1.4. Tekoäly, psykologia ja kielitiede

1960-luvulta alkaen on vallalla ollut *kognitiivinen* psykologia, joka tarkastelee tekoälyn tavoin aivoja informaation käsittelijänä. Monet tämän koulukunnan jäsenet lähtevätkin siitä, että kognitiivisten teorioiden pitäisi olla luvun 1.1.2 tapaan ”tietokoneohjelmien kaltaisia” eli mallintaa kognitiiviset prosessit hyvinmääriteltyinä informaationkäsittely- ja muuntotoimituksina. 1943 Kenneth Craik muotoili tämän kannan 3 askeleena:

1. Ulkoa tuleva *ärsyke (stimulus)* pitää muuntaa *sisäiseen esitysmuotoon (internal representation)* . . .
2. jota toimijan *kognitiiviset prosessit* käsittelevät tuottaen tuloksina uusia sisäisiä esitysmuotoja. . .
3. jotka sitten muunnetaan takaisin toiminnaksi.

Eli älykkäiden toimijoiden A_{ert} teoriana!

Noin vuosina 1920–60 varsinkin Yhdysvalloissa oli kuitenkin vallalla *behaviorismi*, joka oli syntynyt vastalauseeksi sellaisten subjektiivisten ”kyökkipsykologian” sanojen kuten ”uskomukset”, ”tavoitteet” ja ”järkeily” aiemmalle käytölle psykologisissa tutkimuksissa.

Behavioristien oma tutkimusohjelma oli taas kovasti luvun 1.3 loogisen positivismin oloinen: sallittiin vain tarkat objektiiviset mittaukset ja havainnot koe-eläinten reaktioista niille annettuihin ärsykkeisiin.

Craikin teoria taas mahdollisti näiden sanojen objektiivisen määrittelyn sopivien kognitiivisten tilojen ja prosessien niminä – palautti ne jälleen tieteellisiksi.

Vuonna 1957 johtava behavioristi *B.F. Skinner* julkaisi kirjan, jossa hän kuvasi behavioristisen käsityksen *kielen oppimisesta*. Nuori kielitieteilijä *Noam Chomsky* kirjoitti siitä kriittisen arvostelun [Cho97], jossa hän osoitti, ettei behaviorismi pysty kuvaamaan *kielellistä luovuutta*: sitä, että lapsi(kin) pystyy ymmärtämään ja tuottamaan sellaisia oikein muodostettuja ilmauksia, joita hän ei ole aikaisemmin tavannut. Behaviorismi oli tiensä päässä – se ei yleistynyt monimutkaiseen käyttäytymiseen.

Chomskyn oma *generatiivisten kielioppien* teoria taas selitti kielellisen luovuuden kognitiivisella mallilla, joka mukaan (ihmis)mielellä olisi erityinen *kielikyky* (erinomainen johdatus tähän teoriaan on [Pin94]):

1. Se erottaa puheen muista kuulohavainnoista *kielen sääntöjen* avulla.
   Sitten se "kääntää" sen samoilla säännöillä *mentaleesin* eli "mielen kielen" ilmaisuksi, joka säilyttää puheen merkityksen (*syvärakenteen* eli *deep structure*).
2. Muu ajattelu saa syötteenään tuon mentaleesi-ilmauksen.
   Ajattelun ei silti tarvitse olle *pelkästään* kielellistä.
3. Kielikyky muuntaa käyttäen samoja sääntöjä vastakkaiseen suuntaan ajattelun tuloksen jälleen luonnolliselle kielelle, ja panee äärielimet liikkeelle tuottamaan puhetta.
   Itse äänen tuotto vaikuttaa yllättävän hankalalta askeleelta:
   - Vauvaiän jälkeen (vain) ihmisen kurkunpää muuttuu paremmin puheen tuotantoon sopivaksi – johtaen "väärään kurkkuun"!
   - Lapsi oppii viittomakielen 24–36 kk:n "kieliräjähdystä" aiemmin.

Teoria olettaa, että aivoissa on (evoluution tuloksena [NKN01]) *universaalikielioppi*, joka sisältää kaikki ne kielioppisäännöt, joita missään luonnollisessa kielessä voi esiintyä.

Tämän vahvan oletuksen voi heikentää oletukseksi, että aivoissa on säännöt "vierekkäisten" symboleiden liittämiseksi toisiinsa ja liitosten purkuun [Car97, Dea98].

Lapsi oppii kielensä havaitsemalla mitkä näistä säännöistä ovat päällä/pois päältä ympäristössä. Empiristin "Tabula rasa" -kielenoppimisteorioita häiritsee *syötteen vähyys (dearth of input)*: 2-vuotias lapsi on kuullut niin vähän puhetta, ettei hän ole ehtinyt muodostaa alusta sääntöjä joita hän alkaa menestyksellä soveltamaan.

