

<https://helda.helsinki.fi>

Tekoäly ja ihmiskognitio

Lappi, Mikko Otto Tapio

2018-02-15

Lappi , M O T , Rusanen , A-M & Pekkanen , J J O 2018 , ' Tekoäly ja ihmiskognitio ' ,
Tieteessä tapahtuu , Vuosikerta. 36 , Nro 1 , Sivut 42-46 . <
<http://urn.fi/URN:NBN:fi:ELE-1898768> >

<http://hdl.handle.net/10138/270785>

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

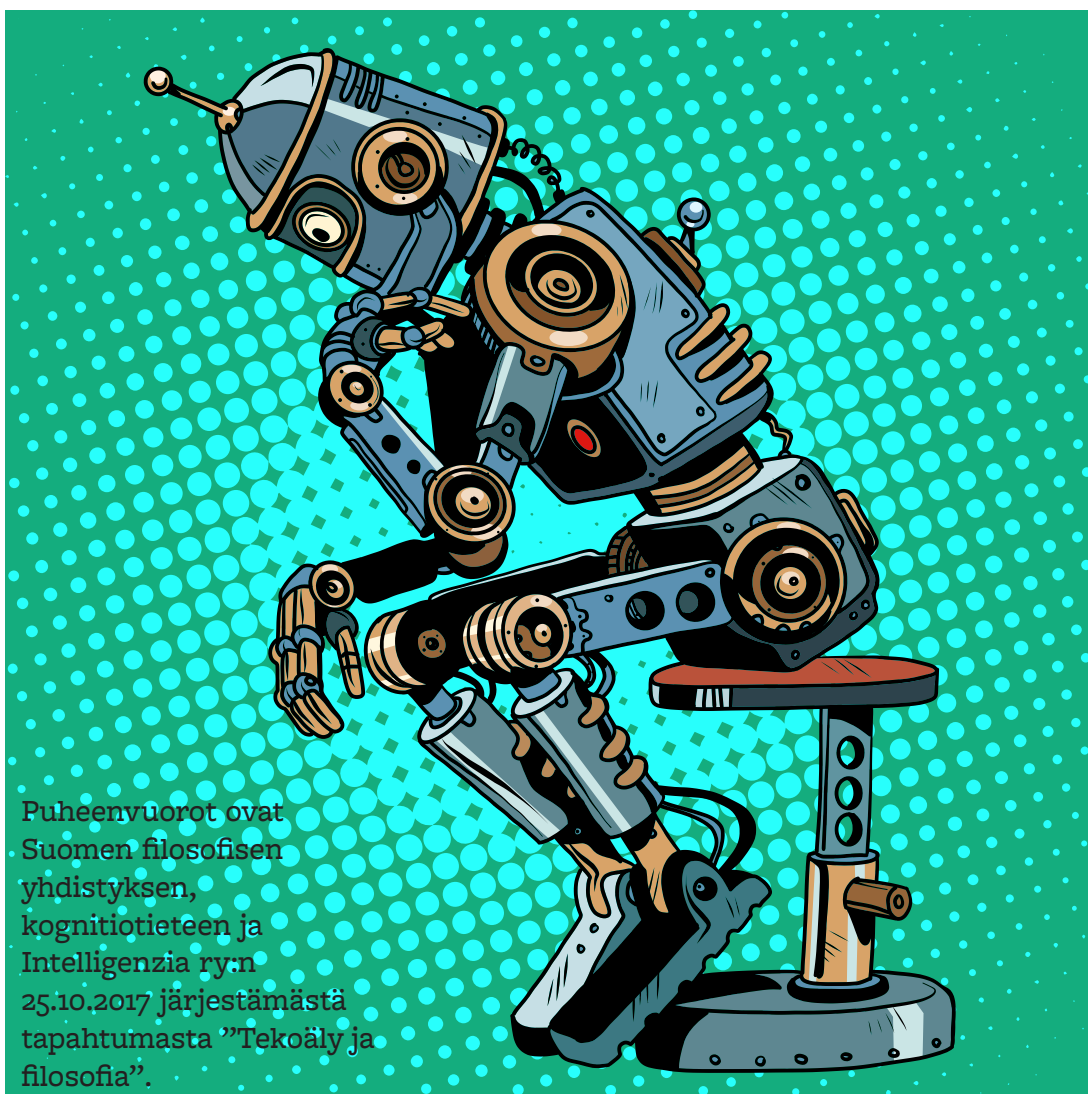
This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

