamaan kilpailukykyä ja arvonlisäystä. Suomesta tavoiteltiin Sipilän hallituksen johdolla kestäväan tietoe kosysteemin esimerkkimaata, jonka nähtiin nojautuvan luottamusyhteiskuntaan, hyvään hallintoon, korkeaan osaamistasoon, rekistereihin ja tietovarantoihin, innovatiivisiin yrityksiin sekä digitalisaation ja tekoälyn vahvaan kehitykseen.

Suomen kansallinen tekoälystrategia laadittiin Tekoälyaika -ohjelmassa, joka päättyi 2019. Tekoäly 4.0 ohjelma käynnistyi marraskuussa 2020, jatkumona 2019 päättyneelle Tekoälyaika-ohjelmalle Suomen tekoälyn käyttöönoton vauhdittamiseksi.

Näistä lähtökohdista aloitettiin ETAIROS-tutkimushanke 2019. Tekoälyn nopean kehittämisen ja laaja-alainen soveltamisen nähtiin vaikuttavan kaikkiin yhteiskunnan osa-alueisiin, ja sen käytön nostavan esiin monia eettisiä, yhteiskunnallisia ja juridisia ongelmia. Siksi tarvittiin tutkittua, empiiristä tietoa tekoälyjärjestelmistä ja niiden yhteiskunnallisista vaikutuksista sekä uusia lähestymistapoja ohjaamaan jatkuvasti kehittyvien teknologioiden kehitystyötä ja sovelluksia niin julkisella kuin yksityisellä sektorilla (Kuva 1).



Kuva 1. ETAIROS-hankkeen tutkimuskysymykset.

³ Valtiovarainministeriö 2018

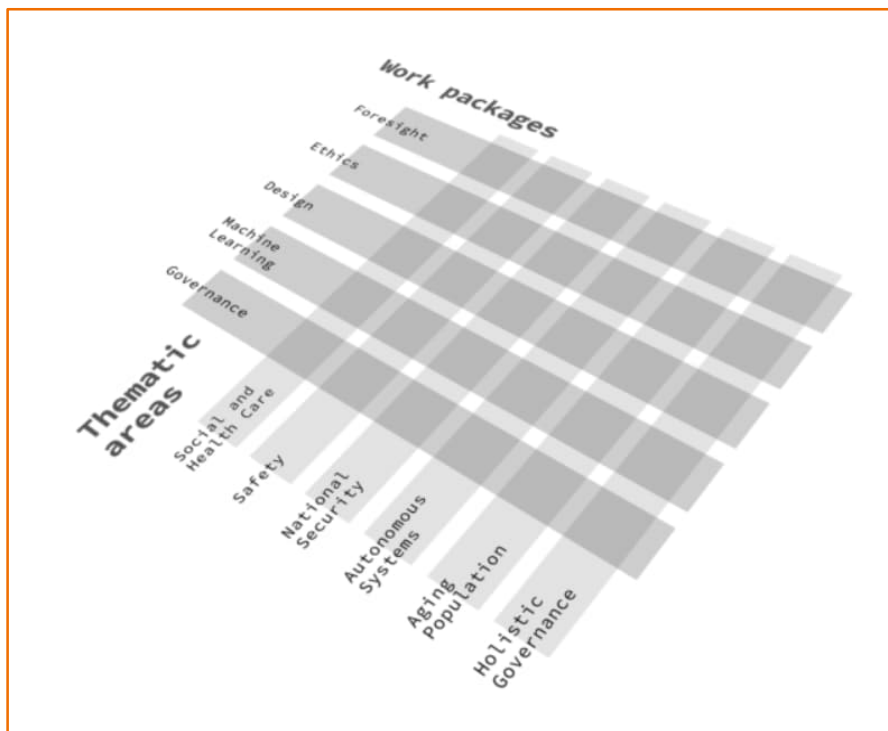
Edellä esitetyt lähtöoletukset osoittautuivat oikeiksi hankkeen aikana ja erityisesti hankkeen toiseen rahoituskauteen osunut generatiivisen tekoälyn (mm. suuret kielimallit) läpilyönti nosti merkittävästi hankkeen keskeisten kysymysten yhteiskunnallista relevanssia.

Haaste nähtiin monitasoisena ja moniulotteisena hallintotavan kysymyksenä: On tärkeää muodostaa eri sektoreiden välinen yhteisymmärrys siitä, miten tekoälyä käytetään ja säännellään. On myös varmistettava hyvä koordinaatio hallinnon eri tasojen välillä sekä löydettävä tasapaino pehmeiden ja kovien hallintomekanismien välillä erilaisissa tekoälyn käyttökonteksteissa.

ETAIROS yhdisti yllä olevat neljä tutkimushaastetta kattavaksi monialaiseksi hankkeeksi, jonka tavoitteena oli:

- Kiinnittää suunnitteluhaasteet ja yhteiskunnalliset huolenaiheet empiirisiin tutkimuksiin.
- Käyttää ennakointimenetelmiä tekoälykehityksen yhteiskunnallisten vaikutusten arvioimiseksi ja integroimiseksi teoreettiseen ja käytännön tutkimukseen.
- Sisällyttää mukaan hallintoon liittyvät näkökohdat, jotta voidaan tarjota politiikan kannalta merkityksellisiä ehdotuksia ja käytännönläheisiä lähestymistapoja tekoälyn liittyvien tulevien toimien suunnitteluun.
- Kehittää yhteiskunnallisesti kestäviä ja käytännönläheisiä ratkaisuja yhteistyössä sidosryhmien (yksityinen, julkinen ja kolmas sektori) ja kansalaisten kanssa yhteisten työpajojen ja avoimen vuoropuhelun foorumien avulla.

Projektin rakenne koostui työpaketeista ja toisiinsa liittyvistä aihealueista (kuva 2).



Kuva 2. ETAIROS-projektin rakenne ja temaattiset alueet.

VTT keskittyi seuraaviin haasteisiin:

- Sovellettava ja toiminnallinen lähtökohta tekoälyn etiikkaan (vrt. sääntö tai periaatepohjainen etiikka)
- Kansalliset hallinnoinnin mekanismit ja eettisen ja vastuullisen hallinnoinnin kehikot
- Eettisyys, riskit, mahdollisuudet ja heikot signaalit käyttötapauksissa
- Mahdolliset tekoälytulevaisuudet: skenaariokehys erilaisista maailmoista
- Julkishallinnon kyvykkyys soveltaa eettistä ja vastuullista tekoälyn hallintaa
- Sääntely-ympäristö ja tekijät, eettistä suunnittelua; tekoälyn käynnistämät sääntelyn muutospainet

VTT keskittyi tutkimuksessaan erityisesti autonomisen liikenteen, tekoälyn julkisten hallintomekanismien, organisatoristen jännitteiden, osallisuutta tukevien DemTech-alustojen ja ikääntyvän väestön tekoälytarpeiden tarkasteluun sekä tulevaisuuksien ennakointiin uusien tekoälyteknologioiden näkökulmasta (Kuva 3).



Tekoälyn etiikka autonomisessa liikenteessä

Arvojen, vaikutusten ja käyttäjien hyväksynnän ymmärtäminen.



Kollektiivinen älykkyys ja DemTech -osallistumisalustat
Miten demokratiateknologiaa käytetään?

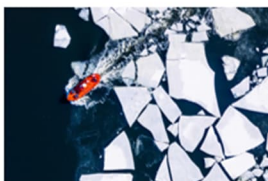


Kestävä tekoälyn hallintomalli

Kattava, osallistava, institutionalisoitu, toteuttamiskelpoinen.



Tulevaisuuden ennakointi
Metaverse ja uudet teknologiat: Eettinen vuoropuhelu ja yhteiskehittäminen.



Organisatoriset jännitteet

Tekoälyn käytön tuomat jännitteet; miten ne ratkaistaan?



Tekoäly ja ikääntyvä väestö
Kasvavat hopeamarkkinat ja tekoälyn eettiset näkökulmat.

Kuva 3. ETAIROS@VTT tutkimushaasteita.

2. Tekoälyn etiikka

2.1 Etiikka kontekstuaalisena ja toiminnallisena käsitteenä

ETAIROS-hankkeen keskiössä olivat kysymykset siitä, miten eettisesti hyväksyttäviä tekoälyjärjestelmiä suunnitellaan, käytetään ja hallitaan – haasteet, jotka koskettavat yhä laajemmin sekä julkista että yksityistä sektoria. Haasteita, joiden merkitys kasvaa nopean ja osin arvaamattoman yhteiskunnallisen ja teknisen kehityksen ristivedossa.

Hankkeen elinkaaren aikana käydyt keskustelut heijastavat nopeaa muutosta. Ensin keskityttiin kysymyksiin vinoutuneesta datasta ja automaattisesta päätöksenteosta, sitten myrkyttyneestä informaatioympäristöstä, tutkijoiden ja organisaatioiden riippuvuudesta tekoälystä, valtakysymyksistä ja teknologiajättien vallasta, sääntelyn mahdollisuuksista, taloudellisesta ja episteemisestä eriarvoisuudesta, älykaupungeista, lopulta kognitiivisesta turvallisuudesta, vaikutuksista ihmisyyteen ja ympäristöön.

Projektin aikana kävi selväksi, että teknologinen kehitys tuottaa jatkuvasti uusia kysymyksenasetteluja ja samalla tekee vanhoista vastauksista ja keskusteluista osin hyödyttömiä. Esimerkiksi sääntelykehityksessä muutos on johtanut siihen, että osa hallintakeinoista, joita keskusteluissa ja lainsäädäntöhankkeissa on kehitelty, eivät enää toimi. Tekoäly uutena, vieraana objektina on tuonut esille rakenteita, jotka ovat muutoin itsestään selviä ja joita ei tulisi mielenkään etsiä.

ETAIROS- työskentely ja ennen muuta ympäröivät haasteet osoittivat väkevästi, että tekoälyn etiikka ei ole eikä voi olla erillinen (akateeminen) saareke, vaan sen on oltava käytäntölähtöistä ja toimija- ja toimintakeskeistä työskentelyä, jossa keskeiset kysymyksenasettelut eivät kumpua ensisijaisesti filosofian traditioista vaan käsillä olevista ongelmista. Samanaikaisesti klassiset filosofian ongelmat, kuten Humen giljotiini⁴ ovat näyttäytyneet uudella tavalla ja löytäneet uusia ratkaisumalleja tekoälyn valossa.

Filosofisessa etiikassa usein käsiteltyt yleiset hyvän, pahan, oikean ja väärän hahmotelmat toimivat vain suuntaa antavina periaatteina, joiden soveltaminen käytännössä luo oman haastekenttensä. Koska konteksti muokkaa vaikutuksia ja seurauksia eri sidosryhmille, on teknologian ja tekoälyn etiikka kontekstisidonnaista, jota ei voida tarkistuslistoilla tai yleisillä periaatteilla ratkaista. Käytännön etiikan tulisi kulkea esimerkiksi hankkeissa ja projekteissa sen koko elinkaaren läpi tietoisena siitä, että myös yhteiskunnat, politiikka, lainsäädäntö ja teknologia kehittyvät ja uusia eettisiä kysymyksiä voi ja usein syntyykin.

2.2 Ihmiskeskeisyys kehittämisen keskiössä

Yhteiskunnallisen muutoksen keskusteluissa ihmiskeskeisyys nostetaan usein tavoitteeksi, mutta mitä sillä oikeastaan tarkoitetaan ja toteutuuko se käytännössä? Millainen arvopohja ja toimijuuteen perustuva etiikka voisivat ohjata tulevaisuuden tekoälyä? Millaisia hallintokehyksiä, -käytäntöjä ja valmiuksia tarvitaan eettisesti kestäväen tekoälyn edistämiseksi?

Tekoälysovellusten ihmiskeskeisyys nähdään organisaatioissa suhteellisen kapeasti yksityisyyden suojan ja datan hallinnan kysymyksinä. Huolenaiheeksi koetaan datan vinoutuminen sekä dataan liittyvien tulkintojen oikeellisuus. Suurimpina riskeinä pidetään tietoturvaa ja yksityisyyttä koskevia tekijöitä ja kontrollin lisääntymiseen, syrjintään, tasa-arvoon sekä kansalaisten vaikutus-

⁴ Saariluoma & Leikas 2020

mahdollisuuksien kapenemiseen liittyviä riskejä. Luonnollisesti myös toiminnan laillisuus sekä tekninen luotettavuus ja turvallisuus koetaan tärkeiksi. Vastuuvollisuuden toteutumisen keinoiksi nähdään teknisten menetelmien varmistaminen ja järjestelmän toiminnan testaus, sekä kansalaisten muutoksenhaun mahdollisuus.

ETAIROS määritteli ihmiskeskeisyyden osallisuuden kautta⁵. Tässä lähestymistavassa ihminen ei ole toiminnan ja suunnittelun ”keskiössä” vain ajatuksen tasolla, tai passiivisena kohteena, vaan osallistuvana toimijana. Oletuksena on, että kansalaisten ja sidosryhmien osallistumiseen ja osallisuuteen perustuva järjestelmällinen lähestymistapa sekä vahvistaisi yleistä luottamusta tekoälyjärjestelmiin ja niitä hallinnoiviin instituutioihin, että parantaisi kollektiivista päätöksentekoa tekoälyyn liittyvissä politiikkatoimenpiteissä, minkä ansiosta julkishallinto voisi tukea teknologian kehittämistä yhteisen edun nimissä. Näin ollen ihmiskeskeisyys ei rajoitu käyttäjälähtöisyyteen, vaan asettaa osallistajat aktiiviseen asemaan ja osaksi keskustelua.

Yhteistyö- ja kansalaisteknologian välineet voivat auttaa toteuttamaan systeemisen ja osallistavan lähestymistavan, sillä ne voivat mahdollistaa laajamittaisen sitoutumisen ja osallistumisen päätöksentekoprosesseihin. Sidosryhmien ja kansalaisten osallistumisteknologioilla on mahdollisuus muuttaa päätöksenteon luonnetta. Kansalaisvaikuttamiseen liittyvät kokeilut tarjoavat kokemusta sellaisista konkreettisista menetelmistä, joilla voidaan kartoittaa erilaisten sidosryhmien tarpeita, tavoitteita ja näkemyksiä tavoilla, jotka eivät aiemmin olleet mahdollisia. Vaikka tällaisia osallistumisen välineitä ei lupaavista tuloksista huolimatta ole onnistuttu hyödyntämään osana demokraattisia perusprosesseja, kokeilut auttavat hahmottamaan uusien kansalaisvaikuttamismenetelmien hyötyjä esimerkiksi ihmiskeskeisen palvelusuunnittelun toteuttamisessa sekä julkisen hallinnon edellytyksiä ottaa uusia menetelmiä käyttöön.

⁵ Sigfrids et al. 2023A

3. Miten ohjata sosioteknistä muutosta kohti eettisesti kestäviä tuloksia?

3.1 Tekoälyn vastuullinen hyödyntäminen Suomen kilpailukyvyyn ja hyvinvoinnin edistämiseksi

Kun puhumme yhteiskunnan ohjauksesta ja tekoälyn sääntelystä, tarkoitamme instituutioita, prosesseja ja sovelluksia sekä erilaisten sääntöjen, asetusten, standardien, menettelyjen ja päätösten tuloksia, jotka tarjoavat puitteet tekoälyn eettiselle kehittämiselle ja käytölle. Millaisia keinoja yhteiskunnalla on tekoälysovellusten kehittämisen ja käytön sääntelyyn? Millaisia ohjeita ja työkaluja tekoälyn soveltamiseen tarvitaan? Miten tekoälysovellusten avulla voidaan palvella kansalaisia?

Nykyiset julkiset ohjauskeinot eivät ehkäise tekoälyn käytön nostattamia sosiaalisia riskejä ja uhkia riittävästi. Julkishallinnon tulisi kehittää kyvykkyyttä ennakoida ja tunnistaa tekoälyn riskejä kokonaisvaltaisesti sekä vastata syntyviin ongelmiin ketterällä tavalla. Siksi tarvitaan tekoälyn välillisten vaikutusten esiin nostamista ja demokraattista keskustelua siitä, millaisissa tekoälytulevaisuuksissa haluamme elää, sekä toimia näiden tulevaisuuksien rakentamiseksi.

Tekoälysovellusten käyttöönotto muuttaa päätöksenteon muotoja ja nostaa esiin uudenlaisia vastuukysymyksiä, jotka tulisi pystyä huomioimaan jo tekoälysovellusten suunnittelussa. Alustavia tekoälyn suunnittelun ja käyttöönoton periaatteita on kyllä muotoiltu, mutta niiden soveltamista vaikeuttaa selkeiden käytäntöjen puute. Tämä teorian ja käytännön välinen kuilu on nähtävissä tekoälyn etiikan pohdinnassa. Euroopan tasolla tuotetuilla ohjeistuksilla ei ole juurikaan vaikutusta kehittäjien käytännön työhön. Juridisista säännöksistä ollaan tietoisia, mutta esimerkiksi kansalaisten ottaminen mukaan kehitystyöhön on vähäistä tai lähes olematonta. Ymmärretään kyllä, että kehitystyön tulee olla "ihmiskeskeistä" ja "käyttäjälähtöistä", mutta hyväksyttävyyden ja haluttavuuden näkökulman tuominen käytännön toimenpiteinä kehitystyöhön puuttuu.

Suomi on edennyt digitalisaatiokehityksessä melko lailla teknologiapainotteisesti. Etenemistapa on ollut erilainen kuin esimerkiksi Japanin Society 5.0-kehityksessä, jossa painopiste ei ole pelkästään teknologisessa vaan myös sosiaalisessa kehityksessä: Japanin verkottunut yhteiskunta suunnittelee ja integroi digitaaliset välineet ihmisten arjesta käsin: mitä ihmiset tarvitsevat ja miten he asemoivat itsensä asiakkaina ja hyvinvointipalvelujen käyttäjinä⁶. Tätä kehitystavoitetta on kuvailtu älykkään ja jopa superälykkään yhteiskunnan käsitteen avulla. Älykkäällä yhteiskunnalla tarkoitetaan visiota yhteiskuntajärjestelmästä, jossa älykästä teknologiaa käytetään ihmisen toiminnan apuvälineenä tukemaan esimerkiksi työvoimapulan korvaamista, ikääntyvän väestön hoivaa, ilmastomuutoksen ja luontokadon torjuntaa tai uusiutuvaa energiaa tuottavia järjestelmiä.

Suomessa on kehitetty valtiotason tekoälystrategiaa, teollisuus 5.0 strategioita (Industry 5.0), älykaupunki -hankkeita sekä kaupunkitason metaverse -hankkeita. Näiden pohjalla on samankaltaista ajattelua, kuin esimerkiksi Japanin Society 5.0 ohjelmassa, mutta toteutuksen suunnittelu on pistemäistä, jos sitä vertaa tavoitteiden toteutumisen edellytyksiin.⁷ Työ- ja elinkeinoministeriön Tekoäly 4.0 -ohjelma pyrkii puolestaan edistämään neljättä teollista vallankumousta, vahvistamaan digitaalisuutta ja talouskasvua keskittymällä erityisesti tekoälyn ja muiden digitaalisten teknologioiden kehittämiseen ja käyttöönottoon yrityksissä. Tavoitteena on nostaa Suomi tekoälyn kehittämisessä kansainväliseen kärkeen. Ohjelman näkökulma, samoin kuin tekoälystrategiainkin, on yritysten tuottavuuskehityksessä kestävä kehitys hengessä. Ne eivät

⁶ Bartoloni et al. 2022; Fukuyama 2018

⁷ Salo-Pöntinen & Saariluoma 2022

luo suuntaa toivottavalle tekoälykehitykselle yhteiskunnallisella tasolla, eivätkä kuvaa ehtoja tai tarjoa eväitä julkishallinnolle älykkään yhteiskunnan strategiseen, ihmiskeskeiseen ja eettiseen kehittämiseen.

Kuten tietopolitiikan yhteistyöryhmän käsikirja tekoälystä päätöksentekijöille⁸ toteaa, Suomen tekoälystrategia on päivitettävä ajantasaiseksi. Sen lisäksi että Suomen ja Euroopan asemaa suhteessa tekoälyyn tulee vahvistaa, on myös olennaista hahmottaa tekoäly kehittyvien teknologioiden raja-alueena, jolla tulee olemaan merkittävä rooli siirtymässä kohti älykästä yhteiskuntaa. Tämä tarkoittaa älykkään yhteiskunnan määrittelyä strategisella tasolla – osana tekoälystrategiaa, jossa vastuullisesti toimiva julkinen hallintotapa luotsaa tekoälyn kehittäjiä ja käyttäjiä siten, että tekoälyn avulla voidaan tuottaa yhteistä hyvää niin yksilöiden, yhteisöjen kuin koko yhteiskunnan kannalta.

Vaikka lähtökohtaisesti Suomessa on hyvät valmiudet ja voimavarat sekä kehittää että ottaa käyttöön tekoälysovelluksia, on tärkeää korostaa, että Suomelta puuttuu tässä hetkessä koherentti yhteiskunnallinen tekoälystrategia. Tämän tarpeen tunnistaminen ja toimien käynnistäminen myös etiikan näkökulmasta on ensiarvoisen tärkeää, mikäli haluamme vaikuttaa muutoksen suuntaan: esimerkiksi EU-tason sääntely ja ohjaus eivät ratkaise sellaisenaan suomalaisen yhteiskunnan erityiskysymyksiä.

Odotettavissa olevia hyötyjä voidaan kasvattaa ja riskejä minimoida koordinoimalla eri politiikkatoimia nykyistä selvemmin. ETAIROS suositteli julkaisemassa *Politiikkasuosituksessa*⁹ viittä politiikkatoimea tekoälyn laaja-alaiseen hyödyntämiseen ja yritysten liiketoiminnan edellytysten vahvistamiseen:

1. Tekoälypolitiikan valtakunnallisen neuvottelukunnan perustaminen. Neuvottelukunnan tehtävänä tulisi olla eettisesti kestävä ja vastuullisen tekoälypolitiikan muotoileminen, toimenpidesuosittelujen antaminen sekä niiden koordinoiminen ja seuranta.
2. Tekoälyn kasvuohjelman laatiminen. Kasvuohjelman tavoitteena tulisi olla teknologisen kehityksen vastuullinen hyödyntäminen, uusien kasvuyritysten tukeminen, palveluiden uudistaminen sekä nämä tavoitteet mahdollistava lainsäädäntö.
3. Monialaisen yhteistyön vahvistaminen ja ekosysteemien kehittäminen. Yrityksiä ja julkisia toimijoita yhdistäviä palvelu-, liiketoiminta- ja innovaatioekosysteemejä tulisi vahvistaa. Ekosysteemit auttavat tunnistamaan ja saattamaan yhteen toisiaan täydentävää osaamista, nopeuttamaan innovaatioiden ja palveluiden luomista ja edistämään uusien kasvuyritysten perustamista.
4. Tekoälyn vaikutusten ennakkoinnin ja arvioinnin tehostaminen. Mahdollisia hyötyjä ja riskejä tulisi arvioida niin etu- kuin jälkikäteen. Samalla kun tekoälyratkaisujen ennakkointia kehitetään, on luotava mahdollisimman yksikäsitteisiä laadunarviointikriteereitä ja vaikutusten seurantamittareita.
5. Alan koulutuksen kehittäminen ja osaamisen vahvistaminen. Keskeisiä toimia tulisi olla laaja-alainen peruskoulutuksen uudistus sekä osaamisen kehittymisestä huolehtiminen niin toisen asteen koulutuksessa, korkeakouluissa kuin elinikäisen oppimisen ratkaisuissa. Lisäksi osaamisen syventäminen tulisi liittää osaksi korkeakoulujen ja oppilaitosten koulutus- ja tutkimusstrategioita.

