

A kogníció alapjai

Kérdésfeltevés

Ahogy folyamatosan fejlődik a mesterségesintelligencia-kutatás, számítógép-tudomány, robotika stb., tanúi lehetünk az egyre szofisztikáltabb technológiai vívmányoknak, olyan információfeldolgozó és -vezérlő rendszereknek, amelyekről a szakértők folyamatosan úgy beszélnek, mint „intelligens ágensekről”, „autonóm robotokról”, „mesterséges-kognitív entitásokról” stb. A mentális nyelvezet használata e technológiai alkotások értelmezésében olyan természetességgel történik, hogy eközben mintha néha feledésbe merülne egy alapvető különbség a mentális fogalmak kétféle értelme között. Hogy pontosabb legyek: gyakran megeshet, hogy egy rendszer viselkedése vagy működése látszólag és megtévesztően az elme jelenlétéről árulkodik. Amikor arról beszélünk, hogy egy számítógépnek memóriája van, hogy képes logikai következtetések levonására, matematikai számításműveletekre, játzsmákra stb., egyértelmű, hogy a jelenséget mentálisan ragadjuk meg, vagyis olyan fogalmakat használunk, amelyek intencionálisak, reprezentációsak, teleológikusak (célorientáltak) stb. Ezzel a szóhasználattal önmagában semmi gond nincs, de könnyen becsaphat, ha nem tartjuk szem előtt, hogy e fogalmaknak kétféle értelme lehet. A számítógép például sakkozik bizonyos értelemben (bizonyos formális szinten), anélkül azonban, hogy ezt a folyamatot kísérné olyasmi, mint a megértés, a versenyszellem, a küzdelem, a győzelem vágya vagy ehhez hasonlók. Ilyen értelemben tehát a számítógép sakkozása egyáltalán nem egy intencionális esemény, és amennyiben a szemantika konstitutív szerepet játszik a sakkbán, a számítógép a szó szoros értelmében nem is sakkozik. Az alapvető különbség tehát nagy vonalakban a valóságos intencionálitás és valóságos elme, másrészt a csak látszólagos vagy csak pszeudomentálitás között húzható meg. Értelmezhetünk egy rendszert mentálisan, anélkül azonban, hogy interpretációnk szó szerint megfelelné annak, ami ontológiailag történik. A probléma persze az, hogy attól, amiért egy rendszer működése értelmezhető vagy kiszámítható, kezelhető stb. úgy, mintha mentális eseményeket tükrözne, vagy azokból kifolyólag történne,¹ még mindig nyitva hagyja a kérdést, hogy az értelmezés megfelel a tényeknek vagy sem.

Ezzel a problémával kapcsolatban a mesterségesintelligencia-kutatásban egy különös jelenséggel találkozunk: egyes teoretikusok szó szerint gondolják azt, hogy bizonyos mesterséges alkotások intrinzikusan rendelkeznek a kognitív élet, az elme, az intelligencia meghatározott formáival. A probléma a szemükben nem az, hogy vajon egyáltalán lehetséges-e mesterséges intelligencia, hiszen számukra egyértelmű, hogy az nemcsak lehetséges, hanem már eleve működik, és konkrét példák is bizonyítják létezését.² Ezen hitet követő kérdés inkább az, hogy az

¹ Vö. Daniel C. Dennett: *The Intentional Stance*. Cambridge 1998. 15.

² Ray Kurzweil: *The Age of Spiritual Machines. When Computers Exceed Human Intelligence*. New York 1999. 15.

eddig mesterséges intelligens rendszerek vajon milyen irányokba fejleszthetők tovább és vajon az eddigi bázis hogyan bővíthető ahhoz, hogy újabb mentális képességeket, performanciákat stb. hozzunk létre. Némely kritikák, amelyek a mesterségesintelligencia-kutatásra irányulnak, inkább azt jelzik, hogy az eddigi alkotások valamilyen formában még hiányosak, de nem azért, mert nélkülözik az intelligenciát a maga teljességében, hanem mert mondjuk, rávilágítanak arra, hogy bizonyos konkrét képességet még nem sikerült létrehozni. Például egy rendszer jó a sakkban, de még nem képes jó minta-felismerésre, vagy nem képes jól fordítani egyik nyelvről a másikra. Ha ebben a formában gondolkodunk, vagyis ha eleve elfogadjuk, hogy a jelenlegi mesterséges alkotások már bizonyítékokkal szolgálnak arra, hogy igenis létrehoztunk valódi kognitív kapacitásokat, de még bizonyos egyéb, szintén elmés képességek létrehozása várat magára, akkor hajlamosak lehetünk azt gondolni, hogy csak idő és munka kérdése, míg létrejönnek azok a szupergépek, amelyek nemcsak elérik akár az emberi intelligencia szintjét, hanem meg is haladják azt. Ebben a keretben pedig máris felmerülhetnek a különböző, a jövőre irányuló aggodalmak, hogy vajon mikor válik problémává például az autonóm robotok jogi kérdése, avagy mikortól kellene aggódnunk, hogy esetleg az emberiség rabjává válik a saját alkotásainak, vagy önmagát sodorja veszélybe azzal, hogy intelligensebb rendszereket alkot.

Egy másik szemszögből tekintve azonban ez az egész diskurzus nemcsak hogy túl korai, hanem mélységes tévedéseken alapszik. A mélyebb kritikák alapján a valódi mesterséges intelligencia kialakítása el sem kezdődött, hiszen nemhogy semmilyen jelenlegi mesterséges alkotásunk nem bír valódi kognitív teljesítményekkel (intrinzikus mentális képességekkel), hanem nem is teljesen biztos, hogy egyáltalán lehetséges, kivitelezhető projektről van-e szó. E mélyebb kritikák pedig jogosan figyelmeztetnek arra, hogy habár egy rendszer viselkedése gyakran azt a látszatot keltheti, hogy az intelligencia vagy az elme által vezérelt jelenségről van szó, a jelenlegi mesterséges alkotások pusztán csak abban az értelemben nevezhetők intelligenseknek, ahogyan mondjuk a számítógép sakkozása az előbbi példában „játásznak” nevezhető. Ezek az utóbbi kritikák szimpatikusak és kijózanítóak lehetnek, de eközben persze folyamatosan szükségesek a megfelelő megalapozások: milyen alapon gondoljuk, hogy x entitás rendelkezik a kognitív élet valamilyen formájával vagy éppenséggel nélkülözi azt? Mit kellene például egy robotnak teljesítenie ahhoz, hogy meggyőzzön arról, hogy nem pusztán pszudomentalitással bír?