TEKOÄLY

Tekoälyalgoritmien kehityksestä, tietokoneiden laskentatehon ja muistikapasiteetin kasvamisesta huolimatta – tai osittain juuri niiden takia – tekoälyyn ja sen kehittämiseen liittyy useita avoimia kysymyksiä. Osa on perustavia ja syvällisiä tieteellisiä kysymyksiä inhimillisen älyn ja koneälyn luonteesta, osa tekoälysovelluksiin liittyviä laajoja yhteiskunnallisia, eettisiä ja juridisia ongelmia.



Puheenvuorot ovat Suomen filosofisen yhdistyksen, kognitiotieteen ja Intelligenzia ry:n 25.10.2017 järjestämästä tapahtumasta ”Tekoäly ja filosofia”.

TEKOÄLY JA IHMISKOGNITIO

OTTO LAPPI, ANNA-MARI RUSANEN JA JAMI PEKKANEN

Tässä kirjoituksessa tarkastelemme tekoälytutkimusta, sen historiaa ja nykytilaa kognitiotieteen näkökulmasta. Se on sikäli ainutlaatuinen, että kognitiotiede on (ainoa) tutkimusala, joka tarkastelee ihmis- ja tekoälyjärjestelmiä täsmälleen samasta näkökulmasta, eli monimutkaisina informaatiota prosessoivina järjestelminä.

Kognitiotieteen laskennalliset menetelmät antavat mahdollisuuden tarkastella, tutkia ja analysoida älyä ja sen ilmenemistä toimintaperiaatteiltaan erilaisissa tiedonkäsittelyjärjestelmissä. Siksi sen käsitteellinen ja metodinen eksaktisuus erottaa kognitiontutkimuksen näkökulman metaforisista puhetavoista – ”aivot ovat kuin tietokone” – tai arkipsykologisista intuitioista kumpuavista pohdiskelusta tekoälyn ja ihmismielen välisestä suhteesta – ”mutta voiko kone *oikeasti* ajatella, olla tietoinen, ymmärtää kieltä jne.?”

Mitä on tekoäly?

Tekoälyalgoritmien toteutukseen käytettävissä olevien järjestelmien laskentateho ja muistikapasiteetti ovat kasvaneet räjähdysmäisesti viimeisten 50 vuoden aikana. Monia älykkyyttä vaativia tehtäviä, voidaan nyt toteuttaa – joskus jopa ihmistä paremmin – koneellisesti. Tyypillisiä esimerkkejä ovat esimerkiksi lautapelien pelaaminen, liikennereittien suunnittelu tai vaikkapa halutun tiedon hakeminen internetistä.¹

1 Joskus sanalla ”tekoäly” viitataan näitä älykkäitä järjestelmiä tuottavaan insinöörیتieteiden osa-alueeseen, jolloin tuotettujen järjestelmien kohdalla puhutaan ”keinoälystä”. Tässä kirjoituksessa viittaamme molempiin sanalla ”tekoäly”. Kun haluamme korostaa tekoälytutkimuksen tuottamaa teknologiaa puhumme koneälystä.

Tällainen luonnehdinta jättää kuitenkin avoimeksi sen, millaista älykkyyttä tekoälyn ”älykkyydellä” tarkoitetaan. Usein esimerkiksi kokeellisessa psykologiassa älykkyydellä tarkoitetaan yleisiä kognitiivisia valmiuksia. Niitä mitataan standardoiduilla päättelytehtävillä². Tekoälyn yhteydessä älykkyydellä ei kuitenkaan viitata tällaiseen niin sanottuun testiälykkyyteen. Tekoälytutkimuksen tavoitteena ei nimittäin ole kehittää koneälyjärjestelmiä, jotka saisivat mahdollisimman korkeita pisteitä nykyisissä psykometrisissä älykkyyesteissä, vaan kehittää järjestelmiä, jotka suoriutuvat älykkyyttä vaativissa tehtävissä.

