# Uhkakuvat voivat sumentaa tekoälykehityksen todellisuutta

HENRIKKI SALO-PÖNTINEN

## JOHDANTO

Tässä katsauksessa esittelen ongelmallisia tulkintoja tekoälyn tulevaisuudesta sekä niihin kytkeytyvien pelkojen vaikutusta suunnitteluajattelun kaventumiseen. Katsauksessani käsittelen erityisesti Catriona Campbellin teosta *AI by Design: A Plan for Living with Artificial Intelligence* (2022), joka kuuluu Chapman & Hall/CRC Artificial Intelligence and Robotics -kirjasarjaan. Campbellin teos on hyvä esimerkki populaarista tietokirjallisuudesta, jonka esittämiä näkemyksiä tekoälyn kehityksestä voi pitää haitallisina tai jopa vaarallisina.

Catriona Campbellin esikoisteos pyrkii tarjoamaan lukijalleen näkökulman tekoälyn kehityksen lähitulevaisuudesta ja sen kaitsemiseen tarvittavista hallinnollisista toimista siinä kuitenkaan täysin onnistumatta. Ongelmallisinta on kirjan perustuminen mediassa esiintyviin artikkeleihin ja osin epäkoherentteihin mielipiteisiin ja väittämiin. Kirjasarja, johon julkaisu kuuluu, antaisi olettaa, että kyse on populaaritieteellisestä teoksesta. Varoisin kuitenkin kutsumasta teosta populaaritieteelliseksi, sillä suuri osa kirjoittajan ajatuksista ei pohjaudu tutkimuskirjallisuuteen. Varoitus pätee erityisesti tekoälyteknologioiden kehityksen kuvaamiseen. Tekoälyyn liittyvien hallinnollisten aiheiden[1] ja tekoälyn periaatelistausten puolesta teos tarjoaa kapean yleiskatsauksen sekä muutaman esimerkin suunnitteluajattelun mahdollisesta hyödyntämisestä hallinnollisten toimien kehittämiseksi.

Teoksen nimi *AI by Design* ja Catriona Campbellin tausta ihmisen ja tietokoneen vuorovaikutukseen (human-computer interaction)

1 Teoksen heikkoutena tästä näkökulmasta tarkasteltuna on, että vaikka kirjoittaja pitää lainsäädäntöä ja standardeja keskeisinä hallinnollisina keinoina, hän ei mainitse näiden parissa tehtyä merkittävää kansainvälistä työtä, kuten EU:n *AI act* ja yleinen tietosuojalaki GDPR (sekä USA:n vastaava tietosuojalaki), YK:n alla tehtyä autonomisten tappavien aseiden hallintaan liittyvää työtä tai kansainvälisen standardointijärjestön IEEE:n tekoälyn eettisen kehityksen standardointityötä. Salo-Pöntinen ja Saariluoma (2022) sekä Sigfrids et al. (2022) tarjoavat katsauksia tekoälykehityksen hallinnollisiin kysymyksiin.

erikoistuneena konsulttina ja suunnittelijana antaa olettaa, että kirja olisi analyyttinen ja tekoälyn suunnitteluun perehtynyt kuvaus hyvin määritellystä aihekokonaisuudesta. Todellisuudessa kirja ennemminkin osoittaa, miten vinoutuneen kuvan tekoälyn kehityksestä voi saada perehtymällä tekoälyä koskevaan keskusteluun valta- ja sosiaalisen median syötteiden kautta, kuten Campbell kertoo tehneensä. Campbell toistaa kirjassaan tekoälykehitykseen usein liitettyjä suuria narratiiveja: tekoäly on itsenäinen toimija, joka kehittyy ja oppii itse, ihmisten tuntema maailma tulee tuhoutumaan itsetietoisen tekoälyn kehittyessä, vain harva viisas ihminen kykenee ennustamaan tekoälyn kehitystä, tekoälyteknologiat on kehitetty viimeisen vuosikymmenen aikana ja teknologia kehittyy muusta yhteiskunnasta erillään. Viimeksi mainitusta hyvänä esimerkkinä on ajatus, jonka mukaan oikeusjärjestelmämme olisivat voimattomia teknologiakehityksen edessä. Narratiivit ovat läsnä heti tekstin alusta, minkä vuoksi lukijan toiveet yleispätevästä näkökulmasta tekoälyteknologian suunnitteluun ja ihmistieteitä hyödyntävän suunnitteluajattelun hyödyllisyyteen kaikkoavat nopeasti.