Ez a probléma nyilván nemcsak a mesterséges intelligenciában merül fel, hiszen egy másik doménben is különösen fontosnak bizonyul. Egyértelműnek tartjuk, hogy az emberek és meghatározott állatfajok elméseknek számítanak, de ha mondjuk, figyelembe vesszük az egész evolúció folyamatát, nem világos, hogy miként húzzuk meg a határt az élővilágban az elmét nélkülöző és az elmével bíró élőek közt (a jelenre, mind a történelemre való tekintettel). Avagy az élet evolúciós történelmében honnantól is számítható az elme kezdete? Mely élőlények számíthatnak elméseknek, és melyek nem?

Nyilván nem pusztán arra a kérdésre gondolok, amit az irodalomban úgy neveznek, mint a más elmék problémája. Egy dolog azon tűnődni, hogy vajon x entitással kapcsolatban honnan tudhatjuk, hogy elmés vagy sem, de egy sokkal alapvetőbb ontológiai probléma, hogy egyáltalán miféle kritériumoknak kell eleget tегyen valami ahhoz, hogy elmésnek számíthasson. E bevezető gondolatok tehát az elmefilozófia, illetve a kognitív tudományok filozófiájának fundamentális kérdéséhez küldenek: mi is az elme jegye, kritériuma?

Állíthatjuk, hogy az elméről alkotott fogalmaink és tudásunk kialakítása elsősorban azokból a sokrétű tapasztalatokból származik, amelyek a Földünkön az evolúciós folyamatok során

kibontakozó elmés rendszerekkel kapcsolatosak, hiszen ez az egyetlen általunk ismert domén, amely számunkra adott, és amelyet tudományosan vizsgálhatunk. (Az ember esetében nyilván nem tekinthetünk el a kulturális evolúciótól sem.) De a filozófiára és tudományokra jellemző általánosság utáni vágy magában hordozza, hogy egyben el is vonatkoztassunk ettől a bázistól. Nemcsak arra vagyunk kíváncsiak, hogy milyen létmóddal bírnak, hogyan működnek a természetben és a mi földi kereteink közt előforduló és az általunk ismert elmés rendszerek (pl. az emberi vagy állati pszichológia), hanem szeretnénk tudni azt is, hogy egyáltalán *hogyan lehetséges* az elme (pl. hogyan lehet esetleg másként is, mint ahogyan ismerjük), és *hogyan kell szükségszerűen legyen* minden más szituációban (ahhoz, hogy egyáltalán elmeként létezhesen). Hogy világosabb legyek, szeretnénk tudni, hogy mi is az esetleges és mi is a szükségszerű vagy konstitutív az elme lényegében. E kérdés tükrözi a filozófiának egy klasszikus attitűdjét, amelynek mentén a filozófia a pontos és szigorú definíciókra törekszik, amelyek kimutatják egy fogalom meghatározásának az ún. szükségszerű és elégséges feltételeit. Ilyen vonatkozásban a filozófia a lényegi belátásokra törekszik azért, hogy egy szóban forgó dolgról gondolatban lehántja az esetlegességeket. Ehhez kötődik a különböző gondolat kísérleteknek egy bizonyos osztálya is, amelyeknek a célja egyáltalán nem a pusztán üres spekuláció, hanem a lehetőségek elgondolásán keresztül a valóság lényegi összefüggéseinek feltárása. (Pl. elgondolható-e, lehetséges-e x y hiányában; ha igen, y nem képez konstitutív elemet x létében.) Továbbá a filozófiai gondolkodás sajátossága a legmagasabb szintű általánosság utáni vágyakozásban ragadható meg. Hogy a témánkkal kapcsolatos példával éljek: a „mi is az elme természete?” kérdésével kapcsolatban elsősorban egyáltalán nem az emberi elmére vagy bizonyos állatok elméjére vagyunk kíváncsiak, hanem a kérdés olyan igazságok után kutat, amelyek érvénnyel bírhatnak az elmének mint jelenségnek minden aktuális és minden lehetséges példájára.

Nyilván a filozófiának ez a klasszikus meghatározása több okból kifolyólag is problémás lehet. A szigorú definíciók formálása részben azért nehéz, mert habár a meghatározások igazságosak akarnak lenni, különböző konkrét eseteket nehezen kezelnek; könnyen szűkösekké, illetve túl engedékennyé válhatnak. A valóság dolgai nem mindig engedelmessé válnak a pontos határvonalaknak, precíz kereteknek, szigorú distinkcióknak. A fogalmaink révén határokat állapítunk meg ahhoz, hogy elkülönítsük és azonosítsuk a világ dolgait, de a határok a valóságban sokszor elmosódtak. Nem minden esetben egyértelmű, hogy valami pontosan mikor kezdődik, hol végződik, honnantól kezdve számít x -nek, avagy mennyi változást és átalakulást engedhet meg úgy, hogy közben lényegén ne essen csorba. Nem világos, hogy az esszenciális látásmód miként kezelheti az átmeneteket, a graduális fokozatokat stb. Továbbá, a klasszikus definíciókkal kapcsolatban máris eszünkbe juthat Wittgenstein játék példája. Lehetséges, hogy a témánkkal kapcsolatban az elme fogalma hasonló a játék fogalmához. Ennek fényében tehát megtörténhet, hogy egyszerűen képtelenség volna egyetlen formulában pontosan kifejezni, hogy melyek azok az egyenként szükségszerű, illetve közösen elégséges feltételek, amelyek megszabják az elme esszenciáját.³ Még ha ez így volna is, ez nem jelenti azt, hogy a lényegi belátások utáni kutatás eleve egy teljes kudarc volna. Elvégre ha a cél (az elme esszenciájának megismerése) talán lehetetlen is lenne, ez nem zárja ki, hogy a próbálkozás során fontos és lényeges összefüggésekre bukkanhassunk.

E dolgozatomban tehát arra szeretnék rákérdezni, hogy mi is az elme természete, de nyilván ezt a nagyszabású kérdést a következőképpen szeretném szűkíteni, illetve specifikálni:

³ Vö. Colin McGinn: *The Character of Mind. An Introduction to the Philosophy of Mind.* 1999. 3.

