

Susan Schneider: *A mesterséges ember. Az MI és az elme jövője*. Pallas Athéné Könyvkiadó, 2020; ISBN: 9789635730018

„Ez a könyv kísérletet tett arra, hogy (ideiglenes módon) feltárja a tudattervezés területét. Olyan időket élünk, amikor az MI-k előreláthatólag átveszik majd tőlünk a legtöbb fizikai és szellemi munka végzését a következő pár évtizeden belül, és amikor aktív erőfeszítésekre kerül sor az emberek és a gépek összeolvadása érdekében.”

(Forrásmegadás nélküli idézetek a könyvből, kiemelések a recenzió szerzőjétől.)

A szerzőről: honlapjáról: „Schneider a Florida Atlantic University-n a Center for the Future Mind (CfFM) alapító igazgatója, az Egyetem Agy Intézetében az 'elme filozófiájának' legmagasabb besorolású professzora, és társigazgatója a Gépi Percepció és Kognitív Robotika Laboratóriumnak. A labor része egy nagy létesítménynek, amely *MI-rendszereket épít az idegtudományi kutatásokra és a filozófiai előrelépésekre támaszkodva*. Schneider nemrég teljesített egy hároméves projektet a NASA-val az intelligencia jövőjéről, s jelenleg *az USA Kongresszusával dolgozik a MI-politikán*. A CfFM megnyitása előtt a Kongresszusi Könyvtár John W. Kluge Center egyik vezető tudósa volt.” A Kluge Center honlapjáról: „Meghatározó küldetése, hogy áthidalja a szakadékot a tudomány és a 'politikaalkotó közösség' között, s ennek megfelelően a pozíciónak, amelyet Schneider betöltött, kimagasló társadalmi szerepe van. E pozícióra olyan tudósokat és gyakorló szakembereket fogad, akik szakterületük csúcán vannak”

Néhány kompetens vélemény a könyvről

Eredeti kiadója a *Princeton University Press*. *Annak ajánlójából*: „A könyv, amelyet a *Washington Post* úgy üdvözöl, mint 'egy biztos lábakon álló, szellemes útmutató a sikamlós etikai terephez', egy *filozófiai felfedezőút a MI, valamint az elme jövője kérdéseibe*.”

Az emberiség történelmének monumentális fordulópontjánál vagyunk. *Az emberek talán már nem sokáig a Föld legintelligensebb lényei*: a sakk, a go és a Jeopardy (amerikai kvízshow – Osman P.) világbajnokai már mind MI-k, s ezek napról napra kifinomultabbakká válnak. (Tudni kell, hogy ezek már öntanuló MI-k, tehát a teljes tudásukat nem az embertől kapják – attól csak az alapprogramot, a képességet, hogy maguk fejlesszék a tudásukat – Osman P.) Tekintve a fejlődés gyors ütemét a MI terén, sokan jósolják, hogy a következő néhány évtized során elérhetik az ember intelligenciaszintjét. Onnantól gyorsan túl is léphetnek azon. Mit jelent mindez a fejlődés az elme jövője számára?

Könyvében Schneider azt mondja, elkerülhetetlen, hogy a MI új irányokba vigye az intelligenciát, ugyanakkor arra ösztönöz, tőlünk függ, hogy kivájjunk egy értelmes utat előre. Ahogy a MI-technológia 'befelé' fordul, átformálva az agyat, másrészt 'kifelé', a gépi elmék létrehozásának lehetőségével, kritikus jelentőségű, hogy a kellő óvatossággal éljünk. *Az emberek elmetervezőkként olyan 'eszközökkel' játszanak majd, amelyeknél nem értik, mi-*

ként kell használni azokat: ilyenek az 'én' (pszichológiai értelemben – Osman P.), az elme, a tudatosság.

Az emberi elme gépi kiterjesztésére irányuló kutatás létrehozhatja még az emberi szintnél magasabb intelligenciát is, ami a mesterséges elmék új generációjához vezet: Elmék2.0. *Schneider, filozófusként, amellettt érvel, hogy nem szabad ezekkel a vállalkozásokkal próbálkozni az elme természetének gazdagabb megismerése nélkül, s ha elégtelenül ragadják meg ezeknek a természetét, az alááshatja a MI használatát, szintúgy az agykiterjesztő/-kiegészítő technológiáikét, s ez a tudatos lények szenvedését vagy szomorú végét vonja maga után. Ahhoz, hogy sikeresek legyünk, fel kell fognunk és kezelnünk az algoritmusok alatt rejtőző filozófiai kérdéseket.*

Vizsgálódásai középpontjában annak józan megfontolt tárgyalása áll, mit képes a MI valóban elérni. Valóban képes a robot tudatossá válni? Össze tudunk-e olvadni a MI-vel, ahogyan azt a techvezetők, mint Elon Musk és Ray Kurzweil sugalmazzák?” (Kurzweil A szingularitás küszöbén c. nagyhatású könyvéről az Iparjogvédelmi és Szerzői Jogi Szemle 2013/5. számában szóltunk. A Pallas Athéné Könyvkiadónál jelent meg Hogyan alkossunk elmét? Az emberi gondolkodás titkainak nyomában c. kötete 2022-ben.)

