

Élet, elme, mesterséges intelligencia

Bevezetés

Ha eddigi ismereteink alapján a természetben minden, elmével bíró, létező élő organizmus, felmerül a kérdés, vajon élettelen rendszerek is rendelkezhetnek elmével? Vajon az élet, mint biológiai szerveződés és sajátos létmód, a pszichológiai képességeknek csak egyik lehetséges, esetleges bázisa, avagy univerzális és szükségszerű alapja? Mi az élet szerepe az elme kialakulásában, természetében és egyáltalában vett létlehetőségében? A tematika szorosan kötődik a mesterségesintelligencia-kutatáshoz, és diskurzusomat idevágó relevanciájának fényében tárgyalom. Ha úgy véljük, hogy elvben lehetségesek a szó szoros értelmében vett gondolkodó gépek, és e rendszerek alatt élettelen technológiai alkotásokra gondolunk, akkor kényszerülve vagyunk az életjelenséget az elme pusztán egyik lehetséges, de nem szükségszerű alapjának tekinteni. Tanulmányomban ezt a nézetet szeretném megkérdőjelezni, rávilágítani egyik alapvető problémájára.

Ami a fogalomhasználatot illeti, az elme kifejezését egyrészt azon mentális (pszichológia, pszichés) képességek gyűjtőfogalmaként alkalmazom, mint az észlelés, gondolkodás, érzés stb. Ezek a képességek azáltal különíthetők el a nem mentális jelenségektől, hogy aktivitásuk során az intencionalitás (reprezentáció) és/vagy a tudatosság (szubjektív tapasztalat) jegyeit hordozzák. Az elme kifejezését azonban kontextustól függően magára az észlelésre, gondolkodásra, érzésre képes rendszerre is alkalmazni fogom, előfeltételezve, hogy az így értett elme nem a testtől elhatárolható és attól lényegében független szellemi szubsztanciára vonatkozik. Továbbá, tanulmányomban azzal az állásponttal dolgozom, miszerint az eddigi empirikus és teoretikus ismereteink fényében jó indokaink vannak úgy vélni, hogy a mentális képességek az ismert világunkban kizárólag megfelelően összetett idegrendszerrel bíró élőlényeknek tulajdoníthatók.¹ Habár úgy vélem, hogy az élet önmagában még nem elégséges az elméhez, felmerül a kérdés, hogy vajon szükségszerű-e ahhoz? Észlelhetnek-e, érezhetnek-e, gondolkodhatnak-e élettelen rendszerek?

A mesterséges intelligencia (= M. I.) kibontakozásának és exponenciális fejlődésének köszönhetően az elme valamilyen formáját hordozó, mesterségesen létrehozott rendszer lehetőségének gondolata egyre erőteljesebben mozgatja az emberi

1 | Antonio Damasio: *Feeling & Knowing: Making Minds Conscious*. Pantheon Books, New York 2021. 22.

kíváncsiságot. Bizonyos gondolkodók pusztán úgy tekintenek az M. I. vállalkozásának termékeire, a számítógépekre, mint amelyek újabb technológiai eszközökként többek között alkalmasak a gondolkodás bizonyos aspektusainak (pl. logikai szerkezetének), az elme működésének, a pszichológiai folyamatoknak egyre szofisztikáltabb modellezésére, számítógépes szimulációjára, reprezentációjára. Mások azonban az M. I.-kutatásban nemcsak az elme tanulmányozásának egy újabb módszerét látják, hanem egy lehetőséget, amely idővel képessé válhat ténylegesen gondolkodó gépek konstrukciójára is.² Az M. I. egyesek által képviselt értelmezésében a számítógép felfedezésével az elme magyarázatának kulcsára is rábukkantunk, hiszen egy mély belátást nyertünk az elme lényegéről, és ezzel együtt az állati és az emberi elme működéséről is. Ebben a megközelítésben a számítógépek feltalálásával arra a hagyományos kérdésre is sikerült egy konkrét választ találnunk, hogy mégis hogyan gondolkozhatnak anyagi rendszerek, hiszen ebben az elméleti keretben az elme szó szerint egy specifikus digitális számítógép, amely a természetben egy biológiai „wetware”-ben (a biológiai agyban) valósul meg, de elvben eltérő anyagi bázisban is működhet.³ Ezáltal felmerül a lehetősége nemcsak az elme számítógépes szimulációjának, hanem tényleges gépi létrehozásának is. Ebben a megközelítésben az elme lényege bizonyos információfeldolgozásban, a gondolkodásnak megfelelő komputációs folyamatok realizálásában rejlik. Az ún. erős M. I. nézete alapján a megfelelő programok tervezése, kialakítása és számítógépekben vagy robottestekben történő implementációja révén az ember idővel és elvben valódi elméket, tényleges értelemben vett mesterséges intelligenciát hordozó rendszereket is alkothat.⁴

Ezt az elképzelést több irányból is megkérdőjelezhetjük. Nem világos például, hogy a számítógépekben vagy a számítógépek analógiája alapján elgondolt információfeldolgozás segíthet-e bennünket az elme intencionalitásának (reprezentációs természetének) magyarázatában. (Lásd pl. John Searle kínaiszoba-argumentumát.⁵) Jelen tanulmányomban azonban az erős M. I.-hez kapcsolódó következő problémára szeretnék összpontosítani: még akkor is, ha a diskurzus kedvéért elfogadjuk, hogy az elme lényege bizonyos információfeldolgozásban rejlik, nem vehetjük adottnak, hogy ez a fajta információfeldolgozás élő és élettelen bázisban is egyaránt megvalósulhat. Ha úgy véljük, hogy elvben lehetséges olyan élettelen gépeknek a konstrukciója, amelyek valódi pszichológiai képességeket is tanúsíthatnak – és a következőkben az M. I.-nek ezen specifikus fogalmával fogok dolgozni –, megtörténhet, hogy egy ilyen koncepció vagy remény többek között azért is problémás, mivel az elme lényegileg igényli az élő *szervezetbe* való beágyazást. Egyáltalán miért gondolnánk, hogy élettelen anyagi rendszerek is tanúsíthatnak pszichológiai képességeket? Tanulmányom első részében a funkcionálisnak és azon belül az elme komputációs elméletének a fényében bemutatom röviden a szemléletet, miszerint a biológia pusztán esetleges szerepet játszik az elme létlehetőségében, majd a következő részben e nézet néhány alapvető problémájára összpontosítok.

