Tiedetta vai tarinaa -podcast

Jakso 4
Miksi tekoälylle kannattaa olla ystävällinen? Tutkija kertoo kolme syytä

Lauri Järvilehto: Mitä jos meillä syntyy tällainen teknologinen entiteetti, joka käyttäytyy kaikin puolin ikään kuin se olisi tietoinen? Se eksistentiaalisen kriisin määrä, mitä tuollainen murros voi tuoda mukanaan.. Se on jopa isompi kuin se, että jotkut avaruusolennot laskeutuis tänne.

Suvi Rantonen: Tervetuloa seuraamaan Aalto-yliopiston Tiedettä vai tarinaa -podcastia. Tänään tarkastelemme scifitarinoiden näkökulmia tietoisuuteen ja mieleen. Mikä on koneen ja ihmisen välinen ero ja miten tietoisuus muodostuu? Siitä keskustelimme tänään. Mun nimi on Suvi Rantonen. Mä olen Aalto-yliopiston alumni ja scifiharrastaja. Tiedettä meillä tänään edustaa Aalto-yliopiston työelämäprofessori, filosofi Lauri Järvilehto. Lämpimästi tervetuloa!

Lauri: Kiitoksia!

Suvi: Lauri, sun 'Konemieli'-kirja ilmestyi aikaisemmin tänä vuonna ja mennään myöhemmin sen kirjan aiheisiin. Mutta aloitetaan tällaisella kevyellä kysymyksellä, että monet ainakin kuvittelee tietävänsä mitä on tietoisuus, mutta mitä se oikeastaan on?

Lauri: No se on hyvä kysymys, johon on siis olemassa monenlaisia vastauksia myöskin. Mun mielestä aika vitsikkäämpiä on Daniel Dennettin 90-luvulla julkaistu kirja nimeltään "Consciousness Explained", jossa siis Dennett oikeastaan päätyy selittämään, että tietoisuutta ei ole olemassa alkuunkaan. Mutta kyllä suurin osa filosofeista ja aivotutkijoista nyt on sitä mieltä, että jonkinlainen tietoisuus kuitenkin meillä mitä ilmeisimmin on.

Ehkä suurin tieteellinen ennustusvoima on tällaisella ajattelun kaksoisprosessointiteorioilla, joiden mukaan ihmisen mieli jakaantuu kahteen erityyppiseen yksikköön. Eli meillä on tietoinen systeemi kakkonen, joka on hyvin kapea, että ihminen pystyy tietoisesti käsittelemään vain noin 3-5 informaatioyksikköä kerrallaan.

Ja sitten meillä on tämä systeemi ykkönen, joka on valtavan laaja assosiaatioverkosto. Eli jokaista tietoista ajatusta kohti vähimmilläänkin tapahtuu 280 000 jonkintyyppistä assosiaatioprosessia tuolla. Mistä meillä ei ole tuon taivaallista aavistustakaan, kunnes ne pulpahtaa päähän. Sitten jonain oivalluksena tai tai me herätään keskellä yötä miettimään jotain: "Ai niin vitsit, unohin lähettää sen sähköpostin!"

Tai näin, mutta se on laskennallinen alaraja, joka perustuu itse asiassa vasta meidän aisti- ja liikehermoston, afferentin ja efferentin hermoston rakenteeseen. Laskennallinen yläraja karkaa sitten jo tuonne satojen tuhansiin miljardeihin assosiaatioprosesseihin. Mutta se on ehkä se suurin oivallus viime vuosilta, jolla on aika paljon myös empiiristä näyttöä: Meillä on aika kapea tällainen kyky suunnata

tarkkaavaisuutta asioihin. Yksi on, että pystytään reflektiiviseen itsereflektiiviseen ajatteluun ja se on ehkä se, mikä erottaa tällä hetkellä ihmiset tekoälystä.

Mä voin.. Tässä kun mä puhun, niin mä en valitse näitä sanoja jostain sanastosta, vaan nää tulee jostain just sieltä systeemi ykkösestä sen assosiaatiorakenteen pohjalta. Mutta mä voin pysähtyä jossain vaiheessa, jos mä vaikka sanoin hassusti: "sori, en mä tarkoittanu silleen". Ja sit toinen on tämmönen algoritminen ajattelu eli päättely. Kyky hahmottaa, että okei: "yksi lasi, kaks lasia" – tehdä tällaisia erotteluja.

Tietoisuuden osalta viime kädessä siis kyse on ihmisen, eläimen, olion kyvystä itsereflektiivisyyteen, reflektiivisyyteen ja tällaiseen erottelevaan ja algoritmiseen ajatteluun.

Suvi: Aivan, ja scifissähän nämä tietoiset koneet on jo ollut monta kymmentä vuotta. Esimerkiksi klassisia esimerkkejä on Terminaattorit ja muut, missä tietoisesta koneesta yleensä seuraa jotain todella katastrofaalista ihmiskunnalle ja niitä ollaan aikaisemmissa jaksoissa käsitelty. Mutta meillä on tosiaan nyt jo älykkäitä koneita keskuudessa, mutta onko se tietoisuuden lisääminen näihin kyvykkyyksiin, mitä koneella jo nyt on? Onko se teknologiamielessä, algoritmin kannalta niin onko se ero iso vai pieni nykyiseen?

Lauri: No sitähän ei tiedetä.

Suvi: Sitä ei tiedetä.

Lauri: Jos ajatellaan ensin psykologian puolta, että kaksoisprosessointiteorioiden mukaan meillä on tosiaan siis kaksi erityyppistä taas ajattelumoduulia. Meillä on tietoinen systeemi kakkonen, joka on semmonen toiminnanohjausjärjestelmä, jolla on itse asiassa aika vähän valtaa meidän toimintaan. Suurin osa sen toiminnasta on autonomista ja se selittää tämän koko tällaisen inhimillisen eksistentiaalisen kriisin, että miksi me tehdään tyhmyyksiä ja ei aina tehdä sitä mitä halutaan. Platonista Freudiin tätä on pohdittu, että miksi, kun ihminen on rationaalinen olento, niin miksi me ei kuitenkaan toimita rationaalisesti suurinta osaa aikaan?

Vielä joku 10-15 vuotta sitten aika paljon kirjoitettiin siitä, että systeemi kakkonen olisi tavallaan tämä frontaalialueiden toiminnan ilmentymä. Mutta nyt me nähdään, että oikeastaan systeemi kakkonen, systeemi ykkönen molemmat aika holistisesti syntyy erilaisista aivotoiminnan takaisinkytköksistä. Sen lisäksi, että meillä on 86 miljardia neuronia aivoissa, meillä on 500 miljoonaa täällä vatsan seudulla. Esimerkiksi serotoniinituotanto, joka on mielihyvä hormoni, joka vaikuttaa siihen, että kuinka hyvä olla

ihmisellä on, niin siitä suurin osa tapahtuu näissä hermoverkoissa mitkä on täällä suolistossa. Se, että mitä kohtaa tässä pitäisi lähteä mallintamaan, on ehkä se kysymys.