Seuraavaksi käyn läpi keskeiset ongelmat Campbellin kirjan tekoälykehityksen kuvauksissa ja liitän ne laajempaan kontekstiinsa. Lopuksi tarkastelen johtopäätösten kautta, miten kyseiset kuvaukset haastavat rakentavan suunnitteluajattelun muodostumista.

## ONGELMALLISISTA TULKINNOISTA UHKAKUVIIN

Selkeys ja yksinkertaisuus ovat hyviä tavoitteita niin tieteellisten teorioiden kehittämisessä kuin missä tahansa kommunikoinnissa. Niiden saavuttaminen ei kuitenkaan ole helppoa. Ensiksi tulee huomioida, että vaikka niiden välillä vallitsee yhteys, selkeys ja yksinkertaisuus eivät ole toistensa synonyymeja. Asian ilmeneminen selkeänä perustuu suureksi osin tietopohjalle, jota aiheen ymmärtämiseksi vaaditaan. Kaikki mikä on selkeää minulle, ei välttämättä ole sitä muille ja vastavuoroisesti toisinpäin. Toiseksi tulisi tietää, mitä selitettävästä ilmiöstä voidaan yksinkertaistaa ja mitä yksinkertaistuksessa kadotetaan.

Catriona Campbell kertoo, että hän luottaa suunnittelijan työssään ilmiöiden yksinkertaisiin ilmauksiin ja selityksiin. Tämä välittyy lukijalle tavassa, jolla Campbell kuvaa tekoälykehitystä. Lukijan annetaan muun muassa olettaa, että kaikki suuret keksinnöt ja loikat tekoälyn saralla ovat tapahtuneet viimeisen muutaman vuoden aikana. Tekoälyteknologioita on kuitenkin kehitetty jo yli 60 vuoden ajan ja monet nykyään käytössä olevista teknisistä ratkaisuista on keksitty vuosikymmeniä sitten. Nykyisen tekoälyratkaisujen saaman medianäkyvyyden syyt löytyvät valtavasti kasvaneista tietomassoista ja laskentatehon kasvusta, jotka ovat mahdollistaneet aiemmin keksittyjen tekniikoiden potentiaalin toteutumisen käytännössä. Nils J. Nilsson antaa erinomaisen katsauksen vuosikymmenten aikana tapahtuneesta tekoälytekniikoiden kehityksestä kirjassaan *The Quest for Artificial Intelligence* (2009).

Campbellin hyödyntämistä yksinkertaistuksista haasteellisin on se, että hän pohjaa teoksensa näennäistieteelliselle *loogiselle evolutiiviselle* jatkumolle muun muassa todetessaan: ”Singularity[2] is the next logical leap forwards”.[3] Jatkumoajattelun mukaan tekoäly kehittyy ensin kapeana tekoälynä, josta se kehittyy yleistekoälyksi ja lopulta supertekoälyksi. Kapealla

2 Singulariteetti on lainasana astrofysiikasta. Se kuvastaa erityisesti mustien aukkojen ominaisuuksiin linkitettyä ilmiötä, jossa käsittämämme fysiikan lait eivät enää päde. Tällöin emme voi myöskään ennustaa, mitä singulariteetin tuolla puolen tapahtuu. Teknologinen singulariteetti on otettu käyttöön kuvastamaan kuvitteellista tilannetta, jossa supertekoäly kehittää itse itseänsä muutamissa sekunneissa niin valtavasti, että ihmisen ymmärrys ei riitä ennustamaan kehityksen suuntaa tai vaikutuksia.

3 Campbell 2022, 16.

tekoälyllä tarkoitetaan, että tekoälyteknologioita voidaan hyödyntää vain hyvin rajatuissa kapeissa tehtävissä, kuten kuvantunnistuksessa. Yleistekoälyllä tarkoitetaan puolestaan tekoälyä, jota voitaisiin hyödyntää toteuttamaan kaikkia ihmistenkin suorittamia tehtäviä ja johon usein liitetään sellaisia piirteitä kuin itsetietoisuus, kyky asettaa omia tavoitteita sekä hallinta omasta oppimisesta. Supertekoäly kuvastaa puolestaan tekoälyä, jonka älykkyys ylittää inhimillisen ymmärryksen rajat. Tarkennukseksi totean, että kaikki tekoäly, jota maailmassa kehitetään, on kapeaa tekoälyä. Hyvin suunniteltuina ja toteutettuina tekoälysovellukset auttavat ihmisiä ylittämään kyvykkyytensä noilla kapeilla tehtäväalueilla: ne mahdollistavat nopeamman, laajemman ja tarkemman informaation prosessoinnin, mikä johtaa nopeampaan ja laadukkaampaan ennusteiden tekemiseen, päätöksentekoon ja ongelmanratkaisuun. Lisäksi ne mahdollistavat tauottoman työtahdin. Juuri näiden ominaisuuksien vuoksi ne koetaan niin hyödyllisiksi.

