

PENTTI O A HAIKONEN

Tekoälyn olemassaolo ja tietoisuus

Tekoälytutkimuksessa 'elämän' käsite tulee esiin tekoelämän (*Artificial Life* tai lyhyesti *A-Life*)¹ ja konetietoisuuden (*machine consciousness*) yhteydessä. *A-Life* -tutkimuksella tarkoitetaan itseään ylläpitävien, muuntuvien ja monistavien olioiden matemaattisten perusteiden tutkimusta tietokonesimulaatioin.

Käsitteenä 'elämä' on laaja ja joustava. Sen eri merkitykset antavat lähtö- ja vertailukohtat myös tekoelämän ja konetietoisuuden pohdinnalle. Biologiassa elämän perusteena on elävä solu, joka ylläpitää itseään aineenvaihdunnan avulla. Arkikielessä elämällä tarkoitetaan usein myös elämän kulkua, kaikkia tapahtumia, joita ihminen olemassaolonsa aikana kokee. Puhutaan myös henkisestä sisäisestä elämästä, joka liittyy tietoisuuteen ja sen sisältöön.

A-Lifen piirissä tekoelämän tutkimus on tavallisesti abstraktia, eikä se yleensä pyri mallittamaan todellisia biologisia soluja. Tämä saattaa olla tutkimuksen vahvuus, koska tulosten uskotaan olevan yleistettävissä monenlaisiin eläviin järjestelmiin. Toisaalta tämä on heikkous, koska tulokset eivät erityisesti liity ainoaan tunnettuun elävään järjestelmään, biologiseen soluun. Keinotekoisia elämää tutkitaan myös biologiassa, jolloin tavoitteena on mallittaa ja simuloida todellisia soluja ja viime kädessä luoda tällaisia keinotekoisesti orgaanisen kemian keinoin. Nykyinen *A-Life* toimii puhtaasti tietokonesimulaatioin ja yhdistelee eri tutkimusalueita kuten soluautomaatit, keinotekoiset neuroverkot, itseorganisointumisen, kompleksisuus ja emergenssi, evoluutio ja geneettiset algoritmit. Huomattakoon, että soluautomaatti on matemaattinen käsite, jolla ei ole mitään tekemistä biologisten solujen kanssa.

Tekoälytutkimuksen perimmäinen haaste on tietoisin koneen rakentaminen. Siinäkin eteen tulevat käsitykset elämästä.

Tietoisuus

Tietoinen systeemi operoi merkityksillä ja ymmärtää, mitä aistien tuottamat signaalit tarkoittavat. Sen sijaan tietokoneen toiminta perustuu symbolien käsittelyyn annettujen sääntöjen perusteella, eikä tietokone tiedä, mihin ulkoisiin merkityksiin nämä symbolit mahdollisesti viittaavat. Jotta systeemi voisi ylipäättään ymmärtää merkityksiä, sillä pitäisi olla fyysikaalisen maailman kytkentöjä, joihin merkitykset voisivat perustua.

Elävän järjestelmän peruskysymys fyysikaaliseen maailmaan on sen oma biologinen olemus ja kyky ylläpitää

itseään. Tähän liittyy systeemin tasapainotilan ylläpito vaihtuvissa olosuhteissa. Tämä taas edellyttää monien järjestelmien ja prosessien yhteistoimintaa ja säätävyyttä toistensa suhteen. Insinööritermein puhuttaisiin prosessien mittauksesta ja takaisinkytketyistä säätösystemeistä. Kehittyneemmissä biologisissa olioissa keskushermosto seuraa systeemin tilaa ja toimintoja jatkuvasti lukuisten anturien, reseptorien, välittämän tiedon perusteella. Myös prosesseissa vapautuvat kemikaalit toimivat indikaattoreina systeemin tilasta ja voivat täten myös osaltaan säätää prosessien toimintaa. Tämä mahdollistaa lopulta biologisen olion, kuten ihmisen, tietoisuuden omasta ruumiistaan ja sen tilasta. Osa systeemin ylläpitoon liittyvästä tiedosta on jatkuvasti tietoisuuden käytettävissä, osa vain silloin, kun jokin ei ole kohdallaan (jano, nälkä, kipu...). Kaiken tämän tiedon ajatellaan olevan edellytyksenä tietoisuuden kehittymiselle.

Tällä perusteella eräät tutkijat, kuten Ziemke², ovat esittäneet, että vain elävä biologinen olio voi olla aidosti tietoinen. Useammat tutkijat (allekirjoittanut mukaan lukien) eivät mene näin pitkälle, sillä tämä oletamus näyttäisi edellyttävän jotain biologisesti elävän olion erikoisominaisuutta (*élan vital*), jota muilla systeemeillä ei voisi olla. Insinöörikokemuksen perusteella oletetaan, että myös keinotekoinen systeemi voi luoda itselleen merkitysjärjestelmän, jos se pystyy havainnoimaan itseään ja ympäristöönsä sopivilla sensoreilla. Edelleen oletetaan, että keinotekoinen säätösystemi on omalta osaltaan riittävä itsetietoisuuden perustaksi, jos se tehdään itseään monitoroivaksi. Tietoista konetta ei nykytutkimuksen mukaan ollakaan tekemässä keinotekoisista elävistä soluista vaan perinteisellä mikropiireihin perustuvalla elektroniikalla.