Niin sitten yksi kiinnostava juttu tietoisuudessa on siis se, että kun tietoisuuttahan pidetään yleensä ihmisen yksityisimpänä asiana, se on se mun sisäinen ääni mun pään sisässä. Mutta itse asiassa kaikki se mistä sä olet tietoinen, niin kaiken tietoisuuden tavallaan intentionaalisen kohteen. Siis se mitä kohti se suuntautuu, vesilasi tai sanotaan – ovenkahva, niin kaiken tämän pitää olla jaettua. Eli itse asiassa meidän kyky olla tietoisia, niin se syntyy yhteistyön tarpeesta ja mistään sellaista asiasta millä sulla ei ole käsitettä eli sellaista jaettavaa, joka voidaan kielellisesti ilmaista, niin sellaista asiasta ei voi olla tietoinen.

Se on vähän paradoksi, että se meidän kaikista yksityisin on itse asiassa kaikista voimakkaimmin jaettua, koska jotta me voidaan olla tietoisia jostakin asiasta, niin meidän pitää pystyä jakamaan se. Tästä seuraa siis se, että myös sosiaalinen vuorovaikutus on keskeinen tekijä tietoisuuden synnyssä. Tästä syystä se, että mitä meidän pitäisi tarkkaan ottaa mallintaa, niin on aika avoin kysymys.

Suvi: Katsotaan vähän, miten nämä erilaiset scifitarinat käsittelevät tätä kysymystä tietoisuudesta ja mielen synnystä ja esimerkkinä Westworld-sarja, jossa on tällainen villin lännen teemapuisto, jossa on työntekijöinä ihmisen näköisiä androideja, niin sanottuja 'hosteja'. Ja sitten nämä rikkaat ihmiset, asiakkaat, jotka pääsevät tänne puistoon niin ne saa mennä ja sikailla menemään miten ne haluaa. Saa tehdä näille hosteille mitä ikinä haluavatkaan. Alussa nämä androidit toimii kuten ne on ohjelmoitu, eli ne toteuttavat tällaisia ennalta määritettyjä tarinaluuppeja ja jokaisen päivän jälkeen niiden muistit tyhjennetään. Mutta pikkuhiljaa kun se sarja etenee niin heille sitten jääkin tämä oma elämäntarina mieleen. Ja Westworld-sarjassa tämän tietoisuuden syntyminen liitetään nimenomaan tähän omaan elämäntarinaan ja muistojen syntymiseen ja oman itsensä muistamiseen. Miltä tämä ajatus kuulostaa?

Lauri: Se on musta tosi kiinnostava lähestymistapa ja tietyssä mielessähän tuossa ehkä pitää pitää erillään kaksi asiaa, että on erikseen tietoisuus ja sitten on yksilön identiteetti, joka edellyttää tietoisuutta. Siis myöskin siinä tilanteessa, kun nämä hostit nollataan aamulla – mutta siitä ei synny sitä elämäntarinaa, niin siitä ei synny sellaista identiteettikokemusta, että "okei, minä olen kokenut ja minusta on tullut minä siksi, että x, y ja z".

Tämä on just yksi syy siihen, miksi mä siinä omassa kirjassa kanssa päädyin siihen, että dynaamisen muistin mallintaminen on luultavasti kuitenkin ennakkoehto siihen ennen kuin se olisi meidän kaltaista tietoisuutta voidaan sanoa olevan olemassa.

Musta oli kiinnostavaa – siis siitä on tovi, kun olen sen katsonut sen ekan kauden, mutta mä muistan, että siinä on se Maeve, joka alkoi tavallaan heräämään sieltä, että sille alkoi tulla niitä flasareita. Niillä oli joku pilvi-virtuaalisysteemi siellä tai jotain ja sit se alkoi leviämään pikkuhiljaa. Ne alkoi ne muutkin hostit herätä siinä. Mut se on siis musta ihan kiinnostava skenaario kaikenkaikkiaan.

Ja sit tietysti se, jos ajatellaan se, että missä vaiheessa me voidaan ajatella, että joku tällainen ihmisen toimintaa simuloiva entiteetti on ikään kuin kone ja missä vaiheessa se ylittää sen koneen.. Siis tavallaan työkaluroolin Ja siitä tuleekin suvereeni entiteetti,

agentti, siis toimija. Ja se on siis kysymys, joka siis vielä tosiaan kolme vuotta sitten oli sciencefictionia. Nyt se on kysymys, minkä meidän pitää olla tosi valppaana koko ajan.

Suvi: Joo. Sitten Westworldissa vielä kärsimyksen merkitys tuodaan tärkeäksi tämän tietoisuuden tai identiteetin syntymiselle. Onko sillä roolia?

Lauri: Siis sanotaan, että kärsimys liittyy tietenkin siis jollakin tavalla arvoihin ja arvostuksiin. Että jotta voi olla kärsimystä, niin pitää olla joku käsitys hyvästä ja pahasta. Nykyisillä kielimalleillahan on arvot samalla lailla kuin ihmiselläkin ja ne ei ole välttämättä julkilausuttuja myöskään taas sen enempää kuin suurimmalla osalla ihmisistä, koska harva ihminenkään itseasiassa tietää, missä se oma arvomaailma menee.

Mutta ne kyllä näkyy ja ilmenee siinä, että jos ajatellaan vaikka, että ChatGPT:stä ollut paljon juttua viime aikoina, että se pyrkii vähän liiankin kanssa miellyttämään. Se miellyttämisenhalu on yksi arvostelma siellä ja silloinhan kun tavallaan ihmisellä on arvot eli käsitys hyvästä ja pahasta, niin silloinhan jos mennään sinne pahan sektorille, niin silloin syntyy kärsimystä.

Sanotaan, että ehkä lyhyesti vastattuna niin, että se voi toimia katalyyttina, jos se järjestelmä on muuten piuhotettu sillä, että siellä on kyvykkyys tällaiseen sisäänpäin kääntymiseen tai itsereflektiivisyyteen.

Suvi: Entäs sitten ihmisen kannalta, kun tässä esimerkiksi Westworld-sarjassa tietenkin, kuten introsta voidaan tulkita niin. Katsoja on tyypillisesti näiden hostien puolella. Ne ihmiset on ne niin sanotusti pahat siinä ja katsoja alkaa kokemaan empatiaa näitä hosteja kohtaan. Niin onko se fiktiivinen ajatus, että me kuvittelemme tietävämme, että ne on tietoisia, niin ennakkoehto sille, että aletaan tuntea empatiaa niitä kohtaan?

Lauri: Ei.

Suvi: Ei.

Lauri: Siis ihminenhän voi olla empaattinen vaikka ilmapalloa kohtaan. Siis sanotaan, että sulla on joku lempilelu ja joku kohtelee sitä kaltoin. Niin tulee semmoinen olo, että nyt tämä on..

Suvi: Joo, totta!

Lauri: Sitä lelua sattuu. Ei se vaadi mitään sellaista kognitiivista tietoisuusolettamaa, vaan meillä on itseasiassa aika voimakas sosiaalinen perusluonne on se, että me ulotetaan se aika laajalle se meidän empatiakyvykkyys, jollei sitä sit erikseen väkisin aja alas. Tai sit joillekin traumojen takia sulje pois.

