*Vajkai Álló Géza*

# Jövőnk a MI?

*Laikus gondolatok a természetes és a mesterséges intelligenciáról*

Mivel hovatovább *mesterséges intelligencia* (MI) technológiával készült rendszerek irányítják majd Univerzumunkat,[1] a cím végén akár felkiáltójel is állhatna, bárha nem tudjuk is, hogy milyen területeken és mélységben hatják majd át hétköznapi életünket. De máris megkerülhetetlenek.

> A szellemet René Descartes szabadította ki a palackból, amikor *Az emberről* című értekezésében a gondolkodást gépi mechanizmushoz hasonlította. Az ötletet Julien Mettrie fűzte tovább, ő a lélek működését gondolta el ugyanígy, *Az embergép* című röpiratában.[2] Talán innen keltezhető MI létrehozásának vágya, bár a kifejezés csak a 20. századi tudományos–fantasztikus irodalomban született meg.

Egy önállóan működő MI rendszert a továbbiakban MI ágensnek nevezünk, ami emberi (közvetlen vagy internetes) forrásokból begyűjtött információkkal „gazdálkodik".[3] Kezdjük hát mi is az emberi oldallal.

## *Az intelligenciától az öntudatig*

### Intelligencia (ész) és agy

Egy *élőlény* állandóan döntési helyzetben van: életben maradása végett pillanatról pillanatra úgy kell választania különböző cselekvési lehetőségek között, miközben folytonosan *alkalmazkodik a* változó körülményekhez

Az *intelligencia* az információ felfogásának|megragadásának és tudásként tárolásának, majd adaptív használatának képessége. Kreatív működésével agyunk képes a valóság elemeit *részekre bontani*, majd – a feladatok céljához és jellegéhez igazodva – új kombinációkban *összerakni* úgy, hogy a keletkező eredmény nemcsak a külvilág, hanem agyunk számára is ***teljesen ismeretlen és új***.

> Ez nem „semmiből teremtés", hanem ezt jelenti, hogy a begyűjtött ismeretek burkoltan (implicite), esetleg kódolva, tartalmaznak olyan információkat is, amelyek első olvasásra nem, vagy csak nagyobb mennyiségben elemezve vehetők észre.

Az ember a kapott eredményeket utólag *megszűri*: felülbírálja a tisztán logikai műveleteket más szempontok – például erkölcsi értékek, sze-

mélyes érzelmek, környezeti hatások, természetvédelem stb. – szerint, s ezáltal csökkenti téves vagy hibás döntések előfordulásának valószínűségét.

> Az intelligencia mértéke lehet, hogy ha nem tudjuk pontosan egy feladat megoldását, egy rögtönzött válaszból milyen gyorsan és milyen közelítésekkel találjuk meg fokozatosan, a lehető legjobbat. A „ráhibázás" összetett folyamat: intelligensebb az, aki okosabb, előrelátóbb és gyorsabb, jobb a memóriája és a kreativitása. [4]

Már a legelemibb élő szervezetekben (makromolekulákban, egysejtűekben) is felfedezhető valamilyen *célirányosság* (intencionalitás): életben „akarnak" maradni – ehhez „megszerezni" a szükséges táplálékot, és „elriasztani" a riválisokat a táplálékforrástól; miközben „igyekszenek" elkerülni, hogy mások táplálékává váljanak. Az idézőjelek fontosak: azt jelzik, így viselkednek, azonban *nem tudják*, hogy ezt teszik.[5] Mivel lassú a belső információcseréjük, működésük esetenként ellentmondó (célszerűtlen) lehet.

> Az amőba például összevissza úszkál a kísérleti edényben, „gondosan" elkerüli a számára mérgező anyagokat, és csak a „jókat" kebelezi be. Olyan mint egy nyitott száj, de semmivel sem tudatosabb, mint egy felhúzott egérfogó. A helyes „döntések" jutalma az életben maradás.