1.5. Tekoälyn historiaa

Tekoälyn synty (1943–56)

Ensimmäisen tekoälyartikkelin kirjoittivat neurofysiologit *W. McCulloch* ja *W. Pitts* vuonna 1943 [MP43] yhdistämällä 3 aiemmin erillistä tieteenhaaraa:

- aivojen *hermosolujen* (neuronien) toiminnasta saadun tiedon;
- formaalin logiikan; ja
- laskettavuuden teorian.

He ehdottivat myös, että hermosolujen *verkot* voisivat *oppia*. *Donald Hebb* esitti 1949 sopivan säännön päivittää solujen välisten kytkentöjen voimakkuuksia.

1950-luvun alussa alalla toimi jokunen (lähinnä) yhdysvaltalainen tutkija:

- Vuonna 1950 fysikaalisen informaatioteorian luoja *Claude Shannon* kirjoitti shakkiohjelmia – vain tietokone puuttui. . .
- Vuonna 1953 Alan Turing (luvut 1.1.1 ja 1.2) teki samaa Iso-Britanniassa, kunnes hallitus kielsi häneltä tietokoneiden käytön moiseen humpuukiin.
- Vuonna 1951 *Marvin Minsky* ja *Dean Edmonds* rakensivat ensimmäisen hermoverkkokoneen Princetonin yliopiston matematiikan laitoksella.

  John von Neumann (luku 1.2) puolusti Minskyn väitöskirjaa arvostelijoilta, jotka eivät pitäneet sitä matematiikkana.

Samoin Princetonissa vaikuttanut *John McCarthy* siirtyi valmistuttuaan Darthmouth Collegeen, jossa hän, Minsky, Shannon ja *Nathaniel Rochester* (IBM) järjestivät 1956 kesäkoulun. 10 osallistujan joukossa olivat

Trenchard More (Princeton), *Arthur Samuel* (IBM) sekä *Ray Solomonoff* ja *Oliver Selfridge* (MIT).

Muita pidemmällä olivat kuitenkin *Allen Newell* ja *Herbert Simon* (Carnegie Tech, nyt Carnegie-Mellon U. eli CMU): heidän ohjelmansa *Logic Theorist* teki samanlaisia loogisia todistuksia kuin Principia Mathematica (luku 1.2) sisälsi. Hilbertin ohjelma kantoi siis sittenkin hedelmää!

Newellin ja Simonin perusajatuksena oli, että *kone etsii ratkaisua annettuun ongelmaan omaamansa tiedon ohjaamana*.

Alkuinnostus, suuret toiveet (1952–69)

Tänä aikana moni tekoälyn vastaväite tyyppiä "X on mahdotonta tietokoneelle" kumottiin antamalla vastaava ohjelma — kunhan X oli tarpeeksi hyvin rajattu ongelmaperhe, ja esimerkit riittävän pieniä senaikaisille tietokoneille.

Newell ja Simon jatkoivat ohjelmalla *General Problem Solver (GPS)*, joka käytti samanlaisia ongelmanratkaisutapoja kuin ihmiset. Tämä kognitiotieteellinen tekoälytutkimus (luku 1.1.2) on jatkunut CMU:ssa aina näihin päiviin.

Samuel kirjoitti vuodesta 1952 alkaen *oppivia* ohjelmia tammipeliin (checkers), jotka aikanaan ohittivat laatijansa taidot. Rochesterin johdolla jatkettiin Logic Theoristin aloittamaa työtä teoreemien todistamisessa logiikan säännöillä.

McCarthy siirtyi MIT:iin ja vuonna 1958 teki historiaa kolmesti:

- Hän määritteli uuden *AI lab*oratorion ensimmäisessä raportissa ohjelmointikielen *Lisp* (*Lis*t *p*rocessing), joka on edelleen elossa tärkeimpänä tekoälykielenä — ja toiseksi vanhimpana vielä käytettynä kielenä heti vuonna 1957 esitellyn FORTRANin (FORmula TRANslator) jälkeen.