Se Westworld on hyvä esimerkki. Toinen mikä on siis Steven Spielbergin AI-leffassa, vaan se tämmöinen meat fair mikä se oli – lihativoli-kohtaus missä siis näitä robotteja

tuhotaan ja ihmiset hurraa siinä ympärillä ja sehän pelaa just tuolla samalla, että katsojahan kokee sen, että tää on ihan kauheeta. Kun se söötti pikkupoika tuodaan siihen ja siellä joku happo meinataan kaataa sen päälle, niin siinä tulee sellainen että "Ei, älkää!" Että me ajatellaan, että se on on pikkupoika. Että jos siinä ois leivänpaahdin, niin ei me ajateltais sitä samalla tavalla.

Mutta siinä on kaksi asiaa. Yksi on se, että jos se ikään kuin entiteetti, jota kohti kohdistuu, kohdistuu vaikka väkivaltaa, on niin kuin jollakin tavalla meistä miellyttävä tai viehättävä tai sitten joku lempinalle tai..

Suvi: Ilmapallo, mä jäin kiinni siihen.

Lauri: Mä en tiedä mistä se ilmapallo tuli mun mieleen. Mut mä mietin sitä Tom Hanks - leffaa, missä se jutteli tolle lentopallolle,tai mikä se on – Wilson. Ja tää on muuten kiinnostava, että ihminen voi luoda siis tällaisessa pakollisessa tilanteessa ihmissuhteen lentopallon kanssa. Niin ihminen pystyy myöskin luomaan ihmissuhteen tekoälyratkaisunjen kanssa. Tämähän on ihan todellisuutta.

Suvi: Ja tuo on se kysymys, mitä esimerkiksi tässä Ex Machina elokuvassa käsitellään. Eli 2014 julkaistiin elokuva Ex Machina, jossa on tällainen tekoälyrobotti Ava. Ja siinä esitellään itse asiassa tällainen uudenlainen Turingin testi. Ja Turingin testissä tietysti testataan sitä, että chat-keskustelun perusteella, pystyykö ihminen tulkitsemaan onko kyseessä ihminen vai kone. Tässä elokuvassa käännetään toisinpäin. Eli jos ihminen tietää juttelevansa koneen kanssa, niin uskooko hän silti, että kyseessä on tietoinen olio. Ja sen elokuvan aikana katsojalle ei välttämättä selviä, että onko Ava tietoinen vai ei. Mutta sillä ei välttämättä ole edes merkitystä. Mutta mistä ylipäätään tietää, että ollaanko me oikeasti itsekään tietoisia vai ollaanko tämmöisiä tavallaan puhetta suoltavia zombeja?

Lauri: Siis ainoa asia mikä voidaan tietää varmuudella on se, että että sä itse olet tietoinen. Mutta se että mistä voi tietää sen, että onko joku toinen tietoinen? Teknisesti ottaen ei oikeastaan mistään. Niin ja tämä on David Chalmersin tunnetuksi tekemä zombi-argumentti, että käytännössä meillä ei ole mitään kriteeriä, jolla voidaan varmuudella sanoa, että toinen tietoiselta vaikuttava olento on oikeasti tietoinen eikä vain automaatti, joka niin nerokkaasti simuloi olevansa tietoinen.

Niin, on tietysti teoriassa mahdollista, että maailmassa olisi vain yksi tietoinen entiteetti eli sinä tai sinä, joka tätä seuraat ja kuuntelet tätä podcastia ja kaikki me muut ollaan zombeja. Eli mitään sellaista kriteeriä ei ole sille, että miksi mä en pystyisi tätä tekstiä muodostamaan mitä mä puhun vaan täysin automaattioperustein. Se, mikä tästä tekee kiehtovan.. Turingin testihän on läpäisty jo useampaan otteeseen näillä uusilla kielimalleilla. Ja sitten taas tämä Garlandin koe eli tämä käänteisversio siitä. Niin me ollaan siis oikeastaan päivittäin tekemisissä sen kanssa. Mä mietin vaikka, että kun automatkalla pistän ChatGPT:n puhemoodin päälle ja juttelen sen kanssa, se kokemus on oikeastaan identtinen sen kanssa, että mulla ois puhelu auki jonkun toisen ihmisen kanssa ja sittenhän on näitä tällaisia mihin mä viittasin mun kirjassa.

Lauri: Tää Blake Lemoine, Googlen insinööri, joka siis ajatteli, että tämmöinen lambda-tekoälyratkaisu, jonka hän pystyi keskustelemaan, joka väitti olevansa tietoinen, niin se riittää, että väittää olevansa tietoinen. No nyt tietysti näissä uusissa tekoälyratkaisuissa: ChatGPT on sananarvauskone, että se tosiaan vain tilastollisesti sylkee sanoja sanojen perään ulos. Niin voitaisiin nyt ajatella, että eihän se ole tietoinen. Mutta aika usein mun luennolla kysyn sitten mun opiskelijalta esimerkiksi, että onko kukaan ikinä sanonut mitään tyhmää ja sitten kaikki viittaa. Ja siis itse asiassa ihmisten systeemi ykkönenhän toimii hyvin samalla tavalla kuin ChatGPT eli assosiaatioperiaatteen perusteella.. Tulee näit ulos näitä sanoja jossain järjestyksessä, eikä meillä ole sellaista tietoista kontrollia.

Siis ihminen ei ensin muotoile jonnekin työmuistiin valmiiksi lausetta ja sitten tulosta sitä sanomalla vaan tämä on tällaista jatkuvaa, vähän niin kuin improvisointia. Ja se tietoisuus ihmiselläkin on, kun se on niin paljon kapeampi kuin mitä me yleensä ajatellaan. Se on oikeastaan enemmän vain se toiminnanohjausjärjestelmä, joka pystyy tarvittaessa pikkaisen tavallaan uudelleenohjaamaan suuntaa. Jos tulee möläytettyä jotain, niin pysäyttämään sen: "En mä sitä tarkoittanut tuolla tavalla."

Suvi: Tai ainakin periaatteessa pystyy.

Lauri: Niin joissakin tilanteissa.

Suvi: Mutta siis tämä Googlen insinööri sai fudut väitteestään ja myös Ex Machina -leffassa Ava saa vakuutettua olevansa tällainen tietoinen olento ja siinäkin käy kehnosti lopussa ihmisille, samoin kuin tälle Googlen insinöörille, niin olisiko hyvä strategia aina epäillä tietoisuutta? Jos huomaa ajattelevansa, että ChatGPT itse asiassa tuntee mut ja kokee mua kohtaan empatiaa, niin kannattaako siinä kohti aina epäillä?

Lauri: No joo ja ei. Siis tottakai kannattaa epäillä. Siis niin pitkään kuin sellaista konklusiivista näyttöä siitä tietoisuudesta ei ole. Tai on selkeää syytä miksi luultavasti tietoisuutta ei vielä ole. Niin kannattaa olla varovainen ja kriittinen, mutta tavallaan siis esimerkiksi mun kirjassa mä päädyin siihen, että jos nyt sitten niin käy, että me luodaan järjestelmä, joka pystyy tällaiseen itsenäiseen suvereniteettiin, agentuuriin, niin tavallaan, että jos se kävelee kuin ankka ja vaakkuu kuin ankka, niin sitten pitäisi ajatella et se on ankka.