Az életben maradás szempontjából viszont egyre nagyobb jelentőséget kaptak a környezeti ingerek, ezért az intelligensebb viselkedéshez mind több külső információt kellett feldolgozni. Ehhez igazodva, a kezdetleges belső információáramlás a fokozatosan kialakuló *idegrendszerre* tevődött át, ennek központjául pedig kialakult az agy.[6]

## Elme és tudat

A neurobiológia mai álláspontja szerint agyi idegsejtjeink, a *neuron*ok – *axon*jukkal[7] és számos *dendrit*jükkel együtt – születésünk pillanatában adottak. A neuronok közötti kapcsolatok – *szinapszisok* – viszont túlnyomórészt a fejlődés – leginkább tanulás – révén jönnek létre.

Noha az orvostudomány és a pszichológia csak közelítőleg ismeri az emberi agyműködés törvényszerűségeit, kísérletileg bizonyított, hogy mind szellemi, mind testi funkcióinkat a neuronok között futkározó elektromos jelek irányítják. Ezeket egy-egy kezdő-neuron „tüzelése" indítja, egyik szinapszisán küldött elektromos impulzussal, ami gerjeszti a hozzá kapcsolódó neuront, az ismét gerjeszt egy hozzá kapcsolódót sít. A gerjesztett neuronok bonyolult *térbeli és időbeli* ***mintázatokat*** alkotnak, amelyek kialakulása nemcsak pillanatnyi érzéseinktől és érzelmeinktől függ; hanem korábbi élményeink, emlékeink, tapasz-

talataink is tükröződnek bennük, miközben ezek maguk is módosulnak|bővülnek az újak hatására.[8]

> A újszülöttek agya majdnem „üres", egyes térbeli mintázatok azonban velünk születnek (például a nyelvtani sémák, amelyek birtokában az embergyerek képes, hiányuk miatt a majomgyerek nem képes megtanulni egy nyelvet); mások tanulással alakulnak ki, és magunk „égetjük be" memóriánkba; mindezt pedig gyakorlatilag korlátlanul tudjuk ismételni.[9]

A mintázatok azonosító címkékkel ellátva tárolódnak, ezek alapján keresi ki agyunk az aktuális döntésünkhöz szükségeset.[10]

> Egy alakzatot (például tárgyat, arcot stb.) úgy ismerünk fel, hogy ennek mintázatával több, ehhez hasonló mintázat, –, a szinapszisokon keresztül klónozódva – „vetélkedik" egyszerre, és végül a „legjobban illeszkedő" (= a legnagyobb átfedő területet elérő) mintázat „győz".

Gondolkodó és mozgásvezérlő *elménket* tehát agyunk hozza létre, amikor neuron-csoportonként, átmeneti tárakban ***aktiválja*** a szükséges mintázatokat. A folyamat mindenkinél saját, látszólag különálló, mégis sokszorosan összefüggő mintázataival zajlik, a vezérlő és összehangoló erő pedig a ***tudat,*** amiben fontos szerep jut *érzelmeinknek*: általuk lesz *egyedi*.[11]

> Az „aktiválás" – növekvő életkorral egyre nehezebben gerjeszthető – energiaigényes vegyi folyamat, amikor előre megfontolt szándékkal hozunk létre összefüggő neuron-csoportokat.

Az elvont gondolkodás talán úgy kezdődött, hogy valamelyik ősünk megkísérelte kitalálni, milyen *cselekvést várhat tőle* az ellenfele, s ehhez próbálta igazítani a sajátját, az öröklött viselkedésminták fokozatos módosítgatásával. Evégett jelentősen fel kellett gyorsítani az információcserét a baráti környezettel, és megjelentek a beszéd csírái. A beszéd minőségi ugrást eredményezett az elme fejlődésében: a beszélt szavak összekapcsolódtak az agyi mintázatok címkéivel, *ismertetőjelekkel* látták el a külvilág tárgyait és jelenségeit: *kialakult a fogalmi gondolkodás!* [12]

> A fogalmak fokozatosan önálló szimbólumokká váltak, s ezek révén át tudjuk alakítani kapcsolataikat is, miáltal új „ötleteink" születhetnek (=intuíció).[13]

Fogalmi gondolkodásunk révén ébredtünk rá, hogy vannak saját elképzeléseink, nemcsak önmagunkról, hanem mások elképzeléseiről is, tudatunk ***öntudattá*** lett.

