

Asia: VM151:00/2018

Lausuntopyyntö kansallisen tekoälyohjelma AuroraAI:n kehittämis- ja toimintasuunnitelmasta vuosille 2019-2023

1. Kohti ihmiskeskeistä yhteiskuntaa

Klikkaa ja lisää otsikko avoimelle kysymykselle

-

2. AuroraAI-toiminta 2019-2023

Klikkaa ja lisää otsikko avoimelle kysymykselle

-

3. Organisaatioiden kytkeytyminen ja toimeenpanon tuki

Klikkaa ja lisää otsikko avoimelle kysymykselle

-

4. Yleiset huomiot, kommentit ja palaute

Klikkaa ja lisää otsikko avoimelle kysymykselle

Tekoälyyn liittyvien eettisten suuntaviivojen kehityksen tulee olla jatkuvaa, sillä myös eettinen ymmärryksemme kehittyy. Tekoälysovellusten kehityssuuntaa ei voida ennustaa, jolloin on tärkeää kyetä vastaamaan myös tulevien sovellusten tarpeisiin. Tästä johtuen tekoälysovelluksiin liittyvissä hankkeissa kuten AuroraAI:ssa tulee huomioida ja tuoda esiin uusin tietämys eettisistä haasteista ja kysymyksistä.

Kuten toimeenpanosuunnitelmassakin korostetaan, tulee tekoälysovelluksia kehittää yhteistyössä hallinnon, yritysten, tutkijoiden ja kansalaisten kesken. Tästä syystä kansallisissa ryhmissä, kuten AuroraAI:n Muutoksen tuki tiimissä, jotka tutkivat ja kehittävät tekoälyä tulisikin olla jäsenenä IT-etiikan asiantuntija, asiantuntijoita joilla on koulutus erityisesti tältä alalta. Suomesta löytyy tällä hetkellä tähän rooliin sopivia tutkijoita esimerkiksi Vaasan, Jyväskylän, Helsingin ja Turun

yliopistoista; toki tutkijoiden painotukset eri yliopistoissa ovat erilaisia. Sopivan henkilön tai ryhmän valitseminen tulee tehdä projektin tavoitteita silmällä pitäen, kuitenkin aina huomioiden myös yleinen etu. Pelkkä eettinen koodi, niin tärkeä kuin onkin, ei riitä takaamaan eettistä kehittämistä, ellei sitä osata soveltaa oikein. Tähän soveltamiseen IT-etiikan tutkijat ja asiantuntijat voivat tuoda merkittävän osaamispanoksen.

AuroraAI kehittämis- ja toimeenpanosuunnitelmassa on kattavasti käsitelty tietoturvaa ja datan yleistä käsittelyä – sekä kansalaisen itsensä että muiden timijoiden taholta – koskevat eettiset periaatteet, joten emme tässä lausunnossa keskity niihin, vaan tekoälysovellusten erityisesti esiin nostamiin eettisiin kysymyksiin. Erityisen ansiokasta on kiinnittää huomiota läpinäkyvyyden, vastuullisuuden ja luotettavuuden tarpeeseen – näitä korostamme myös tässä lausunnossa. Tekoälysovellusten kohdalla täytyy erityisesti pitää mielessä toiminnan läpinäkyvyys, kansalaisten oikeusturva, tekoälyn tuottaman datan pohjalta tehtyjen päätösten perusteltavuus ja oikeellisuus ja järjestelmien toiminnan seuraaminen mahdollisten vinoumien tai vääristymien estäminen järjestelmän toiminnassa. On tärkeää että tekoäly tekee oikeat päätökset oikeista syistä.

Toimeenpanosuunnitelmassa korostetaan läpinäkyvyyttä. Läpinäkyvyyden toteuttaminen tuottaa kuitenkin kehittyneissä tekoälysovelluksissa välttämättä ongelmia. Jos kehittyneitä tekoälysovelluksia (neurooverkkopohjaiset, itseoppivat tai itseään muokkaavat) käytetään, tulee niiden raportoida uskottavalla tavalla toiminnastaan ja tehtyjen ratkaisujen perusteista. Jo normaalin mobiililaitteen tai toimistosovelluksen ohjelmisto on ohjelmistokoodin tasolla mahdoton kenenkään yhden ihmisen kunnolla ymmärtää, puhumattakaan kehittyneistä tekoälysovelluksista ja niiden toiminnasta. Tällöin ohjelmistoon tulee sisällyttää tarvittavat itseraportointiominaisuudet jotta sen toimintaa pystytään seuraamaan.