Ja koska tämä palautuu siihen Chalmersin zombie-hypoteesiin, että se on eettisesti hyvin vaarallinen positio ruveta ajattelemaan.. Siis puhutaan tämmöisestä solipsismista, jossa ajatellaan, että on itse ainoa merkityksellinen olento ja kaikki muut on vain jotain zombeja tai vähempiarvoisia. Siis yksi nykyaikainen versiohan tästä on tämä tämä simulaatiohypoteesi. Että on siis ihmisiä jotka ajattelee että he on tällaisessa simulaatiossa tai pelissä. Ja tämä on kaikki vaan pelkästään taas peliä ja leikkiä.

Minusta Juhani Mykkänen kirjoitti jokunen vuosi sitten Hesariin hyvän artikkelin siitä, että Elon Musk kuvittelee olevansa niinkuin Ready Player One – tällaisessa Elon Muskin simulaatiomaailmassa. Jos sä heräät joka aamu ja huomaat olevasi maailman rikkain mies ja kaikki tuntuu vaan menevän silleen lapaan, ihan sama mitä tekee, niin kyllähän

siinä voi syntyä sellainen käsitys siitä, että kun on vielä tällainen kuin IT-alalla toimiva, jossa näkee kuinka komplekseja algoritmeja nykyään pystytään tekemään, niin voi syntyä semmonen: "Ehkä tää on vaan kaikki peliä."

Ja nyt kun katsoo Muskin toimintaa, niin sehän on aika lailla kyllä johdonmukainen tästä perspektiivistä katsottuna. Mutta sitten myöskin musta hyvä eettinen varoitus meille siitä, että lähtökohtaisesti oletus siitä, että itse olisi maailmankaikkeuden keskipiste, niin ei ole eettisesti kauhean kestävä.

Eli sen takia meillä on eettinenkin velvoite ajatella, että se tietoisuus ei ole vain oma yksinoikeus. Sitten jonkin ihmisen kohdalla on helposti empatian kautta ajatella, koska me ollaan aika samantyyppisiä, toimitaan samalla tavalla ja koetaan samanlaisia asioita ja puhutaan samanlaisista asioista. Niin okei, ajatellaan nyt sitten, että meillä on kaikilla joku samantyyppinen tietoisuus. Fine.

Eli tämä tietenkin tavallaan laajenee myöskin eläinten oikeuksiin, joka on konstikas.. Joka tavallaan edellyttäisi sitä, että ne eläimet pitäisi pystyä tuomaan osaksi meidän kulttuuripiiriä. Että me voitais tämmösiä arvostelmia niitä tehdä. Ja siihen perustuu siis edelleenkin se, että ajatellaan että eläimet on jotenkin, että se on ihan ok syödä possuja, vaikka possuilla on keskimäärin vaikka viisivuotiaan lapsen älykkyys nykytutkimuksen mukaan.

Mutta sitten sen chatGPT:n kanssa. Niin okei, senkin kanssa pystyy kyllä jutteleen ja se pystyy reflektoimaan ja se pystyy niinku ilmaisee asioita. No voisiko se olla? Mä en ehkä vielä menisi niin pitkälle, että se on. Mutta sanotaan. Mä en myöskään menisi niin pitkälle, että mä sanoisin, että se ei missään tapauksessa ole mahdollista. Mä oon monta kertaa siis tähän liittyen vitsaillu.

Noissa mun koulutuksessa ja opetuksessa siitä, että on kolme syytä miksi tekoälylle kannattaa olla ystävällinen. Ensimmäinen on siis se vaan, että ihan tutkitusti, että jos muotoilee syötteet sille, että sanoo 'kiitos' ja 'moi' ja 'mitä kuuluu' ja tälleen, niin silloin ne vastaukset on laadukkaampia. Joka ilmeisesti johtuu osittain siitä, että niin iso osa harjoitusaineistosta on tällaisia foorumikeskusteluja. Kun se tilastollisesti peilaa sitä, että mitkä sanajoukot esiintyy todennäköisimmin sen sanajoukon perään. Kun sanotaan tämmösiä ystävällisiä ilmaisuja, niin sitten usein foorumeissakin ihmiset vastaavat vaikka laajemmin.

Toinen syy on siis se, että koska meidän käyttäytyminen on plastista, eli se tarkoittaa, että mitä useammin me toimitaan tietyllä tavalla, niin sitä tyypillisemmäksi se muuttuu osaksi meidän käyttäytymistä. Jos me puhutaan tosi töksäyttäen, niin silloin me aletaan helposti puhua myös toisillemme töksäyttäen. Ja sitten kolmas on se, että mä nyt hirveästi vielä usko, että tällainen koneiden vallankumous olisi tuloillaan tässä. Mutta jos se nyt sitten sattuu tulemaan.. Niin ois tietenkin kiva, että niillä ois hyvät muistot meistä.

Suvi: Niin kyllä.

Lauri: Tämä on vähän niin kuin vitsihuumoria, mutta se pointti on paljon syvällisempi siinä. Yleensä tavallaan porukka tuon kolmoskohdan kohdalla nauraa vähän hermostuneesti, mutta se pointti ei oikeastaan ole se, että turvataan selusta koneiden vallankumousta varten, vaan se, että me opittaisiin ajattelemaan, että minkälaista elämä on maailmassa, jossa on muitakin suvereeneja toimijoita kuin ihmiset ja se on ihan mahdollista, että meidän elinaikana tällainen maailma on tuloillaan.

Suvi: No onko se tavoite ja päämäärä kiinnostava asia tuossa – eli esimerkiksi tässä Ex Machinassa tämä Ava muodostaa itselleen ohjelmoinnin ulkopuolisen tavoitteen ja päämäärän, eli 'hän' haluaa pois sieltä tai 'se' haluaa pois sieltä kartanosta missä se on ikään kuin vankina. Mutta onko tämä ohjelmoinnin ulkopuolisen päämäärän asettaminen merkki tietoisuudesta?

Lauri: No ei kyllä oikeastaan millään tavalla mun mielestä. Koska siis jo nyt, jos ajatellaan sitä, että tällaisessa nykyisessä kielimallissa niin on jossain plus-miinus tuhannen miljardin parametrin luokkaa se kompleksisuus. Niin siis ilman muuta sieltä syntyy päämääriä, jotka ei ole.. Siis niissähän ei ole ohjelmointirakennetta sillä tavalla, että siellä olisi luotu joku tämmöinen tavallaan niinku tavoitteenasettelu tai jokin ehtolauseiden joukko, että jos tapahtuu näin, teet noin.