vizsgálódásomban egyszerűen arra vagyok kíváncsi, hogy miféle alapvető képességekkel kell rendelkezzen egy létező ahhoz, hogy belépésre tehesen szert a kognitív rendszerek világába. (A kogníciót ebben a dolgozatban az elmével szinonimaként használom, és egy tágabb értelemben kezelem, nem pusztán úgy, mint amely elkülönül az affekciótól vagy percepciótól). E kérdés mögött az az egyértelmű gondolat húzódik meg, hogy a kogníciónak egy igen sokszínű formája lehetséges – gazdagabb, szegényebb, egyszerűbb, komplexebb –, és az elmés létezők nagymértékben különbözhetnek egymástól azzal kapcsolatban, hogy konkrétan miféle adottságokkal rendelkeznek. Gondoljunk például azokra a sokszínű különbségekre, amelyek meghúzódnak a különböző fajok elméje között. Eközben azonban valamilyen, még tisztázatlan szempontok szerint a sokszínű variációk mögött léteznek bizonyos kritériumok, amelyeknek mentén a legalapvetőbb és a legszofisztikáltabb elmés rendszereket is a különbségeik ellenére még mindig egy osztályba soroljuk. Hogyan vélekedjünk azonban ezekről a kritériumokról? Mi tesz egy rendszert kognitív rendszerré? E kérdés tehát arra ösztönöz, hogy keressük az elme fundamentális gyökereit és alapjait, hiszen az elme fogalmának kritériumai kapcsán nyilvánvalóan a minimális feltételek megragadása az elsődleges feladat. Céлом persze nem az, hogy a kérdést e dolgozatban egyszerűen feloldjam, hanem pusztán a téma boncolgatásával, különböző aspektusainak tárgyalásával szeretnék foglalkozni. Úgy vélem azonban, hogy felmutatható egy-
 pár kulcsfontosságú fogalom –, mint például a megtestesülés, autonómia, ágencia, teleológia, normativitás, értelem-alkotás stb. – amelyek az elme fundamentumainak kutatásában a teoretikus keret alapkategóriáit képeznék. De mindezzel kapcsolatban számomra a legizgalmasabb rejtélyt a következő probléma képezi: vajon egy elmés létező konstitúciójában elkerülhetetlen elemet képez-e a biológiai szerveződés, az élet princípiuma? Más szóval: vajon lehetséges a kogníció realizálása élettelen rendszerekben is?

Az elme alapformájának keresése, a „lentől felfele” haladó megközelítés fontossága

Melyek tehát azok a jegyek, amelyek mentén a világban bizonyos entitások, rendszerek úgy érvényesülnek mint elmések, szemben azokkal, amelyek nem bírnak mentális képességekkel?

Az elméről szóló eszmefuttatásunkat kezdetül egy laza kiindulóponttal, amelynek fényében kutatásunk alapfogalmát a következőképpen rögzíthetjük: az elmés folyamatok alatt nagy vonalakban az érzéki, perceptuális, emocionális, akarati, motivációs és gondolati folyamatokra gondolok.

Ez a laza, kezdeti kiindulópont sokféle kérdést implikál, de talán folytathatjuk azzal a problémával, hogy mi az a közös ezekben a folyamatokban, amelyek révén a pszichológia tudományának tárgykörébe csoportosulnak. Az elmés folyamatoknak talán megadhatjuk két olyan általános jegyét, amelyek valamilyen formában egyesítik őket, és megszabják sajátosságukat (amelynek révén elkülönülnek a világ egyéb, nem mentális eseményeitől). Másképp kifejezve, talán ezen képességek egyik közös jegye abban lenne megragadható, hogy mindegyik valamilyen formában (nem feltétlenül egyazon módon) tartalommal bír, irányul valamire, megcéloz, prezentál vagy reprezentál valamit. Ezen állapotok másik általános karaktere a tudatosság, amely egyszerűen azt fejezi ki, hogy e folyamatok (amennyiben tudatosak) egy

szubjektív perspektíva számára mutatkoznak meg, avagy egy szubjektum átéli azokat, tanújává válik azoknak (intencionális korrelátumaikkal együtt). Az intencionalitás mint az elme jegye bizonyos vonatkozásban alapvetőbbnek tűnik, mint a tudatosság, mivel beszélhetünk tudatalatti vagy tudattalan mentális állapotokról. Egy másik vonatkozásban azonban a tudatosság is lényeges szempont, amennyiben úgy vélnénk, hogy egy létező csak akkor rendelkezhet intencionális folyamatokkal, ha egyáltalán képes tapasztalatokra (még akkor is, ha x, y időperiódusokban a tudata inaktív).

Habár a tudat és intencionalitás fontos szempont, önmagában véve nem eléggé specifikus annak megértéséhez, hogy konkrétan miféle képességek szükségesek az elme minimális formáinak realizáláshoz. Elvégre az intencionalitásnak megannyi formája lehetséges, és különböző elmés rendszerek az intencionális folyamatok meghatározott típusaival és sajátosságaival rendelkezhetnek.

Az elmés létezők lényegére irányuló kérdés tehát arra ösztönöz, hogy eltekintsünk mindenféle egyedi, fajspecifikus képességektől. Alapvető feltétel például, hogy eltekintsünk egy emberközpontú megközelítéstől, hiszen célunk, hogy ne használjunk olyan kritériumokat, amelyek mentén aztán oda lyukadunk, hogy meghatározásunk igazságtalan módon kizár az elmések osztályából viszonylag egyszerű, de attól még kognitívnek tekinthető rendszereket. Az elme lényege természetes módon pl. nem a nyelvhasználatban, absztrakt gondolkodás képességében, szociális érzelmek jelenlétében, racionalitásában, morális kompetenciákban stb. rejlik, hiszen, habár ezek mind-mind fontos mentális képességek, tudjuk, hogy nem szükséges feltételei annak, hogy egy létező egyáltalán belépésre tehesen szert az elmések osztályába. Hogy csak pár példával éljek: egy létező érzékelhet és mozoghat a világban annélkül is, hogy szofisztikált módon értelmileg is feldolgozhatná a tapasztalatait, és számtalan élménye és érzete lehet annélkül is, hogy lenne reflektív öntudata, lennének bonyolult fogalmi kapacitásai stb. Habár ezek nyilvánvaló megjegyzések, mégis gyakran megtörténik, hogy valaki egy emberközpontú gondolkodás foglya marad (pl. feltételezi, hogy a gondolkodás bármilyen formájának lehetősége azon múlik, hogy jelen van-e vagy sem egy gazdag, szintaktikailag is strukturált nyelvi rendszer). Ha tehát az elmék természetéről a lehető legtágasabb perspektíva szintjén próbálunk gondolkodni, belátható, hogy keresnünk kell az elme minimális formáját, amely – attól függetlenül, hogy relatív értelemben igen egyszerű, de mégis eleget tud tenni azoknak az elengedhetetlen és szükségszerű feltételeknek, amelyek mentén egyszerűsége még nem akadályá annak, hogy elmének számíton. De hogyan is festene az elme alapformája? (E kérdéssel kapcsolatban beszélhetünk egyrészt a természetben előforduló, az élő organizmusokhoz kötődő elmékről, de talán a mesterséges kognícióról is.)