Siksi tekoälytutkimuksen kohteena oleva älykkyys on olennaisesti laajempi käsite kuin pelkkä testiälykkyys. Se kattaa myös älykkään hahmونتunnistuksen tapaisia kognitiivisia kykyjä, joita ei yleensä pidetä lainkaan tyypiesimerkkeinä älykkästä toiminnasta. Esimerkiksi näköaistin toiminnan ei usein ajatella vaativan älykkyyttä. Arki-intuitio nimittäin on, että näkemisessä on kyse lähinnä siitä, että avaamme silmät ja näemme ympäristön. Vaikka tämä arki-intuitio ei ehkä heijastu kokeelliseen psykologiaan, näkö tutkimus ja älykkyytutkimukset eivät historiallisesti ole varsinaisesti kuuluneet saman tutkimusohjelman, toisin sanoen paradigman, alle. Tekoälytutkimuksen ja kognitiontutkimuksen näkökulma on tässä suhteessa toisenlainen; erityisesti tekoälytutkimuksessa *konenäkö* ja sen vaatima ”älykkyys” on ollut yksi keskeisimpiä tutkimuskohteita. Itse asiassa konenäön tutkimus on osoittanut ihmiskognition tutkijoille, kuinka näköaistin varainen toi-

2 Esimerkiksi ÄO 100 tarkoittaa, että henkilön suoritus vastaa täsmälleen normipopulaation keskiarvoa.

minta todellisuudessa perustuu hämmästyttävän monimutkaiseen ja hienostuneeseen hahmontunnistukseen, eri aistipiirien aistitiedon yhdistämiseen sekä havaintojen ja toiminnan koordinointiin älykkäällä tavalla.

Tekoälyn yhteydessä älykkyydellä, tai täsmällisemmin älykkäällä käyttäytymisellä, usein tarkoitetaankin *joustavaa ja tarkoituksenmukaista toimintaa monimutkaisessa, muuttuvassa ja osittain ennustamattomassa ympäristössä*. Tämä määritelmä korostaa älykkään toimijan (agentin) ja ympäristön vuorovaikutusta ja keskinäistä suhdetta älyn määritelmässä: älykkyys ilmenee nimenomaan siinä, miten toimija kykenee sopeuttamaan tiedonkäsittelyn varaista toimintaansa monimutkaisessa ympäristössä (šakki vaatii enemmän älykkyyttä kuin yhdeksän ruudun ristinolla), joka on osittain ennustettavissa (jolloin ennakointi ja oppiminen on ylipäätään mahdollista).

Kognitio, ”äly”, puolestaan tarkoittaa niiden monimutkaisten prosessien rakennetta ja toimintaa, joihin älykäs käyttäytyminen perustuu. Kognitiontutkimuksen tavoitteena on siis ymmärtää sitä, millaiset prosessit mahdollistavat älykkään toiminnan: Se tutkii luonnonilmiöissä esiintyvää älyä tutkimalla älyn ilmenemistä esimerkiksi ihmisten ja eläinten käyttäytymisessä sekä selvittämällä käyttäytymisen taustalla olevaa kognitiota. Historiallisesta näkökulmasta tekoäly ja ihmisen kognitiotieteellinen tutkimus ovatkin kulkeneet käsi kädessä. Tekoäly kehittyi osin yrityksestä ymmärtää ihmisen ajattelua, ja kognitiotiede puolestaan laskennallisen näkökulman soveltamisesta ihmisälyn tutkimiseen.

Tekoälyn historia

Tekoälyn tutkimus käynnistyi 1950-luvulla, kun joukko John McCarthyn, Marvin Minskyn, Claude Shannonin, Alan Newellin ja Herbert Simonin kaltaisia kognitiontutkimuksen, informaatioteorian, psykologian ja käyttäytymistieteen pioneereja koontui Dartmouthin yliopiston kuuluisaan kesäseminaariin. Heitä yhdisti ajatus siitä, että ihmisen tiedonkäsittelyä voitaisiin kuvata laskennallisesti.

