

# Kompleksisuus tekoälyn perimmäisenä ongelmana

Tuomas Toivonen

19. joulukuuta 2003

## 1 Johdanto

Käyn lyhyesti läpi kompleksisuuden, kombinatorisen räjähdysen ja niiden aiheuttamat ongelmat tekoälylle. Käsittelen ongelmaa sekä symbolisen että adaptiivisen tekoälyn näkökulmasta. Lopuksi kommentoin Pauli Salon hypoteesia ihmismielen aidosta kompleksisuudesta.

## 2 Kombinatorinen räjähdys

Kombinatorisella räjähdyksellä viitataan ilmiöön, jossa syötteiden määrän lineaarinen kasvu johtaa tietyn algoritmin aika- tai tilavaatimusten eksponentiaaliiseen kasvuun. Kyseessä on siis klassinen “moniuloitteisuuden kirous” [Perlovsky 1998].

Seuraavassa kombinatorista räjähdystä tarkastellaan eri tyypisissä tekoälyjärjestelmissä jakamalla ne kahteen luokkaan eli symbolisiin ja adaptiivisiin järjestelmiin. Jako perustuu pikemminkin järjestelmien toteutustapoihin kuin toiminnallisuuteen ja toteutuksesta johtuviin erityispiirteisiin.

### 2.1 Symboliset tekoälyjärjestelmät

Symbolisilla tekoälyjärjestelmillä tarkoitetaan formaalille logiikalle — aksiomiin ja niihin liittyviin sääntöihin — perustuvia järjestelmiä. Puhdasoppisten symbolijärjestelmien pohjana on niihin syötetty apriori tieto eli säännöt. Säännöt pyrkivät kuvaamaan järjestelmän ulkoisen ympäristön ja mallit rationaaliseen toimintaan tässä ympäristössä. Käytännön toteutuksissa sääntöjen määrä kasvaa räjähdysmäisesti [Perlovsky 1998, p. 667]: joko (a) rationaalista toimintaa ei voida yksinkertaistaa perussääntöihin, joista muut säännöt olisivat johdettavissa tai (b) tapauskohtaisesti rationaaliseen toimintaan riittävien sääntöjen johtaminen perussäännöistä on liian hidasta.

### 2.2 Adaptiiviset tekoälyjärjestelmät

Adaptiivisilla järjestelmillä viitataan oppiviin järjestelmiin eli käytännössä erilaisiin neuroverkkototeutuksiin [Perlovsky 1998, p. 667]. Puhdasoppisten adap-

tiivisten järjestelmien tieto on aposteriori eli oppimisprosessin tuloksena peräisin järjestelmän ympäristöstä. Järjestelmien oletetaan kykenevän rationaaliseen toimintaan maailmassa, josta niiden tieto on peräisin. Käytännössä järjestelmien oppimisprosessi toteutuu joko (a) liian hitaana tai (b) rationaalille toiminnalle välttämätöntä tietoa ei kyetä eristämään järjestelmän ympäristöstä.

## 2.3 Yhdistelmäjärjestelmät

Hyödyntämällä sekä symbolisten että adaptiivisten tekoälyjärjestelmien ominaisuuksia voidaan toteuttaa ns. yhdistelmäjärjestelmiä. Näissä järjestelmissä pyritään rajaamaan apriori -sääntöjen kombinatorinen räjähdys oppimisella ja aposteriori -oppimisen räjähdys oppimista edesauttavilla valmiilla säännöillä. Käytännössä yhdistelmäjärjestelmien toteutuksessa ongelmaksi on muodostunut laskennallinen kompleksisuus [Perlovsky 1998, pp. 667-668].

## 2.4 Kombinatorisen räjähdysten hallinta

Kombinatorista räjähdystä on pyritty rajaamaan heuristiikoilla tai tiettyyn ongelmakenttään rajatuilla asiantuntijajärjestelmillä [Salo 2002]. Heuristisessa ratkaisussa säilytetään järjestelmän yleiskäyttöisyys karsimalla rajaamalla kombinatorista räjähdystä rationaalisuuden kustannuksella. Sen sijaan, että käytäisiin täydellisen rationaalisuuden saavuttamiseksi läpi kaikki mahdolliset vaihtoehdot, yritetään saada aikaan yleensä oikea ratkaisu vain osalla vaihtoehdoista. Asiantuntijajärjestelmissä rajoitetaan toiminta-aluetta. Järjestelmä kykenee toimimaan rationaalisesti tietyssä ongelmakentässä, mutta ei sen ulkopuolella laisinkaan.