Edellä kuvattu jaottelu kapeaan tekoälyyn (narrow AI), yleistekoälyyn (artificial general intelligence) ja supertekoälyyn (superintelligence) ei ole kuvaus loogisesta jatkumosta. Se on kuitenkin tärkeä käsitteellinen työväline, jotta voimme määritellä mistä puhumme, kun puhumme tekoälystä. Jaottelun tekee hyödylliseksi se, että meillä ihmisillä on taipumus antropomorfisoida teknisiä artefakteja[4] eli mieltää niihin inhimillisiä piirteitä. Tämä pätee erityisesti autonomista toimintaa – esimerkiksi liikkumisen tai tekstin tuoton muodossa – ilmentäviin artefakteihin. Campbellin teoksen suurin kompastuskivi on, että hän esittää tekoälyn kolmiosaisen jaottelun siten kuin kyseessä olisi luonnon lakeja seuraava vääjäämätön tekoälyn kehityskulku. Samalla hän tulee tehneeksi nipun hyvin kyseenalaisia oletamuksia tekoälystä ja sen kehityksestä.

Selitän olettamuksia enemmän tekstin edetessä, mutta nostan aluksi esiin niin sanotun *kehysongelman* (engl. frame problem), joka asettaa yleistekoälyn kehityksen mahdollisuuden hyvin kyseenalaiseksi. Se tarkoittaa sitä, että todellisessa maailmassa tekoälyratkaisut tulevat kohtaamaan tilanteita ja tekijöitä, joita niihin ei ole ohjelmoitu. Näin ollen niillä tulisi olla kyky sopeutua uusien tekijöiden läsnäoloon, mutta sellaisen kehittäminen ei ole yksinkertainen tekninen haaste eikä välttämättä mahdollista, sillä kaikki maailman asiat eivät ole formalisoitavissa.[5]

Käytetyt termit vaikuttavat osaltaan siihen, miten miellämme keskustelujen aiheet, ja siksi sillä, millä termeillä puhumme asioista, on suuri merkitys. Campbellin tapa puhua tekoälystä sekoittaa usein toisiinsa itsetietoisen yleistekoälyn ja monimutkaiset autonomiset järjestelmät, jotka yhdistelevät useita kapean tekoälyn ratkaisuja monimutkaisen ongelman ratkaisemiseksi. Autonomiset ajoneuvot ovat hyvä esimerkki tällaisesta monimutkaisesta kokonaisuudesta: järjestelmän täytyy kyetä tarkkailemaan dynaamisia ulkoisia olosuhteita, kuten tien muotoja, muita ajoneuvoja ja paikallisista liikennesäännöistä viestiviä merkkejä, ja havaintojensa pohjalta muuttamaan ajoneuvon toimintoja ajotehtävän näkökulmasta tarkoituksenmukaisesti. Monimutkaisuudestaan huolimatta tällaiset järjestelmät ovat edelleen kapeaa tekoälyä.

Sen sijaan, että puhumme Campbellin tavoin yleistekoälystä, joka on inhimillistetty kuvaus, olisi parempi ilmaista tekoälytekniikoiden olevan yleishyödyllisiä. Yleishyödyllisyys tarkoittaa sitä, että tekoälytekniikoita voidaan käyttää lähes rajattomassa määrässä erilaisia kontekseja, vaikkakin jokaisella teknisellä ratkaisulla on omat rajoitteensa, mikä asettaa raamit sille, miten ja mihin tekoälytekniikoita voi ja kannattaa hyödyntää. Näin ollen tekoälyteknologioiden yleishyödyllisyyden ymmärtäminen edellyttää, että huomioimme sekä kontekstisidonnaiset

4 Tekninen artefakti viittaa mihin tahansa ihmisen kehittämään tai hyödyntämään objektiin.

5 Daniel Dennett antaa ongelmasta hyvän kuvauksen kirjoituksessaan ”Cognitive Wheels: The Frame Problem of AI” (1984) ja Hubert Dreyfus kirjassaan *What Computers Still Can't Do* (1992).

tekijät että yleispätevät teknologiakehityksen piirteet.

