Lähettäjän viite:	Tulevaisuusvaliokunta perjantai 29.10.2021 klo 10.00 / E 62/2021 vp / Lausuntopyyntö
Lausunnon pyytäjä:	Tulevaisuusvaliokunta
Vastaaja:	Jaana Leikas, Johtava tutkija, Dosentti, Foresight and data economy
Tarkastaja:	Kalle Kantola, Vice President, Foresight-driven Business Strategies

Aihe: Tekoälyn etiikka

Kiitän asiantuntijalausuntopyynnöstä koskien alla olevaa kokonaisuutta ja tekoälyn etiikkaa:

> E 62/2021 vp Valtioneuvoston selvitys: Komission tiedonanto: Tiedonanto eurooppalaisesta lähestymistavasta tekoälyyn
> E 48/2021 vp Valtioneuvoston selvitys: Komission tiedonanto: 2030 digitaalinen kompassi: eurooppalainen lähestymistapa digitaalista vuosikymmentä varten
> E 86/2021 vp Valtioneuvoston selvitys: Ennakkovaikuttaminen: Euroopan unionin datasäädös
> U 28/2021 vp Valtioneuvoston kirjelmä eduskunnalle komission ehdotuksesta Euroopan parlamentin ja neuvoston asetukseksi tekoälyn harmonisoiduksi sääntelyksi (Artificial Intelligence Act

Tekoälyn eettiset kysymykset ovat moniulotteisia

Tekoälyn odotetaan tuovan merkittäviä taloudellisia hyötyjä oletetun lisääntyvän tehokkuuden ja tuottavuuden kautta. Tekoälystä on povattu myös merkittävää tekijää hyvän (tekoäly-)yhteiskunnan ja YK:n sosiaalisten kehitystavoitteiden edistämisessä. Julkisessa puheessa tekoälyn avulla pyritään luomaan paitsi vaurautta ja kilpailukykyä, myös **oikeudenmukaista ja kestävää yhteiskuntaa**, sekä taloudellista, poliittista ja sosiaalista **tasa-arvoa**.

Tekoälyn suunnitteluun, kehittämiseen ja käyttöönottoon liittyy aina ristiriitaisia kysymyksiä, sillä tuotteiden ja palvelujen suunnittelussa on kyse ihmisen toiminnasta ja tätä toimintaa tukevan teknologian kehittämisestä. Tiedetään, että tekoälyn hyödyntämisellä on suoria ja epäsuoria vaikutuksia kansalaisten kohteluun, oikeuksiin ja mahdollisuuksiin. Erilaisia eettisiä koodistoja onkin luotu mm. turvaamaan tekoälyn käytön kohteena olevien henkilöiden asema ja oikeudet. Eettiseksi koodiksi koottuna arvot ja prinsiipit asettavat tavoitteen välttää toiminnan negatiivisia seurauksia ja tavoitella positiivisia, ja niiden tarkastelun ajatellaan lisäävän sovelluksen hyväksyttävyyttä ja soveltuvuutta käyttötarkoitukseensa. Esimerkiksi Euroopan komission riippumaton, korkean tason tekoälyä käsittelevä asiantuntijaryhmä (AI HLEG) ilmoittaa julkaisemansa Ethics Guidelines for Trustworthy AI (AI HLEG 2019) -asiakirjan eettiseksi tarkoitukseksi "varmistaa perusoikeuksien, periaatteiden ja arvojen kunnioittaminen kehitettäessä tekoälyn käyttöönotto ja käyttöä".

Keskustelun tekoälyn etiikasta ajatellaan helposti käytävän pääsääntöisesti **organisaatioiden sisällä ja organisaatioiden omista lähtökohdista käsin.** Organisaatiot näkevätkin avoimuuden ja läpinäkyvyyden tärkeinä organisaation sisällä, mutta sidosryhmäien osallistaminen kehitystyöhön koetaan haasteellisena. Erityisen vaikeaksi koetaan **kansalaisten osallistaminen** tekoälyn kehitystä koskevaan keskusteluun.

