

ARTO LAITINEN

## Kipua tunteva kone?

Pentti O. A. Haikonen, *Tietoisuus, tekoäly ja robotit*. Art House, Helsinki 2017. 295 s.

**T**ekoäly- ja robotiikkakeskustelussa on erotettavissa kärkeästi pelkistään kahdenlaisia lähestymistapoja. Arkinäkökulmasta robotit ovat koneita, joilta puuttuu lukuisia persoonuuteen vaadittavia piirteitä tietoisuudesta itsetietoisuuteen, kivun kokemuksista välittämiseen ja moraalituntoihin ja aidosta ymmärryksestä autonomiaan. Kybernetiikan ja kognitiotieteen näkökulmasta robotit ja ihmispersoonat ovat tietojenkäsittelysysteemejä, joilla on representaatioita ympäristöstään ja itsestään sekä kyky jossain mielessä autonomisesti saada aikaan haluamiaan muutoksia ympäristössä. Arkinäkökulmasta täytyy tapahtua laadullinen hyppy, jotta tulevaisuuden robotit voisivat olla tietoisia ja kokevia ihmispersoonien tapaan. Vaikka tällainen hyppy olisikin periaatteessa tulevaisuudessa mahdollinen, ainakaan nykyiset koneet eivät ole tietoisia. Kybernetisestä näkökulmasta mitään laadullista hyppyä ei tarvita: tietojenkäsittelysysteemit ovat erilaisia ja niillä on eri vahvuuksia, eikä fenomenaalinen, koettu tietoisuus ole välttämätöntä älykkäille funktioille.

Tekniikan tohtori Pentti O. A. Haikosen *Tietoisuus, tekoäly ja robotit* yhdistelee näitä tuttuja lähestymistapoja. Teoksen mukaan kokemuksellinen tietoisuus on tavoittelemisen arvoista, mutta se vaatii uudenlaista lähestymistapaa; digitaaliset symboliset koneet eivät tule tietoisiksi. Haikonen on väitellyt aiheesta vuonna 1999 ja julkaissut siitä englanniksi kolme kirjaa ja lukuisia artikkeleita. Hän on työskennellyt parikymmentä vuotta Nokian tutkimuskeskuksessa ja vuodesta



2009 alkaen ollut filosofian dosentti Illinoisin yliopistossa Springfieldissä.

*Tietoisuus, tekoäly ja robotit* käy mukaansatempaavasti läpi tekoälyn syntyhistoriaa, suuria odotuksia ja aiempia pettymyksiä. Haikonen ei usko nykyisenkään tekoälyn eksponentiaalisen kasvun jatkuvan vaan ajattelee sen taittuvan S-kuvion kaltaiseksi: edessä on taas nousun laantuminen vaakasuoremmaksi. Haikosen asianhallinta näkyy hienosti esimerkiksi silloin, kun kirjassa esitellään erilaisia käsityksiä aivoista: Puhelinkeskus? Tietokone? Luokittelija? Sääätösystemi? Sisäinen internet ja hakukone? Ennustaja ja simulaattori? Kartesiolainen teatteri homonculuksille?

### Tavoitteena suora havainto ilman symbolista tulkintaa

Haikoselle olennaista on symbolisen merkityksen ja alisymbolisen suoran merkityshavainnon välinen erottelu.

Mielihyvän tai kivun tunnistamiseen tai ulottuvilla olevan ruoan etäisyyden havaitsemiseen ei tarvita symboleja eikä tulkintaa. Tietoinen olento voi kokea ne suoraan (neuronien ja neurosignaalien synnyttäessä tietoisin elämyksen). Nämä alisymboliset, suoraan tunnistettavat sisällöt ovat tietoisuudelle keskeisiä. Symboliset, digitaaliset tietokoneet eivät kuitenkaan ole tietoisia, eivätkä ihmisaivot puolestaan ole digitaalisia tietokoneita. Vaikka tietokoneet, tekoäly ja robotit simuloisivat ihmisten toimintoja ja läpäisisivät Turingin testin, ne eivät itse olisi tietoisia. (Turingin testissä ihminen yrittää kysymyksiinsä saamiensa vastausten avulla erottaa, onko vastaaja ihminen vai tietokone.) Silti tietoisia robotteja voitaisiin Haikosen mukaan rakentaa, jos niiden toimintaperiaatteet sisältäisivät alisymbolista tiedonkäsittelyä.