Az elme minimális, legegyszerűbb alakzatainak keresése azért is fontos, mert felhívja a figyelmünket arra, hogy a mentális folyamatok és képességek között bonyolult hierarchiák, fontossági sorrendek, prioritási listák, avagy konstitutív viszonyok építhetők ki (ez egyaránt érvényes az individuális és az evolúciós-történeti fejlődéstörténetekre is). Tudjuk, hogy egy létező nem rendelkezhet bizonyos mentális képességekkel, ha először nincsenek adva azokat megelőző egyéb, fundamentálisabb kapacitások (egy rendszer például evidens módon nem rendelkezhet percepcióval, ha nem rendelkezik először egy alapvetőbb szenzori-motorikus képességgel). Némely képességek ugyanakkor egymáshoz konstitutív módon viszonyulnak, körkörösén feltételezik egymást, egyfajta összjátékban működhetnek. (Érvelhetünk például azon tézis mellett, hogy ez így van a percepció-mozgás folyamatával vagy a kogníció és

emóció bizonyos összhangjával). Az effajta relációknak a kifürkészése azért fontos, mert tudni szeretnénk, hogy mely mentális képességek szolgálnak alapul az egyéb, rájuk épülő képességeknek. Bizonyos folyamatok ugyanis logikai, történeti és természetes módon is egyszerűen biztosítva kell legyenek ahhoz, hogy lehetőséget engedjenek a rájuk épülő, belőlük kisarjadó további fejleményeknek. Az evolúciós folyamatok során először nem matematikával foglalkozó organizmusok alakultak ki, amelyek a későbbiekben alapvetőbb elmés képességekre is szert tettek. Ez a megjegyzés persze banálisnak hangzik, de mégis tanúi vagyunk a mesterségesintelligencia-kutatás történetének, amely mintha nem vette volna figyelembe a fontossági sorrendet. Elvégre nyilvánvalónak tűnik, hogy az ún. magasabb rendű kognitív képességek csak azt követően és annak talajából fejlődhetnek ki, hogy először adva vannak a világhoz fűződő alapvetőbb szenzori-motorikus viszonyulásmódok.⁴ Hogyan is rendelkezhetne például egy pusztá számítógép egy bonyolult tudással a világról, ha soha nem lépett konkrét érzéki-motorikus kapcsolatba azzal, ami önmagában nem elégséges, de mégis szükségszerű feltétele volna belső szimbólumrendszerének lehorgonyzásához, szemantikájának kialakulásához. Az elme-kutatása tehát először is az elme fundamentális formáira kell összpontosítson, hiszen ha először az elme vizsgálata, modellezése avagy alkotása olyan folyamatokra koncentrál, mint az absztrakt, racionális, elvont kognitív kapacitások, ez hasonló volna egy olyan házépítési projekthez, amelyben először a tetőt szeretnénk felhúzni az alapok lerakása helyett. A kultúra például egy elkerülhetetlen szempont és közeg, amelyet nem lehet figyelmen kívül hagyni, ha az emberi intelligencia a kutatás tárgya. De ha az elme alapjainak vizsgálása a cél, természetesen nem a kultúra, hanem a biológia a kiindulópont.

Neurocentrizmus, az elme komputációs elmélete és a megtestesült elme

A kérdés, hogy mi is az elme természete, eleve magával ragadhatja azt a tévedést, amely szerint az elmét dologiságként vagy szubsztanciaként kezelve elvonatkoztatunk a testiségtől, és absztrakcióink csapdáiba esünk. Az elmefilozófia történelme során egy fontos konceptuális váltás történik a descartes-i rendszertől az elme naturalizációs törekvése irányába. Descartes számára egy alapvető kategóriahiba volna a testnek mentális attribútumokat tulajdonítani, hiszen a mentális jegyek mindig a testtől független, ontológiailag independens elmének, mint különálló és alapvető metafizikai szubsztanciának a tulajdonságaiként, modifikációiként körvonalazódnak. Ez a keret teljes mértékben kizárja a mesterséges intelligencia lehetőségét: bármilyen szofisztikált gépet hoznánk is létre, az sohasem lehetne gondolkodó gép, hiszen fizikai testek képtelenek gondolkodni.

A dualizmus ezen klasszikus koncepciójával szemben a kognitív tudományokban és elmefilozófiában valamilyen típusú materializmus, naturalizmus uralkodik, amelynek fényében nagyjából mindenki egyetért abban, hogy a mentális állapotok nem valamilyen különálló lelki szubsztancia folyamatai, hanem fizikai rendszerek működésének következménye. A kérdés persze az, hogy miként magyarázzuk ezt a tényt?

⁴ Vö. Roodney A. Brooks: *Elephants don't Play Chess*. Robotics and Autonomous Systems 6. Cambridge 1990. 3.

Ha az elme bizonyos fizikai rendszerek képességeiként körvonalazódik, nyilván specifikálnunk kell ezeket a rendszereket. Tudjuk azt, hogy ismereteink szerint a világban pusztán az élő organizmusok körében fordul elő az elme, mégpedig egy sokszínű formában és variációban. Ezzel kapcsolatban azonban a tudomány legtöbbször az elmés folyamatokat úgy szemlélteti, mint amelyek tulajdonképpen az idegrendszeri-agyi aktivitásokhoz kötődnek és ezért csak bizonyos organizmusok számíthatnak kognitív rendszereknek is. Így jutunk el az elmefilozófia egyik alapvető problémájához, amely nem inkább az elme–test kérdése (a szó szoros értelemben), hanem az agy–elme problémája. A kérdések tehát tipikusan a következő formában tevődnek fel: hogyan teszi lehetővé, hogyan realizálja az agy az érzékelést, percepciót, gondolkodást vagy emocionális világot? Ez a gondolatmenet tehát egyenesen elvezet egy újabb dualizmushoz, amely nyilván nem pontosan a klasszikus karteziánus tézis, hanem az agynak és vele együtt az elmének a test többi részétől való leválasztása. Másképp kifejezve, általában a természetben előforduló elmét kizárólag az agyi folyamatokkal szokás asszociálni, eközben pedig a testnek a többi része a kogníció realizálásában pusztán perifériális szerepet kap. A neurocentrizmus tehát egyfajta újabb kortárs dualizmus.⁵