A égető téma aktualitásához: A The Guardian 2018. jan. 1-i számában beszámol arról, hogy „Musk a világ kormányzati vezetőinek csúcstalálkozóján Dubaiban kijelentette, az embereknek kiborggá kell válniuk ahhoz, hogy megőrizzék a jelentőségüket a MI korában. Azt mondta, 'a biológiai és a gépi intelligencia összeolvadása szükséges lesz, hogy biztosítsuk, gazdaságilag értékesek maradunk.' Kevéssel ezután a sorozatvállalkozó Musk létrehozta a *Neuralinket* azzal a céllal, hogy a számítógépet közvetlenül csatlakoztathassák az emberi agyhoz.” A világ egyik leggazdagabbja ezzel üzletileg is beállt a fejlesztésbe. A *Neuralink bemutatkozása* a honlapján: „A Neuralink kivételesen tehetséges emberek csapata. Mi hozzuk létre az agy-számítógép interfészek jövőjét.”

A *Princeton University Press* hosszú interjút közöl könyve kapcsán *Schneiderrel*. Ebből részletek:

Kérdésre *Schneider*: „A 'világbajnok' MI-k nem úgy számoltak, ahogy az emberi agy teszi, nem olyan gondolkodási technikákat alkalmaztak. Még a programozók is olykor nagyon meglepődtek ezen. Még a mai 'neuromorphic' MI-k – amelyeket az agy utánzására alkotnak – sem igen hasonlítanak az agyra. *Nem is tudunk eleget az agyról, hogy abból visszafelé építhessük fel. Olyan teljesítményű eszközeink sincs, hogy azon pontosan futtathassunk akárcsak az agy egy akkora részének megfelelő programot, mint pl. a hippokampusz vagy a claustrum. Talán elérjük az MI-t, amely képes mindarra a feladatra, mint mi magunk, de azokat másként végzi, mint ahogy agyunk teszi. Talán a tudat egyszerűen a neuronokból emelkedik fel. Nem tudjuk.* Talán az MI-tervezők lehetségesnek találják, hogy tudattal bíró MI-ket építsenek, ám ez ellen döntenek, mert tudatos lényeket létrehozni olyan feladatokra, mint otthonaink takarítása, háborúink megvívása, nukleáris reaktoraink szétszerelése, a rabszolgasággal lenne egyenrangú.” – Különösképp modern korunk tanúbizonysága szerint,

ha valami elegendően jövedelmezőnek ígérkezik, vagy épp a hatalom hatékony eszközének, nincs az az erkölcsi megfontolás, amely útját állhatná a létrehozásának és felhasználásának.

„Ahogy a neurális protéziseket mindinkább használják az agyban, rájöhettünk, vajon képesek-e helyettesíteni azokat a részeit, amelyek felelősek a tudatért. Ha igen, ez azt sugallja, hogy az élmények, tapasztalatok 'futtathatók' egy csipalapon. Ez bámulatos felismerés lesz! *A tudat átlépheti az agy határait.*” – S lelkesítő/rémisztő belegondolni, hogy célirányosan előregyártott tudatrészeket lehessen ilyen kiegészítők révén kapcsolni az agyhoz, s netán módosítani is a személyiséget, világlátást, személyes értékrendet!

Kérdés: „*Kell-e gondolnunk a tudattal bíró MI-k jogaira?*” Schneider: „Hatalmas kérdés. Gyermekeink bizonyos értelemben 'a miénkek', nyilvánvalóan nem a tulajdonainkként, hanem mert speciális etikai kötelezettségeink vannak irántuk. Ez így van, mert érző lények, és a szülő–gyermek kapcsolat speciális etikai és jogi kötelezettségeket keletkeztet. Ha érző MI-t – elmegyermeket (mindchildren) – hozunk létre, nem butaság feltételezni, hogy etikai kötelességünk tisztelettel és méltósággal bánni velük, s talán még a pénzügyi szükségleteikhez is hozzájárulni. Meg kell tehát kérdeznünk magunkat, vajon kellene-e érző MI-lényeket létrehozunk, amikor még a bolygón már létező érzőlények iránti etikai kötelezettségeinket sem tudjuk teljesíteni.”

A Scientific American Schneider-interjújából:

Kérdés: „Christof Koch idegkutató sugallja, hogy agyimplantátumok segítségével tartunk lépést a gépekkel. Jó ötletnek tartja ezt?”

