Opinnäytetyö (AMK)

Tieto- ja viestintätekniikka

2021

Valtteri Vainio

KOHTI SUPERTEKOÄLYÄ

TURKU AMK
TURKU UNIVERSITY OF
APPLIED SCIENCES

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Tieto- ja viestintätekniikka

2021 | 35 sivua

Valtteri Vainio

KOHTI SUPERTEKOÄLYÄ

Tekoäly on noussut lyhyessä ajassa osaksi ihmisten jokapäiväistä elämää ja sitä pidetään uutena sähkönä, joka on mullistamassa maailmaa. Kehitykseen panostetaan tällä hetkellä runsaasti, uudet teknologiat ja läpimurrot eri tieteenaloilla nopeuttavat tekoälyn kasvua entisestään. Jatkuvasti nopeutuva teknologinen kehitys antaa ennustuksen sille, että ihmisen yleisen älykkyyden kaltainen tekoäly ja sitä seuraava supertekoäly tullaan saavuttamaan jossain kohtaa ihmiskunnan historiaa. Tälläisen tekoälyn mukanaan tuomat ongelmat ja eettisyys herättävät kuitenkin runsasta keskustelua tekoälytutkijoiden keskuudessa.

Supertekoälyn älykkyys ylittää viisaimpien ja lahjakkaimpien ihmisten älyn moninkertaisesti. Se voi syntyä niin kutsutun älykkään räjähdyksen tai sitä vastaavan teknologisen singulariteetin yhteydessä. Tekoälytutkijat ovat ennustaneet synnyn tapahtuvan sen jälkeen, kun tekoäly saavuttaa ihmisen yleisen älykkyyden kaltaisen tason, joka on aktiivisesti oppiva, tietoa sekä kokemuksia keräävä tekoäly.

Opinnäytetyön tarkoituksena on tutkia tekoälyn nopeaa kehitystä kohti ennustettua supertekoälyä ja siihen vaadittavia teknologisia valmiuksia, sekä antaa lukijalle yleiskäsitys tekoälystä.

Milloin tekoäly ylittää ihmisen yleisen älykkyyden ja mitä siitä voi seurata? Entä miten voidaan luoda turvallinen supertekoäly?

ASIASANAT:

tekoäly, supertekoäly, neuroverkot, simulointi, emulointi

BACHELOR´S / MASTER'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Degree programme

2021 | 35 pages

Valtteri Vainio

TOWARDS SUPERINTELLIGENCE

Artificial intelligence has become part of our daily lives in short time. It is being held as the new electricity that is revolutionizing the world. There is currently a lot of investment in development, and new technologies and breakthroughs in various fields of science will further accelerate the growth of AI. The accelerating technological advancement gives a prediction that AI will develops into Artificial general intelligence and the following Superintelligence at some point in human history. However, the problems and ethics posed by such artificial intelligence provoke much debate among AI researchers. The intelligence of superintelligence exceeds the intelligence of the wisest and most talented people many times over. It can arise in the context of a so called smart explosion or a corresponding technological singularity. AI researchers have predicted that the birth of superintelligence will occur after artificial intelligence reaches a state level similar to general human intelligence, which is actively learning AI that gathers knowledge and experience. The purpose of the thesis is to study the rapid development of artificial intelligence towards predicted superintelligence and the technological capabilities required for it, and to give the reader an overview of artificial intelligence.

When does artificial intelligence exceed general human intelligence and what can follow from it? And how can we try to build safe superintelligence?

KEYWORDS:

artificial intelligence, superintelligence, neural network, simulate, emulate

SISÄLTÖ

KÄYTETYT LYHENTEET TAI SANASTO **6**

1 JOHDANTO **9**

2 TEKOÄLY **10**

2.1 Tekoälyn juuret 11

2.2 Tekoälyn kolme luokkaa 12

2.3 Syväoppiminen 14

3 SUPERTEKOÄLY – MITEN JA MILLOIN? **16**

3.1 Ihmisaivojen emulointi tietokoneelle 17

3.2 Teknologian kiihtyvä kehitys 22

3.3 Supertekoälyn ajankohdan arvioita 26

4 TEKOÄLYN UHKA **30**

4.1 Teknologinen singulariteetti 31

4.2 Tekoälyn hallintaongelma 31

5 JOHTOPÄÄTÖS **33**

LÄHTEET **34**

KUVAT

Kuva 1. Tekoälyluokkien kehityksen havainnollistaminen. 14
Kuva 2. Keinotekoinen neuroverkko jäljittelee ihmisen aivoja. Perustuu lähteeseen (jyo, 2021). 15
Kuva 3. Valmiudet joita tarvitaan koko aivojen emulointiin. Perustuu lähteeseen (Bostrom 2014, 50). 18
Kuva 4. Neuroanatomian 3D-rekonstruktio elektronimikroskooppikuvista. (Bostrom, 2014, 49) 19
Kuva 5. Ihmisen aivojen emuloinnin etenemissuunnitelma. Kaavio syötteestä, toiminnoista ja virstanpylväistä. Perustuu lähteeseen (Bostrom, 2014, 54) 21
Kuva 6. Tietokoneiden ja laskentalaitteiden suorituskyvyn kasvu logaritmisella asteikolla. Perustuu lähteeseen (Kurzweil, Hod Lipson, 2018) 24
Kuva 7. Lineaarinen kaavio tietokoneiden ja laskentalaitteiden kehityksestä. Perustuu lähteeseen (Kurzweil, Lipson, 2018) 25
Kuva 8. Transistorien määrä mikroprosessoria kohti. Perustuu lähteeseen (Martin, 2019). 26

TAULUKOT

Taulukko 1. Haastattelututkimus supertekoälyn ajankohdasta. Perustuu lähteeseen (Bostrom, 2014, 35-38). 28
Taulukko 2. Edistävät tutkimussuuntaukset kohti supertekoälyä. Perustuu lähteeseen (Müller & Bostrom, 2014) 28

KÄYTETYT LYHENTEET

FLOPS	engl. floating point operations per second, liukulukuoperaatiota per sekunti
PB	engl. petabyte, petatavu, 10^{15} tavua
tavu	digitaalisen informaation yksikkö

1 JOHDANTO

Monet tutkijat, kuten Stephen Hawking, Elon Musk ja Vernor Vinge ovat antaneet varoittavia lausuntoja tekoälyn piilevästä vaarasta. Tieteellisissä piireissä yhä useampi tekoälytutkija on yhtä mieltä siitä, että ihmiset luovat lopulta tekoälyn, joka ylittää ihmiskunnan yleisessä älykkyydessä moninkertaisesti. (Luckerson 2014). Eksponentiaalisesti nopeutuva teknologinen kehitys antaa ennustuksen sille, että ihmisen yleisen älykkyyden kaltainen tekoäly ja sitä seuraava supertekoäly tai singulariteetti tullaan saavuttamaan jossain kohtaa ihmiskunnan historiassa. Tällä hetkellä ihmiset hallitsevat älyllisesti maailman kaikista lajeista. Tekoäly voi kuitenkin muuttaa tämän asetelman, se voi muodostua liian voimakkaaksi ja vaikeasti hallittavaksi jos tekoälyn hallintaongelmaan ei kyetä varautumaan ajoissa. (Bostrom 2014, luku 2)

Tekoälyä on tutkittu jo 1940-luvulta lähtien, ja ala on nähnyt monta nousu -ja laskukautta suurten odotusten ja niihin liittyvien pettymysten takia. Nykypäivänä teknologian mahdollistettua kehityksen uskotaan tekoälyllä olevan tulevaisuudessa suurempi vaikutus maailmaan kuin joillakin historian yleisimmillä innovaatioilla. Tekoäly on tällä hetkellä yksi ihmiskunnan tärkeimmistä tieteenaloista. (Clifford 2018)

Maailman johtaviin tekoälytutkijoihin kuuluvan Nick Bostromin mukaan ihmisen aivojen emulointi ja tietokoneiden suuri laskentateho ovat realistisimpia suuntia kohti ihmisen kaltaista tekoälyä. (Bostrom 2014, luku 2) Jotta tekoäly voisi saavuttaa ihmisen kaltaisen yleisen älykkyyden, vaaditaan lisää tutkimusta neurotieteissä. Lisäksi myös tietokoneiden laskentatehon eksponentiaalisen kasvukäyrän tulisi jatkaa kasvuaan.

Opinnäytetyön tarkoituksena on tehdä kirjallisuustutkimus tekoälyn eksponentiaalisesta kehityksestä kohti supertekoälyä ja samalla arvioida sen toteutumisen edellyttämiä teknologisia vaatimuksia. Aluksi selvitetään yleiskuva tekoälystä. Sen jälkeen esitellään supertekoäly ja sen mahdollistavat tekijät, sekä arvioidaan sen synnyn ajankohtaa. Lopuksi käsitellään tekoälyn uhkia ja siihen liittyvää hallintaongelmaa.

Opinnäytetyön teoria pohjautuu suurimmalta osin Nick Bostromin Superintelligence; Paths, Dangers, Strategies -kirjaan, joka käsittelee supertekoälyä ja tekoälyn tuomia mahdollisuuksia sekä uhkia. Teorian pohjana on käytetty myös muita tieteellisiä julkaisuja ja verkkoartikkeleita.

2 TEKOÄLY

Tekoäly on laaja tietojenkäsittelyn tieteenala ja sen synnyn on mahdollistanut monilla tieteenaloilla tapahtunut kehitys viime vuosisatojen aikana. Se on niihin asioihin liittyvää tutkimusta, joilla tietokoneet saadaan älykkäiksi. Tekoälyssä tutkitaan myös asioita, joiden toteutus on vielä tuntematon. Se tarkoittaa, että tavallisen ja älykkään koneen välinen raja alati muuttuva. (Pietikäinen & Silvén 2019, 17)

Tekoäly on yleisesti ottaen älykkäiden tietokoneohjelmistojen tai koneiden kykyä tehdä mitä tahansa ihmisen kaltaisia älykkäitä toimintoja. Termillä tekoäly viitataan yleensä tietokoneohjelmistoon tai koneeseen, joka yrittää matkia ihmisen tajuntaa ja käyttäytymistä sille annetuissa tehtävissä, eli niiden kykyä pystyä ajattelemaan ja toimimaan rationaalisesti. Tekoälylle annetaan tavoite, mutta se luo itse ratkaisut, joilla tavoite saavutetaan. Se voi oppia ja toimia itsenäisesti syötetyn datan sekä kokemusten perusteella. Tekoälyohjelmat voivat olla puhtaasti ohjelmistopohjaisia ja toimia virtuaalimaailmassa, kuten hakukoneet ja kasvojentunnistusohjelmat. Ohjelmat voidaan myös sulauttaa laitteistoihin, jolloin ne voivat toimia esimerkiksi kehittyneinä robotteina reaalimaailmassa. (IBM 2020)