Haikonen on rakentanut oman lähestymistapansa pohjalta robotin, joka kirjassa mainitaan kuin ohimennen. XCR-1-robotti perustuu puhtaasti neuroverkkoihin ja alisymboliseen tiedonkäsittelyyn. Jos sitä paukauttaa, se raportoii kivusta. Onko se tietoinen? Haikonen ei aivan suoraan kiellä tätä, mutta sellainen käsitys kirjasta syntyy: ”XCR-1-robotin jonkinasteisten todellisten tuntemusten olemassaolo vaatii vielä jatkotutkimuksia. Joka tapauksessa XCR-1 lienee tietävästi ensimmäinen robotti, jossa kipu esiintyy alisymbolisena systeemitilana.” (236)

Ei ole aivan selvää, ajatteleeko Haikonen jatkotutkimuksen ratkaisevan, tunteeko XCR-1 kipua, vai selvittääkö tutkimus, kuinka voidaan rakentaa kehittyneempi tunteva ro-

botti. Arvelisin jälkimmäistä. Haikonen näet surutta aiheuttaa koneelleen näitä ”kiputiloja”, mikä kielinee siitä, että hän ei ajattele todellista koettua kiputuntemusta esiintyvän. Hän kysyykin:

”[E]ntäpä, jos jonain päivänä keskuudessamme todellakin olisi tietoisia ja aidosti kipua tuntevia palvelurobotteja? Miten kohtelisimme niitä? Sanoisimmeko niille kiitos, kun ne tekisivät meille palveluksen? Entä olisiko tällaisen robotin lyöminen eettisesti oikein silloin kun se tekisi jotain väärin? Kyllähän se siitä kuitenkin jotain oppisi” (248).

Sitaatista käy ilmi teoksen kepeyteen tähtäävä tyyli, joka joskus naurattaa enemmän ja joskus vähemmän riippuen siitä, mihin pilkka osuu. (Visio älykkäästä kapakasta, joka twiittaa ja houkuttelee asiakkaan kavereita paikalle – ja tarvittaessa kutsuu valmiiksi poliisin – saakoon hieman Douglas Adams -pisteitä.)

### Miksi koemme elämyksiä emmekä hermosignaaleja?

Haikonen käy myös lennokkaasti joskin pikaisesti läpi mielen keskeisiä piirteitä: kognitio ja havaitseminen; näkeminen ja katseleminen; kuuleminen ja kuunteleminen; muisti ja muistaminen; mielikuvitus ja oppiminen; motoriikka; ajattelu ja assosiaatio; sisäinen puhe; äly; logiikka ja päättely; tunteet ja mielialat; kipu ja mielihyvä; hyvät ja pahat; kauniit ja miellyttävät asiat; sekä tunteiden kirjo saavat kaikki omat alalukunsa. Esimerkiksi sisäisen puheen luonne virtuaalisena havaintona on kiinnostavaa nostaa esiin. Kaikkiaan varsin onnistunutta tiivistä katsausta rydytetään osuvilla esimerkeillä poikkeustapauksista.

Luvussa 6 käsitellään tietoisuuden mysteeria. Haikonen vastustaa Daniel C. Dennettin kaltaisia yrityksiä selittää tietoisuus tyystin pois. Toisaalta hän syyttää muita tietoisuuden selityksen tarjoajia substanssidualismista, aineettoman mielen olettamisesta aineellisen ruumiin ja fyysikaalisen todellisuuden

rinnalle. Kanta on sympaattinen, vaikka pyörää ei olisikaan tarvinnut keksiä uudelleen: dualismin ja eliminativismin väliin sijoitettuja näkemyksiä on muotoillut ja kannattanut varsin moni asiaa pohtinut. Vetoaminen prosesseihin tai tapahtumiin on sinänsä kiinnostavaa. Aiheesta voi suositella lisälukemiseksi vaikkapa Helen Stewardin kirjaa *Ontology of Mind. Events, Processes, and States* (1997).