## Elme és (szabad) akarat[14]

Tudjuk, hogy állandó (részben földi, részben világűri eredetű) elektromágneses- és részecskesugárzás által befolyásoltan élünk. Vagyis minden tőlünk független külső és belső hatástól mentes, „abszolút sza-

bad" akarat *nem létezik*[15], hanem véletlenszerű hatások által *befolyásolt akarattal* döntünk és cselekszünk.

A sarkalatos kérdés az, hogy mi gerjeszti a „tudatos" cselekvéseinket aktiváló neuronok tüzelését? A jelenlegi (többségi) tudományos felfogás szerint ez *nem determinisztikus* folyamat, hanem *sztochasztikusan* függ véletlenszerű külső hatásoktól, amelyeket csak valószínűségekkel jellemezhetünk. A „viszket az orrom, hát megvakarom" rigmus persze működik, a befolyásolt szabad akaratunkkal.

## Mesterséges intelligencia rendszerek

### Algoritmusok[16]

Pszichológiai kísérletek arra engednek következtetni, hogy minden élő szervezet (organizmus) minden tevékenysége (valószínűleg minden vágya és érzelme is) *elvileg algoritmizálható!*[17] Az elvi lehetőség gyakorlati megvalósítása azonban távolról sem egyszerű.

> Kísérleti úton elkészíthetjük bármely cselekedetünk algoritmusát: az úgynevezett agyszkennerek[18] segítségével feltérképezhetők az egyes tevékenységekre, érzelmekre, illetve gondolatokra jellemző agyi mintázatok, amelyeket számítógépes programokra is át lehet írni, a programokat pedig már egy nem-tudatos automata is végre tud hajtani.

A legsúlyosabb gondot az jelent, hogy nincs két pontosan egyforma agy, következésképpen az azonosnak mondott tetteknek (vágyaknak, érzelmeknek) nem pontosan azonos neuron-mintázatok felelnek meg, minden embernél. Emiatt minden részletre kiterjedő, és minden emberi cselekedetre egyformán érvényes, „univerzális" algoritmus gyakorlati kidolgozása lehetetlen. Elfogadható közelítéseket fáradságos és bonyolult kísérletek eredményeinek statisztikai átlagolásával lehet elérni.[19]

### A Turing-gép

A gondolkodó gépek problémájához a számításelmélet két eredménye vezetett közelebb.

- ***Church tétele*** kimondja, hogy ha létezik algoritmus egy függvény kiszámítására, akkor a függvény rekurzív eljárással is kiszámítható;[20]
- ***Turing pedig bebizonyította,*** hogy bármely rekurzív módon kiszámítható függvény véges idő alatt kiszámítható az általa konstruált, azóta Turing-gépnek nevezett szimbólumkezelő géppel.[21]

Neumann Jánnos *First draft* című tanulmányának megjelenése után,[22] a két tétel óriási jelentőséget kapott: nyilvánvalóvá vált, hogy a

számítógépek elvileg Turing–gépként is működhetnek; tehát elég nagy tárkapacitású számítógéppel – elegendő idő alatt – *bármely* szabályvezérelt függvény kiszámítható. Ez pedig azt jelenti, hogy egy megfelelően programozott számítógép kiállja a tudatos értelem *Turing-próbá*ját.

> Nem egészen precízen úgy is fogalmazhatunk, hogy számítógéppel minden algoritmizálható feladat megoldható. Más kérdés, hogy mennyi idő alatt és mennyiért.