⁸ Poikola ym. 2024

⁹ Koskimies ym. 2020

3.1.1 CASE OSKI-malli - Julkishallinto tukemassa eettisen tekoäly-yhteiskunnan rakentamista

Laadimme systemaattisen kirjallisuuskatsauksen perusteella¹⁰ tekoälyn julkisen ohjauksen tukemiseksi tarkoitetun OSKI-mallin (Osallistava, Soveltuva, Kokonaisvaltainen ja Institutionalisoitu), sekä malliin liittyvät toimenpide-ehdotukset. Tutkimus perustui ajatukseen olemassa olevien ohjauskeinojen paremmasta soveltamisesta ja edelleen kehittämisestä, jotta julkiset hallintoelimet voivat vastata tehokkaasti tekoälyyn liittyviin haasteisiin samalla varmistaen, että tekoäly hyödyttää yhteiskuntaa laajasti ja tuottaa yhteistä hyvää.

Vaikka lainsäädäntö ja tekniset standardit muodostavat tärkeän osan ohjauskeinoista, ne eivät yksinään riitä ohjaamaan tekoälykehitystä yhteiskunnallisesti tarkoituksenmukaiseen ja hyödylliseen suuntaan. Siksi tarvitaan jatkuvaa eettistä pohdintaa ja arviointia.

Selvityksessä analysoitiin 21 tutkimusartikkelin ehdotukset tekoälyn julkisen ohjauksen kehittämisestä.

Tekoälyn julkisen eettisen ohjauksen tulisi hyödyntää etiikkaan, ihmisoikeuksiin, hyvään hallintoon sekä vastuulliseen tutkimukseen ja innovointiin perustuvia normatiivisia periaatteita politiikanteossa. Ohjauksen tulisi perustua osallistaviin ja joustaviin hallintomekanismeihin ja -käytäntöihin, tarkoituksena tukea päätöksentekoa ja sopeutumiskykyä lyhyen ja pitkän aikavälin tekoälyn yhteiskunnallisiin haasteisiin. Julkisessa päätöksenteon valmistelussa tulisi tiedostaa ohjauskeinojen kehystäminen ja olla refleksiivinen sen suhteen, kuka osallistuu ja pääsee vaikuttamaan tekoälyä koskeviin strategioihin ja päätöksiin. Tekoälyn ohjauksen sosiaalisesti kestävä ja pitkäaikainen kehittäminen edellyttävät institutionalisoitua ja koordinoitua päätöksenteon rakennetta, joka tuottaa tietoa tekoälyn vaikutuksista ja sidosryhmänäkemyksistä, koordinoi tekoälyn ohjausta, ja valvoo normatiivisten periaatteiden toteutumista, tekoälysovellusten turvallisuutta ja riskienhallintaa.¹¹

OSKI-malli kuvaa ideaalitilaa, joka tarjoaa suuntalinjoja kokonaisvaltaiseen mutta joustavasti käyttötilanteittain soveltuvaan päätöksentekoon (Kuva 4).

Sigfrids, A., Nieminen, M., Leikas, J., Karvonen, A., & Pikkuaho, P. (2023). *Julkishallinto tukemassa eettisen tekoäly-yhteiskunnan rakentamista: Katsaus tekoälyn ohjauskeinoin ja politiikkatoimenpiteisiin*. VTT Technical Research Centre of Finland. VTT Technology No. 421. <https://doi.org/10.32040/2242-122X.2023.T421>

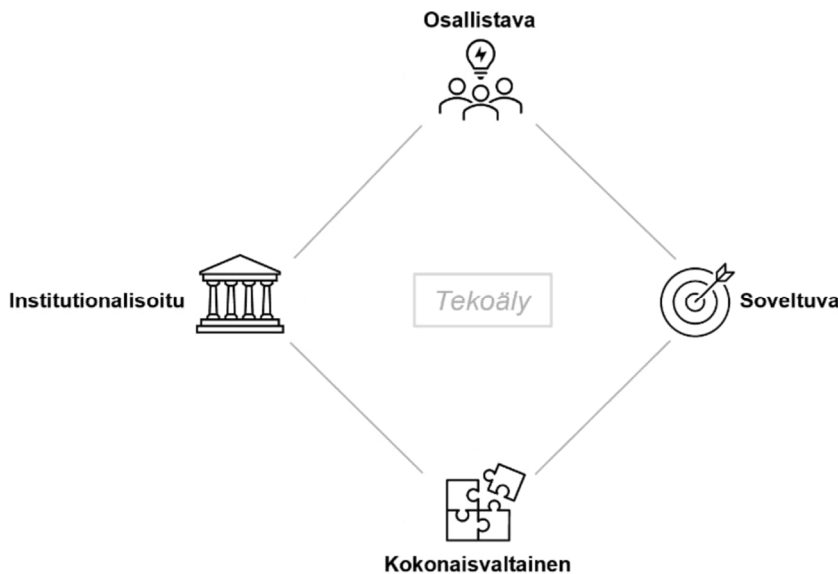
Sigfrids, A., Karvonen, A., Rilla, N., & Lanne, M. (2024). *Governing AI services: a state-of-the-art overview*. Handbook of Services and Artificial Intelligence. Edward Elgar Publishing Ltd, p. 308-329 21 p.

Sigfrids, A., Nieminen, M., Leikas, J., & Pikkuaho, P. (2022). *How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI*. Front. Hum. Dyn. 4:858108. <https://doi.org/10.3389/fhumd.2022.858108>

Koskimies, E., et al. (2020). *ETAIROS Policy Brief: Suomeen tarvitaan koordinoitu, vastuullisuutta korostava tekoälypolitiikka*.

¹⁰ Sigfrids et al. 2022

¹¹ Sigfrids et al. 2023B



Kuva 4. Neljä ideaalia tekoälyn julkiselle ohjaukselle ja koordinaatiolle. (Sigfrids et al. 2023B)

Osallistava. Tekoälyn haasteiden hahmottaminen ja ohjauksen kehittäminen edellyttää yhteistyötä ja yhteiskehittämistä eri sidosryhmien, kuten tutkijoiden, yritysten, hallinnon virkamiesten, kansalaisyhteiskunnan ja kolmannen sektorin kesken. Yhteisten ratkaisujen löytämiseksi tulee arvioida toimenpiteiden vaikutuksia eri sidosryhmille sekä luoda edellytykset käydä keskustelua mahdollisista eturistiriidoista ja eri arvojen toteutumisesta. Eri sidosryhmien näkemysten ja arvojen huomioiminen vuorovaikutteisissa prosesseissa on ketterien ja sosiaalisesti hyväksyttävien ohjauskeinojen ja eettisten periaatteiden suunnittelun ja toimeenpanon lähtökohta. Monialaista yhteistyötä voidaan vahvistaa kehittämällä vastuullisen tekoälyn ekosysteemejä sekä perustamalla toimintaa tukeva organisaatio tai elin.

Soveltuva. Yhteiskunnallisesti tavoiteltavien, kestävien ja eettisesti hyväksyttävien ratkaisujen löytymiseksi voidaan soveltaa vastuullisen tutkimus- ja innovaatiotoiminnan (eng. Responsible Research and Innovation, RRI) käytäntöjä, sekä jo olemassa olevia tekoälyn etiikkaa ja vastuullisuutta tukevia suunnittelumenetelmiä ja työkaluja. Lisäksi pitää tehostaa tekoälyn vaikutusten ennakkointia ja arviointia. Eri sidosryhmien kyvykkyyttä soveltaa näitä työkaluja ja arvioida tekoälyn eettisiä ulottuvuuksia tulee tältä osin tukea esimerkiksi kehittämällä alan koulutusta. Tarvitaan myös konkreettisia käytännön ohjeita ja keinoja eettisten periaatteiden ja ihmisoikeuksien huomioimiseksi tekoälyä koskevassa päätöksenteossa.

Kokonaisvaltainen. Parhaiden ratkaisujen löytämiseksi ja uusien kehittämiseksi tarvitaan muutokseen vaikuttavien tekijöiden kokonaisvaltaista ymmärtämistä sekä tekoälyn systeemisen luonteen ja pitkän aikavälin vaikutukset huomioiva näkökulma. Kokonaisvaltaisen näkökulman omaksuva julkinen hallinto huomioi esimerkiksi tekoälyn sovellusalueet, eettiset ja lainsäädännölliset kysymykset sekä keskeisten toimijoiden näkökulmat, arvot ja intressit. Ratkaisut tulee sovittaa niin pitkäaikaisiin strategioihin kuin paikallisiin tarpeisiin, ja niiden kehittämisessä tulisi huomioida tekoälyn kehittäjien, loppukäyttäjien sekä päätöksentekijöiden näkökulmat.

Institutionalisoitu. Osallistavan päätöksenteon valmistelun järjestäminen tulisi olla valtakunnallisen elimen vastuulla, joka koordinoi tekoälypolitiikkaa ja kehittää ohjauskeinoja, kouluttaa sidosryhmiä, kerää tarpeellisen tiedon ja tuo eri sidosryhmät keskinäiseen vuoropuheluun. Lisäksi tarvitaan vakiintunutta viranomaistahoa valvomaan ja varmistamaan lainsäädännön noudattamista ja tukemaan eettisten normien toteutumista.

3.2 Sosioteknisen muutoksen ennakointi

Teknologian tulevaisuuden kehittymisen ja käytön ennakointi tähtää siihen, että varaudumme teknologian mahdollisesti aiheuttamiin riskeihin yhteiskunnassa, ja luomme teknologian käytölle ja kehittämiselle mahdollisuuksia ja uusia käyttötarkoituksia.

Ennakoinnin menetelmistä ETAIROS:ssa hyödynnettiin horisonttitarkastelua (horizon scanning) ja skenaariomenetelmää. Horisonttitarkastelulla tunnistettiin, tarkasteltiin ja yritettiin ymmärtää nousevia trendejä, mahdollisia ongelmia sekä pitkän aikavälin kehityssuuntia liittyen tekoälyn käyttöön yhteiskunnassa. Käytännössä tämä tarkoitti esimerkiksi uutisissa ja muussa mediassa esiin nostettujen teemojen tarkastelua niin, että käsittelyyn nousivat myös heikoiksi signaaleiksi laskettavat teemat. Heikoilla signaaleilla tarkoitetaan ilmiöitä, jotka eivät vielä ole valtavirtaa. Poimituista teemoista kirjoitettiin blogeja ETAIROS:in verkkosivuille (signal posts¹²). Horisonttitarkastelun yhteydessä tunnistettiin samalla avaintekijöitä tekoälytulevaisuuden sosio-tekni- skenaarioiden rakentamiseen.

Horizon scanning -prosessin tuloksena kirjoitimme kuukausittain signal post -kirjoituksen, jossa esiteltiin tulevaisuutta luotaava, teknologian mahdollisuuksia kuvaava tai huolta aiheuttava signaali. Signaalit poimittiin mediassa esitetyistä aiheista. Tekoälyteknologian tulevaisuuden kehittämiseen ja soveltamiseen vaikuttavien signaalien arviointi suomalaisen yhteiskunnan näkökulmasta edistää osaltaan tekoälyn kehittämistä ihmistä palvelevaksi ja ihmisystävälliseksi.

Tekoäly voi olla merkityksellinen työkalu ihmiskuntaa kohtaavien uhkien tunnistamisessa ja ymmärtämisessä. ETAIROS Signal Post -kirjoituksia innoittaneet signaalit ovat vuosien 2020–2023 aikana olleet esimerkiksi tekoälyn suhde pandemiaan ja Covid-tautiin. Tekoälyteknologiaa on sovellettu pandemian etenemisen ennustamiseen. Samantapaisesti tekoälyteknologiaa on käytetty ilmastonmuutoksen etenemisen ja vaikutusten ennustamiseen.

Toinen ETAIROS signaalikokonaisuus liittyy tekoälyalgoritmien hyödyntämiseen ihmisen elämän mallintamisessa ja suuntaamisessa. Esimerkkinä mainittakoon kansainvälisessä oppilaitoksessa käytetty algoritmi, joka laski oppilaille koulun päättöarvosanan koko kouluajan tulosten ja arvioinnin perusteella. Tulokset, joiden syntyyn tekoälyllä oli merkittävä ohjaava vaikutus, ohjasivat oppilaiden elämää jatko-opiskelupaikkaoikeuden saavuttamisen kautta. Toinen esimerkki on Suomen kansallinen AuroraAI-järjestelmä, jossa tekoäly pystyy ohjaamaan ihmisen valintoja esimerkiksi suosittelemalla lääkärillä käyntiä kerätyn tiedon pohjalta.

Kolmas signaalikokonaisuus liittyy tekoälyyn ja liike-elämään. Vaikka tekoälyä sovelletaan jo yritystoiminnassa, suomalainen pk-teollisuus ei ole omaksunut tekoälyä vielä laajasti hyödykseen. Tekoälyn eettiseen soveltamiseen on haastettu yrityksiä muun muassa Työ- ja elinkeinoministeriön toimesta. Etiikka ja eettisyyden soveltaminen ovat kompleksisia asioita, joiden huomioiminen ei koskaan liene helppoa, ei myöskään liiketoimintamallien osalta.

Neljäntenä signaalikokonaisuutena hahmottuvat algoritmit ja ihmiset. Tekoälyn suunnittelussa, luomisessa ja käytössä tulisi pohtia kansalaisten roolia. Tietävätkö kuluttajat mitä haluavat tekoälysovelluksilta? Mikä on kuluttajamanipulaation rooli niissä tarpeissa, joihin tekoälyä suunnitellaan? Miten markkinoiden vahvat toimijat tosiasiallisesti ohjaavat kuluttajavalintoja ja mitkä ovat pienempien markkinatoimijoiden mahdollisuudet? Miten kuluttajat voivat osallistua tekoälyn suunnitteluun, kehittämiseen ja käyttöön. Algoritmit ja ihminen -kokonaisuuteen sisältyy myös tekoälyn turvallisuus ja riskien hallinnointi. Otsikoihin nousi esimerkiksi vuoden vaihteessa 2020–2021 lapsille tarkoitettu tekoälysovellus, jonka käyttö altisti lapset vaaralliselle sähköiskulle. Kuinka tekoälystä saadaan turvallinen kaikille?

¹² <https://etairos.fi/category/signal-post/>

Merkittävä ilmiö tekoälyteknologian kehityksessä on virtuaalisen todellisuuden kehittyminen. Esimerkiksi Facebook julistautui vuoden 2021 lopussa metaversen kehittäjäksi. Metaverse tuo kansalaisten maailmaan avatarit ja digitaaliset kaksoset. Kehityskuvauksissa nämä ohjelmoidut toisinnot henkilöistä lähetetään esimerkiksi kokoukseen varsinaisen henkilön puolesta.

Kansainvälinen turvallisuustilanne siivitti vuoden 2022 signaalien osalta tarkastelemaan tekoälyn käyttöä rauhan, tiedustelun ja sodan palveluksessa. Pohdimme myös tekoälyn eettisyyttä kahdesta eri näkökulmasta: kuka voi olla tekoälyn eetikko ja kuinka etiikka huomioidaan Suomen Tekoäly 4.0 ohjelmassa. Näkemyksemme on, että tekoälysovellusten tekemisessä ja käytössä tulee huomioida eettisiä kysymyksiä jo mahdollisimman varhaisessa suunnittelun vaiheessa. Näin syntyy vastuullisesti suunniteltu ja toteutettu tekoälyteknologia.

Joulun alla 2021 sekä 2022 tarjosimme sidosryhmillemme joulukalenterin Twitterissä, joka avasi jokaisena joulukuun päivänä ennen joulua yhden signaalin tekoälyn kehityksen kentästä. Vuoden 2023 signaalikirjoituksissa käsitelimme ChatGPT-kielimallin käyttöä opetuksessa. Useat oppilaitokset ovat jo tehneet ohjeistuksia kielimallien käytöstä. Keskeistä on, että kielimallin käytöstä kerrotaan avoimesti ja oppilaille opetetaan siihen liittyviä mahdollisuuksia ja riskejä. Keväällä 2023 Suomessa olleiden vaalien aikaan julkaisimme signaalikirjoituksen tekoälypuolueista. Tämä kirjoitus muokattuna päättyi Helsingin Sanomien mielipidekirjoitukseksi 29.3.2023: "Tekoälypuolueet raikastaisivat demokratiaa".

ETAIROS Signal Post -kirjoituksiin innostaneet signaalit haastavat tekoälyn etiikkaa. Jotta eettisiä kysymyksiä on mahdollista tarkastella ja huomioida, on signaalien tunnistaminen ensiarvoisen tärkeää. Signaalien avulla voimme varautua tulevaisuuteen ja huomioida tekoälyn eettiseen ja vastuulliseen kehittämiseen liittyviä systeemisiä tekijöitä, riskejä ja mahdollisuuksia, sekä uusia menestymisen edellytyksiä. Tästä tulevaisuuden riskien ja mahdollisuuksien arvioinnin merkityksestä kirjoitimme luvun "Horizon scanning for anticipating social impacts of AI technologies" kirjaan "Handbook on Artificial Intelligence and Services" (Edward Elgar). Signal Post -kirjoitukset ovat luettavissa ETAIROS:n kotisivuilla.

3.2.1 CASE JULKISHALLINTO -Tekoälytulevaisuuksien eettisten vaikutusten ennakoiva arviointi

Sosio-tekniset skenaariot ovat työkalu, jonka avulla voidaan tutkia ja suunnitella tulevaisuutta sekä sosiaalisten että teknologisten näkökulmien kautta. Niiden avulla voidaan paremmin ymmärtää ja hallita monimutkaisia muutoksia ja valmistautua tulevaisuuden riskeihin, haasteisiin ja mahdollisuuksiin.

Sosio-tekniset skenaariot:

- Tarkastelevat sitä, kuinka teknologiset innovaatiot ja sosiaaliset muutokset vaikuttavat toisiinsa: miten esimerkiksi uusi teknologia muuttaa työskentelytapoja, tai miten yhteiskunnalliset trendit vaikuttavat teknologian kehitykseen.
- Ennakoivat mahdollisia tulevaisuuden kehityskulkuja: erilaisia vaihtoehtoisia tulevaisuuskuvia, jotta voidaan paremmin valmistautua mahdollisiin muutoksiin ja haasteisiin.
- Voivat osallistaa sidosryhmiä, kuten työntekijöitä, johtajia, asiantuntijoita ja kansalaisia skenaarioprosessiin. Tämä varmistaa, että eri näkökulmat ja kokemukset tulevat huomioiduksi.
- Hyödyntävät eri (tieteen)alojen asiantuntemusta. Monitieteinen lähestymistapa auttaa luomaan kattavamman kuvan tulevaisuuden mahdollisuuksista ja haasteista.

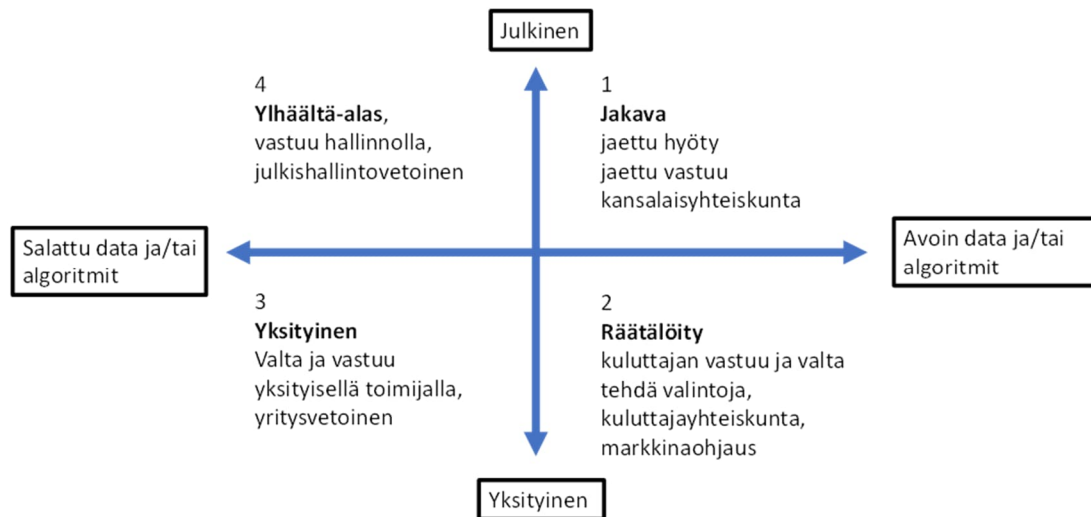
ETAIROS -hankkeessa tarkasteltiin tekoälyn mahdollisia soveltamistapoja eri tavoin kehittyvissä toimintaympäristöissä kahdella tunnistetulla risteävällä soveltamisen ulottuvuudella: yhtäältä algoritmit ja käytetty data ovat joko salattuja tai avoimia käyttäjille sekä toisaalta tekoälyn hyödyntämistä määrittelevät joko julkisen tai yksityisen toimijan valinnat ja intressit.¹³ Näin syntyy neljä ideaalityypistä kenttää (Kuva 5), joissa yhtäältä syntyy jännitteisiä suhteita sovellusintressien välille, mutta toisaalta myös toisiaan vahvistavia dynamiikkoja. Näin tarkastellen tekoälyn soveltaminen muotoutuu näiden kenttien mahdollisista yhdistelmistä ja tasapainotiloista.

Gotcheva, N., & Wessberg, N. (2024). Ethical foresight approach for anticipating societal impacts of AI technologies. Handbook of Services and AI.

Wessberg, N., & Leikas, J. (2021). Tekoälyn tulevaisuuksien eettiset kysymykset sosio-teknisissä skenaarioissa. Futura 3/21, pp. 28-32.

Leikas, J., Johri, A., Latvanen, M., Wessberg, N., & Hahto, A. (2022). Governing Ethical AI Transformation: A Case Study of AuroraAI. Front. Artif. Intell. 5:836557. <https://doi.org/10.3389/frai.2022.836557>

¹³ Wessberg & Leikas 2021



Kuva 5. Skenaariot tekoälyn tulevaisuuden sovellusten ja eettisten ulottuvuuksien tunnistamiseen.

Julkishallinnon rooli suhteessa tekoälyyn on kaksijakoinen. Yhtäältä se pitää sisällään yhteiskunnallisen kehityksen suuntaamisen sekä tekoälyn mahdollisuuksien tukemisen ja riskien hallinnan. Toisaalta se on tekoälyn hyödyntäjä osana julkisen toiminnan digitalisaatiota. Käytännössä tästä juontuu kaksi keskeistä tulevaisuuskysymystä julkishallinnolle: a) Miten julkishallinto voi ohjata laajempaa yhteiskunnallista kehitystä, jossa tekoäly toimii merkittävänä ajurina? b) Miten julkishallinto voi uudistaa toimintaansa hyödyntämällä tekoälyä eettisesti ja kestävästi?

Tekoälyn käyttöönottoa tarkasteltiin edellä esitetyllä tavalla julkishallinnon toiminnan näkökulmasta: Yhtäältä hallinnon tehokkuuden ja kokonaisvaltaisen päätöksenteon akselilla sekä hallinnon kehittämisen ja demokratian vahvistamisen akselilla (Kuva 6). Tekoälyn soveltamisen olennaiset piirteet yhdistäviä ideaalityyppejä voisivat tällöin olla alla esitetyn kuvan mukaisesti "tehokas demokratia", "vaikuttava demokratia", "vaikuttava hallinto" sekä "hallinnon sisäinen tehokkuus". Tarkastelu tekee näkyväksi, kuinka tekoälyn käyttöä voidaan kohdentaa hallinnon sisäiseen kehittämiseen ja/tai sidosryhmäyhteistyön parantamiseen. Oletuksena on tällöin, että vahvistamalla sidosryhmä- ja kansalaisvuorovaikutusta tekoälyn avulla, pystymme vahvistamaan myös demokratiaa ja tukemaan osallistavia valmisteluprosesseja. Hallinnon kehittämiseen keskittyminen edistäisi puolestaan vaikuttavuutta ja parantaisi tuottavuutta.