Tuolloin laskenta ymmärrettiin konkreettisten symbolien manipulaationa, kuten laskutoimituksia tehtäessä. Ajatus oli syntynyt 1930-luvulla, kun

laskennan teorian kehityksen myötä kehkeytyi oivallus siitä, että laskutoimitusten tekemiseen tarvittavat säännöt voitaisiin kuvata symbolien avulla. Jos laskennassa tarvittavat säännöt voitaisiin määrittellä yksikäsitteisesti ja kuvata symbolisesti, niin niitä voitaisiin soveltaa mekaanisesti. Tutkijat ajattelivat, että tällöin ehkä pystyttäisiin rakentamaan kone, joka pystyisi ihmisen kaltaiseen matemaattiseen ajatteluun – ”muodostamaan yleistyksiä” ja ”oppimaan käsitteitä”, kuten McCarthy jälkeenpäin totesi.³

Aluksi tutkijat olivat hyvin optimistisia. Esimerkiksi Minsky, McCarthy ja Shannon kirjoittivat seminaarin apuraha-anomuksessa, että kymmenen tutkijan ryhmä kykenisi ”kahdessa kuukaudessa” ratkaisemaan, kuinka luonnollisen kielen oppimista tai käsitteiden ja abstraktioiden muodostamista voitaisiin simuloida tietokoneiden avulla. Pian tutkijat kuitenkin huomasivat, kuinka monimutkaisia haasteita tekoälyn luomiseen todellisuudessa sisältyy.

Pelkästään jo sääntö- tai logiikkapohjainen ohjelmointi, jossa tekoälyohjelmaan ohjelmoitiin käsin joukko sääntöjä, osoittautui sekä käytännössä kankeaksi että sovellusalueeltaan hyvin kapeaksi. Silloiset ohjelmistot, ensimmäisenä the Logic Theorist, soveltuivat hyvin *täsmällisesti rajattujen tai niin sanottujen hyvin määriteltujen* ongelmien ratkaisuun. Tyypillisiä esimerkkejä tällaisista ongelmista ovat erilaiset pelit, joiden ongelma-avaruus voidaan määrittellä yksiselitteisesti pelien sääntöjen avulla.

Koska tällaisten ongelmien parissa koneälyn nopeus ja raaka laskentateho ovat valttikortteja, on ymmärrettävää, kuinka ”vain” 40 vuodessa sääntöpohjaiset šakkialgoritmit ensin haastoivat ja lopulta voittivat ihmispelaajat. On kuitenkin myös huomattava, että parhaat šakkiohjelmat päihittävät ihmispelaajat niin kauan, kun *niiden ei tarvitse itse siirtää nappuloita*. Nappuloiden siirtäminen nimittäin muuttaa hyvin määritellyn ongelman avoimesti määritellyksi ongelmaksi. Ongelma ei ratkea enää vain pelin sääntöjen avulla, vaan vaatii arvioita nappuloiden koosta, muodosta, väristä, massasta, kitkasta, laudan sijainnista, etäisyyksistä ja katselu- ja siirtokulmista. Lisäksi lautojen koko voi

3 Solomonoff, G. Julkaisematon käsikirjoitus.

vaihdella, samoin nappulat voivat olla erikokoisia tai -muotoisia, ne voivat olla konkreettisia tai pikseleitä tietokoneen näytöllä jne. Kun šakkipeliä tarkastellaan sääntöjen määrittämän abstraktin hakuvaruuden sijaan robotiikan konkreettisesta näkökulmasta, ihmispelaajalle triviaali osatehtävä osoittautui sääntöpohjaiselle – ja myös nykyisille – tekoäly(i)lle vaikeaksi haasteeksi.

