

Az intelligencia kialakítása

Lehmann Miklós
lehmann@ludens.elte.hu

A kognitív folyamatok szimulálására és modellezésére immáron sok és változatos elmélet áll rendelkezésünkre. Azon túlmenően, hogy a mesterséges intelligencia (MI) megvalósításának lehetősége/lehetetlensége körüli vita sem jutott még nyugvópontra, a kutatási irányok nagyjából három jellegzetes típust képviselnek, amelyek egyben a kognitív folyamatok mibenlétéről vallott felfogást is tükrözik. Előadásomban legelőször is ezeket szeretném röviden áttekinteni, majd a harmadik típus kapcsán felvázolnám azt az irányt, amely állásponantom szerint leginkább gyümölcsözőnek tekinthető.

Legegyszerűbbnek tűnik, ha már most megfogalmazom azt a tézist, mely gondolatmenetemben központi helyet foglal el: *a tervezés paradoxona*, hogy egy szigorú értelemben tervezett rendszer nem tekinthető intelligensnek; márpedig a MI kutatások jelenlegi (egyik) általános alpelve éppen a tervezés – mind strukturális, mind algoritmikus értelemben. Ennyiben a MI definíció szerint lehetetlen lenne, hiszen tervezett intelligenciát takar. Feladatunk tehát, hogy megkíséreljük áthidalni e paradoxont.

A matematikai reprezentáció

A korai MI kutatás fő jellemzője, hogy a kognitív képességek matematikai reprezentációjára törekedett. Előfeltevése, hogy a megismerőfolyamatok matematikai komputációs eljárásoknak feleltethetők meg, és mint ilyenek, matematikai rendszereken modellezhetőek. Az elmélet erénye, hogy meglepően sikeres volt: a vizsgált területek nagy részén megfelelő pontossággal volt képes a kogníció szimulálására. Ez azonban túlzott általánosításokhoz vezetett: a szimulációt azonosította a modellezéssel, majd a valós működéssel, így a kezdeti sikereket elégedetlenség követte.

A kutatások kezdetektől fogva karöltve jártak a számítógépes kutatásokkal, ezért szükséges rámutatnunk, hogy a matematikai-elv nem függetleníthető a számítógép-elvtől ("számítógép-analógia"). Ennek tudható be a szimbólummanipulációs elmélet gyors térhódítása is. A ma már klasszikussá vált szerzők, mint Alan Turing, Alonzo Church és Herbert Simon,¹ a gondolkodást szimbólummanipulációval azonosították, amely persze közvetlenül lehetővé tette annak számítógépes megjelenítését.

Ennek kapcsán a következőkre érdemes felfigyelnünk. A kogníció fogalmát e szerzők lényegében a racionális gondolkodás értelmében használták, melyet logikai-matematikai eszközökkel leírhatónak vélték. Ennek köszönhető

¹ Ld. pl. A. M. Turing (1964), H. A. Simon (1982), A. Newell és H. A. Simon (1972). A komputációs paradigma Johnson-Laird (1987a) szerint három fő vonással rendelkezik: a komputációt a belső reprezentációk (1) konstruálására, (2) manipulálására, (3) interpretálására használják.

a kezdeti sikerek is.² Az elmélet akkor kerül szorult helyzetbe, amikor be kívánja illeszteni azokat a folyamatokat is, melyek nem ilyen egyértelműen foglalhatók algoritmusokba.

Másrészt a szimbólummanipuláció korai elmélete nyitva hagyja a szimbólumok és jelentésük problematikáját. E tekintetben ember és gép radikálisan különböző rendszerek: míg az ember számára a szimbólumok kezelése és a jelentésadás egyetlen rendszerbe épül, a gép számára a jelentés extrinzik, azaz az embertől függő. Lehetetlen vállalkozásnak tűnik, hogy a jelentéseket ne magunk, hanem a gép számára határozzuk meg. Az interpretáció sem oldja meg a problémát, csupán áthidalja – számunkra, hiszen, leegyszerűsítve a dolgot, egy gép sosem interpretálja saját működését.