Nem feledkezhetünk meg egy másik, még mindig aktuálisan is érvényesülő dualizmusról, amely a kognitív tudományok klasszikus elmekoncepciójához kötődik. A számítógép-tudomány kialakulásával és fejlődésével párhuzamosan igen nagy lelkesedés kezdte övezni azt a gondolatot, miszerint pontosan az információfeldolgozás képezheti a hagyományos test-elme probléma megoldásának kulcsát. Másképp kifejezve, talán az alapvető rejtély, hogy vajon miként is gondolkodhat egy fizikai rendszer, megoldható a számítógép működés módjának modelljével. Talán az agy azért képes mentalításra (legalábbis azon belül gondolkodásra), mivel információkat dolgoz fel, komputációkat végez, reprezentációkat és szimbólumokat manipulál hasonlóan ahhoz, ahogyan számítógépeink is képesek „okos” programokat futtatni. E feltételezés, ha működne, filozófiai szempontból is hatalmas sikert aratna, hiszen általa képesek volnánk az elmével kapcsolatban egy lényegi rálátásra is. Ha ugyanis ez az elméleti keret valóban helytálló volna, nemcsak azt tudnánk, hogy a természetben előforduló elme hogyan működik, hanem belátást nyernénk az elme lehetséges előfordulásának természetébe is. Az elme komputációs elmélete ugyanis szorosan összefügg a funkcionálizmus filozófiájának azon ötletével, hogy az elme realizálásában nem a konkrét anyag a fontos, hanem a szerveződés és funkcionális. Ebből aztán hamar oda lyukadunk ki, hogy például a konkrét biológia a természetben előforduló elmékkel kapcsolatban pusztán egy kontingens részletkérdés.⁶ Az élet tehát az elme realizálásának kapcsán csak egy esetleges implementációs közeg, miközben az elme megvalósulhat eltérő fizikai entitásokban, amennyiben azok mind megengedik a „szoftverszint” működését. Ezt a fajta dualizmust nyilván nem nevezhetjük neurocentrizmusnak, hiszen alapvető módon pontosan az implementációfüggetlenséget hirdeti. De attól még szintén egyfajta dualizmus, hiszen az elmés folyamatokat absztrakt információfeldolgozásként szemlélteti. Az agyakra és a számítógépekre való koncentráció azonban eltereli a figyelmet a test szerepétől. Először is a természetben sohasem pusztán az agyak érzékelnek, gondolkodnak és alkalmazkodnak a környezethez, hanem egész organizmusok. Az agy szerepe, funkciója, természete csak általt

⁵ Vö. Lawrence Shapiro: *Embodied Cognition*. New York 2011. 26–27.

⁶ Vö. Michael Wheeler: *Cognition's Coming Home: the Reunion of Life and Mind*. = *Fourth European Conference on Artificial Life*. Eds. Phil Husbands–Inman Harvey. Cambridge 1997. 11.

érthető meg, ha figyelembe vesszük, hogy az agy elsősorban egy egész élő organizmusba, az élő organizmus pedig a környezetébe ágyazott. Hasonlóan, a pusztá számítógép, ameddig testetlen, addig nem alkalmas arra, hogy a kogníció működését realizálja vagy imitálja.

Ezt a problémát a megtestesült kogníció paradigmaváltása be is látta, hiszen tagadja a klaszszikus nézeteknek az elméről alkotott testetlen, lecsupaszított, absztrakt felfogását. (E tanulmányom tulajdonképpen ebbe az elméleti keretbe illeszkedik.) A megtestesült kogníció paradigmája hozzájárul, hogy kezdeti kérdésünket finomítsuk. Nem feledkezhetünk meg ugyanis arról, hogy tulajdonképpen nem létezik olyasmi, mint pusztá elme, csakis megtestesült elmék, avagy elmés testek. Mielőtt azonban belevágnánk annak a tárgyalásába, hogy a testnek mi is a szerepe az elme realizálásában, szükségünk van még egypár kiindulóponttra. Először is úgy vélem, hogy a naturalizmusnak valamilyen formája elkerülhetetlen. Más szóval ha a naturalizmus alatt pusztán annyit értünk, hogy fizikai alapok nélkül nem lehetséges elme, akkor ez a gondolat a tanulmányom egyik alapvető háttér-előfeltétele. Ez az önmagában véve egyszerű és nem túl érdekes gondolat azonban egy izgalmas reménnyel is társul: talán nemcsak az a lehetőség áll az emberiség előtt, hogy megértse az elme természetben való előfordulásának működésmódját, hanem egyúttal az is, hogy létrehozzon valódi mesterséges intelligenciát. Ha azonban kiderülne, hogy az erős mesterséges intelligencia nem egy opció, ettől függetlenül felfogható olyan kutatásnak is, amelynek filozófiai relevanciája megkérdőjelezhetetlen. Elvégre ha kiderülne, hogy nem is lehetséges az emberiség számára, hogy valódi kognitív rendszereket hozzon létre, a mérnöki foglalkozás egyben filozofikus, hiszen rávilágít esszenciális problémákra: egyáltalán mire is vágyunk, hogyan is fogjunk hozzá, pontosan miféle képességeket szeretnénk beépíteni egy rendszerbe ahhoz, hogy az elmésként érvényesüljön? Más szóval egyáltalán minnek a megvalósítása szükséges ahhoz, hogy létrehozunk konkrét elmés rendszereket? A mérnöki problémák tehát arra ösztönöznek, hogy a természetben előforduló elmés rendszereket jobban megértsük, nem pusztán úgy, mint a pszichológusok és biológusok, hanem úgy, mint akik a lényegi belátások után is vágyódnak, akik kíváncsiak az elmés létezőkkel kapcsolatban az esetleges és a konstitutív elemekre is. A mesterségesintelligencia-kutatás filozófiája azonban elkerülhetlenül reflektív és kritikus kell legyen a saját vállalkozásával szemben is, mégpedig a következő értelemben: magának a vállalkozásnak a lehetőségét is szükséges megkérdőjelezni, hiszen megtörténhet, hogy végső soron az erős mesterséges intelligencia pontosan azért lehetetlen, mert az intrinzikusan kognitív rendszerek olyan önalkotó, önfenntartó és önszabályozó autonómiával bírnak, amelyet kívülről tervezni eleve ellentmondást rejt magában.

Az elme alapjai

Érzékiség és mozgás

Az elmével kapcsolatban gyakran előtérbe helyezük a tudásszerzést, a valóság reprezentálásának képességét. A mesterségesintelligencia-kutatásra vonatkozóan ez megfelelne azoknak a törekvéseknek, amelyek mentén a tervezők minél nagyobb terjedelmű tudást próbáltak beépíteni a számítógépekbe. Egyes kritikusok azonban, mint például John Searle vagy Steven Harnad jogosan hívták fel a figyelmet arra, hogy a számítógépek szimbólumai nem kódolhatnak valóságos tudást, hiszen a szimbólumoknak, mind a szintaktikai, mind a szemantikai oldala

interpretációdependens, kívülről származik, és nem intrinzikus a számítógép működésében.⁷ Az értelem nem a számítógép szemszögéhez tartozik, hanem a tervezők fejében létezik, mivel a számítógép szimbólumrendszere nem kötődik megfelelően a világhoz, annál az egyszerű oknál fogva, hogy a pusztá számítógép nem tapasztal semmit. A robotikában azonban egyesek pontosan ennek a problémának a feloldását látták: ahhoz, hogy a tudás egy mesterséges rendszerben egyáltalán kialakulhasson, az érzékelés különböző formáival kell kezdeni, amely majd „anyagot” biztosíthat a fogalmak kialakításához. A valóság absztrakt reprezentálása másodlagos az alapvetőbb szenzoros képességekhez viszonyítva.

Úgy tűnik tehát, hogy az elme egyik minimális jegye, egyik szükségszerű alapvető komponense az érzékelés. Ennek az alapnak a hiányában nincs lehetőség feltétel ahhoz, hogy bonyolultabb dolgok is történhessenek. Egy pusztá számítógép „okos” programjai pontosan ezért maradnak pusztán pseudoértelemben intelligensek.

