Asiantuntijalausunto: Tekoälyn eettiset ohjeistukset

Anna-Mari Rusanen
Erityisasiantuntija
JulkICT, Valtionvarainministeriö
anna-mari.rusanen@vm.fi
p. 02955 30360

FT, Yliopistonlehtori
Vastuullinen tutkija
Kognitiotiede, Digitaalisten ihmistieteiden osasto, Helsingin Yliopisto

1. Miksi tekoälyn eettisiä ohjeistuksia tarvitaan?

Perinteistä automaatiota, ja siihen perustuvia sovelluksia, on käytetty jo pitkään niin viranomaistoiminnassa kuin tutkimuksessa, teollisuudessa ja tuotekehityksessä. Se perustuu tietoteknisesti mekaaniseen laskentaan, johon ei liity harkintaa. Uudet algoritmiset, usein kone-oppimiseen ja syväoppiviin verkkoihin perustuvat ns. älykkäät autonomiset menetelmät ovat sekä laskennallisesti että tiedonkäsittelymenetelmiltään monimutkaisempia ja sopeutuvampia. Siten ne voivat suorittaa jonkinasteista harkintaa sisältäviä toimintoja.

Tällä voi olla syvällisiä seurauksia. Nykyteknologia ei vain nopeuta tai tehosta toimintoja, vaan se muovaa niiden taustalla olevia tiedonkäsittelyn käytäntöjä ja rakenteita. Esimerkiksi julkishallinnon osalta perinteinen virastoissa asioiminen voi muuttua fyysisen kaavakkeen täyttämisestä chatbotin kanssa asioimiseksi, virkailijan tapaaminen etäyhteydellä kommunikoimiseksi, ja tarvittavien asiakirjojen käsittely osittain tekoälysovelluksilla toteutettavaksi yksilöä koskevan datan räätälöinniksi. Siten älykkäiden teknologioiden vaikutuksesta toimintoja uudelleenorganisoidaan tavoilla, joiden seurauksia on harkittava tarkoin laajempien yhteiskunnallisten vaikutusten näkökulmasta.

Näitä uusien informaatioteknologioiden vaikutuksia voidaan arvioida ainakin kolmesta eettisestä näkökulmasta (Dignum 2018, Koivisto & al 2018):

- *Eettisten tekijöiden suunnittelu (Ethics by design):* Luonteeltaan eettisen päättelyn tai päätöksenteon kyvykkyyden toteuttaminen osaksi tekoälyjärjestelmää hyödyntämällä teknisiä ja algoritmisia ratkaisuja.
- *Suunnitteluetiikka (Ethics in design):*
  - Autonomisten ja älykkäiden järjestelmien suunnittelu ja sen taustalla olevat yleiset eettiset periaatteet
  - Tekoälyjärjestelmien eettisten vaikutusten suunnittelu ja arviointi, sekä tähän tarvittavien taitojen ja menetelmien kehittäminen järjestelmien sulautuessa osaksi yhteiskunnan rakenteita

• *Ammattietiikka (Ethics for design(ers):*
o Tekoälyn kehittäjien ja käyttäjien eettisen ja moraalisen toiminnan tukeminen järjestelmien tutkimuksen, suunnittelun, toteutuksen, käyttöönoton ja hallinnan näkökulmista.
o Esimerkiksi toimintamallien kehittäminen, codes of conduct, standardit ja sertifiointi.

2. Eettisten ohjeistusten lähestymistavat

Tekoälyn eettisissä ohjeistuksissa on kaksi pääasiallista lähestymistapaa: *periaatteellinen* ja *metodinen. Periaatteellisissa lähestymistavoissa* etiikkaa lähestytään ohjaavien periaatteiden ja periaatekokoelmien kautta. *Metodisissa lähestymistavoissa* eettinen näkökulma nousee suunnittelun tarkastelusta vaiheittain etenevien menetelmien kautta (Koivisto & al 2018). Näitä ovat esimerkiksi Value-Sensitive Design (VSD) (Friedman et al. 2006) ja Care-Centered Value Sensitive Design (CCVSD) (Wynsberghe 2013).

