



ASKO NIVALA

# Onko tekoälyä olemassa?

2010-luvulla ei ole voinut välttyä tekoälyä koskevalta keskustelulta. Teknologiayrittäjä Elon Muskin mukaan tekoäly on ydinaseita vaarallisempi uhka ihmiskunnalle<sup>1</sup>. Samaan aikaan Muskin johtama Tesla kehittää automaattiohjauksella varustettuja autoja, jotka pysyvät ainakin suurimman osan ajasta kaistaviivojen välissä. Muskin suhtautuminen tekoölyyn vaikuttaa ristiriitaiselta, ellei oteta huomioon, että hän kenties tarkoittaa vaarallisella tekoölyllä vahvaa tekoälyä, kun taas Teslan koneoppimiseen perustuva autopilotti edustaa heikkoa tekoälyä<sup>2</sup>. Myös Onnibussin, Uberin ja Foodoran kaltaiset disruptiiviset yritykset hyödyntävät hinnoittelussaan ja liiketoimintamallissaan koneoppimista ja dataa kerääviä kännykkäsovelluksia. Tekoölystä on tullut bisnesmaailman *buzzword*, jonka mainitsemalla start up -yritysten on mahdollista avata enkelisijoittajien kukkarojen nyörit. Tosiasiassa kukaan ei ole vielä kehittänyt keinotekoisia toimijaa, jolla olisi ihmismieleen verrattava tietoisuus. Koneet pystyvät toki tunnistamaan kasvoja ja säveltämään virheettömiä Bachin koraaleja, mutta nämä koneoppimisen sovellukset perustuvat tilastollisten mallien kouluttamiseen tarpeeksi laajalla aineistolla.

**T**ekoälykeskustelua tuntuu luonnehtivan käsitteellisten sekaannusten ohella presentismi ja antropomorfismi. Ihmiset ovat aiemminkin olettaneet juuri oman aikakautensa olevan ratkaisevien teknisten läpimurtojen kynnyksellä. Läpi historian he ovat projisoineet itsensä luomiinsa koneisiin. Kuinka perusteltuja tekoölyyn liittyvät utopiat ja dystopiat ovat historiallisesti tarkasteltuna? Kun tekoälykeskustelua tarkastellaan historiallisesti, voidaan myös hahmottaa, mitä eroa on vahvalla tekoölyllä, symbolisella tekoölyllä ja neuroverkolla.

## Haave tekoölystä on yhtä vanha kuin ihminen

Historiallisesti katsottuna tekoälykeskustelua on luonnehtinut kateettoman innostuksen ja syvän pettymyksen kausien vaihtelu. Ajatus ihmisenkaltaisesti toimivasta mekaanisesta automaatista ei ole uusi vaan pikemminkin yhtä vanha kuin länsimainen kirjallisuus. Homeroksen *Iliaan* mukaan seppäjumala Hefaistos valmisti metallista automaatteja, jotka työskentelivät hänen apunaan.<sup>3</sup> Myös *Talmudin* savesta muovattu Golem voidaan ajatella haaveeksi ihmisen luomasta keinotekoisesta apulaisesta, joka kykenee suorittamaan itsenäisesti helppoja tehtäviä. Itse asiassa Aadam ja hänen jälkeläisensä ovat kenties itsekin eräänlaisia savesta luotuja Golemeja, mutta juutalaisen perinteen mukaan ihmisen luomat Golemit ovat hänestä itsestään poiketen mykkiä ja typeryytensä vuoksi erittäin vaarallisia: ne kääntyivät lopulta luojaansa vastaan.

Vaikka automaateilla on esihistoriansa myyttis-maagisen ajattelun antropomorfismeissa, emme kenties nykyisin pidä magian keinoin animoituja älykkäitä olentoja

varsinaisena tekoölynä. Modernin tekoälykeskustelun olennainen juuri on pikemminkin komputationalismissa ja funktionalistisessa mielenfilosofiassa. Komputationalismiksi kutsutaan kognitiivisen mielenfilosofian suuntausta, jonka mukaan ihmisajattelu palautuu symbolisia representaatioita käsittelevään sääntöjärjestelmään<sup>4</sup>. Siinä missä maagisiin uskomuksiin sisältyy oletus, että eloton materia voidaan herättää taikakeinoilla henkiin, komputationalismi päin vastoin olettaa, että ajattelu voidaan palauttaa laskennaksi, jolloin se voidaan toteuttaa myös koneellisesti.

Komputationalismi on tietokoneita vanhempi mielenfilosofian paradigma. Pikemminkin jotkut tietokoneen taustalla olevat matemaattiset innovaatiot syntyivät varhaisten komputationalistien inspiroimana. Majorcalainen filosofi Ramon Llull (1232–1315), Thomas Hobbes ja Gottfried Wilhelm Leibniz kehittivät vuorollaan ajattelun formaaleja sääntöjä, joiden avulla argumentoiminen voitaisiin palauttaa laskemiseksi. Kiinassa lähetystyötä tehneen jesuiitta Joachim Bouvet'n (1656–1730) innoittamana Leibniz myös tutki binäärilukujärjestelmää, joka on nykyisten tietokoneiden toiminnan perustana. Jälleen tietotekniikan ja maagisen ajattelun historia kietoutuvat yhteen, sillä Leibnizin binäärijärjestelmä otti vaikutteita kiinalaisesta *Muutosten kirjasta* (*Yijing*, noin 800 eaa.) ja sen ennustustaitoon liittyvistä heksagrammeista.<sup>5</sup> Leibniz esitti, että binääriluvuilla toteutettavat yhteen-, vähennys-, kerto- ja jakolaskut voitaisiin laskea mekaanisella laitteella – vieläpä helposti!<sup>6</sup> Blaise Pascal ja Leibniz rakensivat tahoillaan laskukonetta seuraten Wilhelm Schickardin (1592–1635) esimerkkiä, joka rakensi ensimmäisen mekaanisen laskukoneen vuonna 1623<sup>7</sup>.