A kérdés egyben szintaxis és szemantika kapcsolatához is kötődik. A szimbólummanipulációs rendszerek tisztán szintaktikaiak; az algoritmus a szimbólumok egymásba való komputációs átalakítását és kapcsolatrendszerét határozza meg. Az interpretáció itt sem hoz megnyugtató megoldást, mivel a szemantika a gép számára megintcsak külső lesz. Ezen kívül a metamatikai reprezentáció meglehetősen szegényes nyelvet tesz lehetővé. A természetes nyelvek gazdagsága nem egykönnyen implantálható mesterséges rendszerekbe (olyan okok miatt, mint pl. a kontextusfüggőség), ez pedig megkérdőjelezi, hogy a gépi szimbólumfeldolgozás ténylegesen az emberét modellálja-e.

A formális rendszerek tehát jól működnek logikai környezetben, de amennyiben az emberi intelligencia nem írható le tökéletesen algoritmusok segítségével, csődöt mondanak. Nem véletlen, hogy a klasszikus MI legnagyobb sikereit a sakkautomaták terén érte el, ahol megfelelő kapacitással és matematikai elvekkel előkelő helyre került a világranglistán.

A formális elvek használhatóságát kérdőjelezi meg John R. Searle (1996) jól ismert *kínai szoba* argumentuma is. Szerinte a formális elvek egyszerű követése nem hoz létre intelligenciát, mivel erre bárki (és bármilyen szimbólumfeldolgozó gép) képes megértés nélkül. Azaz, megintcsak a jelentésadás problematikájába ütközünk: a megértésre csak az képes, aki a szimbólumok szemantikájával is tisztában van; a szintaxis ismerete ehhez elégtelen.

Más megközelítésből a megértés kérdése a mikrovilágokból való kilépéskor kerül előtérbe. John Haugeland (1996) szerint a MI mikrovilágokban való sikeressége eltűnik a mindennapokra való alkalmazásban, melyre nem adhatók olyan explicit szabályok, mint a jól körülhatárolt területekre. A feladatok fokozódó összetettségével ezek a rendszerek képtelenek megbírkózni – mivel ehhez egyre növekvő háttérinformációra lenne szükség. Az ebben való eligazodás pedig csak meglehetősen kifinomult heurisztikák segítségével válik lehetségessé.³

² Ma már a racionalitást sem látjuk ennyire egyszerűnek. Például az érzelmek racionális gondolkodásban betöltött szerepéről ld. A. R. Damasio (1996).

³ Csak jelzem az ehhez kötődő jól ismert problémákat: a kapacitás végeességét és a valós időben történő működés igényét.

A fizikai reprezentáció

Amikor a kutatások a logika gondolkodásbeli alkalmazhatóságának korlátjaiba ütköztek, a vizsgálódás előterébe kerültek olyan egyszerű jelenségek, mint a hétköznapi viselkedés vagy a tájékozódás. A komputációs elmélet ezek megoldásánál eleinte egyre bonyolultabbá válik, majd működésképtelen lesz. A megoldást sokan a fizikai reprezentációkban látták, melyek sokkal rugalmasabban képesek követni ezeket a jelenségeket.

A hétköznapi képességek megjelenítőinek jellegzetes képviselői a "csevegő" vagy társalgó rendszerek, mint amilyen az ELIZA (Weizenbaum, 1966) vagy később a NENE (Toda, 1986). Ez utóbbi a fizikai reprezentációk előnyeit kihasználva már egy többé-kevésbé teljes személyiség megjelenítését tűzte ki célul, így a gép virtuális modellel rendelkezik saját magáról, saját alternatív állapotairól, valamint reprezentációja van ezen állapotok preferálására szolgáló érvényességi rendszerről. Tudásbázisát a hétköznapi társalgási nyelv és az ehhez kapcsolódó alapvető, de szerteágazó ismeretek adják.⁴

A statikus matematikai modellek szigorú logikai rendjét így sikerül áttörni a fizika flexibilisebb dinamikájával, de egyben a korábbi eredményekről sem kell lemondani.⁵ Ez az elmélet a matematika algoritmusait a fizika differenciálegyenleteire cseréli fel, sőt, ezzel úgy véli, sikerül kiküszöbölni a komputáció-problémát.