Ha az érzékelés egy fundamentális képesség, ezt hajlamosak lehetünk a motorikus kapacitásoktól függetlenül kezelni. Ennek alapján egy dolog, hogy egy rendszer érzékel, másik dolog, hogy esetleg mozogni is tud. Ezek szerint azonban a kettő közti kapcsolat kontingens, és habár a mozgás megfelelő koordinálása feltételezi az érzékelést, talán úgy vélhetjük, hogy ez nem működik fordítva, hiszen elgondolhatunk egy rendszert, amely érzékel, de nem mozog. Egy ilyen rendszer egyszerűen ki volna téve a hatásoknak – érzéki apparátusa pusztán elszenvedné a hatásokat, az inputokat, amelyek érik, anélkül, hogy ő aktívan tudna keresgélgni, kutatni, ide-oda fókuszálni, elfordulni ingerektől, vagy odafordulni ingerekhez. De amikor egy ilyen, mindig is passzív módban működő érzékelést próbálunk elképzelni, tudjuk, hogy eredményünk nem tükrözi az érzékelés valódi formáját, amely szükségszerűen igényel egy aktív szabályozást és egy motorikus bázist (pl. a szemek, a fej vagy az egész test mozgását).⁸

Ebben a keretben azonban látható, hogy egy elmés rendszer eleve csak úgy lehetséges, ha egy környezetbe ágyazott, szituált és megtestesült. Ha a tudás alapjai szükségszerűen az érzéki tapasztalatokban gyökereznek (legalábbis részben), a pusztá számítógépek helyett a mesterséges intelligencia kialakításának reménye a robotikában rejlik.

Ha az érzékelés az aktív keresgélésben, kutatásban, felderítő mozgásban nyerheti el valódi formáját, akkor egy elmés rendszer eleve csak úgy lehetséges, ha egy egész testnek a környezetben való dinamikus tevékenységét implikálja. A valódi tudás elsajátítását nem lehetséges pusztán beültetni, beprogramozni egy rendszerbe, hiszen természetes, hogy maga a rendszer kell érzéki és motorikus interakcióba kerüljön a világgal ahhoz, hogy azt valóban megismerhesse.

A robotika előnye, szemben a mesterségesintelligencia-kutatás egyéb irányzataival elsősorban abban rejlik, hogy belátta ezt az alapot: az elme fundamentuma, legalábbis részben, a szenzori-motorikus képességekben, környezetben történő flexibilis mozgások lebonyolításában, a sikeres eligazodásban, alkalmazkodásban, a környezethez szabott dinamikus viselkedés kialakításában rejlik. Csak ennek a bázisnak az alapjából kiindulva lehetséges az offline kogníció – az itt és most szituációktól való elvonatkoztatás (mint emlékezni a múltra, megjósolni a jövőt, tervezni vagy elképzelni azt, ami érzékszervileg nincs jelen stb.).

⁷ Steven Harnad: *The Symbol Grounding Problem*. Physica D 42. 1990. 4; John Searle: *Minds, Brains and Science*. Cambridge 2003. 37.

⁸ Alva Noe: *Action in Perception*. Cambridge 2004. 2.

Ágencia

Hamar beláthatjuk azonban, hogy ez a kulcsfontosságú bázis még önmagában igen hiányos. Elvégre ez az érzéki-motorikus koordináció és interdependencia (ahhoz, hogy mozogjak, érzékelnem kell, de ahhoz, hogy megfelelően érzékeljek, mozognom kell) feltételezi a kontrol képességét, hiszen e viselkedés aktivitást, szabályozást igényel, tehát olyan rendszert, amely már ágensként érvényesül a világban. A szenzori-motorikus kapacitás tehát a következő kérdést vonja után: mit jelent az ágencia, hogyan azonosítsuk és ismerjük fel a világban azokat a rendszereket, amelyek ágensek?

Az ágenciát nehéz pontosan meghatározni, de van egy intuitív kiindulópontunk, amelynek alapján az ágencia feltételezi, hogy egy test nem pusztán elszenvedi a fizikai hatásokat, hanem saját maga aktív forrása a tevékenységeinek. Nyilván számos fizikai entitásról beszélünk úgy, mint amely meghatározott dolgokat tesz, vagy meghatározott módon viselkedik: mint például a Nap kering a Föld körül. De e példában nyilvánvaló, hogy az aktivitás és „viselkedés” nem az ágenciából fakad, hiszen a csillag nem lehet egy ágens. Másképp kifejezve, bizonyos fizikai testek egyszerűen egymásra hatnak és elszenvednek hatásokat, annélkül azonban, hogy történne bármilyen, számunkra releváns értelemben zajló szabályozás. Tehát csak azok a fizikai rendszerek lehetnek ágensek, amelyek nem pusztán passzívok, hanem valamilyen értelemben képesek legalább egy részleges önszabályozásra. Az önszabályozás azonban csak azáltal lehet kognitív önszabályozás, ha normatív és valamilyen értelemben intenciók, célok vezérlik.⁹

Télosz

Az elme fundamentumainak keresését tehát egy újabb nélkülözhetetlen elemmel kell egészítenünk: *a teleológiával*. A robotikában a téloszra olyan explicit példák szolgálhatnak, mint bizonyos tárgyak keresése, felismerése vagy manipulálása. De tudjuk, hogy ezeknek a céloknak a követése addig a pontig nem lehet intrinzikus a robotban, ameddig nem a rendszer önérdékéből származik. Még ha úgy tűnhet is, hogy egy robot viselkedése célorientált, érezzük, hogy ez csak pseudoértelemben számíthat valódi célkövető aktivitásnak (a robot céljai nem a saját céljai, hanem külső forrásokból származnak). A hiányzó elem egyértelmű: a valódi célszerűség és intenció egy alapvetőbb szerveződés következményéből származhat. A kogníció mint célszerű tevékenység csak akkor nyerheti el valódi formáját, ha e célok valóban tükrözik a rendszer érdekeit, ha valóban saját szükségletekből, motivációkból, vágyakból, késztetésekből erednek. A természetben pontosan az élet princípiumai jelentik azokat az alapokat, amelyekből ezek az autonóm hajtóerők és saját, autonóm érdekek kibontakozhatnak.

Élet és elme

Ebben a pontban elkerülhetlenné válik az élet kérdése, hiszen rákérdezhetünk arra, hogy a biológia doménjében az elmés rendszerek motivációs bázisa honnan is származik, miből fakad. A természetes elme elsődleges motivációs folyamatai egyszerűen az élet egyik alapvető szerveződéséből, az önépítésből, identitásalkotásból fakadnak.

⁹ Xabier Barandiaran–Ezequiel Di Paolo–Marieke Rohde: *Defining Agency. Individuality, Normativity, Assymetry and Spatio-Temporality in Action*. Journal of Adaptive Behavior 2009. 5.