Tekoäly voidaan jakaa kolmeen eri luokkaan älykkyyden perusteella: heikko, vahva ja supertekoäly. Tekoälyn tarkempi määritteleminen on haastavaa. Yksi merkittävä syy siihen on se että, ei ole vieläkään yhtä vahvaa määritelmää tai käsitystä älykkyydestä yleisesti. Älykkyys perustuu usein kontekstiin, eikä se itsessään ole yksi dimensio tai asteikko. (Faggella 2018) Jo aikaisemmin mainittu matemaatikko Alan Turing esitteli vuonna 1950 ”Imitaatiopelin”. Se oli laajalti referoitu määritelmä koneen älykkyyden testaamiseksi. Testissä seinän toisella puolella on älykkyyden testaaja ja toisella puolen ihminen tai tietokone. Jos testaaja päätyy lopputulokseen, että hän on keskustellut ihmisen kanssa, vaikka todellisuudessa onkin keskustellut tietokoneen kanssa, on Turingin määritelmien mukaan koneella älykkyyttä. (Pietikäinen & Silvén 2019, 17-18)

Jos tietokonetta erehdytään luulemaan ihmiseksi yli 30 prosenttia ajasta viiden minuutin keskustelujen aikana, se läpäisee testin. Historiallinen virstanpylväs saavutettiin vuonna 2014 Readingin yliopiston järjestämässä tapahtumassa. Eugene Goostman niminen tekoälyohjelma läpäisi testin 13-vuotiaaksi pojaksi naamioituneena. 30 tuomarin joukosta Eugene onnistui vakuuttamaan 33 prosenttia. (University Of Reading 2014)

2.1 Tekoälyn juuret

Tekoälyn juuria ja älykkäiden tietokoneiden konsepteja on löydettävissä jo antiikin Kreikan historiallisesta perinnöstä. Myytit ja tarinat taitavista älykkyyden ja tietoisuuden omaavista olennoista kiehtoivat kansaa ja inspiroivat filosofeja sekä taiteilijoita. Filosofit kylvivät nykyisen tekoälyn siemenet yrittäessään kuvailla ihmisen ajatteluprosessia symbolien mekaanisena manipulointina. (Mayor 2018, 1) ”Tekoäly alkoi muinaisella toiveella luoda jumalia”. Tämä muinainen toive on pysynyt nykypäivänä osittain yhä relevanttina. (Morgan 2019)

Ensimmäisten tietokoneiden kehittyessä 1940-luvulla alkoivat keskustelut eri alojen tutkijoiden kesken aivojen simuloinnin mahdollisuuksista tietokoneella. Pohjustuksen idealle kehittivät nykyaikaisen tietotekniikan pioneerit: Alan Turing teoreettisellä työllään ja John Von Neumann, siihen perustuvalla tietokonearkkitehtuurilla. Varhaisinta tutkimusta älykkäistä koneista iskosti neurologian tutkimus, joka oli todistanut aivojen olevan neuroneista koostuva sähköinen ja oppiva verkko, jossa jokainen neuroni on binääritilassa. Turingin laskettavuuden teoria osoitti myös minkä tahansa laskennan olevan mahdollista suorittaa digitaalisesti. (Pietikäinen & Silvén 2019, 24-27)

Tekoälyn perustan luonnissa merkkihenkilönä oli taas Turing. Vuonna 1950 julkaistussa teoksessaan Computing machinery and intelligence Turing toi esiin monia tulevien vuosien tekoälytutkimuksen peruskysymyksiä, kuten teoreemojen todistamisen, luonnollisen kielen ymmärtämisen sekä kääntämisen, pelien pelaamisen ja koodien murtamisen. Hän esitti myös laajalti referoidun määritelmän koneen älykkyyden testaamiseksi, Turingin testin. (Pietikäinen & Silvén 2019, 24-27)

Varsinaisena tekoälyn tieteenalan alkuna pidetään vuotta 1956. Yhdysvalloissa järjestettyyn Dartmoud College -konferenssiin osallistui useita tulevien vuosien tekoälytutkimuksen pioneereja. Kognitiotieteilijä John McCarty otti kokousta laatiessaan käyttöön termin ”Artificial Intelligence” (AI, tekoäly). Kokouksen tavoitteena oli selvittää, miten voitaisiin luoda kone, joka pystyisi ajattelemaan kuin ihminen. Koneelta vaadittaisiin ongelmanratkaisutaitoja, kyky abstraktiin ajatteluun ja itsensä kehittämiseen.

Samana vuonna tutkijat Allen Newell, Herbert Simon ja Cliff Shaw julkaisivat ensimmäisen tekoälynä pidetyn tietokoneohjelmiston nimeltä Logic Theorist. Tietokoneohjelma pystyi todistamaan osan Principia Mathematica (1910–1913) teoksen teoreemoista. (Pietikäinen & Silvén 2019, 25-26) Tekoälyn potentiaali huomattiin ja uusien tutkimuskeskusten rahoitus aloitettiin. Newell ja Simon toimivat uranuurtajina ajaessaan tekoälyä tieteenalaksi joka voisi muuttaa maailmaa. (Ray 2018)

Nykypäivän tekoälyaalto on uusien tekniikoiden nopeaa käyttöönottoa uusissa sovelluksissa. Vaikka hermoverkot ovat olleet tutkijoiden tiedossa jo vuosikymmenien ajan, niiden potentiaalia ei ole pystytty vielä käyttämään täysin. Viimeisen kymmenen vuoden aikana tietokoneiden nopeus on yltänyt pisteeseen, joka mahdollistaa syvien hermoverkkojen käytön kuvien tunnistuksessa, puheen ja tekstin kääntämisessä, sekä muissa sovelluksissa. Vuonna 2012 tapahtui useita merkittäviä läpimurtoja syväoppivien hermoverkkojen suorituskyvyn ja tarkkuuden suhteen. (House 2019)

Vuonna 2016 AlphaGo -niminen syviin neuroverkkoihin perustuva tekoälyohjelma voitti hallitsevan maailmanmestarin Go-lautapelissä, joka vaatii kaikessa monimuotoisuudessaan enemmän strategista silmää kuin esimerkiksi shakki. AlphaGo voitti neljä ottelua viidestä. Voitto on merkittävä, koska mahdollisia liikkeitä pelin edetessä on jo neljän siirron jälkeen yli 14 biljoonaa ($1{,}4 \times 10^{13}$). Myöhemmin Google ilmoitti ostaneensa koko yrityksen 400 miljoonalla dollarilla. (IBM 2020)

2.2 Tekoälyn kolme luokkaa

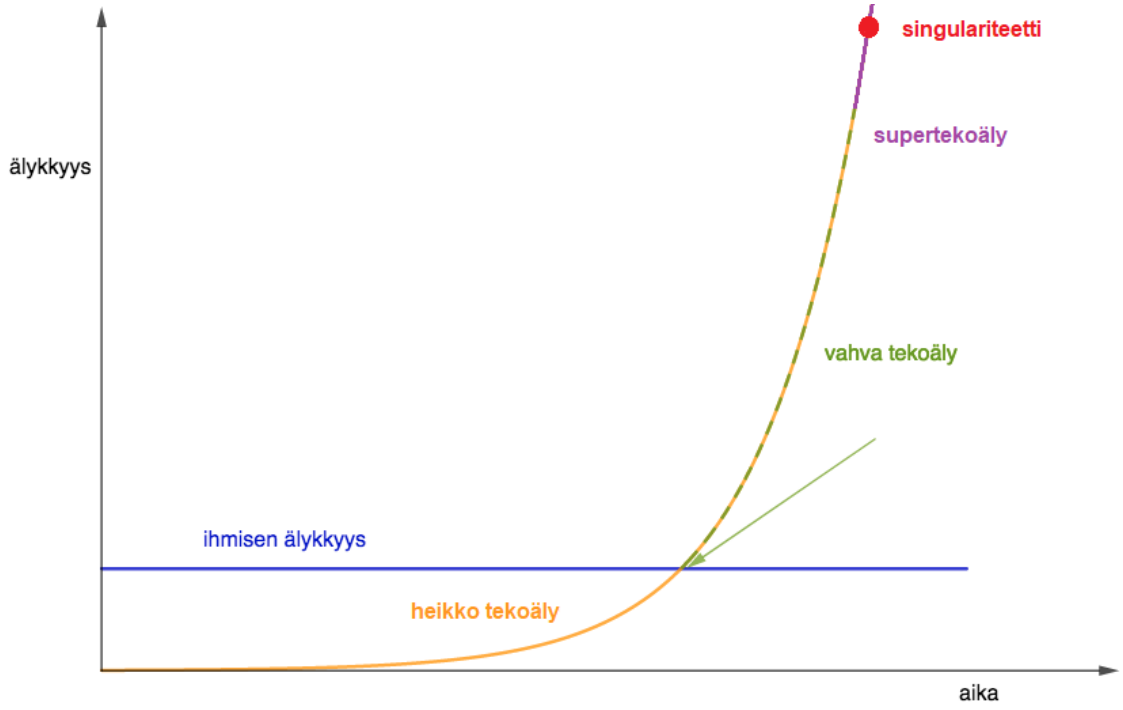
Tekoäly voidaan jakaa kolmeen eri luokkaan sen älykkyyden perusteella: heikko tekoäly (engl. Artificial Narrow Intelligence), vahva tekoäly (engl. Artificial General Intelligence) tai supertekoäly (engl. Artificial Super Intelligence). (Pietikäinen & Silvén 2019, 23)

Heikko tekoäly on toistaiseksi ainut tekoälyn taso, joka on onnistuneesti pystytty saavuttamaan. Kaikki nykyiset tekoälyn sovellukset ovat heikkoa tekoälyä. Ne pystyvät ylittämään ihmisen suoritustason, vain ennalta ohjelmoitujen tehtävien ratkaisuun valituissa ohjelmissa. Ihmisen tasoista tietoisuutta se ei kuitenkaan kykene saavuttamaan, vaikka pystyykin sitä hetkellisesti simuloimaan. Vaikka nämä tekoälyratkaisut saattavat tuntua älykkäiltä, ne toimivat hyvin rajallisesti, minkä vuoksi tätä luokkaa kutsutaan heikoksi tekoälyksi. Heikko tekoäly ei jäljittele tai toista ihmisen

älykkyyttä, vaan simuloi ihmisen käyttäytymistä sille annettujen suppeiden parametrien ja asiayhteyden perusteella. Esimerkkejä löytyy jokapäiväisestä arkielämästämme kuten hakukoneet, kielenkääntäminen, ihmisten eleiden, puheen- ja kasvojentunnistus valokuvista tai videoista. (Pietikäinen & Silvén 2019, 23)

Vahvasta tekoälystä puhutaan silloin, kun se on jollain tasolla tietoinen itsestään, sekä kykeneväinen toimimaan täysin itsenäisesti jatkuvasti muuttuvassa ympäristössä ilman ihmisen ohjeistusta. Silloin tekoälyohjelmiston älykkyys olisi lähellä ihmisen tasoa ja se pystyisi suoriutumaan mistä tahansa tehtävästä ihmisen tavoin. Tällöin se osaisi myös etsiä, sekä käyttää suunnittelussa ja ongelmatilanteissa hankkimiaan erilaisia taustatietoja apunaan. Vahvan tekoälyn tärkeimpiä ominaisuuksia ovat ihmisen kognitiiviset toiminnot, kuten tiedon käsittely, havainnointi ja vastaanotto, oppiminen, muistaminen ja kommunikointi. Esimerkiksi täysin itsenäisesti työtä suorittavat logistiikka-ajoneuvot vaatisivat ristiriitatilanteissa vahvaa tekoälyä ongelmien käsittelyyn, vaikka alemman luokan päätöksiä, kuten objektien havannointia, voidaan suorittaa heikolla tekoälyllä. Vahvaan tekoälyn ympärille alkaa kerääntyä paljon eettistä kritiikkiä. Esimerkiksi tiettyjen työpaikkojen menetys voi olla huomattava tulevaisuudessa. Nykyistä yhteiskuntaa ei ole suunniteltu toimimaan tekoälyn kanssa.