Erityisesti Leibnizin kohdalla komputationalismi ei kuitenkaan rajoittunut laskukoneeseen vaan siihen sisältyi myös käsitteellinen yleiskieli *characteristica universalis* ja loogisen laskennan pohjana toimiva *calculus ratiocinator*. Yhdessä ne kykenisivät formalisoimaan kaiken käsitteellisen ajattelun mekaanisiksi säännöiksi, jotka olisivat yhtä intuitiivisia ja erehtymättömiä kuin matemaattiset todistukset. Tämän tuloksena Leibniz lupasi – hiukan Timo Honkelan *Raubankoneen* (2017) tavoin – koneiden päättävän Eurooppaa repivät uskonsodat, sillä ajattelun matematisointi estäisi teologiset väärintymmärrykset<sup>8</sup>. Ei ole aivan itsestään selvää, mitä Leibniz tarkkaan ottaen tarkoitti *calculus ratiocinatorilla*, mutta kybernetiikan perustajan Norbert Wienerin mukaan Leibniz kuitenkin asetti sillä perustan modernille tietokoneelle, joka ei prosessoisi pelkkiä lukuja vaan kykenee olemaan ”ajatteleva kone” (*the reasoning machine*)<sup>9</sup>.

## Kesät ja talvet vuorottelevat

Tietotekniikan ja tekoälyn historia on täynnä suuria suunnitelmia ja utopistisia lupauksia, joita ei ole kuitenkaan osattu toteuttaa käytännössä. Leibnizin työtä jatkoi Charles Babbage (1791–1871), joka kuvasi ensimmäistä kertaa vuonna 1837 analyttisen koneen. Se oli tarkoitettu yleiskäyttöiseksi mekaaniseksi tietokoneeksi, joka käyttäisi muistinaan reikäkortteja.<sup>10</sup> Analyttistä konetta ohjelmoitaisiin erityisellä konekielellä, jonka on arveltu olleen jo Turing-täydellinen eli kykenevä simuloimaan minkä tahansa muun tietokoneen toimintoja. Babbagen kaavailema tietokone nimittäin käytti pelkkiä syötteitä sisältävien reikäkorttien ohella myös erillisiä käskykortteja, jolloin kone kykeni vaihtamaan tilaansa ja siirtelemään reikäkortteja käskyjen mukaan eteen- ja taaksepäin eli suorittamaan hypyn muistiosoitteesta

toiseen. Analyttinen kone jäi lähes kokonaan suunnitelmien tasolle eikä ole edes selvää, olisiko sitä voitu toteuttaa aikalaisteknologialla. Tästä huolimatta lordi Byronin tytär Ada Lovelace (1815–1852) ehti kirjoittaa analyttiselle koneelle ensimmäisenä tietokoneohjelmaksi pidetyn algoritmin Bernoullin lukujen laskemiseksi<sup>11</sup>.

Vaikka ajatus tietokoneesta on siis esitetty jo aiemmin historiassa, toimivaa laitteistoa saatiin odottaa kauan. Vielä 1900-luvun alkupuolella sana *computer* ei viitannut laitteeseen vaan matalapalkkaisten naisten ryhmiin, jotka suorittivat tieteellisen ja sotilaallisen laskennan tarpeisiin vaivalloisia rutiinilaskuja kuten ballistisia taulukoita. Ennen digitaalisia eli täsmällisillä portaittaisilla arvoilla laskevia tietokoneita oli olemassa myös analogisia tietokoneita, joiden arvot vaihtuivat portaattomasti ja jotka tuottivat tuloksenaan likiarvoja. Saksalaisen Konrad Zusen vuonna 1941 toteuttama Z3 oli jo täysin ohjelmoitava, digitaalinen ja Turing-täydellinen tietokone, joka oli kuitenkin vielä suurelta osin mekaaninen. Brittiläisen Women’s Royal Naval Servicen kryptografiaan käyttämä Colossus oli jo täysin sähköinen muttei kuitenkaan puolestaan Turing-täydellinen tietokone. Vasta 1945 valmistunut ENIAC oli täysin elektroninen ja Turing-täydellinen. Toisen maailmansodan tarpeisiin syntyneitä tietokoneita käytettiin kotirintamalla ballistiseen laskentaan ja tästä syystä myös ensimmäiset tietokoneohjelmoijat olivat tavallisesti naisia<sup>12</sup>.

Tietotekniikan varhaishistoria oli pääasiassa laskukoneiden kehittämistä ennen kuin George Boole (1815–1864) kehitti lausekalkyylin. Boolean algebran operaattorit kuten AND (konjunktio), OR (disjunktio) tai NOT (negaatio) toimivat loogisina konnektiiveina, joita informaatioteorian perustaja Claude Shannon (1916–2001) sovelsi kymmeniä vuosia myöhemmin elektroniikassa<sup>13</sup>. Loogisten porttien avulla digitaaliset

”Kun ihminen järkeilee, hän ei tee muuta kuin ajattelee summaa, joka saadaan kun osia *lasketaan yhteen*, tai ajattelee erotusta, joka saada, kun yksi summa *vähennetään* toisesta. Jos tämä tehdään sanoin, silloin ajatellaan sarjaa kaikkien osien nimistä kokonaisuuden nimeen tai kokonaisuuden ja yhden osan nimistä toisen osan nimeen. [...] Nämä operaatiot eivät koske yksinomaan lukuja, vaan kaikenlaisia olioita, joita voidaan laskea yhteen ja vähentää toisistaan. Sillä kuten aritmeetikot opettavat laskemaan yhteen ja vähentämään *luvuilla*, niin geometrikot opettavat tekemään samaa *janoilla, kuvioilla* (avaruudessa ja pinnassa), *kulmilla, suhteilla, ajoilla, nopeuksilla ja voimilla* ja niin edelleen. Loogikot opettavat samaa *sanasarjoilla*, jolloin kaksi nimeä lasketaan yhteen ja saadaan *sylogismi*, ja *monta sylogismia* lasketaan yhteen ja saadaan *todistus*, ja *sylogismin summasta* eli *johtopäätöksestä* he vähentävät yhden *proposition* ja löytävät toisen. Poliittiset kirjoittajat laskevat yhteen *välipuheita* saadakseen selville ihmisten *velvollisuudet*, ja lakimiehet laskevat yhteen *lakeja* ja *tosiasioita* saadakseen selville, mikä on itse kunkin toiminnassa *oikein* ja *väärin*. Ylimalkaan kaikissa asioissa missä on sijaa *yhteenlaskulle* ja *vähentämiselle* on sijaa myös *järjelle*, ja missä näillä ei ole sijaa, siellä *järjellä* ei ole mitään tekemistä.”