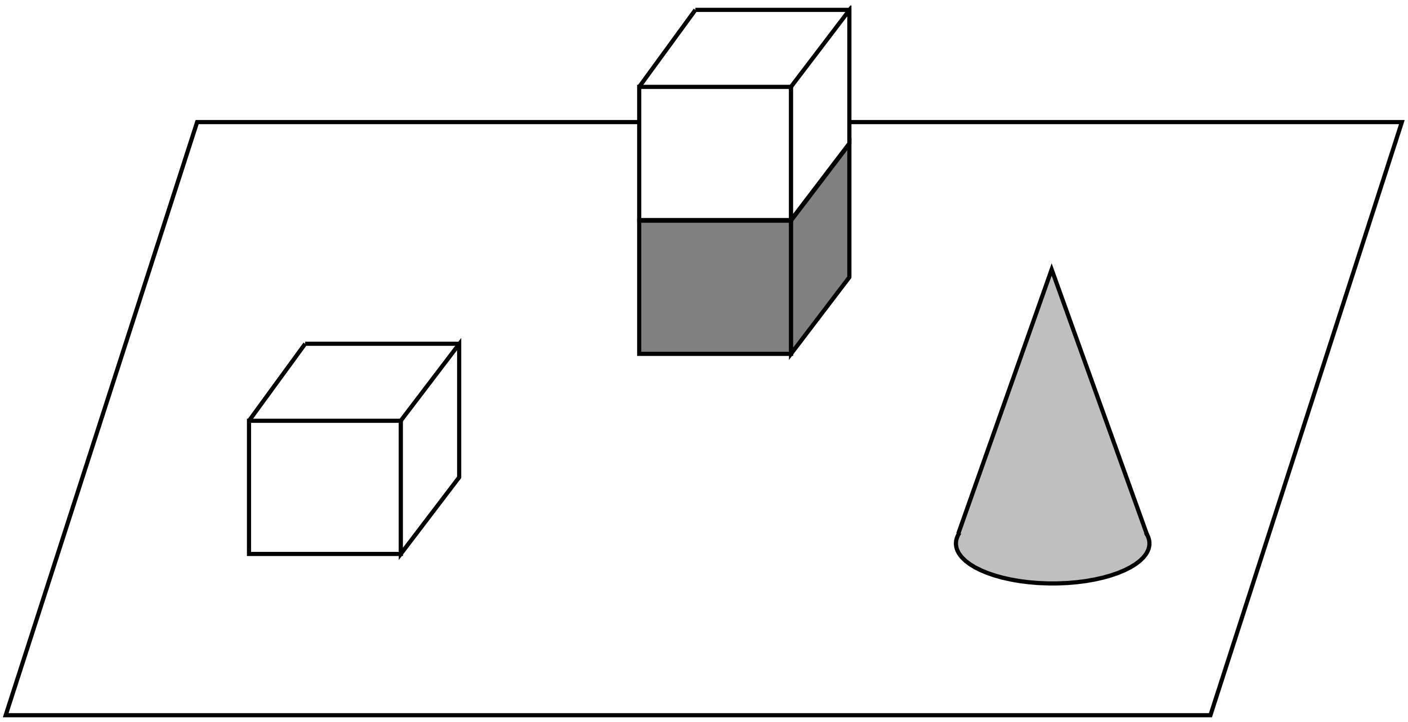
  Kielen perustietorakenne *lista* oli peräisin jo Logic Theoristia varten kehitetystä ohjelmointikielestä IPL.

- Yhdessä muiden MIT:n tietokoneihmisten kanssa hän ideoi *osituskäytön (time sharing)*.
- Hän esitteli — muttei toteuttanut — ehkä ensimmäisen *täydellisen* tekoälyohjelman *Advice Taker*. Perusidea oli Newellin ja Simonin, mutta nyt aksiomatisoitavana olikin matematiikan sijasta *arkimaailma*. Ohjelmaan pystyi myös lisäämään uusia aksiomeja ilman uudelleenohjelmointia, eli erotti toisistaan *tietämyksen esittämisen logiikalla* ja *näiden esitysten käsittelyn deduktiivisilla päättelysäännöillä*.

Samana vuonna myös Minsky siirtyi MIT:iin ja alkoi yhteistyö, joka alkoi rakoilla kun "scruffy" Minsky painotti toimivia ohjelmia logiikan kustannuksella ja "neat" McCarthy taas painotti tietämyksen esittämistä ja loogista päättelyä. Nämä kaksi kantaa ovat ottaneet mittaa toisistaan myöhemminkin — ja täydentäneet toisiaan.

Vuonna 1963 McCarthy perusti Stanford Research Instituten (SRI) ja jatkoi siellä Advice Taker -ideaansa.

MIT:ssa Minskyn oppilaat kehittivät puolestaan ohjelmia käsittelemään (usein simuloituja) *mikromaailmoja*:



Kuva 2: Palikkamaailma (blocks world) [RN95, kuva 1.3].

Palikkamaailmassa tutkittiin menestyksellisesti esimerkiksi *konenäköä*, *oppimista*, *luonnollisen kielen ymmärtämistä* ja *moniosaisten suunnitelmien tekoa*.

Innostus laantuu, toiveet karisevat (1966–74)

Tutkijoiden *ylioptimistiset* odotukset ja ennusteet jäivät toteutumatta, kun ratkaistavat *tehtävät kasvoivat ja monimutkaistuivat* (skaalautumisongelma):

- Monet varhaiset ohjelmat eivät olleet kovinkaan "älykkäitä" vaan suoriutuivat pienistä esimerkkitehtävistään yksinkertaisin syntaktisin manipulaatioin.