On huomioitava – ja ilmoitettava – tekoälyn tyyppi, sen vahvuudet, heikkoudet ja rajoitukset, ja ennen kaikkea mihin tekoälyn käytöllä pyritään. Tekoälyksi kutsutun järjestelmän toimintaperiaatteet on avoimuuden lisäämiseksi – tai mahdollistamiseksi – ilmaistava selkeästi ja yksiselitteisesti. Näin toimimalla kansalaiset, tutkijat ja asiantuntijat pystyvät arvioimaan tekoälyn käytön mahdollisia seurauksia tapauskohtaisesti.

Tekoälysovelluksiin, kuten muihinkaan ohjelmistoihin, ei myöskään tule luottaa sokeasti. Kaikki päätökset niiden tuottaman datan pohjalta on pystyttävä tarvittaessa perustelemaan toiminnan kohteille. Tietojärjestelmätieteessä korostetaan, että kaikki toiminnanohjausjärjestelmät (jollaisia AuroraAI:nkin voidaan katsoa olevan) ovat vain apuvälineitä, eikä niiden tulisi tehdä päätöksiä itsenäisesti ellei tähän ole erityisen perusteltua syytä, ja että niiden tuottajien ja käyttäjien tulee pitää tämä mielessä. Jos päätöksiä automatisoidaan, vain kansalaisen kannalta kaikki tehdyt pyynnöt myöntävät järjestelmät voivat toimia siinä määrin itsenäisesti, että niitä vain pistokokeen omaisesti ja tilastollisesti seurataan. Kaikki päätökset jotka ovat kansalaisten kannalta edes osittain negatiivisia tulee käyttää ihmisen tekemän perustellun arvioinnin kautta (ks. esim. vakuutusyhtiöiden käyttämät tekoälysovellukset).

Etenkin oppivat tekoälysovellukset jotka perustuvat ohjattuun tai ohjaamattomaan koneoppimiseen voivat oppia myös väärin; ja usein tekevätkin niin. Tästä on yleisesti tunnettuja esimerkkejä kuten Amazonin työhakemusten käsittelyjärjestelmän virheet sen painottaessa naiseutta huonona ominaisuutena tai Microsoftin Tayn muuttuminen alle 24 tunnissa rasistiseksi ja misogyniseksi Hitlerin Saksan ihannoijaksi. Lienee monia muita järjestelmiä, joiden toiminnasta emme tiedä, jotka ovat tehneet vastaavanlaisia virheitä, jotka on joko myöhemmin korjattu tai joiden käyttö on kaikessa hiljaisuudessa lopetettu. Tästä syystä on erityisen tärkeää seurata tekoälyjärjestelmien toiminnan kehittymistä ja siitä seuraavia muutoksia sen toiminnassa. Tekoälyjärjestelmien tuottamien tulosten käytössä ei saa tuudittautua uskoon niiden oikeellisuudesta, vaan sen suhteen tulee olla valppaana silloinkin, kun järjestelmä päällisin puolin näyttäisi toimivan moitteettomasti. Itseoppivat ja itsemuokkaavat tekoälyjärjestelmät eivät ole koskaan ”valmiita” samassa mielessä kuin perinteiset tietojärjestelmät.

Yleisenä kommenttina tekoälystä tulee ottaa huomioon, kuten tiedonannossakin korostetaan, että tekoäly ei ole yksi sovellus, vaan vastaavalla tavalla monta erialaista sovellusryhmää kuin mikä tahansa tietotekniikan alan sovellusryhmä. Hakuagentti ja itseohjautuva auto, hoivarobotti ja itsenäisesti toimiva ase ja niitä ohjaavat algoritmit – vain muutaman esimerkin mainitakseni – ovat keskenään kovin erilaisia laitteita ja tekoälysovelluksia, ja kuitenkin kuuluvat sovelluksina käsitteen ”tekoäly” alle. Siksi on tärkeää, että projekteissa pyritään tarkkuuteen erilaisten käsitteiden käytössä.

Vaikka tekoälysovellukset tarjoavat eittämättä monia positiivisia muutoksia yhteiskunnassa, on niiden vaikutusten suhteen syytä myös varovaisuuteen.

Kimppa Kai
Turun yliopisto - Tietojärjestelmätiede/Future Ethics tutkimusryhmä