

Tarvitseeko tietokoneäly moraalifilosofiaa?

Mahdollisuus aikaansaada ihmisaivojakin tehokkaampi päätöksen tekokyky on kiinnostanut ainakin kirjailijoita jo sadan vuoden ajan. Varhaiset tekoälykkäät laitteet oli yleensä rakennettu ihmisen hahmoon eli roboteiksi — tuo nimikin perustuu Carrel Čapekin näytelmään R. U. R. vuodelta 1920.

Tieteiskirjallisuuden ja -elokuvien ystäville tuttuja robotikklassikkoja ovat ainakin Maria, koneihminen Fritz Langin Metropolis-elokuvassa vuodelta 1927, Isaac Asimovin Robbie the Robot (kirja vuonna 1940 ja elokuva 1956) sekä suomalaisen oma Atorox, kuuden itseenäisen kirjan sarja vuosilta 1947 ja 1948.

Asimov loi genren peruskliiseksi jo muodostuneet ”robotiikan kolme perussääntöä”, joiden perustalle rakennettu tekoäly pysyttäisi robotit ikuisesti ihmisten ongelmattomina orjina. Asimovin tarinoissa robotien moraalilait olivat kuitenkin kierrettävissä, ja paradoksit johtivat aina niiden ja ihmisten välisiin vaikeuksiin.

Vain sadassa vuodessa tekninen kehitys on jo miltei ylittänyt viime vuosituhannen par-

haidenkin kirjailijoiden mielikuvituksen. Alexa ja Google Chat pystyvät haastattelemaan toisiaan, vaikka joskus melko hilpein tuloksin. Vuonna 2017 Saudi-Arabia ehti ensimmäisenä jopa myöntämään kansalaisuuden Sophia-robotille, viehättävillä ihmiskasvoilla varustetulle chattibotille.

”Ajattelevia” koneita ei aina rakennettu itsestään liikkuvaan kykenevien laitteiden osiksi. Tekoäly saattoi toimia vaikkapa vain halpahintaisen kioskirakennuksen väsymättömänä suoltajana, kuten Ronald Dahlin mainiossa novellissa Suuri automaattinen kielikone (1954) ja Stanislaw Lemmin moitteetonta runoa työstävä Trurlin Elektrubaduuri (kirjassa Kyberias, 1965).

Ensimmäinen ylivertaisen koneälyn mahdollisuuksia pohdiskeleva teos lienee Kurt Vonnegutin debyyttiromaani Sähköpiano vuonna 1952: *”EPICAC XIV hallitsee samanaikaisesti satoja ellei tuhansia puolia samasta asiasta täysin tasapuolisesti”, koska sillä ”ei ollut lainkaan järkeä sumentavia tunteita eikä se koskaan unoh- tanut mitään”.*

”EPICAC pystyi arvioimaan herkästi räjähtävien pommi- edut verrattuina atomi- aseisiin taktisena tulitukena ja samalla kertaa pitämään mielessään räjähdysaineiden saatavuuden verrattuna fissio- räjähdysaineisiin, työllisyyden kummankin räjähdysaineteollisuuden piirissä, lentokoneiden tuhoutumisnopeuden vihollisen ilmatorjuntatekniikan kehittyessä, ja niin edelleen.”

Vaikka Sähköpianossa päätäjät jo tukeutuivat vapaaehtoisesti ja sokeasti koneen laskelemaan ja päätelmiin kaikissa kansantalouden ja sotatarviketeollisuuden kysymyksissä, viimeistä napin painallusta hyökkäys- tai puolustussodan aloittamiseksi ei kuitenkaan ollut annettu koneen toimivaltaan — vielä.

Arthur C. Clarken 2001 avaruusseikkailussa (kirja sekä elokuva 1968) HAL-päätietokoneelle oli annettu kaikki tiedot sekä täydellinen päätäntävalta kaikkien tutkimusmatkan tavoitteen toteuttamisen kannalta tarpeellisten toimenpiteiden käynnistämiseksi, mikä osoitautui lopulta kohtalokkaaksi ihmismatkustajien kannalta.

Tekoälyn historia liittyy paitsi tietojenkäsittelytieteeseen, läheisesti myös matematiikkaan ja filosofiaan. Modernin filosofian perustajana pidetty René Descartes jo päätteli, että sielun eli älyn valtakunnan olemassa olo ei välttämättä liittyisi mitenkään ihmiskehoon, ja G. W. Leibniz argumentoi järjen ja tahdon ylivertaisuutta uskonnolliseen fatalismiin nähden.

Vuonna 2020 ilmestynyt ja viime vuonna suomennettu Tekoälyn etiikka ei kerro juuri mitään edellä mainitsemastani historiallisesta kirjallisuudesta, vaan keskittyy otsikon aiheeseen uusimpien tutkimusten pohjalta. 228-sivuinen tiivis kirjanen kertoo tuoreimpia esimerkkejä tekoälyn käytöstä — ja väärinkäytöstä.

Kaupallistettu ja tietokoneistettu tilastotiede on nykyisin jo suurta liiketoimintaa. Tekoälyllä voidaan mm. tuottaa nettiin vihapuhetta ja väärää informaatiota sekä luoda ihmiseltä vaikuttavia botteja, jotka oikeasti ovatkin tekoälyohjelmistoja.

2018 Uberin itseohjautuva auto aiheutti jalankulkijan kuoleman. Onko vastuussa auton ohjelmoinnin kehittänyt autonvalmistaja, Uber-yhtiö, auton käyttäjä, jalankulkija vai kenties lainsäätäjä (osavaltio), joka on sallinut tuollaisen toiminnan?

Floridassa käytetään jo COM-PASS-algoritmia päätettäessä ehdonalaiseen vapauteen päästämiseksi, ja on todettu väärin positiivisten ratkaisujen (jotka estävät vapauttamisen) kohdentuvan useammin mustaan kuin valkoiseen väestöön. Siksi on helppoa väittää, että al-

goritmissa on vinouma väärävärysten epäeduksi.

Entä nyt, kun yhä enemmän päätöksiä ja toimia siirretään tekoälylle? ”Koska kone ei ole tietoinen tekemisistään, se ei pysty moraaliseen päättelyyn eikä siten voi olla vastuussa tekemisistään,” kirjoittaja toteaa. EU:n vuonna 2018 voimaan tullut yleinen tietosuojasetus (GDPR) antaa ainakin periaatteessa rekisteröidylle kansalaiselle oikeuden ”saada käyttöönsä automaattisessa

päätöksenteossa merkitsevää informaatiota käytettyä logiikasta”, mutta sekään ei ilmeisesti oikeuta yksittäisen päätöksen perustelujen selvittämistä.

TEKOÄLYN ETIIKKA



MARK COECKELBERGH

Suomentanut
KIMMO PIETILÄINEN

Terra Cognita