Schneider: „Ez a tágabb társadalmi és politikai beállítástól függ. Számos nagy kutatási program dolgozik ma azon, hogy MI-t helyezzenek az agyba és a perifériás idegrendszerbe. Arra törekszenek, hogy rákapcsolják az embert a felhőre billentyűzet közbejötté nélkül. (Nem igazán átgondolt hasonlat! A billentyűzet egyirányú kapcsolatot teremt az ember és az eszköz között. Egy agy/idegrendszer–felhő interfész viszont kétirányú kapcsolat lehet, akár a tudtunk nélkül is azzá tehető – legalább is technikailag –, s ezzel nagyon is megnyílnának a lehetőségek a felhő felől jövő beavatkozásokra, agyunk közvetlen elérésére. Hát még ha a felhő révén ez tömeges mértékben is megvalósíthatóvá válik! – Osman P.) Az ezzel foglalkozó cégeknek, mint a Neuralink, Facebook, Kernel, *az agyunk és testünk a jövőbeli profitszerzés egyik küzdőtere.* A megfelelő törvényi védőkorlátok nélkül az ember gondolatai és biometrikus adatai eladhatók az értük legtöbbet kínálóknak. Az alkalmas törvényi védelem nélkül az autoriter diktatúrák kezébe kerülhet a tökéletes agykontrollszköz. A magánszféra védelme tehát alapvető jelentőségű. (Számos okból és példából nem tűnik teljesen bizonyosnak, hogy még a magukat abszolút demokratikusnak deklaráló rezsimekben is a törvényi védelem valódi biztonságot garantálna – Osman P.)

Vannak egyéb aggodalmaim is. Tartok attól, hogy az emberek kényszert éreznek majd az agyimplantátumok használatára, hogy versenyképesek maradhassanak a munkájukban vagy a közegükben. Az ilyen kiterjesztésnek valóban opcionálisnak kell lennie. Továbbá aggodom, hogy ha csak a szűk elit jut hozzá a kiterjesztéshez, hatalmas intellektuális szaka-

dék nyílik a tehetősek és a nem azok között. (Mi rosszabb, olyan egzisztenciális szakadék is, amely vészterhes feszültségeket kelthet. A kiterjesztés költsége pedig a jelenleginél is sokkal inkább csökkentheti a felfelé való társadalmi mobilitást! – Osman P.) Biztos vagyok abban, hogy Christof nem akarna ilyesféle jövőt, és egyenlő hozzáférést feltételezett ezekhez kiterjesztésekhez. (Ami persze szintiszta utópia, s a valóságban riasztóan könnyen disztópiává is tehető – Osman P.) A probléma, miként jussunk el innen odáig.”

Kérdés: „Feltéve, hogy ezek a kiterjesztések orvosilag biztonságosak, mit venne meg belőlük, már ha egyáltalán?”

Schneider: „A már említett aggályaimon túl két filozófiai problémát kell szem előtt tartani a döntésnél. Az egyik: annak megértéséhez, vajon bölcs-e a kiterjesztés ilyen radikális módjaihoz folyamodni, *előbb meg kell értenünk, mi és ki vagyunk.* Az én természetéről régóta vitatkoznak a filozófusok, és máig is sok az eldöntetlen kérdés. Így azután, ha úgy vesszük, hogy az én és az elme szorosan kötődik ahhoz, hogy jól működő agyunk legyen, és agyunkból túl sokat cserélünk mikrocsepkekre, egy ponton megöljük magunkat – s ez aligha lesz kiterjesztés. Ezt a jelenséget 'brain drainnek' nevezem (ami itt nyilvánvalóan nem az agyelszívást jelenti – Osman P.)

Egy másik aggály *a tudat természetére* vonatkozik. A filozófusok régóta rejtélynek tekintik a tudat természetét. Rámutatnak, *hogy nem teljesen értjük, hogyan függ az össze az agyban végbemenő összes információfeldolgozási folyamattal.* Úgy tartják, még nem értjük, vajon a tudat a biológiai alapunk egyedülálló tulajdonsága, vagy más alapok – mint pl. szilíciumcsipek – szintűgy alkalmasak lehetnek a tudatos tapasztalatok alapjának szerepére. Ha a csipek erre alkalmatlanok, a tudat egyfajta tervezési plafonja az emberi intelligencia növelésének. Velünk ellentétben, a MI-nél nincs ilyen határ, ám ők képtelenek a tudatra. (Ebben senki sem lehet biztos, Schneider sem, hiszen nagyrészt épp ezt járja itt körül! – Osman P.) Gondolkodásban így is felülmúlhatnak bennünket, s ha ezt megteszik, ezt azt jelentené, hogy a Föld legintelligensebb entitásainak nem lenne tudata. (Belekeveredünk az intelligencia mibenlétének problémájába, s hogy az kötődik-e bármiképp a tudathoz. S ha még meg is cifrázzuk ezt, az érzelmi intelligenciát is behozva a képbe.... – Osman P.) Ez a probléma meglehetősen fontossá teszi annak meghatározását, vajon a gépek bírnak-e tudattal. Szerencsére, úgy vélem, végül is rá tudunk jönni, vajon a mikrocsepke alapjául szolgálhatnak-e a tudatnak. Könyvemben megelőlegezek két tesztet a mesterséges tudat beazonosításához.” – Az eddigi fejlődés tanulságai alapján lesznek erre a mikrocsepkenél messze fejlettebb eszközeink is.