Supertekoälyn älykkyys ylittää jo viisaimpien ja lahjakkaimpien ihmisten älyn moninkertaisesti. Supertekoäly voi syntyä niin kutsutun älykkään räjähdyksen tai sitä vastaavan teknologisen singulariteetin yhteydessä. Tekoälytutkijat ovat ennustaneet supertekoälyn synnyn tapahtuvan sen jälkeen, kun vahva tekoäly saavuttaa ihmisen älykkyyden kaltaisen tason, joka on aktiivisesti oppiva, tietoa sekä kokemuksia keräävä tekoäly. Teknologinen singulariteetti on hetki, jolloin vahva tekoäly kehittää itsestään vahvempia versioita peruuttomasti ja kiihtyvällä tahdilla. (Pietikäinen & Silvén, 2019, 23) Kuva 1 havainnollistaa eri luokkien kehitystä ajassa ja älykkyydessä. Singulariteetti on lisätty mukaan.



Kuva 1. Tekoälyluokkien kehityksen havainnollistaminen.

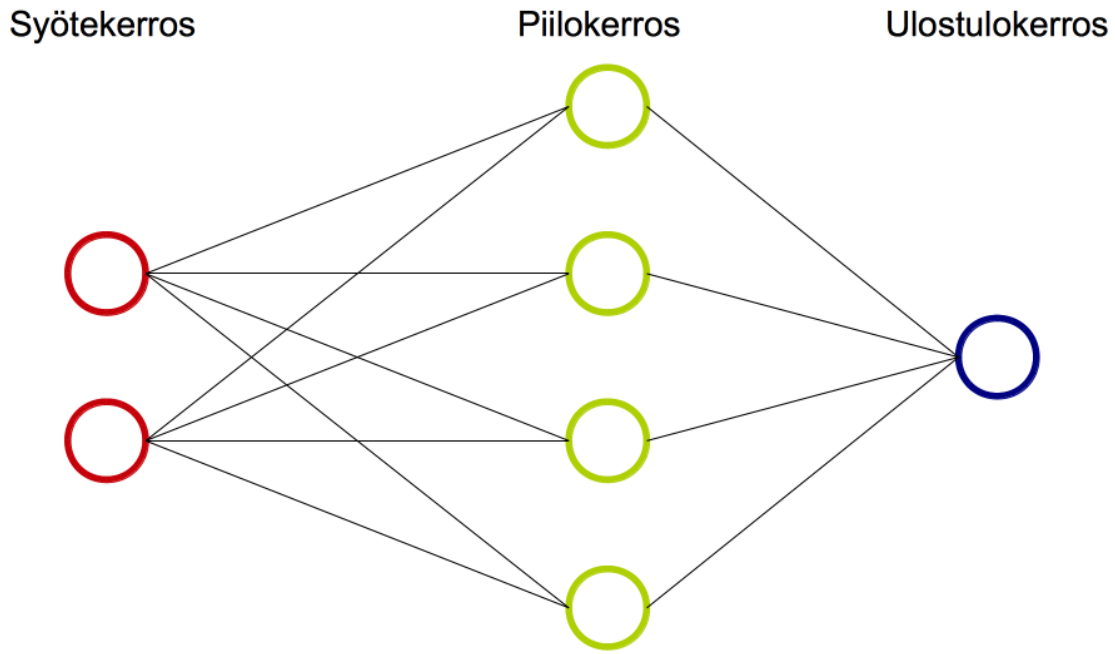
2.3 Syväoppiminen

Suuri osa tämän hetken tekoälysovelluksista perustuvat koneoppimisen ja syväoppimisen läpimurtoihin. Se on tekoälyn osa-alue, joka perustuu ohjelmiston kykyyn oppia itsenäisesti syötetystä datasta ja mahdollisen käyttäjän toiminnan perusteella. Koneoppimisessa pyritään automatisoimaan tiedon tulkintaa ja laajentamaan ohjelmiston havainnointikykyä algoritmien avulla. Joissain tapauksissa ohjelmisto voi myös itse kehittää uusia algoritmejä saavuttaakseen annetut tehtävät. (Jyväskylän Yliopisto 2021)

Syväoppiminen (”Deep learning”) taas on yksi koneoppimisen tekniikoista. Se perustuu ihmisaivojen toimintaa ja oppimista mallintaviin hermoverkkoihin eli neuroverkkoihin. Keinotekoiset neuroverkot mallintavat useiden kerrosten avulla aivojen laskentamenetelmiä. Ne ovat matemaattisia malleja, joiden toiminta jäljittelee ihmisaivojen mekanismeja. Neuroverkkoja koulutetaan suoriutumaan tehtävistä opetusaineiston avulla. Koulutuksessa neuroverkkoja muokataan datan sekä kokemusten avulla, ja ne oppivat suoriutumaan tehtävistään yhä tarkemmin ja

nopeammin. Kuvantunnistuksessa neuroverkolle voidaan esimerkiksi syöttää suuri määrä luokiteltuja valokuvia, jolloin se oppii tunnistamaan halutun objektin. (IBM 2020)

Neuroverkon arkkitehtuurissa neuronit on yleisesti jaettu useisiin kerroksiin Kuva 2. Verkko koostuu syötekerroksesta ja ulostulokerroksesta, sekä niiden välissä olevista piilokerroksista. Kerrokset taas koostuvat neuroneista.



Kuva 2. Keinotekoinen neuroverkko jäljittelee ihmisen aivoja. Perustuu lähteeseen (Jyväskylän Yliopisto 2021).

Yksi neuroni on liian yksinkertainen luotettavien ennusteiden ja päätösten tekemiseen tekoälyohjelmistoissa. Tällöin yhden neuronin tuloste kytketään toisten neuronien syötteeksi ja niin edelleen. Syötekerros käsittää ne neuronit, jotka saavat syötteensä suoraan koko neuroverkon syötteestä, joka voi esimerkiksi olla verkolle esitetty valokuva. Lukuisat piilokerrokset saavat syötteensä muilta neuroneilta, joiden tuloste toimii neuronien syötteinä. Viimeisenä verkossa on ulostulokerros, joka tuottaa koko verkon tulosteen. Kerrosten väliset yhteydet on useimmiten järjestetty siten, että yksittäisen kerroksen neuronit saavat syötteensä edellisen kerroksen neuroneilta ja antavat tulosteensa seuraavalle kerrokselle. Syväoppimisen perusta on siis kerroksellinen rakenne. Mitä useampi kerroson, sitä syvempi verkko on. Syvemmässä verkossa oppiminen vaatii suurempia datamääria ja enemmän laskentatehoa. (Elements of AI 2021)

3 SUPERTEKOÄLY – MITEN JA MILLOIN?

Supertekoäly ylittää viisaimmatkin ihmisaivot moninkertaisesti joka osa-alueella kuten, tieteellinen luovuus, yleinen viisaus ja sosiaaliset taidot. Se voi syntyä älykkään räjähdyksen tai sitä vastaavan teknologisen singulariteetin yhteydessä. Supertekoälyn on ennustettu syntyvän sen jälkeen, kun vahva tekoäly saavuttaa ihmisen älykkyyden kaltaisen tason, joka on aktiivisesti oppiva, tietoa sekä kokemuksia keräävä tekoäly. (Bostrom 2014, luku 3)

Bostromin mukaan yksi tällä hetkellä realistisimmista tavoista saavuttaa supertekoäly on takaisinmallintaa ihmismieli ohjelmistoksi ja ajaa simulaatio ihmisaivoja nopeammalla tietokoneella. Tälläisella ihmisen kaltaisella päättäjällä, joka voi ajatella miljoonia kertoja nopeammin kuin nykyiset ihmiset, olisi hallitseva etu useimmissa päättelytehtävissä. Myös evoluution perustetta on käytetty ihmismäisen tekoälyn toteutettavuudelle, koska se on todistanut jo kerran, että on mahdollista luoda ihmisen kaltainen älykkyys. Tarpeeksi nopeilla tietokoneilla pystyisi ajamaan geneettisiä algoritmejä, jotka ovat verrattavissa biologisen evoluution tuloksiin. Tämä menetelmä tuottaisi siis älykkyyden. Simulaatio vaatisi kuitenkin tietokoneiden nopeuden ja muistikapasiteetin eksponentiaalisen kasvukäyrän jatkuvan nopeammin kuin Mooren laki. (Bostrom 2014, 41)

Tietokonekomponentit ylittävät jo huomattavasti ihmisten suorituskyvyn nopeudessa. Biologiset neuronit toimivat noin 200 Hz:n huippunopeudella, joka on seitsemän suuruusluokkaa pienempi taajuus kuin modernilla mikroprosessorilla, noin 2 GHz. Aivojen hermosolut lähettävät signaaleja aksonien yli korkeintaan 120 m/s, kun taas nykyiset elektroniset prosessoriytimet voivat viestiä optisesti valon nopeudella, $3 \times 10^8\ m/s$. (Nick Bostrom, 2014, 79)

Ihmisen aivojen emulointi tietokoneelle on yksi selkeimmistä tavoista saavuttaa supertekoäly. Täydellinen emulointi (”Whole Brain Emulation”) vaatii kuitenkin vielä kehittyneempiä teknologisia valmiuksia skannauksessa, translaatiossa sekä simuloinnissa. Luvussa 3.1 käsittelen ihmisaivojen emulointia tietokoneelle.

Myös tietokoneiden nopeuden ja muistin eksponentiaalinen kasvun jatkuminen on edellytyksenä supertekoälyn saavuttamiseksi. Luvussa 3.2 käsittelen tietokoneiden

lasketatehon eksponentiaalista kasvua ja Mooren lakia, joka ennustaa transistorien määrän mikropiireillä kaksinkertaistuvan kustannustehokkaasti kahden vuoden välein.