Hahmoteltu tulevaisuuskuva jäsentää myös tekoälyn hyödyntämisen eettisiä haasteita. Nämä voivat liittyä ainakin yhdenvertaisuuteen ja oikeudenmukaisuuteen, luottamukseen ja luotettavuuteen, avoimuuteen ja läpinäkyvyyteen, lainmukaisuuteen, selitettävyyteen, yksityisyyteen ja autonomiaan, datan keräämiseen, käyttöön ja omistajuuteen, sekä osallisuuteen päätöksentekoprosesseissa ja saavutettavuuteen. Eettisten vaikutusten ennakoivan arvioinnin avulla voidaan hahmottaa systemaattisesti ja realistisesti tekoälyn mahdollisia hyötyjä ja haittoja erilaisissa käyttötilanteissa. Ilman systemaattista ennakoivaa vaikutusten arviointia tekoälyn käyttöönottoa ohjailee joko katteeton innostus tai liiallinen varovaisuus ja kyvyttömyys nähdä teknologian tarjoamia hyötyjä sekä niiden saavuttamisen edellyttämiä toimenpiteitä, kuten työtehtävien uudelleenorganisointia tai koulutusta.



Kuva 6. Tekoälyn käyttöönoton tarkastelukulmia julkishallinnossa.

3.3 Osallistava eettinen visioiden arviointi

Kuten edellä on todettu, tekoälyjärjestelmien eettinen arviointi ei ole vain haittojen tarkastelua tai ihmisten oikeuksien suojelua. Sen oleellisena näkökulmana tulisi olla myös ihmisten hyvinvoinnin ja kukoistuksen edistäminen. Koska eettiset vaikutukset riippuvat hyvin pitkälti asiayhteydestä, sidosryhmien osallistuminen ja tasavertainen kriittinen pohdinta ovat avainasemassa mahdollisten tekoälytulevaisuuksien tarkastelussa.

Parhaimmillaan tekoälyjärjestelmien ja -sovellusten *eettisten vaikutusten ennakoiva arviointi* on yhteistyöprosessi, joka edistää teknologian tai palvelun yhteiskunnallista hyväksyttävyyttä. Tässä prosessissa organisaatiot ja sidosryhmät ennakoivat yhdessä teknisten järjestelmien riskejä ja tunnistavat eettisesti merkittäviä vaikutuksia, jotka johtuvat teknologian ja sosiaalisen kontekstin välisestä vuorovaikutuksesta. Teknologisen muutoksen nopean vauhdin vuoksi tällaisista arvioinneista on tullut välttämättömiä. Erityisesti Euroopan komissiolla on ollut tässä keskeinen rooli, sillä se on sisällyttänyt vastuullisen tutkimus- ja kehitystyön periaatteet puiteohjelmiinsa.¹⁴

ETAIROS:ssa toteutettu eettisten visioiden arvioinnin malli keskittyy kollektiivisiin mielikuviin toivottavista tulevaisuuksista. Vaikka visiot harvoin toteutuvat täsmälleen odotetulla tavalla, ne muokkaavat strategisia ja täytäntöönpanoprosesseja esimerkiksi vaikuttamalla resurssien

¹⁴ Karvonen et al. 2025

jakamiseen, vahvistamalla tiettyjä tavoitteita ja muokkaamalla päätöksentekoa. Nopeiden teknologisten muutosten aikana selkeät visiot auttavat eettisen pohdinnan ankkuroinnissa.¹⁵



Kuva 7. ETAIROS-työpajojen kuvitusta.

¹⁵ Karvonen et al. 2025

3.3.1 CASE CITYVERSE – Tampereen kaupunki

Muutamat edelläkävijäkaupungit ovat alkaneet hyödyntää metaverseä älykkään kaupungin kehittämisen välineenä. Tampereen kaupunki on esimerkki tästä lähestymistavasta "CityVerse"-visiollaan - virtuaalisen ja fyysisen maailman yhdistelmällä, joka on suunniteltu edistämään liiketoimintainnovaatioita ja ihmiskeskeistä älykkään kaupungin kehittämistä. CityVerse pyrkii parantamaan kaupunkisuunnittelua, kansalaisosallistumista, kulttuurikokemuksia ja matkailua metaverseympäristöihin integroitujen laajennetun todellisuuden teknologioiden avulla. Vision mukaan CityVerse on "kognitiivinen kaupunki", jossa fyysinen ja virtuaalinen yhdistyy mahdollistaen immersiiiviset kokemukset, sosiaalisen vuorovaikutuksen, oppimisen, viihteen ja kaupallisen toiminnan.

ETAIROS selvitti puolistrukturoitua fokusryhmämenetelmää hyödyntäen kaupungin työntekijöistä ja viranhaltijoista koostuvan joukon käsityksiä CityVerse-konseptin luonteesta ja arvoista Tampereella. Tampereen kaupungin kanssa toteutetussa tapaustutkimuksessa tarkasteltiin, miten osallistavaa eettisen näkemyksen arviointia voidaan soveltaa kehittyviin metaverseteknologioihin kaupunkiympäristössä: miten osallistavia menetelmiä voidaan suunnitella ja toteuttaa tehokkaasti uusien teknologioiden eettisten ulottuvuuksien arvioimiseksi, kun kyseiset teknologiat ovat vielä pitkälti käsitteellisiä ja niiden vaikutukset epävarmoja.¹⁶

Eettisen arvioinnin jäsentämiseksi kehitimme sarjan "paradigmaattisia tapauskuvauksia", joissa ennustetaan mahdollisia CityVerse-toteutuksia kaupungin asukkaille vuonna 2040. Nämä kuvaukset suunniteltiin yhteistyössä kaupungin edustajien kanssa, jotta varmistettiin niiden merkityksellisyys ja uskottavuus. Tapaukset sisälsivät tarkoituksellisesti mahdollisesti kiistanalaisia eettisiä ulottuvuuksia, mutta pysyivät kuitenkin kaupungin ilmoittamien arvojen pohjalta. Keskusteluryhmiä ohjannut keskeinen tutkimuskysymys oli: "Miten CityVerseä voidaan hyödyntää tavalla, joka edistää luottamusta, vahvistaa yhteisön ja yksilön hyvinvointia sekä edistää kestävyyttä?". Keskustelujen tuloksia tutkittiin laadullisella sisällönanalyysillä.

Tuloksemme tarjoavat arvokasta tietoa CityVerse-palvelun tulevaa kehittämistä varten ja herättävät samalla tärkeitä metodologisia kysymyksiä uusien teknologioiden osallistavasta eettisestä arvioinnista. Kun otetaan huomioon kaupunkien suuntautuminen julkisiin arvoihin, visioiden eettinen arviointi on ratkaisevan tärkeää.

Ilmaisemalla visionsa kaupunki antaa eräänlaisen lupauksen ja luo ulottuvuuksia tulevaa eettistä arviointia varten. Kaupungin työntekijöiden

Karvonen, A., Leikas, J., Wessberg, N., Sigfrids, A., & Pöntinen, S. (2025) *Some problems in the ethical impact assessment of emerging technologies and socio-technical visions: case CityVerse*.

In: Alvarez, I., Arias-Oliva, M., Dediu, A.H., Silva, N. (eds) *Ethical and Social Impacts of Information and Communication Technology. ETHICOMP 2025. Lecture Notes in Computer Science*, vol 15939. Springer, Cham.

https://doi.org/10.1007/978-3-032-01429-0_32

¹⁶ Karvonen et al. 2025

ja viranhaltijoiden osallistuminen on olennaisen tärkeää, koska nämä sidosryhmät viime kädessä ohjaavat ja toteuttavat kehitystä.

Tutkimuksen perusteella voidaan todeta, että sosiotekniset visiot toimivat tehokkaammin eettisen keskustelun foorumeina kuin konkreettisina toteutussuunnitelmina. Sosioteknisten visioiden eettinen tarkastelu tulisikin ymmärtää uudella tavalla: ei niinkään konkreettisten tulevaisuuksien arviointina, vaan arviointina visioista itsestään ja siitä, millaisia arvovalintoja ja painotuksia tekstiin on valikoitunut sekä millaisen prosessin kautta visio on syntynyt.¹⁷

3.4 Kansalliset ja kansainväliset sääntelykehdot

Sääntelyn puitteet muodostavat monimutkaisen kentän tekoälyyn liittyvälle toiminnalle. Keskeinen kehikko tekoälyn yhteiskunnallisen ohjauksen kannalta ovat kansainväliset, EU-tasoiset ja kansalliset ohjeet, strategiat ja sääntely. Yhteiskunnallisella tasolla tekoälyn ohjaus rakentuu EU:n tekoälysäädöksen (EU AI ACT) ja sen toimeenpanon sekä kansallisten strategioiden ja ohjelmien kautta. AI ACT asettaa useita velvoitteita tekoälyn kehittäjille ja käyttöönottajille, joiden on kuitenkin noudatettava myös monia muita säädöksiä. Näitä ovat yleinen tietosuoja-asetus (GDPR), datalaki tai digitaalisia palveluja koskeva laki (DSA), joka velvoittaa alustojen ja hakukoneiden tarjoajat puuttumaan esimerkiksi laittomaan sisältöön sivustoillaan. Digitaalisia markkinoita koskeva laki (DMA) taas pyrkii varmistamaan tasapuoliset ja oikeudenmukaiset toimintaedellytykset suurimmilla verkkoalustoilla toimiville yrityksille.

Erilaisten ohjaus- ja strategiasiakirjojen kirjo on siis laaja ja monissa asiakirjoissa tekoälyä käsitellään osana digitalisaation ohjausta. Esimerkiksi EU: n digitalisaatiostrategiaan kuuluva digikompassi määrittelee myös tekoälyn hyödyntämisen suuntaviivoja. Erityisenä haasteena onkin ja tulee olemaan teknologiasääntelyä koskevan oikeudellisen kehyksen toimivuus yhdistettynä sekä digitaalitekniikkaa edeltävältä ajalta peräisin oleviin asiaa koskeviin säännöksiin että tuleviin säädöksiin. Kansallisilta ja kansainvälisiltä tasoilta strateginen ohjaus laskeutuu esimerkiksi kaupungeille sekä sektori- tai teknologiakohtaisiin strategioihin ja sääntelyyn.

ETAİROS on seurannut EU:n tekoöllysäädöksen kehitystyötä ja antanut kehityksestä lausuntoja. ETAİROS järjesti myös Suomen tekoöllykeskuksen Etiikan asiantuntijaryhmän¹⁸ (FCAI EAB) kanssa tutkijoille keskustelutilaisuuden tekoöllyn ja datan sääntelystä EU:n tasolla ja alan tärkeimmistä EU:n sääntelykehyksistä. Sääntelyssä todettiin olevan lukuisia ongelmakohtia, jolloin järjestelmien luotettavuutta on vaikeaa varmistaa.

Oman näkökulmansa yhteiskunnalliseen ohjaukseen tuovat erilaiset tekoälyn eettiset periaatteet ja julkilausumat. Esimerkiksi OECD:n¹⁹ luotettavan tekoälyn periaatteet on rakennettu oikeudenmukaisen kasvun, kestävän kehityksen ja hyvinvoinnin, ihmisoikeuksien ja demokraattisten arvojen, läpinäkyvyyden ja selitettävyyden, turvallisuuden ja vastuullisuuden ja vastuunkannon ympärille. Läpinäkyvyys- ja selitettävyyksvaatimuksilla on pyritty varmistamaan, että järjestelmien toimintaa voidaan varmistaa ja päätöksiä legitimitettiin säilyy. Tähän kuuluu myös datalähteiden, algoritmien ja tulosten ymmärrettävyys niille, joita esimerkiksi tekoälyn suorittama päätöksenteko koskee. Vastuullisuuden periaatenipulla on pyritty varmistamaan, että ihmiset ovat edelleen vastuussa päätöksistä ja heillä on riittävät kannustimet varmistaa päätösten laatu. Lisäksi on pyritty puuttumaan tekoälyn koulutus- ja käyttödatan mahdollisiin vinoumiin ja varmistamaan päätöksenteon reiluus ja oikeudenmukaisuus. Ohjauskehikkoja ja keinoja on siis olemassa, mutta niiden yhteensovittaminen ja toimivuus käytännön ohjauksessa on toistaiseksi vasta kehittymässä.

¹⁷ Karvonen et al. 2025

¹⁸ <https://fcai.fi/ethics-advisory-board>

19 OECD 2024

Tarvitaan käytännön esimerkkejä erilaisten käyttötapausten arvioinnista tukemaan niiden noudattamista käytännössä.

3.5 Julkishallinto tekoälyn hyödyntäjänä

Julkishallinnossa tarvitaan tekoälyn hyödyntämiseen koulutusta ja lisäresursseja. ETAIROS:n tutkimuksen mukaan pohdinta tekoälyn hyödyntämisestä on etenkin pienemmissä kunnissa vielä lapsenkengissä. Tekoäly koetaan aihepiirinä vaikeaksi, eikä sitä välttämättä mielletä ajankohtaiseksi. Erityisesti kunnissa tekoälyä hyödynnetään toistaiseksi vähän, minkä vuoksi on vaikea muodostaa kokonais kuvaa tekoälyn kehittämisestä ja tekoälysovelluksille annettavasta vastuusta ja päätösvallasta.

Eniten tekoälyn käyttöönottoa hidastaviksi tekijöiksi koetaan yhtäältä sääntelyn ja toisaalta ymmärryksen, osaamisen ja koulutuksen puute sekä datan puutteellinen laatu. Myös vähäiset resurssit kunnissa sekä tekoälykeskustelun nostamat eettiset huolenaiheet tunnistetaan hidastavan käyttöönottoa. Tekoälyjärjestelmien hyödyntämiseen ei yleisesti ole järjestetty koulutusta, ja osaamistarpeen ja resurssien suuntaamisen välillä koetaan olevan ristiriitaa. Edelläkävijäorganisaatioiksi mielletään koulutukseen panostaneet.

Hitaasta haltuunotosta huolimatta julkishallinnossa on melko riittävät valmiudet tekoälyn käytön suunnitteluun ja hyödyntämiseen, mutta resurssit sovellusten toteuttamiseen ovat niukat etenkin pienissä koulutuskuntayhtymissä. Tekoälyä suunnitellaan käytettäväksi erityisesti yleishallinnossa, asiakaspalvelussa, taloushallinnossa ja turvallisuutta edistävissä palveluissa. Tekoälystä toivotaan työkalua julkishallinnon asiointin helpottamiseksi ja asiointikokemuksen parantamiseksi, sekä välineitä organisaation sisälle työn tehokkuuden ja työntekijöiden luovan asiantuntijatyön edistämiseksi.

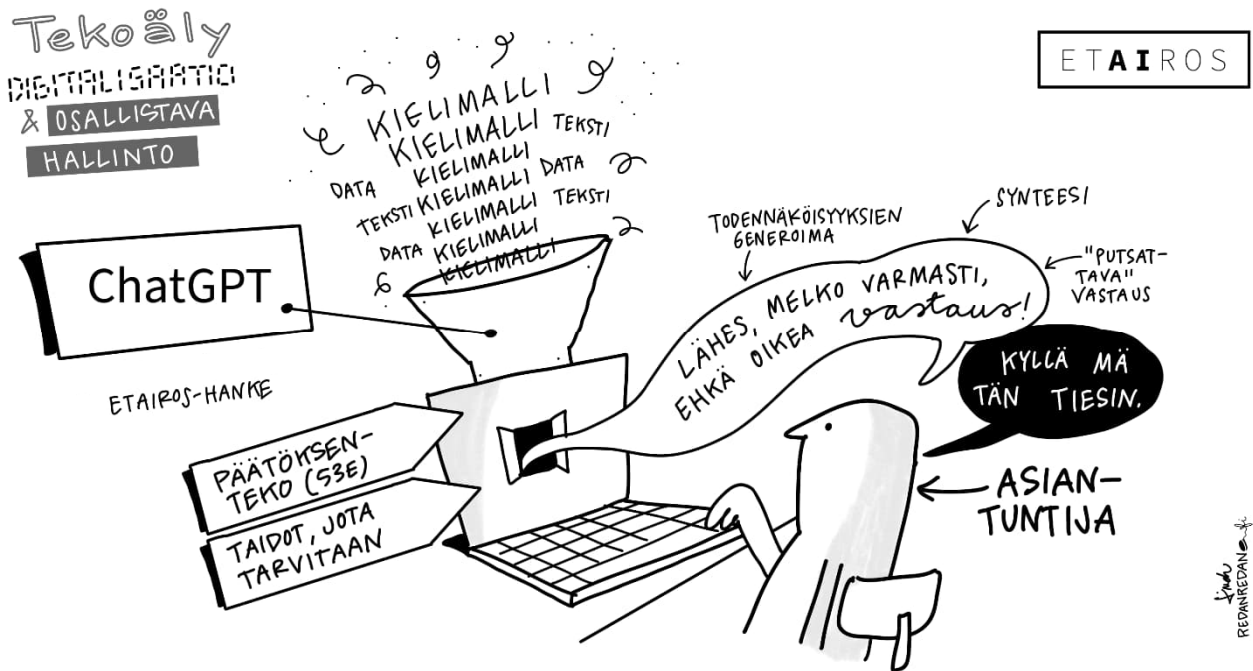
ETAIROS tuki julkishallinnon tekoälyn hyödyntämistä monella tavalla:

- Asiantuntijayhteistyö ja työpajoja julkishallinnon kanssa (mm. kuuden työpajan sarja kielimallien eettisestä hyödyntämisestä VNK:n kanssa)
- Julkiset raportit (8)
- Lainsäädäntötyön tuki ja valmistelu tutkitun tiedon avulla; lausumiset mm. AI ACT:sta
- Digtuetun kansalaisosallistumisen suositukset.

VTT teki ETAIROS-hankkeessa konkreettisiin kysymyksiin paneutuvaa yhteistyötä eri ministeriöiden ja muiden julkishallinnon toimijoiden kanssa, ja vaikutti näin mm. ministeriöiden poliittiseen valmistelutyöhön. Konkreettisia, tarkastelun kohteeksi asetettuja ongelmia olivat erilaiset aineiston käsittelyn ja sisällöntuotannon kysymykset sekä tekoälysovellusten mallintaminen ja kokeilu ministeriöiden tuottaman aineiston pohjalta. Seuraavassa on esitetty joitakin merkittävimpiä VTT:n yhteistyön muotoja julkishallinnon kanssa.

- *Valtioneuvoston ja ministeriöiden edustajien kanssa tarkasteltiin tekoälyn ja erityisesti kielimallien käyttöönoton strategisia ja käytännöllisiä kysymyksiä sekä eettisiä haasteita. Työpajakeskustelut nostivat esiin organisatoristen jännitteiden problematiikan. (Työn tuloksia esitellään lähemmin luvussa 7.4.)*
- Osallistuminen *Valtiovarainministeriön* asettamaan, Suomen digitaaliseen kompassiin 2030 kytkeytyvän Elämäntapahtumien ja liiketoimintatapahtumien digitalisointi ja automatisointi -esiselvityshankkeen tukena toimivaan asiantuntijaryhmään. Asiantuntijaryhmän esitysten pohjalta aktivoidaan ja vahvistetaan kansalaisten osallisuutta sekä luodaan uusia malleja osallistamisrakenteille.

- Osallistuminen Valtiovarainministeriön Tekoälyn ja digitalisaation tutkimuksen asiantuntijaryhmän jäsenenä.
- Osallistuminen Valtiovarainministeriön valtakunnallisen Tulevaisuuden tekoäly -webinaarin järjestämiseen.
- Kommentointi EU-komission White Paperia tekoälytutkimuksen ja -kehityksen suuntaviivoista.
- Lausuminen AI-ACT -lakiehdotuksen kehitystyöstä.
- Osallistuminen Valtiovarainministeriön AuroraAI:n Etiikkaryhmän työskentelyyn sekä AuroraAI:n Osaaminen -teemaryhmään ihmiskeskeisen ja ennakoivan yhteiskunnan näkökulmasta.
- Osallistuminen jäsenenä Eduskunnan Tulevaisuusvaliokunnan Ikääntymisen tulevaisuus -työryhmässä roolinaan tarkastella ja tuoda yhteiseen keskusteluun uusien teknologioiden eettisiä näkökulmia. ETAIROS-tutkijat osallistuivat aktiivisesti Tulevaisuusvaliokunnan tekoälytyöryhmään ja Tekoälyraportin kirjoittamiseen (2022), sekä Hyvinvointi -Pahoinvointi -raportin kirjoittamiseen (2022). Etairoslaiset ovat myös lausuneet TEM:lle digikompassiesityksestä sekä VM:lle digitaalista henkilöllisyyttä koskevasta esityksestä.
- Osallistuminen strategisen tutkimuksen ja sosiaaliturvakomitean tietokumppanuus -yhteistyön tuloksena syntyneen tietokoosteen "Digitalisaatio, työn murros ja sosiaaliturva" kirjoittamiseen ja työn pohjalta järjestettyyn seminaariin asiantuntijapuheenvuorolla.
- Osallistuminen OKM:n rahoittamaan SotePeda -työskentelyyn yhdessä sote-alan koulutusta tarjoavien ammattikorkeakoulujen kanssa, ja yhteisjulkaisuun Tulevaisuuden teknologiat ja tekoälyn etiikka.
- Osallistuminen jäsenenä OKM:n "Tekoälyosaaminen" pyöreään pöytään.
- Virtuaalisen keskustelufoorumin järjestäminen yhdessä valtakunnallisen Digituki -hankkeen kanssa tekoälyn ja digiteknologian käyttöön liittyvistä mahdollisuuksista, unelmista, esteistä ja haitoista.



Kuva 8. ETAIROS-työpajojen kuvitusta.

3.5.1 CASE VNK - Tekoälyn käyttöönoton johtaminen ja organisatoriset jännitteet

Tekoälyn käyttöönotto saattaa luoda organisaatioissa jännitteitä ja ristiriitaisia vaatimuksia. Käyttöönotto liittyy moniin toimintoihin, jotka muuttavat teknisten järjestelmien lisäksi työn luonnetta ja organisointia. Arvoperustan johdonmukaisuus ja jaettu ymmärrys tekoälystä ovat tärkeitä tekoälyn onnistuneen käyttöönoton kannalta.

ETAIROS-hankkeessa toteutetun tutkimuksen mukaan tekoälyn käyttöönottoon liittyviä organisatorisia jännitteitä voidaan odottaa eri syistä, kuten arvoriistiriitojen, muutosta vastustavan organisaatiokulttuurin tai riittämättömien prosessien ja valmiuksien vuoksi. Jännitteet liittyvät yleisesti ottaen tekoälyn toteuttamiseen ja vaikutuksiin.

Tarkastelimme tekoälyn hyödyntämisen organisatorisia jännitteitä i) usean organisaation tekoälykehitysasiantuntijoille suunnatun haastattelututkimuksen sekä ii) ministeriöiden ja valtioneuvoston kanslian virkamiehille suunnattujen työpajojen avulla.

Hankkeessa haastateltiin tekoälykehitysasiantuntijoita 36 suomalaisesta organisaatiosta. Haastateltavista 10 edusti ministeriöitä, 7 yrityksiä, 4 tutkimuslaitoksen edustajaa ja loput kuntia ja järjestöjä sekä muita julkisia toimijoita kuten sairaalaa, poliisia, pelastustointia ja finanssialaa.

