FCAI Finnish Center for Artificial Intelligence

REAL AI FOR REAL PEOPLE IN THE REAL WORLD

Petri Myllymäki
Professori (tekoäly ja koneoppiminen), Helsingin yliopisto
Varajohtaja, tekoälyn tutkimuskeskus FCAI

# Tulevaisuusvaliokunta perjantai 29.10.2021 klo 10.00 E 62/2021 vp / Lausuntopyyntö

## Mitä on tekoäly?

Vaikka tekoälystä ja sen mahdollisuuksista on kohistu jo vuosia, käsitteelle ei ole yleispätevää määritelmää. Yksittäisen ilmiön tai menetelmän sijaan kyse on suuresta joukosta erilaisia tutkimusongelmia ja niiden ratkaisemiseen kehitettäviä teknologioita.

Yksi tapa lähestyä asiaa on määritellä tekoäly **tieteenalaksi**, joka tutkii keinotekoisia järjestelmiä ja laitteita, joiden toiminta vaikuttaa edellyttävän älykkyyttä. Tässä kohtaa tärkeää on huomioida, että itse älykkyydelle ei ole selkeää määritelmää - siksi puhutaan mieluummin älykkään vaikutelman tekemisestä kuin älykkyydestä, koska emme osaa määritellä mitä älykkyys on. Ja koska tällaiset järjestelmät ovat tyypillisesti tietokoneohjattuja, voidaan vielä tarkentaa, että **kyse on älykkäältä vaikuttavan toiminnan automatisoinnista tietokoneilla**.

Vaikka emme osaakaan määritellä mitä älykkyys on, inhimilliseen älykkyyteen voidaan katsoa kuuluvan sellaisia ominaisuuksia kuten esimerkiksi päättelykyky, havainnointikyky ja ennustamiskyky. Tekoälytutkimuksen eri haarat tarkastelevat tällaisia ominaisuuksia kukin eri näkökulmasta. Kokonaisuudessaan ala on kuitenkin hyvin laaja ja dynaaminen, eikä siitä saada osa-alueita listaamalla luotua kovinkaan selkeää kuvaa. Tämän vuoksi yritykset luoda juridista sääntelyä jonkinlaiseen tekoälyn määritelmään perustuen ovat varsin heikolla pohjalla: ensinnäkin tekoälyn osa-alueiden lista ei muodosta millään tavalla järkevää tekoälyn määritelmää, vaan jokainen osa-alue pitäisi silloin erikseen täsmällisesti määritellä, mikä ei ole mahdollista, koska osa-alueillekaan ei ole olemassa selkeitä, yleispäteviä määritelmiä. Lisäksi on ymmärrettävä, että kullakin ajanhetkellä on toki olemassa suuri joukko suosittuja arkkityyppisiä tekoälymenetelmiä, joista asiantuntijoiden kesken vallitsee melko suuri yhteisymmärrys sen suhteen, että ko. menetelmät ovat tekoälyn ytimessä, mutta mitä lähemmäs ns. ”tavallisia” tietokonejärjestelmiä mennään, sitä epävarmemmiksi ja erimielisemmiksi asiantuntijat tulevat. Mikäli siis halutaan määritellä jonkinlainen raja tekoälyjärjestelmien ja muiden

tietokonejärjestelmien välille, uskoakseni kaikki asiantuntijat ovat samaa mieltä siitä, että selkeä rajanveto on mahdotonta, ja lähellä rajaa olevien järjestelmien statuksesta tulee aina vallitsemaan erimielisyys, ja lisäksi tämä epämääräinen rajavyöhyke liikkuu jatkuvasti ajan myötä.