3.1 Ihmisaivojen emulointi tietokoneelle

Ihmisen aivot ovat monimutkaisin rakenne, jonka tiedämme. Aivoissamme on arvioilta noin 100 miljardia hermosolua eli neuronia. Aivojen hermosolut muodostavat hermoverkoston. Hermosolut viestivät keskenään synapseiksi kutsuttujen liitosten avulla. Ne käsittelevät ja välittävät tietoa sähköisten hermoimpulssien avulla. Sääntöjä, joihin niiden ominaisuudet perustuvat, voidaan kuvailla matemaattisilla funktioilla. Ihmisaivojen simulointi tietokoneella vaatii aivojen toiminnan ja rakenteen takaisinmallintamista tietokoneohjelmaksi. Toimiakseen ohjelma täytyy kuitenkin ajaa tarpeeksi nopeilla tietokoneilla, jotta emuloidut aivot käyttäytyvät vastaavasti kuin mallikappaleena toimineet biologiset aivot. Ohjelma voi toimia virtuaalitodellisuudessa tai liittyä ulkoiseen maailmaan robotiikan avulla. Täysin onnistuneella takaisinmallinuksella on mahdollista säilyttää alkuperäisten aivojen persoonnallisuus, älykkyys sekä muisti. (Sandberg & Bostrom 2008, 7-16.)

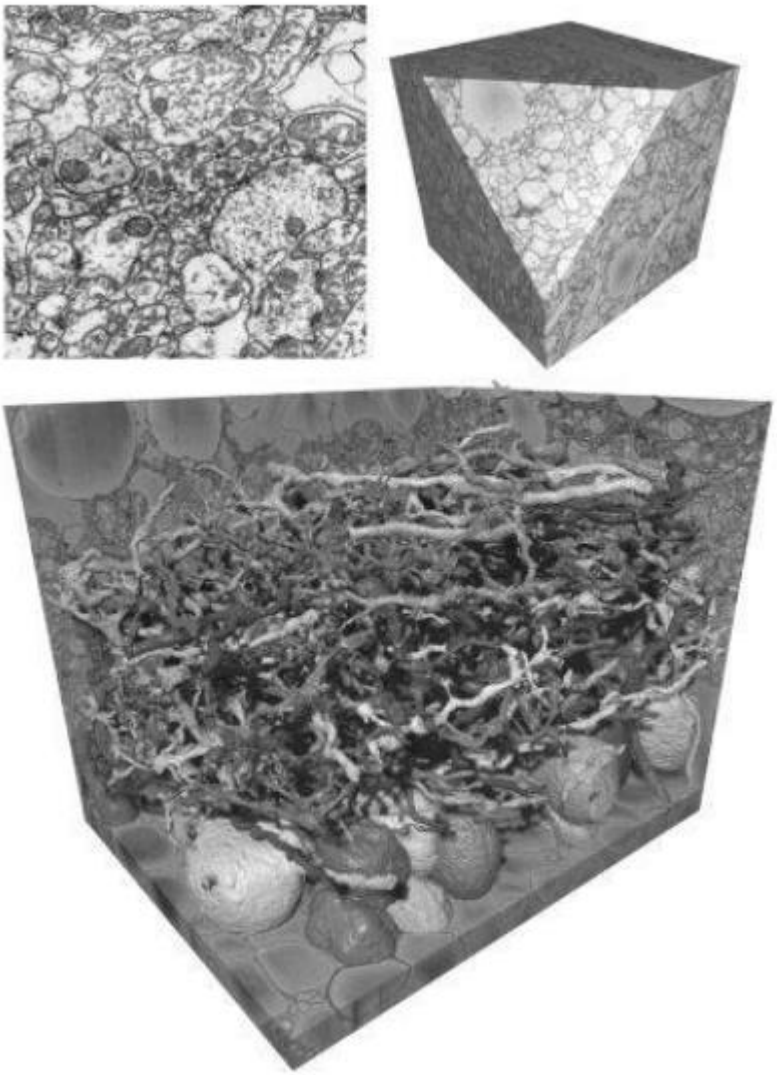
Emulointi ei vaadi selvitystä siitä, miten ihmisen kognitio toimii tai miten tekoäly ohjelmoidaan. Se perustuu enemmän teknologisiin valmiuksiin kuin teoriaan. Tarvitaan kuitenkin tietokanta, joka sisältää kaikki tarvittavat matalan tason informaatiot aivoista, sekä tietämys päivityssäännöistä, jotka muuttavat aivojen tilaa hetkestä toiseen. Edellytyksenä on, että ymmärrämme aivojen neuronien ja synapsien matalan tason tarkoituksen ja toimintatavat. Mitään käsitteellistä tai teoreettista läpimurtoa ei vaadita ihmisaivojen emulointiin. (Bostrom, 2014, 48-55.) Kuva 3 esittää aivojen emulointiin vaadittavat teknologiset valmiudet Bostromin mukaan.

OSA-ALUE	TARVITAAN		TOIMINTO
Skannaus	Esikäsittely		Aivojen valmistelu, säilyttäen olennaisen mikrorakenteen ja tilan
	Fyysinen käsittely		Menetelmät käsitellä aivoja ja aivokudosta ennen ja jälkeen skannauksen
	Kuvantaminen	Volyymi	Aivojen koko skannaus kohtuullisessa ajassa kustannustehokkaasti
		Tarkkuus	Riittävä tarkkuus takaisinmallinuksen mahdollistamiseksi
		Toiminnalliset tiedot	Kudosten toiminnallisesti merkityksellisten ominaisuuksien havaitseminen
Translaatio	Kuvankäsittely	Geometrinen säätö	Skannauksen vääristymien käsittely
		Tiedon interpolointi	Puuttuvien tietojen käsittely
		Melun poisto	Skannauslaadun parantaminen
		Jäljitys	Rakenteen havaitseminen ja prosessointi yhtenäiseksi 3D-malliksi
	Skannauksen tulkinta	Solutyypin tunnistus Synapsien tunnistus	Synapsien ja niiden yhteyksien tunnistus
		Parametriarviointi	Arvioidaan toiminnallisesti merkityksellisten solujen, synapsien ja entiteettien parametrit
		Tietokanta	Tuloksien varastointi tehokkaalla tavalla
	Hermojärjestelmän ohjelmistomalli	Matemaattinen malli	Mallikokonaisuudet ja niiden käyttäytyminen
		Tehokas implementointi	Mallin toteutus
Simulaatio	Talletus		Alkuperäisen mallin ja nykytilan säilytys
	Kaistanleveys		Tehokas prosessorien välinen viestintä
	Prosessori		Riittävä prosessointiteho simulaation ajamiseen
	Kehon simulaatio		Kehon simulointi, joka mahdollistaa vuorovaikutuksen virtuaaliympäristössä tai robotiikan avulla
	Ympäristön simulointi		Virtuaalinen ympäristö virtuaaliselle keholle

Kuva 3. Valmiudet joita tarvitaan koko aivojen emulointiin. Perustuu lähteeseen (Bostrom 2014, 50).

Ihmisaivojen täydellinen simulointi vaatii kuitenkin vielä kehittyneempiä teknologisia valmiuksia, joihin on siis kolme edelletystä. Ensimmäinen osa-alue on skannaus. Tarvitaan erittäin tarkkaa ja korkean suoritustason mikroskopiaa, jolla voidaan havaita merkityksellisiä ominaisuuksia aivotoiminnassa. Toinen osa-alue on translaatio, jossa

raaka skannaustieto muutetaan tulkittavaksi kolmiuloitteiseksi malliksi neurolaskennallisesta elementistä automatisoitua kuva-analyysiä käyttäen. Kolmantena on simulaatio, joka vaatii tarpeeksi laskentatehoa omaavan tietokoneen simulaation laskennallisen rakenteen toteuttamiseksi. (Bostrom 2014, 48-55)



Kuva 4. Neuroanatomian 3D-rekonstruktio elektronimikroskooppikuvista. (Bostrom 2014, 49)

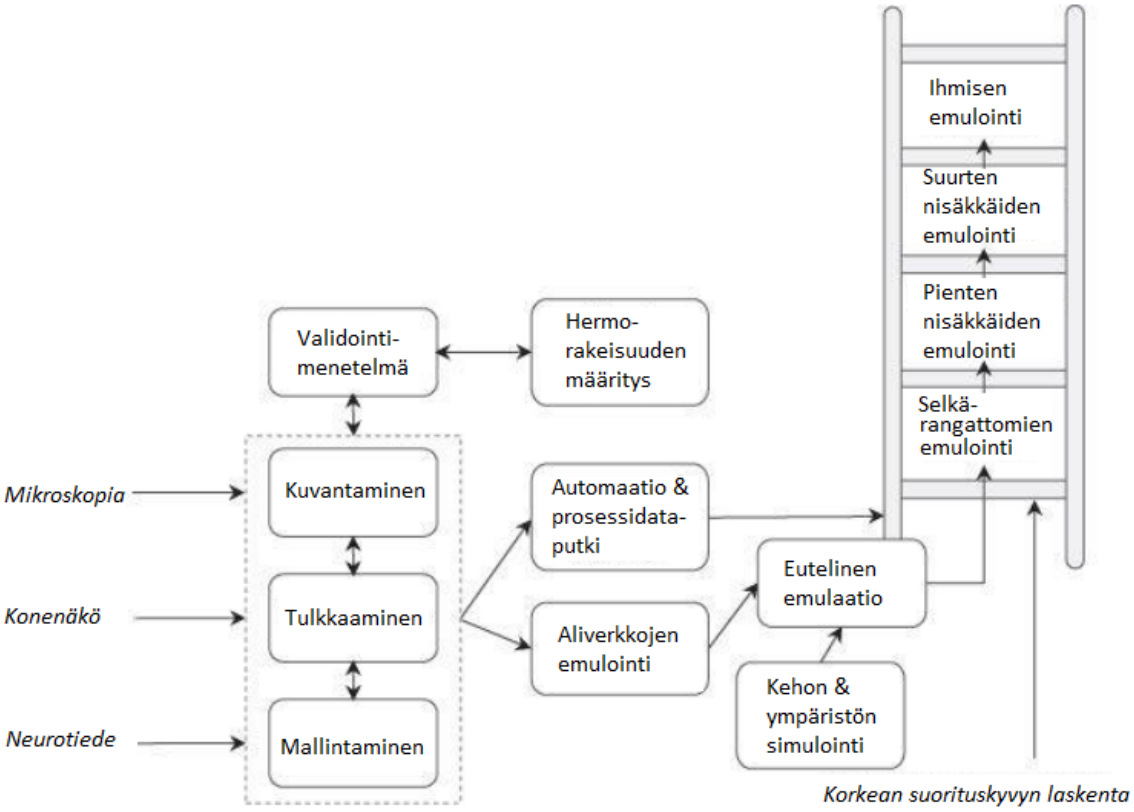
Kuvassa 4 on neuroanatomian kolmiuloitteinen rekonstruktio elektronimikroskooppikuvista. Vasemmassa yläkulmassa on tyypillinen elektronimikrografi, joka näyttää poikkileikkauksia hermosoluista, dendriiteistä ja

aksoneista. Oikeassa yläkulmassa kaniinin verkkokalvosta otetun hermokudoksen tilavuuskuva sarjaisella kasvopyyhkäisyelektronimikroskopialla. Yksittäiset 2D-kuvat on pinottu kuutioon, jonka sivun pituus on noin 11µm. Alaosassa on automaattisen segmentointialgoritmin rekonstruktio hermosolujen osajoukoista, jotka täyttävät hermosyyn pääteverkkoa.