Haastattelujen analyysin pohjalta löydettiin kolme erilaista, useita teemoja sisältävää jänniteperhettä. Nämä kytkeytyivät organisaation arvoihin, tekoälyn implementointiin sekä tekoälyn vaikutuksiin. Tekoälyn käyttöönotto on käytännöllinen prosessi, jota ohjaa yleensä kolme lähtökohtaa. Ensinnäkin se voi olla teknologialähtöistä eli riippua siitä, minkälaisia uusia teknologisia ratkaisuja kehitetään. Toiseksi tekoälyn käyttöönotto on datalähtöistä. Tällöin pyritään kiinnittämään huomiota siihen, minkälaisella datalla tekoäly toimii ja mitä dataa organisaatiolla on käytössään sekä millaista dataa sen tulisi kerätä. Kolmanneksi tekoälystrategian toimeenpanossa korostuu organisaation muutos. Tällöin toimeenpanon kohteena ovat muun muassa prosessien ja johtamisen muutokset ja käyttöönoton osaamisen kasvattaminen. Käytännössä tekoälyn käyttöönotto organisaatiossa vaatii erityisesti 1) johtamista ja koordinaatiota, 2) ratkaisukeskeisyyttä, 3) ajattelutapojen muutosta sekä 4) merkityksellisuuden kokemuksen.²⁰

Tämän jälkeen hyödynnettiin ja kehitettiin *sparrausklinikan konseptia*, jossa virkahenkilöt ja tutkijat törmäytetään uuden tiedon ja ymmärryksen saavuttamiseksi. Valtioneuvoston johdolla usean eri ministeriöiden virkahenkilöiden kanssa toteutetut työpajat järjestettiin syksyllä 2024 ja keväällä 2025. Niissä keskityttiin dialogisuuden ja yhteiskehittämisen avulla yhteisen ymmärryksen luomiseen ja

Lanne, M., Nieminen, M., & Leikas, J. (2025). *Organisational tensions in introducing socially sustainable AI*. AI & Soc. <https://doi.org/10.1007/s00146-025-02293-y>

Leikas, J., Karvonen, A., Wessberg, N., & Sigfrids, A. (2025). *Tulokset korostavat tarvetta avoimelle keskustelulle ja selkeille linjauksille tekoälyn käytössä* | Kuntalehti

Karvonen, A., & Leikas, J. (2023). Tekoälyjärjestelmien vastuullinen ja eettinen suunnittelu ja käyttö. *Digitalisaatio, työn murros ja sosiaaliturva-tietokooste*. Suomen Akatemia. ISBN 978-951-715-950-0

Leikas, J. (2022). Tekoälystä vastuullista. Teoksessa: Aillisto, H., et al. (2022). *Tekoälyratkaisut tänään ja tulevaisuudessa*. Tulevaisuusvaliokunnan julkaisu 1/2022. Eduskunta.

²⁰ Lanne et al. 2025

tutkimustiedon välittämiseen tekoälykokeilujen ja strategisen kehittämisen yhteydessä, päätyen erityisesti lainvalmisteluun liittyviin erityiskysymyksiin. Tutkimukselliseksi näkökulmaksi nousi lopulta *organisatoriset jännitteet*. Niitä nostettiin esiin tarkastelemalla tekoälyn ja erityisesti kielimallien käyttöönoton strategisia ja käytännöllisiä kysymyksiä sekä eettisiä haasteita.

Työpajakeskustelujen kautta kerätystä aineistosta nousi kymmenen jännitettä, jotka liittyvät organisaation arvoihin, käytön toteutukseen ja käytön vaikutuksiin (Kuva 9).

Toiminnan ohjaus, päätökset	Käyttöönotto ja käyttö	Käytössä olo ja seuranta
Arvoihin liittyvät jännitteet Laatu vs. tehokkuus Korkeat vs. matalat odotukset Painotetaan inhimillistä toimijaa vs. koneen suorituskykyä	Toteutukseen liittyvät jännitteet Pitkäjänteinen kehitys vs. kokeilut Sujuvoittaa vs. mutkistaa työtä Herättelee vs. korvaa ajattelua Toimintakulttuuri edistää vs. vastustaa	Vaikutuksiin liittyvät jännitteet Tasainen laatu vs. vinoumat Paras mahdollinen teknologia käytössä vs. kasvava riippuvuus Luottamus kasvaa vs. vähenee

Kuva 9. Tekoälyn/kielimallin käyttöönoton jännitteet organisaatiossa.

Laatu vs. tehokkuus

- Tehokkuus on tavallisesti keskeisin arvo, jonka pohjalta tekoälyyn siirtymisen päätös tehdään. Samalla tavoitellaan myös parempaa laatua. Laadun ylläpitäminen on vaikeaa tehokkuuden noustessa päätavoitteeksi.

Korkeat vs. matalat odotukset

- Odotukset tekoälyn tuomasta tehokkuudesta ja laadusta ovat usein korkealla ja tavoitellaan ideaalitulannetta, jossa tekoäly hoitaa prosessit ja nostaa työntekijän tehoa. Käytännössä toteutuksen kuitenkin ajatellaan tapahtuvan minimiresurssein ja hyödyt voivat jäädä lunastamatta.

Inhimillinen toimija vs. koneen suorituskyky

- Tekoäly tehostaa ja nopeuttaa päätöksentekoa tuoden ja analysoiden enemmän tietoa, mutta se voi myös painottaa ”väärä” asioita ja oikaista prosesseissa kuten keskusteluvaiheessa uhaten demokraattisen päätöksenteon ja oikeusvaltioperiaatteiden ydintä. Tekoälyn ja ihmisen tekemät virheet ovat erilaisia ja kykymme ymmärtää tekoälyn virheitä on rajallinen.

Kokeilut vs. pitkäjänteinen kehitys

- Tekoälysovelluksen käyttöönoton kehittämistä tehdään hankerahalla, jolloin saadaan polkaistua käyntiin kokeilu. Näin tekoälyä päästään testaamaan. Eri tarpeiden

kartoittamiseen, prosessien toiminnalliseen suunnitteluun ja parantamiseen tarvitaan kokonaisvaltaisen, laaja-alaisen ja jatkuvan tekoälyhyödyntämisen kehitystyötä.

Mutkistaa vs. sujuvoittaa työtä

- Tehostaessaan toimintaa ja tuodessaan mukaan esim. lisää dataa ja analyysimahdollisuuksia ja generoidessaan kysymyksiä tekoäly voi samalla lisätä työmäärää. Käyttöönotto ilman riittäviä kyvykkyyksiä ja resursseja lisää kognitiivista kuormitusta ja teknologiestressiä. Tarvitaan sujuvoittamista siten, että ihmisen toiminta mahdollistuu ihmiselle ominaisin tavoin.

Herättelee vs. korvaa ajattelua

- Tekoäly voi herätellä ihmistoimijaa huomaamaan eri näkökulmia mm. esittämällä erilaisia kysymyksiä ja tuottamalla analyysejä. AI työparin toimintalogiikka kuitenkin pysyy osin piilossa työntekijältä, jolloin luottamus analyysien oikeellisuuteen ei ole aukoton. Tekoäly voi myös ”tehostaa” toimintaa niin, että ihmisen hidas ajattelu korvataan ja oikaistaan prosessissa suoraan tekoälyn tuottamaan lopputulemaan.

Toimintakulttuuri edistää vs. vastustaa

- Tekoälyn hyödyntäminen toteutetaan nykyisessä organisaatiokulttuurissa, mutta kulttuuri voi myös hitaasti muuttua. Kulttuuri voi joko edistää tai vastustaa (myös kumpaakin yhtä aikaa) tekoälyn käyttöönottoa. Tekoälyn mahdollisuudet tulkita kulttuuria ovat kysymysmerkki.

Tasainen laatu vs. vinoumat

- Tekoälyllä kyetään yhtenäistämään päätöksentekoa (mm. kriteerit), mutta kielimallin taustalla kenties vaikuttavat jonkinlaiset arvot ja näkemykset, joita ei saada näkyviksi. Mallin avulla voidaan päätyä uskottavan näköisiin tuloksiin ilman, että analyysi lopulta sisältää oikeita asioita. Malliin ja sen lopputuloksiin on vaikea luottaa.

Paras mahdollinen teknologia käytössä vs. kasvava riippuvuus

- Julkishallintoa tehostavan teknologian käyttö on vastuullista; silti voidaan joutua riippuvaiseksi malleista, joiden koodaus- ja kehitystausta ei ole omissa käsissä. Teknologian myötä nykyiset hyvin toimivat prosessit voivat rapautua eikä toimintaa kyetä enää myöhemmin suorittamaan ilman ko. teknologiaa.

Luottamus julkishallintoon pysyy vs. heikkenee

- Tekoälyä hyödynnettäessä hallinnan läpinäkyvyys osin heikkenee (musta laatikko), mutta toisaalta päätöksentekokriteereitä voidaan kenties avata tarkemmin, jolloin päästään aiempaa yhtenäisempään ja johdonmukaisempaan lopputulemaan.

Tunnistetut jänniteryhmät osoittavat, mihin näkökohtiin organisaatioiden olisi kiinnitettävä erityistä huomiota ottaessaan käyttöön tekoälyä. Tämä puolestaan tukee organisaatioiden käytäntöjen ohjaamista ja johtamista. Tunnistetut jännitteet ovat epäilemättä melko yleismaailmallisia ainakin julkishallinnon yhteydessä. On kuitenkin hyvä huomata, että niiden arvo on myös siinä, että lähteenä ovat toimineet organisaatioiden jäsenet. Jännitteet kertovat todellisista ja relevanteista kysymyksistä organisaation sisällä. Ne synnyttivät runsaasti keskustelua kertoen tunnistettujen ilmiöiden merkityksestä osallistujille.

Viimeisen työpajan ja loppukeskustelun yhteydessä löydökset validoitiin, osallistujien korostaessa erityisesti jännitteitä laatu vs. tehokkuus, toimintakulttuuri edistää vs. vastustaa, luottamus julkishallintoon kasvaa vs. vähenee sekä paras mahdollinen teknologia käytössä vs. kasvava riippuvuus.

Onnistunutta tekoälyn käyttöönottoa tulee johtaa ja koordinoida *kokonaisvaltaisesti*. Näin kyetään toteuttamaan tekoälyn mahdollistama systeeminen muutos. Tällä hetkellä esimerkiksi vain harva julkinen organisaatio johtaa ja koordinoi tekoälyn kehitystä organisaatiossaan riittävällä tavalla. Ongelma on myös johtamiseen ja koordinaatioon liittyvien toimintamallien puute.

Tekoälyn teknologinen ja datalähtöinen kehittäminen tulisi liittää mahdollisimman laajasti myös kehittämisvaiheessa siihen, että sillä pyritään *toiminnalliseen muutokseen*, kuten parempiin palveluihin. Monella julkisella organisaatiolla on teknologialähtöisiä tekoälyn kehittämishankkeita, joissa kehitetään ”teknistä potentiaalia” toivoen, että joku toimija hyödyntäisi tehtyjä teknologisia malleja.

Kolmas keskeinen tekoälyn toimeenpanon ongelma on *ajattelumallien muuttumattomuus*. Tekoälyn onnistunut käyttöönotto vaatii, että toimintaa ajatellaan johtamisessa uudella logiikalla työntekijöiden itseohjautuvuutta korostaen. Valitettavan usein johtamista toteutetaan vanhalla tavalla hierarkkisesti, vaikka samalla halutaan hyödyntää tekoälyn käyttöönottoa. Tällöin johtaminen on pikemminkin tehokkaan tekoälyn käyttöönoton rasite kuin sen promoottoriedistäjä.

Tekoälyn toimeenpanossa tulee myös huomioida *merkityksellisyys*. Kysymys on pyrkimyksestä ratkaista tekoälyn avulla ihmisille merkityksellisiä ongelmia merkityksellisellä tavalla. Johtamisen tehtävä on pyrkiä vähentämään tekoälyyn liittyviä, merkityksellisyyden kokemuksia heikentäviä tekijöitä niin työntekijöiden kuin asiakkaiden kannalta. Esimerkiksi osallistavat kehittämisprosessit antavat työkaluja tähän. Kun kehitysprosesseissa pystytään vastaamaan kysymykseen, miksi jokin muutos toteutetaan, mikä työntekijöiden rooli ja tehtävä muutoksessa on, annetaan mahdollisuus merkityksellisyyden kokemukselle. Merkityksellisyyden kokemusta vastaavasti vähentävät tilanteet, joissa työntekijät eivät tiedä toiminnan tavoitetta, vastuut ja roolit toiminnassa ovat epäselkeitä tai tavoitteiden saavuttamista ei seurata.

Tekoälyn luomat transformaatiot luovat jännitteitä organisaatioihin, joita eettinen keskustelu usein nostaa pinnalle. Tekoälyn käyttöönottoon liittyviä organisatorisia jännitteitä voidaan odottaa erityisesti, kuten arvostusriitojen, muutosta vastustavan organisaatiokulttuurin tai riittämättömien prosessien ja valmiuksien vuoksi. Jännitteet liittyvät yleisesti ottaen tekoälyn toteuttamiseen ja vaikutuksiin.²¹ Tekoälyn vastuullinen käyttö organisaatiossa edellyttää sääntöjä ja ohjeita, mutta myös joustavuutta ja kontekstisidonnaisia lähestymistapoja. Yleensä suhteellisen muodollisia ja joustamattomia vaatimuksia on vaikea muuntaa monenlaista sovellettavuutta edellyttäväksi käytännön toimiksi. Tämä aiheuttaa jännitteitä abstraktien periaatteiden ja yksityiskohtaisten määritelmien ja toimien välillä. Organisaation toimintaa ohjaavat erilaiset arvot saattavat aiheuttaa ristiriitoja myös sen suhteen, mitä tekoälyn avulla tulisi ensisijaisesti saavuttaa. Samoin voi olla ideologisia eroja siinä, ajatellaanko ensisijaisesti yksilön etuja ja koettuja hyötyjä vai yhteiskunnan hyötyjä. Ristiriitoja voi ilmetä myös arvojen selkeydessä.

Koska jännitteet rakentuvat ja säilyvät osittain viestinnän ja vuorovaikutuksen kautta, johtamisessa tulisi rakentaa luottamusta ja avoimuutta, jotta jännitteitä voidaan hallita rakentavasti. Organisaation johdon olisi pyrittävä tunnistamaan jännitteet ja käytävä avoimesti keskustelua tekoälyn kielteisistä ja myönteisistä vaikutuksista sekä pohdittava ennakoivasti riskejä ja mahdollisuuksia erilaisista yhteiskunnallisista näkökulmista, sisältäen niin sidosryhmät, asiakkaat, kansalaiset kuin erilaiset järjestöt. Jännitteiden tunnistaminen ja niiden pohjalta uuden kehittäminen on osa organisatorista kehitystä.

²¹ Lanne et al. 2025

3.6 Luottamus tekoäly-yhteiskuntaan edellyttää osallisuuden kokemusta

Tekoälyyn täytyy voida luottaa. Luottamus syntyy panostamalla tekoälyn avulla tehtävien päätösten läpinäkyvyyteen ja selitettävyyteen. Luottamusta parantavat vastuuvollisuuden toteutuminen sekä tekoälyn tulevaisuuden kehitykseen ja vaikutuksiin liittyvien kysymysten selvittäminen oleellisten ihmisryhmien ja organisaatioiden kannalta. Luottamusta testaa viime kädessä kansalaisten kokemus.

Kansalaisilla on oikeus yksityiskohtaiseen ja kattavaan tietoon siitä, millä perusteella tekoäly on päätyntä tiettyyn ratkaisuun ja mitä tietovarantoja se on hyödyntänyt henkilöä koskevissa kysymyksissä. Päätöksentekoprosessin tulisi olla mahdollisimman selkeä kansalaisten kannalta. Tekoälyn potentiaali ja käyttötavat ihmisten arjessa ja työnteossa ovat vasta hahmottumassa, vaikka tekoälyjärjestelmät ovat jo monessa toiminnassa mukana tiedostamme sitä tai emme.

Kansalaisten tietoisuus omista oikeuksistaan ja vastuistaan koetaan merkittäväksi. Julkishallinto voi edistää sidosryhmäyhteistyötä toimimalla vuorovaikutuksessa tekoälyn käyttäjien kanssa sekä huomioimalla käyttäjäkokemuksia. Tutkimuslaitokset, teknologiatarjoajat ja oma organisaatio koetaan merkittävimpinä yhteistyökumppaneina, mutta kansalaisia on keskusteluun osallistettu toistaiseksi hyvin vähän. Keskustelu tekoälyn käytön tavoitteista, toivottavista ja ei-toivottavista vaikutuksista sekä käytön riskeistä tulisi viedä myös käyttäjärajapintaan. Tätä kysymystä tarkastelemme mm. artikkelissa "*Governing Ethical AI Transformation: A Case Study of AuroraAI*" (2022), jossa tapaustutkimuksena oli kansallinen tekoälyohjelma AuroraAI.

Kenttäkokeilun avulla (N= >500), jonka toteuttivat VTT, Sitra ja Digi- ja väestötietokeskus yhdessä testattiin, kuinka hyvin kansalaisten osallistamista ja kuulemista varten kehitetyt, uudenaikaiset digitaaliset osallistumisalustat soveltuvat kuulemiseen julkisten palveluiden kehittämisessä. Yhteiskehittämisellä viitataan tyypillisesti kansalaisten ja palvelujen käyttäjien osallistumiseen palvelujen suunnitteluun, kehittämiseen ja arviointiin. Tavoitteena on, että erilaisten käyttäjäryhmien ja kansalaisten osallistuessa lopputuloksena on parempia palveluita, tehokkaampaa palvelukehittämistä ja läpinäkyvämpää päätöksentekoa. Digialustat voivat vahvistaa yhteiskehittämistä tarjoamalla kansalaisille helpomman kanavan relevanttiin tietoon sekä matalamman, usein ajasta ja paikasta riippumattoman tavan osallistua puhelimen tai tietokoneen välityksellä. Osallisuus yhteiskehittämiseen viittaa siihen, että erilaiset sidosryhmät ja kansalaiset otetaan mukaan palvelusuunnittelun prosessiin siten, että osallistumisella on vaikutusta. Osallisuustyö perustuu ajatukseen, että henkilöillä, joihin päätös vaikuttaa, on sekä oikeus olla mukana päätöksentekoprosessissa, että tietää, millä tavalla heidän osallistumisensa vaikuttaa päätöksentekoon ja on huomioitu siinä.²² Kenttäkokeilu painottui suunnitteluosallisuuteen, ts. asioiden valmisteluun liittyvään vuorovaikutukseen. Suunnitteluosallisuudessa kansalaiset osallistuvat palvelusuunnittelijoiden ennalta valitseman aihealueen suunnitteluun kertomalla kokemuksiaan, tarpeitaan ja näkökulmia siitä, minkälaisia elementtejä tulevan palvelun tulisi sisältää.

ETAIROS-hankkeen tuloksissa nousi esille tekoälyn toimeenpanon tärkeys julkisessa hallinnossa. Toimeenpanoon liittyvät niin teknologiset kuin organisaation ja toiminnan muutoksesta nousevat sekä datalähtöisen päätöksenteon ulottuvuudet. Nämä kaikki on otettava huomioon etiikan ja hallinnan kysymyksiä ratkaistaessa.

²² International Association for Public Participation 2025. IAP2 core values for public participation.
<https://www.iap2.org/page/corevalues>

ETAIROS kehitti lainsäädäntöön perustuvan sääntelyn rinnalle pehmeän ohjauksen muotoja tekoälyn eettiseen hyödyntämiseen. Hankkeessa luotiin yleisiä toiminta- ja suunnittelumalleja, joiden varassa tekoälypohjaisia järjestelmiä ja sovelluksia voidaan ottaa käyttöön ja hyödyntää eettisesti kestäväällä tavalla.

Tutkimuskirjallisuuden systemaattisen katsauksen avulla selvitettiin millaisia hallinnan malleja ja kehitysehdotuksia esitetään ja mistä periaatteista ja käytännöistä tekoälyn sääntely- ja koordinaatiomekanismit koostuvat. Millaisiin periaatteisiin nojaten ja millaisia keinoja hyödyntäen julkishallinto voisi ohjata tekoälyn kehittäjiä ja hyödyntäjiä omaksumaan eettisiä ja vastuullisia menettelyitä toimintaansa? Tavoitteena oli tukea tekoälyn eettistä ja vastuullista käyttöä edistävää yhteiskunnallista ohjausta kuvailemalla, tiivistämällä ja syntetisoimalla saatavilla olevaa tietoa tekoälyn ohjauksen ja koordinaation tavoitteista, haasteista, toimenpiteistä ja kehittämis ehdotuksista. Katsauksessa selvitettiin:

- Nykyiset tekoälyn ohjauksen ja koordinaation kehitysehdotuksia esittävien tutkimusten ongelmanmäärittelyt ja ehdotukset.
- Julkishallinnon toimenpiteiden rooli näissä ehdotuksissa.
- Osallisuuden ja dialogin voimistamisen keinot ja periaatteet ehdotuksissa.

Selvityksemme osoittaa, että tutkimuskirjallisuudessa esitellään sidosryhmien välistä keskustelua, ennakkointia ja arviointia oleellisina osina tekoälyn hallinnan mekanismien kokonaisuutta, mutta konkreettisia keinoja näiden toteuttamiseksi ei juurikaan esitetä. Julkishallinnon tulisi omaksua erilaisia osallistavan päätöksenteon muotoja voimistaakseen eettisten ja vastuullisuusperiaatteiden toteutumista tekoälyn kehittäjien ja käyttäjien keskuudessa.

Kirjallisuuskatsauksen perusteella todettiin, että tekoälyn ohjaus ei nykyisellään ole riittävää, sillä toimijat eivät ilman selkeitä ohjauskeinoja tule toimimaan yhteisen hyvän puolesta järjestelmässä, jossa vallitsee organisaatiokohtainen voitontavoittelun logiikka. Merkittävänä muutosta edellyttävänä aiheina ovat 1. julkisen hallinnan luonne suhteessa nopeasti muuttuvaan ja yhteiskuntaa monialaisesti muuttavaan teknologiaan 2. keinot yhteisesti jaettujen arvojen ja tavoitteiden juurruttamiseksi toimintaa ohjaaviksi tekijöiksi, mikä tarkoittaa jonkinlaista uutta tasapainoa pehmeiden ja kovien ohjauskeinojen välillä. Selvityksen pohjalta julkaistiin artikkelit "How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI" (2020), "Human-centricity in AI governance: A systemic approach" (2023), sekä "Governing AI services: a state-of-the-art overview" (2024).

Tämä OSKI-malli rakentuu neljän keskeisen näkökulman ympärille: Osallisuus, Sovellettavuus, Kokonaisvaltaisuus ja Institutionaalisuus.²³

²³ Sigfrids ym. 2023B

3.6.1 CASE ETPK - Osallisuus ja osallistuminen

Suomen julkinen hallinto sai 2021 Taloudellisen yhteistyön ja kehityksen järjestö OECD:ltä palautetta siitä, että kansalaiset saavat osallistua päätöksentekoon liian vähän, liian kapeasti sekä liian myöhään.²⁴ Digitaaliset osallistumisalustat voivat tarjota käyttäjäystävällisiä ja helposti lähestyttäviä osallistumisen väyliä. Osallistumisalustojen käyttöön liittyy kuitenkin vielä merkittäviä haasteita.