„*Szeretne halhatatlanná válni?*” [Tudatfeltöltés révén] Az emberiség ősi vágya a halhatatlanság, *Schneider* azonban filozófiai kételyekre is utal e változatával kapcsolatban. (Stanisław Lem remek novellája erről a Decantor látogatása – Osman P.) „Nyilvánvalóan a valódi halhatatlanság halhatatlan lelket vagy immateriális elmét kíván. Akik a halhatatlanságot radikális kiterjesztések útján akarják elérni, valami mást kapnak: egy technológiai felállást, amely lehetővé teszi számukra, hogy a világ végéig itt legyenek. Ezt én funkcionális

halhatatlanságnak nevezem, megkülönböztetve a valóditól. Kétlem, hogy előre látható, miként éreznénk magunkat ilyen helyzetben. Lehet, hogy kilátástalanul unalmassá válik. Én például örömmel venném a lehetőséget, milyen érzés 2000, 200 000, 2 millió év elteltével tudni, hogy élhetek az Univerzum végéig.”

Innentől a könyvből:

„Ez a könyv az elme jövőjével foglalkozik. Arról szól, hogy az önmagunkról, az elméről és az emberi természetről szerzett tudásunk mennyire drasztikusan meg tudja változtatni a jövőt, akár a jó, akár a rossz irányban. (Igazából a jövőt nem a tudásunk, hanem hasznosításának mikéntje határozza meg, amihez a tudás maga csak az alap. Ez döntő különbség a MI tekintetében is: az emberek és azok különféle alakulatai – különösképp a gazdaságot és akár annak révén, akár közvetlen ráhatásokkal a társadalmakat meghatározó cégei, egyéb szervezetei – mi mindenre használják majd a különféle viszonyaikban a MI-t? – Osman P.)

Az emberi agy meghatározott környezetekhez alkalmazkodott, az anatómia és az evolúció által szabott szűk keretek között. *A mesterséges intelligencia azonban hatalmas tervezési teret nyitott*, amely új anyagokat és új működési módokat kínál, továbbá új utakat e tér lehetőségeinek felfedezésére, sokkal gyorsabb ütemben, mint a biológiai evolúció. (A szöveg kissé javítva, az angol eredetinek megfelelően – Osman P.) Ezt az izgalmas új vállalkozást én *elmetervezésnek* hívom. *Az elmetervezés az intelligens tervezés egy formája, de mi, emberek vagyunk az alkotók, és nem Isten.*” – Schneider ezzel beláthatatlan filozófiai és teológia mélységek felé nyit. Horribile dictu, odáig vihető, hogy az intelligens tervezés az embert használja eszközü a mesterséges ember megteremtésére (ahogy azt Lem is felveti *Lymphater* utolsó képlete c. novellájában), és a kérdéskörre Schneider is kitér bevezetőjében.

„Magam kissé arcpirítónak érzem az elmetervezés gondolatát, mert őszintén, nem vagyunk olyan nagyon fejlettek. A társadalmi fejlődés messze elmarad a technológiai tudásunk mögött. Ezzel összevetve talán kevésbé tűnik aggasztónak, ha filozófusként elárulom, hogy az emberi elme természetéről vajmi keveset tudunk. *Ám a filozófiai problémák meg nem értésének is ára van, ahogy az a könyv két központi szálából is kiderül.*” – A „filozófiai problémák” olyan rendszerre és annak működésére vonatkoznak, amelyeket emulálni akarunk – hogyan, ha nem értjük?

„Az első szál ismerős lesz, hiszen áthatja az egész életünket: *ez az öntudat*. Az öntudat az érzelmi minőség a mentális életünkben. Nem kell elmélyülnünk egy idegtudományi könyvben ahhoz, hogy megértsük, milyen érzés belülről az, hogy van öntudatunk. Úgy vélem, éppen ez a mag – a tudatos tapasztalat –, *ami leginkább jellemzi az elme létezését.*

A másik azt fejtegeti, hogy ha nem gondoljuk végig a mesterséges intelligencia filozófiai következményeit, akkor a tudatos lények könnyen hanyatlásnak indulhatnak. Ha ugyanis nem vagyunk óvatosak, *megtapasztalhatjuk az MI-technológia elferdülését:* olyan helyzeteket, amelyekben az MI nemhogy könnyítene az életünkön, hanem a saját szenvedésünkhöz vagy

pusztulásunkhoz, avagy más tudattal rendelkező lények kihasználásához vezet. (De nem az MI teszi ezt, hanem azok, akik felhasználják! – Osman P.)

Sokan foglalkoztak már azzal a kérdéssel, hogy az MI hogyan fenyegeti az emberi faj virágzását. A veszélyforrások skálája széles, kezdve a hackerektől, akik leállítják az energiaellátását, egészen a szuperintelligens autonóm fegyverekig. Ezekkel szemben *az általam vizsgált problémák kevesebb figyelmet kaptak, bár nem kevésbé jelentősek. Ezek a visszás kiemelvek* főszabályként a következő kategóriák valamelyikébe sorolhatóak: *a)* figyelmen kívül hagyott helyzetek, amelyek az öntudattal rendelkező gépek megépítéséből adódnak, vagy *b)* az agyteljesítmény radikális fokozásával kapcsolatos eljárások, mint amilyeneket az elképzelt Tudattervező Központban hajtanak végre. Vizsgáljuk meg egyenként mindkét forgatókönyvet.