Hasonló eredményre jut a konnekcionista irányzat is. Az intuíció kérdését vizsgálva például, annak algoritmizálhatatlansága miatt Paul Smolensky (1996) arra a megállapításra jut, hogy szimbólummanipuláció helyett szubszimbolikus szinten kellene a megoldást keresnünk. A formális leírás nem alkalmas olyan képességek modellezésére, melyek meglehetősen köznapiak (mint pl. az intuíció). A konnekcionizmusnak azonban nincs szüksége teljes formális leírásokra, tehát elképzelhető, hogy könnyebben megbírkózik ezekkel a feladatokkal.

Csakhowy ezen a ponton a következő problémával kell szembenéznünk. Valójában a fizikai rendszerek sem képesek meghaladni a komputáció kérdését, mindössze annyit sikerül elérniük, hogy azt bonyolultabbá és szofisztikáltabbá tegyék. Ez pedig a rendszer különleges korlátaira hívja fel a figyelmet: minél "intelligensebbé" teszünk egy rendszert, annál speciálisabb lesz, azaz annál jobban szűkül felhasználhatósági köre. Ahelyett, hogy általános gépeket alkotnánk (ahogyan azt korábban pl. Simon az Egyetemes Problémamegoldóval kívánta), szakértő rendszereket és célgépeket hozunk létre. A lényeg azonban nem változott: továbbra is programokat írunk és átalakítási képleteket táplálunk a rendszerbe⁶ – Searle tehát továbbra sem lenne elégedett.

⁴ M. Toda (1987). A hétköznapi képességek és jelenségek matematikai modellezhetőségéről P. N. Johnson-Laird a következőképpen nyilatkozik: "...a mindennapi élet értelmezése nem egy logikai kalkulus formális következtető szerepétől függ..."; azaz annak megragadása lehetetlen ilyen formában (P.N. Johnson-Laird, 1987b)

⁵ A fizikai reprezentáció egyik legjobb kifejtése található Roger Penrosenál (1993), aki, bár igencsak szkeptikus a MI megvalósíthatóságát illetően, az emberi intelligencia törvényeit a kvantummechanika területén véli felfedezni.

⁶ Jó áttekintés található a MI kutatások irányelveinek változásairól Casimir A. Kulikowskinál (1988)

A kezdeti nyelvi modellezés, a környezet modellálása, a heurisztikák alkalmazása – ami a fizikai reprezentációk megoldása az emberi és gépi intelligencia szimulálására – tehát valójában nem lépnek túl a szimuláción, nem visznek sokkal közelebb sem a természetes intelligencia megértéséhez, sem a MI megvalósításához.⁷ Más, olyan modell irányába kell a kutatást folytatnunk, amely reális alternatívát jelenthet a logikai, komputációs megoldások számára.

Komputáció, vagy...

Az utóbbi évtizedben e problémák hatására alternatív elgondolások jelentek meg. Nem kívánom most ezeket teljesen áttekinteni, de érdemes egy példán felvázolni azokat az új szempontokat, melyek ezen elméleteket nagy általánosságban meghatározzák.

A komputációs eljárásokkal helyezkedik szembe Jaakko Hintikka játékelméleti modelljével.⁸ A kogníció véleménye szerint jobban modellálható a játékok strukturális jellemzői segítségével; így jobban képes például azt a hosszú sor aktivitást bevonni elméletébe, melyek mind a belső kognícióban, mind a külvilággal való kapcsolatban egyaránt megjelennek a játékok és a kognitív tevékenységek során. A játékok modellezésére nem található egyértelmű matematikai háttér, bár néhány megszorítással közelíteni lehet hozzá (pl. általános stratégia helyett a matematika nyerőstratégiát alkalmaz). Lényeges különbség még, hogy sikerül meghaladni a kogníció eddig többé-kevésbé szolipszista felfogását: a játékok sokszor kooperációt kívánnak.