Heuristisella ratkaisulla on kyetty rakentamaan käytettävän tehokkaita järjestelmiä, mutta käytännössä niiden rationaalisuus on ollut hyvin rajallista [Salo 2002]. Tulokset asiantuntijajärjestelmien osalta ovat olleet parempia, mutta myös niissä mallien käyttö laajempien ongelmien ratkaisuun on törmännyt kombinatoriseen räjähdykseen. Voidaankin sanoa, että kombinatorinen räjähdys voidaan välttää vain ns. "leluongelmissa" [Perlovsky 1998, Salo 2002].

## 3 Kolmogorov-kompleksisuus

Kolmogorov-kompleksisuus eli algoritmaalinen informaatioteoria pyrkii selvittämään informaation todellisen kompleksisuuden. *Dictionary of Algorithms and Data Structures* määrittelee Kolmogorov-kompleksisuuden seuraavasti:

Pienin mahdollinen määrä bittejä, johon merkkijono voidaan tiivistää menettämättä informaatiota. Määritetään suhteessa kiinnitettyyn, yleiskäyttöiseen purkumenetelmään kuten universaaliin Turingin koneeseen [NIST 2000].<sup>1</sup>

---

<sup>1</sup>"The minimum number of bits into which a string can be compressed without losing information. This is defined with respect to a fixed, but universal decompression scheme, given by a universal Turing machine." Käännös kirjoittajan.

Keskeistä määritelmässä on informaation tiivistäminen ja yleisen purkumenetelmän käyttö. Esimerkiksi toistoa sisältävä merkkijono on vain näennäisesti kompleksi sillä se voidaan tuottaa yksinkertaisella, toiston suorittavalla ohjelmalla. Samoin esimerkiksi alkulukujen sarja ei ole kompleksi, koska se voidaan tuottaa lyhyellä ohjelmalla. Merkkijono on todellisesti kompleksi vain, jos sen tiivistetyn esityksen ja purkumenetelmän toteuttavan ohjelman yhteiskoko ei ole merkittävästi alkuperäistä merkkijonoa pienempi.

Yleisen purkumenetelmän käyttö tekee merkkijonojen kompleksisuudesta vertailukelpoista. Esimerkiksi kahden merkkijonon kompleksisuutta voidaan vertailla laskemalla yhteen molempien merkkijonojen tiivistetyn esityksen ja alkuperäisen merkkijonon tuottavan ohjelman pituus. Merkkijono, jonka tiivistetyn esityksen ja ohjelman pituuksien summa on suurempi, on kompleksisempi. Oletuksena on, että molemmat ohjelmat ovat ajettavissa samalla arkkitehtuurilla [Wikipedia 2003, Salo 2002]. Ohjelmat on siis vertailua varten kiinnitetty yleiskäyttöiseen purkumenetelmään.

Matemaattisesti kiinnittämätön ja kiinnitetty kompleksisuus voidaan määritellä seuraavasti [Salon ja Lappi 2003, pp. 10-11]:

$$\begin{aligned} \text{Kiinnittämätön: } IC(x) &= \min\{|p| : f(p) = x\} \\ \text{Kiinnitetty: } IC_{UTM}(x) &= \min\{|ip| : fi(p) = x\} \end{aligned}$$

Määritelmässä  $IC$  on merkkijonon  $x$  informaation sisältö (information content) eli kompleksisuuden aste,  $p$  merkkijonon tuottava ohjelma ja  $|p|$  ajettavan ohjelman pituus. Kiinnittämättömässä määritelmässä merkkijonon kompleksisuutta osoittaa lyhimmän mahdollisen purkumenetelmän pituus siten, että pidempi purkumenetelmä osoittaa kompleksisempia merkkijonoja. Kiinnitettyssä määritelmässä  $UTM$  on universaali Turingin kone ja  $i$  on purkumenetelmien toteuttamiseen käytetyn määrittelykielen toteutus. Käytännön esimerkkinä purkumenetelmä voisi olla toteutettu tietyllä ohjelmointikielellä kuten C tai Java. Tällöin  $|i|$  olisi hieman yksinkertaista ohjelmointikielen universaalille Turingin koneelle suunnatun toteutuksen koko.