Campbell on hyvä kirjoittaja ja hän onnistuu luomaan argumenteistaan koherentin mielikuvan, joka lukijan on helppo hyväksyä. Valitettavasti argumenttien välinen koherenssi ei riitä, kun on tarkoitus kuvata todellista teknologian kehitystä ja luoda tulevaisuuden kuvia, joiden varaan olisi mahdollista kehittää strategista hallinnollista työtä. Sen vuoksi on tärkeä huomata, että Campbellin tekoälyteknologiaa koskevien argumenttien sisällöt ovat oikeasti pistemäistä teknisten artefaktien ja valittua näkökulmaa tukevien näkemysten luettelemista. Näitä ovat muun muassa luonnollisen kielen prosessointi (NLP), kuvantunnistus ja -manipulointi, parviäly sekä Elon Muskin Neuralink-yrityksen popularisoima aivokäyttöliittymä. Näistä erityisesti aivokäyttöliittymien mahdollisuuksia kuvaa virheellisesti sekä Muskin yritys että Campbell kirjassaan.

On virheellistä väittää, että aivokäyttöliittymät mahdollistavat ajatusten lukemisen, kuten Campbell ja Musk antavat olettaa. Aivoalueen verenkierto voi esimerkiksi viestiä, että henkilö yrittää liikuttaa raajaa, mutta sen avulla ei voi saavuttaa varsinaista ajatuksen sisältöä, toisin sanoen mielensisältöä. Mielensisältöjen selittämistä aivotoimintojen avulla kutsutaan neurotieteissä mereologiseksi virhepäätelmäksi, josta Bennet ja Hacker ovat kirjoittaneet teoksessaan *Philosophical Foundations of Neuroscience* (2003). Esimerkin on tarkoitus osoittaa, kuinka teknologioiden hyvin vaikuttavia, jopa sensaatiomaisia toimintoja – kuten tekoraajojen hyödyntäminen aivokäyttöliittymällä ehdottomasti on – käytetään usein hyödyksi argumentoitaessa teknologisen kehityksen rajoittamattomuuden puolesta. Esitettyjen toimintojen ja väitettyjen mahdollisuuksien välillä ei kuitenkaan ole mitään tieteellistä yhteyttä, vaikka niin annetaan olettaa.

Campbell ei erottele, milloin hän puhuu tekoälyn kehityksen asiantuntijasta ja milloin äänekkäästä mielipiteiden lausujasta. Tärkeimmät tekoälyn teknisiin näkökulmiin perehtyneet asiantuntijat, joiden näkemyksiä Campbell hyödyntää, ovat filosofi Nick Boström, tietojenkäsittelytieteilijä Stuart Russell ja matemaatikko Alan Turing. Heidän lisäkseen Campbell pohjaa suuren osan argumentoinnistaan mediassa hyvin näkyvien, mutta ei aiheen asiantuntijoiden, kuten Elon Muskin ja Ray Kurzweilin lausuntoihin sekä tieteisfiktioon. Campbell käyttää hyödykseen myös Stephen Hawkingin lausuntoja tekoälykehityksen haasteista. Hawking on ehdottomasti merkittävä tieteentekijä teoreettisen fysiikan ja tieteen popularisoinnin saralla, mutta populistisissa tekoälyä käsittelevissä teksteissä hänen lausuntojaan hyödynnetään usein virheellisesti viimeisenä auktoriteettina, joka kertoo totuuden tekoälyn vaaroista.