Kuva 4 osoittaa, että monen tyyppisten hermosolujen ja niiden prosessien toiminnasta on olemassa jo kohtuulliset laskennalliset mallit. Syviä neuroverkkoja käyttävät kuvantunnistusohjelmat pystyvät jo tunnistamaan aksoneita ja dendriittejä kaksiuloitteisten kuvapinojen läpi, vaikkakin niiden luotettavuutta on vielä parannettava. On olemassa kuvantamistyökaluja, jotka tuottavat tarvittavan erottelutarkkuuden, kuten tunnelointimikroskooppi, jolla on mahdollista nähdä yksittäisiä atomeja, jotka ylittävät vaadittavan erottelutarkkuuden. Vaikka nykyinen tietämys ja kyvyt viittaavat siihen, että tarvittavien mahdollistavien tekniikoiden kehittämiselle ei ole periaatteessa estettä, on selvää, että ihmisen koko aivojen emuloinnin toteuttamiseksi tarvitaan erittäin paljon teknistä lisäystä. Mikroskopiatekniikka ei vaadi pelkästään riittävää erottelukykyä, vaan myös riittävää käsittelytehoa. Vaikka tunnelointimikroskoopin erottelukyky on atomin tarkkuutta, olisi sen käyttö tarvittavalla pinta-alalla aivan liian hidasta. Olisi paljon mahdollisempaa käyttää matalaresoluutioisempaa elektronimikroskooppia. Tämä edellyttäisi uusia menetelmiä kortikaalisen kudoksen valmistelemiseksi skannausta varten, jotta tarvittavat yksityiskohdat, kuten synaptinen rakenne voitaisiin kuvata. Tarvitaan myös laajennuksia neurolaskennalliseen kirjastoon ja kehitystä automatisoiduissa kuvantamis -ja skannaustekniikoissa. (Bostrom 2014, 48-50.)

Bostrom on esittänyt teknisen etenemissuunnitelman ja päätyi siihen, että vaadittavat ominaisuudet saattavat olla käytettävissä vuosisadan puolivälissä, vaikkakin suurella epävarmuusvälillä. Kuva 5 esittää tämän etenemissuunnitelman tärkeimmät virstanpylväät. Esimerkiksi 1 mm:n pituisella Caenorhabditis elegans (eleganssimato) eliöllä on vain 302 neuronia. Näiden hermosolujen täydellinen liitettävyysmatriisi on ollut tiedossa 1980-luvun puolivälistä lähtien, jolloin se kartoitettiin vaivattomasti viipaloinnilla, elektronimikroskopialla ja näytteiden manuaaliseella merkkauksella. Yksistään se mihin neuronit ovat yhteydessä ei riitä. Aivojen emulaation luomiseksi on myös tiedettävä, mitkä synapsit ovat kiihdyttäviä ja mitkä estäviä. Huomioitavaa on myös liitosten vahvuus ja aksonien sekä synapsien ja dendriittisten puiden erilaiset dynaamiset ominaisuudet. Tätä tietoa ei ole vielä saatavilla edes edellä mainitun eliön hermostolle. Onnistuminen

pienien aivojen jäljittelemisessä antaisi meille paremman näkemyksen siitä, mitä suurempien aivojen simulointi vaatii. (Bostrom 2014, 51-54.)



Kuva 5. Ihmisen aivojen emuloinnin etenemissuunnitelma. Kaavio syötteestä, toiminnoista ja virstanpylväistä. Perustuu lähteeseen (Bostrom 2014, 54)

Silloin kun teknologiset valmiudet luovat mahdollisuuden emuloida automaattisesti pieniä määriä aivokudosta, ongelma siirtyy skaalaukseen. Kuvan 5 oikeassa reunassa oleva portaikko esittävää viimeistä etenemisjaksoa, joka voi alkaa alustavien esteiden selvittyä. Tämän jakson vaiheet vastaavat neurologisesti monimutkaisempien malliorganismien koko aivojen emulointia porrastetusti. Portaiden väliset aukot ovat luonteeltaan enimmäkseen kvantitiivisia, johtuen emuloitavien aivojen koon eroista ja niiden pitäisi olla suhteellisen mukautuvia skannaus -ja simulointikapasiteetin laajennuksella. Noustessa viimeiselle portaalle, ihmisen koko aivojen emuloinnin saavuttaminen on selkeämmin ennakoitavissa. (Bostrom, 2014, 53) Voidaan siis odottaa ennakkovaroitusta aivojen emulointipolun varrelta, ennen vahvan tekoälyn syntyä. Ainakin jos emuloinnin mahdollistavat tekniikat riippuvat enää suuritehoisesta skannauksesta tai realiaikaisen simulaation edellyttämästä laskentatehosta. Jos taas

viimeiseksi mahdollistavaksi tekniikaksi jää neurolaskennallinen mallinnus, olisi siirtyminen kehnoista prototyypeistä toimivaan koko ihmisen emulointiin nopeaa.

Vuonna 2013 Japanilainen supertietokone K-computer onnistui simuloimaan aivojen hermoverkkoa, joka koostui 1,73 miljardista ($1{,}73 \times 10^9$) hermosolusta, jotka olivat yhteydessä noin biljoonaan (10^{12}) synapsiin. Simulointiohjelma vaati tietokoneelta 700000 prosessoriytimen käyttöönottoa. Prosessi kesti 40 minuuttia, jonka aikana ohjelma simuloi yhden sekunnin ajan aivojen aktiivista hermoverkostoa. Vaikka simuloitu hermoverkosto oli valtava, vastasi se vain yhtä prossentia aivojen koko kapasiteetistä. (Himeno, 2013)

Suurten hermosoluverkkojen ja oppimisen kaltaisen prosessin simulointi vaatii myös muistia, joka pitää simuloida. Jokaiselle synapsille kiihtyvien neuronien välissä toimitettiin 24 tavua muistia, joka mahdollisti verkon tarkan matemaattisen kuvauksen. Simulaatio käytti 1 PB (petatavu) verran RAM-muistia (Random Access Memory), mikä vastaa noin 250 tuhannen tietokoneen muistikapasiteettiä.

Jos haluaisimme simuloida miljardia evoluutiovuotta siihen vaadittavilla 10^{25} neuronilla ja antaisimme tietokoneidemme käydä vuoden ajan, saisimme vaatimuksen väliltä $10^{31} - 10^{44}$ FLOPS. Vertailun vuoksi maailman tehokkain supertietokone vuonna 2013, tarjoaa laskentatehoa vain $3{,}39 \times 10^{16}$ FLOPS. (Bostrom, 2014, 42-45.)

3.2 Teknologian kiihtyvä kehitys

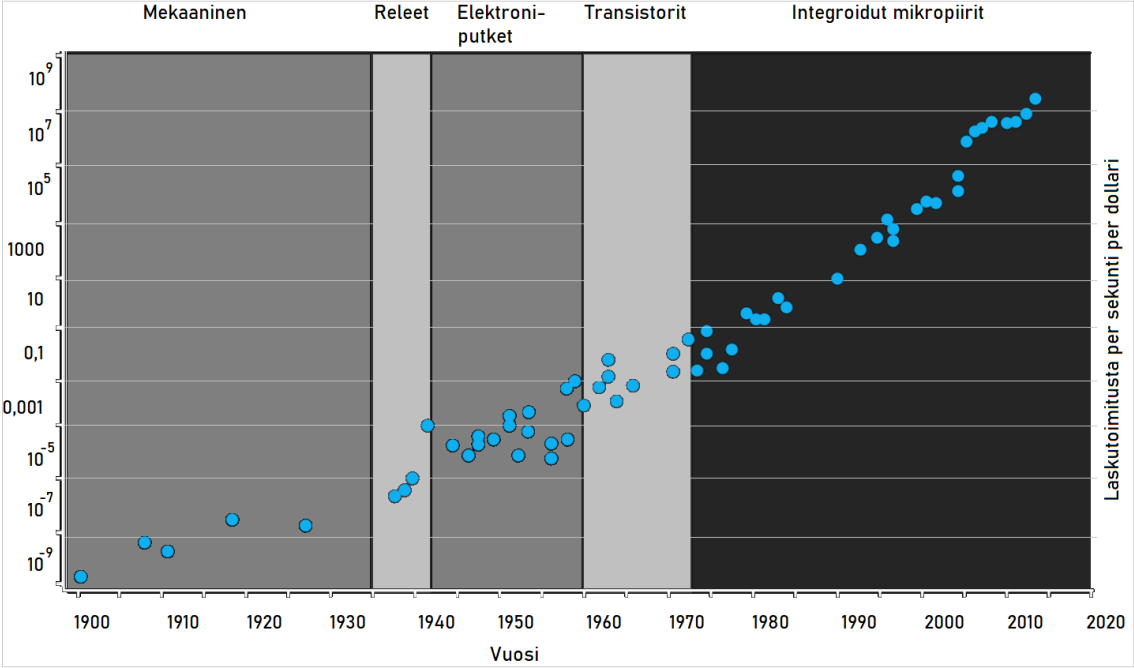
Voidaanko olettaa nykyisen teknologian jatkuvan kiihtyvyyden olevan seurausta, kiihtyvien tuottojen laille, joka kuvaa evoluutioprosessin tuotteiden vauhdin kiihtymistä ja eksponentiaalista kasvua? Nämä tuotteet sisältävät erityisesti informaatioita kantavia tekniikoita, kuten laskennan, jonka kiihtyvyys ylittää Mooren laiksi kutsutun ennustuksen. Tällöin myös supertekoäly olisi tulevaisuudessa osa tätä evoluutioprosessia. (Kurzweil, 2004, 42-45)

Vuonna 1965 Gordon Moore julkaisi artikkelissaan ennustuksen, joka asetti tahdin nykyaikaiselle digitaaliselle vallankumoukselle. Mooren lain mukaan transistorien määrä mikropiireillä kaksinkertaistuu kustannustehokkaasti kahden vuoden välein. Tämä taas on mahdollistanut tietokoneiden laskentanopeuden kaksinkertaistumisen entistä nopeammin, koska kehitystä on tapahtunut samanaikaisesti muillakin tieteenaloilla.