Yllä annettujen merkkijonoja käsittelevien esimerkkien lisäksi Kolmogorov-kompleksisuuden ideaa voidaan käyttää yleisesti järjestelmien kompleksisuuden arviointiin. Järjestelmä on monimutkainen vain, jos sen ylätasoa näennäistä kompleksisuutta ei voida selittää alempien tasojen yksinkertaisilla malleilla.

Esimerkiksi symbolisen tekoälyn tapauksessa näennäisen kompleksi järjestelmä redusoituu formaalin logiikan keinoin tapahtuvaan syntaktiseen manipulointiin. Kompleksisuutta symbolisiin järjestelmiin tuovat syötetyt säännöt eli apriori -tieto. Kokemus käytännön järjestelmistä on osoittanut sääntöjen syötön johtavan kombinatoriseen räjähdykseen, joka osoittaisi ympäristön olevan todellisesti ja järjestelmän vain näennäisesti kompleksi.

Tilanne on vastaava adaptiivisen tekoälyn malleissa. Järjestelmä redusoituu yksinkertaisiin periaatteisiin pohjautuvaan neuroverkkoon ja kompleksisuuden lähteenä on järjestelmän oppima aposteriori -tieto. Oppimisen kombinatorisen räjähdysten aiheuttaa informaatio järjestelmän ulkopuolelta.

## 4 Ihmisen aito kompleksisuus

### 4.1 Ihmismielen modulaarisuus

Ihmismielen toiminnallisuutta mallinnettaessa voidaan tehdä jako itsenäisiin, tietyn toiminnallisuuden toteuttaviin yksikköihin (moduulit) ja keskitettyihin prosesseihin [Salo ja Lappi 2003, p. 2]. Kognitiivisessa neurotieteessä yksiköt hahmottuvat anatomisina ja arkkitehturaalisesti erikoistuneina neurobiologian tasolla. Psykologia eristää mielestä erillisiä toiminnallisuuksia rakenteellisten yksikköjen sijaan. Keskeistä molemmissa näkemyksissä on yksikköjen riippumattomuus toisistaan. Esimerkiksi fysiologiset vauriot tiettyyn anatomisesti eristettyyn yksikköön näkyvät usein tietyn toiminnallisuuden vauriona.

Kognitiotieteen malleissa yksikköjako toteutetaan yhdistelemällä neurotieteellisiä ja psykologisia malleja, mutta yksikköjen raja- ja eritys on tarkempi. Mieli jaetaan toiminnallisiin yksikköihin ja näitä yhdistäviin keskusprosesseihin. Malli toimii hypoteesina, joka pyritään todentamaan toteutettuna tekoälyjärjestelmänä. Yksikköjaon toteutusta rajoittaa joukko parametreja, joista tärkeimpinä jokaisen yksikön on oltava (a) keskittynyt tiettyyn toiminnallisuuteen eli yleensä tietyn tyyppisen informaation käsittelyyn (domain specificity) ja (b) toiminnallisuuden toteutuksen on oltava riippumaton muista yksiköistä tai keskusprosesseista eli toimittava yksikön oman informaation varassa (information encapsulation) [Salo ja Lappi 2003, pp. 2-4].

Kognitiotieteen mallien yksittäiset yksiköt ovat ”tyhmiä”. Ne toteuttavat tietyn toiminnallisuuden automaattisesti ja deterministisesti: määritellystä syötteestä ennustettavissa oleva tulos. Erillisen yksikön algoritmaalinen mallinnus ja toteutus on mahdollista, koska keskitytään tiettyyn toiminnallisuuteen ja informaation määrä on rajoitettu. Kombinatorinen räjähdys vältetään samoin keinoin kuin asiantuntijajärjestelmiä toteutettaessa: järjestelmä tehokkaasti ja rationaalisesti aihepiirinsä sisällä.