Campbell tekee hallaa Boströmin, Russellin ja Turingin näkemyksille tekoälyn kehitykseen liittyvistä haasteista liittämällä heidän kokonaisuuksista irrotetut sitaatit tukemaan Campbellin kuvaamaa tiedostavan yleistekoälyn tulemista. Todellisuudessa heidän näkemyksensä kuvastavat niin kutsuttua *kontrollin ongelmaa*, joka on hyvin relevantti haaste ilman oletusta yleis- tai supertekoälystä. Lyhyesti selitettynä kontrollin ongelma liittyy siihen, että tekoälyä kehittävät ihmiset ovat kaikkien ihmisten tapaan rajallisen kapasiteetin toimijoita, eivätkä he voi tyhjentävästi ilmaista mielekkäitä tavoitteita, joita sofistikoituneiden tekoälyjärjestelmien olisi tavoiteltava. Todellisuus, jossa tekoälyartefaktien on tarkoitus toimia, on paljon monimutkaisempi kuin yksittäiset kehittäjät voivat ymmärtää. Näin ollen kehittäjät saattavat tietämättään luoda järjestelmiä, jotka tuottavat ei-toivottuja tuloksia. Russell on käsitellyt tätä aihetta hyvin kattavasti kirjassaan *Human Compatible* (2019), jossa hän myös tarjoaa ongelmaan erinomaisia ratkaisuehdotuksia.

## ALARMISMI JA MYYTTI SISÄPIIRISTÄ

Campbellin teoksen tekninen anti tyhjentyy Ray Kurzweilin teoksen *The Singularity is Near* (2005) ja Nick Boströmin kirjan *Superintelligence* (2014) kapealle tulkinnalle. Nämä kaksi teosta ovat hyvin vaikutusvaltaisia Campbellin kirjan kaltaisissa tekoälyn kuvauksissa. Campbell muun muassa pohjaa yleistekoälyn kehittymisen ”aika-

taulun" Kurzweilin ennustukselle. Lisäksi kirjan ensimmäisen luvun otsikko "Sleepwalking into Singularity" on Boströmin usein hyödyntämä metafora, jolla hän kuvaa kontrollin ongelmaa ja keinoja sen muodostumiselle. Näin ollen otsikko herättää odotukset analyyttisesta tavasta käsitellä kontrollin ongelmaa. Campbellin otsikko kuitenkin pettää lukijansa, sillä hän ei esittele mitään reittiä tai tapoja, joilla kontrollin ongelma tai yleistekoäly voisivat toteutua. Luvun lopussa hän vain on yhtäkkiä hyväksynyt oletamuksen, että yleistekoälyn kehittyminen tulee tapahtumaan ja nopeasti.

Boström on Oxfordin yliopiston *Future of Humanity* -instituutin johtaja ja käytännöllisen etiikan professori. Hän kyllä puhuu kirjassaan superälystä ja sen kehittymisen mahdollisuuksista, mutta vahvan akateemisen filosofin tapaan hän eksplikoi, mitä perusolettamuksia hänen kuvaamansa skenaariot sisältävät. Boström myös pohjustaa ajatuksiaan toteamalla ensimmäisessä luvussa, että on hyvin mahdollista, että kaikki mitä hän kirjassaan väittää, voi osoittautua täysin vääräksi, mutta aiheet, joita hän esittelee, on silti tärkeää pohtia läpi. Kurzweilin teos ei puolestaan ole tieteellinen teos, vaan enemmänkin yksilön ajatuksia kuvaava ja pohtiva tutkielma, joka pohjaa singulariteetin argumenttinsa fysikalistiselle käsitykselle mielestä. Kurzweilin mukaan supertekoäly kehittyy, kun ihminen saa skannattua aivonsa digitaaliseen muotoon. Hän ajattelee, että tällöin myös ihmisen mieli tai tietoisuus siirtyy digitaaliseen muotoon ja on vapaa ihmiskehon rajallisuudesta. Tämä on kuvaus niin kutsutusta aivojen täydellisestä emuloinnista (whole brain emulation).

Kurzweilin ajatus on monin tavoin ongelmallinen. Keskeisin ongelma liittyy fysikalistiseen käsitykseen mielestä, jonka mukaan mieli on yhtä kuin aivojen toiminta, ja mielentilat ovat yhtä kuin aivotilat. Fysikalistinen käsitys on ristiriidassa sen kanssa, minkälainen ymmärryksemme on mielestä. Vaikka ihmisen mielen ja kehon (sisältäen aivot) välillä on yhteys, ei voida sanoa, että aivot ovat yhtä kuin mieli tai tietoisuus. Muun muassa mielensisältöjä ei voida selittää aivotoiminnoilla,[6] eikä ihmisen tietoisuuden sijaintia tai syntymekanismeja ole kyetty selvittämään. Fysikalismin mukainen päätelmä tekee siis hypyn nykytietämyksestämme. Tähän kaikkeen voi tutustua syvemmin David Chalmersin tunnetuksi tekemän *tietoisuuden vaikean ongelman* (the hard problem of consciousness) kautta. Vaikka tietoisuuden ongelmaa ei Kurzweilin tapaan pitäisi olennaisena, kirjoittajan tulisi tietää, ettei hänen esittämilleen väitteille ole empiirisiä todisteita. Toisin sanoen olisi hyväksyttävä ja todettava ääneen, että kyse on olettamuksista, joita ei ole todistettu. Tätä Campbell ei tee.