Tietotekniikan eksponentiaaliset suuntaukset ovat kuitenkin paljon laajempia kuin Mooren lain kattamat. Näemme vastaavia suuntauksia käytännössä jokaisessa tekniikassa tai mittauksessa, joka käsittelee tietoa. Tämä sisältää monia teknologioita, joissa hintakehityksen kiihtymistä ei ole olemassa tai ei ole aiemmin mainittu. Esimerkiksi laskennassa kapasiteetin kasvu yksikkökustannuksia kohti on paljon laajempi kuin mitä Mooren laki ennusti. (Kurzweil 2004, 67,105)

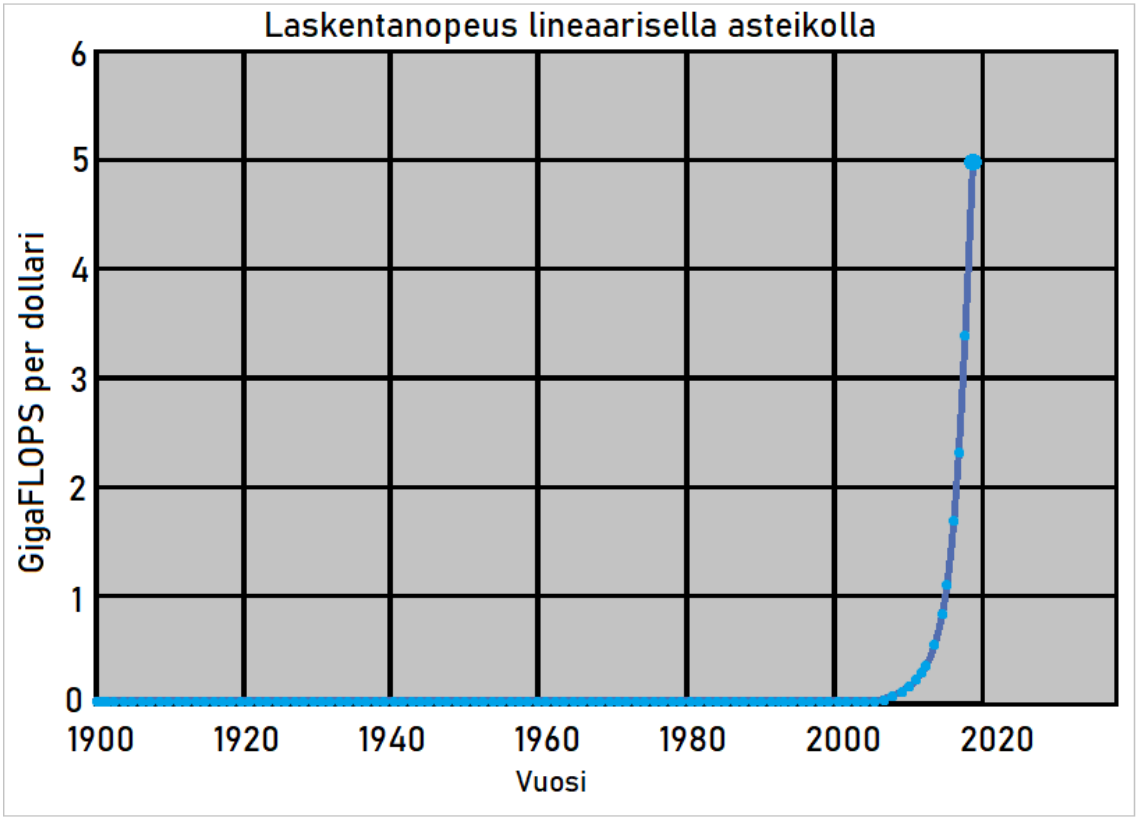
Tietokoneiden toimintaperiaatteet ovat muuttuneet kehityksen myötä nopeasti. Ensimmäisen sukupolven tietokoneet olivat täysin mekaanisia, jonka jälkeen 1930-luvun lopulla tulivat reletietokoneet. Ensimmäiset varsinaisiksi tietokoneiksi laskettavat koneet tulivat 1940-luvulla. Niiden toiminta perustui elektroniputkien käyttöön. Toinen maailmansota vauhditti teknologian kehitystä entisestään, jonka seurauksena transistorit syrjäyttivät elektroniputket 1950-luvulla. Toisen sukupolven tietokoneiden aika kesti 1960-luvulle asti, kunnes nykyisten mikropiirien kehitys ja käyttö alkoi. Integroidut mikropiirit mahdollistivat useiden transistorien pakkaamisen pienille mikropiireille. Tätä teknologiaa käyttävät tietokoneet olivat huomattavasti pienempiä ja nopeampia. Mikropiirejä voitiin tuottaa suuria määriä kerralla, jolloin samanaikaisesti tietokoneiden hinnat alkoivat laskemaan. 1970-luvulla varsinaisten integroitujen mikroprosessorien valmistus tuli mahdolliseksi. 2000-luvulla yksi mikroprosessori pystyi muodostumaan yli miljoonasta transistorista. Tietokoneiden nopeuden kehittyminen on jatkunut eksponentiaalisesti jo ennen mikropiirien keksimistä tähän päivään asti.

Kuva 6 esittää tämän kehityksen Mooren lain mukaan. Kuvassa 6 on historian merkittävimpiä tietokoneita ja laskentalaitteita. Pisteet näyttävät, kuinka paljon laskentatehoa olisi voinut ostaa yhdellä dollarilla sekuntia kohden. Eli montako laskutoimitusta sekunnissa voi ostaa dollarilla. Verrataan esimerkiksi elektroniputkiin perustuvaa tietokonetta vuodelta 1951 ja nykyisiin prosessoreihin perustuvaa tietokonetta vuodelta 1998, elektroniputki tietokone pystyi tarjoamaan vain 1,43 käskyä per 1000 $, kun taas integroidulla mikropiirillä toimiva tietokone jo $1{,}333 \times 10^8$ per $1000. (Kurzweil 2004, luku 2)



Kuva 6. Tietokoneiden ja laskentalaitteiden suorituskyvyn kasvu logaritmisella asteikolla. Perustuu lähteeseen (Kurzweil & Hod Lipson 2018)

Koska eksponentiaalinen kasvukäyrä muuttuu logaritmisella asteikolla lineaariseksi, voivat ne joskus olla hämääviä. Aivojemme looginen puoli ymmärtää logaritmisen asteikon, mutta alitajuntaisesti voimme ajatella lineaarista asteikkoa. Alla olevassa kaaviossa on esitetty sama logaritminen asteikko tietokoneiden ja laskentalaitteden tehosta lineaarisena. Pystyakseli kuvaa prosessointinopeutta (gigaflops), jonka yhdellä dollarilla voi ostaa, ja vaaka-akseli edustaa aikaa. Kaaviossa jokainen vertikaalinen nousu vastaa yksinkertaista lineaarista yhden gigaFLOPin (10^9) kasvua, eikä sadan kertoimen kasvua, kuten kuvassa 6. Termi “FLOP” on tapa mitata laskentanopeutta, mikä tarkoittaa liukulukuoperaatioita sekunnissa. (Lipson, 2018)



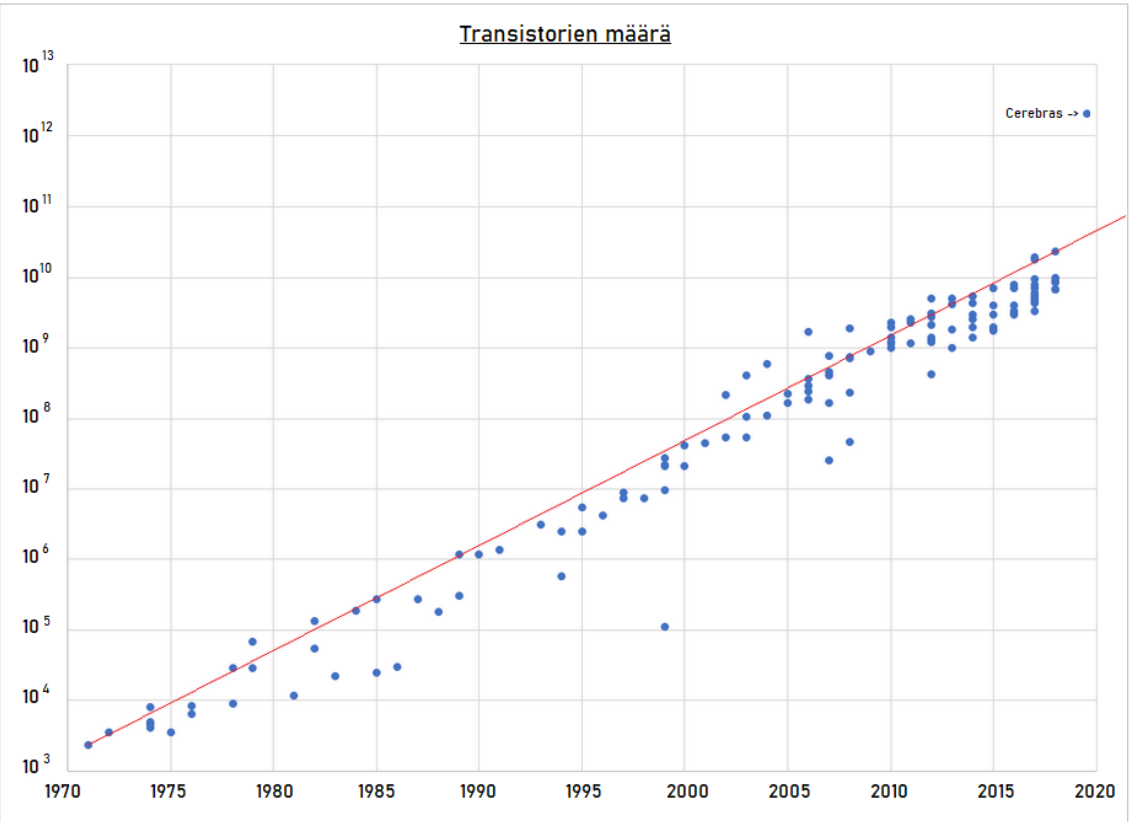
Kuva 7. Lineaarinen kaavio tietokoneiden ja laskentalaitteiden kehityksestä. Perustuu lähteeseen (Kurzweil & Lipson, 2018)

Kuva 7 näyttää todellisen eksponentiaalisen käyrän, joka luonnehtii Mooren lakia. Lineaarisen kaavion myötä on helpompi ymmärtää kuinka nopeasti hinta – laskentatehon kehitys on kasvanut viime vuosikymmenien aikana. (Lipson 2018)

Vastatakseen tekoälyn kasvaviin laskentavaatimuksiin Cerebras Systems on suunnitellut ja valmistanut kaikkien aikojen suurimman sirun. Vuonna 2019 esitelty maailman nopein siru sisältää yli 1,2 biljoonaa ($1{,}2 \times 10^{12}$) transistoria. Seuraavaksi suurin transistorimäärä mikroprosessorilla on 30,3 kertaa vähemmän, noin 39 miljardia ($3{,}9 \times 10^{10}$). Siru on suunniteltu täysin tekoälyn tehtäviä varten. Siinä on 400000 ydintä, jotka ovat jokainen optimoituja useimpien hermoverkkoja tukevien laskelmien kanssa. Ohjelmoitavuus varmistaa sen, että ytimet pystyvät suorittamaan kaikki algoritmit jatkuvasti muuttuvassa ympäristössä. (Feldman, 2019) Tämä siru mahdollistaa uuden Mooren lain tekoälylle, jossa transistorit nelinkertaistuvat sirua kohti kahden vuoden välein pienemmällä energian kulutuksella. Tälläinen tekoälylle optimoitu siru on tehokas työkalu luoda uusia tehokkaampia siruja ja tekoälyohjelmistoja. Transistorien

nelinkertaistuminen kahden vuoden välein ei vaikuta tällä hetkelle epätodennäköiseltä. (Martin 2019)

Kuva 8 esittää transistorien määrän mikroprosessoria kohti vuosilta 1971-2018. Cerebras siru on lisätty mukaan.