Haikoselle tietoisuus itse on mitä on riippumatta siitä, millaisilla mittareilla sitä voidaan havaita. Valonmittaaminen hiukkasina tuottaa tulokseksi hiukkasia ja aaltoliikkeenä aaltoliikettä – valo itse on eri asia kuin mittatulokset. Elämyksiä ei voi mitata, ne vain koetaan.

”Kaikki tietoiset havainnot ovat siis todellisia tai [kuten sisäisessä puheessa] virtuaalisia aistihavaintoja, jotka muodostavat ainakin lyhytaikaisen muistikuvan ja jotka voidaan raportoida. Tietoisia havaintoja vastaavia neurosignaalikuvioita voidaan havaita aivoista omilta aistialueiltaan. Selitettäväksi jäävätkin nyt vain tietoisuuden peruskysymykset: miten nämä todellisia tai virtuaalisia aistihavaintoja välittävät neurosignaalikuviot voivat ihmiselle itselleen näyttytyä laadullisina, subjektiivisina elämyksinä? Miksi havaintoihin liittyvät signaalit näyttytyvät sisäisesti väreinä, toiset ääninä, makuina, kosketuksina ja niin edelleen? Miksi emme havaitse näitä neurosignaaleja sellaisenaan kuin ne ovat, ja miksi näitä subjektiivisia elämyksiä ei voida instrumenteilla havaita?” (199–200).

Vaikka ei ole instrumenttia, joka tuottaisi mittaustulokseksi elämyksiä, tästä ei voi päätellä, ettei olisi mitattavaa ilmiötä, elämystä. Vain kokevat aivot itse voivat niitä havaita.

### Minä ja minä, selitys ja selitys

Haikonen vetoaa usein ”selittämiseen” vastustaessaan kilpailevia näkemyksiä. On epätoivottavaa, että jokin kanta ei selitä jotain toista asiaa tai että jokin kanta olettaa se-

littämättömän lähtökohdan. Tällä osittain retorisella tempulla leimataan tietyt näkemykset epätieteellisiksi. Haikosen oma ”selitys” tietoisuudelle on oikeastaan sen kuvaus: tietoisuus on laadullista (kvalioita sisältävää) havaitsemista.

Haikonen itse kysyy, eikö tarvita selitystä

”ihmeellisestä prosessista, jolla aivot muuttavat neurosignaalit kokemuksiksi havaitusta ulkomaailmasta väreineen, äänineen ja makuineen? Selitystä ei tarvita, koska tällaista ihmeellistä prosessia ei aivoissa ole, eikä tarvita. Neurosignaalikuviot näyttytyvät suoraan kokemuksena, koska muutakaan vaihtoehtoa ei ole. Signaalit eivät näyttäydy signaaleina, koska aivoissa ei ole niiden havaitsemiseen sopivia aistielimiä. Tämän takia emme voi punnita neuroneita tai neurosignaaleja itsenään, vaan tarkastelemme ulkomaailmaa neurosignaalien välityksellä.” (202–203)

Haikonen vertaa tätä radion kantoaaltoon, jota ei kuulla, mutta jonka välityksellä kuullaan musiikkia. Kiistaa tarkastelun ja ulkomaailman suhteesta akselilla realismi–konstruktivismi voisi tältä pohjalta punnita tarkemminkin, samoin sitä, millainen ”selitys” Haikosen selitys oikeastaan on.