**Thomas Hobbes, *Leviathan* (1651). Suom. Tuomo Aho. Vastapaino, Tampere 1999, 57.**

piirit kykenevät toteuttamaan näitä loogisia operaatioita binäärisille syötteille. Norbert Wienerin perustama kybernetiikka syntyi läheisessä yhteydessä informaatioteoriaan. Nykyisessä lehtimieskielessä kybersodankäynnillä tarkoitetaan tietohyökkäyksiä ja internetissä julkaistua propagandaa. Täsmällisemmin ymmärrettynä kyberneettiset järjestelmät perustuvat palautesilmukkaan, jonka avulla ne pystyvät säätelemään itse itseään ja pysymään annettujen raja-arvojen puitteissa. Wiener sovelsi kybernetiikkaa tähtäystään automaattisesti korjaaviin ilmatorjuntatykkeihin mutta havaitsi myös orgaanisen elimistön säätelevän omaa toimintaansa pyrkien homeostaasiin eli tasapainoon. Esimerkiksi ylikuumentunut keho alkaa hikoilla, jotta se voi säädellä ruumiinlämpöä. Ymmärrettävästi kybernetiikka on vaikuttanut merkittävästi niin tekoälyn teoriaan kuin robotiikkaankin, vaikka se on niitä paljon laajempi tutkimusala.<sup>14</sup>

1950- ja 1960-luvuilla tietotekniikka kehittyi nopeasti, ja aikalaiset ajattelivat ihmisen tasoisen ellei paremmankin tekoälyn valmistuvan muutamien vuosien kuluessa. Tutkimusalue nimeltä *Artificial Intelligence* perustettiin vuonna 1956 Dartmouth Collegessa järjestetyssä kesäkoulussa. Klassinen symbolinen tekoäly perustui oletukselle, jonka mukaan tekoälyä olisi kehitettävä ylhäältä alaspäin niin, että tietokone suorittaa sille eksplisiittisesti annettuja sääntöjä älykkään tuloksen saavuttamiseksi. Tyypillinen esimerkki tästä klassisesta lähestymistavasta ovat erilaiset haku- ja päätöspuut (*decision tree*). Esimerkiksi vuosina 1966–1972 kehitetty Shakey-robotti hyödynsi reittiensä valinnassa A\*-algoritmia, jota monien tietokonepelien tekoäly edelleen käyttää lyhimmän mahdollisen reitin laskemiseen kahden solmun välillä. Myös voittoa shakkikone Deep Blue perustui edelleen kovakoodattuihin shakin sääntöihin, laajaan avauskirjastoon ja tehokkaaseen heuristiikkaan hakupuun vaihtoehtojen karsimiseksi.

Komputationalismin kanssa kilpaileva konnektionistinen paradigma teki myös läpimurron 1950-luvulla. Siinä missä komputationalismi ymmärtää mieltämisen laske- misena, konnektionismi olettaa mielen olevan verkosto, jossa monimutkaisemmat toiminnot rakentuvat emergentisti (alhaalta ylöspäin) yksinkertaisten toimintojen yhdistelmistä. Vuonna 1958 psykologi Frank Rosenblatt keksi perseptronin. Se oli ensimmäinen neuroverkko – sähkö- johdoista ja potentiometreistä koostuva fyysinen laite, jota käytettiin kuvantunnistukseen. Klassinen komputationalismi seuraa Turingin koneen teoreettista mallia, jonka muistilaitte sisältää ohjelman eli joukon ohjeita, joiden perusteella kone suorittaa siirtymät tilasta toiseen. Neuroverkkoa taas ei ohjelmoida käsky käskyltä, kuten tavanomaisia sekventiaalista laskentaa suorittavia tietokoneita, vaan neuroverkko perustuu ihmisaivoja jäljitellen massiiviseen määrään rinnakkaisia ”neuroneita”. Ohjelmoinnin sijasta keinotekoisia neuroverkkoja ”koulutetaan” esimerkiksiaineistolla syötteitä ja tulosteita. Verkon onnistuneeseen koulutukseen tarvitaan tilastollisesti edustava aineisto. Vaikka neuroverkolle ei anneta eksplisiittisiä sääntöjä, joiden mukaan sen tulisi ennustaa syötteen mukainen tuloste, sen arkkitehtuuriin sisältyy algoritmi, joka korjaa koulutuksen aikana neuronien välisiä painotuksia.

Perseptroni oli siis pohjimmiltaan piiri, joka summasi syötteitä oppimiensa painotusten mukaan. Vaikka neuroverkkoja mallinnetaan nykyään tavallisesti tietokoneiden avulla, ei neuroverkkoa tule sekoittaa konnektionismin kanssa kilpailevaan komputationalismiin. Yhtä lailla tietokoneita käytetään ihmisaivojen ohella vaikkapa meteorologisten ilmiöiden mallintamiseen, mutta tästä ei seuraa, että meteorologiset systeemit olisivat itsessään tietokoneita.

Perseptroniin, kuten tekoälyyn yleensäkin, ladattiin aikanaan kovia odotuksia. Rosenblatt julisti *The New*

”Jos sen sijaan olisi sellaisia koneita, jotka muistuttavat meidän ruumistamme ja jäljittelevät meidän toimintojamme siinä määrin kuin on asiallisesti mahdollista, meillä olisi aina kaksi täysin varmaa keinoa todeta, etteivät ne suinkaan ole aitoja ihmisiä. Ensiksikään ne eivät koskaan pystyisi käyttämään sanoja eivätkä yhdistelemään muitakaan merkkejä niin kuin me teemme ilmaistessamme ajatuksia toisillemme. Voidaan kyllä hyvin kuvitella puhetta tuottava kone, joka tuottaa sanoja ja vieläpä sellaisia, jotka vastaavat sen elimissä muutosta aiheuttavien esineiden aktioihin: jos sitä vaikkapa kosketetaan johonkin kohtaan, niin se kysyy, mitä sille halutaan sanoa, tai jos sitä kosketettaisiin johonkin toiseen kohtaan, niin se huutaa, että sille tehdään pahaa, ja niin edelleen. Mutta se tuskin osaisi järjestellä sanojaan eri tavoin vastatakseen mielekkäästi kaikkeen siihen, mitä sen läsnäollessa puhutaan, mihin kaikkein typerimmätkin ihmiset pystyvät. Toiseksi, vaikka sellaiset koneet suorittaisivat useita tehtäviä yhtä hyvin kuin me tai jopa paremmin kuin kukaan meistä, ne taatusti epäonnistuvat joissakin muissa, mikä paljastaisi sen, ettei niiden toiminta perustu tietämiseen vaan pelkästään elimien järjestykseen.”