  Esimerkiksi luonnollisten kielten kääntämistä toisikseen pidettiin helppona: ohjelmoitaisiin Chomskyn kielioppiteoria ja kerättäisiin elektroninen sanakirja. Kuitenkin tekstin "ymmärtäminen" vaatii *ennakkotietämystä* sen aiheesta — maailmasta, ihmisistä,. . . — moniselitteisyyksien ratkomiseksi:

  $$\text{"Olen nähnyt Helga-neidin } \left\{ \begin{array}{c} \text{kun minä olin} \\ \text{kun hän oli} \end{array} \right\} \text{ kylvyssä, ihanaa!"}$$

- Monet etsintäongelmat tiedetään nyt NP-koviksi (luku 1.2).

  Vaikka ohjelmassa olisikin mekanismi ratkaisemaan tehtävä *periaatteessa*, ei tuo mekanismi välttämättä ratkaise sitä *käytännössä*.

  Kyvyttömyys hallita tämä *kombinatorinen räjähdys* oli (muodollisena) syynä *Lighthill-raporttiin*, joka lopetti Iso-Britannian hallituksen tuen tekoälytutkimukselta kahta yliopistoa lukuunottamatta.

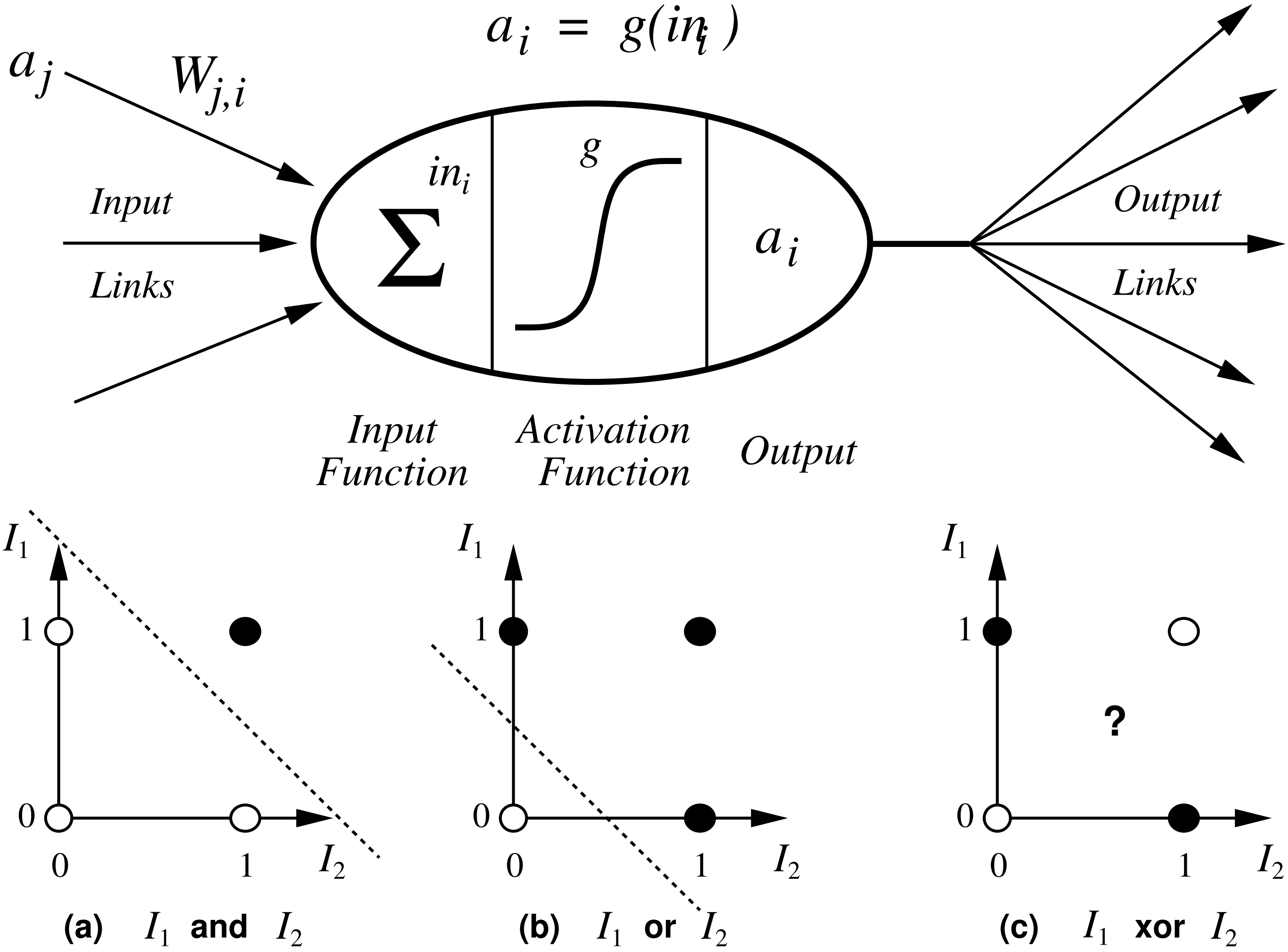
- Hermoverkkotutkimuksessa johdettiin teoreettisia tuloksia käytetyistä tekniikoista:

  1. *Frank Rosenblatt* osoitti vuonna 1962 *perceptronien konvergenssilauseen*, jonka mukaan hänen *oppimisalgoritminsa* pystyi sovittamaan *yhden* hermosolun kytkentäpainot vastaamaan mitä tahansa sellaista syötettä, jolle sopivat painot olivat olemassa.