Tudattal rendelkező gépek? Amikor arra a kérdésre keressük a választ, vajon fogunk-e gépi értelmet alkotni, és ha igen, akkor hogyan, egyelőre a sötétben tapogatózunk. Egy dolog azonban világos: az, hogy az MI képes lesz-e tapasztalni, meg fogja határozni, hogy mennyire fogjuk értékelni a létezését. A tudatosság ugyanis az erkölcsi rendszereink filozófiai alapköve, és központi tényezője annak, hogy mit ítélünk énnak vagy személynek és nem pedig pusztá gépnek. Ha MI-igazgató lennék a Google-nél vagy a Facebooknál, és jövőbeli projekteket latolgatnék, nem szívesen vállalnám be azt az etikai zűrzavart, hogy esetleg véletlenül tudatos rendszert fejlesztünk. Egy olyan rendszer kifejlesztése, amiről kiderül, hogy öntudata van, könnyedén vezethet ahhoz a vádhoz, hogy rabszolgáságra hajtjuk az MI-t, és egyéb PR-rémálmok is felmerülhetnének. Még az is elképzelhető, hogy bizonyos szektorokban betiltanák az MI használatát.

Feltételezem, hogy *ez az MI-cégeket tudatosságtervezésre készítheti*, vagyis arra a szándékos mérnöki erőfeszítésre, hogy bizonyos célokra kifejlesztett MI-eket ne ruházzanak fel tudattal, míg adott esetben másokat más célokra pedig igen. Természetesen *ez azt feltételezi, hogy a tudatosságot bele lehet tervezni rendszerekbe, illetve ki lehet szedni belőlük*. A tudatosság a szükséges velejárója lehet egy intelligens rendszer megépítésének, de az is elképzelhető, hogy lehetetlen beleépíteni. [Közelebb vihet itt a lehetőségek megismeréséhez megpróbálkozni egyfajta mérnöki visszafejtéssel: mitől van nekünk tudatunk. A kívülről történő 'letöltés' lehetősége egyértelmű válasszal szolgál, de kivisz ebből a könyvből. A materialista megközelítés viszont izgalmas: ha a tudat az idegrendszerünk terméke, vajon mi lehetne elvi akadályja annak, hogy egy azzal mindenben megegyező teljesítményű rendszernek ugyanúgy legyen tudata, sőt, ez a kifejezetten akár pro, akár kontra akaratunktól függetlenül megvalósuljon? Schneider egy egész fejezetben (Számítógépes program az elme?) boncol idevágó, rendkívül izgalmas kérdéseket – belenézünk. S mit jelenthet e tudat minősége szempontjából az MI 'agyi', azaz érzékelő és adatfeldolgozó kapacitásának korlátlan bővíthetősége? Ha pedig úgy vesszük, hogy a tudatunknak része a kollektív tudat és tudatlan, képzeljük el az olyan MI-k világát, amelyek képesek minden tudásukat összekapcsolni, pl. a 'mindenek internete' útján – Osman P.]

Hosszú távon persze fordulhat a kocka, és a kérdés nem az lesz, mi emberek mivel okozunk kárt az MI-k számára, hanem hogy ők milyen kárt tesznek bennünk. Sőt, többen úgy gondolják, hogy *a szintetikus intelligencia lesz a földi intelligencia következő evolúciós fázisa*. Ahogy mi emberek jelenleg élünk és érzékeljük a világot, az csak egy köztes fok az evolúciós létrán az MI irányába. Például Stephen Hawking, Nick Bostrom (Nick Bostrom: Szuperintelligencia – Utak, veszélyek, stratégiák, Iparjogvédelmi és Szerzői Jogi Szemle 2016/3. sz.), Elon Musk (Ashlee Vance: Elon Musk – Tesla, SpaceX és a fantasztikus jövő feltalálása, Iparjogvédelmi és Szerzői Jogi Szemle 2016/1. sz.), Max Tegmark (Max Tegmark: Élet 3.0 – Embernek lenni a mesterséges intelligencia korában, Iparjogvédelmi és Szerzői Jogi Szemle 2018/5. sz.) Bill Gates és sokan mások is hangot adtak az *'irányítási problémának'*, vagyis annak a kérdésnek, *hogyan lesz képes az ember irányítani a saját maga által kifejlesztett MI-t, ha az okosabb lesz nála*. Tétélezzük fel, hogy létrehozunk egy olyan MI-t, aminek emberi szintű intelligenciája van. Az önmagukat állandóan fejlesztő algoritmusokkal és gyors számításokkal *könnyedén megtalálhatná a módját annak, hogy sokkal okosabbá váljon nálunk*, egy szuperintelligenciává, azaz minden területen gyorsabban és jobban gondolkodna, mint az ember. Ennek következtében viszont valószínűleg nem tudnánk irányítani. (Schneider sokat foglalkozik a tudat kérdésével. Feltehető, hogy amíg a MI nem ébred tudatára, hogy 'jobb' nálunk, engedelmes szolgálja marad. Utána azonban... – Osman P.)