**René Descartes, *Metodin esitys (Discours de la Méthode, 1644). Teokset I.***  
**Suom. Sami Jansson. Gaudeamus, Helsinki 2001, 117–168 (154–155).**

*York Timesin* artikkelissa vuonna 1958, että perseptroni kykenisi tulevaisuudessa kävelemään, puhumaan, näkemään, olemaan tietoinen ja vieläpä lisääntymään<sup>15</sup>! Myös aina valpas *Länsi-Savo* uutisoi vuonna 1960:

”Aakkoset osaava kone

Mark I Perceptron on nimeltään Cornell Aeronautical Laboratory Inc.-toiminimen rakentama kokeellinen kone, joka tuntee aakkoset sen jälkeen kun ihminen on ne sille opettanut. [...] Mark I Perceptronille aakkoset opetetaan suunnilleen samoin kuin lapselle. Opettaja panee tietyn kirjaimen koneen valokennosilmän eteen ja pitää siinä kunnes kone on suorittanut valinnan. Heti tämän jälkeen kirjain poistetaan. Jos kone tekee virheen, opettaja korjaa sen valvontalaitteilla ja panee koneen vastaamaan oikein.”<sup>16</sup>

*Länsi-Savon* artikkeli oletti perseptronin oppivan aakkoset samalla tavoin kuin lapsi oppii ne, kunhan verkon kouluttaja korjaa riittävän monta kertaa sen tulosteita. Neuroverkkoja sovelletaan edelleen esimerkiksi merkkien optisessa tunnistuksessa (OCR), mutta olisi silti liioiteltua ajatella perseptronin kaltaisen alkeellisen summauspiirin ”osaavan aakkoset”.

Rosenblattin saama julkisuus ja liioitellut väitteet ärsyttivät erityisesti hänen opiskelukaveriaan Marvin Minskyä, joka oli itsekin tehnyt tutkimusta perseptronien parissa. Minsky julkaisi Seymour Papertin kanssa neuroverkkojen rajoituksista teoksen *Perceptrons* (1969). Minsky esimerkiksi huomautti, että Boolean logiikan XOR-konnektiivia ei voida toteuttaa perseptronilla, mikä teki sen käyttöalueesta varsin rajallisen. Tämä on totta perseptronin kaltaisten yksikerroksisten verkkojen kohdalla, mutta ongelma ei kuitenkaan koske nykyisiä monikerroksia neuroverkkoja, jotka osaavat myös itse optimoida painotuksiaan takaisinkytkennän avulla ja käyttävät myös useita kerroksiaan eräänlaisena muistina. Tietehistorioitsijat ovat esittäneet, että Minskyn osin liiankin ankara kritiikki neuroverkkoja kohtaan johti painopisteen siirtymiseen takaisin aiempaan komputationalismiin ja ylhäältä–alas-suunnitellun symbolisen tekoälyn kehittämiseen.<sup>17</sup>

1970-lukua luonnehti yleisesti pettymys tekoälyn lunastamattomiin lupauksiin ja myös tutkimusrahoituksen leikkaukset. Jaksoa kutsutaan tekoälytalveksi (*AI winter*). Perseptronissa havaitut puutteet johtivat konnektionistimista luopumiseen, kun yksinkertainen summauspiiri ei oppinutkaan 1960-luvun kuluessa puhumaan, lukemaan ja lisääntymään. Konnektionistinen paradigma nousi kuitenkin 1980-luvulla uuteen kukoistukseen, kun David Rumelhartin ja James McLellandin johtama PDP-tutkimusryhmä esitteli, kuinka neuroverkon painotuksia säättävää *backpropagation*-algoritmia voisi soveltaa monikerroksisiin neuroverkkoihin<sup>18</sup>. Tämä oli ratkaiseva edistysaskel suhteessa perseptroniin. 1980-luvulla saatavilla ollut laskentateho ei kuitenkaan vielä riittänyt näiden neuroverkkojen täysimittaiseen hyödyntämiseen, joten ne painuivat unohduksiin. Komputa-

tionalistinen paradigma koki yhtä lailla takaiskuja, kun esimerkiksi kylmän sodan kannalta olennainen kielten välinen konekääntäminen venäjstä englanniksi ei onnistunutkaan toivotulla tavalla. Monimutkaisiin symbolisiin IF–THEN-ehdolauseisiin perustuvat 1980-luvun asiantuntijajärjestelmät toivat symbolisen tekoälyn hetkeksi parrasvaloihin. Asiantuntijajärjestelmät eivät kuitenkaan kyenneet oppimaan uusia asioita, joten niiden kirjoittaminen oli hyvin hidasta, eivätkä ne kyenneet reagoimaan yllättäviin tilanteisiin. 1980-luvun loppua luonnehti tekoälyn takatalvi. 1990-luvulla kiinnostus tekoälyyn lisääntyi uudelleen kulminoituen siihen, että Deep Blue voitti shakin hallitsevan maailmanmestarin Garri Kasparovin vuonna 1997.