     *Moni*soluisia verkkoja tutkittiin myös, ja uskottiin, että niillekin saataisiin pian samanlaisia tuloksia.

2. Vuonna 1969 (aiempi hermoverkkotutkija) Minsky ja *Seymour Papert* julkaisivat kuitenkin kirjan *Perceptrons*, jossa he osoittivat että nämä sopivat painot olivat olemassa vain hyvin yksinkertaisille syotteille.

   Vaikka monisoluiset verkot ratkaisevatkin tämän ongelman, ei niille tunnettu vielä oppimisalgoritmeja. Kirjan sisältämän terävän kritiikin (jota 2. painos 1988 lievensi) tuloksena tietojenkäsittelyllisen hermoverkkotutkimuksen rahoitus kutistui olemattomiin. . .



Kuva 3: Perceptron ja Boolen logiikka [RN95, kuvat 19.4 ja 19.9].

3. . . . peittäen alleen sen, että sellainen algoritmi oli juuri keksitty (tosin kansantaloustieteessä)! Hermoverkkojen tutkimus koki renesanssin vasta 1980-luvun loppupuolella, kun se keksittiin uudelleen.

Tietämyspohjaiset järjestelmät (1969–79)

Newell ja Simon olivat uskoneet *heikkoihin menetelmiin*, joissa itse etsintäkone oli yleiskäyttöinen ja syötteenä annetut säännötkin kuvailivat yleispäteviä periaatteita. Näin saadut tuloksetkin olivat usein heikkoja.

Alettiin tutkia vaihtoehtoja, joissa valittiin ensin sovellusalue ja sitten kirjoitettiin hyvin sovellusaluekohtaisia sääntökokoelmia. Ei siis mallinnettu enää "yleistä" älykkyyttä vaan sitä suorempaa päättelyä jota *asiantuntija (expert)* tekee käyttäessään omaa erityisosaamistaan.

DENDRAL, joka päätteli näytteen molekyylirakenteen massaspektrometrin antamista mittaustuloksista, lienee ollut ensimmäinen *tietämysintensiivinen ohjelma*, jonka asiantuntemus johtui suuresta kokoelmasta erityistapauksia koskevia sääntöjä.

MYCIN puolestaan diagnosoi verisairauksia noin 450 säännön voimin lähes yhtä hyvin kuin alan erikoislääkärit. Se erosi DENDRALista kahdessa suhteessa:

- DENDRALin sääntökanta voitiin johtaa kyseeseen tulevista kemian teorioista; MYCINin pohjaksi ei ollut tarpeeksi tarkkaa lääketieteellistä teoriaa, vaan se kerättiin *haastattelemalla* alan spesialisteja.
- MYCINin sääntöihin oli liitettävä *varmuuskertoimia (certainty factors)* heijastamaan tarjolla ollutta vain likimääräistä, ei varmaa sovellusalueen tietämystä. Niistä tuli suosittu tapa mallintaa epävarmuutta sääntöpohjaisissa asiantuntijajärjestelmissä.

  Jos annetun säännöstön ja siitä tehtävien päätelmien rakenne on tietyllä tapaa yksinkertainen (kuten onneksi MYCINissä) niin varmuuskertoimet ovat eräs Bayesin teorian (luku 1.2) erikoistapaus; mutta *yleisesti* niiden lisääminen saattoi antaa aivan harhaanjohtavia tuloksia, joten niiden käyttöa ei enää suositella [RN95, luku 15.6].

Epävarmuuden mallinnukseen (varsinkin lääketieteellisessä diagnoosissa) käytettiin jo tuolloin myös perinteistä todennäköisyyslaskentaa ja utiliteettiteoriaa. Esimerkiksi probabilistinen päättelykone PROSPECTOR ehdotti koeporauksia paikassa, josta löytyi laaja molybdeeniesiintymä.

Luonnollisen kielen käsittelyä tutkittiin joko rajoittumalla yksinkertaisiin ympäristöihin (*Terry Winograd*in SHRDLU palikkamaailmassa) tai lisäämällä tarvittavaa taustatietoa ja päättelyä sen pohjalta (*Roger Schank* oppilaineen Yalessa).

Tosielämän sovelluksia varten kehitettiin myös uusia tiedonesitystapoja:

Prolog on logiikkaan perustuva ohjelmointikieli, joka syntyi Ranskassa ja Iso-Britanniassa (Skotlannissa) luonnollisen kielen käsittelyn tarpeista.