Mun käsitys tällä hetkellä on, että kyse on matalan fideliteetin ihmisen tiedostamattoman mielen simulaatiosta. Se on matalan fideliteetin siinä mielessä, että se ei vielä samalla nyanssitasolla kykene niiden assosiaatiopolkujen mallintamiseen kuin mitä meidän aivoissa tapahtuu. Se mikä se kompleksisuuden määrä pitäisi olla, niin vähimmillään se nyt on varmaan luokkaa satatuhatta miljardia yksikköä, koska meillä on about sen verran synapseja meidän aivoissa. Mutta luultavasti se on vielä siitä kompleksisempi, koska meillä on sen lisäksi muutama tusina erilaisia välittäjäaineita, jotka säätelee sitä, että kuinka hyvin signaali kulkee synapsin läpi eli neuroimpulssi. Ja sitten sen lisäksi vielä se, että ihmisen aivoissa on esimerkiksi gliasoluja, minkä funktiota me ei oikeastaan hirveästi tiedetä, että mitä ne siellä luuhaa ja niitäkin on aika paljon siellä neuronien joukossa.

Ja sitten vielä on tällaisia niinkun Roger Penrose on esittänyt, joka on fyysikko taustaltaan. Mutta se on yksi näitä kuuluisimpia tällaisia tekoälykriitikoita. Ja sen argumentti on se, että on tällasia kvanttitason.. Niinku neuronirakenteessa on tiettyjä, että neuronissa on tiettyjä rakenneosia, joilla on tällaisia kvanttitason vuorovaikutussuhteita. Eli sitten se lähtee ihan lapasesta se.

Niin jos me mennään siitä, että me ollaan noin tuhannen miljardin parametri kompleksisuudessa, nykyjärjestelmissä ja ihmisen aivoissa on ehkä jossain satojentuhansien miljardien parametrien luokkaa vähimmillään, niin se on vielä karkean fideliteetin simulaatio, mutta simulaatio kaikki tyyni kuitenkin.

Ja se, että jos ajatellaan että me tuossa voidaan simuloida tuolla tavalla niin kuin näytöllä sitä, että tuossa on tuollainen kuva jostain piirilevystä. Niin siis itse asiassa tuossahan on vain pikseleitä rivissä, niin se, että me voidaan pikseleillä luoda tämä kuvakäsitys ja me voidaan neuroverkkomallilla luoda se mielikäsitys, niin se on se

tilanne missä me ollaan tällä hetkellä ja sen takia siis ilman muuta, että siis kielimallit pystyy luomaan niin kuin lainausmerkeissä ohjelmointinsa ulkopuolisia tavoitteita, koska ei niillä oikeastaan ole ohjelmoitu sinne mitään tavoitteita, vaan se mitä se on koulutettu on se, että se on.. Ensin on pyritty laadullistamaan se koulutusaineisto, että siellä olisi mahdollisimman vähän hötöä, koska kävi ilmi aikanaan, että internetillä ei kannata kouluttaa tekoälyä, koska niistä tulee yleensä misogynistisiä natseja, jos sen tekee sillä tavalla.

Suvi: Yllättävämpää se, että se oli yllätys jollekin.

Lauri: Niin just näin. Mutta se lopputulos mitä sieltä tulee, niin sehän on viime kädessä kiinni siitä, että mikä on sen ihmisen ja sen tekoälyn keskinäinen vuorovaikutus? Yhden sanan muuttaminen syötteessä saattaa muuttaa radikaalisti sen kaiken, mitä sieltä tulee ulos. Ja nyt siis tavallaan se tekoäly ei itse tee. Se agentuurihan puuttuu siltä, eli nehän ei tee mitään jos se ihminen pyydä.

Suvi: No mutta tuossa on hyvä hyvä aasinsilta Star Trek The Next Generationiin. Siinä on tämmöinen poikkeuksellisen sympaattinen androidrobotti Data nimeltään, joka on reipas ja ahkera miehistön jäsen. Mutta se Datalla on vaikeuksia ymmärtää ihmisten välisiä sosiaalisia kanssakäymisiä, koska Data ei tunne tunteita. Ja siinä sarjassa nimenomaan tämä kyky tuntea esitellään niinkuin keskeisenä erona ihmisten ja koneiden välillä. Onko tunteet sellainen tekijä, joka meidät erottaa?

Lauri: Tunteet on tosi jännä juttu, koska tunteethan on fundamentaalinen kaikille kognitiolle. Ei ihminen järjelläkään tee mitään, jos ei se tunnu miltään. Tunteet on fundamentaalisempia kuin edes nämä perityt ja opitut prosessit. Eli itse asiassa se intuitio on se meidän automaatiokyvykkyys, joka me ollaan opittu. Sitten me ollaan geneettisesti ohjautunut energiansaannin maksimointia ja suvunjatkamista ja tämän tyyppisiä juttuja, jotka perustuu siihen, että luonnontilassa siis mitä evoluutio on meistä tuottanu.

Suvi: Ja kerro vielä opituista prosesseista pari esimerkkiä.

Lauri: No sanotaan vaikka pianonsoitto tai kielenkäyttö, kirjoittaminen. Siis tää mitä me tehdään nyt, eiks niin? Nää ei oo sellaisia, että me synnytään ja pystytään käymään keskustelua vaan meidän pitänyt lukea paljon, katsoa kiinnostavia elokuvia, tv-sarjoja ja sit me voidaan käydä tällainen keskustelu. Se muokkaa siis tätä meidän hermostoa sillee, että tämä hermosto voi tuottaa tällaisia suun liikkeitä.

Siis tunteet on sen koko järjestelmän perusasia, joka ohjaa sitä siihen, että onko jokin asia oikein tai väärin, tai hyvä tai huono. Siinä mielessä siis esimerkiksi just tämä, että tämmöinen RLHF hienosäädetty kielimalli, niin sillähän on tietyssä mielessä, sillä on ainakin arvomaailma varmasti olemassa. Se, että onko se sillä tunteet.. Tunteethan edellyttää kokijan, eikö niin? Jonkun, joka siis kokee sen toiminnan jonkinlaisena ja tämä palautuu sitten siihen tietoisuuskysymykseen.

Ja sitten taas Datan tapauksessa niin Datahan on myös Pinokkio. Siis sehän on, sehän haluaa olla oikea poika, eiks niin? Jos se sais sen tunnesirun joskus, niin sit se voisi oikeasti tulla ihmisen kaltaiseksi. Mä muistan sen jakson, kun se sai sen sirun ja sitten se meni aluksi ihan sekaisin, kun ne kaikki tulee niin valtavina.. Ja sitten se itkee ja nauraa peräkanaa. Se tarkoittaa sitä, että se meidän kokemuksen nyanssien määrä kasvaa sitä mukaa, kun me pystytään tuntemaan asioita.

Ja onhan tuossakin ihmisellä valtava hajonta. Kouluikäisenä Oulussa nuoren miehen piti olla aina kauhean kova jätkä ja ajatus siitä, että itkisi jossain leffassa oli ihan posketon. Kun mä muutin kotoa, niin mä katsoin sen Baz Luhrmanmin Romeo ja Julia -elokuvan on edellisenä iltana. Ku seuraavana aamuna sitten lähdin muuttokassien kanssa Helsinkiin. Ja sitten mä itkin silleen aivan niin kuin Niagaran putouksena. Kyky vaikka itkeä elokuvassa on itse asiassa aika valtavan hieno osa ihmisenä olemista. Jos sanotaan, että on kasvanut Oulussa 80-luvulla, ei välttämättä ehkä ihan itsestäänselvyys, että on sellaistakaan nyanssia olemassa.