Míg tehát a gondolkodás racionalitásába vetett hit most sem változott (bár ez a racionalitás sokkal tágabb értelemmel bír, mint a logikai racionalitás), addig a formalizálás zártságát sikerült fellazítani. A környezeti modellálás, a nyelvi reprezentáció megmaradt lényeges összetevőként, de hozzájuk csatlakozott az emberek közti interakció egyfajta szimulációja.

De vajon tényleg sikerült-e megkerülnünk a komputáció problémáját? A helyzet az, hogy amennyiben csak szimulálni kívánjuk a kogníciót, akkor igen; de ha intelligenciát szeretnénk létrehozni, akkor nem. Gépeink csak a programnak képesek engedelmessé válni, és tegyük bármannyire is bonyolulttá a programot, nem lesznek intelligensek. Úgy vélem, csak abban az esetben neveznénk valóban intelligensnek egy rendszert, ha képes lenne *eltérni* programjától (pl. a környezeti változásokhoz alkalmazkodni). Az intelligencia tehát váratlanul érne: a programtól való eltérésben, a benne megjelenő elvárásainknak való ellentmondásban nyilvánulna meg. Mondhatni, a MI kutatás téves úton jár: programokkal kívánja elérni azt, amit csak a programok ellenében valósíthatna meg.

Éppen ezért úgy gondolom, nem beszélhetünk az intelligencia – a címben is kifejezett – kialakításáról, csupán *előidézéséről*. Azonban a helyzet mégsem annyira reménytelen, mint amilyennek tűnik.

⁷ Természetesen ennek másik lényeges oka, hogy az intelligenciafogalom mibenlétéről korántsem teljes a konszenzus – erre azonban most nem térek külön ki.

⁸ A Church-tézissel szembeszegezett Hintikka-tézis így hangzik: “Minden racionális kognitív tevékenység eljátszható” (idézi Johan van Benthem, 1990).

Jelenlegi eszközeinkkel csak a programok útján tudunk dolgozni, és ez, mint láttuk, megkérdőjelezi sikerességünket. De esetleg programjaink flexibilisebbé tételével elindulhatunk egy másik irányba is, amely a program szerepének radikális változásával járna: eddig a program "tökéletességétől" vártuk az intelligenciát, egy más irány viszont igyekezne kiszorítani a programot központi helyéről.

Kézenfekvő, hogy amennyiben "előidézéstről" beszélünk, vegyük fontolóra a fejlődéseméleti megoldást. Ekkor a következőképpen lehet felvázolni a MI megvalósításának lehetőségét: megfelelő kiinduló feltételeket kell biztosítanunk, valamint egy olyan keretet, amely kellőképpen rugalmasan képes a tényleges megvalósulást irányítani. A folyamatban a programozás elsődleges szerepét át kell, hogy vegye a tanulás és az alkalmazkodás. Erre áthozhatjuk például az evolúciós elméletekből a szelekciós elvet, amely a lehetséges alternatívák és a környezeti viszonyok meghatározottságait illeszti össze.⁹ Durván leegyszerűsítve, csak annyit tehetünk, hogy létrehozzuk az intelligencia *lehetőségét*; a többi már magának a rendszernek a dolga.

Megpróbálom mindezt egy lingvisztikai analógiával megvilágítani. Noam Chomsky (1985) generatív grammatikája a nyelvhasználatot egy "univerzális nyelvtanból" eredezteti, amely az összes potenciális nyelv alapját adja meg. A konkrét megvalósulási folyamat a fejlődési elvek és a környezeti hatások (azaz az adott nyelvi környezet) eredője (lényegében ez már előfeltételezi a fejlődéseméleti megközelítések egy fajtáját). Mindez – viszonylag egyszerű elvek alapján – olyan komplex rendszert eredményez, mint a mai élő nyelvek. Úgy gondolom, ezzel analóg módon dolgozhatnánk a MI kutatásban is, hiszen ez lehet az, ami sikeresen oldja meg a mindennapiság problémáját – míg a logikai alapokon létrehozott mesterséges nyelvek, legyenek bármennyire is bonyolultak, képtelenek erre.