Sisäinen elämä

Konetietoisuuden tutkimuksen kohteeksi on tullut myös 'sisäisen elämän' käsite (*inner life* tai *mental life*). Esimerkiksi Chalmers tarkoittaa sillä aivojen tietojenkäsittelyn elämyksellistä osaa, jonka yksilö itse subjektiivisesti kokee tietoisuutensa sisältönä³.

Mutta mitä olisi koneen sisäinen elämä? Nykyteknikka on tuottanut monimutkaisia laitteita, joiden toimintaa on tullut tavaksi kutsua älykkääksi, yleensä ilman



sen kummempaa perustetta. On myös kyetty valmistamaan hyvinkin ihmisenkaltaisia robotteja, joita tietokoneohjelmiin ja mikroprosessoreihin perustuva tekoäly ohjaa. Tällainen tekoäly on vain kokoelma suunnittelijansa luomia sääntöjä: se ei ole robotin omaa älyä tai tietoisuutta. Tietokoneohjelman suorittuminen mikroprosessorissa, niin kuin se nykyisin tapahtuu, ei ole sellaista 'sisäistä henkistä elämää', jota voisi kohtuudella verrata ihmisen ajatusvirtaan; robotilla voi olla valot päällä, mutta taatusti ketään ei ole kotona. Teollisuusrobottien yhteydessä tämä ei yleensä ole ongelma, mutta tilanne on toinen viihde-, pelastus- ja hoivarobottien kanssa. Seuraisrobotit, joilla ei ole mitään sisäistä elämää, ovat tylsiä. Pelastus- ja hoivarobotit voivat olla hyödyttömiä tai suorastaan vaarallisia, jos ne eivät kykene ymmärtämään tilanteita, suunnittelemaan toimintaansa sekä ennakkoimaan toimintansa seuraamuksia. Tällaisten robottien pitäisi tuntea empatiaa ja ymmärtää käyttäjänsä tarpeita eri tilanteissa.

Tekoälytutkimuksen haasteena onkin sellaisten systeemien kehittäminen, jotka voisivat olla tietoisia omasta olemassaolostaan ja pohtia sitä. Konetietoisuuden eräänä edellytyksenä nähdään systeemin kyky mielikuvitukseen ja oman historian sekä mahdollisen tulevaisuuden tarkasteluun⁴. Kun mukaan otetaan emotionaalinen arviointi ja tavoitteellisuus, voidaan ehkä päästä samantapaisiin ajatusprosesseihin kuin ihmiselläkin on. Koneella olisi tällöin mielikuvitus ja omaehtoinen sisäinen elämä. Tällaisella koneella olisi myös kipuun ja mielihyvään verrattavia tiloja, joiden vaikutukset olisivat samantapaisia

kuin ihmiselläkin. Kone saattaisi käyttäytyä ikään kuin se kokisi todellista kipua ja myös ilmoittaa tämän sanallisesti. Ulkopuolisen tarkkailijan olisi kuitenkin vaikea tietää, josko raportoitu kipu olisi todellista, samanlaista kuin me tunnemme. Näin päädytään kysymyksiin, jotka toistaiseksi ovat vain filosofisia: Miten tällaista konetta tulisi kohdella? Olisiko eettisesti oikein tuottaa kipua ja mielipahaa tietoisille roboteille?

Viitteet

- 1 Ks. Levy 1993.
- 2 Ziemke 2007.
- 3 Chalmers 1996 .
- 4 Esim. Aleksander 2005; Haikonen 2003 & 2007.

Kirjallisuus

- Aleksander, I., *The World in My Mind, My Mind in the World. Key Mechanisms of Consciousness in People, Animals and Machines*. Imprint Academic, Exeter 2005.
- Chalmers, D. J., *The Conscious Mind. In Search of a Fundamental Theory*. Oxford University Press, Oxford 1996.
- Haikonen, Pentti O A, *The Cognitive Approach to Conscious Machines*. Imprint Academic, Exeter 2003.
- Haikonen, Pentti O A, *Robot Brains. Circuits and Systems for Conscious Machines*. John Wiley & Son, New York 2007.
- Levy, Steven, *Artificial Life. A Report from the Frontier Where Computers Meet Biology*. Vintage Books, New York 1993.
- Ziemke, Tom, What's Life Got to Do with It? Teoksessa *Artificial Consciousness*. Toim. A. Chella & R. Manzotti R. Imprint Academic, Exeter 2007.