Avointen ongelmien ratkaiseminen edellyttää ongelma-avaruuden rajaamista, ongelman määrittelyä ja ratkaisustrategioiden valikointia. Nämä puolestaan edellyttävät kykyä erottaa oleellinen epäoleellisesta (niin sanottu *frame*-ongelma) sekä hahmottaa päättelyn, ongelmanratkaisun ja päätöksenteon epävarmuustekijöitä. Lähes kaikki päättely tai ongelmanratkaisu on ehdollista, toisin sanoen sen toimivuus perustuu epävarmoihin olettuksiin tapahtumien todennäköisyyksistä. Siksi siinä, missä šakissa epävarmuustekijöiden määrä on hallittavissa, toisin sanoen ne liittyvät yksinomaan vastustajan mahdollisiin valintoihin, todellisessa maailmassa epävarmuustekijöiden määrä on suurempi ja niiden huomioon ottaminen on vaikeampaa. Kolmanneksi, sääntö- ja logiikkapohjainen tekoäly ratkaisee ongelmia soveltamalla valmiiksi ohjelmoituja loogisia sääntöjä. Sen kyky oppia uutta tai soveltaa jo oppimaansa uusiin tilanteisiin on rajallinen. Esimerkiksi nerokkainkaan sääntöpohjainen šakkialgoritmi ei pysty pelaamaan pokeria, saati soveltamaan kognitiotaan talon remontointiin tai voileivän tekemiseen. Šakin maailmanmestari ihmisaivoineen sen sijaan pystyy.

Tutkijat kamppailivat pitkään näiden ongelmien kanssa. Vaikka näitä ongelmia ei ole vielä kukaan ratkaistu, tekoälyn tutkimus kuitenkin liikahti huomattavasti eteenpäin 1980-luvulla koneoppimisen kehityksen myötä. Koneoppimisella tarkoitetaan lyhyesti sitä, että valmiiksi ohjelmoitujen sääntöjen sijaan ohjelmistot oppivat itse maailmasta havaintojen (käytännössä yleensä tutkijan valitseman ja esikäsittelmän numeerisen aineiston) perusteella päättämään haluttuun lopputulokseen. Koneoppimisessa on kyse siis sellaisten algoritmien rakentamisesta, jotka kykenevät datan perusteella oppimaan ja muodostamaan sitä koskevia yleistyksiä ja ennusteita. Tällaiset algoritmit toimivat siten, että ne rakentavat niin sanotusta opetusaineistosta kuvauksen tai sisäisen mallin,

joka tehokkaasti tiivistää informaatiota aineistosta. Tämän perusteella ne pystyvät tekemään päätelmiä opetusaineistoon kuulumattomista tapauksista.

Koneoppimisen kehitys oli tekoälyn kannalta keskeistä siksi, että se tarjosi tavan rakentaa itsenäisesti oppivia koneita. Myös keinotekoisien hermoverkkojen tutkimuksen eli konnektionismin kehitys 1980-luvun taitteessa oli nykyisen tekoälyn kehittymiselle tärkeää. Se johti ensimmäiseen varsinaiseen ”neuroverkkobuumiin”, jolloin kiteytyivät monet alan perusolettamukset, käsitteet ja tekniikat, joille nykyiset syväoppimisen menetelmät perustuvat. (Esimerkiksi keinotekoisien hermoverkkojen rinnakkaislaskennan periaatteet, joiden pohjana on eräänlainen äärimmilleen yksinkertaistettu signaalinprosessointimalli biologisen hermosolun toiminnasta, määriteltiin jo 1960–70-luvuilla.)

Vaikka tekoälyn kehityksessä olikin 1980-luvun jälkeen hiljaisempi vaihe, tietokoneiden laskentatehon kasvun myötä hermoverkkosimulaatiomallien koko ja sisäinen monimutkaisuus lisääntyivät. Lisäksi etenkin 2000-luvulla rinnakkaislaskentamenetelmät kehittyivät, kun laskennassa alettiin hyödyntämään keskusprossessorin sijaan tehokaiden grafiikkakorttien ominaisuuksia. Samoin algoritmien parissa tehtiin koko ajan tärkeää kehitystyötä. Kaikki nämä edistysaskeleet yhdessä mahdollistivat sen, että viimeisen kymmenen vuoden aikana simuloitujen hermoverkkojen koko ja sisäinen monimutkaisuus – yhdessä massiivisten verkkojen oppimista tehostavien oppimismenetelmien kanssa – kehittyivät niin pitkälle, että verkkoja voitiin alkaa soveltaa entistä luonnollisempiin ja dynaamisempiin tehtäviin. Tällaisia tehtäviä olivat esimerkiksi valokuvien luokittelu jonkin kategorian perusteella, käsikirjoitetun tekstin muuttaminen konekirjoitukseksi tai 1980-luvun alun Atari-konsolipelien pelaaminen.