Vaikka 2000-luvun koneoppiminen perustuu moniin muihinkin toimiviksi havaittuihin tilastollisiin malleihin alkaen 1700-luvulla kehitetystä bayesiläisestä tilastotieteestä, uuden sukupolven koneoppimisen leimallinen piirre on kuitenkin ollut uudelleen herännyt kiinnostus neuroverkkojen mahdollistamaan syväoppimiseen (*deep learning*). Syväoppiminen soveltaa edelleen Rumelhartin ja McLellandin 1980-luvulla kehittämiä algoritmeja mutta käyttää vain paljon isompia neuroverkkoja. Monet klassisen symbolisen tekoälyn algoritmit ovat toki myös edelleen käytössä, mutta nykyinen innostus tekoälyyn on perustunut varsin yksinkertaisesti saatavilla olevan digitaalisen datan määrän eksponentiaaliseen lisääntymiseen ja laskentatehon kasvuun. Nämä seikat suosivat nimenomaan konnektionistista paradigmaa, joka hyötyy graafiseen laskentaan erikoistuneiden suorittimien kyvystä rinnakkaiseen laskentaan. Tavallisissa prosessoreissa (CPU) on yleensä kaksi tai neljä rinnakkaista ydintä, kun taas grafiikkasuorittimissa (GPU) on satoja rinnakkaiseen laskentaan erikoistuneita ytimiä.

Aika näyttää, seuraako myös 2000-luvun kahden ensimmäisen vuosikymmenen tekoälybuumia vastaava krapula kuin aiemmin. On täysin mahdollista, että myös nykyinen tekoälyn kehitys pysähtyy useiksi vuosiksi tai vuosikymmeniksi tietylle tasolle ja ihmiset pettyvät, kun koneet eivät sittenkään opi ajattelemaan ja toimimaan itsenäisesti kuten Homeroksen kuvittelemat Hefaistoksen automaattit.

## Tekoälyä vai koneoppimista?

Tekoälyn määrittäminen ei ole mitenkään helppoa, sillä ’älykkyydelle’ ei liene olemassa yhtä kaikkien hyväksymää määritelmää. Älykkyyden määrittely on perustunut usein normatiivisiin valtarakenteisiin, joissa esimerkiksi miehet on määritelty naisia ja lapsia älykkäämmiksi, eurooppalaiset muita ihmisiä älykkäämmiksi tai ihmiset eläimiä ja koneita älykkäämmiksi. Feminismi, postkolonialismi ja posthumanismi ovat haastaneet näitä oletuksia ja osoittaneet älykkyyden määritelmien taustalla olevia kulttuurihistoriallisia ehtoja. Vaikka miespuoliset korkeakoulutetut insinöörit pitävät usein abstraktien symbolien manipulointia korkean älykkyyden merkinä, koneilla on silti edelleen suuria vaikeuksia selviytyä monista yk-



Kuva 1. Berthe Morisot, *Kehto* (Le Berceau, 1872). Kuvantunnistukseen koulutetun neuroverkon mukaan tässä maalauksessa on: *mosquito net* (69,5 %), *cradle* (21,1 %), *gown* (2,3 %), *hoopskirt* (1,4 %) tai *abaya* (0,8%). Kuva:Wikimedia Commons.

sinkertaisista tehtävistä, jotka ovat pienille lapsille vaivatonta. Moravecin paradoksin mukaan koneita on paljon helpompi opettaa todistamaan matemaattisia lausekkeitä tai pelaamaan shakkia kuin opettaa niille edes alkeellisia motorisia taitoja kuten esimerkiksi kävelemään portaat ylös tai solmimaan kengännauhat.<sup>19</sup>

Usein tutkimuksessa tehdään myös erottelu vahvaan ja heikkoon (tai laajaan ja suppeaan) tekoälyyn. Vahva tekoäly kykenisi toimimaan millä tahansa elämän alueella ja sillä olisi ihmisen kaltainen tietoisuus, tahto tai ainakin ymmärrys siitä, mitä se on tekemässä. Heikko tai suppea tekoäly taas tarkoittaa yhden tarkasti rajatun tehtävän suorittamiseen räätälöityä ohjelmistoa. Perinteisen symbolisen tekoälyn ohella se sisältää myös varsin kirjavan joukon erilaisia koneoppimisen ja syväoppimisen ratkaisuja. Mediassa kirjoitetaan usein tekoälystä tekemättä erottelua vahvaan ja heikkoon tekoälyyn sekä koneoppimiseen. Hyppy Teslan autopilotista tai go-peliä voitokkaasti pelaavasta neuroverkosta aidosti tietoiseen koneeseen voi olla pitkä – jopa mahdoton.

Erona symboliseen tekoälyyn koneoppiminen on tilastotieteeseen pohjautuvaa mallien laskemista ja syötteiden arviointia mallien pohjalta. Se ei poikkea millään olennaisella tavalla vaikkapa sään ennustamisesta tieto-

konemallin avulla. Kuitenkaan sääennusteista ei puhuta tavallisesti tekoälynä. Yhtä lailla esiluokitetulla kuva-aineistolla opetettu neuroverkko voi ennustaa, että tietty binäärilukujen jono (eli digitaalinen kuvatiedosto) on suurella todennäköisyydellä tietynlainen kuva. Vaikka esimerkiksi esineitä valokuvista tunnistavat verkot ovat varsin kehittyneitä, ennusteen onnistuminen riippuu tietenkin täysin koulutusdatasta, sillä neuroverkot eivät edelleenkään ymmärrä sen paremmin kielellisten ilmausten kuin kuvien kontekstia. Kuvassa 1 on Berthe Morisot'n impressionistinen maalaus *Kehto* (Le Berceau, 1872), jonka syötin valokuvilla koulutettuun neuroverkkoon. Ohjelman mukaan kuvassa on 69,5 prosentin todennäköisyydellä hyttysverkko, mutta 21,1 prosentin varmuudella esine on kehto. Tulos on samaan aikaan sekä hyvä että huono. Huvittavaa kyllä, kone erehtyi pitämään kehdon päällä olevaa huntua hyönteisverkkona. Toisaalta kehto oli toiseksi vahvin arvaus. Lisäksi käytetty kirjasto on koulutettu valokuvilla eikä maalauksilla, joten neuroverkko menestyi luokitustehtävässä olosuhteisiin nähden jopa yllättävän hyvin.