Toisinaan tekstin yksityiskohtien seuraaminen vaatii pysähtymistä. Virke ”myöskään homunculusta tai muutakaan havaitsijaa ei tarvita, sillä systeemi itse on havaitsija” (205) yrittäneen sanoa, että ”havaitsija tarvitaan, mutta homunculusta tai muuta *systeemin ulkopuolista* havaitsijaa ei tarvita, sillä systeemi itse on havaitsija”. Jos ja kun näin on, systeemin rooliin havaitsijana liittyvät ongelmat vaivaavat myös Haikosen ehdotusta. Jos havaitsemista tapahtuu, oletettavasti joku tai jokin toimii havaitsijana ja jotain havaitaan.

Itsetietoisuuden tai minuuden tapauksessa havaitsija (subjektiminä) ja havaittava (objekti-minä) ovat yksi ja sama minä. ”Minuuteen liittyy havaintokokemus siitä, että olemme itsemme sisällä”, Haikonen tiivistää (205). Hän ei py-

## ”Digitaalisista koneista ei ole tietoisiksi, mutta ihmiskyborgeista olisi.”

sähdy pohtimaan ”itsensä sisällä olemista”. Se saattaa olla mielekäs ilmaus, jos sisällä oleva ”minä1” on eri asia kuin ”minä2”, jonka sisällä se on. Galen Strawson on huomauttanut, että aivan kuten ”linna” voi viitata pelkkään rakennukseen tai rakennukseen pihamaineen ja muureineen vallihautojen sisäpuolella, myös ”minuus” voi olla samalla tavalla kaksimerkityksinen.

Filosofisesti kiehtova osio on visio tietoisuuden siirtämisestä koneeseen. Digitaalisista koneista ei ole tietoisiksi, mutta ihmiskyborgeista olisi: jos ihmisaivoihin kytkettäisiin elektroniset vara-aivot (aistimiseen), joita henkilö alkaisi käyttää omiensa rinnalla, ja jos henkilö pikkuhiljaa siirtyisi toimimaan vara-aivojen varassa ja varsinaiset aivot saisivat kuolla pois, henkilö voisi jatkaa tietoista elämäänsä. Haikonen vertaa

tätä ihmiseen, jolla on aluksi yksi työhuone, sitten kaksi huonetta rinnan ja lopulta vain toinen huone: samaan tapaan tietoisuus voisi kenties muuttaa aivoista toiseen.

### Kyttäys, rahastus ja pisteytys

Kirjan puolitoista viimeistä lukua käsittelevät hyvin eri rekisterissä tekoälyn ja erityisesti tulevaisuuden mahdollisten tietoisten robottien herättämiä kysymyksiä työllisyydestä (SARS- tai ebolaepidemian puhjetessa robotit olisivat kipeästi tarpeen), seksistä (jos seksirobotit kykenevät saamaan orgasmeja, riittävätkö niille ihmiset, vai tarvitsevatko ne itse seksilelujä ja tietoisten robottien vastuuta (rangaistusten mielekkyys edellyttää kykyä tuntea mielipahaa).

Viimeinen luku on otsikoitu apokalypsiksi. Se tuo esiin, millaista

kyttäystä, valvontaa ja rahastusta (digitaalinen, tietoisuuteen kykenevätön) tekoäly ja jokaisesta saatavilla oleva runsas data tuo tullessaan. Sairasta, kauhistuttavaa ja hallitsematonta, maalailee Haikonen.

Lopun kauhukuvat on vedetty aika leveällä pensselillä. Joitain aitoja filosofisia kysymyksiä kirjassa ehkä hieman oioetaan, eivätkä vastakkaiset kannat aina saa ansaitsemaansa huomiota. Keskustelu vapaasta tahdosta myös hiukan ontuu. Ilmeisesti kirjoittaja on olettanut, että tahto ei voi olla vapaa, mikäli mikään vaikuttaa siihen lainkaan. Kirjan omaperäisin ja paras anti onkin konetietoisuuden ja samalla ihmistietoisuuden luonteen kuvauksessa. Kaikkiaan Haikosen teos on oikein suositeltava näkemyksellinen kokonaiskatsaus tietoisuuteen, tekoälyyn ja robotteihin.