Boströmin ja Kurzweilin taustasta on hyvä tietää, että he ovat posthumanismin (Boström) ja transhumanismin (Kurzweil) keulahahmoja. Kumpikin ajatussuuntaus kannattaa lähtökohtaisesti rajatonta ihmisen keinotekoista kohennusta teknologisten ratkaisujen avulla. Boströmin edustama kanta on jokseenkin varovaisempi ja analyyttisempi kuin transhumanistien kanta. Hänen mukaansa mitään keinoja ei tulisi lähtökohtaisesti sulkea pois, vaan tarkastella tilannekohtaisesti. Trans- ja posthumanistit hyödyntävät alarmismia – eli perusteettomien uhkakuvien ja kiireellisyyden tunteen luomista – omien argumenttiensa toteuttamisen tarpeellisuuden ja välttämättömyyden perustelemiseksi. Alarmismin tapaisen uhkakuviin perustuvan argumentoinnin tarkoitus on kaventaa mahdollisten toimien näkymiä argumentoijan esittämien näkymien hyväksi, eli kyseessä on suunnittelua ja keskustelua tyrehdyttävä kommunikoinnin tapa.

Myös Campbell hyödyntää alarmismia kommunikoinnin keinona. Hän joko käyttää uhkaavaa sävyä tehokeinona tai tekstin tummasävytteisyys heijastaa hänen omia pelkojansa tekoälykehitystä kohtaan. Alarmismi välittyy erityisen selkeästi tulevaisuuden skenaarioissa, joita Campbell tarjoaa lukijalleen. Hyvin lyhyessä

6 Pertti Saariluoma on käsitellyt tätä aihetta kattavasti kirjassaan *Foundational Analysis* (1997).

skenaarioiden valikointiperusteiden kuvauksessaan Campbell kertoo, miksi hän ei luo kuvauksia apokalyptisista robottisodista: ” – – there are probably enough apocalyptic movies out there that cover the topic of battling with AI for control of the earth”.[7] Jo se, että näkee tällaisen perustelun tarpeellisena, kertoo paljon kirjoittajan näkemyksien luonteesta. Tästä huolimatta yksi Campbellin esittelemistä skenaarioista on apokalyptisen robottisodan jälkeisen maailman kuvaus, toinen sisältää viittauksia robottisotaan ja kolme muuta skenaariota ovat kuvauksia siitä, kuinka tekoäly dominoi ihmisiä tavalla tai toisella. Campbell ei siis vain pyörrä omia sanojansa, vaan tarjoaa lukijalleen pelkästään dystooppisia tulevaisuuden fantasioita, joita hän ei vaivaudu selittämään. Hän vain toteaa tulevaisuuden skenaarioidensa olevan ”rooted in fact”.[8]