ETAIROS toteutti yhdessä Sitran ja Digi- ja väestötietoviraston (DVV) kanssa kokeilun, jossa pilotoitiin digitaalisia osallistumisalustoja. Kokeilu tuki kansalaisosallistumisen kehittämistä Elämäntapahtumalähtöisen digitalisaation edistämishjelmassa (ETPK; VN/33803/2023), ja se nivottiin osaksi läheisen kuoleman elämäntapahtuman suunnittelua. Kokeilun pääasiallisena tavoitteena oli lisätä ymmärrystä alustojen hyödyistä ja käyttöönoton haasteista julkisessa palvelukehittämisessä sekä laatia kokeilun oppeihin perustuen suosituksia kansalaisosallistumisen voimistamiseksi ETPK-ohjelmassa.

Kokeiluun osallistui yli 500 henkilöä marras-joulukuussa 2024. Kokeilussa oli kaksi pääkysymystä. Yhtäältä haluttiin selvittää, miten digialustoilla voidaan tukea kansalaisosallistumisen toteutumista ja mitä käytännön haasteita ja mahdollisuuksia niihin liittyy. Tavoitteena oli arvioida digitaalisen osallistumisprosessin toteutumista sen suhteen, miten alustat toimivat ja miten osallistujat kokivat alustojen käytön, keitä prosessiin osallistui, miten puntarointi onnistui ja minkälaista tietoa prosessista syntyi. Tämän lisäksi haluttiin selvittää, miten kansalaisten osallistumista voidaan vahvistaa ETPK-ohjelmassa: miten kokeilussa mukana olleet palvelukehittäjät ja ETPK-ohjelman keskeiset henkilöt suhtautuvat kokeilussa käytettyjen tai vastaavien digialustojen käyttöönottoon palvelukehittämisessä ja minkälaisia toimenpiteitä tulisi tehdä kansalaisosallistumisen edistämiseksi, kuinka hyvin kansalaisten osallistamista ja kuulemista varten kehitetyt, uudenaikaiset digitaaliset osallistumisalustat soveltuvat kansalaisten kuulemiseen julkisten palveluiden kehittämisessä, ja millä tavalla digitaaliset osallistumisalustat voisivat vahvistaa kansalaisten osallistumista julkisten palveluiden elämäntapahtumalähtöiseen kehittämiseen.²⁵

Sigfrids, A., Latvanen, M., & Hantula, K. (2025). Digitaalinen kansalaisosallistuminen julkisessa palvelukehittämisessä. Kenttäkokeilun loppuraportti. VTT Technology No 437.

Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers*. <https://doi.org/10.3389/frai.2023.976887>

²⁴ OECD (2021). Civic Space Scan of Finland, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/f9e971bd-en>.

²⁵ Sigfrids, Latvanen & Hantula 2025

Vaikka digitaalisia osallistumisalustoja on ollut käytössä jo 1990-luvulta lähtien, ne eivät ole juurtuneet vakiintuneeksi osaksi päätöksenteon instituutioita. Syitä tähän on useita. Alla on esitetty joitakin keskeisiä kirjallisuudesta tunnistettuja digitaalisen osallistumisen haasteita (Taulukko 1).

Taulukko 1. Digitaalisen osallistumisen haasteita.

Osin teknisesti ratkaistavissa olevat haasteet	
Puutteellinen saavutettavuus	Edellyttävät teknisiä ratkaisuja.
Sisällön vaikeaselkoisuus	Alustat eivät ole esteettömiä kaikille (esim. näkövammaiset, kielivähemmistöt).
Verkkokeskustelun nopeus	Käsiteltävät aiheet ja kieli eivät ole helposti ymmärrettäviä.
Polarisaatio	Nopeatempoinen vuorovaikutus voi heikentää argumentaation syvyyttä ja lisätä polarisaatiota.
Tietoturva, anonymiteetti ja yksityisyys	Verkkoalustat voivat kärjistää mielipiteitä, mikä vaikeuttaa rakentavan vuoropuhelun ja yhteisymmärryksen muodostamista erityisesti moniarvoisissa kysymyksissä. Huoli tietojen väärinkäytöstä vähentää osallistumishalukkuutta.
Kulttuuriset haasteet	
Puutteelliset digitaidot	Edellyttävät kulttuurista, poliittista tai kognitiivista muutosta.
Ihmisten vähäinen kiinnostus politiikkaan	Monilta kansalaisilta – erityisesti iäkkäiltä tai vähemmän koulutetuilta – puuttuu tarvittavat ICT-taidot.
Luottamus poliitikkoihin ja hallintoon	Osallistuminen ei motivoi etenkin nuorempia tai syrjäytyneitä ryhmiä.
Tietämättömyys osallistumismahdollisuuksista	Usko vähäisiin vaikuttamismahdollisuuksiin vähentää osallistumishalukkuutta.
Aiheet eivät resonoi arjen kanssa	Kansalaiset eivät tunne alustoja tai niiden käyttötarkoitusta.
Vähäinen aktiivisuus sisällön tuottamisessa	Kansalaiset kokevat osallistumisen irralliseksi omasta elämästään.
Päätöksentekijöiden asenteet osallisuusprosesseja kohtaan	Suurin osa osallistujista on passiivisia tarkkailijoita. Käyttäjät seuraavat keskustelua, mutta eivät osallistu aktiivisesti.
	Päätöksentekijät eivät hahmota kansalaisten osallistumisen itseisarvoa ja mahdollisia hyötyjä.
Institutionaaliset haasteet	
Digitaalinen eriarvoisuus	Edellyttävät rakenteellista ja strategista muutosta hallinnossa.
Vaikutusten epäselvyys	Teknologiset (heikko pääsy internettiin), sosiaaliset ja alueelliset erot vaikuttavat osallistumisen tasa-arvoon.
Institutionaalinen integraatio puutteellista	Osallistumisen yhteys päätöksentekoon on usein hämärä tai puuttuu kokonaan.
Poliittisten toimijoiden moninaisuus ja ristiriidat	Alustat eivät ole vakiintuneet osaksi hallinnon päätöksenteon prosesseja.
Lainsäädännölliset ja hallinnolliset esteet	Siilot ja osa-optimointi vievät edellytykset osallisuusprosessien vakiintumiselle.
Strategian ja toimintamallien puute	Esim. saavutettavuus, tietosuojat, datan jakamisen lainsäädäntö ja kielipolitiikka vaativat erityistoimia, joiden vuoksi alustoja ei voi ottaa käyttöön.
Organisaatiokulttuuri ja resurssit	Ilman selkeitä tavoitteita ja arviointikeinoja alustojen käyttö jää irralliseksi.
	Julkisen hallinnon kyky hyödyntää osallistujadataa ja tukea vuorovaikutusta on rajallinen.

Tulokset osoittavat, että digitaaliset osallistumisalustat soveltuvat hyvin sidosryhmien tarpeiden ja ehdotusten kartoittamiseen elämäntapahtumalähtöisessä palveluiden kehittämisessä. On hyvä huomioida, että:

- Kansalaisosallistumisen arvo tulee tunnistaa eri vaiheissa palvelukehitystä, ja osallistumispisteet tulee suunnitella etukäteen.
- Alustojen käyttöönotto vaatii organisaatioilta selkeitä työskentelytapoja ja resursointia – näitä on tärkeää kehittää ja kirkastaa.
- Vaikuttavuuden arvioimiseksi tarvitaan yhteisiä mittareita ja menetelmiä, jotka tukevat digitaalisten osallistumismenetelmien hyödyntämistä osana palvelumuotoilua.
- Tarvitaan ekosysteemimäinen rakenne, joka tukee osallistamisen systemaattista kytkemistä päätöksentekoon.

Alustojen hyödyntäminen vaikuttavaan kansalaisten osallistumiseen edellyttää oikea-aikaista ja suunnitelmallista kytkemistä suunnittelu- ja päätöksentekoprosesseihin. Tämä vaatii organisaation strategista sitoutumista, operatiivista valmiutta, moniammatillista yhteistyötä ja riittäviä resursseja.²⁶

4. ETAIROS eettisenä kompassina

4.1 Vaikuttavuus syntyy käytännön näkökulman kautta

ETAIROS toimi väylän avaajana ja keskeisenä vaikuttajana tekoälyjärjestelmien eettisessä kehitystyössä. Hankkeessa tehty tutkimus- ja vuorovaikutustyö työ on nostanut tekoälyn eettisen ohjauksen kysymykset näkyvästi suomalaiseen yhteiskunnalliseen keskusteluun ja tarjonnut konkreettisia ratkaisuja ja työkaluja niin päätöksentekijöille kuin yrityksille.

Kuuden vuoden tutkimushankkeen aikana ETAIROS:n vaikutus on nähtävissä mm.:

- Suomen julkisyhteisöillä, kuten ministeriöillä ja kaupungeilla, on paremmat edellytykset rakentaa hallintokehyksiä sosiaalisesti kestäväälle ja osallistavalle tekoälylle.
- Tiedeyhteisöillä ja teollisuudella on tietoa tekoälyn merkityksellisistä sovelluskohteista sekä sen erityisistä haasteista ja mahdollisuuksista.
- Suunnittelijoille on tarjolla eettisen suunnittelun ja arvioinnin puitteita ja työkaluja.
- Päättäjillä ja teollisuudella on ymmärrystä tekoälyn pehmeän sääntelyn tarpeellisuudesta lainsäädännön rinnalla ja tekoälyteknologioiden aiheuttamista sääntelyn muutoksista.
- Kansalaisyhteiskunnalla on ymmärrystä mahdollisuuksista tietoon perustuvaan julkiseen keskusteluun ja vaikuttamiseen.

ETAIROS toimi poikkeuksellisten myllerrysten sävyttämässä ajassa: pandemiaa seurasi tekoälyn todellinen läpimurto ja turvallisuustilanteen muutos. Niin tekoälyn kuin yhteiskunnallisen kehityksen suuntaviivat näkyvät edessämme epävarmoina, mutta merkittävänä. Hankkeen keskeiset

²⁶ Sigfrids, Latvanen & Hantula 2025

kysymykset – kuinka suunnitella, käyttää ja hallita eettisesti hyväksyttäviä tekoälyjärjestelmiä – haastavat edelleen ja yhä suuremmissa määrin niin julkisia kuin yksityisiäkin toimijoita.

ETAIROS-hanke edisti merkittävästi yhteiskunnallista keskustelua tekoäly-yhteiskunnan kehittämisestä ja kehityksen ohjauksesta monitahoisesti ja useissa asiantuntijaryhmissä. Hanke on myös luonut yhteistä ”kieltä” insinööritieteiden, ihmis- ja yhteiskuntatieteiden, oikeustieteen ja hallintotieteen kesken ja tuonut etiikkaa pois ”filosofian siilosta” yhteiseksi keskustelun aiheeksi. Eikä vähäisimpänä yhteistyö julkishallinnon kanssa on saanut virkahenkilöt keskustelemaan ja innostumaan aiheesta, joka ei ole heidän tavanomaisessa agendassaan.

Projektin aikana tapahtunut kehitys toi korostetusti esiin, kuinka uusien teknologioiden hyötyjä ei voida irrottaa niiden käytön riskeistä ja haitoista. Teknisen kehityksen nopeus vie huomion siitä, että uusien teknologioiden mielekäs, hyödyllinen ja suotuisa käyttöönotto ja soveltaminen ovat (useimmiten) vaivalloisia ja hitaita prosesseja. Haviteltujen hyötyjen ja todellisten kustannusten välillä on usein epäsuhta. Hanke työsti välineitä tehdä epäsuhtaa näkyväksi ja käsitellä sitä.

Vaikka tekoälyn etiikkaan liittyviä pulmia osataan nyt nimetä aiempaa paremmin, niiden juurisyyt liittyvät myös laajempiin yhteiskunnallisen tasa-arvon ja hyvinvoinnin kysymyksiin. Jos näitä ei ratkaista, eivät myöskään tekoälyn etiikan ongelmat poistu.

Tekoälyn yleistyessä näkökulmia etiikkaan tulee jatkuvasti lisää. Kysymykset tekoälyn vaikutuksista demokratiaan ja kansalliseen turvallisuuteen ovat kansallisesta näkökulmasta yhä keskeisempiä toimintaympäristön, niin sisäisen kuin ulkoisen, muutoksen hallinnassa. Myös teknologiariippuvaisuuteen ja strategiseen autonomiaan liittyvät kysymykset korostuvat tänä päivänä mm. suurten kielimallien yleistymisen myötä.

4.2 Vuorovaikutus sidosryhmien kanssa: innovaatiofoorumit ja työpajat

Teemakohtaisten tarkastelujen avulla pyrimme kohti yleistettävää ymmärrystä tekoälyn eettisistä kysymyksistä. Tekoälyn käyttötapauksia kehitettiin ja innovoitiin yhdessä innovaatiokumppanien kanssa ennakoinnin, suunnittelun ja hallinnan näkökulmasta. Näin yhdistettiin ETAIROS:n suoraan ajankohtaisiin tutkimus- ja kehitystarpeisiin ja eri viranomaisten meneillään olevaan toimintaan. Tunnistimme temaattisten alueiden haasteita ja haimme niihin konkreettisia ratkaisuja sidosryhmätyöpajojen avulla.

Vuorovaikutus sidosryhmien – ennen kaikkea valtion- ja julkishallinnon toimijoiden kanssa oli hankkeen keskeinen menetelmällinen tapa toimia. ETAIROS toteutti ns. *yhteisinnovaatiofoorumien* (co-innovation forums) mallia, jossa päättäjät, yksityisen sektorin toimijat, suunnittelijat ja teknologiakehittäjät keskusteleval yhdessä tekoälyn eettisesti kestävästä hyödyntämisestä koskevista kysymyksistä. Forumeissa käytiin vilkasta keskustelua tekoälyn mahdollisuuksista, mutta erityisesti sen käyttöön liittyvistä ajankohtaisista ongelmista ja vielä ratkaisemattomista kysymyksistä. Osa innovaatiofoorumeista järjestettiin yhdessä Business Tampereen kanssa, mm. AI Finland -tapahtuman yhteydessä, sekä SuomiAreenan kanssa. Konkreettisia tarkastelukohteita olivat mm. ihmisten elämäntapahtumia tukeva kansallinen tekoälyohjelma AuroraAI, turvallisuus yksilön kokemuksena sekä julkisen sektorin toiminnassa ja kansallisena haasteena, tekoäly riskienhallinnan, tiedolla johtamisen ja päätöksenteon näkökulmista, sekä tulevaisuuden liikkuminen ja liikenne. Eri näkökulmia avasivat Valtiovarainministeriön, Työterveyslaitoksen, Tampereen kaupungin, Digi- ja väestötietoviraston, Liikenne- ja viestintäministeriön sekä ETAIROS:n asiantuntijat.

Nämä yhteiskehittämisen menetelmin toteutettavat forumit mahdollistivat yhdessä oppimisen. Niiden avulla ETAIROS toimi yhteisen innovoinnin alustana, jossa oli mahdollista keskustella

käytössä olevista ja tulevista tekoälyn käyttötapauksista ja ratkaisuvaihtoehtoista ja arvioida niitä monitieteisesti ja monitoimijaisesti.

Yli 20 sidosryhmätyöpajassa käsitelimme tekoälyn vaikutusten arviointia, vastuullisia ekosysteemejä, kestäviä digitaalisia hankintoja, kansallista (kyber)turvallisuutta, AI-lainsäädäntötyötä, älykästä raitiovaunua, itseohjautuvia ajoneuvoja, toimintakykyä ja terveyttä seuraavia bioantureita, terveydenhuollon chatbottia, itsemääräämisoikeutta suhteessa kansalaisista kerättävään dataan sekä julkishallintoa ja ministeriöitä kielimallien hyödyntäjänä. Olemme tarkastelleet, miten tekoälyä kehitetään, millaiset arvot ohjaavat kehitystä, millaisia yhteiskunnallisia haasteita tekoälyn avulla pyritään ratkaisemaan tai haluttaisiin ratkaista, millaisia vaikutuksia kehitystyöllä on yksilöiden, yhteisöjen ja yhteiskunnan näkökulmasta ja millaisia ohjauskeinoja tekoälyn hyödyntämiseen kaivataan.

ETAIROS hyödynsi keskusteluissa tehokkaasti digitaalisia alustoja. Covid-19-pandemian vuoksi kevään 2020 tilaisuudet toteutettiin verkkokeskusteluina käyttämällä Howspace-alustaa. Hyvin pian keskeisiksi työskentelytiloiksi nousivat Teams- ja Zoom -alustat. Yhteiset työpajat olivat diskursiivisia, ja niissä hyödynnettiin luovuutta ja ryhmähenkeä tukevia menetelmiä. Niiden avulla olemme yhdessä muodostaneet tarkoituksenmukaisia kysymyksiä erilaisiin käyttötapauksiin ja etsimään vaihtoehtoisia näkökulmia ja ratkaisuja.

Tutkijayhteisössä kuuden vuoden aikana ETAIROS toi yhteen niin insinööri-, yhteiskunta- ja oikeustieteilijät kuin hallinnon ja yritysmaailman toimijoita. Hanke tuotti työkaluja muun muassa eettiseen hallintaan julkishallinnolle ja yrityksille sekä kehitti osallistumisen ja osallisuuden digitaalisia alustoja. Case-yhteistyötä tehtiin esimerkiksi terveydenhuollon, puolustusministeriön, Sitran, Valtioneuvoston kanslian sekä Oulun ja Tampereen kaupunkien kanssa. Lisäksi tarjottiin asiantuntija-apua ministeriöille, yrityksille ja järjestettiin lukuisia puhujatilaisuuksia, julkaisuja sekä podcasteja.

4.3 Vaikuttavuus syntyy yhteistyöstä ja yhteiskehittämisestä

ETAIROS:n tavoitteisiin päästiin kokoamalla luovaan tutkimus- ja innovaatioprosessiin niin tutkijat, julkiset toimijat ja päätöksentekijät kuin yritykset ja kansalaisetkin. Keskeiset teemakumppanit edustivat eri viranomaisia ja toivat tekoälyn temaattiset, olemassa olevat ja tulevat käyttötapaukset yhteiseen keskusteluun. Myös hankkeen sidosryhmien organisaatiot ja yritysten edustajat osallistuivat yhteiseen pohdintaan.

Hankkeen aikana järjestettiin lukuisia työpajoja ja seminaareja, sparrausklinikoita ja demokratiakorjaamoja, kehitettiin politiikkasuosituksia ja julkaistiin ratkaisukortteja tekoälyn vastuulliseen hyödyntämiseen. ETAIROS oli mukana rakentamassa koulutusmateriaaleja sekä osallistui lukuisiin kansallisiin ja kansainvälisiin keskusteluihin ja työryhmiin, joissa edistettiin tekoälyn etiikkaa ja vastuunjakoa.

Yhteistyö eri organisaatioiden ja verkostojen kanssa on ollut laajaa. Hankkeen tutkijat ovat osallistuneet aktiivisesti mm. kansainvälisiin konferensseihin, asiantuntijapaneelisiin ja julkaisufoorumeihin. Tieteellisen tutkimuksen lisäksi hankkeen asiantuntijat ovat esiintyneet laajasti mediassa, osallistuneet podcasteihin ja julkaisseet blogeja sekä mielipidekirjoituksia, joissa on pohdittu tekoälyn vastuullisuutta, sääntelyä ja yhteiskunnallisia vaikutuksia.

4.4 Vaikuttavuus syntyy julkisten toimijoiden muuttuvan roolin tukemisesta

ETAIROS toteutti konkreettisiin kysymyksiin paneutuvaa yhteistyötä eri ministeriöiden ja muiden julkishallinnon toimijoiden kanssa, ja vaikutti näin ministeriöiden poliittiseen valmistelutyöhön. Konkreettisia, tarkastelun kohteeksi asetettuja ongelmia olivat erilaiset aineiston käsittelyn ja sisällöntuotannon kysymykset sekä tekoälysovellusten mallintaminen ja kokeilu ministeriöiden tuottaman aineiston pohjalta. ETAIROS:n tutkijat olivat Valtiovarainministeriön Tekoälyn ja digitalisaation tutkimuksen asiantuntijaryhmän jäsenenä järjestämässä valtakunnallista *Tulevaisuuden tekoäly -webinaaria*, kommentoimassa *EU-komission White Paperia* tekoälytutkimuksen ja -kehityksen suuntaviivoista sekä lausumassa *AI-ACT -lakiehdotuksen* kehitystyöstä. Samoin osallistuimme aktiivisesti Valtiovarainministeriön *AuroraAI:n* Etiikkaryhmän työskentelyyn sekä *AuroraAI:n* Osaaminen -teemaryhmään ihmiskeskeisen ja ennakoivan yhteiskunnan näkökulmasta. Olimme myös jäsenenä *Eduskunnan Tulevaisuusvaliokunnan* Ikääntymisen tulevaisuus -työryhmässä roolinaan tarkastella ja tuoda yhteiseen keskusteluun uusien teknologioiden eettisiä näkökulmia.

Osallistuimme aktiivisesti Tulevaisuusvaliokunnan tekoälytyöryhmään ja *Tekoälyraportin* kirjoittamiseen (2022), sekä Hyvinvointi - *Pahoinvointi -raportin* kirjoittamiseen (2022). Etairolaiset ovat myös lausuneet TEM:lle digikompassiesityksestä sekä VM:lle digitaalista henkilöllisyyttä koskevasta esityksestä. Keväällä 2020 osallistuimme OKM:n rahoittamaan *SotePeda* -työskentelyyn yhdessä sote-alan koulutusta tarjoavien ammattikorkeakoulujen kanssa. Järjestimme neljä viikkoa kestäneen verkkokeskustelun digitaalisuuden vaikutuksista sote-alaan, keräsimme aineistoa ja laadimme alkuvuodesta 2021 ilmestyvän yhteisjulkaisun *Tulevaisuuden teknologiat ja tekoälyn etiikka*. Syksyllä 2020 järjestimme yhdessä valtakunnallisen Digituki -hankkeen kanssa kaksi viikkoa kestäneen virtuaalisen keskustelufoorumin tekoälyn ja digiteknologian käyttöön liittyvistä mahdollisuuksista, unelmista, esteistä ja haitoista.

4.5 Eettisen suorituskyvyn mittaristo

Luottamusta voidaan rakentaa myös johtamisen työkalujen avulla. Indikaattorit ja niiden mittaaminen ovat yksi johtamisen järjestelmä. ETAIROS-hankkeessa kehitettiin eettisen suorituskyvyn mittaristoa (ERL Ethical Readiness Level) organisaation eettisen kypsyyden arviointiin (Taulukko 2).

Esimerkkinä käytettiin NASAn kehittämää teknisen suorituskyvyn indikaattorijärjestelmää²⁷ (TRL Technology Readiness Level). Eettistä suoritustasoa arvioidaan ennalta valittujen kypsyydentasoluokkien parametrien perusteella, minkä jälkeen organisaatiolle annetaan ERL-luokitus. Eettisen kypsyyden valmiustasoja on yhdeksän. ERL 1 on alhaisin ja ERL 9 on korkein valmiustaso.

ETAIROS:n ERL-mittariston avulla organisaatio voi arvioida eettistä suorituskykyään tarkastelemalla järjestelmiä, joiden avulla se pyrkii ottamaan eettiset kysymykset huomioon päätöksiä tehtäessä.