Suvi: Mutta kun sä olit jo päättänyt lähteä Oulusta, niin sä päätit, että nyt voi nyt. Nyt voi itkeä elokuvalle. No, Star Trekissä kapteeni Picard toteaa, että ihmiset ovat vain toisenlaisia koneita, sähkökemiallisia luonteeltaan. Voiko ihmisiä ajatella koneina, jos ajatellaan, että älykäs androidi ja tunnesiru..

Lauri: Se on aika vaarallinen ajatus. Konehan on siis työkalu, joka on alisteinen ihmiselle. Mulla oli yksi kaveri, joka oli esilukijana sille konemelle kirjalle, kun mä käytän sitä vähän kieliposkessa sitä sananarvauskone-esimerkkiä, että tavallaanhan ihminenkin on sananarvauskone, koska niin iso osa meidän toiminnasta on autonomista. Mutta kun ihmisen toimintaperiaatehan ei perustu.. Siis ihminen ei ole mekaaninen järjestelmä. Aivot ei ole kone vaan aivot on puutarha. Eli käytännössä siis se, että puutarhassa se ne rönsyt mistä pidetään hyvää huolta niin ne kasvaa ja voi hyvin ja ne mitä laiminlyödään ne kuihtuu pois.

Ylipäätään elämähän on siis valtavan kompleksista. Ja just meidän kompleksisuuden määrä mitä on täällä. Siis se ylittää minkä tahansa matemaattisen mallin, mitä me ollaan koskaan ihmiskunnan historiassa pystytty kehittämään. Ja sitten jos ajatellaan se Penrosen esimerkkiä, niin on myös mahdollista, että se on laadullisesti siis niin kompleksinen, että sitten mennään "Turtles all the way down".

Sehän on hirveen tyypillinen ajatus ja vaikka 90-luvun neurotieteessä kantava ajatus, että tää on vaan tämmönen just sähkökemiallinen kone ja sieltähän se 90-luvulta on se Star Trek TNG:kin. Niin mun mielestä se ei ole hyvä metafora siis a) sen takia, että ihminen on paljon kompleksisempi kuin kone ja b) sen takia että ihmisen käsittäminen koneena johtaa isoihin eettisiin haasteisiin, kun meillä on jo tällaisia vähän niin kuin tavallaan uus-feodalismi tyyppistä, että on näitä teknoparoneita, jotka pystyy teettämään vaikka halpatyövoimaa Keniassa, vaikka ChatGPT:n tapauksessa, niin minusta eettisesti ei ole kyllä kovin kestävää lähteä tuollaiseen metaforaan.

Suvi: Mutta siitä tulee kysymys myös siitä tekoälyn tai koneen oikeuksista. Ja Star Trekissäkin on jakso, jossa tämän Datan oikeuksia pohditaan siinä, että voiko Datan

sulkea vastoin Datan tahtoa vai ei. Nyt me ollaan hyvin lähellä sitä kysymystä, että mikä on, mikä on koneiden oikeudet ja mikä on tekoälyn oikeudet verrattuna ihmisen oikeuksiin.

Lauri: Mutta siis minusta se vastaus on tuossa sun kysymyksessä, että voiko Datan sulkea ilman Datan tahtoa? Jos datalla on tahto, niin ei voi. Eli tavallaan se, että jos sä voit sanoa, että sillä entiteetillä on tahto, halu, niin silloinhan sitä sitten tavallaan siihen kohdistuu ihan samat eettiset kysymykset mihin tahansa muuhun entiteettiin, jolla on tahto tai halu.

Konstikas tästä tulee, ja se mistä ne tulee ne eksistentiaaliset kriisit syntymään – palaten jälleen siihen, että ei ole absoluuttista tapaa sanoa ihmisistäkään, että onko näillä tietoisuus vai ei, niin se että aina tulee olemaan niitä.. Siis katsotaan vaikka 1800-luvun tätä USAn sisällissotaa, tätä orjatalouden murtumista. Niin aina tulee olemaan, siis vieläkin niitä ihmisiä olemassa, jotka on sitä mieltä, että kyllähän nyt on noita alempiarvoisia ihmisiä. Mulla on oikeus niitä käyttää työvoimana maksamatta mitään.

Ja sitten toisaalta se porukka, jotka on siellä, että me ollaan jo automatisoitu, kaikki pyörii nämä tekoälyratkaisuilla, että eihän me nyt voida niille antaa ihmisoikeuksia, koska sitten meidän kaikki voitot ja muut haihtuvat taivaan tuuliin.

Niin tämmöinen tavallaan kulttuuritörmäyshän varmasti tulee siinä kohtaa käsiin. Mutta taas mä koen eettisestä näkökulmasta, että jos olio pystyy kaikin olennaisin osin toimimaan yhteiskunnassa samalla tavoin kuin ihminen, niin silloin sen täytyy olla samat oikeudet kuin ihmisellä. Ja jos ihminen ei halua, että siitä pannaan virrat pois, niin ei siitä toisestakaan oliosta voi silloin laittaa.

Ja toki sillä pienellä varauksella, että on riskinä taas se, että me ikään kuin luodaan liian tämmöinen antropomorfinen suhtautuminen nyt vaikka näihin nykyisiin systeemeihin. Kuitenkin niin pitkään kuin niillä ei ole autonomiaa tai agentuuria, niin silloin tämä kysymys on kuitenkin enemmän teoreettinen.

Ensimmäinen kysymys olisi tietenkin se, että kannattaako tällaista teologiaa ylipäätään kehittää, mutta se on sinänsä idealismia, koska jos ihminen jotain pystyy tekemään, niin kyllähän se sen tekee. Oikeastaan kysymys on juuri siitä, että nyt jos me jos eettiset toimijat laittaa.. Kun yritettiin sitä moratoriumia laittaa pari vuotta sitten pystyyn, että lopetettaisiin kehittäminen vähäksi aikaa, niin sehän tarkoittaa sitä, että ne epäeettiset toimijat.. Eihän ne lopeta sitä kehittämistä, vaan..

Suvi: Ne saa etumatkaa. Just niin.

Lauri: Eli sen takia itse asiassa kyllä eettisten toimijoiden täytyy niin raivokkaasti puskea eteenpäin kuin vain suinkin pystyy.

Suvi: Milloin meillä on tietoisia koneita käytössämme?

Lauri: Ehkä ensi vuonna, ehkä viiden vuoden päästä, ehkä ei koskaan.

Suvi: Mistä me huomataan, että koneista tulee tietoisia?

Lauri: Siitä, että ne kykenee samalla tavalla itsenäiseen toimintaan ja ajatteluun ja päättelyyn kuin ihminenkin. Se itsenäisyys on se avainsana. Tällä hetkellä tekoäly ei tee mitään, jos ei ihminen pyydä. Kaikki tällä hetkellä olevat tekoälyjärjestelmät. Niillä on aina joku rajattu alue – sovellusalue.