Tekoälyn nykytila ja tulevaisuus

Viimeiset 5–10 vuotta ovatkin olleet niin sanotun kapean tekoälyn täsmällisesti rajattujen ongelmien kehityksen kultakautta. Erityisesti on kohistu niin sanotuista *deep learning* (DL) -pohjaisista järjestelmistä. DL on eräänlainen yhdistelmä koneoppimista, konnektionismia, dynaamisten systeemin teoriaa ja bayesilaista matematiikkaa.

”Syvyys” Deep Learning -termissä viittaa siihen, että järjestelmät pystyvät hyödyntämään useita, käytännössä muutamasta muutamaankymmeneen, käsittely- tai analyysikerroksia⁴. Ne eroavat tässä suhteessa aiemmista hermoverkkomenetelmistä. Tällainen monikerroksinen rakenne mahdollistaa sen, että järjestelmä pystyy oppimaan sille syötetystä aineistosta, toisin sanoen datasta, piirteitä eräänlaisina piirrehierarkioina konkreettisista (esim. nappuloiden sijainnit) abstrakteihin (aseman strateginen tasapaino). DL-pohjaisten sovellusten etu on se, että ne kykenevät oppimaan aineiston säännönmukaisuuksia ”tavallista” koneoppimista paremmin ja käyttämään tehokkaammin etenkin suuria aineistoja.

Viime vuosina DL-pohjainen tehokas ja joustava informaation tiivistäminen on onnistuttu yhdistämään tavoitteelliseen toiminnanohjaukseen, etenkin niin sanotun *vahvistusoppimisen* (engl. *reinforcement learning*) avulla. Toisin kuin monissa perinteisissä ohjattuun oppimiseen perustuvissa hermoverkkojen oppimisalgoritmeissa, vahvistusoppimisessa tekoälyjärjestelmälle ei opeteta suoraan haluttuja toimintamalleja, vaan sitä opetetaan yksinomaan positiivisen tai negatiivisen palautteen kautta (esim. peli joko voitettiin tai hävittiin). Esimerkiksi suurta huomiota herättänyt AlphaGo Zero oppi pelkästään itseään vastaan pelaamalla strategioita, joita ihmispelaajat eivät Go-pelin vuosituhantisen historian aikana ole keksineet. Sillä ei siis ollut käytössään harjoitusvaiheessa massiivista aineistoa korkeatasoisista ihmispeleistä, joista se olisi voinut yleistää voittamisen kannalta ”hyvien siirtojen kokoelman”. Sen sijaan, AlphaGo Zero oppi pelin mestariksi harjoittelemalla itsenäisesti ja päivittämällä hermoverkkokerroksiaan hankkimansa palautteen perustella.

AlphaGo Zeron yhteydessä onkin paljon pohdittu sitä, mitä tästä kyvystä pitäisi päätellä. On selvää, että aiempiin AlphaGo:n versioihin verrattuna AlphaGo Zero on tehokkaampi. Lisäksi, ja kognitiivisesta näkökulmasta ehkä tärkeämmin, se ei käytä mallina *ihmisen* asiantuntijuutta oppiakseen esimerkiksi Go-pelin pelistrategiaa. Sen sijaan se pystyy suoraan toimimaan ”Go:n maa-

ilmassa” ja saavuttamaan ihmispelaajia korkeamman tason. AlphaGo Zeron kehittäjät kirjoittivatkin *Nature*-lehden artikkelissaan, että tässä mielessä AlphaGo Zero onkin ensimmäinen aidosti ”yli-inhimillinen” tekoälyjärjestelmä (Silver ym., 2017)⁵.

Toinen tärkeä seikka on se, että esimerkiksi šakkiin verrattuna Go on peli, jossa perinteiset siirtosarjojen laskentaan perustuvat algoritmit joutuvat nopeammin vaikeuksiin. Pelissä on mahdollisia – siis sääntöjen sallimia – siirtoja paljon enemmän, jolloin sen hahmottaminen, mitkä mahdollisista siirroista ovat oleellisia, korostuu. Tämä sivuaa perinteistä frame-ongelmaa: Jos Go on tässä mielessä pelinä sensitiivisempi kyvyille keskittyä olennaiseen kuin esimerkiksi šakki, niin osoittako AlphaGo Zeron tapaisten järjestelmien menestys, että koneet alkavat lähestyä ihmisälyä tämän hienovaraisen kyvyn osalta? Avaako AlphaGo Zero oven uuden polven tekoälylle, joka alkaa pärjätä myös ”tosimaailman” tehtävissä, kuten voileipien tekemisessä esimerkiksi hotdog-kioskeissa tai autonomisten avoneuvojen ohjaamisessa?