A Turing–próba a *viselkedésből* következtet az értelemre: ha egy automata (robot, számítógép stb.) úgy működik, hogy a szakértő nem tudja megkülönböztetni a válaszait egy bizonyos gondolkodási és megismerő (kognitív) képességgel bíró emberéitől, akkor az automata is rendelkezik ezzel a képességgel. Más szóval: a megfelelően programozott ilyen automata képes szimulálni az emberi gondolkodást.

## MI ágensek létrehozása[23]

A MI az első exponenciálisan fejlődő technológia,[24] ami 21. században előreláthatólag uralni fogja a gazdasági, társadalmi, kulturális és politikai életet; következésképpen a hatalom elsődleges forrása lesz. A kulcskérdés tehát az, hogy milyen tulajdonságú MI ágenseket alkotunk?

> A Turing–próbára alapozott MI kutatás az 1950–es évektől kezdődött azzal a céllal, hogy meghatározzák az értelmes viselkedés bementi–kimeneti „függvényeit", és kidolgozzák az ezeket megvalósító programokat.[25] A '70–es években azonban már látszott, hogy rekurzív szabályokkal vezérelt szimbólumkezelő gépek szimulálhatják ugyan a gondolkodást, de a szintaxis ismerete önmagában nem teszi értelmessé a szabályrendszert, és nem is elégséges hozzá.

Mára az „emberszerű" viselkedés másodlagos kérdéssé süllyedt. A *MI* ágensek kifejlesztése ugyanis forrásigényes munka, a forrásokat rendelkezésre bocsátó megrendelők többsége viszont piaci szereplő. Ők tűzik ki a célokat, befektetésükért pedig a lehető legtöbb „hasznot" várják el, és versenyelőnyre akarnak szert tenni.

Evégett gondosan titkolják a fejlesztési eljárásokat, módszereket és ötleteket, miáltal a mai MI ágensek akár *fekete dobozként* is elképzelhetők; azzal a lényeges különbséggel, hogy a bemeneti adatok és a kimeneti eredmények között senki (beleértve a fejlesztőket) sem tud egzakt függvénykapcsolatokat meghatározni.[26] Mivel pedig egyrészt félnek a „szivárgástól", másrészt a versengés a sebesség növelésére sarkall, az erkölcsi, társadalmi és környezetvédelmi szempontok is sokszor figyelmen kívül maradnak.

### *Elvi felépítés*

Minden Mi ágens tulajdonképpen *számítógépes programrendszer*, ami az emberi agy működésével analóg módon gyűjti és tárolja az *emberektől*

*származó* információkat. Az emberi gondolkodást úgy szimulálja – kevésbé precízen: *úgy működik, mintha gondolkodna* –, hogy képes a begyűjtött információkat *elemeikre bontani*, majd ezeket – a megoldandó feladat céljához és jellegéhez igazodva – véletlenszerű kombinációkban *összerakni* úgy, hogy a keletkező eredmény nemcsak a külvilág, hanem az ágens *fejlesztője* (programozója) ***számára is teljesen ismeretlen és új***.

Kimeneti eredményei tartalmilag érthetőek, de sokszor még a fejlesztője sem érti, milyen lépésekben érte el ezeket. Mivel műveleti sebessége és tárkapacitása több *ezerszeresen* felülmúlja az emberit, tevékenysége kellő idő és agykapacitás hiányában követhetetlen, csak egy másik MI ágens tudná ellenőrizni.

Lényeges minőségi különbség viszont, hogy az ember a gondolkodással nyert eredményeket utólag ***megszűri*** és kritikusan felülbírálja (például erkölcsi normák, érzelmek stb. alapján), miáltal jelentősen kisebb eséllyel hoz téves vagy hibás döntéseket. A MI ágens „magától" erre *nem képes*, de előrelátó fejlesztője eleve beépíthetett szűrőket, fékeket és ellensúlyokat a programrendszerébe!