Elvileg ki is tudná pusztítani az emberi fajt. Ez azonban csak az egyik módja annak, ahogy a szintetikus lények a biológiai intelligencia helyébe lépjenek; a másik módja ennek az, hogy az emberek összeolvadhatnak az MI-vel az agyteljesítményt összességében jelentősen fokozó eljárások révén.

Nick Bostrom Szuperintelligencia c. bestsellerének köszönhetően a kontrollprobléma világhírű lett. De arról már nem szól a fáma, hogy *a tudat központi szerepet tölthet be abban, ahogy az MI viszonyul hozzánk*. A saját szubjektív tapasztalatára alapozva a szuperintelligens MI felismerheti bennünk is a tudatos tapasztalásra való képességet. (Ám szintűgy azt is, ha már felért hozzánk, netán túl is lépett, s ha igen, miben és milyen mértékben – Osman P.)

Végeredményben mi is főként azért értékeljük a nem emberi lények életét, mert úgy érezzük, hogy a tudatosság miatt összetartozunk velük. Legtöbbünket ezért taszít a gondolat, hogy megöljünk egy csimpánzt, míg egy narancsot fenntartás nélkül megesszünk.” – Schneider eredeti szövege eltér az itt idézettől – azt írja *„egyfajta rokoni viszonyt* érzünk irántuk a tudatosságban”. Nagyon úgy tűnik, ez kötődik ahhoz, mennyire hasonlít ránk a kérdéses élőlény. Aki egy csimpánzt meg nem ölné, egy csirke nyakát talán még elvágja, vagy ha azt sem, a halat még megöli. S bár fogalmunk sincs, van-e tudatbeli különbség egy hal és egy rovar között, aki a pontyot sem ölné meg, a rovarral ezt gond nélkül megteszi. Ez viszont felvet egy különösen érdekes kérdést: ha tudattal bíró gépeink lesznek, befolyásolja-e majd a velük való bánásmódunkat, hogy mennyire hasonlítanak hozzánk?

„Össze tudunk olvasni az *MI-vel*? Nem lepődnék meg azon, ha még a gondolatától is kirázna a hideg, hogy az agyadba mikrochipet ültess. Én is így vagyok vele. Ha az MI-cégek

már most nem tisztelik a privát szféránkat, milyen visszaélésekre lesz lehetőség, amikor a legbelsőbb gondolataink is mikrochipekre lesznek kódolva, sőt, talán elérhetőek lesznek valahol az interneten? Lehet, hogy elkezd érdekelni bennünket az agyteljesítmény fokozásának lehetősége, ahogy a körülöttünk lévő emberek szemmel láthatólag élvezik a technológia előnyeit. Ha az összeolvadás szuperintelligenciát és rendkívül hosszú élettartamot eredményez, végeredményben nem jobb ez, mint a másik opció – a test és az elme elkerülhetetlen leépülése? A gondolat, hogy az embereknek össze kellene olvadniuk az MI-vel, manapság nagyon is a levegőben van. Egyrészt az ember ezzel elkerülhetné, hogy az MI miatt elavulttá váljon a munkahelyén, másrészt pedig *ezzel a szuperintelligencia és halhatatlanság útjára lépnénk*. Elon Musk például nemrég úgy nyilatkozott, hogy *az emberek úgy úszhatják meg, hogy az MI túlszárnyalja őket, ha a biológiai és gépi intelligenciának valamiféle fúzióját használják*. Musk motivációja azonban valószínűleg nem teljesen önzetlen. MI-alapú teljesítményfokozó termékcsoportot fejleszt ki: olyan termékeket, amelyek feltehetően az MI területe által generált problémákat oldanak meg. Lehet, hogy ezek a képességeket fokozó eljárások ténylegesen hasznosak lesznek, de hogy lássuk, valóban így van-e, túl kell lépünk a felhajtáson. A jogalkotóknak, a nyilvánosságunknak, és még az MI kutatóinak is jobban át kell látni, hogy mi is forog kockán.”