## Tekoälyn on aina lopulta voitettava ihminen

Voiko tällaista tilastollisten mallien perusteella tapahtuvaa laskemista ja ennustamista pitää osoituksena siitä, että koneella olisi kyky tunnistaa esineitä enemmän tai vähemmän tarkasti? Neuroverkko on oppinut myös pelaamaan vaikkapa Super Mario Brosin itsenäisesti läpi.<sup>20</sup> Jos sama tehtäisiin symbolisella tekoälyllä, ihmisen täytyisi opettaa koneelle eskplisiittisiä IF–THEN-sääntöjä, joita sen tulisi noudattaa *ad hoc* kussakin pelin vaiheessa. Käytännössä oppiva neuroverkkokin kuitenkin edelleen vain laskee ison määrän iteraatioita läpi, minkä jälkeen se pystyy arvioimaan, mitä sen kannattaisi kulloinkin tulostaa (*output*), kun se kohtaa tietynlaisen syötteen (*input*). Tämä tuloste sitten ohjaa pelihahmoa ruudulla tai kenties myös oikeaa autoa tiellä tai aseistettua dronea taistelulentällä.

Riippuu älykkyydelle annetusta määritelmästä ja filosofisesta paradigmasta, tuleeko tällaista toimintaa pitää älykkäänä vai ei. G. W. F. Hegel arvosteli Leibnizin kombinatorista kalkkyliä *Logiikan tieteesä* (Wissenschaft der Logik, 1816) väittäen, että se tekee ”järjellisestä kuollutta ja käsitteetöntä”<sup>21</sup>. Toisaalta konnektivismin kannattaja voisi vastata Hegelille, että nyky-ymmärryksen mukaan nimenomaan orgaaniset eliöt toimivat neuroverkkojen tavoin. Uusmaterialistifilosofi Manuel DeLanda on esittänyt teoksessaan *Philosophy and Simulation. The Emergence of Synthetic Reason* (2011) mielenkiintoisen vertailun alkeellisten eliöiden ja keinotekoisien neuroverkkojen välillä. DeLandan mukaan jopa yksinkertaisilla eliöillä on sisäinen malli ympäristöstään, mutta tämä malli on hyvin karkea eikä siihen sisälly minkäänlaisia symbolisia mielteitä. Monisoluisilla eliöillä kuten meduusalla on yksinkertainen hermosto, jonka avulla ne voivat oppia säätelemään vastettaan haitallisiin tai haittomiin ärsykkeisiin. Hyönteisten hermosto on vielä

tätä monimutkaisempi niin, että yksittäisistä hermoista alkaa muodostua tiettyihin tehtäviin erikoistuneita laajempia kokonaisuuksia. DeLandan mukaan juuri keinotekoiset neuroverkot simuloivat uskottavalla tavalla tällaista alkeellista älykkyyttä, jota tapaamme hyönteisillä. DeLanda käy kirjassaan läpi useita keinotekoisilla neuroverkoilla tehtyjä kokeita, jotka synnyttivät emergentisti luonnossa tapaamiamme vastaavia tuloksia. DeLanda kuitenkin korostaa, että keinotekoiset neuroverkot eivät muistuta fysiologisesti oikeaa hermostoa, vaikka niiden toimintaperiaate mallintaa sitä funktionaalisesti.<sup>22</sup>

Turingin koneen käsitteen keksinyt Alan Turing osallistui myös tekoälyn mahdollisuutta koskevaan keskusteluun. Turingin mukaan 'älykkyyden' käsite on liian epäselvä, jotta koneiden älykkyyttä voitaisiin mitata yksiselitteisesti. Niin sanotun Turingin testin tunnetussa käytännön kokeessa tietokone ja ihminen keskustelivat ihmishavainnoijan kanssa. Mikäli ihminen ei kykene erottamaan päätteen kautta keskusteltavaa konetta ihmisestä, tietokonetta on pidettävä älykkäänä. Alkuperäisessä artikkelissaan Turing tosin esittää matkimispelin (*imitation game*) koekasetelman hiukan toisella tavalla. Turing ajattelee matkimispelillä koetta, johon osallistuu mies, nainen ja kuulustelija. Miehen tehtävän on yrittää vakuuttaa kuulustelija siitä, että hän on nainen. Turing kehottaa meitä nyt kuvittelemaan, että miehen tilalle vaihdetaan tietokone, joka yrittää vakuuttaa kuulustelijalle olevansa nainen.<sup>23</sup> Toinen tärkeä tekoälyn mahdollisuuteen liittyvä ajatuskoe on John Searlen "kiinalainen huone". Searle kehottaa meitä ajattelemaan ihmistä, joka istuu suljetussa huoneessa ja ottaa huoneen ulkopuolelta vastaan kiinankielisiä viestejä. Tällä ihmisellä on huoneessaan tarvittavat ohjeet siitä, kuinka rakentaa luonteva kiinankielinen vastine viesteihin, vaikka hän ei itse ymmärrä lainkaan kiinaa. Ulkopuolisille viestit kuitenkin vaikuttavat siltä, kuin huoneessa oleva henkilö osaisi kiinaa, sillä ohjeet ovat tarpeeksi yksityiskohtaiset tekemään vastauksista luontevaa kiinaa. Searle pyrkii tällä analogialla osoittamaan, että sääntöjä seuraava kone ei tosiasiaa ymmärrä viestien merkityksiä. Vaikka tekoäly kykenisi siis vaikuttamaan ulospäin älykkäältä ja läpäisemään niin sanotun Turingin testin, Searlen mukaan se ei silti ole tosiasiaa älykäs.<sup>24</sup>

Toisaalta niin sanotun heikon tekoälyn suoritusten arviointiin liittyy usein *AI effect*, tekoälytutkimuksen saavutusten vähättely. Paradoksaalisesti vaadimme tekoälyltä aina täydellistä suoritusta, jotta sen voi katsoa osaavan jonkin taidon. Jos Teslan autopilotti kolaroi, sen ei katsota hallitsevan auton ohjausta. Mutta myös ihmiskuljettajille voi sattua onnettomuuksia. Useampikin pelkkiä luvallisia shakkisiirtoja tekevä ohjelma oli jo kehitetty heideggeriaani Hubert L. Dreyfusin kirjoittaessa kuuluisaa teostaan *What Computers Can't Do* (1972). Dreyfus tuntuu kuitenkin implisiittisesti oletavan, että mikäli tekoäly pelaa shakkia huomattavasti nopeammin kuin ihminen, sen ei voi katsoa osaavan shakkia lainkaan<sup>25</sup>. Kun Deep Blue sitten lopulta voitti taitavimmankin ihmispelaajan, maalia siirrettiin ja tekoälyn kriitikot totesivat sen voittaneen *brute*