Tekoälystä toivotaan myös välinettä julkisten organisaatioiden sisälle edistämään työn tehokkuutta ja työntekijöiden luovaa asiantuntijatyötä. Edellä mainittu AI HLEGAI toteaa, että koska tekoäly voi parhaimmillaan olla ihmiskunnan kukoistusta tukeva teknologia, juuri yhteiskunnallisesta näkökulmasta tulisi ymmärtää, miten tekoälyä voidaan hyödyntää, jotta sillä edistetään ihmisen hyvää ja parannetaan ihmisten elämänlaatua.Siksi **erityisesti julkishallinnon alalla tarvitaan tutkimusta, joka tunnistaa eri yhteisöjen kulttuuriset herkkyydet ja tuo nämä mukaan tekoälykehitykseen**. Kun julkishallinnossa kehitetään tekoälyratkaisuja, otetaan samalla vahvasti kantaa siihen, millaista tulevaisuuden yhteiskuntaa halutaan rakentaa. Siksi teknologian haltuun ottamista koskevissa kysymyksissä tarvitaan selkeitä eettisiä näkemyksiä, joiden perusteella teknologian päämäärät voidaan valita.

Riskien tarkastelua yhdessä arvojen ja hyvän elämän tarkastelun kanssa

Artificial Intelligence Act -esitys lähtee liikkeelle ongelmakeskeisesti ja riskien tarkastelu on esityksen keskiössä. Tavoitteena on, että tekoäly voitaisiin nähdä teknologiana, jonka käytössä voidaan olla varmoja siitä, että ihmisille ja ympäristölle aiheutuvan vahingon riski hallitaan ja monimutkaiset arvot ja eettiset kompromissit ratkaistaan laajemman yleisen edun mukaisesti. Sääntelyn tarkoituksena on ensisijaisesti varmistaa, että tekoälyjärjestelmiä kehittävät organisaatiot ovat tietoisia siitä, mitä tekevät, ja että ne dokumentoivat valintansa ja varmistavat, että valitut ratkaisut ovat asianmukaisia.

Tekoälyjärjestelmä määritellään esityksessä suhteellisen väljästi, mutta soveltamisala on väljästä määrittelystä huolimatta suhteellisen rajattu ja sääntelykohteet vaikuttavat perustelluilta: tekoälyjärjestelmillä voi olla merkittäviäkin vaikutuksia esimerkiksi yksilöiden oikeusasemaan ja terveyteen. Tällaisissa järjestelmissä riskit ovat suuria ja sääntely nähdään tarpeelliseksi. Jos prosessivaatimuksia ei olisi, tekoälyjärjestelmiä satettaisiin kehittää organisaatioissa ilman dokumentaatiota ja laadunvarmistusta.

Soveltamisalueiden ja tarkastelutilanteidenn rajaaminen helpottaa eettisten kysymysten tarkastelua jonkin verran, mutta toisaalta on myös tiedostettava se, että **kehitystä ohjaavat arvot kytkeytyvät väistämättä laajempiin yhteiskunnallisiin suuntaviivoihin**. Siksi esimerkiksi Euroopan komission mukaan **inhimillisten arvojen** tulisi olla keskeisessä asemassa tekoälyjärjestelmien kehittämisessä, käyttöönotossa, käytössä ja valvonnassa. Käyttötapausten riskipohjaisen tarkastelun ohella eettistä keskustelua tulisi käydä sekä ihmisen hyvään ja hyvinvoinnin olemukseen liittyvistä tekijöistä että tekoälyn suhteesta yhteiskuntaan ja yhteiskunnan muutokseen. Kun asetusehdotus koskee vain korkean riskin tekoälyjärjestelmiä, esimerkiksi suurin osa internetpalveluista nähdään matalan riskin järjestelminä, jolloin ne jäävät asetuksessa ehdotetun sääntelyn ulkopuolelle. Ei kuitenkaan ole mahdotonta, että osa esimerkiksi sosiaalisen median sisällönvalinta-algoritmeista tuottaisi käyttäjilleen psykologista tai mielenterveydellistä haittaa.