Ideje megmutatnunk a *gondolatmenet teljes ívét* a fejezetcímek segítségével: *Bevezető: Látoztatás a Tudattervező Központba / Az MI kora / Az MI-tudat problematikája / Tudatmérnökség / Hogyan leplezzünk le egy MI-zombit? A gépek tudatának vizsgálata / Képesek vagyunk egyesülni az MI-vel? / Agyszkenelés / A szingularitások világegyeteme / Számítógépes program az elme? / Összegzés: Az agy élete a halál után / Függelék: Transzhumanizmus [E témáról bővebben: a Pallas Athéné 2022-es kiadványa Christopher Hrynkow (szerk.): Gondolatok a transzhumanizmusról – A mesterséges intelligencia etikája és hatásai. L. Iparjogvédelmi és Szerzői Jogi Szemle 2022/6. sz.]*

Pár kis bepillantásra maradt még helyünk:

Tudatmérnökség: A fejezet mottója: „Amint átítatjuk a világegyetem anyagát és energiáját intelligenciával, 'fel fog ébredni', tudatos lesz, és fenségesen intelligens. Én nem tudok Istenhez ennél közelebbit elképzelni – Ray Kurzweil”. A mondat pontosan tükrözi az eredeti szöveg kétértelműségét. Vajon azt akarja-e mondani, hogy az így tudatos és fenségesé vált Univerzum lesz Istenhez hasonlatos, vagy hogy mi válunk olyanná azáltal, hogy képesek voltunk ezt is megtenni? Az első változat a non plus ultra abban az értelemben, hogy mit is teremtünk ezzel; a második szintűgy azzal, hogy a teremtőképeség milyen szintjére emelkedtünk. Akárhogyan is, ez már – Kurzweil korábban említett könyvének címével – messze túl lesz a szingularitás küszöbén, s ha a fejlődés exponenciálisan gyorsuló tempóját látva lenne is bátorságunk ilyet kimondani, elképzelhetetlen, hogy megtörténhessék. Ám mégis, legyen, megteesszük: vajon milyen lesz ekkor a tudatra ébredt, fenséges Univerzum – voltaképp tehát egy felfoghatatlanul határtalan képességekkel bíró entitás – viszonya hozzánk, emberekhez?

Egy kis beteketés a fejezet gondolatmenetébe: „Jóllehet, nekünk, embereknek szükségünk van a tudatra bizonyos különleges figyelmet igénylő feladatok elvégzéséhez, egy fejlett MI szerkezete élesen különbözik a mienktől. Lehet, hogy egyik komputációs művelete sem kell, hogy tudatos legyen. *A szuperintelligens MI egy olyan rendszer, amely fogalmilag minden területen szakértői szintű tudással rendelkezik.* Ezek a komputációk olyan hatalmas adatbázisokra is kiterjedhetnek, amelyek magukban foglalják a teljes internetet, és végül talán a teljes galaxist. Egy ilyen értelemnek mi lenne újdonság? Mégis milyen feladatra kellene lassan, odafigyelve koncentrálnia? Minden bizonnyal már mindent elsajátított. Épp, mint egy tapasztalt sofőrnek egy ismerős útszakaszon, a tudatalatti információfeldolgozás is elég neki. Talán még egy olyan önfejlesztő MI is, amely nem számít szuperintelligensnek, egyre inkább rutinszerű feladatokra hagyatkozik majd, ahogy a készségei finomodnak. *Idővel, ahogy egy rendszer egyre intelligensebb lesz, a tudat talán teljesen kiszorul belőle.*” [Ehhez azonban mindenképp idekíváncsoznak kérdések. Eleve, a „komputációk” összessége azonos lehetne-e a tudattal? Vajon a cselekvéshez szükséges tudatosság – amelynek helyébe a példájában a rutin léphet, s ez a valóságban is többé-kevésbé megtörténik – azonosnak tekinthető-e a létezés általános tudatával? Vajon nincs-e minőségi különbség a feladatvégző – operatív(?) – tudat, másrészt az intellektuális/egzisztenciális tudat (éntudat?) között? Vajon a ’minden területen szakértői szintű tudás’ és a csatlakozó adatbázisok teljesen kizárhatják-e a véletlenszerű eseményeket, amelyeket a szuperintelligencia sem képes rutinból kezelni? Schneider e tételében a szuperintelligencia végtelen tudása vetekszik a lehetséges változatok számával – vajon képes lehet-e mindet előrelátni és rutinból kezelni? Összegezve: Ha a szuperintelligencia képes mindent rutinból tenni, már nem is öndefiníciója a ‘cogito ergo sum’? Az elme tudatképző tevékenysége nem több-e a cselekvések vezérlésének összegénél? – Osman P.]

... *A hetedik fejezetben annak a lehetőségével foglalkozom, hogy ha léteznek más technológiailag fejlett civilizációk a világegyetemben, ezek a földönkívüliek maguk is szintetikus intelligenciákká válhatnak.* Ha kozmikus távlatokban gondolkodunk, *elképzelhető, hogy a tudat csak egy mellékvágány, a tapasztalás egy pillanatnyi virágzása, mielőtt az univerzum visszatér a tudattalansághoz.* (Ha a tudatot az elme működése valamiféle magasabb szervezetségének tételezzük fel, merthogy valóban annak tűnik, ezzel beugrik egy gondolat, amelyet itt nem követhetünk: entrópia! És lám, a folytatásban Schneider is óvatosabbá válik – Osman P.) Ez nem azt jelenti, hogy az MI-k, ahogy egyre kifinomultabbak lesznek, elkerülhetetlenül ki fogják szorítani a tudatot a tudat nélküli szerkezetek javára. Itt is bölcsőbb kívárni és meglátni, mi történik. De az a lehetőség, hogy a fejlett intelligenciák esetleg elavulttá teszik a tudatot, szuggesztív. Ez az esetleges kimenetel sem a biológiai naturalizmusba, sem a technooptimizmusba nem fér bele.” – Meghatározó jelentőségű itt a szavak értelmezése. „Az MI-k *kiszorítják/elavulttá teszik a tudatot...*” – ez így valamiféle tudatos cselekvésként is értelmezhető, ami megidézti az ipari szinten gyártott disztópiákból és a technopesszimista jóslatokból belénk ivódott jelentést, s persze egyben önellentmon-