2 | John Haugeland: *Artificial Intelligence: The Very Idea*. The MIT Press, Cambridge 1985. 2.

3 | Adam Frank, Marcelo Gleiser, Evan Thomspson: *The Blind Spot. Why Science Cannot Ignore Human Experience*. The MIT Press, Cambridge 2024. 183.

4 | John Searle: *Minds, Brains and Programs*. Behavioral and Brain Sciences 3/3. 1980. 417.

5 | *The Chinese Room Argument*. <https://plato.stanford.edu/entries/chinese-room/>. (Utolsó megtekintés: 2024. 9. 23.)

Az élet kérdése az elme komputációs elméletének fényében

Első benyomásra úgy tűnik, hogy az életet és elmét konstituáló folyamatokról viszonylag egy jól elhatárolható fogalmi apparátussal rendelkezünk. Nagy vonalakban fogalmazva, amíg az élet meghatározó jegyei a metabolizmus, homeosztázis, fejlődés, szaporodás, addig az elme meghatározó képességei az észlelés, affekció és kogníció. Ennek a viszonylag világosnak mutakozó általános distinkciónak azonban van egy témánkba vágó következménye is. Amikor ugyanis más, lehetséges elméről spekulálunk, gondolatban nem nehéz elképzelnünk egy olyan létezőt, amely nem tanúsítja az előbbi életfolyamatokat, de birtokában van bizonyos mentális képességeknek. Könnyen elképzelhetünk pl. egy élettelen gondolkodó robotot, és innen kiindulva nemcsak a fogalmi, de a metafizikai lehetősége is plauzibilisnek tűnhet. Első benyomásra nem tűnik valószínűnek, hogy az életjegyek az elme minden lehetséges előfordulásában szükségszerűek lennének. A gondolkodásra képtelen élőlény és az élettelen, gondolkodó létező intuitíven egyaránt lehetségesnek tűnik.

Az élet és elme esetleges viszonyát tápláló gondolat másik forrása nem az intuíciónál vagy gondolatkísérletekből, hanem bizonyos tudományos és filozófiai teoretizálásból fakad. A következőkben egy elterjedt szemléletmódot mutatok be, amelyhez egy bizonyos fejlődéstörténeti kontextus is fűződik. Habár az élet megjelenése körülbelül 4 milliárd évvel ezelőtt egy nagy határkövet jelentett az összetettség alakulásában, az élő rendszerek hosszú ideig kizárólag egysejtűek voltak.⁶ Az Univerzum történetében a többsejtű organizmusok felbukkanása tehát egy viszonylag kései fejlemény. Nagymértékű konszenzus övezi a nézetet, miszerint az elme emergenciája bizonyos összetettségi fokot tanúsító, idegrendszeri alapot is igényel. Habár a biológiában egy szivacs állatnak számít, idegrendszer nélkülözve problémássá válik mentális képességeket is tulajdonítani neki. Ha ezzel a feltételezéssel egyetértünk, akkor egy alapvető különbséget tehetünk az elmével rendelkező és azt nélkülöző élőlények között, hiszen nem minden organizmus rendelkezik idegrendszerrel vagy egyáltalán idegsejtekkel. Ha a történetünk fényében az organizmusok evolúcióját megpróbáljuk az Élet fájának képén keresztül szemlélni, akkor gondolatban elképzelhetjük azokat az ágazatokat, amelyek az elmét hordozó és azt nélkülöző élőket reprezentálják (ezek az ágazatok nyilván magukban foglalják mind a jelenlegi, mind a kihalt fajokat). Az uralkodó tudományos álláspont alapján tehát az elmével rendelkező élőlények kategóriája az állatvilág egy alcsoportját képezi.

Ez a történet azonban könnyen elvezethet ahhoz a további gondolathoz is, miszerint az elme, ahogyan azt ismerjük, nem az *egész* élőlénynek *mint* élőlénynek a környezethez viszonyuló bizonyos alkalmazkodóképességeire, hanem az élőlény egy komponensének, mégpedig az agyának vagy az idegrendszerének a működésére vonatkozik. Az élőlény részekre bontása révén ugyanis elsőre megalapozottnak tűnik azt állítani, hogy miközben pl. a máj, a tüdő, a szív azokért a pusztá biológiai, nem mentális, nem reprezentációs képességekért felel, mint az emésztés, lélegzés,

6 | David Christian: *Maps of Time: An Introduction to Big History*. University of California Press, Berkeley 2004. 16.

vérkeringés-fenntartás, addig a gondolkodást, érzékelést, érzést biztosító pszichológiai képességek lókusza az idegrendszerben, azon belül is inkább a központi idegrendszerben összpontosul. A gondolatmenet alapján az elme bizonyos élőlényeknek az idegrendszerében zajló specifikus információ feldolgozására vonatkozik – pl. érzéki adatok befogadására, feldolgozására, elraktározására, manipulációjára –, amely műveletek megkülönböztethetők ezeknek az élőlényeknek olyan egyéb szerveikhez kötődő, elmétlen biológiai folyamataiktól, mint az emésztés, lélegzés, vérkeringés stb.

Az elmével bíró élőlények esetében tehát alapvető különbségeket tehetünk az anyagi aspektusai, (konkrét, materiális építőelemeik) életfolyamataik (pl. metabolizmusuk) és a mentális-információfeldolgozó képességeik között, de az előbbi gondolatok alapján az agyuk (elméjük székháza) és a szervezetük egyéb komponensei között is. A számítógépes funkcionalizmus perspektívájában az elme többszörösen megvalósulhat, és elvben élettelen rendszerekben is működhet, vagyis a mentálisra jellemző információfeldolgozás nem feltétlenül igényli az élő létmódba való ágyazódást, a biológiának sem a specifikus materialitására, sem a sajátos organizációjára való tekintettel. Mindaz, ami biológiaiának tekinthető, az elmének pusztán egyik lehetséges, esetleges implementációs bázisaként körvonalazódik. Milyen megfontolások engednek ilyen következtetésekhez? Vajon, ebben a perspektívában hogyan kellene elgondolnunk azt a lehetséges élettelen rendszert, amely viszont ettől függetlenül elvben pszichológiai képességeket hordozhatna?

Az elme komputációs elmélete kapcsán gyakran találkozhatunk azzal a kijelentéssel, miszerint a megfelelően programozott számítógép elvben valódi, mentális állapotokat hordozhat. Ez az állítás azonban a hamarosan kifejtett okok miatt önmagában véve máris problematikusvá válik, ha ebben a kontextusban egy testetlen, vagyis érzéki-motorikus képességeket nélkülöző számítógépre gondolunk, amely nincs beágyazva pl. egy robottestbe és azon keresztül a környezetébe. Az óvatosság kedvéért tehát fogalmazunk inkább úgy, hogy a funkcionalizmus talán védhetőbb verziójának szellemében a potenciális mesterséges intelligens rendszerünk egyik központi eleme, az első és legfontosabb komponense az a mesterséges agya, az a processzora, amely az élőlények agyához hasonlóan képes lehet az elmére jellemző információfeldolgozásra. A funkcionalizmus alapján nem tűnik plauzibilisnek, hogy minden lehetséges agy mint az elme információfeldolgozó központja biológiai alkotóelemeket vagy akár életfolyamatok realizálásának bázisát is igényelné. Először is, ahhoz hasonlóan, hogy az óra lényege az időmérés, de az időmérésre különböző anyagokból előállított órák is képesek, a funkcionalizmus alapján valószínűnek tűnik, hogy az érzékelésnek, gondolkodásnak, érzelmeknek a folyamatai sem feltétlenül igénylik pontosan ugyanazokat a biológiai alkotóelemeket, amelyek az élőlények idegrendszerére jellemzők. Az elme komputációs elmélete alapján az agysejtek pusztán lehetséges adathordozó, közvetítő, információkat manipuláló egységek, de ezeket a szerepeket elvben más anyagi egységek is elláthatják. Hozzászoktunk, hogy valamilyen különbséget tehetünk az információ és az információt hordozó médium között, hiszen ugyanazok az adatok eltérő fizikai bázisban is tárolhatók, közvetíthetők, manipulálhatók. Miért ne lehetnének képesek a pszichológiához köthető releváns információfeldolgozásra olyan rendszerek, pl. bizonyos számítógépek, amelyek sejtek helyett mondjuk szilikoncsipekből tevődnek össze?

Az elme többszörös megvalósíthatóságának tipikus diskurzusa inkább az elme anyagi bázisának lehetséges variálhatóságára összpontosít, amely kontextusban biológia alatt általában meghatározott típusú anyagi komponensekre (pl. bizonyos szerves elemekre) és nem az életre jellemző sajátos organizációra (metabolizmusra, homeosztázisra) gondolunk. Az általunk ismert agyak azonban élő szervek is, amelyek nemcsak pszichológiaiainak tekinthető aktivitásokat folytatnak, hanem az élet esszenciális jegyeit is hordozzák. Hogyan gondolkodjunk erről a karakterükről, ha a lehetséges elmékről is spekulálni kezdünk? Habár az elme, ahogyan azt ismerjük, nemcsak mentális információfeldolgozást folytat, hanem eközben életfolyamatokba van ágyazva, a funkcionalizmus perspektívájában az utóbbi tulajdonságot tévedés lenne minden lehetséges elmének a szükségszerű aspektusának vagy bázisának titulálni. Ebben a szemléletben az elme esszenciájában nem az életet meghatározó folyamatok (metabolizmus, homeosztázis) a szükségszerűek, hanem az információk tárolásának, közvetítésének, manipulálásának a képessége a mérvadó. Mivel tudjuk, hogy a számítógépek élettelen létmódban működve is képesek információkkal operálni, elsőre plauzibilisnek tűnik azt gondolni, hogy a mentálisra jellemző komputációk sem feltétlenül igénylik az életfolyamatok bázisát. Mi több, az élő agyra hajlamosak lehetünk úgy tekinteni, mint amely materialitásából és életmivoltából kifolyólag egy rendkívül törekeny és túlságosan sebezhető, érzékeny rendszer. Miért ne lenne az elmét jellemző információfeldolgozás az idegrendszerrel ellentétben sokkal robusztusabb, erősebb, fizikailag kevésbé érzékeny hardverekbe is implementálható, olyanokba, amelyeket nem fenyegetnek az élőkre jellemző külső és belső hatások, veszélyek?

Még akkor is, ha úgy vélnénk, hogy a mentális folyamatok, ahogyan azokat ismerjük, centralizált, fejben zajló folyamatok, az élő agyak nem az egész organizmusok nélkül érzékelnek, gondolkodnak és hoznak döntéseket a világban. Felmerül tehát a kérdés, hogy az élőlények elméjének létében és működésében mi is a szerepe annak a szervezetnek és testnek, amelybe az agyuk ágyazva van, és ezzel párhuzamosan milyen értelemben lenne szükséges egy lehetséges mesterséges agynak is (pl. szilikonsípekből álló processzornak) valamilyen testre, amelyhez kapcsolódik?

Először is, ebben a szemléletben beismerhetjük, hogy az élőlények esetében a testnek mint az agyon túli szervezetnek kulcsfontosságú funkciói vannak az elme lételemében, hiszen a test teszi lehetővé a különböző érzéki információs csatornákat, a világban folytatott akciók kivitelezhetőségének a lehetőségét, és a test biztosítja az elme létehez és működéséhez szükséges energiaforrást is. Ebben a perspektívában az élőlény testének szerepe úgy jelenik meg, mint amely az elmét támogatja. Ha azon gondolkodunk, miféle testek szükségesek a mesterséges elméknek és miért, akkor arra az észszerű konklúzióra juthatunk, hogy az érzékelés és mozgás képessége nélkülözhetetlen a valódi intelligencia lételemében, és ezért az így értett M. I. egyrészt feltétlenül igényli a testet, mint egy érzéki-motorikus rendszert. Pontosabban fogalmazva, a gondolatmenet fényében a puszta számítógépek, bármennyire is szofisztikáltak legyenek, amíg nem kapcsolódnak a környezetükhöz a megfelelő érzéki-motorikus módon, vagyis nem robotokba vannak ágyazva, addig eleve elvágjuk a szálát annak, hogy ténylegesen gondolkodó rendszerek lehessenek. Miért vélekednénk így? Egyrészt az érzékelés nyilvánvalóan szükséges a világra és a saját testre vonatkozó információk elsajátításához és ezáltal a gondolkodás anyagának biztosításához, másrészt az érzékelés nemcsak egy passzív folyamatnak bizonyul, hanem vélhetőleg feltételezi

az érzékelő test aktív, felderítő, kereső, kutató tevékenységét is.⁷ Ugyanakkor a fizikai világban folytatott mindenféle akcióknak a kivitelezéséhez nyilvánvalóan szintén szükség van a testre. Plauzibilis tehát úgy vélni, hogy az elme, legyen az természetes vagy mesterséges, igényli az így körvonalazott érzéki-motorikus beágyazást, és ezért egy olyan testet, amely ezt lehetővé teszi a számára.

Továbbá, elismerhetjük, hogy az élőlények esetében az agyon túli testnek nem csak érzéki-motorikus funkciói vannak, hiszen a testnek a biológiai/metabolikus folyamatai biztosítják az elme működéséhez és fenntartásához szükséges energiát is. Ebben a pontban felmerül a biológiai szerveződés szerepe, de értelmében redukálódik az elmét fenntartó energiabiztosításra. Ebben a képletben az elmétlen biológiai szervek azért fontosak az élőlények elméjének létlehetőségében, mert a megfelelő tápanyagokat biztosítják az agynak, az elme székhelyének.

Ha azonban a testnek vagy az élő szervezetnek az említett, elmében betöltött funkcióira gondolunk – érzéki inputok szolgáltatása, az energia biztosítása, a cselekedeteknek a mozgások révén történő kivitelezése –, azonnal látható válik, hogy ezeket a funkciókat elvben élettelen rendszerek is elláthatják. Első benyomásra úgy tűnik, hogy a környezet érzékelésére és a világban folytatott aktivitásokra egy élettelen robot is képes lehet. Másodszor, habár beláthatjuk, hogy egy mesterséges intelligens rendszer működésének is szüksége van energiára, nyilvánvaló, hogy az energiabiztosításnak nem csak az élőkre jellemző metabolizmus az egyetlen lehetséges módja. Egy robot a saját akkumulátorát is feltöltheti egy rendelkezésre álló áramforrásból, tehát az energiához való hozzáférés nem csak életfolyamatokon keresztül történhet. Mivel a testnek az említett funkciói első benyomásra nem igényelnek esszenciálisan biológiai megalapozást, ezért az élet ebben a pontban sem tűnik kulcsfontosságúnak a lehetséges elmékre vonatkozó spekulációkban.

Az élet mint az elme alapja

Az eddigi szempontok alapján a testnek az elmében betöltött szerepe kimerül az elme támogatásában. Úgy is fogalmazhatunk, hogy ebben a szemléletben a test az elméért létezik: hozzájárul a világból származó adatok biztosításához, lehetővé teszi az elmének az energiához való hozzájutást, illetve az elme intencióinak a viselkedés révén történő végrehajtását. A test ezen funkcióit tehát felfoghatjuk úgy, mint amelyek az elme lehetőségében szükségszerű támaszok, tehát érvelhetünk azon gondolat mellett, miszerint a robotika elkerülhetetlen a valódi (ténylegesen gondolkodni képes) M.I. potenciális konstrukciójában. Másképp fogalmazva, e gondolatmenet alapján az elme úgy természetes, mint potenciálisan mesterséges formájában igényli az említett megtestesülést, testbe való ágyazódást.

Amikor a mesterséges intelligencia-kutatásban arra törekedünk, hogy gépek végezzenek el intelligensnek vélt műveleteket, feladatokat, legyenek azok absztrakt logikai-matematikai műveletek vagy akár konkrétabb, a fizikai környezetbe megvalósítani kívánt akciók, megjelenik a logika, amelyben a számítógépes hardver és a robottest az előzőleg specifikált programokhoz képest másodlagos, azoknak alárendelt szerepben,

7 | Alva Noë: *Action in Perception*. The MIT Press, Cambridge 2004. 1.

a programok támogatására és kivitelezésére szolgáló bázisként érvényesülnek. Egy pár konkrét példa talán jobban rávilágít a szóban forgó logikára. Ha azt szeretnénk, hogy egy gép végezzen el egy matematikai műveletet vagy egy robotautó vezessen el a város egy adott pontjába, mindkét esetben egy előre specifikált célt és műveletet gondolunk el, majd ennek megfelelően elmélkedünk a feladatok végrehajtását biztosító fizikai rendszerekről, amelyeken keresztül és amelyek által realizálhatóvá válhatnak a célok megvalósításainak konkrét lépései. Ebben az értelemben a hardver alá van rendelve annak a programnak, amely azt vezéri. Ebben a perspektívában az élőlény agyának és testének szerepe is abban az értelmében jelenik meg, mint amely az élőlény elméjét támogatja, és amely azért szükséges, hogy az élőlény kogníciójának és intenciói végrehajtásának biztosíthasson egy megfelelő materiális bázist. Ebben a kontextusban a hangsúly és a fókusz bizonyos kognitív vagy valamilyen értelemben vett intelligens operációk, műveletek realizálásának hogyanjára esik, és inkább annak a megértésére koncentrálnunk, hogy egy rendszer miként, milyen lépéseken keresztül valósíthat meg bizonyos célokat.

A mesterségesintelligencia-kutatásban tehát hajlamosak lehetünk az elme és anyagi bázisának kapcsolatáról az említett egyoldalú viszonyban gondolkodni, amelynek során a test értelme úgy jelenik meg, mint az elme támogatója és „szolgálója”. Ha úgy véljük, hogy egy mesterséges elmének a létrehozása első lépésben azoknak a megfelelő programoknak a konstrukciójáról szólna, amelyeknek a hardver és egyéb komponensek pusztán egy implementációs és a programok végrehajtásához szükséges közegét képezik, akkor az elme és anyagi bázisának viszonyáról pontosan az említett fentről lefelé irányuló egyoldalú szemléletben gondolkodunk. Semmi kétségem afelől, hogy mind az élőlények megértésében, mind a mesterségesintelligencia-kutatásban figyelembe kell vennünk a testnek mint a kogníció támaszának, „szolgálójának” szerepét és értelmét. A problémát azonban az így felfogott perspektíva egyoldalúságában látom. A következő sorokban ezt a kritikát igyekszem kibontani.

Miközben az említett kontextusban a fókusz bizonyos kognitív vagy információfeldolgozó műveletek realizálására esik (ezek megvalósulásainak hogyanjára), a diskurzusban könnyen eltekinthetünk attól az esszenciális kérdéstől, hogy mégis hogyan bontakozhatnak ki bizonyos rendszerekben ilyen vagy olyan specifikus célok, intenciók, motivációk. Ha azonban e kérdésnek megfelelő figyelmet szentelünk, láthatóvá válik, hogy a biológiában magának az élő testnek a sajátossága képezi az élőlény számára a saját kognitív céljainak (motivációinak, késztetéseinek, szándékainak) az alapvető forrását, és ilyen értelemben magának a testnek mint élő testnek a saját szükségletei határozzák meg az élőlény alapvető kognitív „feladatait”. (Gondoljunk azokra az egyértelmű példákra, mint a szomjúság, éhség, fáradtság, fájdalom stb., amelyek révén a test jelzi az elmének, hogy milyen problémák megoldására van szükség.) Az élővilágban nemcsak a test van az elméért, hanem az elme is a testért, amelyet támogatnia, védelmeznie, fenntartania kell.⁸ Ennek a szoros interdependenciának a gondolata és jelentősége tipikusan nem jelenik meg a mesterségesintelligencia-kutatásban. Mit jelent tehát az élő testnek a célok (szándékok, késztetések, motivációk) emergenciájában betöltött szerepe, és hogyan kötődik ez a kérdés a mesterséges intelligencia bizonyos szálaihoz?

8 | Antonio Damasio: *The Strange Order of Things. Life, Feeling, and the Making of Cultures*. Pantheon Books, New York 2018. 52.

Amikor egy számítógépnek vagy robotnak a tervezői egy célkitűzést fogalmaznak meg – valamilyen feladatnak mint végrehajtandó programnak a formájában –, teljesen egyértelmű, hogy ez a cél nem a gépnek a saját igényeiből fakad, amelyet a természetből kifolyólag önmaga határozott meg vagy generált, hiszen ezek a célkitűzések a tervezők elméjére utalnak vissza. Még abban az esetben is, amikor a gépek úgymond „maguktól” tanulnak vagy „maguktól jönnek rá” megoldásokra, műveleteiket nyilván nem a saját motivációik vezérlik, hanem bonyolult, külsőleg meghatározott és kezdeményezett tanulóprogramokat futtatnak, olyanokat, amelyek a tervezőik és felhasználóik érdekeit tükrözik. A mesterséges intelligencia alkotása során az a célunk, hogy gépek végezzenek el olyan műveleteket, operációkat, amelyeket bizonyos értelemben racionálisnak, intelligensnek vagy okosnak gondolunk, de ebben a projektben maguk a tervezők szabják meg a gépek alkotóelemeit, működésmódjait, feladataikat és azok megvalósításának lépéseit. Amikor arról beszélünk, hogy a gépek most már maguktól is képesek számtalan, bizonyos értelemben intelligensnek tekinthető műveletre, és már vannak is körülöttünk ilyen értelemben vett autonóm rendszerek, nyilvánvaló, hogy ebben a kontextusban önmeghatározás alatt nem azt értjük, hogy a gépek most már saját célokat is képesek generálni. A jelenlegi gépi autonómia pusztán annyit jelent, hogy a gépek most már rendkívül összetett műveletekre is képesek viszonylag minimális, online külső beavatkozással. Amikor valaki egy autonómnak tekinthető gépkocsiban üldögel, miközben az autó eddig nem tesztelt utakon folyamatosan döntéseket hoz, hogy milyen irányba forduljon, milyen gyorsan haladjon, mikor álljon le, stb., a gép tulajdonképpen csak azokat az algoritmusokat követi, amelyeket döntésmechanizmusai kívülről határoztak meg a számára. Nyilvánvaló, hogy ez a megállapítás csak a mesterséges intelligencia azon területén képezhet egy komoly teoretikus problémát, ahol felmerül a tényleges mesterséges elmék létrehozásának a diskurzusa, hiszen jó okunk van feltételezni, hogy minden lehetséges elmével bíró létező nemcsak egy teleologikus (célszerűen működő) rendszer, hanem egy olyan létező, amely szükségszerűen *a saját intencióival* rendelkezve képes valamilyen szintű önszabályozásra, és nem lehet kizárólag egy tőle külső, externális ágens céljainak alárendelve, hiszen ez utóbbi esetben eszközzé redukálódik. (Mint ahogyan az előbbi példánkban az autonóm gépkocsival ez történik.) Az autonóm gépkocsi ugyanis csak abban az értelemben rendelkezik egy bizonyos fokú autonómiával, hogy önmagától is képes bizonyos műveletekre, amely műveletek azonban egyáltalán nem a saját céljaiból erednek. Az utóbbi esetben ugyanis a célok külsődlegesek, és nem magában a rendszerben bontakoznak ki. E gondolattal kapcsolatban fontosnak tartom hangsúlyozni a „külső cél”-nak és a „saját cél”-nak a distinkcióját.⁹

Először is, a mindenféle céloktól (intencióktól, szándékoktól, késztetésektől) mentes elme fogalma azért problematikus, mert egy ilyen elképzelt elme képtelen volna bármiféle mentális és testi akciónak a kezdeményezésére, és érzéki-kognitív képességeinek így nem maradna semmiféle szervezőereje, értelme, funkciója és iránya. Továbbá, nézetem alapján a teleológia (célszerűség) affektív mozgatórugókat igényel, hiszen ha egy létező nem rendelkezne semmiféle érzellemmel, érzéssel, motivációval, akkor egy teljesen indifferens állapotban lehetetlenné válna, hogy bármiféle mentális

9 | Willem F. G. Haselager: *Robotics, Philosophy and the Problem of Autonomy*. Pragmatics & Cognition 13/3. 2005. 519.

jellegű vagy abból fakadó testi aktivitást kezdeményezzen.¹⁰ Egyáltalán miért és miről gondolkodna, egyáltalán miért és hogyan kezdeményezhetne bármiféle fizikai cselekedetet vagy mentális aktivitást egy olyan létező, amely teljesen indifferens módon viszonyul a léthez? Még ha értelmes is volna elképzelni, hogy ebben a szituációban érzéki és kognitív képességei továbbra is működhetnének, az affektív mozgatórugók nélkül ezek a képességek relevancia nélkül maradnának. Az intelligencia szokásos definíciója problémák megoldására való képességet jelent, de hogyan lehetne egy, a léthez teljesen indifferensen viszonyuló létezőnek bármiféle problémája, ha alaphoz semminek sincs számára semmiféle jelentősége, fontossága vagy súlya?

Másodszor, fontos tudatosítanunk, hogy a teleológia feltételezi a *saját* célokkal való rendelkezés képességét is, különben a saját szándékok birtoklása nélkül csak egy kívülről kontrollált eszköz képével maradunk, amely legfeljebb csak származtatott vagy metaforikus, de nem tényleges és intrinzikus értelemben maradhatna célszerű. Ilyen értelemben a teleológia feltételezi az autonómiának mint a saját célokkal való önrendelkezésnek és a belőle fakadó önszabályozásnak az értelmét. Addig a pontig tehát, ameddig egy számítógép vagy egy robot képtelen saját intencióinak a fényében vezérelni a viselkedését, és ezért pusztán követi azokat az utasításokat, amelyek a külső mérnökök fejében megfogalmazott ötletekre és érdekekre utalnak vissza – mint ahogyan feltételezhetően minden eddigi alkotásunk működik –, addig mindvégig kívülről szabályozott elméleten eszköz marad, bármennyire is szofisztikáltak a különböző képességei. Ezeknek a képességeknek és aktivitásoknak a mozgatórugói, a kezdeményező célok ugyanis nem magukban a rendszerekben bontakoznak ki (nem belőlük fakadnak), hanem külső forrásokra vezethetők vissza (a mérnökök és programozók céljaira).

De akkor mit jelenthetne egy gépnek egyáltalán saját motivációkat biztosítani, olyanokat, amelyek magának a gépnek számíthatnának? A megfogalmazás a következő okok miatt már alaphoz paradoxálisnak tűnhet. A dilemma a következő: ha az alkotók próbálják meghatározni azokat a célokat, indítékokat vagy motivációkat, amelyeket a gépekbe szeretnének „ültetni”, mégis hogyan lehetnének ezek a kívülről megszabott célok végül mégiscsak a gépek saját céljai? Ha egy mérnök határozza meg, hogy miféle célt kövessen a gép, akkor hogyan beszélhetnénk a gépnek a saját, nem a tervezőjétől származó céljáról? Ha viszont ezt a problémát elkerülve nem a mérnök határozza meg a gép céljait, hanem valahogyan a gépnek kellene saját öncélokat generálnia, eleve megkérdőjeleződik a gépnek a kívülről történő meghatározása és ezáltal az alkotásának az értelme és lehetősége. Ez a probléma visszaküld a kérdéshez, hogy mégis honnan származnak az élők esetében azok a motivációk, célok, indítékok stb., amelyeket magukban fedeznek fel, amelyeket magukénak ismernek fel, és amelyek mentén viselkedésüket szabályozzák?

A kérdéssel kapcsolatban kiindulhatunk a következő észrevételből. Amikor egy mesterséges intelligens rendszer lehetőségén spekulálunk, tipikusan elmerülünk abban a kérdésben, hogy milyen anyagi bázisra lenne szüksége egy elmének, ahhoz, hogy kognitív funkcióit realizálhassa, és könnyen megfeleldkezhetünk a probléma fordított oldaláról, hiszen általában nem kérdezzük rá arra, hogy egyáltalán milyen

10 | Giovanna Colombetti: *The Feeling Body. Affective Science Meets the Enactive Mind*. The MIT Press, Cambridge 2014. 1.

tulajdonságokkal rendelkező rendszernek lenne szüksége elmére. Miközben egy elmével bíró élőlény esetében a pszichológiai képességei, az elméjének a jelenléte és megfelelő működése kritikusak a túlélése szempontjából, egy robotnak vagy számítógépnek a fizikai integritása és léte nem függ ilyen szorosan a programjainak a jelenlététől és megfelelő aktivitásától. A számítógépről minden applikációt törölhetünk, amely művelet önmagában véve még nem veszélyezteti a számítógép egyáltalában vett létét, a hardverének az épségét. Ha azonban az élőlények evolúciójára gondolunk, és a természetet metaforikusan az elmével bíró élőlények tervezőjének tekintjük, azonnal belátjuk, hogy a természet nem azért hozott létre élő testeket, hogy mentális információfeldolgozásnak biztosíthasson megfelelő platformot vagy bázist, hanem pontosan fordítva, azért „találta ki” a pszichológiai képességeket, hogy bizonyos életformák lehetőségének „biztosíthasson” teret. A természetben elsősorban nem a test van az elméért, hanem az elme van a testért, pontosabban az élő testért, amelyet a túlélése érdekében „támogatnia kell”. Evolúciós értelemben az életfenntartás és nem a kogníció az elsődleges cél, hiszen a kogníció képessége az élet támaszára lett „kitalálva”. A természetben az elme „szoftvere”, vagyis az élőlény elméje (pszichológiai fakultásai) elsősorban az élőlény testéért, szervezetének fenntartásáért és támogatásáért létezik, hiszen magának a testnek a saját létmódjából, saját igényeiből és szükségleteiből kifolyólag válik kulcsfontossá az elme szabályozóerejének „extra” jelenléte, amelynek segítségével az élőlény eligazodhat a környezetében, megtalálhatja a táplálékát, elkerülheti a veszélyforrásokat, és ezáltal egyáltalán túlélhet. Az érzékelésre, gondolkodásra, észsre elsősorban és alapvetően, evolúciós perspektívában azért van szükség, hogy az élet, amely vélhetőleg egy ideig elme nélkül is létezett, a komplexitásából kifolyólag továbbra is „fennmaradhasson”. Habár bizonyos életformák feltételezhetően elme nélkül is létezhetnek, mások létmódja igényli az elme sajátos szabályozóerejét.¹¹ A fának nem kell helyet változtatnia, hiszen mozdulatlanul is hozzáfér a fényhez vagy a vízhez, és ezért nincs szüksége szofisztikáltabb észlelésre és akciókra ahhoz, hogy túlélése lehetővé váljon. (Persze, bizonyos értelemben minden élőlény aktív, de feltételezhetően nem minden adaptív viselkedési formának a magyarázata igényel pszichológiai állapotokra történő hivatkozást.) Ezzel szemben, bizonyos állatoknak olyan mozgásokat kell rendszeresen kivitelezniük a környezetükben pl. a táplálékszerzés, menekvés, biztonságkeresés céljából, amelyben elkerülhetlenné válik a környezetük megfelelő észlelése és azok a vágyak, emóciók, motivációk, amelyek segítségével eltávolodhatnak a veszélyektől, megbirkózhatnak az útjukba kerülő akadályokkal, és egyéb életfenntartó akciókat hajthatnak végre. A gépek konstrukciója kapcsán azonban tipikusan olyan rendszerekről gondolkodunk, amelyek létük és integritásuk szempontjából nem igénylik azokat a programokat, amelyeket támogatnak, és amelyekre éppen ezért nem maguknak, hanem externális ágenseknek van szükségük, akik e programok révén kontrollálhatják a gépeket, mégpedig a saját maguk érdekeinek a fényében. Az így felfogott gépi létmódban nem jelenik meg a testnek az elmét igénylő karaktere, illetve a testnek a saját célok, intenciók, motivációk emergenciájában betöltött szerepe. Ezáltal háttérbe szorul az elmének a testet támogató, szolgáló, védelmező, integritását biztosító szerepe, amely viszont az élőlények esetében az elme elsődleges és alapvető funkciója.

11 | Alvaro Moreno, Jon Umerez, Jesus Ibañez: Cognition and Life: The Autonomy of Cognition. *Brain and Cognition*. 1997/34. 115.

Az evolúció ugyanis a pszichológiai képességeket elsősorban az élő test túlélésének, fenntartásának „érdekében találta ki”. Az élet esetében az elme, legalábbis alapvető formájában az életért van, és bizonyos életformáknak van szükségük az elmére, ahhoz, hogy túlélhessenek. A számítógép esetében azonban nem magának a hardvernek az önfenntartó érdekei miatt válik fontossá számára a programok léte, hiszen a gépnek mit sem számít, hogy van-e bármiféle működőképes applikációja. Ezzel szemben az elme valamilyen formájával megáldott állat léte abszolút függ az elméjének a jelenlététől és megfelelő működésétől, hiszen a mentális szabályozás hiányában képtelenné válik a környezetben való alkalmazkodásra és túlélésre. Ha egy állat nem talál rá a táplálékra, mert elveszíti az érzékelését, élete máris kockán forog. Ezzel szemben, ha egy robot képtelen feltölteni magát energiával, mert szenzorainak meghibásodása miatt nem találja meg a szükséges töltőállomást, működésének szünetelése vagy leállása nem fenyegeti automatikusan a saját létét. A mesterséges intelligencia filozófiájában fontos tudatosítani a következő, az élet és a gépi létmód közti alapvető különbséget. Habár kijelenthetjük, hogy a gépeknek és az élőlényeknek is egyaránt szükségük van energiára, ahhoz, hogy aktívak lehessenek és működhessenek, a gépek léte, ellentétben az élőkkal, nem igényli a saját aktivitásukat. A számítógépet kikapcsolhatom, és ez nem veszélyezteti a létét, de egy élőlény inaktivitása a saját halálához vezet. Az élőlényeknek tehát, szemben a gépekkel, nemcsak azért van szükségük energiára, hogy aktívak lehessenek, hiszen aktívak kell, hogy legyenek, ahhoz, hogy egyáltalán létezzenek.¹² Létmódjuk ugyanis folyamatosan igényli az adaptív önrekonstrukciót, az önszabályozó metabolizmust. Ez a létmód érthetővé teszi, hogy miként alakulhat ki a világban egy olyan perspektíva, amelyben a dolgok számítanak, mégpedig magának a rendszernek.

Az élet tehát az általunk ismert, elmével rendelkező rendszerek esetében nemcsak az elmét támogató energetikai biztosítás szempontjából lényeges, hiszen az élő szervezet egésze és sajátossága fekteti le azt a kezdeti motivációs alapot és bázist, amely az élőlény elméjének egyáltalán funkciókat, értelmet, irányokat szab. Habár valóban igaz, hogy az organizmus elméje (az információfeldolgozás szintje) igényli a testet a már említett értelemben (test nélkül nem lenne a mentális aktivitásoknak bázisa vagy szubjektuma), a képlet fordítva is érvényes, hiszen az idegrendszer aktivitása és ezen keresztül az elme működése is az élő test igényeit „szolgálja”. Az agy pontosan és elsősorban az élet „fenntartásáért” szabályozza a szervezet belső aktivitásait, illetve az egész testnek a környezetben való eligazodását és alkalmazkodását. Nem pusztán önmagától szabja meg, generálja azokat a drive-okat, motivációkat, célokat, amelyek révén egyáltalán aktivitásokat folytat, hiszen az élő szervezet szükségletei szabják meg számára azokat az alapvető „feladatait”, amelyek érzéki/kognitív/affektív aktivitásainak egyáltalán értelmet és irányt biztosíthatnak. Az elme elsődleges és alapvető értelme ugyanis pontosan bizonyos életfolyamatok kiterjesztésében, bizonyos életfolyamatok létlehetőségében, támogatásában, szabályozásában rejlik.¹³ Az idegrendszeri és az idegrendszeren túli biológiai folyamatok tehát a kifejtett értelemben kodependens

12 | Tom Ziemke–Tom Froese: Enactive Artificial Intelligence: Investigating the Systemic Organization of Life and Mind. *Artificial Intelligence* 173/3–4. 2009. 473.

13 | Markus Gabriel: Could a Robot Be Conscious? Some Lessons from Philosophy. In: *Robotics, AI, and Humanity*. Springer, Cham Switzerland 2021. 64.

módon, kölcsönhatásban szerveződve igénylik egymást. Ilyen vonatkozásban nemcsak az élet van az elméért, amelyet energiaforrásként támogat, hanem az elme is az életért, amelyet támogatnia és fenntartania kell. Az evolúciós pszichológusok folyamatosan emlékeztetnek bennünket arra, hogy a legszofisztikáltabb észleleti, kognitív és affektív képességeink is valamilyen módon, eredetüket, funkciójukat és értelmüket illetően, a túlélés és alkalmazkodás gondjaival való megbirkózás köré épülnek. Leegyszerűsítve: habár pl. a szeretetnek vagy kötődésnek mindenképp léteznek olyan tapasztalati dimenziói, amelyek megértésében elkerülhetetlen pl. a kulturális aspektus, hiszen a tanulás, a szokások, a narratívák stb. befolyásolják azt, ahogyan a szeretetet átéljük, a szeretet eredetének megértésében nem feledkezhetünk meg a darwini alapokról és logikáról sem: szükségünk van egymásra, ahhoz, hogy túléljünk, és ezért szükségese az az affektív vonzerők, amelyek a túlélést biztosító kooperációnak helyet teremthetnek.¹⁴

Felmerül tehát a kérdés, hogy egyáltalán van-e értelme az élettelen létmódban szerveződő gépek esetében olyan lehetséges igényeikről, szükségleteikről, problémáikról beszélnünk, amelyek az ő esetükben nem kötődhetnek a túlélés gondjaihoz, lévén, hogy élettelen rendszerek? Ha az agynak a vele rendelkező élőlény számára életfenntartó, életbiztosító szerepe van, ugyanakkor motivációs bázisának forrása maga az élő szervezet egészéből és sajátosságaiból fakad, az élő test elmében betöltött szerepének a témáját nem mellőzhetjük a mesterségesintelligencia-kutatásban. Ha ugyanis az élőlény motivációs forrásai elsősorban az élő szervezetnek abból a létmódjából fakadnak, amelyhez konstitutív módon tartozik az önalkotás (a metabolizmus révén zajló folyamatos önrekonstrukció) és az ehhez fűződő öngondoskodás szükségessége és képessége, de ezt az alapot esetlegesen véljük a ténylegesen gondolkodni képes mesterséges elmék lehetőségében, jogosan merül fel a kérdés, hogy milyen módon lehetnének motiváltak bármiben is az élettelen rendszerek.

Life, Mind and Artificial Intelligence

Keywords: life, mind, teleology, autonomy, artificial intelligence

At the heart of my work lies the theoretical problem of whether life as a unique organization and mode of being (metabolism, homeostasis) is only a contingent or rather a necessary ground for the emergence of the mind. I will briefly show that life as a special mode of being plays a crucial role in grounding psychological teleology (purposefulness) and autonomy (self-regulation) and then express the doubt according to which it is not clear how this role can be provided by inanimate machines. In my study I will draw attention to a fundamental problem with the notion of an inanimate machine capable of thinking.

14 | Anna Machin: *Why We love: The New Science Behind Our Closest Relationships*. Pegasus Books, New York 2022. 9.