Erillisiä yksikköjä yhdistämään oletetaan keskitettyjä prosesseja, joista muodostuu ihmismielen olemus — tietoisuus, tavoittelisuus ja rationaalisuus. Keskitettyjä prosesseja on yritetty toteuttaa mm. yllä esitetyillä symbolisen tai adaptiivisen tekoälyn malleilla. Käytännössä tuloksena on ollut kombinatorisen räjähdysten halvaannuttamia järjestelmiä, joista on saatu käyttökelpoisia vain rajaamalla yleiskäyttöistä. Keskitettyjen prosessien toteuttaminen rajattuna järjestelmänä muistuttaa kuitenkin parametreiltaan enemmän yksittäistä yksikköä ilman merkittävää laadullista tai määrällistä eroa. Ihmismielen keskitettyjen prosessien ideallit — yleistävyys, tehokkuus ja rationaalisuus — eivät toteudu.

### 4.2 Kompleksisuus ja kognitio

On osoittautunut mahdottomaksi ymmärtää, miten ihmismielen keskitetyt prosessit ovat sekä tehokkaita että rationaalisia [Salo ja Lappi 2003, pp. 8-9]. Salo ja Lappi esittävät, että syynä on ongelman luontainen kompleksisuus, ei väärä algoritmi (virheelliset mallit). Kuitenkin ihmismieli kykenee toimimaan luontaisessa ympäristössään tehokkaasti ja rationaalisesti. Salo ja Lappi selitystä

mukaillen ongelman ratkaisu voidaan esittää seuraavasti:

1. Ihmisen toimintaympäristö on luontaisesti kompleksi. Ympäristössä toimiminen on laskennallisesti kompleksia, joten rationaalista toimintaa ei voida toteuttaa tehokkaasti.
2. Toimintaympäristön kompleksisuus kierretään hahmottamalla ihmismieli tavallaan valtavana asiantuntijajärjestelmänä. Ihmisen toiminta on sopeutunut ympäristöön ja on siten näennäisen rationaalista. Ts. ihmisen toiminta on selvästi ympäristön asettamia parametreja (fysiikan lait jne.) rajallisempaa, mutta silti näiden parametrien sisällä rationaalista. Tekoälyjärjestelmien ns. kehysongelma ratkeaa, koska ihmismieli on "sisäisesti kehystetty". Toimintaympäristön laskennallinen kompleksisuus on hallittavissa, koska sitä ei ole tarpeen nähdä rajattomina vaihtoehtoina vaan ihmismielelle rajattuna ilmentymänä.<sup>2</sup>
3. Rajattu ilmentymä voidaan laskennallisuuden sijaan mallintaa yksinkertaisena taulukuselvänä, joka on tehokas. Oikein laadittu taulu tuottaa rationaalista toimintaa.
4. Ihmismielen taulu on muodoltaan todellisesti kompleksi. Täten sitä ei voi tiivistää eli keskitettyjen prosessien yksinkertaiset mallit eivät voi toisintaa ihmismielen toimintaa.

### 4.3 Ympäristö kompleksisuuden lähteenä

Kolmogorov-kompleksisuuden valossa vaikuttaisivat sekä symbolisen että adaptiivisen tekoälyn mallit pohjimmiltaan yksinkertaisilta ja mallien näennäinen kompleksisuus olisi ympäristöstä saatavan todellisen kompleksisuuden tulosta. Salo'n hypoteesi esittää, että ihminen on aidosti kompleksi organismi. Koska tekoälyjärjestelmien mallit ovat vain näennäisesti komplekseja, eivät ne voi saavuttaa ihmisen kognitiivisen toiminnan tasoa. Toisin sanoen, jos aidosti kompleksista järjestelmästä rakennetaan vain näennäisesti kompleksi malli, menetetään järjestelmän toiminnan kannalta välttämätöntä informaatiota.