²⁷ NASA TRL <https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/>

Taulukko 2. ETAIROS ERL-mittaristo

ERL taso	ETAIROS indikaattori
1	Käyttötapausten alustavat ideat. Eettisten kysymysten peruseriaatteet havaittu.
2	Eettiset huolenaiheet ja vahvuudet tunnistettu sidosryhmien ja teknologia-asiantuntijoiden kanssa.
3	Eettiset huolenaiheet ja vahvuudet konkretisoitu sidosryhmien ja asiantuntijoiden kanssa.
4	Eettiset huolenaiheet ja vahvuudet validoitu esimerkiksi laboratoriossa, skenaarioissa, malleissa tai prototyypissä.
5	Eettiset huolenaiheet ja vahvuudet testattu pilotissa.
6	Eettisten kysymysten analyysi syvennetty etiikan asiantuntijoiden kanssa.
7	Eettisten kysymysten lopulliset testit. Täydellinen eettinen demonstraatio sidosryhmien kanssa.
8	Eettinen analyysi loppuun suoritettu.
9	Varmistetaan eettisten kysymysten jatkuva hallinta. Jatkuva parantaminen, eettisten kysymysten tunnistaminen, analysointi ja hallinta.

5. Julkaisut ja media

6.1 VTT:n vertaisarvioidut artikkelit

ETAIROS tuotti yli 120 tieteellistä vertaisarvioitua artikkelia (ks LIITE1).

VTT:n julkaisujen (66 julkaisua) julkaisutiedot on esitetty alla:

- 1 Ailisto, H., Myllymäki, P., Tarkoma, S., Kämäräinen, J.-K., Röning, J., Salakoski, T., Solin, A., Saariluoma, P., Mikkonen, T., van Gils, M., Väänänen, K., Puolamäki, K., Ylén, P., Roos, T., **Leikas**, J., Honkela, A., Kutila, M., Ruotsalainen, L., Ylikoski, P., & Linturi, R. (2022). Tekoälyratkaisut tänään ja tulevaisuudessa. Eduskunnan tulevaisuusvaliokunnan julkaisu.
- 2 Bhatia, R., **Rilla**, N., Laitinen E., & Ferreira, G. (2023). Benchmarking report on gender equality in science, technology and innovation (STI) dialogues.
- 3 Bhatia, R., **Rilla**, N., Laitinen E., Ferreira, G., & Milhazes, C. (2023). Integration of gender in international science, technology and innovation (STI) collaboration: learning from international feminist policies.
- 4 Bhatia, R., **Rilla**, N., Sanchez Nieminen, G., Merisalo, M., & Salonen, T. (2022). From persisting gender inequality to inclusion in research and innovation content: challenging the gender equality norm in the STI fields.
- 5 Christensen, M.V., **Nieminen**, M., Altenhofer, M., Tangcoigne, E., Mejlgaard, N., Griessler, E., & Filacek, A. (2020). What's in a name? Perceptions and promotion of responsible research and innovation practices across Europe. Science and Public Policy.
- 6 Gianni, R., **Lehtinen**, S., **Nieminen**, M. (2022). Governance of Responsible AI: From Ethical Guidelines to Cooperative Policies. Frontiers in Computer Science.
- 7 **Gotcheva**, N., & **Wessberg**, N. (2024). Ethical foresight approach for anticipating societal impacts of AI technologies. Handbook of Services and AI.
- 8 Hallamaa, J., **Leikas**, J., Malkavaara, M., & Vesterinen O. (2020). Tulevaisuuden teknologiat ja tekoälyn etiikka. In: P. Sihvo & A. Koski (Eds.) Eettinen toimintamalli - Osaamista tulevaisuuden koulutukseen ja sote-alan työhön. Karelia ammattikorkeakoulun julkaisuja B:65. Joensuu. ISBN 978-952-275-314-4, ISSN- L 2323-6876.
https://www.theseus.fi/bitstream/handle/10024/347276/B65_Eettinen_toimintamalli.pdf?sequence=7&isAllowed=y
- 9 Hallamaa, J., **Leikas**, J., Sihvo, P., Malkavaara, M., Vesterinen, O., Parviainen, T., & Koski, A. (2020). Virtuaalinen eettinen foorumi 2 : Raportti foorumin toteutuksesta ja tuloksista. Karelia- ammattikorkeakoulun julkaisuja C, Raportteja: 68.
- 10 **Heinisuo**, E., Kuoppakangas, P. & Stenvall, J. (2025). Navigating AI Implementation in Local Government: Addressing Dilemmas by Fostering Mutuality and Meaningfulness. (2025) *Information Systems Frontier*. <https://doi.org/10.1007/s10796-025-10599-x>
- 11 **Karvonen**, A., Kujala, T., Kärkkäinen, T., & Saariluoma, P. (2023). Fundamental concepts of cognitive mimetics. Cognitive Systems Research.
- 12 **Karvonen**, A., Kujala, T., & Saariluoma, P. (2020). Types of mimetics for the design of intelligent technologies. Advances in Intelligent Systems and Computing.
- 13 **Karvonen**, A., & **Leikas**, J. (2023). Tekoälyjärjestelmien vastuullinen ja eettinen suunnittelu ja käyttö. Teoksessa: Tietokooste 2023 - Digitaalisuus, työn murros ja sosiaaliturva - tietokooste. Suomen Akatemia. ISBN 978-951-715-950-0 [Digitalisaatio, työn murros ja sosiaaliturva-tietokooste \(aka.fi\)](#)

- 14 **Karvonen, A., Leikas, J., Wessberg, N., Sigfrids, A., Pöntinen, S.** (2026). Some Problems in the Ethical Impact Assessment of Emerging Technologies and Socio-Technical Visions: Case CityVerse. In: Alvarez, I., Arias-Oliva, M., Dediu, A.H., Silva, N. (eds) Ethical and Social Impacts of Information and Communication Technology. ETHICOMP 2025. Lecture Notes in Computer Science, vol 15939. Springer, Cham. https://doi.org/10.1007/978-3-032-01429-0_32
- 15 **Karvonen, A., & Saariluoma, P.** (2023). Cognitive mimetics and human digital twins: Towards holistic AI design. ERCIM News.
- 16 **Koskimies E., & Kinder, T.** (2022). Mutuality in AI-enabled new local public service solutions. Special issue on AI technologies. Public Management Review. <https://www.tandfonline.com/doi/pdf/10.1080/14719037.2022.2078501>
- 17 **Koskimies, E., Nieminen, M., Stenvall, J., Hallamaa, J., Leikas, J., & Salo-Pöntinen, H.** (2020). ETAIROS Policy Brief: Suomeen tarvitaan koordinoitu, vastuullisuutta korostava tekoälypolitiikka. https://etairos.fi/wp-content/uploads/2021/02/etairos-politiikkamietinto_suomeen-tarvitaan-koordinoitu-vastuullisuutta-korostava-tekoalypolitiikka.pdf
- 18 **Koskimies, E., Stenvall, J., Kinder, T., Leikas, J. & Nieminen, M.** (2022). Artificial intelligence and Public Innovations. In: H. Väyrynen, N. Helander & H. Jalonen (Eds.) Public Innovation and Digital Transformation. Routledge. ISBN 9781032137414 <https://doi.org/10.4324/9781003230854>
- 19 Kuha, S., Xiong, E., Hammarén, M., Ylisirniö, M., Haverinen, J., Suominen, J., **Leikas, J., & Kanste, O.** (2023). Privacy protection, data security and protection in technology utilization in home care services for the older people: A survey study of managers. Finnish Journal of EHealth and EWelfare, 15(3), 338–352. <https://doi.org/10.23996/fjhw.127242>
- 20 **Laitinen, E., Sanchez Nieminen G., Rilla, N. & Lähteenmäki-Smith, K.** (2024). Gender, justice and the green hydrogen transition: moving to inclusive policymaking.
- 21 **Lanne, M., & Leikas, J.** (2021). Ethical AI in the re-ablement of older people – Opportunities and challenges. Gerontechnology (2021) 20:2 <https://doi.org/10.4017/gt.2021.20.2.26-473.11>
- 22 **Lanne, M., Nieminen, M., & Leikas, J.** (2025). Organisational tensions in introducing socially sustainable AI. AI & Soc. <https://doi.org/10.1007/s00146-025-02293-y>
- 23 **Lanne, M., & Palomäki, K.** (2022). Tensions in sustainable business: burdens or enablers? LUT Scientific and Expertise Publications: Research Reports.
- 24 **Lanne, M., & Palomäki, K.** (2023). Stakeholder engagement in developing sustainable business.
- 25 **Leikas, J.** (2020). Aesthetic Wellbeing and Ethical Design of Technology. In: Rousi, R., **Leikas, J., & Saariluoma, P.** (Eds.). Emotions in technology design: from experiences to ethics. Springer.
- 26 **Leikas, J.** (2020). Ikäteknologian kehittämiseen kaivataan todellista keskustelua ihmisen hyvästä. Vanhustyö 2/2020, p. 6-9. https://vtkl.fi/wp-content/uploads/2020/03/Vanhustyö_0220.pdf
- 27 **Leikas, J.** (2022). Tekoälystä vastuullista. Teoksessa: Ailisto, H., Myllymäki, P., Tarkoma, S., Kämäräinen, J-K., Röning, J., Salakoski, T., Solin, A., Saariluoma, P., Mikkonen, T., van Gils, M., Väänänen, K., Puolamäki, K., Ylén, P., Roos, T., Leikas, J., Honkela, A., Kutila, M., Ruotsalainen, L., Ylikoski, P., & Linturi, R. (2022). Tekoälyratkaisut tänään ja tulevaisuudessa. Eduskunnan tulevaisuusvaliokunnan julkaisu 1/2022. Eduskunta. pp. 114-120. https://www.parliament.fi/FI/naineduskuntatoimii/julkaisut/Documents/tuvj_1+2022.pdf
- 28 **Leikas, J.** (2022). Eettisesti kestävä teknologia. Kannanotto teoksessa Seikkula, A., & Petäjäjärvi, K. (toim.) Ikääntymisen tulevaisuuksia 2060 - taidevälitteinen asiantuntijadialogi uuden ajattelun lähteenä. Eduskunnan Tulevaisuusvaliokunnan julkaisu 2/2022.
- 29 **Leikas, J.** (2023). Digihyvinvointi kasvattaa pahoinvointia - Lisääntykö epätasa-arvo uusien digiteknologioiden myötä? Teoksessa: Miten Suomi voi nyt ja tulevaisuudessa? Eduskunnan tulevaisuusvaliokunnan julkaisu 6/2022. Eduskunta. pp. 86-96. <https://www.eduskunta.fi/FI/naineduskuntatoimii/julkaisut/Documents/TUVJ-6-22.pdf>

- 30 **Leikas, J., & Floréen, P.** (2022). A Systematic Exercise Tool Helps Researchers Ponder the Ethical Implications of AI. ERCIM News No. 131. Special theme: Ethical Software Engineering and Ethically Aligned Design, pp. 11-12. [A Systematic Exercise Tool Helps Researchers Ponder the Ethical Implications of A \(ercim.eu\)](https://ercim.eu)
- 31 **Leikas, J., Gotcheva, N., Nieminen, M., Koivisto, R., Rilla, N., Sigfrids, A.** (2019). Ethical AI for the Governance of the Society. Conference Poster, presented at FCAI AI Day, Espoo, Finland, 26 Nov 2019.
- 32 **Leikas, J., Halkoaho, A., & Lanne, M.** (2024). Informed Consent in the age of smart technologies. Finnish Journal of eHealth and eWelfare. FinJeHeW 2024;16(3) 296-308. <https://doi.org/10.23996/fjhw.137992>
- 33 **Leikas, J., Johri, A., Latvanen, M., Wessberg, N., & Hahto, A.** (2022). Governing Ethical AI Transformation: A Case Study of AuroraAI. *Front. Artif. Intell.* 5:836557. <https://doi.org/10.3389/frai.2022.836557>
- 34 **Leikas, J., Koivisto, R., & Gotcheva, N.** (2019). Ethical framework for designing autonomous systems. Journal of Open Innovation: Technology, Market, and Complexity (ISSN 2199-8531). Special Issue on Technology Driven Innovation, Research Management and Policy Making. J. Open Innov. Technol. Mark. Complex. 2019, 5, 18; <https://www.mdpi.com/2199-8531/5/1/18/htm>
- 35 **Leikas, J., Nieminen, M., Koskimies, E., & Stenvall, J.** (2021). Tekoäly julkishallinnon työkaluna. Kuntalehti. <https://kuntalehti.fi/uutiset/tekniikka/asiantuntija-artikkeli-tekoaly-julkishallinnon-tyokaluna/>
- 36 **Leikas, J., Rousi, R., Vilpponen, H., & Saariluoma, P.** (2024). Super-intelligent society for the silver segment: Ethics in design. In: Olsson, T., Sahlgren, O., Parviainen, J., Westerstrand, S., Harviainen, J.T., Laitinen, A., & Rantala, J. (Eds.) Proceedings of the Conference on Technology Ethics 2024 (Tethics 2024), Tampere, Finland, November 6-7, 2024. p. 96-106. https://ceur-ws.org/Vol-3901/paper_7.pdf
- 37 **Leikas, J., Sigfrids, A., Stenvall, J., & Nieminen, M.** (2020). Good Life Ecosystems – Ethics and Responsibility in the Silver Market. In: Rauterberg M. (Ed.) Culture and Computing. HCII 2020. Lecture Notes in Computer Science, vol 12215. Springer, Cham, pp. 105-122. https://doi.org/10.1007/978-3-030-50267-6_9
- 38 **Leikas, J., Vilpponen, H., Saariluoma, P.** (2021). Boomerit digiuudistusten pomppulaudalla. Maaseudun Tulevaisuus, 29.11.2021.
- 39 **Miettinen, J. J., Mäkinen, M., Leikas, J., Jutila, T., & Veko, T.** (2020). Vastuullinen toiminnanohjausjärjestelmän kehittäminen kotihoidossa [Responsible development of an enterprise resource planning system in home health care]. Finnish Journal of EHealth and EWelfare, 12(1), 70-83. <https://doi.org/10.23996/fjhw.77883>
- 40 Myllyoja, J., **Rilla, N., & Lima-Toivanen, M.** (2022). Strengthening futures-oriented agenda for building innovation ecosystems. European Journal of Futures Research.
- 41 **Nieminen, M., Gotcheva, N., Leikas, J., & Koivisto, R.** (2019), Ethical AI for the Governance of the Society: Challenges and Opportunities. Proceedings of Third Annual Seminar on Technology Ethics 2019, Turku, Finland, October 23-24, 2019. Rantanen, M. M. & Koskinen, J. (eds.). CEUR-WS, 7 p. (CEUR Workshop Proceedings, Vol. 2505), pp. 20-27.
- 42 **Nieminen, M., Lanne, M., & Leikas, J.** (2022). Organizational tensions in the introduction of socially sustainable AI. Strategic Research - Scientific Conference: A fair, just and sustainable society 12.-13.10.2022, Helsinki.
- 43 **Nieminen, M., Karvonen, A., Mervaala, E., Pyykkönen, S.** (2025). Osallisuus, dialogi, päätöksenteko – uuttuvan yhteiskunnan tulevaisuuskestävä ohjaus.
- 44 **Rousi, R., Leikas, J., & Saariluoma, P.** (Eds.) (2020). Emotions in Technology Design: From experience to Ethics. Human-Computer Interaction Series. Springer. ISBN-10: 3030534820, ISBN-13: 978-3030534820. DOI <https://doi.org/10.1007/978-3-030-53483-7> <https://link.springer.com/book/10.1007/978-3-030-53483-7>

- 45 Rousi, R., **Leikas**, J., & Saariluoma, P. (2020). Feelings Matter. Human - Computer Interaction Series. Human-Computer Interaction Series. Springer. ISBN-10: 3030534820, ISBN-13: 978-3030534820. DOI https://doi.org/10.1007/978-3-030-53483-7_1
- 46 Rousi, R., Saariluoma, P., & **Nieminen**, M. (2022). Editorial: Governance AI ethics. Frontiers in Computer Science.
- 47 Saariluoma, P., Cañas, J., & **Karvonen**, A. (2020). Human digital twins and cognitive mimetic. Advances in Intelligent Systems and Computing.
- 48 Saariluoma, P., & **Karvonen**, A. (2024). Theory languages in designing artificial intelligence. AI and Society.
- 49 Saariluoma, P., **Karvonen**, A., Cañas, J., & **Leikas**, J. (2024). Working with human thinking in developing intelligent technologies. 44th IBIMA Computer Science Conference, 27-28 Nov 2024, Granada, Spain Working with Human Thinking in Developing Intelligent Technologies | International Business Information Management Association (IBIMA)
- 50 Saariluoma, P., **Karvonen**, A., & Kujala, T. (2021). Conceptual basis of cognitive mimetics for information engineering. Journal of Physics: Conference Series.
- 51 Saariluoma, P., **Karvonen**, A., & Sorsamäki, L. (2021). Human digital twins in acquiring information about human mental processes for cognitive mimetics. Frontiers in Artificial Intelligence and Applications.
- 52 Saariluoma, P., & **Leikas**, J. (2019). Ethics in designing autonomous systems. In: T. Ahram et al. (Eds.): Human Interaction and Emerging Technologies. Proceedings of the IHiet 2019, August 22-24, Nice, France, AISC 1018, pp. 47–52, 2020. ISBN 978-3-030-25628-9; ISBN 978-3-030-25629-6 (e-book). https://doi.org/10.1007/978-3-030-25629-6_8
- 53 Saariluoma P., & **Leikas**, J. (2020). Designing Ethical AI in the Shadow of Hume's Guillotine. In: Ahram T., Karwowski W., Vergnano A., Leali F., Taiar R. (eds) Intelligent Human Systems Integration 2020. IHSI 2020. Advances in Intelligent Systems and Computing, vol 1131, pp. 594-599. Springer, Cham https://doi.org/10.1007/978-3-030-39512-4_92
- 54 Saariluoma, P., & **Leikas**, J. (2020). Ethics in designing intelligent systems. Advances in Intelligent Systems and Computing.
- 55 Saariluoma, P., Myllylä, M., & **Karvonen**, A. (2022). Mental contents in designing AI ethics. Lecture notes in computer science.
- 56 Saariluoma, P., Myllylä, M., & **Karvonen**, A. (2023). Human digital twins in interaction design: from abstract to concrete.
- 57 **Salonen**, T., **Rilla**, N., & **Pikkuaho**, P. (2022). Eettinen tekoäly autonomisissa järjestelmissä. VTT Technology No 418.
- 58 Sanchez Nieminen, G., **Rilla**, N., Bhatia, R., Merisalo, M., Ferreira, G. (2022). Moving away from traditional STI: How to improve gender equality and inclusivity in STI?
- 59 **Sigfrids**, A., **Karvonen**, A., **Rilla**, N., & **Lanne**, M. (2024). Governing AI services: a state-of-the-art overview. Handbook of Services and Artificial Intelligence. Edward Elgar Publishing Ltd, p. 308-329 21 p. <https://doi.org/10.4337/9781035301973.00029>
- 60 **Sigfrids**, A., Latvanen, M., & Hantula, K. (2025). Digitaalinen kansalaisosallistuminen julkisessa palvelukehittämisessä. Kenttäkokeilun loppuraportti. VTT Technology No 437.
- 61 **Sigfrids**, A., **Leikas**, J., Salo-Pöntinen, H., & **Koskimies**, E. (2023A). Human-centricity in AI governance: A systemic approach. Frontiers. <https://doi.org/10.3389/frai.2023.976887>
- 62 **Sigfrids**, A., **Nieminen**, M., **Leikas**, J., & **Pikkuaho**, P. (2022). How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI. Front. Hum. Dyn. 4:858108. <https://doi.org/10.3389/fhumd.2022.858108>
- 63 **Sigfrids**, A., **Nieminen**, M., **Leikas**, J., **Karvonen**, A., & **Pikkuaho**, P. (2023B). Julkishallinto tukemassa eettisen tekoäly-yhteiskunnan rakentamista: Katsaus tekoälyn ohjauskeinoihin ja politiikkatoimenpiteisiin. VTT Technical Research Centre of Finland. VTT Technology No. 421. <https://doi.org/10.32040/2242-122X.2023.T421>

- 64 Vilpponen, H., **Leikas**, J., & Saariluoma, P. (2020). Designing digital well-being of senior citizens. 13th International Conference on Human System Interaction (HSI), Tokyo, Japan, 2020, pp. 40-44, <https://doi.org/10.1109/HSI49210.2020.9142655>
- 65 **Wahlström**, M., **Tammentie**, B., **Salonen**, T., **Karvonen**, A. (2024). AI and the transformation of industrial work: Hybrid intelligence vs double-black box effect. Applied Ergonomics.
- 66 **Wessberg**, N., & **Leikas**, J. (2021). Tekoälyn tulevaisuuksien eettiset kysymykset sosio-teknisissä skenaarioissa.) Futura 3/21, pp. 28-32.

6.2 ETAIROS-podcast 10 pulmaa tekoälystä

Keväällä 2022 toteutimme yhdessä Suomen Podcastmedian kanssa 10 keskustelun podcast-sarjan "10 pulmaa tekoälystä". Sarjassa selvitettiin, miten tekoälyyn liittyviä pulmia voitaisiin ratkaista. Kanssakeskustelijoina olivat tekoälyä eri perspektiiveistä tarkastelevat ansioituneet tutkijat ja vaikuttajat, joiden kanssa pureuduttiin tekoälyn hyötyihin sekä keinoihin hillitä tekoälyn luomia riskejä. Loppuvuodesta 2022 toteutimme näkyvyyttä lisäävän kampanjan Twitterissä. Podcast nousi oman kategoriansa sijalle 7 Suomessa ja hyvällä "pitokäyrällä": keskimäärin $\frac{2}{3}$ kuulijoista kuunteli jakson alusta loppuun.

Podcastin toisen tuotantokauden ensimmäiset jaksot julkaistiin marras-joulukuussa 2023. Niissä keskusteltiin tutkijoiden, virkamiesten ja yritysjohtajien kanssa tekoälyn kehittämistä ohjaavista mielikuvista, vastuullisuudesta, sääntelyn kipukohdista, algoritmisesta syrjinnän ehkäisemisestä sekä tekoälyn moraalista.



Kuva 12. ETAIROS-hankkeen tuottama pod-cast -sarja "10 pulmaa tekoälystä".

Pod-cast -sarjan vastaavina toimittajina olivat ensimmäisellä tuotantokaudella Taina Kalliokoski Helsingin yliopistosta ja Pietari Pikkuaho VTT:ltä, sekä toisella tuotantokaudella Antero Karvonen VTT:ltä ja Taina Kalliokoski. <https://open.spotify.com/show/26gopYUn5GUW2z5CjwQ3o9>

- Kaksoissiirtymä 21.6.2023 & 28.6.2023 Antero Karvonen, Nina Wessberg & Mari Myllylä. Yhteispodcast ORSI-hankkeen kanssa.
- Onko tekoäly taianomainen, ehtymätön puuropata? Antero Karvonen, Jaana Hallamaa & Arno Solin
- Kun tekoäly hoitaa hommat, kuka kantaa vastuun? Antero Karvonen, Anna Seppänen, CoHumans & Marinka Lanne
- Tekoälyn sääntely on hampaatonta – vai onko? Antero Karvonen, Mika Viljanen & Meeri Haataja, Saidot

- Miten moraali saadaan sisään tekoälyyn? Antero Karvonen, Jyrki Nummenmaa & Kai-Kristian Kemell, HY
- Ei kai tekoäly nyt ketään syrji? Antero Karvonen, Jaakko Peltonen & Marko Latvanen, DVV
- Millaista on kasvaa, elää ja ikääntyä tekoälyn kyllästyneessä maailmassa? Antero Karvonen, Raine Haikarainen & Jaana Leikas
- Miten vaalitaan luottamusta mediaan kun tekoäly suoltaa 100 feikkiuutista minuutissa? Antero Karvonen, Niina Viitanen (IS) & Johanna Törn-Mangs (YLE)
- Voisitko asua tekoälykaupungissa, jossa rikollisuutta kitketään kasvojen tunnistusteknologialla? Antero Karvonen, Sanni Pöntinen & Teppo Rantanen (Tampereen kaupunki)
- Datauunarit – kuka louhii ja jalostaa tekoälyn raaka-aineet? Antero Karvonen, Karoliina Snell & Marja Alastalo (DataLit-hanke)
- Millaisen tekoälytulevaisuuden haluamme? Antero Karvonen, Sirkka Heinonen & Nina Wessberg

6.3 VTT:n toisen rahoituskauden (2022-2025) mielipidekirjoitukset, blogit ja populaariartikkelit

ETAIROS/VTT osallistui kuuden vuoden aikana lukuisiin tapahtumiin, kutsuluentoisiin, pyöreään pöydän keskusteluihin, paneelisiin ja koulutuksen järjestämiseen. Aiheina olivat mm. vastuullinen tekoälykehitys, tekoälylainsäädäntö, tiedolliset oikeudet, avattaret ja ihmisyyden, tulevaisuuden kaupunki ja metaverse, tekoäly ikääntymisen haasteiden ratkojana, tekoäly työelämässä, tekoälyn edistäminen PK-yhtiöissä, tekoäly ja terveysdata sekä tekoälyn uhat. Alla on esitetty tutkijoiden artikkeleja ja blogeja toisen rahoituskauden ajalta.

