

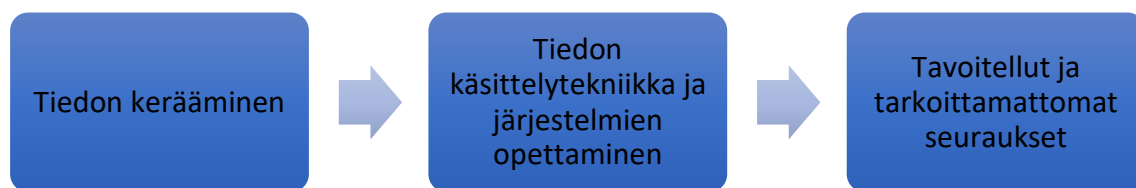
## Tekoälyn eettisten kysymysten jäsentelyä

Antti Kauppinen, Helsingin yliopisto

5.6.2018

Kuten tietopoliittisen selonteon valmistelussa on käynyt ilmi, tekoälyjärjestelmien kehittämiseen ja käyttöön liittyy hyvin monenlaisia eettisiä kysymyksiä. Olisi kenties hyödyllistä kokonaiskuvan muodostamisen kannalta, jos näitä kysymyksiä voisi jotenkin johdonmukaisesti luokitella ja ryhmitellä. Esitän tässä yhden mahdollisen mallin, joka perustuu teknologisen prosessin vaiheisiin.

Kun jonkun ongelman ratkaisuun käytetään tekoälyteknologiaa, voimme karkeasti erottaa prosessissa seuraavat vaiheet:



Ehdotukseni on, että voimme mielekkäästi jakaa suurimman osan tekoälyyn liittyvistä eettisistä kysymyksistä liittämällä ne näihin vaiheisiin. Toki on myönnettävä, ettei niiden välillä ole näin selkeitä eroja – esimerkiksi järjestelmän käyttö vaatii usein jatkuvasti uutta dataa, ja havaitut seuraukset vaikuttavat käyttötapoihin. Mutta näin pääsee ainakin alkuun.

### *1. Datan kerääminen ja järjestelmän opettaminen: Kuinka tekoälyjärjestelmiä tulisi kehittää?*

Tämän ensimmäisen pääkysymyksen alakysymyksiä ovat mm. seuraavat:

- Onko asioita, joista ei pidä kerätä informaatiota tekoälyn opettamiseksi tai soveltamiseksi? (Tässä rajoitusten perusteena voisivat olla yksityisyys tai vallankäytön mahdollisuuksien rajoittaminen)
- Millainen auktoriteetti kansalaisilla tulee olla itseensä liittyvän tiedon suhteen? (Perustelut liittyvät jälleen yksityisyyteen ja vapausoikeuksiin)
- Millaisen datan tulisi olla kaikille avointa?
- Kuinka välttää vinoumat, jotka juontavat ongelmallisesta datasta?
- Missä määrin algoritmien tulee olla läpinäkyviä tuleville käyttäjille?
- Onko tekoälysovelluksia, joita on väärin edes yrittää kehittää? (Tämä on ennen kaikkea kysymys ns. superälystä.)
- Kuinka järjestelmien kehittäjiä tulisi kouluttaa eettiseen toimintaan?

### *2. Tekoälyjärjestelmän käyttö: Miten ja mihin tarkoituksiin tekoälyjärjestelmiä tulisi käyttää?*

Toisen pääkysymyksen alaan kuuluvat mm. seuraavat alakysymykset:

- Mitkä asiat tulee jättää ihmisten päätettäväksi? (Tässä perusteena voivat olla esimerkiksi kansalaisten oikeus vaatia oikeutusta joillekin heitä koskeville päätöksille tai tekoälyjärjestelmien tekniset rajoitukset esim. liikenteessä tai sairaanhoidossa)
- Kenen tulee olla vastuussa tekoälyjärjestelmän toiminnasta ja mahdollisista virheistä?

- Kuka saa päättää, mihin tekoälyä käytetään tai ei käytetä? (Kuuluvatko jotkut asiat esim. demokraattisen kontrollin piiriin?)
- Pitääkö tekoälyjärjestelmien käyttö joihinkin tarkoituksiin kieltää kansainvälisillä sopimuksilla tai kansallisella lainsäädännöllä? (Tässä nousevat kysymykset autonomisista asejärjestelmistä tai demokraattisten prosessien manipuloinnista à la Cambridge Analytica)
- Mitkä käyttökohteet julkisen vallan tulisi priorisoida? (Vrt. Ranskan AI-raportin painopistealueet)

### *3. Tavoitellut ja tarkoittamattomat seuraukset: Kuinka tekoälyn käyttöönoton aiheuttamia muutoksia tulisi arvioida ja kuinka niihin pitäisi reagoida?*

Kolmannen pääkysymyksen alaan kuuluvat mm. seuraavat kysymykset:

- Millä kriteereillä järjestelmän hyviä ja huonoja seurauksia tulisi arvioida? (Tähän kuuluvat esimerkiksi kysymykset siitä, milloin on sopiva käyttää pelkästään taloudellisia mittareita, ja milloin ja minkälaisia hyvin- ja pahoinvointiin liittyviä mittareita tarvitaan.)
- Kuinka kansalaisten tasavertaisuus varmistetaan, mikäli tekoälyn laajamittainen käyttöönotto muuttaa työnteon luonnetta ja jakautumista? Kuinka ylipäänsä hyötyjen ja riskien tulisi jakautua? (Tähän liittyvät isot empiiriset kysymykset esimerkiksi verotuksen, omistusoikeuden ja koulutuksen muutostarpeista)
- Millainen paino uusien teknologian ympäristövaikutuksille tulisi antaa?
- Millä kriteereillä uuden teknologian aiheuttamia kulttuurisia muutoksia tulisi arvioida?

Kuten sanoin, kaikki tekoälyn etiikkaan liittyvät kysymykset eivät mahdu tähän jaotteluun. Ennen kaikkea sen ulkopuolella jää kysymys tekoälyjärjestelmien itsensä moraalisesta statuksesta. Jos kuitenkin on IEEE:n ja Laitisen & Kauppisen kanssa samaa mieltä siitä, ettei (nykyisen kaltaisilla) tekoälyjärjestelmillä ole oikeuksia, kysymys siitä kuinka niitä tulisi kohdella ei ole ainakaan ajankohtainen.

On hyvä muistaa, että nämä eettiset kysymykset limittyvät usein prudentiaalisten (omaa etua koskevien) kysymysten kanssa. Esimerkiksi kysymys siitä, kuinka meidän tulisi tulevaisuudessa kouluttaa ihmisiä on osin eettinen, mutta pääosin varmaankin prudentiaalinen (kansallinen etu vaatii kansainvälisen kilpailukyvyyn edistämistä).