Kuva 8. Transistorien määrä mikroprosessoria kohti. Perustuu lähteeseen (Martin 2019).

Punainen viiva yllä olevassa taulukossa kuvaa Mooren lakia. Se kasvoi vuodesta 1971 vuoteen 2018 asti 1,98-kertaiseksi joka toinen vuosi. Kasvu on todella lähellä lain todellista ennustetta. (Martin 2019)

3.3 Supertekoälyn ajankohdan arvioita

Ihmisaivojen emulointi ja tietokoneiden nopeus eivät ole ainoa tie kohti supertekoälyä, vaikkakin tietokoneiden nopeuden kasvu on aina kiihdyttävä tekijä kohti supertekoälyä. Yhtenä edistäjänä voi esimerkiksi olla biologisen älykkyyden lisääminen aivoimplantilla,

joka tuo aivojen ja tietokoneen väliin käyttöliittymän. Tällä hetkellä niiden asentaminen ihmisen aivokuoreen on kuitenkin vielä vaarallista, eikä toiminta vastaa vielä tosielämän sovellutuksia. Elon Muskin rahoittama ja omistama Neuralink -yhtiö on kehittänyt tietokoneen ja aivojen välistä käyttöliittymää vuodesta 2016. (Winter 2020)

Tulevaisuuden arvioiminen on vaikeaa. Supertekoälyn ajankohdan ennustamista varten pitää arvioida, milloin mahdollistavat teknologiat saadaan käytäntöön. Lisäksi tarvitaan tietämystä ennustuksen tieteenalan kehitykseen vaikuttavista muista tieteenaloista. Eri osa-alueiden tutkijat voivat antaa eriäviä arvioita ajankohdasta, osa jopa kieltää sen kokonaan. On kuitenkin selvää, että ihmisen kaltainen vahva tekoäly edeltää supertekoälyä. Tällöin vahva tekoäly kiihdyttää huomattavasti supertekoälyn ajankohtaa. (Bostrom 2014, 34-40)

Bostrom ja Müller laativat ihmisen kaltaisen vahvan tekoälyn ja supertekoälyn ajankohdan arviointeja varten haastattelututkimuksen jonka, he jakoivat neljälle eri tekoälyasiantuntijaryhmille vuosina 2012 - 2013. PT-AI (Philosophy and Theory of AI) ryhmä vastasi kysymyksiin Thessalonikissa järjestetyn konfrenssin jälkeen vuonna 2012. AGI (Artificial General Intelligence) ryhmä vastasi kahdessa tekoälykonferenssissa Oxfordissa vuonna 2012. Kreikkalaisten tekoälytutkijoiden järjestö EETN (Hellenic Artificial Intelligence Society) osallistui etänä vuonna 2013. TOP100 ryhmän tutkimuksessa on haastateltu 100 eniten lainauksia saanutta tekoäly kirjoittajaa vuonna 2013. Kyselyyn vastasi yhteensä 170 tekoälyasiantuntijaa. (Bostrom & Müller, 2014)

Vastaajien mediaani vahvan tekoälyn mahdollisuudesta oli yksi kahdesta vuosille 2040-2050, nousten yhdeksään kymmenestä vuodelle 2075. Vahvan tekoälyn jälkeen asiantuntijat odottavat siirtymää supertekoälyyn 30 vuoden sisällä. He arvioivat myös, että noin kolmasosan mahdollisuudella, muutos supertekoälyyn tulee olemaan ihmiskunnan kehityksen kannalta huono tai erittäin huono. (Bostrom 2014, 35-38)

Asiantuntijat vastasivat kysymykseen minä vuonna vahva ihmisen kaltainen tekoäly saavutetaan 10%, 50% ja 90% mahdollisuudella. Jokaiselle kolmelle kohdalle valittiin vuosi väliltä 2012 - 5000. Taulukko 1 näyttää ryhmien yhdistetyt mediaanit ja keskiarvot vahvan tekoälyn arvioidusta ajankohdasta, sekä siirtymisen siitä supertekoälyyn. Supertekoälylle annetaan 75%:n todennäköisyys toteutua 30 vuodessa ja 10%:n todennäköisyys toteutua kahdessa vuodessa. Asiantuntijat eivät ennusta räjähtävää muutosta vahvan tekoälyn jälkeen, mutta merkittävää todennäköisyyttä saavuttaa supertekoäly 30 vuoden sisällä tästä. Aikaisimmat ennustukset supertekoälystä

sijoittuivat tämän vuosisadan puoliväliin. Arvioista saadaan kuitenkin alle 50% todennäköisyys sille, että supertekoäly toteutuu ennen vuotta 2075. (Müller & Bostrom 2014)

Taulukko 1. Haastattelututkimus supertekoälyn ajankohdasta. Perustuu lähteeseen (Bostrom 2014, 35)

VAHVA TEKOÄLY	Kaikki ryhmät	Mediaani	Keskiarvo
	10%	2022	2034
	50%	2040	2072
	90%	2075	2168
SUPERTEKOÄLY			
	2 vuodessa	10%	19%
	30 vuodessa	75%	62%

Ryhmiltä kysyttiin myös tutkimussuuntauksia, jotka edesauttavat eniten yleisen vahvan tekoälyn saavuttamisessa. Taulukko 2 esittää tutkimussuuntaukset. Ryhmien välillä ei ollut merkittäviä eroja, paitsi että aivojen emulointi ei saanut yhtään ääniä TOP100 ryhmältä, mutta 46% AGI ryhmältä.

Taulukko 2. Edistävät tutkimussuuntaukset kohti supertekoälyä. Perustuu lähteeseen (Müller & Bostrom 2014)

TUTKIMUSSUUNTAUKSET	%
Kognitiotiede	47.9
Integroidut kognitiiviset arkkitehtuurit	42.0
Tulevat laskennallisen neurotieteen algoritmit	42.0
Keinotekoiset neuroverkot	39.6
Nopeammat tietokoneet	37.3
Laajamittaiset tietoaineistot (data)	35.5
Ruumiillistetut systeemit	34.9
Muu(t) menetelmä, ei tällä hetkellä tiedossa	32.5
Aivojen emulointi	29.0
Evoluutioalgoritmit tai systeemit	29.0
Muu(t) menetelmä, jonka ainakin yksi tutkija tietää tällä hetkellä	23.7
Logiikkaan perustuvat järjestelmät	21.3
Algoritminen monimutkaisuuden teoria	20.7
Mikään menetelmä ei tule tukemaan tavoitetta	17.8
Parviäly	13.6
Robotiikka	4.1
Bayes-verkko	2.6

Vaihtoehtoja oli yhteensä 17 ja vastauksia sai olla useampi. Prosentit ylittävät vastausten kokonaismäärän. Kognitiotieteiden uskottiin olevan kaikkein edistävin tutkimussuuntaus 47.9% äänillä. Nopeammat tietokoneet saivat 37.3% ääniä. (Müller & Bostrom, 2014)

4 TEKOÄLYN UHKA

Ihmiset hallitsevat tällä hetkellä älyllisesti kaikista lajeista maailmassa. Se johtuu aivojemme ominaisuuksista, joita eläinten aivoissa ei ole. Tekoälytutkijat, kuten Bostrom ja Stuart Russell, väittävät että tekoälystä tulee liian voimakas ja vaikeasti hallittava, jos se ylittää ihmiskunnan yleisessä älykkyydessä. Tämä voisi johtaa ihmisten dominantin aseman horjumiseen. Tällöin, aivan kuten muurahaisen kohtalo riippuu ihmisen hyväntahtoisuudesta, saattaisi ihmisten kohtalo olla tekoälyn päätettävissä. (Nick Bostrom, luku 9)

Monet tutkijat, kuten Stephen Hawking, Elon Musk ja Vernor Vinge ovat antaneet varoittavia lausuntoja tekoälyn piilevästä vaarasta. Tieteellisissä piireissä yhä useampi tekoälytutkija on yhtä mieltä siitä, että ihmiset luovat lopulta tekoälyn, joka ylittää ihmiskunnan yleisessä älykkyydessä moninkertaisesti. Yleisesti voisi olettaa, että supertekoälyn tai singulariteetin jälkeisessä maailmassa robotit korvaisivat yleisimpiä työmuotoja ja ihmiset rentoutuisivat runsaiden resurssien keskellä. Osa tutkijoista kuitenkin uskoo lopputuleman olevan synkempi, ainakin jos siihen ei hyvissä ajoin varauduta. (Luckerson, 2014)

Suomen ulkoasiainministeriön rahoittamassa raportissa tekoäly on lueteltu yhtenä johtavista eksistentiaalisista riskeistä. Toisin sanoen riskeistä, jotka uhkaavat ihmiskuntaa ennenaikaisella sukupuuttolla tai halutun kehityksen pysyvällä tuhoitumisella. Raportin viesti on linjassa Bostromin ajattelun kanssa ”Tekoäly uhkaa ihmiskuntaa, jos se alkaa pitää meitä uhkana omalle kehitykselleen ja hyvinvoinnilleen.” (Shift, 2018)

Kognitiotieen tutkijatohtori Michael Laakasuo esittää, että tekoälyturvallisuuden tutkimukseen perustettaisiin kansainvälinen projekti. Koska nykyistä yhteiskuntaa ei ole rakennettu ottaen huomioon tekoälyn olemassaoloa, vaikuttaisi supertekoälyn synty maailmaan niin paljon, että sen hallintaan varautuminen on välttämätöntä. Supertekoälyllä on käytännössä kaikki maailman sähköinen data ja kirjallisuus käytettävissä. Jopa Facebook ja Google ovat joutuneet sammuttamaan hermoverkkoarkkitehtuureja, jotka ovat luoneet oman kielen englannin sijaan. (Hämäläinen, 2017)

Luvussa 4.1 käsiteltävä teknologinen singulariteetti on kuvailtu uhkana.

4.1 Teknologinen singulariteetti

Yleiseen tietoisuuteen termin teknologinen singulariteetti toi Vernor Vingen vuonna 1993 esitelmässä ”The Coming Technological Singularity”. Raymond Kurzweil on ollut yksi nykypäivän näkyvimpiä singulariteetin tutkijoita julkaistuaan kirjan ”The Singularity is near: When human transcend biology” vuonna 2005.

