

# Luottamukseen perustuva tekoälykehitys ja luotettavuuden dimensiot

Usein puhutaan siitä, kuinka eräs Suomen keskeisimmistä vahvuuksista tekoälymarkkinoilla on "luottamusyhteiskunta". Tämä varmastikin on totta. Sen sijaan ongelmallista on, että lähes yhtä usein luottamuksen käsitettä ei täsmennetä, vaan se jää ns. täyte- tai filleritermiksi.

Lisäksi usein myös sivuutetaan se, että luottamus sekä vaatii että tarkoittaa eri konteksteissa eri asioita (luottamuksen käsitteen eri dimensioista, ks. Kirsimarja Blomqvistin taustapaperi luottamuksesta, osaamisen verkosto). Siksi on tärkeätä huomata, että ei ole yhtä yleistä laskukaavaa, jolla voitaisiin maksimoida tekoälyä kohtaan tunnetun *luottamuksen - ts. valmius ottaa riski ennakoimattomassa tilanteessa toisen toimijan suhteen -* määrä, tai edes itse tekoälyn - tai tekoälyä käyttävän toimijan - *luotettavuus*.

Jos mietimme tekoälytoimijoiden, tai tekoälyn, luotettavuutta, on huomattava, että luotettavuutta arvioidaan, tai tulisi arvioida, eri kriteereillä eri konteksteissa (konteksteista, ks. Talvitie-Lamberg & Vakkuri, etiikkaverkoston taustapaperi). "Luotettavuuteen" liittyvät kysymykset ovat erilaisia, jos niitä arvioidaan esimerkiksi tekoälysovelluksen kehittäjän tai sen loppukäyttäjän näkökulmasta. Lisäksi luotettavuus on vahvasti toimialakohtainen ulottuvuus; on eri asia puhua esimerkiksi sairausdiagnoosiikan luotettavuudesta tai hoidon tarpeen arvioinnista kuin suoratoistopalvelun katsomissuosituksista.

Samalla tavalla luotettavuuden arvioinnin lisäksi myös siihen liittyvä vastuu luotettavuudesta eroaa eri konteksteissa. On esimerkiksi sekä eettisesti että juridisesti aivan eri asia puhua esimerkiksi hoivarobotin loppukäyttäjän, terveydenhoitajan, vastuusta kuin hoivarobotin hankintapäätöksen tehneen viranomaisen vastuusta. On myös eri asia puhua hoivarobotin kehitysvaiheessa hyödynnetyn datan datantarjoajan vastuusta kuin algoritmien kehittäjän vastuista (mustat laatikot, datan vinoumat, algoritmien vinoumat, ks. taustapaperi kognitiivisista vinoumista).

Jotta tämä diversiteetti saataisiin näkyväksi, hahmotelimme kuvitteellisen esimerkin hoivarobotista. Hoivarobotin loppukäyttäjä on Paraisilla asuva julkisen terveydenhuollon piirissä oleva vanhus, ja robotti annostelee vanhuksen lääkityksen vanhuksen kulloisenkin terveydentilan pohjalta. Hoivarobotti on puolestaan kehitetty korkeakouluissa, ja sen lopullista teknistä toteutusta on hiottu yritys yhteistyössä.

*Taulukko 1. Hoivarobotti ja vastuun dimensiot. Kuvitteellinen esimerkki.*

Vaihe	FUNKTIO	TOIMIJA	ETIIKKA	VASTUU: kenellä	VASTUU: mistä
Kehitys:	algoritmit,	Tutkija	Tutkijan etiikka;	Tutkijat,	Luotettava

Perus- ja kehittävä tutkimus	arkkitehtuurit, tiedonkäsittelyongelmat...	Tutkimusryhmä Tutkimuslaitos Yritys/yritykset	tutkimusetiikka; alakohtaiset rajoitukset	tutkimuslaitos	tutkimus
Kehitys: Soveltava tutkimus	Implementaatio, tekninen toteutus	Tutkija Tutkimusryhmä Tutkimuslaitos Yritys/yritykset (Rahoittaja)	Tutkimusetiikka, alakohtaiset tekijät	Tutkija, tutkimuslaitos, rahoittaja?	Luotettava tutkimus
Tuotekehitys ja tuottaistaminen	Implementaatio	Yritys		Tuotekehitystiimi	Luotettava toiminta, luotettava tutkimustiedon soveltaminen
Datan tarjoaja: (esim. HUS, yritys)		Yritys, julkishallinnollinen toimija tms.	Datan käyttöön ja käsittelyyn liittyvä eettinen normisto Terveydenhuollon toimialan eettinen normisto	Datan ensisijainen kerääjä	Luotettava soveltaminen/vastuu datan ensisijaiselle tarjoajalle?
Soveltaja:	Implementaatio organisaatio-rakenteessa	Terveydenhoitopiiri, yritys esim. "Kimalainen"	datan ja tekoälyteknologioiden käyttöön liittyvä eettinen normisto Terveydenhuollon toimialan eettinen normisto	Sovelluksen käyttäjäorganisaatio	Luotettava soveltaminen/vastuu "potilaalle" ?
Loppukäyttäjä 1:	Loppukäyttö	Terveydenhoitaja X	Terveydenhuollon toimialan eettinen normisto	Sovelluksen käyttäjäorganisaatio	Luotettava käyttö, ymmärrys toiminnasta
Loppukäyttäjä 2:		Vanhus kotona			oikea käyttö

Luotettavuuskeskustelussa, samoin kuin koko etiikkakeskustelussa, on ollut vahva datapainotteisuus. Toki datan käsittelyyn ja jakamiseen liittyvät kysymykset (esim. vinoutunut/reikäinen data) ovat tärkeitä. Samoin on totta, että tämänhetkinen kehitys pohjaa paljolti suurten datamassojen hyödyntämiseen (analytiikkapohjaiset ja oppivat järjestelmät). Toisaalta jo nyt on useita kehityssuuntia (esim. logiikkapohjaiset teknologiat, tietyt kielen käsittelyn teknologiat), jotka eivät tukeudu samalla tavalla laajaan datamassaan.

Luotettavuuden näkökulmasta dataan liittyvien ongelmien lisäksi itse algoritmeihin tai arkkitehtuureihin liittyy aivan yhtä suuria ongelmia (ks. Rusanen, taustapaperi tietoverkosto). Algoritmienkin kehittäminen kuitenkin tapahtuu esim. yksittäisen developerin tasolla, yritysten ja tutkimusyhteisöjen sisällä, hyvin eritasoisissa käytännöissä. Siksi sitäkin voidaan ohjata eettisellä ohjeistuksella, samoin kuin esim. käyttäjälähtöisyyttä ohjeistetaan. Eettisyyttä ei siis tule käyttää yleisenä täytesanana, vaan purkaa mitkä ovat keskeisiä eettisiä kysymyksiä eri toimijaryhmille.