- Karvonen, A., & Leikas, J. Tekoälyn eettinen soveltaminen edellyttää johtamista
- Lanne, M. Sosiaalinen kestävyys on liiketoiminnan kulmakivi
- Lanne, M. Yhteistyöllä kohti sosiaalista kestävyyttä
- Leikas, J. Teknologiausko, luottamus ja demokratia
- Leikas, J. Does ChatGPT make us lazy?
- Leikas, J. Jalkapallosta, roboteista ja tasa-arvosta
- Leikas, J. Ten common challenges for AI—should we be worried?
- Leikas, J. Tekoälyn 10 yleistä haastetta – pitääkö olla huolissaan?
- Leikas, J., Palo, T., & Hiltunen, S. Suomalaisilla yrityksillä etulyöntiasema kasvavilla hopeamarkkinoilla.
- Leikas, J., Wessberg, N., Karvonen, A., & Sigfrids, A. Digitaalinen demokratia ja sanojen voima.
- Snell, K., Floreen, P., Guckelsberger, C., Hallamaa, J., Leikas, J., & Schütte, B. Ethics is not a 15-minute box-ticking exercise
- Wessberg, N. Agenda 2030 on kohta nurkan takana – missä mennään ratkaisuihin?
Agenda 2030 on | VTT (vttresearch.com)
- Wessberg, N. Helsingin Sanomat / mielipide Tekoälypuolueet raikastaisivat demokratiaa - Mielipide | HS.fi
- Saariluoma, P., Karvonen, A. & Leikas, J. Aamulehti / Lukijalta: Tekoälyn sääntely on välttämätöntä
- Wessberg, N. What we learned from AuroraAI: the pitfalls of doing ethics around unsettled technologies

Lähteet

- Bartoloni, S., Calò, E., Marinelli, L., Pascucci, F., Dezi, L., Carayannis, E., Revel, G.M., & Gregori, G.L. (2022). Towards designing society 5.0 solutions: The new Quintuple Helix - Design Thinking approach to technology. *Technovation*, Elsevier, vol. 113 (2022). <https://doi.org/10.1016/j.technovation.2021.102413>
- Fukuyama, M. (2018). Society 5.0: Aiming for a new human-centered society. *Japan Spotlight*, 27.5 (2018): 47-50. https://www.hitachihyoron.com/rev/archive/2017/r2017_06/trends/index.html
- International Association for Public Participation (2025). IAP2 core values for public participation. <https://www.iap2.org/page/corevalues>
- Karvonen, A., Leikas, J., Wessberg, N., Sigfrids, A., & Pöntinen, S. (2025) Some problems in the ethical impact assessment of emerging technologies and socio-technical visions: case CityVerse. In: Alvarez, I., Arias-Oliva, M., Dediu, A.H., Silva, N. (eds) *Ethical and Social Impacts of Information and Communication Technology. ETHICOMP 2025. Lecture Notes in Computer Science*, vol 15939. Springer, Cham. https://doi.org/10.1007/978-3-032-01429-0_32
- Koskimies, E., Nieminen, M., Stenvall, J., Hallamaan, J., Leikas, J., & Salo-Pöntinen, H. (2020). ETAIROS Policy Brief: Suomeen tarvitaan koordinoitu, vastuullisuutta korostava tekoälypolitiikka. https://etairos.fi/wp-content/uploads/2021/02/etairos-politiikkamietinto_suomeen-tarvitaan-koordinoitu-vastuullisuutta-korostava-tekoalypolitiikka.pdf
- Lanne, M., Nieminen, M., & Leikas, J. (2025). Organisational tensions in introducing socially sustainable AI. *AI & Soc.* <https://doi.org/10.1007/s00146-025-02293-y>
- OECD (2021). Civic Space Scan of Finland, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/f9e971bd-en>
- OECD (2024). Principles for trustworthy AI. <https://oecd.ai/en/ai-principles>
- Poikola, A., Markkanen, J., & Parkkila, J. (2024). Käsikirja tekoälystä päätöksentekijöille – Tietopolitiikan yhteistyöryhmä <https://tietopolitiikka.fi/2024/11/14/kasikirja-tekoalysta-paatoksentekijoille/>
- Saariluoma P., & Leikas, J. (2020). Designing Ethical AI in the Shadow of Hume's Guillotine. In: Ahram T., Karwowski W., Vergnano A., Leali F., Taiar R. (eds) *Intelligent Human Systems Integration 2020. Advances in Intelligent Systems and Computing*, vol 1131, pp. 594-599. Springer, Cham https://doi.org/10.1007/978-3-030-39512-4_92
- Salo-Pöntinen, H., Saariluoma, P. (2022). Reflections on the human role in AI policy formulations: how do national AI strategies view people? *Discov Artif Intell* 2, 3 (2022). <https://doi.org/10.1007/s44163-022-00019-3>
- Sigfrids, A., Latvanen, M., & Hantula, K. (2025). Digitaalinen kansalaisosallistuminen julkisessa palvelukehittämisessä. *Kenttäkokeilun loppuraportti. VTT Technology No 437.* <https://publications.vtt.fi/pdf/technology/2025/T437.pdf>
- Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023A). Human-centricity in AI governance: A systemic approach. *Frontiers.* <https://doi.org/10.3389/frai.2023.976887>

- Sigfrids, A., Nieminen, M., Leikas, J., Karvonen, A., & Pikkuaho, P. (2023B). Julkishallinto tukemassa eettisen tekoäly-yhteiskunnan rakentamista: Katsaus tekoälyn ohjauskeinoihin ja politiikkatoimenpiteisiin. VTT Technical Research Centre of Finland. VTT Technology No. 421. <https://doi.org/10.32040/2242-122X.2023.T421>
- Sigfrids, A., Nieminen, M., Leikas, J., & Pikkuaho, P. (2022). How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI. *Front. Hum. Dyn.* 4:858108. <https://doi.org/10.3389/fhumd.2022.858108>
- Valtiovarainministeriö (2018). Eettistä tietopolitiikkaa tekoälyn aikakaudella selonteko. <https://vm.fi/tietopoliittinen-selonteko>
- Wessberg, N., & Leikas, J. (2021). Tekoälyn tulevaisuuksien eettiset kysymykset sosio-teknisissä skenaarioissa.) *Futura* 3/21, pp. 28-32.

LIITE 1 ETAIROS-julkaisut, kaikki partnerit

Publications

Year of publication	Type of publication	Authors of publication	Name of publication	Name of journal/series
2025	A1	Salo-Pöntinen, Henriikki; Saariluoma, Pertti	Constructing a Methodological Discourse for Using Emotion Recognition in AI Ethics : From Behaviorist to Cognitive Approaches	International Journal of Human-Computer Interaction
2025		Myllylä, Mari; Aromaa, Susanna; Čaić, Martina; Salo-Pöntinen, Henriikki; Tammentie, Bastian; Viljakainen, Anna	Designing Meaningful Industrial Work for Hybrid Human-Technology Environments	
2025	A3	Bhatia, Riina; Rilla, Nina	Inclusiveness in research and innovation settings	
2025	A1	Lanamäki, Arto; Viljanen, Mika; Väyrynen, Karin; Bennett Moses, Lyria	Legal Compliance and the Open Texture of Law	Journal of the association for information systems
2025	D4	Nieminen, Mika; Karvonen, Antero; Mervaala, Erkki; Pyykkönen, Sinikukka	Osallisuus, dialogi, päätöksenteko - Muuttuvan yhteiskunnan tulevaisuuskestävä ohjaus	
2025	A1	Kalliokoski, Taina	Tiedossa hyvä elämä? Tekoälyn eettiset ohjeet sosiaalieettisessä tarkastelussa	Teologinen Aikakauskirja
2025	G5	Zimina, Elizaveta	Transparent RDF Question Answering and Summarisation	Tampere University Dissertations - Tampereen yliopiston väitöskirjat
2025	D4	Rannisto, Pasi-Heikki; Pekkola, Elias;	Työyhteisöjen kriisivalmiuden ja resilienssin vahvistaminen	
2024	A3	Viljanen Mika	A Horizontal Meta-effect? Theorising Human Rights in the AI Act and the Corporate Sustainability Due Diligence Directive	
2024	A1	Saariluoma, Pertti; Myllylä, Mari; Karvonen, Antero; Luimula, Mika; Aho, Jami	A human digital twin for the M-Machine	Discover artificial intelligence
2024	A1	Wahlström, Mikael; Tammentie, Bastian; Salonen, Tuisku Tuuli; Karvonen, Antero	AI and the transformation of industrial work: Hybrid intelligence vs double-black box effect	Applied Ergonomics
2024	A1	Myllylä, Mari; Saariluoma, Pertti	Erroneous thinking on climate change	Journal of Environmental Planning and Management
2024	A3	Myllylä, Mari; Tolonen, Jonna	Formation of artistic identity : Applying themes of narrative identity development in two artists' life stories	
2024	A1	Viljanen, Mika; Salminen, Jaakko; Hurmerinta-Haapää, Anna	Functional Contracting: Re-Conceptualizing Business Contracts in the Face of New Forms of Production	Georgia journal of international and comparative law
2024		Laitinen (she/her), Essi; Sanchez Nieminen, Giovanna; Rilla, Nina; Lähteenmäki-Smith, Kaisa	Gender, justice and the green hydrogen transition: moving to inclusive policymaking	
2024	A3	Sigfrids, Anton; Karvonen, Antero; Rilla, Nina; Lanne, Marinka	Governing AI services: a state-of-the-art overview	
2024	A3	Salminen, Jaakko; Rajavuori, Mikko; Viljanen, Mika	Greenhouse gas emissions in global value chains: governance, regulation and liability	
2024	E1	Saariluoma, Pertti; Manninen, Alpo	Harhainen ajattelu, disinformaatio ja yhteiskunnan toimintakyky	Kanava
2024	A1	Leikas, Jaana; Halkoaho, Arja; Lanne, Marinka	Informed Consent in the age of smart technologies	Finnish Journal of eHealth and eWelfare

2024	A2	Saariluoma, Pertti	Kognitiivisen sodankäynnin kysymyksiä	Tiede ja ase
2024	A1	Zimina, Elizaveta; Järvelin, Kalervo; Peltonen, Jaakko; Ranta, Aarne; Stefanidis, Konstantinos; Nummenmaa, Jyrki;	Linguistic summarisation of multiple entities in RDF graphs	Applied Computing and Intelligence
2024	A4	Saariluoma, Pertti; Cañas Delgado, José J.; Myllylä, Mari	Problems of Content-Based Cognitive Ergonomics	AHFE International
2024	D4	Suominen, Arho; Pihlajamaa, Matti; Deschryvere, Matthias; Nieminen, Mika; Lehenkari, Janne; Bhatia, Riina; Lähteenmäki-Smith, Kaisa; Martins, Jorge Tiago; Rilla, Nina; Hyytinen, Kirsi; Merisalo, Maria; Valkokari, Katri; Hajikhani, Arash	Shaping the Future of Finnish Innovation: The Evolution of Finnish Innovation Policy and the Path Ahead	
2024	A4	Leikas, Jaana; Rousi, Rebekah; Vilpponen, Hannu; Saariluoma, Pertti	Super-intelligent society for the silver segment: Ethics in design	CEUR workshop proceedings
2024	D2	Karvonen, Antero; Leikas, Jaana	Tekoälyjärjestelmien vastuullinen ja eettinen suunnittelu ja käyttö	
2024	A1	Saariluoma, Pertti; Karvonen, Antero	Theory languages in designing artificial intelligence	Ai and society
2024	A3	Hallamaa, Jaana; Janhunen, Tomi; Nummenmaa, Jyrki; Nummenmaa, Timo; Saariluoma, Pertti; Zimina, Elizaveta	Toward Designing Ethically Acceptable AI Security Systems Through Agent Modeling	
2024	A4	Zimina, Elizaveta; Järvelin, Kalervo; Peltonen, Jaakko; Ranta, Aarne; Nummenmaa, Jyrki	TraQuLA: Transparent Question Answering Over RDF Through Linguistic Analysis	Lecture Notes in Computer Science
2024	D2	Rilla, Nina; Nieminen, Mika	Yhteiskunnalliset yritykset yhteiskunnan ja työn muutosagentteina	Työpoliittinen aikakauskirja
2023	A1	Friman Johanna	Abuse of the Veto and Disruption of International Peace and Security Mechanisms in the United Nations	Retfaerd: Nordisk Juridisk Tidsskrift
2023	D4	Bhatia, Riina; Rilla, Nina; Laitinen (she/her), Essi; Gomez Ferreira, Gabriela	Benchmarking report on gender equality in science, technology and innovation (STI) dialogues	
2023	D1	Karvonen, Antero; Saariluoma, Pertti	Cognitive Mimetics and Human Digital Twins : Towards Holistic AI Design	ERCIM News
2023	A3	Hallamaa, Jaana	Creating Irrelations:The #MeToo Phenomenon in the Light of Recognition Theory	Routledge Studies in Contemporary Philosophy
2023	D4	Rilla, Nina; Sigfrids, Anton; Lehtinen, Santru	D4.2 Toolbox on implementing RRI on organizational and system levels: RRI Solution Cards - Challenges and solutions for implementing institutional change in RPOs and RFO	
2023	D2	Leikas, Jaana	Digihyvinvointi kasvattaa pahoinvointia - Lisääntykö epätasa-arvo uusien digiteknologioiden myötä?	Eduskunnan rulevaisuusvaliokunnan julkaisu

2023	D4	Salonen, Tuisku-Tuuli; Rilla, Nina; Pikkuaho, Pietari	Eettinen tekoäly autonomisissa järjestelmissä	VTT Technology
2023	A4	Peltonen, Jaakko; Xu, Wen; Nummenmaa, Timo; Nummenmaa, Jyrki;	Fair Neighbor Embedding	Proceedings of Machine Learning Research
2023	E1	Friman Johanna	Folkrätt i krig - hur kan den vara till hjälp?	Fredsposten
2023	A1	Karvonen, Antero; Kujala, Tuomo; Kärkkäinen, Tommi; Saariluoma, Pertti	Fundamental Concepts of Cognitive Mimetics	Cognitive systems research
2023	A1	Myllylä, Mari; Tolonen, Jonna	Graffiti and Street Art Research : An Outsider Perspective	Nuart journal
2023	A3	Viljanen Mika	Horisontaalivaikutuksen uudet kuteet? Tekoälysäädöksen ja yritysvastuuvollisuusdirektiivin ihmisoikeus- ja perusoikeusvaikutuksien jäljillä	
2023	A3	Viljanen Mika	How to Ensure Safe Navigation: Navigation Safety Regulation in MASS	
2023	A4	Saariluoma, Pertti; Myllylä, Mari; Karvonen, Antero	Human digital twins in interaction design : from abstract to concrete	
2023	A1	Sigfrids, Anton; Leikas, Jaana; Salo-Pöntinen, Henriikki; Koskimies, Emmi;	Human-centricity in AI governance : A systemic approach	Frontiers in artificial intelligence
2023	A4	Lu, Chien; Lauritano, Giacomo; Nummenmaa, Timo; Peltonen, Jaakko	Human-Environment Relationships in Alba: A Typological Analysis of Player Engagement in Steam Reviews	Conference of Digital Games Research Association
2023	E1	Hallamaa, Jaana	Ihmistieteellinen tutkimus teknistyneessä yhteiskunnassa	
2023	B1	Haikarainen, Raine	Ilmailun turvallisuustutkiminta ja kriminalisointipaineet	Teologinen Aikakauskirja
2023	A3	Nordling, Nadja; Leponiemi, Ulriika; Kolehmainen, Jari; Rannisto, Pasi-Heikki	Innovaatioekosysteemit sosiaali- ja terveysalan kehittämisessä: Toiminnan tasot ja yhteistyön organisointi	
2023		Bhatia, Riina; Rilla, Nina; Laitinen, Essi; Ferreira, Gabriela; Milhazes, Catarina	Integration of gender in international science, technology and innovation (STI) collaboration: learning from international feminist policies	
2023	D4	Sigfrids, Anton; Nieminen, Mika; Leikas, Jaana; Karvonen, Antero; Pikkuaho, Pietari	Julkishallinto tukemassa eettisen tekoäly-yhteiskunnan rakentamista: Katsaus tekoälyn ohjauskeinoihin ja politiikkatoimenpiteisiin	VTT Technology
2023	A1	Liuski, Tiia; Aalto, Janne; Kalliokoski, Taina	Kehittyvä sotilasteknologia: Haaste sotilaan etiikalle ja sotilaspapille kouluttajana?	Teologinen aikakauskirja
2023	A1	Tedeschi Miriam, Viljanen Mika	Lost in Transduction: From Law and Code's Intra-actions to the Right to Explanation in the European Data Protection Regulations	Law and Critique
2023	D4	Rilla, Nina; Salonen, Tuisku-Tuuli	Mapping on gender equality in STI bilateral and multilateral agreements	
2023	A1	Viljanen Mika	Menikö juna jo? Tekoälyn sääntelemisen mahdollisuuksista	Lakimies
2023	D1	Haikarainen, Raine	Miksi jaettu kertomuksia tarvitaan autonomisen teknologian aikana?	Teologia.fi
2023	A3	Leponiemi, Ulriika; Nordling, Nadja; Rannisto, Pasi-Heikki; Kolehmainen, Jari	Näkökulmia sosiaali- ja terveysalan innovaatioekosysteemien johtamiseen	

2023	A1	Myllylä, Mari; Cañas Delgado, José Juan; Saariluoma, Pertti	On Conspiracy Thought Models in Thinking Climate Change	European Journal of Sustainable Development
2023		Myllylä, Mari; Cañas, José Juan; Saariluoma, Pertti	On Conspiracy Thought Models in Thinking Climate Change	
2023	B1	Kalliokoski, Taina	Philip Butler. Black Transhuman Liberation Theology: Technology and Spirituality. London: Bloomsbury 2020. [Kirja-arvostelu]	Teologinen Aikakauskirja
2023	E1	Kalliokoski, Taina	Pääkirjoitus: Teknologia - elämänmuoto, tehostuva käytäntö vai puhdas väline?	Teologia.fi
2023	A1	Viljanen Mika	Safety by simulation: theorizing the future of robot regulation	AI and Society
2023	C2	Rannisto, Pasi-Heikki; Leponiemi, Ulrika; Nordling, Nadja; Kolehmainen, Jari	Sosiaali- ja terveysalan innovaatioekosysteemit	
2023	A3	Rannisto, Pasi-Heikki; Rintamäki, Timo	Sote-alan innovaatioekosysteemiin liittymisen odotukset arvolupausten kautta tarkasteltuna	
2023		Lanne, Marinka; Palomäki, Katariina	Stakeholder engagement in developing sustainable business	
2023	A1	Salminen Jaakko, Rajavuori Mikko, Viljanen Mika, Storsjö Isabell, Halonen Kirsi-Maria, Kaave Piia, Tuominen Tomi	Strateginen resilienssi ja ylikansalliset arvoketjut: kohti uutta talousoikeutta	Lakimies
2023	A1	Kuha, Suvi; Xiong, Essi; Hammarén, Mira; Ylisirniö, Minna; Haverinen, Jari; Suominen, Jarno; Leikas, Jaana; Kanste, Outi	Teknologian hyödyntämiseen liittyvä yksityisyydensuoja, tietoturva ja -suoja ikääntyneiden kotipalveluissa: kyselytutkimus johtajille	Finnish Journal of eHealth and eWelfare
2023	D1	Kalliokoski, Taina	Tekoälymurros ja teologit	Crux
2023	B1	Salo-Pöntinen, Henrikki	Uhkakuvat voivat sumentaa tekoälykehityksen rodellisuutta	Teologinen aikakauskirja
2023	A1	Hallamaa, Jaana; Kalliokoski, Taina	Vastuun kantaminen sosioteknisen järjestelmän epäonnistumisissa	Teologinen Aikakauskirja
2023		Rauterberg, Matthias; Saariluoma, Pertti	What are the Drivers in Cultural Development	Communications in Computer and Information Science
2023	D4	Rilla, Nina; Myllyoja, Jouko; Oksanen, Juha; Naumanen, Mika; Jalonen, Kari; Lillberg, Petteri; Rautiainen, Anni; Djakonoff, Vera; Kostilainen, Harri; Kainulainen, Sakari; Klemelä, Juha; Ala-Kauhaluoma, Mika; Salmi, Eeva	Yhteiskunnan kehittäjät ja uudistajat : Yhteiskunnallisten yritysten sosiaaliset innovaatiot.	Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja
2023	B1	Kalliokoski, Taina	Älyteknologioiden kehittämisessä tarvitaan teologista ymmärrystä	Teologinen Aikakauskirja
2022	D1	Leikas, Jaana; Floréen, Patrik	A systematic exercise tool helps researchers ponder the ethical implications of AI	ERCIM News
2022	A1	Viljanen Mika, Parviainen Henni	AI Applications and Regulation: Mapping the Regulatory Strata	Frontiers in Computer Science
2022	A1	Hallamaa, Jaana; Kalliokoski, Taina	AI Ethics as Applied Ethics	Frontiers in computer science
2022	A3	Koskimies, Emmi; Stenvall, Jari; Kinder, Tony; Leikas, Jaana; Nieminen, Mika	Artificial intelligence and public innovations	