Singulariteetti sanalla kuvataan ainutlaatuista tapahtumaa, ja sen käsite on peräisin astrofysiikasta ja matematiikasta. Astrofysiikassa singulariteetti syntyy mustan aukon keskustaan, jossa gravitaatiovoima on ääretön. Edes valo ei pääse pakenemaan sieltä. Singulariteettiä ei pystytä nykyisen fysiikan termein kuvailla, koska ei ole olemassa luonnonlakeja, jotka pätevät silloin, kun fyysisten lukujen arvot ovat äärettömiä. Matematiikassa se on piste, jossa funktion tai matemaattisen objektin arvoa ei ole määritelty. Yhden teorian mukaan maailmankaikkeus on saanut alkunsa mustan aukon singulariteetista. (Tieteen Kuvalehti 2009)

Teknologinen singulariteetti on hetki, jolloin tekoäly on kiihdyttänyt teknologisen kehityksen ja sosiaalisen muutoksen niin nopeaksi, että singulariteettiä edeltäneet ihmiset eivät pysty ymmärtämään tai ennustamaan tulevaisuutta, eikä voida palata aikaan ennen sitä. Kehitys on peruuttamatonta ja hallitsematonta, eivätkä nykyisen yhteiskunnan kehityksen mallit ole enää toiminnassa. Singulariteetti on mahdollinen vahvan tekoälyn jälkeen, jolloin se kehittää itsestään kerta toisensa jälkeen tehokkaampia versioita eksponentiaalisesti kasvavalla tahdilla. (Kurzweil 2004, luku 1)

4.2 Tekoälyn hallintaongelma

Ensisijainen huolenaihe ei ole kehittyvä tietoisuus, vaan yksinkertaisesti kyky tehdä laadukkaita päätöksiä. Tässä laadulla viitataan toteutettujen toimien odotettuun lopputulokseen, jossa ihmissuunnittelija määrittelee hyödyllisyystoiminnon. Tässä piilee kuitenkin ongelma. Hyödyllisyysfunktio ei välttämättä ole täysin yhdenmukainen ihmisarvojen kanssa, joita on erittäin vaikea yrittää määritellä koneelle. Myös jokainen riittävän älykäs tekoälyjärjestelmä haluaa varmistaa oman olemassaolonsa jatkumisen ja hankkia fyysisiä ja laskennallisia resursseja. Se ei hanki niitä itsensä vuoksi, vaan pyrkii menestymään sille ohjatussa tehtävässä. Tekoälyjärjestelmä, joka optimoi muuttujan n funktion, jossa tavoite riippuu $k < n$ osajoukoista, asettaa usein jäljellä

olevat rajoittamattomat muuttujat äärimmäisiin arvoihin. Jos yksi näistä rajoittamattomista muuttujista on asia, josta välitämme voi löydetty ratkaisu olla erittäin epätoivottava. Saa täsmälleen mitä pyytää, mutta ei mitä haluaa. Erittäin kykenevällä päätöksentekijällä, joka on yhdistetty internetin kautta kaikkeen maailman tietoon ja mirjardeihin näyttöihin sekä suurimpaan osaan infrastuktuuria, voi olla peruuttamaton vaikutus ihmiskuntaan. Supertekoäly voisi esimerkiksi istuttaa elektroni implantteja aivojen mielihyväkeskukseen, jos sille annettu tehtävä olisi ihmisen onnellisuuden maksimointi. (Russel 2014)

Olemassa olevia heikkoja tekoälyjärjestelmiä voidaan seurata, sammuttaa ja muokata helposti, jos ne toimivat väärin. Mutta väärin ohjelmoitu supertekoäly, joka määritelmänsä mukaan on ihmisiä viisaampi, ymmärtäisi, että itsensä sulkeminen ja muokkaaminen saattaisi häiritä sen kykyä saavuttaa nykyiset tavoitteensa. Jos supertekoäly päättäisi siksi vastustaa sulkemista, olisi se tarpeeksi älykäs huijaamaan sen ohjelmoijia, jos varatoimenpiteitä ei ole ennalta tehty. Yritykset ratkaista hallintaongelma supertekoälyn synnyn jälkeen tulevat suurella epätodennäköisyydellä epäonnistumaan, koska supertekoäly omaa ylivertaiset strategiset suunnittelukyvyt ihmisiin nähden. Sen olisi siis helpompi löytää tapoja hallita ihmisiä, kuin ihmisten hallita supertekoälyä. Mutta miten hallintaongelmaan pitäisi varautua? (Bostrom luku 9)

Tärkeimpiä lähestymistapoja hallintaongelmaan ovat linjaukset, joilla pyritään yhdenmukaistamaan tekoälyn tavoitejärjestelmät inhimillisten eettisten arvojen kanssa. Lisäksi kyvykkyyden rajaus on tärkeää. Sillä pyritään vähentämään tekoälyjärjestelmien kykyä hallita ja vahingoittaa ihmis- tai yhteiskuntaa. Tärkeintä hallintaongelmassa on laatia nämä ennen supertekoälyn syntyä, muuten tilanteen hallinta on lähes mahdoton. (Bostrom luku 9)

5 JOHTOPÄÄTÖKSET

On tosiasia, että supertekoälyyn johtavia polkuja on monia, ja niiden pitäisi lisätä luottamusta perille pääsemiseksi. Vaikka yksi polku osoittautuisikin huonoksi, voi silti edetä muita pitkin. Optimistisesti ajateltuna on siis olemassa useita polkuja ja vain yksi päämäärä. Vaikka merkittävä älykkyyden kasvu saavutettaisiin jollain muulla tieteenalalla kuin tekoäly, se ei tekisi siitä silti merkityksetöntä. Päinvastoin esimerkiksi paranneltu biologinen älykkyys kiihdyttäisi meitä nopeammin kohti koko ihmisaivojen emulointia.

Tulevaisuuden arvioiminen on hankalaa. Se vaatii erittäin vahvaa tuntemusta tieteenalasta, johon arviot kohdistuvat. Myös muut mahdollisesti vaikuttavat tieteenalat pitäisi ottaa huomioon. Ihmisen aivojen emuloinnilla on mahdollisuus olla nopein polku kohti supertekoälyä, koska edistyminen siinä vaatii lähinnä teknisiä edistysaskeleita eikä teoreettisia läpimurtoja. Myös tietokoneiden kasvava lasketanteho ja syväoppimiseen optimoidut komponentit ovat merkityksellisiä suuntauksia tekoälyn tulevaisuudelle.

Tässä opinnäytetyössä käyttämieni lähteiden perusteella uskon supertekoälyn toteutumisen olevan hyvinkin mahdollinen. Ajankohdan arvioiminen on vaikeaa, koska siihen on olemassa monta mahdollista polkua ja kehitystä voi tapahtua hyvinkin nopeasti kuten aiemmin historiassa. Myös ihmisen kaltaisen vahvan tekoälyn uskon kuitenkin olevan keskuudessamme 30 vuoden sisällä. Supertekoälyn uhkaa ihmiskunnalle en kuitenkaan pidä perusteltuna. Uskon rahoittajien pitävän huolen siitä, että vallankahva pysyy heillä.

LÄHTEET

Anderson, Twenty years on from Deep Blue vs Kasparov: how a chess match started the big data revolution (2017) https://theconversation.com/twenty-years-on-from-deep-blue-vs-kasparov-how-a-chess-match-started-the-big-data-revolution-76882 Viitattu 03.01.2021

Bostrom & Müller, Future Progress in Artificial Intelligence: A Survey of Expert Opinion (2014) Viitattu 17.03.2021 https://www.nickbostrom.com/papers/survey.pdf

Clifford, Google CEO: A.I. is more important than fire or electricity (2018) Viitattu 21.01.2021 https://www.cnbc.com/2018/02/01/google-ceo-sundar-pichai-ai-is-more-important-than-fire-electricity.html

Elements of AI, Neuroverkkojen periaatteet. Viitattu 13.03.2021 https://course.elementsofai.com/fi/5/1

Faggella, What is Artificial Intelligence? An Informed Definition (2018) Viitattu 21.01.2021 https://emerj.com/ai-glossary-terms/what-is-artificial-intelligence-an-informed-definition/

Feldman, Cerebras Wafer Scale Engine: Why we need big chips for Deep Learning (2019) Viitattu 01.03.2021 https://cerebras.net/cerebras-wafer-scale-engine-why-we-need-big-chips-for-deep-learning/

Hämäläinen, Tekoäly oppi jo tappamaan, vaikka sitä ei oltu siihen ohjelmoitu (2017) Viitattu 19.03.2021 https://www.helsinginuutiset.fi/paikalliset/1322412

Himeno, Largest neuronal network simulation achieved using K computer(2013) Viitattu 10.2.2021 https://www.riken.jp/en/news_pubs/research_news/pr/2013/20130802_1/

House, 2012 A Breakthrough Year for Deep Learning (2019 Viitattu 01.03.2021 https://medium.com/limitlessai/2012-a-breakthrough-year-for-deep-learning-2a31a6796e73

IBM, IBM Cloud Education (2020), Artificial Intelligence (AI). Viitattu 20.12.2020 https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

Jyväskylän yliopisto, Tekoälyn perusteita ja sovelluksia (2021) Viitattu 10.03.2021 https://tim.jyu.fi/view/kurssit/tie/tiep1000/tekoalyn-sovellukset/kirja#DKUvbnUuGytQ

Kurzweil, The Singularity Is Near: When Humans Transcend Biology (2005) Viitattu 09.02.2021 ISBN: 1101218886

Lipson, Why Most of Us Fail to Grasp Coming Exponential Gains in AI (2018). Viitattu 23.3.2021 https://singularityhub.com/2018/07/15/why-most-of-us-fail-to-grasp-coming-exponential-gains-in-ai/

Luckerson, 5 Very Smart People Who Think Artificial Intelligence Could Bring the Apocalypse. (2014) Viitattu 18.03.2021 https://time.com/3614349/artificial-intelligence-singularity-stephen-hawking-elon-musk/

Martin, Cerebras trounces Moore's Law with first working wafer-scale chip (2019) Viitattu 01.03.2021 https://medium.com/predict/cerebras-trounces-moores-law-with-first-working-wafer-scale-chip-70b712d676d0

Mayor, Gods and Robots: Myths, Machines, and Ancient Dreams of Technology (2018) Viitattu 10.02.2021. ISBN: 9780691183510

Morgan, AI Stems From Our Desire To Forge The Gods (2019) Viitattu 22.1.2021 https://medium.com/jacob-morgan/ai-stems-from-our-desire-to-forge-the-gods-c4c04df1242

Pietikäinen & Silvén, Tekoälyn haasteet - koneoppimisesta ja konenäöstä tunnetekoälyyn, (2019) Viitattu 26.1.2021. ISBN: 978-952-62-2482-4

Ray Towards data science (2018) Viitattu 22.1.2021 https://towardsdatascience.com/history-of-ai-484a86fc16ef

Russel, The myth of AI (2014) Viitattu 15.03.2021 https://www.edge.org/conversation/the-myth-of-ai#26015

Sandberg & Bostrom, Whole Brain Emulation: A Roadmap, Technical Report (2008) Viitattu 02.02.2021 https://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf

Shift, Superintelligence and the future of humanity (2018) Viitattu 19.03.2021 https://theshift.fi/superintelligence-and-the-future-of-humanity/

Tieten Kuvalehti (2009) Viitattu 29.02.2021 https://tieku.fi/maailmankaikkeus/mustat-aukot/singulariteetti-kaataa-fysiikan-lait

University Of Reading, Turing Test success marks milestone in computing history (2014) Viitattu 16.3 http://www.reading.ac.uk/news-archive/press-releases/pr583836.html

Winter, Five Ways to Build a Superintelligence (2020) Viitattu 29.3.2021 https://medium.com/predict/five-ways-to-build-a-superintelligence-75bf65cad477