On luonnollisesti selvää, että esimerkiksi autenttiset liikennetilanteet tai niitä säätelevät säännöt ja normistot ovat huomattavasti monimutkaisempia kuin Go-pelin säännöt. Ne eivät koostu vain ajoneuvon ohjaamiseen liittyvistä säännöistä tai virallisista juridisista lakipykäliin tai autokoulujen opetusmateriaaleihin kirjatuista säännöistä. Ajotilanteisiin liittyvät fysiikan lait tai liikennekulttuuriin liittyvät epämuodolliset sosiaaliset tai käytännölliset säännöt ovat olennainen osa liikenteen sääntöjärjestelmä, jonka tunteminen on edellytys kyvyille tulkita viralliset säännöt. Aivan samalla tavalla voileipäresepti ei – vastoin usein esitettyä metaforaa – ole riittävä säännöstö, algoritmi, leivän toteuttamiseen. Sen sijaan voileivän tekeminen vaatii huomattavasti kontekstisidonnaista tiedon valikointia ja päättelyä ”rivien välistä”.

On joka tapauksessa varmaa, että monet tekoälysovellusten kehittämiseen osallistuvat suuryritykset eivät käytä massiivisia kehitysresurssejaan

4 ”Syvyys” ei siis viittaa siihen, että kehityskulkuja välttämättä laskettaisiin ”syvälle” hakuvaruudessa, kuten perinteisissä šakkialgoritmeissa.

5 Itse asiassa jo 1990-luvulla niin sanottu TD-gammon-vahvistusoppimisalgoritmi teki saman backgammonissa. Sen seurauksena esim. monet ”parhaaksi tiedetyt” aloitukset kyseenalaistettiin.

pelkästään inhimillisen uteliaisuuden tyydyttämisiksi. Tavalla tai toisella aineistosta itsenäisesti oppivia tekoälyohjelmia, jotka Alpha Zeron tavoin voivat kehittää *ihmisen ajattelusta poikkeavia* ratkaisuja ja käyttäytymismalleja, tullaan soveltamaan entistä laaja-alaisemmin kuluttajatuotteiden, so-
tateollisuuden ja erilaisten perus- ja ylellisyyspalvelujen alalla.

Käytännössä siis on mahdollista, että kone, jonka tiedonkäsittely poikkeaa meidän tiedonkäsittelystämme, ohjaa autoa, päättää asuntolainoista, tekee suunnitelmat terminaaliptilaan hoitotoimenpiteistä, ohjaa miljardien sijoituspäätöksiä tai lentää ”parviälyksi” yhdistettyä taistelulentokonelaivuetta. Kun keinotekoiset järjestelmät tekevät käytännössä päätöksiä ihmisten elämästä (ja kuolemasta), nousee esiin eettisiä ja juridisia kysymyksiä oikeuksista ja vastuista niin järjestelmien kehityksessä kuin hyödyntämisestä jonkin ryhmän etujen ajamisessa (joitakuuta toisia vastaan).

Mitä annettavaa ihmiskognition tutkimuksella on koneälyn tutkimukselle?

Useat tekoälyn ja koneoppimisen kehitystä ajaneet ajatukset ovat peräisin ihmisen älykkään käyttäytymisen tutkimuksesta⁶. Inhimillinen äly tulee jatkossakin olemaan vahva inspiraation lähde tekoälyn teknisessä kehityksessä⁷. Toisaalta tekoälyn tavoitteena ei ole simuloida tai toisintaa nimenomaan *ihmisälyn* ominaispiirteitä, vaan kehittää ylipäättään mahdollisimman älykkäitä konejärjestelmiä. Mitä enemmän ihmisälystä *poikkeavaa* älykkyyttä tällaisiin laitteisiin rakennetaan, sitä tärkeämpää on ymmärtää syvällisellä tavalla sekä

ihmisen että koneen ajattelua. Monet – Googlea ja Elon Muskia myöten – ovat huolissaan tekoälyn hybridi- tai symbioosikehitykseen liittyvistä riskeistä (kuten kommunikaatio-ongelmista tai mahdollisista systemaattisista kognitiivisista vireistä).