Tästä näkökulmasta keskustelussa tekoälyn etiikasta tulisi painottaa sosiaalisiin konteksteihin ja ihmisten käyttäytymiseen liittyviä kysymyksiä: **kansalaisten turvallisuuden** *ja* **yhteiskunnan luottamuksen säilymistä, palvelujen tasa-arvoista saatavuutta**, **kuulluksi tulemisen mahdollisuutta** *sekä* **ratkaisujen läpinäkyvyyttä** *ja* **oikeutusta ihmisarvon, hyvinvoinnin ja kestävän kehityksen näkökulmasta**. Näiden kysymysten operationalisointi ja istuttaminen tekoälyn konteksteihin vaatii analyyttisyyttä ja uusia työkaluja.

Lainsäädännön rinnalle tarvitaan pehmeää ohjausta

Lainsäädännön ja ns. ”kovan” sääntelyn rinnalle tarvitaan

- erilaisia käytännön ajattelutyökaluja ja käsitteellisiä kehikkoja, jotka auttavat yritys- ja julkisia toimijoita ajattelemaan moniulotteisia vastuullisuuteen ja etiikkaan liittyviä kysymyksiä,
- eettistä johtamista ja innovatiivisia työkaluja ja kehitysalustoja tulevaisuuden tarpeiden ja tekoälyn vaikutusten ymmärtämiseen
- kansalaisyhteiskunnan integroimista aidosti tekoälykehitykseen.

Niin viime aikoina esitetty tekoälyä käsittelevä kirjallisuus kuin esimerkiksi OECD (the Council on Open Government) patistavat panostamaan **avoimuuteen ja sidosryhmien osallistamiseen tekoälykeskustelussa**. Sidosryhmien osallistamisen päätöksentekoon nähdään lisäävän hallinnon vastuuvelvollisuutta, laajentavan kansalaisten vaikutusmahdollisuuksia ja vaikutusvaltaa päätöksenteossa, kehittävän kansalaisvalmiuksia, parantavan päätöksenteon tietopohjaa, vähentävän täytäntöönpanokustannuksia ja hyödyntävän laajempia verkostoja innovaatiotoiminnassa ja poliittisessa päätöksenteossa (OECD 2017 & 2020; myös esim. Arana-Catania 2021). OECD:n luottamusarvioinnin mukaan (2021) Suomessa on korkea luottamus julkiseen hallintoon, mutta kokemus vaikutusmahdollisuuksista ja päätöksenteon ymmärtämisestä on suhteellisen matala. Tarvitaan siis lisää avoimuutta sekä osallistavampaa valmistelua ja päätöksentekoa poliittisen syklin eri vaiheissa.

Tämä edellyttää oikeanlaista organisaatioiden eettistä agendaa. Organisaatioilta edellytetään siksi **sitoutumista** tiettyjen (Eurooppalaisten) arvojen noudattamiseen, **avoimuutta** ratkaisujen valinnoissa ja ratkaisujen **selittämistä**. Ennen kaikkea edellytetään uskottavuutta, hyvin perusteltuja, realistisia ja toteuttamiskelpoisia näkemyksiä. Tätä ymmärrystä tukemaan tarvitaan **lisää empiiristä ja selittävää tutkimusta tekoälyn käytöstä, läpinäkyvyyttä tutkimustietojen jakamisessa, sekä tietoa parhaista käytännöistä käytön riskien hallitsemiseksi.**