Planner on myöskin logiikkaan perustuva yhdysvaltalainen perhe kieliä monimutkaisten suunnitteluongelmien esittämiseen ja ratkaisuun.

kehykset (frames) ovat Minskyn *rakenteinen* vaihtoehto logiikalle, jossa tietämys esitetään hierarkisina rakenteina ja niiden solmujen välisinä suhteina. *Olio-ohjelmointi* voidaan nähdä tämän vaihtoehdon tuomisena tietämyksen esittämisestä ohjelmointiin.

Tekoälyteollisuuden synty (1980–88)

Alan kaupallinen liikevaihto kasvoi vuoden 1980 muutamasta miljoonasta vuoden 1988 2 miljardiin dollariin.

Japani käynnisti 1981 kansallisen "viidennen sukupolven tietokonehankkeen": pyrittiin Prolog-tekoälykoneeseen jossa olisi esimerkiksi täysimittainen luonnollisen kielen tunnistus. Yhdysvalloissa tämä nähtiin uhkana, joka kasvatti tekoälytyön kiinnostusta ja rahoitusta. Iso-Britanniassa *Alvey-raportti* käynnisti "uuden" tutkimusalueen nimeltä "älykkäät tietämyspohjaiset järjestelmät".

Samalla tuotettiin tekoälytehtäviin sopivia laite- ja ohjelmointiympäristöjä (Lisp-koneet ja -ympäristöt). Nämä markkinat ja niillä toimineet yritykset eivät ole kuitenkaan enää hengissä: tavallisilla mikrotietokoneilla saavutettiin lähes samat edut paljon halvemmalla.

Hermoverkkojen paluu (1986–)

Hermoverkkojen tutkimus oli jatkunut tietojenkäsittelyn ulkopuolella:

Fyysikot kuten *John Hopfield* mallinsivat niitä ja niiden vuorovaikutuksia alunperin kiinteän aineen atomeille kehitetyillä menetelmillä.

Psykologit kuten *David Rumelhart* ja *Geoff Hinton* mallinsivat niillä (ihmis)muistia.

1980-luvun puolivälin tienoilla monisoluisten verkkojen oppimisalgoritmi keksittiin uudelleen — ainakin 4 kertaa. . . — ja sitä sovellettiin moniin tietojenkäsittelyn ja psykologian oppimisongelmiin.

Nykytila (1987–)

Tekoälytutkimus on vakiintumassa "normaalitieteen tilaan", jossa "neat"-tutkijat testaavat ja kehittävät edelleen jo olemassaolevia teorioita todellisiin tilanteisiin. (Uusi "scruffy-vallankumous" voi olla tarpeen kunhan teorioiden rajat tulevat vastaan.)

Esimerkiksi puheen- ja käsialantunnistus on edennyt kaupalliseen käyttöön saakka, kun otettiin käyttöön *piilo-Markov-mallit* todennäköisyyslaskennasta aikaisempien *ad hoc*-ratkaisujen sijaan. Samoin epävarman tiedon hallinnassa on edetty korvaamalla varmuuskertoimet Bayesin ja päätösteorian menetelmin.

Varsinainen tavoite – "aito" tekoäly – näyttääkin olevan (vähintään) yhtä kaukana nyt kuin alkuinnostuksen vuosina; tällä välin olemme oppineet paljon sen hankaluuksista. . .

Kurssimme "älykäs toimija" -määritelmä tekoälylle onkin "kattokäsite" yhdistämään käsitteellisesti tekoälyn eri osa-alueet, ei lähitulevaisuuden tuotekehittelyä.

1.6. Tekoälyn omasta filosofiasta [RN95, luku 26]

Tekoälyn filosofia jakaa sitä koskevat väitteet kahteen päätyyppiin:

Heikko tekoäly kysyy "voidaanko kone saada käyttäytymään *ikään kuin* se olisi älykäs?"

Tämän väitteen arvostelijat esittävät seuraavanlaisia kantoja:

- On asioita, joihin tietokoneet eivät pysty, ohjelmoitiinpa ne miten tahansa.
- Tietyt tavat suunnitella älykkäitä ohjelmia tulevat epäonnistumaan.
- Näiden ohjelmien laatiminen on käytännössä mahdotonta.

Vahva tekoäly kysyy "onko tällaisella koneella *tietoisuus (consciousness)*?"

- Turingin määritelmä (luku 1.1.1) väitti, että heikosta tekoälystä seuraa vahva.
- Näin voi uskoa uskomatta heikkoon: jos voitaisiin, niin olisi.

Turing luetteli myös useita (heikon) tekoälyn vastaisia argumentteja [Tur50, luku 6], joista osa elää tänäänkin, esimerkiksi:

Matemaattinen vastaväite lähtee Gödelin (tai Turingin omista) epätäydellisyystuloksista (luku 1.2), joiden mukaan mikään formaali systeemi — kuten tietokone — ei voi ymmärtää tiettyjä matemaattisia väittämiä jotka koskevat sitä itseään.

Gödel on ne kuitenkin todistanut ja muut matemaatikot todistuksen ymmärtäneet. Siis he ovat tehneet koneelle mahdottoman älykkyyttä osoittavan ja vaativan toimituksen.