*force* -menetelmällä eli oikeita siirtoja sitkeästi yrityksen ja erehdyksen kautta hakemalla, mikä ei olisi osoitus todellisesta älykkyydestä.<sup>26</sup> Itse asiassa jo Claude Shannon osoitti matemaattisesti, ettei shakkia pelaava kone voi mitenkään kokeilla mekaanisesti pelin kaikkia siirtoja, sillä pelattavissa olevia shakkipelejä on noin  $10^{120}$  ja shakkia pelataan vieläpä aikaa vastaan<sup>27</sup>. Deep Bluen heuristiikka oli sikäli älykkäästi suunniteltu, että se osasi suunnata rajallisen laskentakapasiteettinsa ja pelaikansa juuri kussakin tilanteessa olennaisten siirtojen puntarointiin.

Johtaako *AI effect* lopulta faustiseen tilanteeseen, jossa tekoälyn on ylitettävä ihmisen kyvyt kaikilla aloilla vain todistaakseen, että se ylipäänsä hallitsee ne? Dreyfus käyttää autolla ajamista ja suuntavilkun kaltaisten merkkien ymmärtämistä esimerkkinä siitä, mitä on heideggerilainen maailmassaoleminen, johon koneet eivät voi lähtökohtaisesti kyetä<sup>28</sup>. Dreyfusin esimerkki on sikäli kiintoisa, että nimenomaan autojen automaattisia ohjausjärjestelmiä kehitetään tällä hetkellä raivokkaasti. Googlen tytäryhtiön Waymon sekä Uberin testaamat itseohjautuvat autot hermostuttavat ja pelottavat monia ihmiskuljettajia, vaikka ne osaisivatkin ottaa toisten kuljettajien käyttämän suuntavilkun huomioon. Vaikka autoissa istuu testejä valvova ihmismatkustaja, ne ovat joutuneet jo monenlaisten väkivaltaisten hyökkäysten ja vandalismin kohteeksi. Taustalla on osin pelko siitä, että robottiautot tekevät ihmiskuljettajat kokonaan tarpeettomiksi.<sup>29</sup> Mikäli koneajaminen todella lunastaa lupauksensa, on mahdollista, että ihmisten ajamia takseja ei enää kohta tunneta tai ihmiskuljettajilta kielletään tulevaisuudessa jopa kokonaan pääsy liikenteeseen. Vähintäänkin pilkulleen sääntöjen ja nopeusrajoitusten mukaan ajavat robottiautot luultavasti ärsyttävät joitakin ihmiskuljettajia. *AI effectin* vuoksi suostumme myöntämään koneiden osaavan ajaa autoa vasta niiden tehdessä sen paremmin kuin ihmiset.

Shakki on suljettu ja formaali merkkijärjestelmä, jota koneen on helppo hallita. Auton ajaminen taas tapahtuu avoimessa maailmassa, jossa koneen on opittava reagoimaan hyvin epätodennäköisiin tilanteisiin. Itsestään ajava auto ei kuitenkaan edusta vahvaa tekoälyä eikä sellaista sen toteuttamiseksi edes tarvita. Vaikka usein kuulee puhuttavan tekoälyn etiikasta ja pohditaan, millaisia valintoja robottiauton tulisi tehdä onnettomuustilanteessa, robottiauton ajo-ohjelmisto seuraa edelleen ihmisen laatimaa algoritmia tai datasta rakennettua tilastollista mallia. Kyky tehdä todennäköisyyksiin perustuvia ennustuksia ja valintoja ennalta annettujen vaihtoehtojen välillä ei tarkoita vapautta. Vaikka itseohjautuvan auton kapea-alainen tekoäly on oppinut tulkitsemaan toisten kuljettajien antamia suuntamerkkejä oikein, se ei silti ole Heideggerin tarkoittamassa mielessä olemukseltaan avoin eksistentiaalinen olento. Itseohjautuva auto ei voi yhtäkkiä päättää lopettaa sille määrätyn taksikuskin työtä ja todeta, että sen tarkoituksena on pikemminkin rallin jokamiesluokassa kilpaileminen. Vain vahva tekoäly voisi tehdä tällaisia vapaita valintoja.

Neuroverkkojen ja koneoppimisen alueella työskentelevät tutkijat eivät tavallisesti edes pyri itsetietoisien älyllisen olennon luomiseen. Kuten todettua, ”tekoälyn” kehitys on ollut 2000-luvulla kaikkein nopeinta aloilla, joissa on pikemminkin kysymys suuriin datamääriin perustuvien tilastollisten mallien rakentamisesta. Tekoälyn määrittäminen on selvästi sidoksissa myös käsityksemme siitä, mitä ihmisenä oleminen on. Humanistisesta näkökulmasta katsottuna itseään ajavan auton olemus perustuu siihen, että se on ihmisen määrittämä väline; se ei ole itsetarkoituksellinen olento, joka voisi itse asettaa itselleen tarkoituksen. Tästä syystä robotti-

autolla ei ole myöskään mitään työntekijöille kuuluvia oikeuksia eikä eettisiä velvollisuuksia. Toisaalta jotkut filosofiset perinteet epäilevät myös sitä, onko ihmisenkään vapaa ja itsetarkoituksellinen olento vai sittenkin vain kausaalilakien ehdollistama luontokappale tai pitkälle kehittynyt eläin. Jos ihmistä itseään ajatellaan komputaationaalisenä koneena eikä vapaana henkiolentona, ihmisenkaltaisesti ajattelevan tekoälyn luomisen kynnyks vastavasti madaltuu. Entäpä jos myös inhimillisessä älyssä on pohjimmitaan kysymys kognitiivisten mallien rakentamisesta suureen datamassaan (kokemukseen ja havaintoihin) perustuen?