A metabolizmus mint az élő szervezetek sajátos jegye az autonómia egy alapvető formája.¹⁰ Az anyagcsere révén az élő rendszerek folyamatosan fenn kell tartásuk ezt az önépítő mechanizmust ahhoz, hogy egyáltalán létezhesse. Ez egy bonyolult, szofisztikált folyamat, amely nagy vonalakban a következő lépések sorozatából áll: a tápanyagok azonosítása, a szervezetbe való beépítése, hasznos energiává való átalakítása, a salakanyagok eltávolítása, majd az energia felhasználása a körforgás fenntartása céljából.¹¹

Az élő szervezet fenntartása azonban egy veszélyekkel telített, ingtag folyamat, amely csak bizonyos keretek között működhet. A megélhetőség kereteit olyan különböző faktorok határozzák meg, mint a hőmérséklet értéke, tápanyagok jelenléte, oxigénmennyiség stb. Ezeket a tényezőket megmérhetjük, és akkor kiderül, hogy hol húzódnak meg azok a határok, amelyek mentén az élet működtetése a leghatékonyabb, illetve ahol a legbizonytalanabb. A homeosztázis az organizmusoknak pontosan arra a képességére vonatkozik, amelynek révén a belső környezetüket képesek valamilyen szinten szabályozni, hogy fenntartsák azt az egyensúlyállapotot, amelyben az élet működhet.¹²

Ha adva van ez az autonóm, öngenerált és önfenntartott folyamat, amely az élő szervezet sajátos jegye, külső megfigyelőként megállapíthatjuk, hogy ezen identitás fényében miféle tényezők lehetnek negatív vagy pozitív hatással rá (mi lehet destruktív vagy építő jellegű körülmény, hatás, aktivitás). Ehhez hasonló külsődleges megállapításokat azonban élettelen rendszerek identitásával kapcsolatban is elgondolhatunk. Az élő identitását azonban pontosan az különbözteti meg más dolgok identitásától, hogy maga az élő is érzékeny (bizonyos szinten) olyan belső és külső faktorokra, amelyek identitására és működésére nézve jó vagy rossz hatással lehetnek és ezeknek fényében képes normatív regulációkra, alkalmazkodóképességekre, adaptációkra.¹³ A biológiában ezek a tevékenységek képezik a természetes kogníció gyökereit. Már egy egysejtű baktérium létmódjával kapcsolatban beszélhetünk tehát a törődés, az érdekelttség valamilyen kezdetleges formájáról, az indifferencia hiányáról, még akkor is, ha ezt a létmódot nem kísérik szubjektív tapasztalatok. Más szóval bármilyen kifejezést is használnánk, de láthatjuk, hogy az élet nem pusztán passzívan marad fent, hanem olyan, mintha első perctől kezdve „küzdene saját túléléseért és jólétéért”, mintha már az élet eleve fel lenne ruházva egy olyan perspektívával, amelynek fényében számítanak dolgok.

Ha tehát azokat az alkalmazkodó, regulatív aktivitásokat, amelyeknek a mentén az élő rendszer fenntartja és működteti saját identitását, a kogníció alapjaiként és legkezdetlegesebb formáiként körvonalazzuk, érthetővé válik, hogy nem lehetséges elme egy olyan létmódban, amelyet nem jellemez a törődés, amelyből kivonható lenne az affekció és az értékelés.¹⁴

Mindebből a képletből érthetővé válik, hogy az elme, legalábbis eredetére való tekintettel nem pusztán alárendelődik az élet szerveződésének, hanem szervesen hozzátartozik ehhez a létmóddhoz. Abban a percben, ameddig a kogníciót úgy szemléljük, hogy elvonatkoztatunk az élő organizmus szerveződésétől és létmódjától, egyszerűen attól a kerettől vonatkoztatunk el, amely irányt, relevanciát, értelmet biztosítana az elme számára. Ha a korai életformákról

¹⁰ Vö. Evan Thompson: *Mind in Life. Biology, Phenomenology and the Sciences of Life*. Cambridge 2007. 44.

¹¹ Vö. Antonio Damasio: *Self Comes to Mind. Constructing the Conscious Brain*. New York 2010. 40.

¹² Uo. 41.

¹³ Ezequiel Di Paolo: *Autopoiesis, Adaptivity, Teleology, Agency*. Phenomenology and the Cognitive Sciences 4(2005). 10.

¹⁴ Giovanna Colombetti: *The Feeling Body. Affective Science Meets the Enactive Mind*. Cambridge 2014. 2.

radikális is volna úgy beszélni, mint amelyek elméseknek számíthatnak a szó bemutatott, tágas értelmében, mindenképp felfoghatóak úgy, mint amelyek az elmés létmód gyökereit képezik, avagy az elme fundamentumait építik ki.

Ha az evolúciós időben előreugrunk, a kezdeti képletünk gazdagodik, hiszen az életnek és elmének sokszínű, már komplexebb, gazdagabb alakzatai, szerveződése bontakoznak ki. A szubjektív tapasztalat emergenciájával (amelynek eredete még kérdéses) a sokféle regulatív szabályozásnak bizonyos aspektusai már egy átélt módban is felcsillannak a tanúként is érvényesülő élő entitások tudatában. Ebben a pontban a törődés már egy újabb szintre lép, hiszen olyan formát ölt, amely valóban kínokkal, szorongásokkal, késztetésekkel, szenvedésekkel, élvezetekkel, kielégüléssel, fájdalmakkal stb. színezett. A test átélt testként is megjelenik, amely nem pusztán a sokszínű érzéki aspektusát öleli fel (kinesztézia, proprioceptivitás), hanem az élet homeosztatis és egyben motivációs működését is tükrözi (gondoljunk a hőérzetre, éhségre, szomjúságra, kimerültségre stb.). Az emocionális világ szintén elnyeri értelmét az élet létmódjának fényében: a félelmek jelzik a veszélyeket és menekésre készítetnek, az agresszió hozzájárul az akadályok leküzdéséhez és a célok eléréséhez, a kíváncsiság a környezet felderítését segíti elő stb. Ami pedig a szűkebb értelemben vett kogníciót illeti – emlékezni a múltra, felismerni a környezet dolgait, eltervezni a következő lépést, tanulni a hibákból stb. – mindez szintén elnyeri funkcióját a bizonytalan, ingatag életmód fenntartásának fényében.

Nyilvánvaló, hogy ez a levezetés célja egyáltalán nem az, hogy a pszichológiának minden formája és aspektusa végső soron redukálható lenne biológiai folyamatokra. Elvégre számos emberi motiváció vagy mentális folyamat igen eltávolodhat az alapvető életfenntartás bázisától. Ahhoz azonban, hogy eljussunk az elméhez, amely már a kultúra közegébe ágyazódik, amelyet már igen absztrakt tudásformák is érdekelnek, amelyet leköt a művészet és a mikrofizika, először az alapvetőbb folyamatok megértése a fontosabb, hiszen, mint, ahogyan említettem, a háztető az alapok nélkül a levegőben maradna.