Suunnitteluajattelun esittämisen ja hyödyntämisen näkökulmasta Campbellin teos on hyvin ongelmallinen siinä, että kirja implisiittisesti trivialisoi kaikkien muiden kuin insinöörien osaamisen itse teknisten artefaktien suunnittelussa. Tämä tapahtuu kuvaamalla tekoälyn tekninen kehitys viisaiden sisäpiirin toimintana (sisäpiiriin kuuluakseen tulisi ymmärtää yleistekoälyn olevan tulossa). Koen tarpeelliseksi tuoda esiin, että tällaista sisäpiiriä ei ole olemassa, vaan se on pelkkä tekoälyyn liittyvissä populaarikirjoituksissa usein toistuva mielikuva. Kirjan eri osiot antavat myös ymmärtää, että jos henkilö ei ole tekninen osaaja, hänen tulisi kyseenalaistamatta hyväksyä sisäpiirin kuvaamat kehityssuunnat. Mikäli henkilö ei hyväksy sisäpiirin näkemyksiä, hän on hölmö. Campbell viittaa tällaisiin hölmöihin lainaamalla Elon Muskia: ”– – it feels that anyone who doesn’t believe AI could be a[n existential] threat is ‘way dumber than they think they are’”.[9] Sisäpiirin ulkopuolisten rooliksi jää suunnitella sosiaalinen konteksti, jossa teknisiä artefakteja kehitetään ja hyödynnetään. Tämä on sekä virheellinen että erittäin vahingollinen tapa mieltää teknologian kehitys. Se on yhteydessä Campbellin teoksen pohjavireeseen, joka kuvaa teknologian ihmisistä ja yhteiskunnasta erillään kehittyvänä todellisuutena. Tällainen kuvaus ei ole uusi ilmiö, sillä jo 1900-luvun loppupuolella kehittyi tieteen ja teknologiatutkimuksen suuntaus, joka pyrkii luomaan kriittisiä näkökulmia teknologiakehitykselle ja sen deterministisille kuvauksille. Yksi suuntauksen pioneereista, Sheila Jasanoff, tekee kirjassaan *The Ethics of Invention* (2016) hyvän analyysin harhakäsityksistä ja virheellisistä olettamuksista, joille deterministiset ja teknokraattiset teknologiakäsitykset usein perustuvat.

Onneksi teknokraattinen ja deterministinen käsitys teknologian kehittymisestä ei ole yleispätevä kuvaus insinöörien ja teknologian kehittäjien ajatuksista. Tekoälyn kontekstissa on viimeisen viiden vuoden aikana noussut kansainvälisiä aloitteita, jotka alleviivaavat tarvetta suunnitella tekoälyteknologioita luonnon-, insinööri-, humanististen ja sosiaalitieteiden näkökulmia yhdistävänä yhteistyönä. Esimerkkejä yhteistyöstä eri alojen välillä on Stanfordin yliopiston ihmiskeskeisen tekoälyn instituutti, Wienin manifestin yhteydessä julistettu digitaalisen humanismin ohjelma ja Japanissa alkunsa saanut suunnitteluparadigma ”Yhteiskunta 5.0” (Society 5.0). Näiden lisäksi Saariluoma, Cañas ja Leikas antavat teoksessaan *Designing for Life* (2016) kattavan kuvauksen suunnitteluajattelusta ja -metodologiasta, joka ottaa kaiken teknologian kehittämisen lähtökohdaksi teknologian roolin osana ihmisten toimintaa ja elämää. Tästä näkökulmasta käsin ihmisten rooli teknologian kehityksessä ei kavennu tuotteiden käyttäjiksi ja työvoimaksi, jonka tulee sopeutua vääjäämättömään tulevaan, vaan ihmisten elämä nähdään syynä sille, miksi teknologiaa kehitetään.

Toinen Campbellin teoksen heikkous suunnittelun näkökulmasta on, että se nojaa yksilöitä korostavaan ajatteluun. Campbell toteaa johtopäätöksissään:

7 Campbell 2022, 77.
8 Campbell 2022, 77.
9 Campbell 2022, 11.

> Artificial Intelligence is truly godlike, the more I read and listen and watch, the more I realise that I don't truly understand the scale of AI and never will. It's too vast. There is no latter-day Victorian polymath who can comprehend the intellectual vastness of AI disciplines. If there is, that polymath will be an AI.[10]

Sitaatti on myös hyvä kuvaus ahdistuksesta, jonka median kautta välittyvä näkymä tekoälyyn voi saada aikaan. Campbellin ahdistus näyttäytyy muun muassa lauseessa, jonka mukaan singulariteetti on maailmalle ja yhteiskunnillemme polttavampi kysymys kuin ilmastonmuutos.[11] Tällaiset lausahdukset ovat irrallaan todellisuudesta ja vaarallisia, jos ne saavuttavat auktoriteetin aseman.

## JOHTOPÄÄTÖKSET

Campbellin kirja on parhaimmillaan osoitus siitä, että meillä on huutava tarve moninaistaa näkemystämme teknologia-asiantuntijoista sekä kehittää kykyä puhua teknologiasta ja sen myötä myös tekoälyn kehityksestä. Campbellin ihannoima Elon Musk voi olla äänekäs miljardööri, mutta hän ei ole tekoälyasiantuntija siinä mielessä, että hänen mielipiteidensä pohjalta olisi kannattavaa rakentaa tulevaisuuden visioita, saati yhteiskunnallisia strategioita. Tekoälyartefaktien kehitys ei ole yksittäisten ihmisten kampanja, kuten Campbellin lausahdus ja kirjan anti antaisi olettaa. Se on monen eri osaamisalan asiantuntijoiden työskentelyn yhdistämistä yhteisen todellisuutemme kehittämiseksi.