Viime vuosina julkista keskustelua on hallinnut periaatteellinen lähestymistapa. Tällä hetkellä löytyy jo yli 70 erilaista eettistä ohjeistusta[1], mukaan lukien EU:n asiantuntijoiden (”The Guidelines to promote Trustworthy AI”), yleishyödyllisten järjestöjen, kuten IEEN:n, erilaisten yritysten (Google, Microsoft jne.) sekä akateemisten yhteisöjen suositukset.

Suositukset sisältävät edelleen lähinnä periaatteita, joiden toivotaan muuttuvan käytäntöjä ohjaaviksi normeiksi. Kuitenkaan tiettävästi missään valtiossa (tai Euroopan unionissa) eettisille ohjeistuksille ei ole toistaiseksi annettu vahvempaa regulatiivista roolia.

Vaikka eri suosituksissa on jonkin verran painotuseroja, niissä toistuvat yleensä tietyt teemat. Ensiksi, suosituksien tavoitteena on yleensä minimoida mahdolliset tekoälysovellusten eettiset tai yhteiskunnalliset haittavaikutukset tavalla, joka perustuu yleensä laajempaan ihmisoikeusajatteluun tai eettiseen viitekehykseen. Tällaisia ovat esimerkiksi pyrkimys estää syrjiminen, yhteiskunnan polarisoituminen tai epäoikeudenmukaisuus. Toiseksi, usein normistoissa EU:n suositusten tapaan pääpaino on ”luotettavan” (engl. ”trustworthy”) tekoälyn kehityksessä. Siten normeissa korostetaan mm. vastuiden, teknisen turvallisuuden ja selitettävyyden tapaisia näkökulmia.

Sinänsä valtaosa ohjeistuksien teemoista eivät itsessään ole tekoäly- tai teknologiaspesifejä, vaan pätevät mihin tahansa eettiseen arviointiin. Varsinaiset tekoälykehitykselle erityiset aihepiirit nousevat lähinnä ns. vastuullisuuden, läpinäkyvyyden, selitettävyyden, jäljittettävyyden ja ymmärrettävyyden normistoista (Morley & Floridi 2019). Niihin myös liittyy eniten käsitteellisiä epäselvyyksiä. Ei esimerkiksi ole täysin selvää, mitä ”läpinäkyvyyden” tai ”selitettävyyden” vaatimus tiukasti ottaen tarkoittaa (Rusanen 2019, vrt. Lipton 2016), millaisilla kriteeristöillä sitä tulisi arvioida ja millaisia vaikutuksia sillä on esim. nykyiseen lainsäädäntöön.

Tutkijat ovat myös nostaneet esiin sen, että läpinäkyvyyden tai selitettävyyden kriteeristöjä tulisi arvioida suhteessa toimintoihin ja toimijoiden roolituksiin (Kelmer & Kolkman 2018). Se, millaista läpinäkyvyyttä vaaditaan jonkin sovelluksen kehittäjän menetelmiltä, poikkeaa läpinäkyvyydestä, jota vaaditaan vaikkapa sovellusta käyttävältä viranomaiselta.

On myös huomattava, että usein tekoälysovelluksiin liittyvät eettiset ongelmat nousevat ensisijaisesti tekoälysovellusten vaikutuksista käytäntöihin ja toimintoihin, eivät niinkään itse ohjeistuksista. Siksi

[1] mm. useiden eri valtioiden viranomaiset tai julkishallinnollisten asiantuntijat (Montreal Declaration, Lords Select Committee, European Commission’s High-Level Expert Group), akateemiset yhteisöt (Future of Life Institute, IEEE, AI4People etc.) ja yritykset ovat julkaisseet omat eettiset suosituksensa.

eettisen tarkastelun painopiste on siirtymässä hiljalleen periaatteellisista lähestymistavoista *metodisiin, ja yleisistä abstrakteista periaatteista kohti konkreettisia, käytännöistä ja tapausesimerkeistä nouseviin eettisten haasteisiin.* Tästä näkökulmasta tekoälyn eettisessä arvioinnissa ei ole kyse vain puhtaasta filosofisesta analyysista, vaan ihmisen ja koneen tiedollisesta ja tiedonkäsittelyllisestä yhteistoiminnasta ja siitä seuraavista toiminnan yhteensovittamisen haasteista. Siksi tekoälyn eettisissä ohjeistuksissa on koneen ohella väistämättä huomioitava myös ihmisen ja koneen erilaisuuden, sekä laajemman ympäröivän sosiaalisen todellisuuden vaikutus.