Mikäli tällainen epämääräisyys kuulostaa kummalliselta, muistettakoon että tekoäly on siis tieteenala, eikä tieteenalojen välisiä rajoja ylipäänsäkään yritetä määritellä tarkasti, eivätkä tutkijat ylipäänsä koe tällaisia yrityksiä kovin mielekkäiksi: missä täsmälleen menee teologian, historian tai filosofian raja (nykyään tai vaikkapa 100 tai 500 vuotta sitten?), tai taloustieteen ja matematiikan, tai fysiikan ja kemian? Kysymykseen ei tarvita täsmällistä vastausta, koska kukin tieteenala/tiedeyhteisö määrittelee itsensä jatkuvasti uudelleen: jos esimerkiksi tekoälytutkija haluaa julkaista artikkelin tekoälylehdessä, vertaisarviointiin kuuluu, että alan asiantuntijat arvioivat muiden tekijöiden joukossa myös artikkelin relevanssia tekoälytutkimuksen kannalta. Prosessi on dynaaminen, koska vertaisarvioijat ottavat huomioon nimenomaan tämänhetkisen tilanteen - 30 vuotta sitten relevantti aihe ei enää välttämättä ole tieteenalan kannalta kiinnostava.

Yksi tekoälyn määritelmään liittyvä käsitepari on **vahva** ja **heikko tekoäly**. Jos ihmisälyn kaltaista tekoälyä osattaisiin luoda, sitä kutsuttaisiin vahvaksi tekoälyksi, millä tarkoitetaan sitä, että samaa tekoälyjärjestelmää pystyttäisiin soveltamaan moniin erilaisiin ongelmiin. Nykyiset tekoälyjärjestelmät eivät kuitenkaan ole ihmisen tavoin vahvasti älykkäitä, ja useimmat tekoälytutkijat eivät edes yritä ymmärtää tai jäljitellä ihmisen ajattelua vaan hyväksyvät sen, että keinotekoinen älykäs toiminta voi perustua aivan eri periaatteille kuin luonnollinen äly.

Nykyiset tekoälyt ovatkin järjestään niin sanottuja heikkoja/kapeita tekoälyjä, ja ne keskittyvät kukin yhdenlaisen ongelman ratkaisemiseen: yksi luokittelee roskapostia, toinen pelaa shakkia, kolmas kääntää auton pyöriä. Kaikki toimivat eri periaatteilla, eikä yhden osa-alueen kehittyminen välttämättä edistä muiden osa-alueiden kehitystä. On huomattava, että heikko tekoäly ei suinkaan tarkoita, että tällainen järjestelmä olisi hyödytön, päinvastoin, vaan heikko tekoälyjärjestelmä voi olla äärimmäisen hyödyllinen siinä kapeassa tehtävässä mihin se on rakennettu. Meidän on vaan hyvin vaikeaa ymmärtää sitä, että shakin maailmanmestarin voittava tekoälyohjelma on täysin hyödytön kaikissa muissa tehtävissä, koska vahvan ihmisälyn tapauksessa kyky suoriutua hyvin yhdessä vaikeassa tehtävässä viittaa kykyyn, joka yleistyy muihin samankaltaisiin tehtäviin.

Nykyiset tekoälyjärjestelmät perustuvat usein **koneoppimiseen**, joka on tärkeä ellei tärkein tekoälyn osa-alue. Termi saattaa tässäkin ohjata harhaan, koska kyseessä ei ole samanlainen oppimisprosessi kuin ihmisillä, vaan joukko menetelmiä/algoritmeja, joita tekoälyn asiantuntijat voivat käyttää säätämään tietokoneohjelman toimintaa kerättyjen esimerkkien perusteella niin, että syntyvä ohjelma kykenee suorittamaan sille annetun yksinkertaisen, selkeästi määriteltävän tehtävän. Koneoppimisen merkitys on siinä, että toisin kuin perinteisessä ohjelmoinnissa, tietokoneohjelma säätyy toimimaan samoin kuin annetuissa esimerkeissä, ilman että ihmisohjelmoijan täytyy ohjelmoida tietokoneohjelman jokainen askel käsin. Jos siis halutaan esimerkiksi tunnistaa valokuvassa esiintyvän ihmisen nimi, ohjelmoijan ei tarvitse miettiä mitkä kuvissa esiintyvät piirteet saattaisivat erottaa Sauli Niinistön Paavo Väyrysestä, vaan riittää antaa tarpeeksi suuri määrä esimerkkejä (kuvia, ja niihin