Salo'n hypoteesi pyrkii kiinnittämään ihmisen kompleksisuuden lähteeksi ympäristön. Ihmisen geneeissä on rajallinen määrä informaatiota, joten geenit itsessään eivät voi toimia kompleksisuuden lähteenä [Salo 2002, p. 57]. Geenit kuitenkin tarvitsevat aktivoituakseen tietynlaisen, tarkkaan rajatun ympäristön. Ilman soveltuvaa ympäristöä geneeistä ei kehity ihmistä tai ihmisen henkistä kehitystä tapahdu. Ihmisen kompleksisuus syntyy siis geenien vuorovaikutuksesta ympäristön kanssa.

---

<sup>2</sup>Vrt. sosiologi Pierre Bourdieun konsepti ihmisen toimintaa määrittävästä habituksesta [Bourdieu 1977]. Habitus kattaa ihmisyksilön uskomukset ja tottumukset ohjaten arkitointia ja tehden siitä pitkälti tiedostamatonta. Ihminen on arkitoinnassaan habituksen ohjaama "unissakävelijä". Toiminnan tiedostaminen vaatii tietoista ponnistusta. Habitus kehystää toiminnan taulukuselväksi ja vain erityistilanteissa ihminen tietoisesti valitsee, ts. laskee parhaan vaihtoehdon.

Ihmisen kompleksisuuden tarkka muoto ei ole merkityksellistä. Ihmisyksilöiden välillä syvärakenteet ovat invariantteja, ts. yhteisiä ja jaettuja. Eri ihmisyksilöiden välillä kompleksisuus on siis samankaltaista johtuen yhteisestä ja jaetusta genotyypistä ja ympäristöstä. Kompleksisuuden muodolla ei ole merkitystä, koska ihmisen toiminnan ei täydy olla objektiivisesti rationaalista vaan riittää, että toiminta on subjektiivisesti rationaalista "ihmisyden" piirissä. Kompleksisuuden ihmisyksilössä A aiheuttama toiminta on rationaalista ihmisyksilölle B, koska toiminnan tulkinnan lähteenä on sama kompleksisuus.

## 5 Yhteenveto

Sekä symbolien käsittelylle että neuroverkkoihin perustuvat tekoölyjärjestelmät ovat vain näennäisesti komplekseja. Kummassakin tapauksessa aitoa kompleksisuutta ovat joko apriori –säännöt tai aposteriori –oppiminen ja itse järjestelmä on yksinkertainen ideoiden maailma tai *tabula rasa*, tyhjä taulu.

Salon hypoteesi esittää ihmisen aidosti kompleksina järjestelmänä, jonka kompleksisuuden lähteenä on ympäristö. Koska ihminen on aidosti kompleksi, eivät geenit voi toimia yksinkertaistuksena, jonka päälle ympäristön kompleksisuus ladottaisiin joko sääntöinä tai oppimisena.

Ihminen on syntyessään kompleksi ja selviää siksi vaihtelevassa arkiympäristössään. Tekoölyjärjestelmät ovat syntyessään yksinkertaisia ja eivät voi täten omaksua kompleksisuutta ympäristöstään saati toimia tässä ympäristössä rationaalisesti.

## Viitteet

- [Bourdieu 1977] Bourdieu, Pierre. *Outline of a Theory of Practice*. Trans. Richard Nice. Cambridge: Cambridge University Press, 1977.
- [NIST 2000] National Institute of Standards and Technology. Kolmogorov complexity. In *Dictionary of Algorithms and Data Structures*. <http://www.nist.gov/dads/>
- [Perlovsky 1998] Perlovsky, Leonid I. Conundrum of Combinatorial Complexity. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 666-669.
- [Salo 2002] Salo, Pauli. *Philosophy of Artificial Intelligence*. Lecture notes. Helsinki: Department of Computer Science, University of Helsinki, 2002.
- [Salo ja Lappi 2003] Salo, Pauli ja Otto Lappi. *Why there is yet no theory of non-modular cognition?* Helsinki: Department of Psychology, University of Helsinki, 2003.

[Wikipedia 2003] Wikipedia. Algorithmic information theory. In *Wikipedia*, <http://www.wikipedia.org/>.