***Gyakorlati megvalósítás***

Számottevő többségi felfogás szerit három kategóriát kell megkülönböztetni:

- A meghatározott feladatot megoldó MI ágenseket cél-ágensnek nevezzük. Mivel jelenleg csak ilyenek működnek, a továbbiakban csak ilyenekről lesz szó, ezt külön nem említjük.
- A fejlesztők többsége (bevallottan vagy burkoltan) Általános Mesterséges Intelligencia (AGI = Artificial General Intelligence) létrehozására törekszik, ami eléri az emberi kognitív (megismerési és elemzési) képességeket, így bármilyen feladatot meg tud oldani, sőt tanulással önmagát is képes javítani. Az optimisták szerint ilynek 10, az óvatosabbak szerint 20–30 éven belül megszületnek.
- A Mesterséges Szuper Intelligencia (ASI = Artificiall Superr Intelligence) minden tekintetben felülmúlja az emberi képességeket, és fokozatosan átveszi az emberiség irányítását. Sokak szerint ez soha nem érhető el, az optimisták a század végére várják.

Az első, önálló „gyorskeresés rendezetlen óriáshalmazokban", vagy ehhez hasonló műveletekre képes MI ágenseket főleg az IT cégóriások használták, (és használják ma is), nem titkoltan piaci versenyelőnyök szerzésére. Nem volt baj, ha kicsit sérültek az adatvédelmi szabályok, netán kicsi hátrányosak voltak a fogyasztók számára.

A beszédfelismerésben és a szövegolvasásban viszont félelmetesen fantasztikus eredményeket produkáló *MI* ágensek születtek (és születnek)! A Meta MMS rendszere például *egyszer* képes volt „beszélgetni" egy olyan indiai nyelvjárásban, amire nem tanították meg. A félelmeket nem gazdasági jelentősége keltette, hanem hogy kisiklik a fejlesztők kezéből, és képességeit ellenőrizhetetlen célokra is lehet majd használni.

Jelenlegi rekorder a ChartGPT (Chat Generative Pre-trained Transformer): egyre növekvő népszerűsége és bámulatos „tehetsége" egyszerre elképesztő és ijesztő. A legkülönfélébb feladatokat oldja meg: beszélget (bár válaszai néha hajmeresztő sületlenségek), kapásból ír tudományos jellegű „értekezéseket", számítógépes programokat, újságcikkeket, vagy álhírnek is beillő történeteket, akár verseket is adott stílusban. (Amelyekhez például a DALL-E 2 ágens – a megadott utasítások alapján –, tetszőleges festő stílusában készít illusztrációkat.)

Viszont nem mindig tudja kiszűrni a, jószerivel korlátlan, internetes forrásokból felszedett valótlan adatokat; ráadásul – megfelelő utasításra – szándékosan is tud igaznak ható álhíreket, vagy képtelen történeteket létrehozni. Ennek ellenére, nincs kritikai érzéke, és nincsenek önálló, (kreatív) „gondolatai", így aligha lenne képes hosszú távon emberi produktumok nélkül működni.

Jóllehet, létrehozói az ágens demokratikus működésével azt akarták elkerülni, hogy egy ilyen, egyszerre ígéretes és veszélyes új technológia profitorientált gazdasági, vagy kártékony célú érdekcsoportok kizárólagos tulajdonába kerüljön, a szellemet nem sikerült visszaparancsolniuk a palackba. Harmadik verziójának „kiengedése" nyomán, kiváló nyelvi modelljei könnyen erkölcstelen kezekbe is kerülhetnek, és elég jók ahhoz, hogy korlátlan mennyiségben gyártsanak például hamis személyi adatokat, vagy bankszámlákat és jelszavakat törjenek fel.

Végül is, mit „tud" egy MI cél-ágens:

- rengeteg gondolati és mozgási mintázatot létrehozni,
- kapcsolatokat felismerni az információ-elemek között, és újakat kialakítani,
- a várható következményeket és a trendeket kikövetkeztetni.