## Viitteet & Kirjallisuus

- Catherine Clifford, Elon Musk: 'Mark my words – A.I. is far more dangerous than nukes'. *CNBC* 2018. Verkossa: [www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html](http://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html) (haettu 11.6.2019).
- Vahvan ja heikon tekoälyn erottelusta ks. esim. John Searle, *Minds, Brains, and Programs. Behavioral and Brain Sciences*. Vol. 3, No. 3, 1980, 417–424.
- Homeros, *Ilias*, 18:414; Pamela McCorduck, *Machines Who Think*. A. K. Peters, Natick 2004. Omaa (auto) tahtoa tai ajattelua (*ménitis*) tarkoittava termi αὐτόματος esiintyy Homeroksen kuvauksessa Hefaistoksen itseään liikkuttavista mekaanisista kolmijaloista. Homeros, *Ilias*, 18:376; Henry George Liddell & Robert Scott, *A Greek-English Lexicon* (1843). Uus. p. Oxford, Clarendon Press 1940. Verkossa: [www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0057:entry=au/to/matos](http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0057:entry=au/to/matos) (haettu 17.6.2019).
- Hubert L. Dreyfus olettaa kaikenlaisen sellaisen ajattelun olevan komputationalismia, mikä pyrkii löytämään ihmis toiminnan taustalta erehtymättömiä sääntöjä. Dreyfus aloittaa siksi komputationalismin historian jo Platonin *Euthyfron*-dialogista. Hubert L. Dreyfus, *What Computers Can't Do. A Critique of Artificial Reason*. Harper & Row, New York 1972, xv.
- Gottfried Wilhelm Leibniz, Kaksittaisuista. *Filosofisia tutkielmia*. Toim. Tuomo Aho ja Markku Roinila. Suom. Tuomo Aho. Gaudeamus, Helsinki 2011, 281–287.
- Gottfried Wilhelm Leibniz, Das dyadische Zahlensystem (De Progressione Dyadica, 1679). *Herrn von Leibniz' Rechnung mit Null und Eins*. Käänt. Isä Franz X. Wernz. Siemens Aktiengesellschaft, Berlin & München 1966, 42–47 (46).
- Michael R. Williams, *A History of Computing Technology*. Prentice Hall, Englewood Cliffs 1985, 123–140.
- Timo Honkela, *Rauhankone. Tekoälytutkijan testamentti*. Gaudeamus, Helsinki 2017.
- Norbert Wiener, *Cybernetics or the Control and Communication in the Animal and the Machine* (1961). The M.I.T. Press, Cambridge, Mass., 1985, 12.
- Williams 1985, 182–191.
- James Essinger, *Adan algoritmi. Kuinka lordi Byronin tytär Ada Lovelace käynnisti digiajan* (Ada's Algorithm, 2012). Suom. Tapani Kilpeläinen. Vastapaino, Tampere 2016.
- Jennifer S. Light, When Computers Were Women. *Technology and Culture*. Vol. 40. No. 3. 1999, 455–483.
- Vaikka sanaa ”informaatio” käytetään nykyisin synonyyminä kaikelle tiedolle, Shannon tarkoitti informaatiolla viestin yllättävyyden mittaamista. 6-sivuisen nopan heittäminen tuottaa enemmän informaatiota kuin kaksipuoleinen kolikko, koska todennäköisyys saada kruuna tai klaava on jopa 50 prosenttia, kun taas nopan heittäminen tuottaa kuusi eri mahdollisuutta eli enemmän informaatiota. Informaatio ei siis liity Shannonilla viestin sisältöön vaan sen ennustettavuuden ja toisteisuuden mittaamiseen.
- Kybernetikasta ja informaatioteoriasta, ks. esim. N. Katherine Hayles, *How We Became Posthuman. Virtual Bodies in Cybernetics, Literature, and Informatics*. University of Chicago Press, Chicago 1999; Ronald R. Kline, *The Cybernetics Moment, or, Why we Call our Age the Information Age*. John Hopkins University Press, Baltimore 2015.
- New Navy Device Learns by Doing. Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser. *The New York Times*, 8.7.1958; Mikel Olarazan, A Sociological Study of the Official History of the Perceptrons Controversy. *Social Studies of Science*. Vol. 26. No. 3. 1996, 611–659 (621).
- Aakkoset osaava kone. *Länsi-Savo* 6.10.1960. Verkossa: [digi.kansalliskirjasto.fi/sanomalehti/binding/1027268?page=7](http://digi.kansalliskirjasto.fi/sanomalehti/binding/1027268?page=7) (haettu 12.6.2019).
- Olarazan 1996.
- David Rumelhart, Geoffrey Hinton ja Ronald Williams, Learning Representations by Back-propagating Errors. *Nature*. Vol. 323, 1986, 533–536; David Rumelhart ja James McClelland, *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*. MIT Press, Cambridge 1987.
- McCorduck 2004, 456.
- Luigi/O, YouTube. Verkossa: [www.youtube.com/watch?v=Xsf4AuejeKA](http://www.youtube.com/watch?v=Xsf4AuejeKA), (haettu 17.6.2019).
- G. W. F. Hegel, *Wissenschaft der Logik* (1816). *Hauptwerke Band 4*. Meiner, Hamburg 1999, 108. Hegelin kritiikistä komputationalismia kohtaan ks. Paul Redding, *Mathematics, Computation, Language and Poetry. The Novalis Paradox*. Teoksessa *The Relevance of Romanticism. Essays on German Romantic Philosophy*. Toim. Dalia Nassar. Oxford University Press, New York 2014, 221–238.
- Manuel DeLanda, *Philosophy and Simulation. The Emergence of Synthetic Reason*. Continuum, London 2011, 80–110.
- Alan Turing, Computing machinery and intelligence. *Mind*. Vol. 59, No. 236, 1950, 433–460.
- Searle 1980.
- Dreyfus 1972, 12–19.
- McCorduck 2004, 204, 433.
- Claude Shannon, Programming a Computer for Playing Chess. *Philosophical Magazine*. Vol. 41, No. 314, 1950, 256–275.
- Dreyfus 1972, 173.
- Simon Romero, Wielding Rocks and Knives, Arizonans Attack Self-Driving Cars. *The New York Times* 31.12.2018. Verkossa: [www.nytimes.com/2018/12/31/us/waymo-self-driving-cars-arizona-attacks.html](http://www.nytimes.com/2018/12/31/us/waymo-self-driving-cars-arizona-attacks.html) (haettu 17.6.2019).