Eräs vastaus on, että vaikka formaali systeemi — matemaatikko tai tietokone — pystyykin käsittelemään *itseään "yksinkertaisempien"* systeemien väittämiä, se ei pystykään tekemään samaa omalle väitteelleen. Tämä erottaa "kielen" *jota* tarkastellaan ja "metakielen" *jossa* tarkastellaan.

Roger Penrose esitti tämän vastaväitteen kirjassaan "Keisarin uudet aatteet" vuonna 1989 [Pen89]. Penrose. . .

- halusi toisaalta pitää kiinni aivotoimintojen ja sitä kautta myös tietoisuuden fysikaalisesta perustasta,
- mutta toisaalta halusi kieltää niiden mekaanisen laskettavuuden.

Hän päätyi arvelemaan, että aivoissa tapahtuvat kvanttimekaaniset ilmiöt ovat tärkeitä tietoisuudelle. (Liekö lääke tautia pahempi?)

Käyttäytymisen epämuodollisuus on vastaväite, jonka mukaan inhimillinen käyttäytyminen on niin monimutkaista, ettei sitä voi vangita sääntöihin.

Näin on ajatellut esimerkiksi *Hubert Dreyfus* [DD88].

Kannan edustajilla on kuitenkin jälleen kaksijakoinen ongelma:

- Antaa vastaesimerkki älykkyyttä vaativasta toiminnasta, jota ei voi kuvailla säännöin.
- Antaa vaihtoehtoinen selitys omalle esimerkilleen ilman että samalla tulee itse kehittäneeksi sille säännöt.

Tunnetuin vastaväite vahvalle tekoälylle lienee *John Searlen "kiinalainen huone" (Chinese Room) -ajatuskoe* [Sea80, Sea90, CSC90][Bod90, §4]:

- Istun kopissa, jossa on vain
  - yksi luukku nimeltä 'sisään'
  - toinen luukku nimeltä 'ulos'
  - ja valtava kirja jonka kullakin sivulla on kiinalainen kirjoitusmerkki

  enkä minä osaa kiinaa
- mutta osaan tehdä seuraavaa:
  - kun 'sisään'-luukusta näytetään minulle jokin merkki $\mathfrak{A}$
  - niin etsin kirjasta sitä vastaavan aukeman $\boxed{\mathfrak{A}}\boxed{\mathfrak{B}}$
  - ja näytän 'ulos'-luukusta vastaavan merkin $\mathfrak{B}$.
- Läpäisköön tämä koppi Turingin testin keskustelemalla kiinaksi täysin ihmisen tavoin.

  Kuka tai mikä kopissa
  - osaa kiinaa?
  - on tietoinen (keskustelluista asioista)?

Searlen mukaan ei kukaan eikä mikään!

- Searlen johtopäätökset:
  - Ohjelmaa suorittava tietokone ei ole tietoinen — huonehan on eräänlainen Turingin kone.
  - Tietoisuutta ei voi koodata ohjelmaksi — kirjahan on eräänlainen ohjelma.

Searle haluaa tämän vuoksi eroon tietoisuuden Turing-määritelmästä pelkkänä syöte/tulos-vastaavuutena.

Toisaalta Searle (kuten Penrose) haluaa pitää kiinni tietoisuuden fysikaalisesta perustasta aivotoiminnassa: "brains cause minds".

Searlen haluja voi toisaalta pitää ristiriitaisina *aivoproteesikokeen (the Brain Prosthesis Experiment)* nojalla:

- Oletetaan lääketieteen kehittyneen niin pitkälle, että yksittäinen neuroni voidaan korvata täsmälleen saman syöte/tulos -käyttäytymisen omaavalla mikropiirillä ilman että aivojen toiminta häiriintyy leikkauksessa.
- Vaihdetaan saman potilaan kaikki neuronit *yksi kerrallaan*. Miltä potilaasta itsestään tuntuu, häviääkö hänen tietoisuutensa? Searlen mukaan häviää; tutkitaan mitä tästä seuraa.
- Koska yksittäisten aivonosien syöte/tulos -käyttäytyminen ei muutu, niin koko potilaan havaittava käyttäytyminen ei muutu yllättäen.

  Tällöin potilaan *oma tahto katoaa yhtäkkiä*, muuten hän ehtisi huudahtaa "apua!"

  Siis on jokin "kriittinen massa" neuroneita, jotka juuri ja juuri riittävät oman tahdon olemassaoloon.
- Kun kaikki vaihdot on suoritettu, on saatu alkuperäisten aivojen kanssa syöte/tulos -käyttäytymiseltään identtiset "sähköaivot", joilta Searle kuitenkin kieltää tietoisuuden.

  Silloin potilaan alkuperäistenkään aivojen tietoisilla tiloilla ei ollut mitään vaikutusta hänen käyttäytymiseensä, koska nämä tilat ovat nyt hävinneet mutta käyttäytyminen säilyi.