Eettisyyttä tukevat käytännön toimet

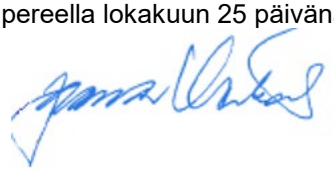
Euroopan komission mukaan tekoälyn käyttöön liittyvää **vastuuvelvollisuutta** voidaan edistää usean eri tekijän, kuten teollisuusstandardien kehittämisen, sertifikaattien, auditointimenetelmien ja vaikutusten arvioinnin avulla. Organisaatioidenn sisäisten ja ulkoisten tarkastajien avulla voidaan arvioida algoritmeja, tietoja ja suunnitteluprosesseja. Tekoälyjärjestelmän kielteiset vaikutukset voidaan tunnistaa ja niistä voidaan ilmoittaa. Ilmoittamisen tulisi olla turvallista ja siihen tulisi reagoida. Lisäksi tulisi huolehtia, että käytössä on helppokäyttöisiä mekanismeja asianmukaista muutoksenhakua varten.

Myös erilaiset tekniset menetelmät voivat edistää vastuuvelvollisuuden toteutumista. Näitä ovat esimerkiksi järjestelmiin sisäänrakennettu yksityisyydensuoja ja turvallisuus. Järjestelmän toiminnan testaus tekoälyjärjestelmän opettamisen ja käyttöönoton aikana ennen todellista käyttötilannetta edesauttavat osaltaan vastuunkantoa. Asiaa voivat varmistaa myös kirjalliset käytännesäännöt. Lisäksi innovaatioiden kehittämisen mahdollistamiseksi voidaan hyödyntää niin sanottuja ”hiekkalaatikkoja” eli innovaatioalustoja, joissa on mahdollista kehittää ja kokeilla innovaatioiden toimintaa lainsäädännöstä vapautettujen tai muutettujen tilojen avulla. (emt.)

Myös **suunnittelun ja suunnitteluosaamisen kehittäminen** on oleellista. Käyttäjien osallistuminen kehitystyöhön on tärkeää. Organisaatio voi järjestää sidosryhmäpaneeleja toimintatapojen arvioimiseksi, sekä osallistaa sukupuolen, kulttuuritaustan tai osaamisen suhteen monimuotoisia ryhmiä keskusteluun.

Kuntasektorilla koetaan ETAIROS-tutkimushankkeen (www.ETAIROS.fi) tutkimusten mukaan tekoälyn käyttöönottoa hidastaviksi tekijöiksi 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. Kunnissa ei juurikaan ole järjestetty koulutusta tekoälyjärjestelmien hyödyntämiseen, ja osaamistarpeen ja resurssien suuntaamisen välillä koetaan olevan ristiriitaa. Edelläkävijäorganisaatioiksi mielletään ne harvat, jotka ovat panostaneet koulutukseen.

Tampereella lokakuun 25 päivänä 2021



Jaana Leikas, Johtava tutkija, Dosentti
Teknologian tutkimuskeskus VTT Oy

Viitteet

1. AI HLEG (2019). Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence, European Commission, April 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
2. Arana-Catania, M., Lier, F. A. V., Procter, R., Tkachenko, N., He, Y., Zubiaga, A., & Liakata, M. (2021). Citizen Participation and Machine Learning for a Better Democracy. Digital Government: Research and Practice, 2(3), 1-22.
3. OECD (2017). Recommendation of the Council on Open Government. Paris: OECD.
4. OECD (2020), Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave, OECDPublishing, Paris, https://doi.org/10.1787/339306da-en
5. OECD (2020), Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave, OECDPublishing, Paris, https://doi.org/10.1787/339306da-en Tekoälyohjelman ohjausryhmä ja sihteeristö. 2019. Edelläkävijänä tekoälyaikaan - Tekoälyohjelman loppuraportti. Työ- ja elinkeinoministeriön julkaisuja 2019:23.
6. OECD (2021), Drivers of Trust in Public Institutions in Finland, OECD Publishing, Paris, https://doi.org/10.1787/52600c9e-en.