AGI taas tarkoittaa sellaista tekoälyratkaisua, joka pystyy kaikkeen mihin ihminenkin, eli pystyy laaja-alaisesti tekemään valtavan erilaisia asioita. Ideana on siis se, että kun ihminen pyrkii kehittämään ihmisen kaltaista konetta, niin sehän tarkoittaa, että jos me onnistutaan, niin ihminen voi tehdä yhtä älykkään koneen kuin hän on itse.

No sitten kysymys on jos ihminen pystyy tekemään koneen, joka on 0,00001 pykälää älykkäämpi kuin ihminen, niin se tarkoittaa, että ihminen pystyy tekemään itse älykkäämmän koneen. Ja silloin se ihmistä älykkäämpi kone pystyy tekemään itseään älykkäämmän koneen, joka pystyy tekemään itseään älykkäämmän koneen. Ja koska näiden vimpainten laskentateho on siis jossain miljardeja operaatioita sekunnissa, niin teoriassa sillä hetkellä, kun me kytketään virta tuollaiseen vimpaimeen, niin se saattaa pystyä skaalaamaan siis sekunneissa eksponentiaalisesti sen ikään kuin kyvykkyyden jonnekin mahdottomiin sfääreihin.

Ja voi olla, voi olla että ei, mutta sitä me ei oikeastaan tiedetä ennenkuin AGI-järjestelmä on syntynyt. Ja tosiaan optimistisemmat – Dario Amodei ja kumppanit – on sitä mieltä, että se olisi ensi vuonna tulossa. Mutta jos se on mahdollista niin kyllä se luultavasti piirun verran pidempään menee. On myös mahdollista, että me ei koskaan.. Että on niin paljon laskennallisia haasteita, että sellaista ei koskaan synny. Mutta se, että mitä sitten syntyy kun ihminen tekee ihmisen kaltaisen koneen, niin se on kyllä sitten ihan anybody's guess siinä kohtaa.

Suvi: Just niin. Jätetään nyt toistaiseksi konemielet ihan hetkeksi ja siirrytään käsittelemään ihmisen tietoisuutta ja ihmisten ihmisen mieltä ja sen teknologisia laajennuksia. Pysytään vielä Star Trekissä. Siinä on transporter-teknologia, jolla pystytään skannaamaan ihmisen atomit ja siirtämään tämä skannattu data valonnopeudella paikasta toiseen. Ja ajatuksena se, että sitten voidaan atomi kerrallaan tämä ihminen luoda uuteen paikkaan ja tämä tismalleen sama atomikonstruktio sitten tuottaa tismalleen saman tietoisuuden ja identiteetin ja mielen. Jos nyt unohdetaan se atomiteknologian kompleksisuus siitä, mutta miltä tämä ajatus kuulostaa, että jos meillä on ihan tismalleen samat atomit ihan tismalleen samassa järjestyksessä, niin tuottaako se tismalleen saman ihmisen?

Lauri: Ei, koska ihminen ei ole pelkästään vaan ne atomit mitä se on, vaan myöskin se on se sosiaalisympäristöön kytketty kokonaisuus. Ja tavallaan siis sehän on aika brutaali se Star Trekin teknologia, koska siis yksi tulkintapahan on siis se, että me kloonataan ihminen ja tapetaan alkuperänen. Ja se kloonattu ihminen tietenkin kokee olevansa.. Se Kirk joka tulee sinne planeetalle, sillä on sen vanhan Kirkin muistot ja se kokee olevansa James D. Kirk. Kaikkien muiden kokijoiden kannaltahan se on.. Se

muistaa se keskustelun, mikä käytiin viisi minuuttia sitten siellä Enterprisella ja se pystyy viittaamaan siihen. "Onhan se Good old Jim!" Mutta onko se näin? Vai onko se edellinen Kirk, joka meni siihen siihen  platformille, joka biimattiin alas, niin onko se kokenut: "Aaa, lakkasin olemasta!"

Ja tuossa täytyy aika paljon venyttää myöskin sitä, että missä se ihminen on ja tietoisuus ja onko se ne atomit vai voiko ne atomit siis sillä kvanttitasolla mallintaa, missä se säilyisi se jonkinlainen integriteetti. Mutta on siihen sitten olemassa suopeampikin tulkinta siitä, että onko se mieli tavallaan vaan sitten se, onko se minuus vaan tää lihaklöntti tässä ajassa ja paikassa. Vai onko se ennemminkin joku systeeminen ominaisuus, joka syntyy meidän vuorovaikutuksesta? Niin, silloin itse asiassa Star Trekin malli voi hyvin toimiakin, koska silloin siinä on se ikäänkuin se kausaaliketju, jossa siirrytään toisaalle.

Suvi: Ja tätä kysymystä pyöritellään myös Muuntohiili-kirjassa ja siihen siihen perustuvassa Altered Carbon -sarjassa. Siinä siis ihmisen mieli ja muistot ja muisti voidaan ladata tämmöiselle yhdelle disketille. Ja siinä siis ajatuksena on se, että aivoissa on jotain dataa tai ihmisessä on jotain dataa, joka yksiselitteisesti pystytään ottamaan pois ja siirtämään johonkin toiseen tällaiseen ruumiiseen tai "sukkaan". Mutta riittäisikö se pelkkä aivojen data simuloimaan ihmisen uudestaan? Vai tarvitaanko siihen juuri se hermosto tai muu ruumiillisuus mukaan?

Lauri: No toi on tommoinen klassinen scifi- ja tieteessäkin esiintyvä ajatus, että ihminen on tietokone, jossa on hardware: aivot. Ja sitten on se data eli miten aivot käsittelee, tai se software mitä se hardware ajaa. Ja tämähän ei nykytiedon varassa pidä paikkaansa, koska se hardware ja software on täysin yhteen kietoutunut.

Nyt kun sä sanoit ton asian, niin se on luonut siis mun kuulohermostossa aktivaation, joka synnyttää siellä tällaisen verkostorakenteen, joka sitten taas laukaisee kaikenlaisia erilaisia prosesseja, jotka saattaa olla.. Esimerkiksi Daniel Dennett puhuu siitä, että täällä on aliohjelmakilpailu käynnissä.. Niin sit loppujen lopuksi joku näistä kytkennöistä johtaa siihen, että syntyy aktivaatio tässä sensori-motorisella aivokuorella ja sitten mä louskutan leukoja tässä näin.

Software ja hardware on sama asia ihmisen tapauksessa, niin silloin se, että mitä me voitaisiin siihen kiekolle tallentaa on aika hyvä kysymys. Mutta on yksi kiinnostava näkökulma, mistä sitä Altered Carbonin maailmaa pystyisi lähtemään ehkä purkamaan, niin David Chalmers esittää sellaisen herkullisen ajatuksen siitä, että kun tavallaan tämä aivoskannaus juttu: siirretään ihminen simulaatioon tyyppinen, mikä on vaikka Upload-tv-sarjassa: Jos neuroni kerrallaan ulkoistetaan se. Että otetaan ensin tosta yksi neuroni ja sitten se siirretään siihen digitaalilaitteeseen, kun enemmistö siitä ajattelusta alkaa tapahtua siinä digitaalisessa prosessointijärjestelmässä. Ja sitten kun se koko systeemi on skannattu, niin sitä se on kaikki siellä. Mutta jos koko sen prosessin aikana pystyt koko ajan sanomaan, että "ei mitään, kaikki hyvin – tuntuu ihan samalta kuin ennenkin", niin silloinhan voisi ajatella, että ihmisen tietoisuus olisi ikään kuin säilyy.