Näiden riskien kartoittaminen ja niihin vastaaminen vaatii ymmärrystä erilaisten kognitiivisten järjestelmien yhteensovittamisesta. Tämä puolestaan vaatii perustutkimusta erilaisten kognitiivisten järjestelmien toiminnasta. Siksi on ehkä tärkeää – *ja ehkä tärkeämpää kuin koskaan aiemmin* – tutkia sitä, miten älykkyyks ilmenee niin ihmis-, eläin- kuin koneälyssä nimenomaan kognitiotieteen näkökulmasta. Se tarjoaa mahdollisuuden tutkia tekoälyn kognitiivisia rakenteita, niiden seurauksia ja myös mahdollisia riskejä sellaisesta näkökulmasta, jossa erilaisten toimijoiden muodostamien järjestelmien monimutkaisuus otetaan aidosti huomioon. Mallit, joilla kognition tutkimuksessa nykyisin operoidaan, ovat usein monitasoisia laskennallisia agentti-ympäristötoiminnanohjausmalleja, joissa painopiste on informaationprosessoinnissa, sen valikoinnissa ja muokkaamisessa.

Kognitiotiede ei ole vain tekoälyjärjestelmien tai vain ihmisaivojen tutkimusta, vaan se avaa myös mahdollisuuksia tutkia ja ymmärtää älyn ja tiedonkäsittelyn luonnetta ylipäänsä. Siksi kognitiotiede on tekoälytutkimukselle tärkeää nimenomaan kognition eri muotojen empiirisenä tieteellisenä tutkimuksena – ei yksinomaan teoreettisena pohdiskeluna kaikista mahdollisista tekoälyyn tai ylipäänsä älykkäisiin järjestelmiin ja ihmismieleen liittyvistä kysymyksistä.

6 Monet koneoppimisen/ ”tekoälyn” nykyteknikat ovat alun perin syntyneet nimenomaan kognitiotieteen tai yleisemmin käyttäytymistieteiden parissa. Esimerkiksi AlphaGo:n ja AlphaGoZeron hyödyntämä vahvistusoppiminen on muunnos alunperin behavioristien esittämistä ajatuksista. Konnektionismista spekulointiin jo paljon ennen perseptro-neja. Samoin viime aikoina objektin tunnistuksessa esille nostettu ”attentio” perustuu aiemmille malleille ”foveoinnista”, joka saattaa osoittautua spatiaalista invarianssia tehokkaammaksi tavaksi toteuttaa tiettyjä objektin tunnistukseen liittyviä toimintoja. Asiantuntijat myös ennakoivat, että kontrollimallien hyödyntämistä niin robotiikassa kuin tekoälyssä tullaan lisäämään.

7 Tekoälyn kehitys on ollut tärkeätä myös ihmiskognition tutkimuksen kannalta. Keinotekoisiiin hermoverkkoihin perustuvia tekoäly ja sen algoritmien kehittäminen ovat tarjonneet välineitä myös sen hahmottamiselle, miten – hahmottamistutkimukseen ja yleistämiseen perustuvat – tiedonkäsittelyjärjestelmät voivat toimia laskennallisesti.

Kirjallisuus

- Silver, D., Schrittwieser, J. Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A. Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G. Graepel, T. ja Hassabis, D. ”Mastering the game of Go without human knowledge”, *Nature*, 550, 354–359 (19 October 2017).
Solomonoff, G. ”Ray Solomonoff and the Dartmouth Summer Research Project in Artificial Intelligence, 1956”, Julkaisematon käsikirjoitus.

Otto Lappi on dosentti, Anna-Mari Rusanen tutkijatohtori ja Jami Pekkanen tohtorikoulutettava kognitiotieteessä Helsingin yliopiston Digitaalisten ihmistieteiden osastossa.