Konklúzió: vissza a robotokhoz

Az eddigiek fényében arra a következtetésre jutottunk, hogy a kogníció realizálásához először is szükség van alapvető érzéki folyamatokra, amelyek konkrét „anyagot”, inputot képezhetnek a magasabb rendű mentális feldolgozások számára. De az érzékelés igényel egy motorikus bázist, különben egy olyan passzív folyamat volna, amelyben az érzékelő rendszer pusztán ki volna téve azoknak a hatásoknak, amelyek érik. Az érzékelés és mozgás interdependenciája azonban egy egész testnek a környezethez való dinamikus tevékenységét implikálja. Ha azonban az érzékelés és mozgás eleve egy aktív, szabályozott folyamat, feltételez valamilyen intenciókat és ágenciát. Az ágencia azonban teleológiát igényel, olyan célkövető aktivitást, amely normatív (különben véletlenszerűvé válna). A célkövető képesség továbbá csak egy még alapvetőbb motivációs bázisra épülhet, amely a rendszer önértékéből, szükségleteiből fakad. Ez azt jelenti, hogy egy kognitív rendszer csak úgy lehetséges, ha eleve számítanak neki dolgok, ha eleve törődéssel fordul a világhoz (céljai különben alaptalanokká vagy külsődlegesekké válnának). Mint láttuk, az élet organizációjából kifolyólag érthető, hogy a biológiai rendszerek esetében pontosan az autonóm identitásalkotásból és -fenntartásból vezethető le az intrinzikus

motivációs és célkövető bázis (legalábbis az elme korai formáira való tekintettel). De hogyan vélekedjünk e gondolatmenet fényében a mesterséges intelligencia lehetőségéről?

A robotika első benyomásra megoldást képezhetne a hagyományos szimbólum-lehorgonyzás problémájára, amely annak a belátása volt, hogy a pusztá információfeldolgozás szintje a számítógépekben nem kapcsolódik a világhoz, és ezért a számítógép szemantikája csakis interpretációfüggő maradhat. Magyarán: egy számítógépnek nem lehet tudása, gondolata, hite, hiszen nem tapasztal semmit. A robotok azonban eleve bizonyos triviális értelemben megtestesült rendszerek, amelyek el vannak látva érzéki és motorikus apparátussal, és képesek interakcióba lépni a világgal. Ez a tény tehát lehetőséget biztosítana arra, hogy a robot valóban megismerhesse a környezetét, hogy a világról saját maga sajátíthasson el reprezentációkat, és a tanulás révén intrinzikus értelemre tehessen szert (imitálva például azt a folyamatot, ahogyan a gyerekek is tanulnak az individuális fejlődésük révén). A megtestesülés és az érzéki-motorikus bázis azonban még mindig nem képez elég alapot ahhoz, hogy jelenlegi robotjainkról úgy vélekedjünk, mint amelyek a valódi intelligencia folyamatait realizálják. A hiányzó elem (többek között) tulajdonképpen a saját motivációs bázis, hiszen érezzük, hogy alaptalan volna azt gondolnunk, hogy robotjainkat bármi is érdekelné, vagy bármi is számítana nekik. Ilyen értelemben a jelenlegi mesterséges rendszerek nem lehetnek valódi ágensek és valódi célkövető entitások, hiszen céljaik, mivel nem saját szükségleteikből és érdekeikből fakadnak, pusztán a tervező mérnökök céljai. Ha a motivációkról értekezünk, a hagyományos szoftver-hardver distinkció megkérdőjeleződik, hiszen motivációkat programozni a gépben azt jelentené, hogy újabb információvezérlő szoftverekkel egészítenénk ki a meglévő bázist. Az élet esetében azonban a motivációk szorosan összekapcsolódnak az élő test konkrét anyagi szerveződésével, a metabolizmus, a homeosztázis dinamikus folyamataival, azzal az alapvető szerveződéssel, hogy az élő organizmus létmódjára való tekintettel feltétlenül és megállás nélkül aktívan szabályoznia kell saját életfolyamatait ahhoz, hogy egyáltalán létezhesen. Ezzel szemben a robotok passzívak, statikusak és identitásuk fenntartásáért nem kell „aggódniuk”. Egy még alapvetőbb probléma alapján azonban identitásuk eleve külsőleg származik, és nem maguktól, belsőleg generált. Matriális konfigurációik külsőleg determináltak, és önmaguk identitása nem függ attól, hogy éppen mit tesznek. Másképp kifejezve, teljesen inaktívak lehetnek, ez akkor sem befolyásolja fennmaradásukat. Ki lehetnek kapcsolva, avagy kifogyhatnak energiájukból, ez nem jelenti azt, hogy létük veszélyben forog.¹⁵ Végző soron egy paradoxon bontakozik ki magának a mesterségesintelligencia-kutatásnak a logikájában. Talán eleve ellentmondásos, ha erős mesterséges intelligenciát próbálunk létrehozni, hiszen maga a vállalkozás feltételezi, hogy kognitív identitást hozunk létre külső tervezet és szabályozás révén (amely az autonómia bizonyos formáját eleve ellehetetleníteni látszik). Még ha ez a kérdés feloldható is volna, szembe kell néznünk egy következő problémával: vajon miként is számíthatna egy rendszernek bármi is, ha élettelen, és ebből kifolyólag nem érinthetik azok a kihívások, veszélyek, bizonytalanságok, ingtag körülmények, amelyek az élők létét fenyegetik?

¹⁵ Vö. Tom Ziemke–Tom Froese: *Enactive Artificial Intelligence: Investigating the Systemic Organization of Life and Mind*. Artificial Intelligence 173(2009). 35.

The Grounds of Cognition

Keywords: embodied cognition, artificial intelligence, life, agency, autonomy, teleology

In this study I am concerned with one of the most fundamental problems of the philosophy of mind and cognitive science, namely the question regarding the mark or criteria of mind. I want to specify this broad question in the following way: in my study I would like to find out what those fundamental capacities are that can be shown as necessary for an entity in the world to count as cognitive. We have an intuition that there are criteria according to which some entities in the world, although can hugely differ in complexity, nonetheless can all be part of the class of minded beings. But how can we specify the mark of the cognitive? I think that there are some crucial concepts, like embodiment, agency, teleology, normativity, motivational basis, sense-making etc., which can be regarded as the fundamental categories in the theoretical frame of studying the grounds of cognition. In this respect I am following the paradigm of embodied cognition, more precisely the enactive approach. For me, the most exciting question in this field consists in the following problem: is life a necessary, constitutive element in being minded? Is cognition possible in systems which are not living?