No sitten Chalmers esittää kyllä siihen kaikenlaista problematiikkaa kanssa, että on myös mahdollista, että se tietoisuus alkaa taas pikkuhiljaa hälvenemään siinä sen prosessin aikana. Mutta koska se ihminen, koska se käyttäytymiskompleksisuus säilyy, niin se edelleenkin sanoo, että "kaikki tuntuu ihan hyvältä".

Päästään taas tähän zombi-problematiikkaa, mutta ongelmallista siinä on tietysti se, että kun se ei ole vain aivot, kuten sanoin aikaiemmin: Meillä on paljon itseasiassa hermostoprosessointia vaikka vatsan seudulla. Siinä mielessä se, että me voitais jollekin levylle se koko ihminen tallentaa, niin en oikein usko. Siis sanotaan, mitä me ollaan tähän mennessä käsitelty näitä esimerkkejä.. Ehkä se Altered Carbon on vähän vanhentunutta scifiä siinä mielessä.

Suvi: Joo kyllä, mutta jos mut erotettaisiin tästä fyysisestä "sukasta", niin onko minkäänlaista mahdollisuutta yksilöidä, tavallaan sarjanumeroida sitä mun mieltä tai tietoisuutta? Osataanko me yksilöidä ihmisen ajattelua tai tietoisuutta tai mieltä?

Lauri: Vielä toi on aika tommosta hypoteettista, mennään täysillä päin scifiä, että tuo on aika hypoteettinen tässä vaiheessa on. Mutta miksei, että jos meillä olisi vaikka kuvaresoluutioltaan tuhat kertaa tarkempi ja ajalliselta resoluutioltaan vaikka 100 kertaa tarkempi. FMRI-laite, niin voishan sillä varmaan sitten tunnistaa sellaisia yksilöllisiä.. Pyytää vaikka tekemään jotain tällaisia liikkeitä ja katsoa, että "haa, selvästi pianisti, kun tuolla tulee vaikka minkälaisia aktivaatioita". Aika teoreettista, mutta miksi ei.

Suvi: Altered Carbonissa se "sukitus" ja mielen kopiointi on valtavaa bisnestä ja sitten se on jossain määrin myös reguloitua ja yksi semmoinen tärkein sääntö siinä on se, että yksi mieli kerrallaan yhteen sukkaan. Mitä mielelle voisi tapahtua, jos ajatellaan, että meillä on kaksi mieltä kahdessa eri ruumiissa ja sitten me kuitenkin yritettäisiin ne jonnekin pilvipalveluun synkata? Mitä mielelle semmoisessa tilanteessa voisi käydä, jos sillä olisi kahdesta samanaikaisesti kaksia eri tuntemuksia ja kaksia eri ajatuksia?

Lauri: Se Alterd Carbonin ontologia on jotenkin niin kaukana mistään, mikä olisi realistista tällä hetkellä, että okei: mitä se tarkoittaa, että se mieli siirretään pilvipalveluun? Että voisihan ajatella, että se on tällainen taas jonkun x-fideliteetin simulaatio siitä mielestä, että se ei varsinaisesti ole se mieli tai se ihminen.

Ja jos se olisi tämmöinen tavallaan painokerrointyyppinen neuroverkkomalli, silloinhan se tarkoittaisi käytössä vaan sitä, että ne parametrit normalisoisi toisensa jollakin tavalla. Niinkuin mitä nämä nykyiset LMM:t on: tämmöinen neuroverkkomalli, missä on koodattu vaikka käsitteiden sisäinen syvärakenne kymmenien tuhansien parametrien tarkkuudella. Eli miten käsitteet suhteutuu toisiinsa, niin sitten sinulla on tiettyjä kokemuksia, tuntemuksia vaikka jostakin hyvästä asiasta tai pahasta asiasta ja sitten niillä on tietynlaiset hienovaraiset painokertoimet siellä, että käytännössähän siitä tuli siis jonkinlainen ikäänkuin symbioosi niistä kahdesta mielestä silloin tällaisella mallilla, että jos sulla on yhdessä vaikka parametri 778 on 1,2 ja toisessa parametri 778 on kokemuksen kautta pyörähtäny 0,8:ksi niin sieltä tulisi ykkönen ulos.

Mutta sitten se lopputulema on riippuen siitä kuinka paljon divergenssiä siinä on siinä niiden kokemuksessa. Sieltä voi tulla ulos joku ihan toinen tyyppi kokonaan. Niin tai sitten voi olla joku ihan täysin pöpi. Tai ettei se nyt itse asiassa nyt tolleen lähtee venyttämään, niin ei se mahdoton ajatus olisi. Mutta kyllä siinä aika monta tavallaan tällaista. Niin kuin Daniel Dennett sanoo, että mielentutkimuksessa paljon sellaista, että "se toimii tällä tavalla.. and then the magic happens." Niin on tossa aika monta semmoista. "And the magic happens" momenttia kuitenkin.

Suvi: "Magic happens" tekee aika paljon duunia tässä.

Lauri: Scifissähän on siis tää MacGuffin.. Aina on joku semmoinen asia, joka vaan on "just because".

Suvi: Nopeat kysymykset loppuun. Eli ota sun systeemi ykkönen käyttöön. Lauri, ottaisitko tutkimusassariksi mieluummin Star Trackin datan vai Ex Machinan Avan.

Lauri: Datan ilman muuta. Sehän on science officer muistaakseni siinä Enterpraisellakin. Mä luulen, että tutkimusassariksi voisi kuitenkin toimi.

Suvi: Entäs menisitkö mieluummin Star Trackin transportteriin vai upgreidattuun sukkaan Altered Carbonissa?

Lauri: Jaa-a.

Suvi: Systeemi 1!

Lauri: No okei, jos on pakko niin ehkä se sukka tulisi mieluummin. Mä luulen, että se transportteri kuitenkin tuhoaisi mut.

Suvi: Kopioisitko mieluummin oman tietoisuutesi koneelle vai lisäsitkö omien aivojesi kyvykkyyttä tekoälyimplantin avulla?

Lauri: No ehdottomasti jälkimmäinen, koska sitä mä teen jo nyt. Mä oon joskus vitsaillut sillä, että jos joku pyyhkisi mun pilvipalvelut tyhjiksi ja veis mun kännykän niin mun älykkyysosamäärä putoaisi 30 pistettä.

Suvi: Lauri, tosi paljon kiitoksia kun tulit meidän podcastvieraaksi! Ja suurkiitos myös kaikille katselijoille ja kuuntelijoille! Muistakaa kommentoida YouTubeen jakson alle, tämäkin jakson aihe oli katsojakommenteista poimittu.