2022	D1	Hallamaa, Jaana	Conflicting values:a challenge in the design of AI systems	KANBrief
2022	B1	Mäntymäki Matti, Minkkinen Matti, Birkstedt Teemu, Viljanen Mika	Defining organizational AI governance	AI and ethics
2022	A1	Salminen Jaakko, Sobel-Read Kevin B., Viljanen Mika, Eller Klaas Hendrik	Digital Platforms as Second-Order Lead Firms: Beyond the Industrial/Digital Divide in Regulating Value Chains	European Review of Private Law
2022	A1	Malik Hanna Maria, Viljanen Mika, Lepinkäinen Nea, Alvesalo-Kuusi Anne	Dynamics of Social Harms in an Algorithmic Context	International journal for crime, justice and social democracy
2022	B1	Rousi, Rebekah; Saariluoma, Pertti; Nieminen, Mika	Editorial : Governance AI ethics	Frontiers in Computer Science
2022	A1	He, Chengxin; Duan, Lei; Dong, Guozhu; Nummenmaa, Jyrki; Wang, Tingting; Pang, Tinghai	Efficient mining of concept-hierarchy aware distinguishing sequential patterns	Knowledge-Based Systems
2022	G5	Myllylä, Mari	Embodied mind and mental contents in graffiti art experience	JYU Dissertations
2022	A4	Silvennoinen, Johanna; Saariluoma, Pertti	Emotional Information Space in Designing AI Technologies	AHFE international
2022	A1	Cañas-Bajo, Jose; Silvennoinen, Johanna; Saariluoma, Pertti	Emotional Responses of Finnish and Spanish Audiences to Culturally Loaded Films	Projections: the journal for movies and mind
2022	A1	Myllylä, Mari T.; Saariluoma, Pertti	Expertise and becoming conscious of something	New Ideas in Psychology
2022		Myllylä, Mari; Saariluoma, Pertti	False thinking and climate change	
2022		Bhatia, Riina; Rilla, Nina; Sanchez Nieminen, Giovanna; Merisalo, Maria; Salonen, Tuisku-Tuuli	From persisting gender inequality to inclusion in research and innovation content: challenging the gender equality norm in the STI fields	
2022	A1	Gianni, Robert; Lehtinen, Santtu; Nieminen, Mika	Governance of Responsible AI: From Ethical Guidelines to Cooperative Policies	Frontiers in Computer Science
2022	A2	Leikas, Jaana; Johri, Aditya; Latvanen, Marko; Wessberg, Nina; Hahto, Antti	Governing Ethical AI Transformation: A Case Study of AuroraAI	Frontiers in Artificial Intelligence
2022	A2	Sigfrids, Anton; Nieminen, Mika; Leikas, Jaana; Pikkuaho, Pietari	How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI	Frontiers in Human Dynamics
2022	E1	Myllylä, Mari; Saariluoma, Pertti	Ilmastokeskustelun vaaralliset harhakäsitykset	Kanava
2022	A4	Hallamaa, Jaana	Improving AI Systems Through Trade-Offs Between Values	Lecture Notes in Computer Science
2022	A1	Araújo, Tiago Brasileiro; Stefanidis, Kostas; Pires, Carlos Eduardo Santos; Nummenmaa, Jyrki; da Nóbrega, Thiago Pereira	Incremental Entity Blocking over Heterogeneous Streaming Data	Information (Switzerland)
2022	C2	Malik Hanna Maria, Lepinkäinen Nea, Alvesalo-Kuusi Anne, Viljanen Mika	Justice, Power and Resistance Volume 5 (2022): Issue 3 (Dec 2022): Special Issue: Algorithmic harms	Justice, Power and Resistance
2022	D1	Haikarainen, Raine	Lentoturvallisuuden toimintatavat tärkeitä malleja myös muille:Ilmailu turvallisuusajattelun edelläkävijä	Turvallisuus & riskienhallinta
2022	B1	Saariluoma, Pertti	Mental Content and Content-Based Cognitive Ergonomics	Ergonomics International Journal
2022	A4	Saariluoma, Pertti; Myllylä, Mari; Karvonen, Antero	Mental Contents in Designing AI Ethics	Lecture Notes in Computer Science

2022	B3	Sanchez Nieminen, Giovanna; Rilla, Nina; Bhatia, Riina; Merisalo, Maria; Gomes Coelho Ferreira, Gabriela	Moving away from traditional STI: How to improve gender equality and inclusivity in STI?	
2022	A4	Lu, Chien; Peltonen, Jaakko; Nummenmaa, Timo; Nummenmaa, Jyrki	Nonparametric Exponential Family Graph Embeddings for Multiple Representation Learning	Proceedings of Machine Learning Research
2022		Nieminen, Mika; Lanne, Marinka; Leikas, Jaana	Organizational tensions in the introduction of socially sustainable AI	
2022	B1	Kalliokoski, Taina	Panu Raatikainen (toim.). Tekoäly, ihminen ja yhteiskunta: filosofisia näkökulmia. Helsinki: Gaudeamus 2021. [Kirja-arvostelu]	Teologinen Aikakauskirja
2022	A4	Hallamaa, Jaana; Kalliokoski, Taina	Placing blame in multi-agent systems	Lecture Notes in Computer Science
2022	A4	Myllylä, Mari	Psychological and Cognitive Challenges in Sustainable AI Design	Lecture Notes in Computer Science
2022	A2	Salo-Pöntinen, Henriikki; Saariluoma, Pertti	Reflections on the human role in AI policy formulations : how do national AI strategies view people?	Discover Artificial Intelligence
2022	B1	Malik Hanna Maria, Lepinkäinen Nea, Alvesalo-Kuusi Anne, Viljanen Mika	Social harms in an algorithmic context	Justice, Power and Resistance
2022	B1	Friman Johanna	Sodan lait ja hiljaisuus	Lakimies
2022	A1	Myllyoja, Jouko; Rilla, Nina; Lima-Toivanen, Maria	Strengthening futures-oriented agenda for building innovation ecosystems	European Journal of Futures Research
2022	E2	Haikarainen, Raine	Taivaalla opittua: Ilmailu ja turvallisuusajattelun kehittyminen	
2022	A1	Viljanen Mika	Technology matters: how algorithm and artificial intelligent technology features affect harms reduction efforts	Justice, Power and Resistance
2022	D4	Ailisto, Heikki; Myllymäki, Petri; Tarkoma, Sasu; Kämäräinen, Joni-Kristian; Röning, Juha; Salakoski, Tapio; Solin, Arno; Saariluoma, Pertti; Mikkonen, Tommi; van Gils, Mark; Väänänen, Kaisa; Puolamäki, Kai; Ylén, Peter; Roos, Teemu; Leikas, Jaana; Honkela, Antti; Kutila, Matti; Ruotsalainen, Laura; Ylikoski, Petri; Linturi, Risto	Tekoälyratkaisut tänään ja tulevaisuudessa	Eduskunnan tulevaisuusvaliokunnan julkaisu
2022	D4	Leikas, Jaana	Tekoälystä vastuullista	
2022	B3	Lanne, Marinka; Palomäki, Katariina	Tensions in sustainable business: burdens or enablers?	LUT Scientific and Expertise Publications: Research Reports
2022	A1	Johanna Friman	The dark lawmaking of human security	Academy of European
2022	A1	Andrushchenko, Mykola; Sandberg, Kirsi; Turunen, Risto; Marjanen, Jani; Hatavara, Mari; Kurunmäki, Jussi; Nummenmaa, Timo; Hyvärinen, Matti; Teräs, Kari; Peltonen, Jaakko; Nummenmaa, Jyrki	Using parsed and annotated corpora to analyze parliamentarians' talk in Finland	Journal of the Association for Information Science and Technology
2022	A4	Saariluoma, Pertti; Möttönen, Tapani; Onikki-Rantajaaskö, Tiina	What can linguistics do to technology design?	Lecture Notes in Networks and Systems

2022	D2	Viljanen Mika, Salminen Jaakko, Rajavuori Mikko	Yksityisoikeudellinen vastuu yritysvastuuvetoiteesta	
2022	E1	Kalliokoski, Taina	Ymmärrän, kritisoin, rakastan?	Sosiaalietiikka.fi-blogi
2021	A4	Lu, Chien;Buruk, Oğuz & Oz;Hassan, Lobna;Nummenmaa, Timo;Peltonen, Jaakko	"Switch" up your exercise: An empirical analysis of online user discussion of the Ring Fit Adventure exergame	CEUR Workshop Proceedings
2021	A4	Salo-Pöntinen, Henrikki	AI Ethics : Critical Reflections on Embedding Ethical Frameworks in AI Technology	Lecture Notes in Computer Science
2021	E1	Leikas, Jaana, Vilpponen, Hannu, Saariluoma, Pertti	Boomerit digiuudistusten pomppulaudalla	Maaseudun tulevaisuus
2021	A4	Saariluoma, Pertti; Karvonen, Antero; Kujala, Tuomo	Conceptual Basis of Cognitive Mimetics for Information Engineering	Journal of Physics : Conference Series
2021	A4	Lu, Chien;Peltonen, Jaakko;Nummenmaa, Timo;Nummenmaa, Jyrki;Järvelin, Kalervo	Cross-structural Factor-topic Model: Document Analysis with Sophisticated Covariates	Proceedings of Machine Learning Research
2021	A4	Strahl, Jonathan; Peltonen, Jaakko; Floreen, Patrik	Directing and Combining Multiple Queries for Exploratory Search by Visual Interactive Intent Modeling	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
2021	A4	Strahl, Jonathan; Peltonen, Jaakko; Florén, Patrik	Directing and Combining Multiple Queries for Exploratory Search by Visual Interactive Intent Modeling	Lecture Notes in Computer Science
2021	A1	Peltonen, Jaakko	Edge-Path Bundling: A Less Ambiguous Edge Bundling Approach	IEEE Transactions on Visualization and Computer Graphics
2021	A4	Xu, Yifan;Duan, Lei;Xie, Guicai;Fu, Min;Li, Longhai;Nummenmaa, Jyrki	Efficient Mining of Outlying Sequential Behavior Patterns	Lecture Notes in Computer Science
2021	A1	Fransberg, Malin; Myllylä, Mari; Tolonen, Jonna	Embodied graffiti and street art research	Qualitative Research
2021	A4	Myllylä, Mari	Empathy in Technology Design and Graffiti	Lecture Notes in Computer Science
2021	A1	Saariluoma, Pertti	Etiikka, aivot, mieli ja tulevaisuusdiskurssi	Futura
2021	A4	Saariluoma, Pertti; Alanen, Hanna-Kaisa; Rousi, Rebekah	Fashion Technology : What Are the Limits of Emerging Technological Design Thinking? Finland	Advances in Intelligent Systems and Computing
2021	A3	Stenvall, Jari ; Kuoppakangas, Päivikki; Talonen, Antti Paavali		
2021	A4	Xiao, Shan;Duan, Lei;Xie, Guicai;Li, Renhao;Chen, Zihao;Deng, Geng;Nummenmaa, Jyrki	HMNet: Hybrid Matching Network for Few-Shot Link Prediction	Lecture Notes in Computer Science
2021	A3	Saariluoma, Pertti; Karvonen, Antero; Sorsamäki, Lotta	Human Digital Twins in Acquiring Information About Human Mental Processes for Cognitive Mimetics	Frontiers in Artificial Intelligence and Applications
2021	A4	Saariluoma, Pertti	Human Research in Technology Design	Lecture Notes in Computer Science
2021	A1	Saariluoma, Pertti	Hume's guillotine and intelligent technologies	Human-Intelligent Systems
2021	D2	Johanson, Jan-Erik;Liljeroos-Cork, Johanna;Stenvall, Jari;Pekkola, Elias;Rossi, Paula	Julkisen toiminnan luonne ja sen uudistamisen tavat	Vaasan yliopiston raportteja
2021	E1	Haikarainen, Raine	Kohti korkeaa luotettavuutta	maan suola
2021	A1	Kuparinen, Olli;Peltonen, Jaakko;Mustanoja, Liisa;Leino, Unni;	Lects in Helsinki Finnish: a probabilistic component modeling approach	LANGUAGE VARIATION AND CHANGE

2021	A4	Saariluoma, Pertti; Salo-Pöntinen, Henrikki	Lost People : How National AI-Strategies Paying Attention to Users	Advances in Intelligent Systems and Computing
2021	A1	Janhunnen, Tomi	On neighbourhood singleton-style consistencies for qualitative spatial and temporal reasoning	Information and Computation
2021	A4	Janhunnen, Tomi;	On Syntactic Forgetting Under Uniform Equivalence	Lecture Notes in Computer Science
2021	A1	Kinder, Tony;Stenvall, Jari	Public value and public services in the post-virus economy	Public Sector Economics
2021	A1	Kinder, Tony;Stenvall, Jari;Six, Frédérique;Memon, Ally	Relational leadership in collaborative governance ecosystems	Public Management Review
2021	D4	Hjelt, Jonna;Mehari, Yohannes;Mykkänen, Mikko;Pekkola, Elias;Stenvall, Jari	RRIL - Learning Programme of the course: Gender Equality	dia-e-logos Working Paper
2021	D4	Hjelt, Jonna;Mehari, Yohannes;Mykkänen, Mikko;Pekkola, Elias;Stenvall, Jari	RRIL-Presentation of the course: Ethics in Responsible & Sustainable Innovation	dia-e-logos Working Paper
2021	A1	Stratigi, Maria;Pitoura, Evangelia;Nummenmaa, Jyrki;Stefanidis, Kostas	Sequential group recommendations based on satisfaction and disagreement scores	Journal of Intelligent Information Systems
2021	A1	Janhunnen, Tomi	Solution Enumeration by Optimality in Answer Set Programming	THEORY AND PRACTICE OF LOGIC PROGRAMMING
2021	B1	Hallamaa, Jaana; Haikarainen, Raine; Kalliokoski, Taina	Tekoälyteknologiat ja turvallisuus	Tieteessä tapahtuu
2021	A1	Peltonen, Jaakko;	The Role of Interactive Visualization in Fostering Trust in AI	IEEE Computer Graphics and Applications
2021	A4	Lu, Chien;Koskinen, Elina;Leorke, Dale;Nummenmaa, Timo;Peltonen, Jaakko	The World Is Your Playground: A Bibliometric and Text Mining Analysis of Location-Based Game Research	Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering
2021	A4	Anttila, Heidi; Niemelä, Marketta; Anttila, Minna; Pekkarinen, Satu; Hallamaa, Jaana; Koskinen, Jani	Towards Ethical and Sustainable Technology-Supported Ageing at Home in Finland ♦ KATI Programme	CEUR Workshop Proceedings
2021	A4	Anttila, Heidi; Niemelä, Marketta; Anttila, Minna; Pekkarinen, Satu; Hallamaa, Jaana; Koskinen, Jani	Towards ethical and sustainable technology-supported ageing at home in Finland – KATI programme	CEUR Workshop Proceedings
2021	A4	Buruk, Oğuz 'Oz';Salminen, Mikko;Xi, Nannan;Nummenmaa, Timo;Hamari, Juho	Towards the next generation of gaming wearables	
2021	A3	Hallamaa, Jaana	What could safety research contribute to technology design?	Lecture Notes in Computer Science
2020	A4	Lehtonen, Miikka;Lu, Chien;Nummenmaa, Timo;Peltonen, Jaakko	Adoption of requirements engineering methods in game development: A literature and postmortem analysis	Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST
2020	A3	Leikas, Jaana	Aesthetic Wellbeing and Ethical Design of Technology	Human-Computer Interaction Series

2020	A4	Janhunen, Tomi; Sioutis, Michael	Allen's Interval Algebra Makes the Difference	Lecture Notes in Computer Science
2020	A1	Bomanson, Jori; Janhunen, Tomi; Niemelä, Ilkka	Applying Visible Strong Equivalence in Answer-Set Program Transformations	ACM TRANSACTIONS ON COMPUTATIONAL LOGIC
2020	A1	Bomanson, Jori; Janhunen, Tomi	Boosting Answer Set Optimization with Weighted Comparator Networks	Theory and Practice of Logic Programming
2020	A1	Laine, Pertti; Kuoppakangas, Päivikki; Stenvall, Jari	Change Strategy (CS): Is There Any?	Nordic Journal of Business
2020	D4	Ringbom, Henrik; Viljanen, Mika; Poikonen, Jussi; Ilvessalo, Saara	Charting Regulatory Frameworks for Maritime Autonomous Surface Ship Testing, Pilots, and Commercial Deployments	Report for the Finnish Ministry of Transport and Communications
2020	A1	Kuoppakangas, Päivikki; Lindfors, Juha; Stenvall, Jari; Kinder, Tony; Talonen, Antti	COVID-19 triggering homecare professionals' change of attitudes towards e-Welfare	Finnish Journal of eHealth and eWelfare
2020	A4	Vilpponen, Hannu; Leikas, Jaana; Saariluoma, Pertti	Designing digital well-being of senior citizens	
2020	A4	Saariluoma, Pertti; Leikas, Jaana	Designing Ethical AI in the Shadow of Hume's Guillotine	Advances in Intelligent Systems and Computing
2020	A4	Saariluoma, Pertti; Leikas, Jaana	Designing ethical AI in the shadow of Hume's guillotine	Advances in Intelligent Systems and Computing
2020	C2	Rousi, Rebekah; Leikas, Jaana; Saariluoma, Pertti	Emotions in Technology Design : From Experience to Ethics	Human - Computer Interaction Series
2020	C2	Rousi, Rebekah; Leikas, Jaana; Saariluoma, Pertti	Emotions in Technology Design: From experience to Ethics	Human-Computer Interaction Series
2020	A4	Chien, Lu; Peltonen, Jaakko	Enhancing Nearest Neighbor Based Entropy Estimator for High Dimensional Distributions via Bootstrapping Local Ellipsoid	Proceedings of the AAAI Conference on Artificial Intelligence
2020	A4	Saariluoma, Pertti; Leikas, Jaana	Ethics in Designing Intelligent Systems	Advances in Intelligent Systems and Computing
2020	A1	Grahn, Hilkka; Kujala, Tuomo; Silvennoinen, Johanna; Leppänen, Aino; Saariluoma, Pertti	Expert Drivers' Prospective Thinking-Aloud to Enhance Automated Driving Technologies : Investigating Uncertainty and Anticipation in Traffic	Accident Analysis and Prevention
2020	A4	Stratigi, Maria; Nummenmaa, Jyrki; Pitoura, Evangelia; Stefanidis, Konstantinos	Fair Sequential Group Recommendations	
2020	A4	Leikas, Jaana; Sigfrids, Anton; Stenvall, Jari; Nieminen, Mika	Good Life Ecosystems: Ethics and Responsibility in the Silver Market	Lecture Notes in Computer Science
2020	A1	Kinder, Tony; Six, Frédérique; Stenvall, Jari; Memon, Ally	Governance-as-legitimacy : are ecosystems replacing networks?	PUBLIC MANAGEMENT REVIEW
2020	A4	Kalliokoski, Taina; Hallamaa, Jaana	How AI Systems Challenge the Conditions of Moral Agency?	Lecture Notes in Computer Science (LNCS)
2020	A1	Tirronen, Anniina; Stenvall, Jari; Kinder, Tony; Tienhaara, Piia; Rossi, Paula	How to measure wellbeing in outcomes-based commissioning?	Lex localis: journal of local self-government
2020	A4	Saariluoma, Pertti; Cañas, Jose; Karvonen, Antero	Human Digital Twins and Cognitive Mimetic	Advances in Intelligent Systems and Computing
2020	A4	Saariluoma, Pertti	Hume's Guillotine in Designing Ethically Intelligent Technologies	Advances in Intelligent Systems and Computing
2020	A4	Saariluoma, Pertti	Hume's Guillotine Resolved	Lecture Notes in Computer Science
2020	D1	Leikas, Jaana	Ikäteknologian kehittämiseen kaivataan todellista keskustelua ihmisen hyvästä	Vanhustyö
2020	B2	Rousi, Rebekah; Leikas, Jaana; Saariluoma, Pertti	Introduction : Feelings Matter	Human - Computer Interaction Series

2020	A1	Pekkola, Elias;Leponiemi, Ulrika;Stenvall, Jari	Kuinka arvioida kansallisia kokeilukokonaisuuksia : kokemuksia ja käsitteellisiä avauksia	Focus Localis
2020	A4	Cui, Dingshan;Duan, Lei;Wang, Xinao;Nummenmaa, Jyrki;Qin, Ruiqi;Xiao, Shan	LOCATE: Locally Anomalous Behavior Change Detection in Behavior Information Sequence	Lecture Notes in Computer Science
2020	A4	Lu, Chien;Li, Xiaozhou;Nummenmaa, Timo;Zhang, Zheyang;Peltonen, Jaakko	Patches and Player Community Perceptions: Analysis of No Man's Sky Stream Reviews	Conference of Digital Games Research Association
2020	A4	Lu, Chien;Peltonen, Jaakko;Nummenmaa, Jyrki;Järvelin, Kalervo	Probabilistic Dynamic Non-negative Group Factor Model for Multi-source Text Mining	
2020	A4	Strahl, Jonathan;Peltonen, Jaakko;Mamitsuka, Hiroshi;Mamitsuka, Hiroshi;Kaski, Samuel	Scalable Probabilistic Matrix Factorization with Graph-Based Priors	Proceedings of the AAAI Conference on Artificial Intelligence
2020	A4	Strahl, Jonathan; Peltonen, Jaakko; Mamitsuka, Hiroshi; Kaski, Samuel	Scalable Probabilistic Matrix Factorization with Graph-Based Priors	Proceedings of the AAAI Conference on Artificial Intelligence
2020	A4	Araújo, Tiago Brasileiro;Stefanidis, Konstantinos;Santos Pires, Carlos Eduardo;Nummenmaa, Jyrki;Da Nóbrega, Thiago Pereira	Schema-agnostic Blocking for Streaming Data	
2020	A4	Colange, Benoît;Peltonen, Jaakko;Aupetit, Michaël;Dutykh, Denys;Lepinat, Sylvain	Steering Distortions to Preserve Classes and Neighbors in Supervised Dimensionality Reduction	Advances in Neural Information Processing Systems
2020	A3	Myllylä, Mari	The Good, the Bad and the Ugly Graffiti	Human - Computer Interaction Series
2020	D2	Hallamaa, Jaana; Leikas, Jaana; Malkavaara, Mikko; Vesterinen, Olli	Tulevaisuuden teknologiat ja tekoälyn etiikka	Karelia-ammattikorkeakoulun julkaisuja. B, Oppimateriaaleja ja kokoomateoksia
2020	A4	Karvonen, Antero; Kujala, Tuomo; Saariluoma, Pertti	Types of mimetics for the design of intelligent technologies	Advances in Intelligent Systems and Computing
2020	A3	Saariluoma, Pertti	User Psychology of Emotional Interaction : Usability, User Experience and Technology Ethics	Human - Computer Interaction Series
2020	A1	Miettinen, Janika; Mäkinen, Markus; Leikas, Jaana; Jutila, Tuula; Veko, Tiia	Vastuullinen toiminnanohjausjärjestelmän kehittäminen kotihoidossa	Finnish Journal of eHealth and eWelfare
2020	A4	Lu, Chien;Peltonen, Jaakko;Nummenmaa, Timo;Li, Xiaozhou;Zhang, Zheyang;Xiaozhou, Li	What makes a trophy hunter?: An empirical analysis of Reddit discussions	CEUR workshop proceedings
2020	A1	Christensen, Malene Vinther; Nieminen, Mika; Altenhofer, Marlene; Tangcoigne, Elise; Mejlgård, Niels; Griessler, Erich; Filacek, Adolf	What's in a name? Perceptions and promotion of responsible research and innovation practices across Europe	Science and Public Policy
2020	D4	Hallamaa, Jaana; Leikas, Jaana; Sihvo, Päivi; Malkavaara, Mikko; Vesterinen, Olli; Parviainen, Tiina; Koski, Aia	Virtuaalinen eettinen foorumi 2 : Raportti foorumin toteutuksesta ja tuloksista	Karelia-ammattikorkeakoulun julkaisuja C, Raportteja: 68

2019		Leikas, Jaana; Gotcheva, Nadezhda; Nieminen, Mika; Koivisto, Raija; Rilla, Nina; Sigfrids, Anton	Ethical AI for the Governance of the Society	
2019	A4	Nieminen, Mika; Gotcheva, Nadezhda; Leikas, Jaana; Koivisto, Raija	Ethical AI for the Governance of the Society: Challenges and Opportunities	CEUR Workshop Proceedings
2019	D4	Koivisto, Raija; Leikas, Jaana; Auvinen, Heidi; Vakkuri, Ville ; Saariluoma, Pertti; Hakkarainen, Jenni Maria; Koulu, Riikka	Tekoäly viranomaistoiminnassa - eettiset kysymykset ja yhteiskunnallinen hyväksyttävyys	Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja