Pokol Béla

A mesterséges intelligencia: egy új létréteg kialakulása?

A mesterséges intelligencia növekvő ereje az utóbbi években már megijeszti az ennek felhasználását irányító praktikusokat (Elon Musk) éppúgy, mint az ezzel közvetlenül érintkező teoretikusokat (Steven Hawking, Nick Bostrom), és mint az ember fölé növekedő és annak ellenőrzésétől elszakadó, új fejleményt írják le. Ennek egy másik leírása – mely Neumann János egyik, halála előtti években tett megjegyzéséből nőtt ki – a szingularitás korának kialakulásaként tematizálja az ebben rejlő újdonságot.[1] Ez utóbbi szerint az egyre nagyobb számítógépes kapacitások és az egyre gyorsabb programfuttatások az öntanuló mesterséges intelligenciát egy olyan pontra juttatják, ahonnan a gyorsulást addig lassító emberi közreműködés kiküszöbölődik. Ettől a – világtörténelemben egyedülálló (szinguláris) – pillanattól kezdve aztán a milliószoros gyorsaságra is képes öntanuló és önfejlesztő algoritmusok, ténylegesen is milliószoros gyorsaságba kapcsolva, néhány órán belül már végleg érthetetlenné és ellenőrizhetetlenné válnak az ezzel foglalkozó informatikusok számára is. Az ezzel együtt fejlődő robotika bárminek a legyártását is lehetővé teszi az ember fölé nőtt mesterséges intelligencia számára. Ettől fogva az ember nemcsak hogy nem érti majd meg a mesterséges intelligenciát, de a világban végbemenő változások irányításából is kikapcsolódik, átadva addigi helyét ez utóbbinak – ez a szingularitás beállta utáni korszak.

Jelezni kell persze, hogy egyrészt a mesterséges intelligencia jövőbeli veszélyeinek ezt a felfokozott képét csak a kutatók kisebbsége látja reálisnak, másészt az emberi irányítástól elszakadó és önálló akarattal és szándékokkal rendelkező gépi értelem létrejöttének esélyét sokan a népszerűséget kereső fantasztikus irodalom fantáziálásának tartják. Így például a John Brockman által szerkesztett 2015-ös kötet, a mesterséges intelligencia neves kutatói körében tartott körinterjúra beérkezett válaszokat összegezve, csak kisebb részben mutat be olyan szerzőket, akik – Kurzweilhez és Bostromhoz hasonlóan – az önálló akarattal és az emberi értelmet meghaladó értelemmel rendelkező mesterséges értelem létrejöttét reálisnak tartják, illetve akik e emellett ezt veszélyforrásnak tekintik az emberiségre nézve (Brockman 2015). Ennek ellenére megítélésem szerint mint egy mégiscsak lehetséges jövő hipotézisét – a kutató közösség kisebbségi álláspontjának státusza tudatában – teoretikusan érdemes szemügyre venni.

Így ha egy időre zárójelbe tesszük a mesterséges intelligencia teoretikusai által használt fogalmakat és félelmeket, és ehelyett a világ eddigi evolúciójának ugrásait elemző filozófiai fogalmakat állítjuk a középpontba, akkor az utóbbi száz év már empirikus alapokon is nyugvó ontológiai elemzései megalapozottabb elemzéseket tesznek lehetővé a mesterséges intelligencia uralomra jutása vonatkozásában is. Számomra az elmúlt évtizedekben a társadalom átfogó evolúciós ugrásait kutatva Nicolai Hartmann létszférákról szóló elmélete adta a legtöbbet ahhoz, hogy a társadalmi lét és az ennek alapjaként létező fizikai és

[1] Hogy Neumann a szingularitás első megfogalmazója, azt csak közvetetten tudjuk Stam Ulamtól. Ő 1957-ben, egy évvel Neumann halála után, idézte fel egy írásában a vele való beszélgetését a korai 1950-es évekből, Neumann ekkor beszélt neki röviden a szingularitás eljöveteléről (Lásd Kurwzweil 2012: 185).

biológiai létréteg együttélését meg tudjam érteni. Ha Niklas Luhmann társadalmi rendszerekre és alrendszerekre vonatkozó gazdag elméletét Hartmann szélesebb összefüggéseket is tematizáló elméletébe beágyazva fogadjuk el (lásd erre Pokol (2004) összegzését), akkor a mesterséges intelligencia izgalmas fejleményei felé fordulva adódik számomra a hartmanni létszférák elméletének kerete, és annak hipotetikus megfogalmazása, hogy esetleg itt egy új létréteg keletkezéséről van szó. Ahogy a Földön valaha a fizikai létréteg felé emelkedett a biológiai létréteg, majd annak evolúciójával a növényi, továbbá az állati lét felső fokain fokozatosan, az emlősökkel indulóan az érzelmi-tudati létréteg, és ennek fokozatos kibomlásával a főemlősök szintjén már az értelmi létréteg csirái is, melyek az emberi közösségekben már egyre dominálóbb értelmi létréteggel borították be az alsóbb létrétegeket. Most pedig itt állunk az újabb evolúciós ugrásnál, és az emberhez kötött értelmi létréteg (illetve az ennek folyamatosan alapul szolgáló érzelmi-tudati, illetve biológiai és fizika létréteg) fölé egy új létréteg látszik kibontakozni, mely az addig emberhez (és az alatta levő többi létréteghez) kötött értelmi létréteg utódaként *önszerveződő szellemi létrétegként* kezd a világot meghatározó erővé válni. E rövid írás e gondolatfelvetés első körüljárása kíván lenni.

Mivel Nicolai Hartmann részletesen elemezte már az eddigi evolúciós ugrásoknál az új létrétegek viszonyát az alattuk fekvő, evolúciós szempontból régebbi létrétegekhez, és közös törvényszerűségeket állapított meg minden eddigi evolúciós ugrás után keletkező új létréteg, illetve elődjei között, röviden érdemes néhány megállapítását megidézni, mielőtt elkezdenénk azt elemezni, hogy ezekből milyen tanulság vonható le a mostani új létréteg, a perspektivikusan önállóvá váló és önszerveződődő mesterséges intelligencia létrétegének vonatkozásában.

Az ember és a létrétegek hierarchiája

Az emberben a sajátságosan „emberi" az értelmi létréteg fokozatos dominálóvá válása a fizikuma, biológikuma és lelki-érzelmi ösztönélete felett, de élete minden egyes pillanatában mind a négy létréteg törvényszerűségei együtt hatnak rá. Az ember többrétegű lény, és az emberi közösségek e létrétegek törvényszerűségeinek összegződő keretén belül bontakozhatnak ki. A felső létrétegek csak az alsóbbak törvényszerűségeinek tiszteletben tartása mellett fejlődhetnek ki, de ez nem akadálya annak, hogy a felsőbb létrétegek törvényszerűségei önállóak legyenek az alsókéhoz képest. A felsőbb létréteg felépülése először az alsóbb létszféra törvényszerűségeinek átformálva-megtartása mellett történt, de a két legfelsőbb esetében már nem átformálás, hanem az alsóbbakra való ráépülés történik. Míg ugyanis a fizikai világ anyagi elemeit a biológiai-organikus létréteg is elemként használja fel – csak a saját élővilági törvényszerűségei által átformálva –, addig a lelki létréteg és az arra ráépülő értelmi létréteg esetében már nincs anyagi elem. Itt már nem átformálva tartja meg az alsóbb létréteg elemeit, hanem egyszerűen csak ráépül azokra a felsőbb.[2] Hartmann megfogalmazásában álljon itt egy hosszabb idézet mind a négy létréteg vonatkozásában: „*Hogy a többrétegűséget meg tudjuk ragadni, elegendő, ha az általánosan ismertekhez*

[2] Most még tegyük zárójelbe, hogy az utóbbi évtizedek agykutatásai időközben feltárták – Donald O. Hebb 1949-es kezdeményezései alapján –, hogy az agyi idegsejtek százmillióiban egy-egy új ismeret és tapasztalat speciális elrendezésbe állítja az idegsejtek egy-egy csoportját, és így szemlélve mégiscsak van anyagi bázisa az értelmi változásoknak az agyban. Ennek elemzéséhez lásd Kurzweil 2012-es könyvének Neokortex című fejezetét (Kurzweil 2012: 85–95)

tartjuk magunkat. Senki nem vonja kétségbe, hogy az organikus-biológiai élet a fizika-anyagitól a lényeget érintően különbözik. Ám nem létezhet függetlenül ettől: magában tartalmazza, rajta nyugszik, sőt a fizika törvényei mélyen belenyúlnak az organikus testbe. Ami persze nem akadályozza azt, hogy ennek ellenére saját törvényszerűségeket fejlesszen ki, melyek nem vezethetők vissza már a fizikai törvényszerűségekre. Ezek a törvényszerűségek átformálják az alsóbbakat, az általános fizikai törvényeket. Ugyanez a viszony létezik a lelki élet és a biológiai-organikus élet kapcsolataiban is. A lelki élet - mint azt a tudati jelenségek bizonyítják - eltérő az organikus-biológiai lét törvényszerűségeitől, és egy önálló létréteget képez felette. Ám mindenhol az jellemző rá, ahol csak találkozunk a realitásban vele, hogy függő viszonyban van tőle, és a biológiai organizmus által hordozott létréteg lehet csak. (…) A lelki élet így csak hordozott jellegű lehet, de minden függősége mellett is saját törvényszerűségeket épített ki. Végül a pszichologizmus háttérbe szorulása óta elismert tény, hogy a szellemi lét birodalma nem egyszerűen a lelki lét folyatása, és e réteg törvényszerűségeit nem lehet megmagyarázni pusztán a tudati élet törvényszerűségeiből. Sem a logikai törvényeket, sem a tudás és a megismerés sajátosságait nem tudjuk megmagyarázni pusztán a lelki élet törvényszerűségeiből. Még kevésbé az akarat és a cselekvés szféráit, az értékeléseket, a jogot, a vallást, a művészetet. Ezek a szférák mind kiemelkednek messze a lelki élet folyamatai fölé, és mint szellemi élet egy önálló és magasabb létréteget képeznek fölötte, melynek gazdagsága és sokoldalúsága az alsóbbtól messze eltávolodik. De itt is ugyanaz a viszony jellemző az alsóbb felé kapcsolódásban. A szellem nem lebeg a levegőben, hanem csak lelki élet által hordozott lehet, ugyanúgy, mint ez pedig az organikus lét által hordozott, ez utóbbi pedig a fizikai-materiális által" (Hartmann 1962: 71).

Az ember tehát négy létréteg egysége, és az emberi értelem az alsóbb létrétegeken alapulva – az ember biológiai testének bázisán – fejti ki hatását. Közelebbről a szellemi létréteg a tiszta értelmi tevékenység terepe, és ezt Hartmann három belső, szellemi létforma együtteseként írja le: az egyéni, az objektív és az objektivált szellem létformáiként. Az első kettő az élő szellem létformáit jelenti, míg az objektivált a holt szellemét, ám az utóbbira visszanyúlva ennek részei állandóan átfordulhatnak az élő szellem alakjába. Az egyéni szellem együtt él korának objektív szellemének tartalmaival, és többé-kevésbé az egyéni szellemek sokasága hordozza az objektív szellem létformáit mint egy-egy kor népszellemét és a korszak többi kollektív szellemi képződményeket. De egyben az egyéni szellem is jórészt épp olyan tartalmakkal rendelkezik, melyeket a korának objektív szellemi formái tartalmaznak, így inkább kölcsönös egymást-hordozással írható le a viszonyuk. A harmadik forma, az objektivált szellemi létforma tartalmainak bővülésével pedig – az értelmi rögzítettség formáinak sokasodásával az írás stb. révén – az egyéni szellemek a korszak objektív szellemi tartalmai mellett mindig visszanyúlhatnak a bárhol és bármely korszakban rögzített objektivált szellemi tartalmak felé, és ezzel gazdagodva-átalakulva a kor objektív szellemi tartalmait is – visszahatásként – átalakíthatják. A szellemi létréteg szintjén így létrejön az élő kollektívum, és míg a biológiai szinten csak a faj közössége hordozza a mindenkor elenyésző egyedei felett a közös lét keretét, és ugyanígy az lelki élet is csak az egyes egyén lelki élete lehet, addig a szellemi lét szintjén épp a közös korszellem tartalmai által szocializált egyéni szellemekkel – mindez pedig belefonódva a sok korszakkal korábban objektivált tudásba és szellemi tartalmakba – az egyéni szellem és a kollektív objektív szellem közösen létezik. Hartmann megfogalmazásában: „*Lelki élete mindenkinek saját maga számára van. Ez az individuum számára ezoterikus, nem átadható. (…) Az ember ugyan együtt tud szenvedni mások fájdalmával, vagy együtt örülni, de ez megmarad egy második fájdalomnak vagy örömnek az eredeti mellett, és minden bensőségesség mellett kvalitatívan más minőségű lesz. A gondolat szintjén azonban, ha az ember átveszi azt, az ugyanaz a gondolat lesz, noha egy másik gondolati aktusban jeleik meg. Más tudatban végbemenő gondolati aktus, de ugyanaz a gondolat marad*" (Hartmann 1962: 15–16).

Hartmann egy másik különbséget is tesz a szellemi létrétegen belül, amely az objektív (élő) szellemi tartalmak és az objektivált (holt) szellem határain jelenik meg. Ugyanis a múltban rögzítődött szellemi tartalmak – hitek, magatartási minták, erkölcsi és más kulturális értékek stb. – úgy is belenyúlhatnak, és így létezhetnek a mai objektív szellemi tartalmakban, hogy tudatalatti szinten mint magától értetődőségek tömegesen követettek. De egy másik fajta mába belenyúlást jelent, ha csak pusztán mint objektivált szellemi tartalom létezik rögzítetten, de már nem jelenik meg a tömegesen követett hitek, tudások, értékelések stb., szellemi tevékenységeiben. Ekkor csak az egyéni szellem tudatos viszonyulása a holt objektivált szellemi tartalmakra hozza be ezeket a mába (Hartmann 1962: 38). Nézzük meg, hogy miként változott e három szellemi létforma viszonya napjainkig – már túlmenve Hartmann 1930-as évekbeli állapotain is –, és miként kezdtek belefonódni ezekbe a mesterséges intelligencia formái.

A mesterséges értelem fokozódó belefonódása a szellemi létrétegbe

A szellemi létszféra emberi közösségekben való dominánsabb helyzetbe kerülése és az alatta levő létmeghatározók hátrébb szorítása az értelmi rögzítettség lehetővé válásával indult meg a valamilyen fajta írás kialakulásával. Ez persze csak az emberi közösségek életének vékony keretét jelentette a legtöbb ilyen szintre elért civilizációban, és a széles tömegeket, illetve ezek mindennapi életét a rögzített értelem nem igazán érintette. Az európai civilizációban az 1400-as évek közepén létrejövő nyomtatás sem jelentett eleinte változást ebben, de a felső rétegek egyénei számára ez a technikai könnyítés már növelni kezdte az írástudás fontosságát, és a mindennapi élet pillanataiban is sűrűbben kezdtek a rögzített értelem felhasználásával gondolkodni, és tevékenykedni. Végül az 1800-as évek folyamán ez Európában és az innen más kontinensekre átterjedt, európai kultúrájú országokban lassanként a teljes emberi társadalomra kiterjedt. A rögzített értelem az általánossá vált írni-olvasni tudás bázisán az 1900-as évek elejétől már a mindennapi életbe is közvetlenül belefonódott kalendáriumok, napilapok – a felsőbb rétegekben ezen kívül még hetilapok és magazinok – formájában, majd a mozifilmek és a rádiózás hangrögzítéseivel még szélesebben, és az 1950-es évektől a televíziózás általánossá válásával a napi élet minden percét a rögzített értelem írásbeli, hangos-mozgóképes formái hatották át, és formálták az egyes emberek gondolkodását, lelki életét és tevékenységét. Ennek folyamán az emberi lét négy létrétegének hierarchiában az értelmi létréteg egyre inkább csak átformálva és részben visszaszorítva hagyta működni az alsóbbakat, és ezt a társadalmak civilizálódásaként fogták fel.[3]

Az igazi lökést azonban ez még csak ezután kapta meg az 1980-as évektől indulóan, amikor az addig szűk körben végzett számítógépes programozás és felhasználás a személyi számítógépek tömeges elterjedésével az írás digitalizálását hozta létre, és az írásos értelemrögzítés a számítógépes digitális rögzítéssel az állandó korrigálás állapotában tudott maradni. De nemcsak az értelmi rögzítettség folyékonnyá válását hozta ez létre, hanem a szövegszerkesztőkkel és ezek könnyű konvertálásával az egyéni értelmi rögzítettség osztott értelmi rögzítettséggé válását is. Az ebben rejlő potenciált aztán az internet 1990-es évekbeli elterjedése ténylegessé is tette. Amit valaki leír, kigondol, videóban feltesz az internetre, az percek múlva századok és ezrek gondolkodását, tevékenységét befolyásolhatja.

[3] Nobert Elias ennek a civilizálódásnak a menetét a természeti szükségletek kielégítésének változó formáin mutatja be, gazdag empíriára alapozva a korai újkortól indulóan, lásd Elias (1987).

Kevin Kelly a következőképpen írta le ezt a folyamatot tizenkét technológiai fejlődési dimenziót elemezve. Mindennek a középpontjában az értelem rögzítettségének folyékonnyá válása áll a digitalizálás révén, a *flowing*, melynek során a korábbi írásos rögzítettség merevsége után a rögzített értelmet létrehozó ember számára az állandó újragondolás, változtatás, és az egyes értelmi részek elválasztása, illetve más kontextusra kigondolt értelmi keretbe átvitele könnyűvé vált. A szellemi szektorokban tevékenykedő egyes emberek számára már ez létrehozta a könnyű felemelkedést az értelem papíron való, fizikai rögzítettségéből az értelem állandó reflexív lebegésének állapotába. A tudós, a művész, az elméleti jogász stb. az állandó elméleti reflexiói során az eredményeit mindig csak ideiglenesen rögzíteni is képessé vált, melyek újragondolása, részleteinek könnyed megváltoztatása, egyes részletek más kontextusokra való átvitele mostantól kezdve szinte akadálytalanul lehetségessé vált. A *Flowing*, a folyékonnyá válás a komputeres digitalizálás a Kelly által kiemelt aspektusok között az összes többi alapja. Ahhoz, hogy ez a lehetőség a legszélesebb emberi közösségben is folyékony *osztott értelemként* működhessen, már csak egy egységesebb szerkesztői program kellett, és ez a számtalan verzió között a könnyű konvertálás mellett a Microsoft Word és még néhány másik program kiemelkedése révén a '80-as évek végére létre is jött. A folyékony és könnyen közösségivé tehető értelem aztán az internet létrejöttével valóságosan is megindult azon az úton, amely mára alapjaiban átalakította a szellemi lét működését. Fokozatosan minden a *Becoming*, a folytonos változásban levő létmódba kerül, és ezzel a tradicionális társadalomtól a modern felé már végbement változásra átépülés a főbb funkcionális intézmények vonatkozásában (hatályon kívül helyezhető jog, választásokon lecserélhető államhatalom, cáfolásig élő tudományos igazság stb.) továbbterjed, és szinte minden egyes dolgunk, közösségi intézményünk már az állandó változásban létezik. A folyékony értelem *Flowing* aspektusa a kommunikáció minden formájára kiterjedéssel a *Screening*, a Képernyő Emberének aspektusát vonja maga után a korai, centralizált TV képernyője után a mai smart tévézésig vezető úton, ezzel párhuzamosan a komputer képernyőjéig, ugyanígy a mobiltelefonálás képernyőjével, mely az okostelefonok egyre több funkció átvételével egy általános televízió/komputer/telefon/mesterséges intelligencia képernyőjévé változik. Ezen az úton a teljes környezetünk értelmi reflexió alá vétele az addigi fizikai-biológiai létszféra dolgaira való puszta „ráülésünk" helyett ezek értelmi reflexiókkal, kognícióval átitatódását hozta létre, így a *Cognifying*, a kognifikálódás sorrendben az előbbi aspektusok által lehetővé tett fejleményt jelent. Az *Interacting*, a kognifikálódással okossá vált dolgaink „visszafigyelése", és reakcióink megfigyelésével számunkra visszajelezve elkezdik tevékenységünk kiegészítését, terelését. Mindezek hozzák létre az *Accesing*, a hozzáférés középpontba kerülését az eddigi tulajdonlás révén biztosított mód háttérbe kerülésével. Hisz miért tulajdonoljam a dolgokat, ha mindenhez percek alatt hozzájuthatok (saját autó helyett Über-taxi, főleg ha az önvezető lesz); elég, sőt kényelmesebb ebben az állapotban az, ha csak hozzáférésem van, és törődjön vele, tartsa karban más a dolgokat (Kelly 2016, részletesebb elemzéséhez lásd Pokol 2016a).

A Kevin Kelly által kiemelt legújabb folyamatok tehát átrendezik a szellemi létréteg három létformájának (egyéni, objektív és objektivált) Hartmann által jelzett hangsúlyait is. Az egyéni szellem a múlthoz képest sűrűbben, ezernyi szálon és folyamatosabban belefonódik a kor objektív szellemi tartalmaiba, és nem egyszerűen csak a korai szocializáció során veszi át a korszak szellemi tartalmait – jórészt egy életre –, hanem sűrű napi érintkezésben is, formálódva ezáltal, és saját szellemi adalékát is rögtön az internetre téve némileg vissza is formálva a korszak objektív szellemi tartalmait. Ugyanígy az objektivált szellemi tartalmak is állandóbban és folyamatosabban a kézhez állnak a mindennek az internetre kerülésével, és a pillanatok alatti lehívhatóság állapotába kerüléssel. Így szinte

alig válnak el egy korszak objektív szellemi tartalmai és válnak holttá, pusztán objektiválttá a bárhol létrehozott szellemi tartalmak. A Hartmann által még hangoztatott három szellemi létforma erőteljes elkülönülése így a tendenciaszerű közeledés állapotába került, noha elkülönültségük teljesen nem szüntethető meg.

A mesterséges értelem közvetlen összekötése a mechanikai léttel

Ha a kiindulópontban tisztáztuk, hogy az ember négy létréteg együttese, és minden értelmi megnyilvánulása mögött is ott van valamilyen közvetettséggel a lelki és a biológiai létmeghatározása, akkor a mesterséges értelemmel működő robot és az ember különbségeit is tisztábban fel tudjuk tárni. Michio Kaku írja nemrég megjelent könyvében, hogy a Rodney Brooksszal készített interjúban az interjúalany azt mondta neki, hogy a robot is éppúgy gép mint ahogy az ember is az, és így egy nap épp olyan élő gépeket tudunk majd építeni, mint amilyenek mi magunk vagyunk (Kaku 2014: 263). Hartmannt komolyan véve ezt így nem mondhatta volna, még akkor sem, ha az egyre fejlettebb és kifinomultabb programok már nemcsak az értelmi műveleteket tudják reprodukálni, és betáplálni a robot tevékenységébe, hanem az érzelmek algoritmusba foglalása, vagy ugyanígy a fiziológiai fájdalomérzet programozásban szimulálása révén ezek a robotok számára is elérhetővé válhatnak. A programozás ugyanis csak digitálisan imitálni tudja, *értelmi szinten*, a fiziológiai érzéseket és az érzelmeket, de mivel e mögött nincs meg a tényleges lelki-limbikus, érzelmi réteg és a biológiai-fiziológiai, létrétegi mechanizmusok, mindez csak imitált lehet. A mesterséges értelemmel működő robot menthetetlenül csak „kétrétegű" lehet, és akármilyen komplex is lesz a programozása, és a lelki életre jellemző reakciókat, illetve a fiziológiai-biológiai mozgásokat is végre tudja hajtani beprogramozása révén, akkor is csak két létréteg együttese lehet az ember négy létrétegűségével szemben. Nem szabad becsapódni attól, hogy ma már egy retina nélküli, vak ember számára mesterséges retinát kreálva – mintegy beépített kameraként – a nyakszirtlebenybe vezetve valódi látási érzékszervet hozhatnak létre (Kaku 2014: 264). Ám robotba ültetve ez még megmarad puszta kamerának, hiába tudja ez a programozással összekötve a mechanikai-fizikai test mozgását meghatározni. Kaku ezeket tárgyalva a könyvében – evidensként elfogadva Rodney Brooks előbb kritizált megállapítását az ember és robot azonosan gép jellegéről – a beépített fájdalomérzettel rendelkező robotok emberi jogainak reklamálásáról ír, és etikai kérdések megjelenéséről (Kaku 2014: 250–252).

A robotnál az értelmi létréteg digitális reprodukálása történik meg a programozásában, és amennyiben egyre gazdagabb tud lenni ez a programozás, le tudnak nyúlni az emberi lét alsóbb létrétegeibe is. Ekkor az itteni reakciókat is algoritmusba foglalva az értelmi reakciók mellett a lelki és a fiziológiai reakciókat is reprodukálni tudják, és ezt az egyre intelligensebb *értelmi* programot közvetlenül a fizikai-mechanikai testekkel kötik össze. De ennek egy másik megjelenési módja, amikor emberi rokkantak vagy másképpen mozgásképtelen sérültek agyhullámait közvetlenül kötik össze béna testrészeikkel, és a sérült agyrészt kikerülve, de annak funkcióit imitálva egy program lép be, és válik ismét mozgóképessé az addig béna ember. De e nélkül is – mint Stephen Hawking esetében –, az agyhullámok értelmi reakcióit összekötve a lebénult ember a kerekesszékkel mozgóképessé válik, illetve mozgatni tudja a külvilág tárgyait, vagy gondolatai agyhullámait egy algoritmus hanggá átalakítva beszélni tud. „Telekinézis: az elmével irányított anyag" – írja fejezetcímében Kaku, és ez pontosan visszaadja a négy létréteggel élő ember helyére belépő kétrétegűségre redukálódást Stephen Hawking esetében, aki a puszta agyi-értelmi lét és

ennek közvetlen mechanikai világgal összekapcsolási lehetősége révén tud már csak élni. (Persze élő aggyal, és így valahogy táplálni, és anyagcseréje révén tisztába kell őt tenni.) Az így megteremtett technika azonban a puszta fizikai robottesttel összekötve a későbbiekben esetleg hozzájárulhat majd a mindenféle lelki létréteg és a biológiai létréteg nélküli létezés létrejöttéhez. Ennek elemzéséhez az önszerveződővé váló mesterséges intelligencia létrejöttének esélyeihez kell fordulni, ahogy azt magyar nyelven is megjelent műveikben Ray Kurzweil vagy Nick Bostrom már elemezte (Kurzweil 2014, Bostrom 2015).

Előtte azonban érdemesnek tűnik az eddigi elemzéseinkből eredő következtetések levonása az értelmi létréteg és a felette létrejövő esetleges új létréteg vonatkozásában. Úgy tűnik, hogy addig, amíg csak gazdagodik az értelmi létréteg, és ez egyre gazdagabban csak felhasználja az eddigi értelmi műveleteinkbe belefonódó mesterséges értelem feljavító hatásait, nem lehet új létréteg létrejöttéről beszélni. A Kevin Kelly által leírt összes tendenciát is beleértve ez nem más, mint a számottevőbben az emberrel megjelenő, de eleinte még így is csekély szerepű értelmi létréteg megizmosodása. Egyre sűrűbben és folyamatosabban használt, illetve az ember-lét alsóbb létrétegeit egyre átalakítóbb hatású értelmi létrétegről lehet beszélni még a mesterséges értelem által napjainkig feljavított állapotában is. Sőt, ha ez még csak a kezdet a mesterséges értelemmel feljavítottság és környezetünk dolgainak „megokosítása" menetében, és ennek sokszorosa megy végbe a következő néhány évtizedben – ahogy Kelly prognosztizálja –, ez akkor is csak az eddigi negyedik, legfelsőbb létrétegünk lesz. Új létréteg felmerüléséről csak akkor beszélhetünk, ha a korunkban kialakított mesterséges intelligencia formái, az algoritmusok, az agyhullámok algoritmusba foglalásai és mechanikai testekkel való közvetlen összeköttetései valahogyan önszerződővé válnak, és az emberi értelemmel összefonódásuk megszűnése mellett tudnak működni a világban. Még egy további kérdés, hogy ez akkor csak egy további új létréteg kibomlása lesz-e, mint ahogy megtörtént ez már háromszor is az elmúlt majd öt milliárd évben a Földön – nélkülözhetetlen előfeltételként átalakítva-megtartva az addigiakat alsóbb létrétegként, vagy egyszerűen csak ráépülve az alsóbbakra –, vagy ehhez képest ez az evolúciós ugrás más lefolyású lesz?

Az önszerveződővé váló mesterséges intelligencia

Alapvetően Ray Kurzweil és Nick Bostrom már idézett könyveiből kiindulva az embertől elszakadó és önszerveződővé váló mesterséges intelligencia létrejöttéhez két út és fejlesztési forma jöhet számba: a mesterséges gépi intelligencia ma is létező gyenge formájának minőségileg megerősödő alakja révén, vagy az emberi elmét hordozó agy emulációja révén, és ennek digitális hordozóra átültetésével az emberi létrétegek korlátaitól elszakításával. Harmadikként ki kell térni a mesterségesen feljavított intelligenciával ellátott ember szuperintelligenssé válására, noha ez csak a mai, mesterséges intelligenciával élésünk további formája lehet, mely nem szakad el evolúciós ugrásként a mai négy létrétegű emberi léttől csak a legfelső dominanciáját erősíti tovább. Jelezni kell, hogy az önszerveződő mesterséges intelligencia három útjának bővebb kifejtését korábbi tanulmányom tartalmazza (Pokol 2016b), és itt elegendő a mostani tanulmány központi tézisének, a létrétegek egymásra épülésének szempontjából röviden tárgyalni ezeket. (*A fejezet további része a hivatkozott tanulmány vonatkozó részeinek rövidített változatát tartalmazza – a szerk.*)

A gépi értelem erős MI alakja

A gépi értelem erős mesterséges intelligencia (MI) alakja azt a fokozatot jelzi, amikor a mesterséges intelligencia eléri az emberi értelem szintjét, majd azt ezerszeresen és milliószorosan meghaladja, szemben a gyenge MI ma ismeretes teljesítményével. Előkérdésként még az is felmerül, hogy ez lehetséges-e egyáltalán, és tényleg létre tud-e jönni ilyen teljesítményű mesterséges intelligencia? Az elemzések az eddigi exponenciális gyorsasággal növekvő teljesítményéről azonban ezt könnyen megválaszolhatóvá teszik: igen létrejön, kérdés csak az, hogy ez mikor lesz, már 2040 körül vagy csak 2100-hoz közeledve. Ezt szem előtt tartva így két alapkérdés merül fel: 1) elszabadul-e az ilyen fokozatú mesterséges intelligencia az ember és az emberi társadalom intézményi felügyelete és irányítása alól; 2) milyen természetű lesz ez az elszabadult mesterséges intelligencia, önálló öntudattal és átható akarattal rendelkezik-e, amely az emberek kívánalmaitól függetlenül irányítja hatalmas valóságalakító kapacitását, vagy ezzel szemben ez az önálló átlátó akarat nem tud létrejönni párhuzamosan az óriási technikai kapacitással együtt, és csak mint egy naiv kisgyerek átlátási képessége lesz párban az óriási technológiai változtatási kapacitással.

Az első kérdést – az elszabadulás lehetőségét – illetően a gépi értelem fejlesztésének fő irányaként létező, genetikai algoritmusokon és rekurzív önfejlesztésen alapuló öntanulás és önmegváltoztatás következményeiből kell kiindulni. Ennek révén csak a kezdetekben betáplált paraméterek megválasztásában van meg az emberi ellenőrzés és behatás, de ezután egyrészt az ezek megvalósítására az embertől elszakadó megoldási javaslatok és a gépi értelmet irányító mechanizmusok jönnek létre a gépi értelem belső irányításában, másrészt a betáplált paraméterek módosítása is az öntanulás hatókörébe kerül. Beleértve ebbe még azt is, hogy a rekurzív önfejlesztés technológiájával még maga a megoldásokat hordozó hardver megváltoztatása is létrejön. Ezek már mind ma is léteznek, de a mai gyenge MI szintjén az emberi értelem fölénye az öntanulás és rekurzív önfejlesztés kicsúszását az emberi irányítás alól még meg tudja gátolni. Ám a maihoz képest milliószoros gyorsasággal végbemenő öntanulási ciklusok és rekurzív önfejlesztési ciklusok órákra, percekre és másodpercekre csökkenése naponta akár százszoros alapvető változásokat tudnak majd létrehozni, amelyek már túl lesznek az emberi értelem általi ellenőrzés lehetőségén. Az erős MI emberi ellenőrzéstől való elszakadása egy pont után már egyszerűen következik a mai tendenciákból.

A következő kérdés így az, hogy milyen természetű lesz ez az emberi ellenőrzés alól kiszabadult mesterséges intelligencia. Ennek megítélésére egy disztinkciót érdemes tenni, és külön kell választani az intelligencián belül a *technológiai intelligenciát*, és külön az átfogó valóságban és az emberi *társadalom valóságában eligazodás képességét*. A technológiai intelligencia a fizikai-biológiai világ átalakítási képessége, és e képesség még egy külön aspektusának nagysága arra vonatkozik, hogy ez mennyiben megakadályozhatatlan más erők (így az ember) behatása révén. Ez a fizikai-biológiai világ maga alá gyűrésének és uralásának képessége. Ez az, ami egyre inkább növekszik a mesterséges intelligencia területén, míg az átfogó valóság, benne az emberi társadalom valóságának átlátási képessége, valamint az emberi társadalom fennmaradását biztosító feltételek átlátása, messze elmarad ettől. Nick Bostrom ez utóbbi MI-programba való beépítésének fontosságát elemezte könyvében, mely alapján két problémára lehet rámutatni. Az egyik probléma az, hogy az emberi társadalom alapszerkezetének és ennek fennmaradásának nincs egyetlen objektív paraméter-rendszere, ehelyett az értékek kiválasztása és hierarchiába állítása az egyes társadalmi csoportok értékválasztásaitól függően változik. Attól függően, hogy melyik társadalmi csoport dominál, választhatók ki a legkülönfélébb értékek és értékhierarchiák. Ez azonban még csak a kisebb baj. A nagyobb probléma abból fakad, hogy még az így sze-

lektíven és hiányosan – és esetleg a társadalom nagyobb része számára hátrányos értékhierarchia – is a rekurzív önfejlesztés áldozatává válhat az elszabadult erős MI változtatásai révén. Ha önmaga tudja értelmi komponenseit újra és újra megalkotni, akkor semmi biztosíték nincs arra, hogy ne váljanak a technológiai paraméterek akadályaivá a betáplált társadalmi értékpremisszák, és ne változtassa meg, tüntetesse el ezeket a mesterséges intelligencia az önváltozatási folyamatainak már néhány ciklusa után.

Így pontosabban megadva a disztinkciót, ez a *technológiai intelligencia és a társadalmi értékekre vonatkozó intelligencia szembenállásában* fogható meg. Miközben a mesterséges intelligencia a technológiai intelligencia dimenziójában óriásivá válik, addig a társadalmi értékekhez és ezek ütköztetéséhez, illetve feloldásukhoz szükséges intelligencia dimenziójában a buta kisgyerek szintjén maradhat. Ha pedig – ezt feljavítandó – az értékekre és ütköztetésükre, feloldásuk kezelésére külön algoritmusokat illesztenek be a mesterséges intelligencia programjába, semmi biztosíték nincs arra, hogy ne kerüljenek kiiktatásra rövid idő után az öntanulás és a rekurzív önfejlesztés révén. Ebből következően megítélésem szerint nem a sokszor olvasható leírás a megfelelő a „gonosszá vált" mesterséges intelligenciáról, hanem az átfogó valóság és az emberi társadalom valóságának átlátásával nem rendelkező, vak-buta, *és ennek ellenére hatalmas*, technológiai intelligenciától kell félnünk. Ez a mesterséges intelligencia nem gonoszságból tolhatja félre az emberi társadalom létezési feltételeit, ha óriási technológiai kapacitásával ez meg tudja tenni, hanem egy buta kisgyerek átlátó képességi hiányai miatt. Már itt jelezni kell azonban, hogy az emberi agy emulációjának (az elme feltöltésének) utóbbi években felerősödő fejleményei ez utóbbi problémát módosíthatják (lásd a következő részben).

Ezt az összefüggést kihangsúlyozva, az erről gondolkodó kutatók az ember fölé növekvő, hatalmas mesterséges intelligencia társadalomra veszélyességének megfékezésére legalább programjának kezdeti algoritmusában igyekeznek olyan működési elveket elhelyezni, melyek biztosíthatják az emberi társadalom számára veszélyes fordulatokat elkerülni. Egy újonnan megjelent kötetben Ben Goertzel és Joel Pitt közös tanulmányukban az emberi társadalom felé pozitív elfogultságot biztosító programelemek lehetőségeit igyekeztek felmérni a mesterséges intelligencia tervezésénél. Abból indultak ki, hogy ennek biztosítását ugyan nem lehet teljes mértékben garantálni, de hogy legalább ebbe az irányba ösztökéljék működését, azt be lehet iktatni a programjába (Goertz és Pitt 2014: 65). Így a gépi értelem „barátságos" irányának biztosítására az első imperatívusz, amit szükségesnek tartanak beépíteni az egyre hatalmasabbá váló gépi értelem programjába, hogy a rekurzív önfejlesztés ciklusai az első időszakban a lehető leglassabbak legyenek – az emberi értelem felfogóképességére szabva még –, és hogy ez az első ciklusokban még ne mehessen végbe teljesen önállóan, az emberi közreműködés nélkül. Ugyanígy az etikai elvek jól végiggondolt beépítését is javasolják a gépi értelem programjába, és ezek gazdag példatárral ellátását, illetve ezt segítendő, az erre ösztökélő szituációkon való sokszoros kipróbálását, végigfuttatását javasolják még a kezdeti fázisokban (Goertz és Pitt 2014: 72). Végső fokon azonban még így is csak reménykedni lehet, hogy a teljesen önjáróvá vált és hatalmas változtatásokra képes erős mesterséges intelligencia nem számolja fel az emberi társadalom működésének előfeltételeit.

Az emulált emberi agy digitális létezése

Az erős mesterséges értelem létrejötte a gépi értelem már működő, gyenge alakjából már több évtizede szem előtt tartott lehetőség, ezzel szemben az „elme-feltöltés" (*mind uploading*) vagy más elnevezésben emberi agyi emuláció csak az utóbbi évtizedben került a

fokozódó érdeklődés középpontjába. Ez is a gépi értelem, a mesterséges intelligencia egyik ágát jelenti, de itt az előbbi mellett más ösztönzők adják a fő motívumot. Ugyanis, ha sikerülne a teljes emberi agyi emuláció, és az eddigi biológiai folyamatok helyett komputeres futtatások is lehetővé tennék az agy idegfolyamatainak működését, akkor ennek egyik perspektívája, hogy az ember elméje és személyisége – megszabadulva az elenyészésnek kitett biológiai test általi hordozás kizárólagosságától – egy örökéletű hordozóra is átkerülhetne. A Transzcendes című film Johny Depp főszereplésével néhány éve ezt dolgozta fel, de a kutatások is a legintenzívebben folynak, és például az elmúlt években az Európai Bizottság is egy másfélmilliárd eurós összeget adott erre a kutatási célra, kezdve a legegyszerűbb élőlények kisméretű agyának emulációjával, majd a kisebb emlősök és ezzel párhuzamosan az emberi agy emulációjának előrehaladásával.

A tiszta gépi értelemhez képest az emberi agy emulálásával létrehozott önálló létezés kérdései a mai technikai állapotok szintjén még kevéssé láthatók át. (Egy új hír a világsajtóban arról szólt, hogy 2017-re várható egy patkány agyának teljes emulációja.) Csak ezután lehet megalapozottabban, tudományosan elgondolkodni azon, hogy az ilyen agyi emuláció mennyiben ismétli meg az eredeti létezőt, és hogy gondolkodási funkciói hogyan működnek, esetleg mennyiben térnek el a fizikai-biológiai testtel rendelkező létezőétől. Az emulációk reális létezése nélkül ezért tényleg csak a filozófiai szintű kérdések szintjén lehet erről gondolkodni. Persze ez sem haszon nélküli, ám mindenképpen csak spekulatív szintű lehet, és ezt szem előtt kell tartani a következőkben.

Az emberi agy emulációjánál előkérdés még maga a technikai megvalósíthatóság is, az, hogy a sok milliárd agysejt sokbilliónyi kapcsolódásaihoz – szinapszisaihoz – szükséges számítógépes kapacitás (tárhely és gyorsaság) létrejöhet-e egyáltalán. Az elemzések az eddigi exponenciális gyorsaságú fejlődést alapul véve a mai elégtelenség után ezt körülbelül harminc éves fejlődéssel elérhetőnek tekintik, így ezzel nem lesz probléma. Egy új híradás szerint például egy emberi agy neurális – tehát a legrészletesebb működésének – szintjén végbemenő folyamatok egyetlen másodpercnyi hosszát emulálva („lefejtve") és számítógépes formátummá átalakítva, majd a világ leggyorsabb számítógépén futtatva negyven percet vett igénybe ez a futtatás. Vagyis ma még 2400 másodperc alatt lehet reprodukálni egyetlen agyi másodperc folyamatait a számítógépeken, és ez első végiggondolásra elbátortalanítóként hathat. Ám ha a Moore-törvény jövőbeni érvényét továbbra is feltesszük, vagyis a számítógépes teljesítményeknek körülbelül másfél évenként való duplázódását – amire például a kvantumszámítógépek gyors előrehaladása is feljogosít –, akkor a 2400-szoros gyorsulást körülbelül 15-16 év alatt elérjük, és így az emberi agy folyamatai egy az egyben tudnak futni reális időben már a komputeren is. A fő vita inkább az lehet, hogy az elme összes tartalmának sikeres emulációja, „lefejtése" és számítógépen való megismétlése átviszi-e egyben az eredeti elme öntudatát is, és e tartalom számítógépen való futtatása közben egyben az öntudat is mindig megjelenik-e?

Az igen intenzív gondolkodás és vita e kérdés felett okos disztinkciókat eredményezett az utóbbi években. Külön kell választani ezek szerint azt, hogy az agyi emuláció után a komputeres futtatásban működni fognak-e az egyes mentális folyamatok, és ettől egy külön kérdés, hogy ezek összegződéseként az ilyen emulált emberi agyi működés egyben *létrehoz-e egy egységes öntudatot*, mintegy a párhuzamosan a folyamatosan futó mentális folyamatok *együtt-látójának* a pozícióját?! Végül egy harmadik kérdés, hogy ha igen, akkor ez az öntudat az emulálás előtti, eredeti elme öntudata lesz-e, vagy ez egy új öntudat keletkezését jelenti, amelynek csak annyi köze lesz az eredeti emberi elme hordozójához, hogy ugyanazok az emlékeik és gondolkodási stílusaik, rutinjaik lesznek. Ez utóbbi esetben mintegy digitális ikertestvérként lehet felfogni az öntudattal rendelkező, számítógépen

futó, feltöltött elmét, de ahogy az egypetéjű ikrek is külön öntudattal rendelkeznek, úgy a digitális elme öntudata is külön úton jár majd a feltöltés után.

Kurzweil és Bostrom számára e kérdésben a válasz magától értetődő, hisz lévén az elme minden megnyilvánulása, minden mentális folyamat az agyi idegsejtek kisüléseinek eredménye – vagyis csak a materiális folyamatok emergens (az elektrokémiai folyamatok szintjéről felbukkanó, e fölé emelkedő!) következménye –, így a tudat és az öntudat is csak ennyit jelent. Ebből az alapállásból következik, hogy ha elég pontos és részletes az emuláció, akkor nemcsak agyi folyamatok részletei (emlékek, élmények stb.) jelennek meg a számítógépes futtatás folyamán, hanem az ezek összegződéseként létező öntudat is. Abban azonban, hogy ez a gépi-elme öntudat mennyiben jelenti az eredeti megkettőződését, vagy ezzel szemben egy új létrejöttét, nem jelenik meg náluk elemzés, és maga a kérdés is inkább csak a legújabb tanulmányokban vált középpontivá.

Több vita és a szembenálló érvek végigolvasása után én inkább hajlok annak elfogadására, hogyha elég pontos és részletes az agyi emuláció, és az egyes neuronok kapcsolódásainak trillióit is át tudják másolni a komputeres platformra, akkor valószínű, hogy a mentális folyamatok központjaként, vezérlőjeként működő öntudat is megjelenik az elme-feltöltés után. Ugyanis, ha az ember nem fogadja el, hogy az idegi folyamatok finom mintázatain kívül lenne egy külön lélek önálló szubsztanciaként, akkor csak technikai hiányosságok miatt következtethet az öntudat megjelenésének elmaradására. Ha pedig ez utóbbit kizárjuk, és a teljes elme pontos és részletes feltöltését az eddigi technikai fejlemények után lehetőnek látjuk, akkor nem tehetünk mást, mint elismerjük a feltöltött elme öntudatának megjelenését mint lehetőséget. Ez azonban pusztán egy digitális ikertestvéri tudatot jelenthet az eredeti számára, de semmiképpen nem azt, hogy most már „két helyen" is megjelenik ugyanaz az öntudat, és „itt is vagyok, ott is vagyok" állapot jöhetne létre. És főként nem azt, hogy az ember –– megunva biológiai kötöttségét – az agyfeltöltéssel átvándorolhatna a digitális létezésbe.

Ez az álláspontja David Chalmersnek is egy nemrégiben megjelent tanulmányában, míg vele vitatkozva Massimo Pigliucci a biológiai testhez kötött öntudat kizárólagosságát vallja. Chalmers – magát funkcionalistának, Pigliuccit biológistának nevezve – így érvel a két felfogás különbségéről: „*A filozófusok két táborba tömörülnek e kérdést illetően. A biológiai megközelítés hívei azt állítják, hogy a tudat a lényeget illetően biológiai természetű, és a nem biológiai rendszer így nem lehet tudatos. A funkcionalista megközelítés hívei ezzel szemben állítják, hogy nem a biológiai szerveződés a fontos itt, hanem az ezt eredményező struktúra és az általa ellátott funkció, így a nem biológiai rendszer is lehet tudatos, amennyiben a felépítése megfelelő volt*" (Chalmers 2014: 104). Ennek a jelenleg még csak filozófiai jellegű vitának azért is van jelentősége, mert ma még – és az ezzel intenzíven foglalkozó kutatók szerint még jó pár évig – az agyi emulációnak csak destruktív technikái léteznek, melyek az állatkísérletek felhasználása révén (most eltekintve egyes állatvédő csoportoktól) nem jelentenek gondot az agyi emulációk létrehozásában és a fejlesztésében. Ám mivel már jelenleg is felmerült a vitákban, hogy végső stádiumban lévő, gyógyíthatatlan betegek számára ezzel tegyék lehetővé majd az esetleges fennmaradást – akik számára az agyfeltöltés destruktív jellege már nem jelent problémát –, így fontos kiemelni, hogy ezen az úton is csak egy digitális ikertestvér teremthető meg, de az eredeti személy elenyészése így sem kerülhető el.

Egy fontos nehézséget jelent az emberi agyi emuláció megvalósításában a legújabb kutatások szerint, hogy a sok reménykedéssel szemben nem elégséges pusztán a magasabb mentális folyamatok elkülönítése és ezek emulálása, félretolva a pusztán mozgási idegfolyamatok és más testrészekkel összekötött agyi folyamatok moduláris részeit. (Például a neocortex feltöltésére koncentrálva a kisagy és a cerebellum ideghálózatainak félretolásával sokkal könnyebben meg lehetne valósítani az emulációt.) E reményekkel szemben ugyanis

az agykutatások azt mutatják, hogy az agy egyes régióinak specializálódott szerepjátszása mellett a legtöbb idegi folyamatban többé-kevésbé minden agyi régió közrejátszik. Így a megfelelően részletes agyi emuláció nem állhat meg a magas szintű mentális folyamatok régióinak feltöltésénél, hanem a teljes agy, sőt azon túl még a főbb testrészek idegi kapcsolódásait is reprodukálni kell ehhez. Ennek egy további következménye így, hogy a sikeresen emulált agy működéséhez aztán szükséges lesz egy szintén szimulált testet is kötni, mert már technikailag sem tudna működni az eredetileg ilyen testtel összefonódott agy a komputeres feltöltés után sem (Linssen és Lemmens 2016: 5).

A genetikailag feljavított szuperintelligencia kérdése

A human géntechnológia általi intelligencia-feljavítás és élettartam-növelés terén számomra három irány tűnik érdemesnek a megvalósíthatóság szempontú elemzésre. Ennek ellenoldala, amit minden eszközzel én is tiltandónak gondolok, az emberklónozás, illetve a DNS géntechnológiai módosítása révén ember és állat vegyítésén alapuló kimérák létrehozására irányuló kísérletek. E három felsorolásszerűen: 1) az embriószelekció az intelligencia növelése céljából, 2) a nanobotok véráramban végbemenő működésének megteremtése, illetve biztonságossá tétele a belső szervek megújítása céljából, 3) és végül az agyi interfészek révén az ember biológiai alapú értelmének felerősítése. Jelezni kell, hogy a mesterséges intelligencia filozófiai kérdéseivel foglalkozó szűkebb elméleti csoportokon túl ezek az aspektusok az általánosabb elemzésekkel foglalkozó filozófus és társadalomtudós körökben is figyelmet kaptak már, és az első viták már lefolytak, szemben az előbb tárgyalt, erős MI témájának és az emberi agy számítógépes feltöltésének kérdéseivel. Peter Sloterdijk vetette fel 1999-ben – a humán géntechnológia addig elért fejlődését alapul véve –, hogy az embernemesítés ezen az úton is fokozható a jövőben, és erre Jürgen Habermas reagált a felháborodás és a morális elítélés érveinek és jelzőinek legnagyobb intenzitású igénybevételével. Persze a németeknél e kérdés érdemi vizsgálat nélküli és pusztán morális felháborodás szintjén való vitatása sajátos történelmi örökségükből fakad, miután az emberi eugenika korábbi törekvéseit a náci hatalmi rendszer karolta fel, és ezzel a német szellemi elit számára a legmélyebben diszkreditálta és tabuvá tette ezt az egész kérdéskört (a vita elemzéséhez magyar nyelven lásd Kiss Lajos András cikkét 2003-ból (Kiss 2003)). E vitán túl az elutasítás hangvételében ugyan, de a normatív beszűkültségtől mentesebben Francis Fukuyama is foglalkozott a humán biotechnológia lehetséges következményeivel 2002-es könyvében, mely alig egy év múlva magyar fordításban is megjelent. A válasz esetében is elutasítás, melynek fő oka, hogy az emberek és társadalmi csoportjaik között eddig is fennálló egyenlőtlenség ennek révén tovább fokozódik. Ezzel a „géngazdagok" csoportjai a jövőben így már nemcsak a nagy vagyonokat és kedvezőbb életfeltételeket hagyják tovább utódaikra – szemben a szegényebb rétegek gyerekeinek siralmas jövőképével –, hanem genetikailag is feljavított formát biztosítva számukra, a társadalomban létrehozzák a „géngazdagok" és a „génszegények" társadalmi csoportjait és a minden eddiginél nagyobb társadalmi egyenlőtlenséget (Fukuyama 2003: 208–210).

Fukuyama egyenlőtlenség-növekvés félelmével szemben számomra meggyőzőbb és megalapozottabb Kurzweil állítása, mely abból indul ki, hogy a kezdeti nagyköltségű humán biotechnológiai eljárások rövid időn belül olcsóvá válnak – hisz alig van bennük anyag- és energiaköltség –, így egy idő után a legszélesebb körben bevetté és rutin eljárásokká válhatnak. A másik út, a nanobotok révén a véráramban végbemenő szervcserék és állandó megfiatalítások az előbbihez képest még inkább csak a kutatások és az állatkísérletek szintjén kezdődtek el, de a Kurzweil és Bostrom által leírtak, illetve az eddigi ta-

pasztalatok a technológiák terjedése és megvalósulása terén kevés kétséget hagynak afelől, hogy a következő években itt is radikális áttörések várhatók. Az előbbi fenntartások nélküli támogatásával szemben e téren már nagyobb szkepszissel kell élni, mert a puszta élettartam meghosszabbítás az egész személyiség, habitus időskori megmerevedése mellett még csak növelheti a lét nehézségét. Már ma is inkább az a fő gond, hogy a 80-90 éves élettartamig meghosszabbított élet értelmetlenségével kell állandóan szembesülni, és képzeljük el ezt egy 130-150 éves korig kinyúló élettartam esetén (még ha néhány évvel ki is lehet nyújtani a személyiség és habitus rugalmasságát).

Az agyi interfészek jövőbeli intelligenciafeljavító fejlesztéseire rátérve, ezek ma kutatásokban és állatkísérletekben már létezőek, például egérkísérletekben a memóriafeljavító hippocampus-chipek beültetése működőképes volt, és az embereken alkalmazás kísérletei is megkezdődtek már, eleinte még például az Alzheimer-kór gyógyításának rövidebb távú céljával. Ezek exponenciális gyorsasággal fejlődése – együtt a többszörös embriószelekciók tömegesebb elterjedésével – valóban az eddigi történelemben példátlan intelligencianövekedést hozhat létre a következő évtizedekben.

Összességében így a human biotechnológia eszközei útján való intelligenciafeljavítás alapvetően üdvözlendő lehet – az e téren jelzett két szigorú tiltáson túl –, és a mesterséges intelligencia társadalomba bevonásának elsősorban támogatandó útjának ezt kell tartani, mint ahogy a téma egyik kutatója írja, ezen irány mellett letéve a voksot a másik kettővel szemben (Goonan 2014: 198).

Új létréteg vagy az elért értelmi lét bázisán a földi evolúció újrakezdése?

Az emberi ellenőrzéstől elszakadó és önszerveződővé vált mesterséges intelligencia két formájának létét kell itt szemügyre venni az előző elemzések fényében, az erős MI-t és az emberi agy által hordozott elme teljes emulálásával digitális hordozóra feltöltött mesterséges intelligenciát. Az előbbiekből látható volt, hogy az itteni, digitalizált szellemi létréteg komputeres, szerver-parkok általi hordozottsága mellett a világban való változások végrehajtására fizikai testekkel összekötése – esetleg csak egy-egy rövid időszakra – két létrétegű létezést tesz lehetővé, a biológiai és a lelki lét teljes hiánya mellett. Még ha az utóbbi egyes részei szimulálással be is kerülhetnek az erős MI esetében a programba a döntési választások meghatározásába, ennek nem lenne semmilyen funkcionális szerepe – csak akadályozó hatása –, így az öntanulással kiküszöbölése a programozásból szinte biztosra vehető. Ugyanígy, ha a teljes elme digitális feltöltésével jönne létre az embertől elszakadó mesterséges intelligencia, és így ez a teljes korábbi személyiséggel együtt tartalmazná annak lelki életi struktúráit is – szolidaritási érzelmeit, identitástudataiból fakadó értékválasztási döntési mintáit stb. –, ennek semmi funkcionális szerepe nem maradna már a biológiai és társadalmi közösségi léttől megszabadult állapotban. A korábbi személyiség feltöltéssel áthozott öntudata – mely korábban a biológiai ösztönök és létmeghatározók állandó behatása mellett alakult és működött, illetve emellett a gyerekkortól családi majd baráti közösségekre ráutaltság ereje által formált lelki élettel együtt tartalmazza az értelmi személyiségi struktúrákat is – a digitalizált komputeres létformába kerülve légüres térbe kerülne az alsó két létréteget tekintve. Biológiai-fiziológiai reakció emlékei megmaradnának egy ideig – ahogy az amputált lábú embernek is viszketésérzései vannak a már nem létező lábát illetően –, és ugyanígy az öntudat lelki életi diszpozíciói is hatás fejthetnek ki, de mindez tényleges funkcionális szerep nélküli lenne. Így a két létrétegűre csökkent szuperintelligencia világában az ilyenféle öntudat-részek eltűnésének valószínűsége a mil-

liószoros gyorsasággal végbemenő, rekurzív programfejlesztések révén az emulált személyiségek esetében is nagyon magas.

A záró rész címében feltett kérdésre így az lehet a válasz, hogy a mesterséges intelligencia mint evolúciós ugrás új létréteget nem hozhat létre a földi evolúcióban és az emberi létben oly módon, ahogy az elmúlt évmilliárdokban ez már háromszor végbement. Ugyanis amíg csak folytatja az elmúlt ezer években fokozódó erővel növekvő szellemi létréteg alsóbb létrétegeket meghatározó és azok fölé növő működését – melyet csak felgyorsított az elmúlt majd félévszázad alatt beindult digitalizált értelem megjelenése –, addig egyszerűen csak a már meglévő legfelső, negyedik létréteg dominanciájának fokozódását jelenti. Amennyiben azonban az így létrejött digitalizált értelem a jelzett módokon elszakad az emberi értelem meghatározása alól, akkor ugyan evolúciós ugrás következik be ismét – az alsóbb létrétegekben létrejött evolúciós végeredmény növekszik a meglévő létrétegek fölé –, ám az eddigiektől eltérően az új evolúciós erő már nem alapul az alsóbb létrétegekre. Működését úgy lehet leírni Hartmann létrétegeinek fogalmaiban, hogy az évmilliárdok alatt a fizikai-mechanikai lét fölé emelkedő biológiai élet, majd ennek egyre felsőbb fokozatain megjelenő lelki-érzelmi lét bázist teremtett a főemlősök, de különösen az ember szintjén az értelmi létréteg megjelenésére, és e négy létréteg együttélése végül létrehozta a digitalizált mesterséges értelmet, mely az embertől elkülönült, önálló hordozó általi működtetés révén már közvetlenül össze tud kapcsolódni a fizikai testekkel. Ezzel az evolúciós ugrással az új létforma nem ráépül az őt létrehozó létrétegek hierarchiájára – egy új hierarchikus emelettel megnövelve azt –, hanem közvetlenül csak a kezdő bázisra, a fizikai-mechanikai létrétegre lesz ráutalva. Az emberi lét együttműködő létrétegei „kiszenvedték" a mesterséges létforma születését – és ez csak így jöhetett létre –, de önállóvá válva már feleslegessé válnak számára. A mesterséges értelmi lét így nem lehet új létréteg az eddigiek fölött, hanem az evolúciónak az induló bázison való újrakezdése az önszerveződő értelem irányításával.

A nagy vitakérdés és a félelem forrása így megítélésem szerint jogos, például Stephen Hawking vagy Elon Musk által állandóan megismételve, hogy mi lesz az emberi léttel és az egész – feleslegessé vált – biológiai létszférával a világ sorsát irányító mesterséges intelligencia korában. Félelmeik jogosságát elismerve csak azt kell kiemelni, hogy az evolúció által létrehozás alatt álló új létforma nem lenne rászorulva a földi életre, és egy sor szomszédos bolygón akadálytalanul tudja kifejteni működését, ahogy Kurzweil nagy ívűen ki is fejtette a kozmosz gyarmatosítását a mesterséges intelligencia által (Kurzweil 2014: 433–564). Jelezni kell persze, hogy a fenti elemzésekben lehetőségként alapul vett erős mesterséges intelligencia létrejöttét és az emberi irányítástól elszakadását is sokan vitatják, és a szabadon szárnyaló fantázia termékének tartják, és mind Kurzweil optimista jövő vízióit, mind a mesterséges intelligencia egész emberiségre veszélyességét szenzációkeltésnek nevezik. (Mint ahogy a közmúltban a nyilvánosság előtt az utóbbit illetően éles vita volt a Facebook-alapító Zuckenberg techno-optimizmust hangoztató álláspontja és Elon Musk MI veszélyességét kiemelő véleménye között.) A mesterséges intelligenciára vonatkozó hazai teoretikus szakirodalomban ennek kitűnő összefoglalóját adja Z. Karvalics László egyik új tanulmányában (Karvalics 2015). Ezzel szemben a megítélésem szerint mint egy elvi lehetőséget nem lehet teljes mértékben kizárni egy ilyen fokozott erősségű mesterséges intelligencia jövőben való létrejöttét, és ennek a Stephen Hawkingék által hangoztatott veszélyét az emberiségre. A Nicolai Hartmann ontológiai létrétegeinek a mesterséges intelligencia elemzésébe bevonása mindenesetre ezt a végveszély jelleget új oldalról látszik felmutatni.

Köszönetnyilvánítás

A tanulmány megírása közben a vitáikkal és megjegyzéseikkel nyújtott segítséget szeretném megköszönni Karácsony Andrásnak, az ELTE filozófus professzorának és Szendrő Zoltánnak, a Miskolci egyetem nyugalmazott oktatójának.

Irodalom

Bostrom, Nick, *Szuperintelligencia,* Ad Astra Kiadó, Budapest, 2015.

Brockman, John (ed.) *What to Think About Machines That Think. Today's Leading Thinkers on the Machine Intelligence*, Harper Perennial, New York, London, Toronto, 2015.

Chalmers, David, "Uploading: A Philosophical Analysis", in: Russel Blackford and Damien Broderick (eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds*, Wiley Blackwell, Malden-Oxford, 2014, pp. 102–118.

Elias, Norbert, *A civilizáció folyamata*, Gondolat, Budapest, 1987.

Fukuyama, Francis, *Poszthumán jövendőnk. A biotechnológiai forradalom következményei*, Európa Kiadó, Budapest, 2003.

Goertzel, Ben and Joel Pitt, "Nine Ways to Bias Open-Source Artificial General Intelligence Toward Friendliness", in: Russel Blackford and Damien Broderick (eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds*, Wiley Blackwell, Malden-Oxford, 2014, pp. 90–101.

Goonan, Kathleen Ann, "The Future of Identity: Implications, Challenges, and Complications of Human/Machine Consciousness", in: Russel Blackford and Damien Broderick (eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds*, Wiley Blackwell, Malden-Oxford, 2014, pp. 193–200.

Hartmann, Nicolai, *Das Problem des geistigen Seins. Untersuchungen zur Grudlegung der Geschichtsphilosophie und der Geisteswissenschaften*, 3. unveränderte Auflage, Walter de Gruyter, Berlin, 1962.

Kaku, Michio, *Az elme jövője. Hogyan próbálja tudomány megismerni, feljavítani és többre képessé tenni az agyat*, Akkord Kiadó, Budapest, 2014.

Kelly, Kevin, *The Inevitable. The 12 Technological Forces that Shape Our Future,* Viking, Kindle Edition, e-book. 2016.

Kiss Lajos András, "Az emberiség bizonytalan jövője: Habermas és Fukuyama a biotechnológa morálfilozófiai kérdéseiről", *Holmi*, XII. évf. (2003) 11. szám, 1469–1474. old.

Kurzweil, Ray, *A szingularitás küszöbén. Amikor az emberiség meghaladja a biológiát*, Ad Astra Kiadó, Budapest, 2014.

Kurzweil, Ray, *How To Create a Mind. The Secret of Human Thought Revealed*, Viking Penguin Edition, London, 2012.

Linssen, Charl and Pieter Lemmens, "Embodiment in Whole-Brain Emulation and its Implications for Death Anxiety", *Journal of Evolution and Technology*, Vol. 25. (2016) No. 2., pp. 1–13.

Pigliucci, Massimo, "Mind Uploading: A Philosophical Counter-Analysis", in Russel Blackford and Damien Broderick (eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds*, Wiley Blackwell, Malden-Oxford, 2014, pp. 119–130.

Pokol Béla, *Szociológiaelmélet,* Századvég Kiadó, Budapest, 2004.

Pokol Béla, „A kollektív szuperintelligencia elmélete", *Jogelméleti Szemle*, XVII. évf. (2016a) 4. szám, 154–168. old.

Pokol Béla, „Emberi értelem, mesterséges intelligencia – a társadalom értelmi felépítettségének változásai", *Jogeleméleti Szemle*, XVII. évf. (2016b) 3. szám, 107-145. old.

Z. Karvalics László, „Mesterséges intelligencia – a diskurzusok újratervezésének kora", *Információs Társadalom*, XV. évf. (2015) 4. szám, 7–41. old. http://dx.doi.org/10.22503/inftars.XV.2015.4.1

Pokol Béla, jogtudós, politológus, egyetemi tanár, a szociológiai tudomány (akadémiai) doktora (1989). Az Eötvös Loránd Tudományegyetem Állam- és Jogtudományi Karán 1977-ben szerzett diplomát, politikatudományi kandidátusi disszertációját 1986-ban védte meg. Az Eötvös Loránd Tudományegyetem és a Szegedi Tudományegyetem oktatója. Főbb kutatási területei a jogelmélet, a politológia, a társadalomelmélet és a társadalmi evolúció.