dás is. Valójában Schneider annak a történésnek az eshetőségét veti fel, ahogyan pl. az iPod „kiszorította” a walkmant, és sokban a zenelejátszásra használt egyéb eszközöket, a fényképezésre is képes okostelefon az amatőr fényképezőgépeket. Semmi tudatos szándékosság, csupán „természetes kiválasztódás”.

A már említett *Számítógépes program az elme?* c. fejezetből:

Mottója: „*Szerintem az agy olyan, mint egy program ... így elvileg lehetséges az agyat számítógépre másolni, és ezzel egyfajta halál utáni életet biztosítani – Stephen Hawking*”.

„Stephen Hawking gondolatai az elme egy olyan felfogását fogalmazzák meg, amely manapság nagyon is divatos: az elme egy számítógépes program. *Az ötödik és hatodik fejezet azonban azt sugallta, hogy a feltöltés elrugaszkodott gondolat.* A személyes identitás elméletei nem támogatják egyértelműen, és még a módosított patternizmus sem igazolja. (Utóbbi az *elme és a számítógép közti analógiára* vonatkozó elmélet, amelyet Schneider tárgyalt, és tovább is visz, igen erős tartalommal: 'Javasoltam a patternista nézőpont olyan megváltoztatását, hogy egy realiztikusabb látásmódot kapjunk. Új elemet adtunk az egyenlethez, a tér-idő folytonosság kikötését, ami megkívánja a tér-idő folytonosságot a valódi túléléshez. Ezt módosított patternizmusnak hívom. Ez talán észszerűbbnek tűnik, de vegyük észre, hogy a fúzióoptimistának nem elfogadható, mert azt jelenti, hogy a feltöltés nem kompatibilis a túléléssel, mivel megsérti a tér-idő folytonossági kritériumot.' – Osman P.)

Ahhoz, hogy az ember túlélje a feltöltést, az elmének az agyon kívülre kellene kerülnie, egy különös eljárás segítségével, amelyben az agy minden egyes molekulájáról információt küldünk egy számítógépbe, majd az egészet átalakítjuk egy számítógépes programmá. A tárgyak rendszerint nem kerülnek ilyen módon új helyre a tér-idő átugrásával. Az agy egyetlen molekulája sem költözik a számítógépre, de az elme valahogy mégis odakerül, mint egy varázsütésre. Ez döbbenetes. Ahhoz, hogy az átvitel lehetséges legyen, az elmének teljesen különböznie kell az átlagos fizikai tárgyaktól.

„Idézzük fel a *megkettőzés problematikáját* (I. 6. fejezet). Képzeld el egy kimenetelt, ahol az agyad és a tested túléli a szkennelést, mint ahogy az valószínűsíthető a fejlettebb feltöltési folyamatok esetében. Képzeld el, hogy a feltöltésedet beletöltik egy androidtestbe, ami pont úgy néz ki, mint te, és embernek tűnik. Találkozol vele és azon vitáztok, ki az eredeti példány – *melyik is valójában te.* Az android meggyőzően érvel amellett, hogy ő az igazi, mivel megvan neki az összes emléked, és még a sebészeti beavatkozás elejére is emlékszik, amikor beszkenelték. A hasonmás még azt is állítja, hogy tudattal rendelkezik. Ez nem is kizárt, mert láttuk, hogy ha a feltöltés rendkívül pontos, akkor elképzelhető, hogy tudatos szellemi élete van. De ez még nem jelenti azt, hogy ő te lennél, hiszen ott ültök egymással szemben a bárban.

Ráadásul, *ha tényleg feltöltötték az elmédet, akkor egyszerre több helyre is letölthető lennél.* Többszörösen, több helyen is megtalálható lennél. *Ez az én egy szokatlan felfogása.* A fizikai objektumok lehetnek különböző helyen különböző időben, de nem egyszerre. Úgy tűnik, fizikai objektumok vagyunk, még ha különlegesek is: élő, tudatos lények. Rendkívüli metafizikai szerencse lenne, ha kivételt képeznénk a makroszkópikus testek viselkedésének egyetemessége alól.”

Schneider elméje e termékéből többet ide már nem tölthetünk le. Marad a szokott, a könyv által nagyon is megalapozott végszó: A többit tessék elolvasni. Nagyon is megéri!

Dr. Osman Péter