Jos teknisten artefaktien kehittäminen mielletäisiin kirjassa edellä kuvatulla tavalla ja tekoälyn kehitystä tarkasteltaisiin yleishyödyllisyyden, ei yleistekoälyn kautta, pääsisi Campbellin kirjan hyvät puolet ansaitsemaansa asemaan. Hyvistä puolista keskeisin liittyy tekoälykehityksen aiheuttaman työtehtävien uudistumisen ja mahdollisten työtehtävien loppumisen aiheuttaman työvoimasiirtymän oikeudenmukaiseen suunnitteluun. Tässä näkökulmassa Campbellin tapa käsitellä mahdollista työtehtävien vähenemistä merkityksellisyyden kokemuksen kautta on raikas tuulahdus monesti käytetyn, yksinkertaistetun ja sisällöllisesti tyhjän ajatuksen ”kun ihmisten ei tarvitse tehdä töitä, he voivat tehdä kaikkea muuta” rinnalle. Tummasävytteisestä otteestaan huolimatta Campbellin keskeisin viesti on, että teknologioihin liittyvien sosiaalisten ulottuvuuksien suunnittelu on mahdollista ja tärkeää. Niin se onkin. Sitä ei kuitenkaan voi mielekkäästi erottaa itse teknisten artefaktien suunnittelusta tai pohjustaa virheellisille käsityksille, kuten Campbellin kirja tekee.

TM Henrikki Salo-Pöntinen on kognitiotieteen väitöskirjatutkija Jyväskylän yliopistossa, informaatioteknologian tiedekunnassa. Väitöstyössään Salo-Pöntinen tarkastelee ja jäsentää tekoälyn eettisen suunnittelun taustaoletuksia ja käytäntöön viemisen ennakkoehtoja. Hän tekee tutkimustaan osana Suomen Akatemian rahoittamaa strategisen tutkimuksen neuvoston ETAIROS-hanketta.

• henrikki.b.pontinen@jyu.fi

10 Campbell 2022, 147.
11 Campbell 2022, 16.

## KIRJALLISUUS


Bennett, Max R. & Peter M. S. Hacker (2003). *Philosophical Foundations of Neuroscience*. Malden MA: Blackwell Publishing.

Boström, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Campbell, Catriona (2022). *AI by Design: A Plan for Living with Artificial Intelligence*. Chapman & Hall/CRC Artificial Intelligence and Robotics Series. Boca Raton: CRC Press.

Chalmers, David (1995). Facing up to the Problem of Consciousness. *Journal of Consciousness Studies* 2:3, 200–219.

Dennett, Daniel (1984). Cognitive Wheels: The Frame Problem of AI. *Minds, Machines and Evolution*. Toim. Christopher Hooway. Cambridge: Cambridge University Press, 129–150.

Dreyfus, Hubert L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: The MIT Press.

Jasanoff, Sheila (2016). *The Ethics of Invention: Technology and the Human Future*. New York: W. W. Norton & Co.

Kurzweil, Ray (2005). *The Singularity is Near: When Humans Transcend Biology*. New York: Viking Books.

Nilsson, Nils J. (2009). *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press.

Russell, Stuart J. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking Press.

Saariluoma, Pertti, José J. Cañas & Jaana Leikas (2016). *Designing for Life: A Human Perspective on Technology Development*. London: Palgrave Macmillan.

Saariluoma, Pertti (1997). *Foundational Analysis: Presuppositions in Experimental Psychology*. London: Routledge.

Salo-Pöntinen, Henrikki & Pertti Saariluoma (2022). Reflections on the Human Role in AI Policy Formulations: How Do National AI Strategies View People? *Discover Artificial Intelligence* 2, Article 3. https://doi.org/10.1007/s44163-022-00019-3.

Sigfrids, Anton, Mika Nieminen, Jaana Leikas & Pietari Pikkuaho (2022). How Should Public Administrations Foster the Ethical Development and Use of Artificial Intelligence? A Review of Proposals for Developing Governance of AI. *Frontiers in Human Dynamics* 4. https://doi.org/10.3389/fhumd.2022.858108.