liitetty haluttu tulos, eli kuvassa esiintyvän ihmisen nimi), niin koneoppiva ohjelmisto säätää toimintaansa haluttuun suuntaan. On tärkeää huomata, että näin syntyvä tekoälyohjelma edustaa kapeaa tekoälyä: jos esimerkkimateriaali on poliitikkojen kasvokuvia, syntyvä ohjelma osaa luokitella niitä, mutta ei erotella vaikkapa kissaa koirasta, tai kertoa kuka valokuvissa hymyilee ja kuka ei, tai vaikkapa tunnistaa shakkilaudan asetelmaa kamerakuvan perusteella. Koneoppimisen merkitys piilee prosessin monistettavuudessa ja yleiskäyttöisyydessä: tekoälyn asiantuntijat osaavat rakentaa tekoälyjärjestelmiä kaikkiin sellaisiin tehtäviin, joissa sopivaa esimerkkidataa on runsaasti saatavilla. Toisaalta on hyvä muistaa, että lopputulos koneoppimisessakin on edelleen vain tietokoneohjelma, joka suorittaa koodinsa mukaisia yksinkertaisia laskentaoperaatioita annetun kapean tehtävän ratkaisemiseksi, eikä mikään yleiskäyttöinen keinotekoinen äly. Avain on siis siinä helppoudessa, jolla tällaisia tietokoneohjelmia voidaan nykyään luoda ilman että kaikkea tarvitsee ohjelmoida käsin.

Yhteenvetona voidaan siis todeta, että sanaa ”tekoäly” ei pitäisi koskaan käyttää yksikössä, ellei sanalla viitata tieteenalaan. Mediassa sanaa käytetään kuitenkin jatkuvasti subjektina tyyliin ”Tekoäly oppi löytämään kasvaimia keuhkokuvista yhtä hyvin kuin ihmislääkärit”, mikä on harhaanjohtavaa monestakin syystä. Ensinnäkin jokaisessa uutisessa on aina kyse erillisestä tekoälyjärjestelmästä, ei yhdestä ja samasta tekoälystä, joka oppisi uusia taitoja. Toiseksi on korostettava, että mikään tekoälyjärjestelmä ei itse asiassa opi itsenäisesti samaan tapaan kuin ihmiset, vaan koneoppimismenetelmät ovat työkaluja uusien tekoälyjärjestelmien luomiseen, ja osaavien tekoälyn asiantuntijoiden käsissä niitä käyttäen voidaan luoda yhä hyödyllisiä tekoälyjärjestelmiä, mutta aktiivinen toimija prosessissa viime kädessä on ihminen, ei tekoäly (itse asiassa, edellä mainitun uutisen tapauksessa parempi otsikko voisi olla ”Lääkärit onnistuivat automatisoimaan tylsän keuhkokuvien tarkastelutehtävän tekoälytutkijoiden avulla”). Kolmanneksi edellä mainitun kaltaiset uutiset luovat sellaisen vaikutelman, että tekoälyjärjestelmillä voitaisiin korvata ihminen, esimerkiksi lääkäri, mutta yksinkertaisten tehtävien automatisointi ei tähän riitä, ja lääkärien työhän on valtaosin jotain aivan muuta kuin keuhkokuvien analysointia. Koska tekoälyjärjestelmien ”älykkyys” on kuitenkin edelleen kovin kapeaa/heikkoa, paras hyöty tekoälystä saadaan siis silloin kun hyväksytään tekoälyn rajoitukset ja luodaan sellaisia järjestelmiä, jotka parhaalla mahdollisella tavalla tukevat ja avustavat ihmisiä, sen sijaan että yrittäisivät korvata heitä.

Helsingissä 27.10.2021



Petri Myllymäki