Felmerül a kérdés: vajon a generatív grammatika paraméter-rögzítése lenne a komputáció-probléma megoldása? Erre nem mernék megnyugtató választ adni. Mindenesetre esélyes alternatívának látszik, mivel nyitott egy sor előre lefektetett kombinatorikus lehetőség egészen széles spektrumára. A nyelvi és fejlődéseméleti modellek felexibilitása azt sugallja, hogy talán sikerülhet megoldani a programozás kérdését egy olyan elvvel, mely csupán a legalapvetőbb konstrukciós keretek meghatározására szorítkozik. A továbbiakat "magának a gépnek" kell elvégeznie.

Át kell tehát fogalmaznunk a kérdést és a választ: létrehozható-e a MI? helyett, előidézhető-e a MI? Az előbbire a válasz nem – ha jelenlegi eszközeinket és kutatási irányainkat vesszük alapul. Az utóbbira a válasz igen – ha képesek vagyunk meghatározni azokat a flexibilis kereteket, melyek lehetővé teszik, valamint elméletünkbe be tudunk építeni egyfajta fejlődéseméleti megközelítést is. Ekkor a konstrukciós elvek és feltételek meghatározása megmarad ugyan programozási feladatnak, magának a gépnek a működése azonban túl fog lépni a megírt programon.

⁹ Vö. M. Piattelli-Palmarini (1996)

Hivatkozott irodalom

- van BENTHEM, J.(1990): Computation versus Play as a Paradigm for Cognition. In *Language, Knowledge and Intentionality* (ed. L. Haarparauta, M. Kusch, I. Niiniluoto), Akateeminen Kirjakauppa, Helsinki
- CHOMSKY, N.(1985): *Generatív grammatika*. Európa Bp.
- DAMASIO, A. R.(1996): *Descartes tévedése*. Aduprint Bp.
- HAUGELAND, J.(1996): A kognitívizmus jellege és kézenfekvő volta. In *Kognitív tudomány* (szerk. Pléh Csaba)
- JOHNSON-LAIRD, P.N.(1987a): How could consciousness arise from the computations of the brain? In *Mindwaves* (ed. C. Blakemore, S. Greenfield), Basil Blackwell, Oxford
- JOHNSON-LAIRD, P.N.(1987b): The comprehension of discourse and mental models. In *Language and Artificial Intelligence* (ed. M. Nagao), Elsevier Science Publishers BV. North-Holland
- KULIKOWSKI, C.A.(1988): Artificial intelligence, modelling and simulation. In *Artificial Intelligence, Expert Systems and Languages in Modelling and Simulation* (ed. C. A. Kulikowski, R. M. Huber, G. A. Ferraté), Elsevier Science Publishers BV. North-Holland
- NEWELL, A. és SIMON, H.(1972): *Human problem solving*. Prentice Hall, Englewood Cliffs
- Penrose, R.(1993): *A császár új elméje*. Akadémiai Bp.
- PIATTELLI-PALMARINI, M.(1996): Evolúció, szelekció és megismerés. In *Kognitív tudomány* (szerk. Pléh Csaba), Osiris Bp.
- SEARLE, J.(1996): Az elme, az agy és a programok világa. In *Kognitív tudomány* (szerk. Pléh Csaba), Osiris, Bp.
- SIMON, H.(1982): *Korlátozott racionalitás*. Közgazdasági, Bp.
- SMOLENSKY, P.(1996): A konnekciónizmus helyes kezeléséről. In *Kognitív tudomány* (szerk. Pléh Csaba), Osiris Bp.
- STERNBERG, R.J.(1990): *Metaphors of Mind*. Cambridge University Press
- TURING, A.(1950): Computing machinery and intelligence. In *Mind*, 59.
- TODA, M.(1987): NENE, a chetting machine: a tool to investigate information integration in the mind. In *Language and Artificial Intelligence* (ed. M. Nagao), Elsevier Science Publishers BV. North-Holland