3. Valtionvarainministeriö ja eettisten ohjeistuksien valmistelu

Valtionvarainministeriön JulkICT- osastolla on kesäkuussa 2019 alettu valmistella metodipohjaisesta näkökulmasta soveltuvien eettisten ohjeistusten valmistelua. Lisäksi VM:ssä selvitetään syksyllä 2019 eettisen ohjeistuksen suhdetta lainsäädännöllisiin tarpeisiin. Tässä valmistelutyössä keskeistä on tekoälyn eettisyyteen liittyvien kysymysten monitasoisuuden huomioiminen.

Valmistelu on jatkoa viime hallituskaudella tietopolitiikan ja tekoälystrategian työlle. Sen aikana valmistui *Tietopoliittinen selonteko*, jossa nostetaan esiin tekoälyn ja tiedon käytön eettiset kysymykset. Vastauksessaan selontekoon eduskunta edellytti, että valtioneuvosto jatkaa yhteistyössä eri hallinnonalojen ja asiantuntijaorganisaatioiden kanssa tietopolitiikan ja tekoälyn eettisen näkökulman käsittelemistä selonteossa esitettyjen periaatteiden konkretisoimiseksi.

Tiedon ja tekoälyn käyttöön liittyvien eettisten kysymysten käsittelyssä ja siihen liittyen ohjeistuksen valmistelussa kyse on sektorirajat ylittävästä tietopoliittisesta toimenpiteestä. Tietopolitiikan toteuttaminen ja sen koordinointi on käynnistetty valtiovarainministeriössä, ja sen tehtävänä on luoda poikkihallinnolliselle kokonaisuudelle hedelmällinen ja osallistava yhteistyö. Siten eettinen ohjeistus valmistellaan osana tietopolitiikan toteutusta ja rakenteita.

Antti Rinteen hallitusohjelmassa on kirjaus, että Suomeen luodaan ohjeistus tekoälyn eettisestä käytöstä (Liikenneverkon kehittäminen s. 112). Koska hallitusohjelmassa ei ole perusteltu tai kirjoitettu auki tarkemmin, mitä eettisellä ohjeistuksella sisällöllisesti tai rakenteellisesti tarkoitetaan ja tavoitellaan, tätä tullaan tarkentamaan työn edetessä. Niihin voivat sisältyä mm. hallituksen linjaukset, toimenpide-ehdotukset ja sektorikohtaiset ohjeet. Lisäksi tulee arvioida ohjeistuksen suhde muihin toimiin kuten lainsäädäntöehdotuksiin.

Ohjeistusten valmistelu toteutetaan laaja-alaisessa verkostoyhteistyössä poikkihallinnollisesti sekä yhteiskunnan eri toimijoiden, kuten yliopistojen ja asiantuntijaorganisaatioiden (mm. FCAI), kansalaisjärjestöjen ja elinkeinoelämän edustajien kanssa. Työssä jatketaan tietopoliittisen selonteon valmistelun laajalti osallistavaa toimintatapaa. Lisäksi valmistelussa huomioidaan niin pohjoismaisella, EU-tasolla kuin laajemmalla kansainvälisellä tasolla käytävä keskustelu tekoälyn, ja laajemmin uusien teknologioiden, eettisestä arvioinnista ja ohjeistusten painopisteistä.

Ohjeistuksen tavoitteena on luoda laaja-alainen yhteiskunnallinen ymmärrys ja yhteisesti muodostettu näkemys tekoälyn käytön, nousevien teknologioiden ja tiedon käytön eettisistä linjoista. Rinteen hallitusohjelman vahvasta ilmastotavoitteesta johtuen työskentelyssä korostuu ekologisen, kestävän kehityksen periaatteet muiden vastuulliseen tekoälyyn liittyvien eettisten suositusten painopisteinä. Lisäksi eettinen ohjeistus sisältää tekoälyn lisäksi myös muita laajempia tiedon hyödyntämiseen liittyviä eettisiä kysymyksiä, ja sen valmistelussa otetaan huomioon jo tehty työ[2].

[2] Tekoälystrategia, tietopoliittinen selonteko, VN:n selvitykset, EU-komission valmistelu, eettiset ohjeistukset jne.