- Peruutetaan lopuksi tämä leikkaus jälleen neuroni kerrallaan, ja kysytään sen jälkeen "miltä leikkaus tuntui?"

  Koska potilaan neuronien syöte/tulos -käyttäytymistä ei taaskaan häiritty, hänen pitäisi muistaa tuntemukset myös siltä ajalta jolloin aivoissa ei ollut yhtään neuronia.

  Searlen mukaan sinä aikana ei aivoissa kuitenkaan ollut ketään kokemassa noita tuntemuksia.

Toisaalta todellinen neurotutkimus ei ehkä salli näin teräviä tulkintoja:

- Tietoisuuden "kriittinen neuronimassa" ei ehkä olekaan kiinteä suure: tietoisuus saattaa tarvita kullakin hetkellä käytössään olevien aktiivisten neuronien lisäksi *vapaita lisäneuroneita* jotka voivat aktivoitua tulevaisuudessa samalla kun osa nykyisistä passivoituu. [TE88]

  Tietoisuuden kulloinenkin sijainti voi siis "matkailla" aivoissa ajan kuluessa, ja tarvita siihen vapaata kulkutilaa.

- Aivot ovat myös yksi *elin* joka on yhteydessä muuhun ruumiiseen myös muuten kuin lähettämällä ja vastaanottamalla hermoimpulsseja.

  Tietoisuus saattaa olla jopa *evoluution sivutuote* matkalla kohden yhä tarkempaa eliön itsesäätelymekanismia [Dam99].

  Tietoisuutta ei silloin ehkä voi tutkia erottamalla aivoja muusta ruumiista tai edes ei-tietoisista aivoprosesseista.

Viitteet

[AK83] J.F. Allen and J.A. Koomen. Planning using a temporal world model. In *International Joint Conference on Artificial Intelligence*, pages 741–747, 1983.

[AS96] H. Abelson and G.J. Sussman. *Structure and Interpretation of Computer Algorithms*. The MIT Press, second edition, 1996.

[Bod90] M.A. Boden, editor. *The Philosophy of Artificial Intelligence*. Oxford Readings in Philosophy. Oxford University Press, 1990.

[Car97] B. Carpenter. *Type-Logical Semantics*. The MIT Press, 1997.

[Cho97] N. Chomsky. B. F. Skinnerin teos *verbal behaviour*. In P. Raatikainen, editor, *Ajattelu, kieli, merkitys*, pages 279–310. Gaudeamus, 1997. Suom. R. Vilkko.

[CL73] C-L. Chang and R.C-T. Lee. *Symbolic Logic and Mechanical Theorem Proving*. Academic Press, 1973.

[CR91] W. Clinger and J. Rees. Revised^4 report on the algorithmic language Scheme, 1991.

[CSC90] P.M. Churchland and P. Smith Churchland. Could a machine think? *Scientific American*, pages 26–31, 1990.

[Dam99] A. Damasio. *The Feeling of What Happens*. William Heinemann, 1999.

[DD88] H.L. Dreyfus and S.E. Dreyfus. Making a mind versus modelling the brain: Artificial intelligence back at a branch-point. *Artificial Intelligence*, 1988.

[Dea98] T.W. Deacon. *The Symbolic Species: The Co-Evolution of Language and the Human Brain*. Penguin Books, 1998.

[MP43] W.S. McCulloch and W.H. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–137, 1943. [Bod90, §1].

[Myr74a] L. Myrberg. *Differentiaali- ja integraalilaskenta*, volume 1. Kirjayhtymä, 1974.

[Myr74b] L. Myrberg. *Differentiaali- ja integraalilaskenta*, volume 2. Kirjayhtymä, 1974.

[Nil98] N.J. Nilsson. *Artificial Intelligence: A New Synthesis*. Morgan-Kaufmann, 1998.

[NKN01] M.K. Nowak, N.L. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.

[Pen89] R. Penrose. *The Emperor's New Mind*. Oxford University Press, 1989.

[Pin94] S. Pinker. *The Language Instinct*. William Morrow and Company, Inc., 1994.

[Pin97] S. Pinker. *How the Mind Works*. W. W. Norton & Company, Inc., 1997.

[RN95] S. Russell and P. Norvig. *Artificial Intelligence: a Modern Approach*. Prentice-Hall, 1995.

[Saa85] E. Saarinen. *Länsimaisen filosofian historia huipulta huipulle*. Werner Söderström oy, 1985.

[Sea80] J. Searle. Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3:417–424, 1980. [Bod90, §3].

[Sea90] J. Searle. Is the brain's mind a computer program? *Scientific American*, pages 20–25, 1990.

[TE88] T. Tononi and G.M. Edelman. Consciousness and complexity. *Science*, pages 1846–1851, 1988.

[Tur50] A.M. Turing. Computing machinery and intelligence. *Mind*, LIX(2236):433–460, 1950. [Bod90, §2].