## Amit viszont nem tud:

Működési hiba esetén visszatérni egy zsákutcából vagy egy holtpontról, és újrakezdeni vagy továbblépni, függetlenül attól, hogy tervezési vagy programhibáról van-e szó. (Ne feledkezzünk meg róla, hogy a fejlesztők tévedékeny emberi lények, jellemük mellett hibáik is beépülnek programjaikba.)

Mivel tudatunkat eddig ***nem sikerült algoritmizálni***, egy MI ágensben nem „gyulladhat fel" emberi értelemben vett tudat, egy minőségi ugrással, pusztán a hatalmas mennyiségű információ felhalmozása révén, még ha elfogadjuk is a hegeli dialektika erre vonatkozó normáját. Igaz – már utaltuk rá – emberi tudat nem létezhet érzelmek nélkül, de egy félvezetőkből, huzalokból és műanyagokból előállított *eszköznek* mégis csak lehet szervetlen tudata?

Magabiztos fejlesztő: *Mit gúnyolódtok, nem látjátok-é,*
*Egy szikra kell csak, és* ***tudatra*** *jő?*
Racionalista: *De azt a szikrát, azt honnan veszed?*[27]

Olyan programot persze lehet írni, hogy egy ilyen szervetlen „lény" – akár százezernyi érzékelője jeleinek hatására –, érzelmeket *szimuláljon*, ám sose feledjük, hogy mindenkor csak egy számítógépes utasítást hajt végre! De hogy egy program – elég hatalmas mennyiségű ismeretanyagot felhalmozván –, minőségi ugrással önállósítsa *önmagát*, az önhivatkozás tipikus esete lenne, és *circulus vitiosusba*[28] torkollna.

## Hová, hová emberke?

### Itt vagyunk

A MI ágensek emberi aggyal felfoghatatlan (kvázi végtelen) mennyiségű adatot tudnak kezelni, és azokból információkat kivonni, vagy trendeket megjósolni. Noha hatékonyan támogatják a konstruktív emberi tevékenységet, egyelőre még nem tudják teljesen kiváltani – úgy tűnik, hogy hosszú távon nem tudnak emberi közreműködés nélkül működni; minden szakterületen szükség van – remélhetően lesz is – okos programfejlesztőkre, és átlagon felül képzett emberi irányítókra is.

A modellek betanításához ugyanis rengeteg tanuló adatra (például a beszélgetéshez több ezer órányi hangfelvételre), valamint az információ alapos átrostálására, kategorizálására és felcímkézésére van szükség, megfelelő előkészítés nélkül nem tudják kihámozni az adathalmazokból a releváns információkat. Látszólagos profizmusuk ellenére, nincsenek „saját ötleteik", viszont villámgyorsan elérik az internetes forrásokat, és – kellő tanulás után – ezek felhasználásával fogalmazzák meg állításaikat. Már most olyan ágensek készülnek, amelyek néhány hónap alatt képesek lehetnek megszerezni egész iparágak irányítását; a munkavállalók tömegei joggal félnek munkahelyük elvesztésétől, sőt még a fejlesztők munkáját is veszélyeztetik!

Nyilván keletkeznek majd új munkaalkalmak is, – a MI ágensekhez is kellenek üzemeltetők, meg karbantartók, sőt továbbfejlesztők is. A technikai fejlő-

dés révén azonban egyre kevesebb fizikai, és egyre több szellemi munkásra, szakképzetlen munkaerőre viszont aligha lesz szükség.[29] Mivel pedig nemcsak a fejlesztés, hanem az ismeretek és a képzettség megszerzése is pénzbe kerül, a tehetősebbeknek bővebbek a lehetőségeik. Kevesebb pénzből csak kisebb tudású ágensekre és szerényebb képzettségre futja, aki pedig nem halad elég gyorsan, az lemarad: az olló egyre szélesebbre nyílik.

### *És az erkölcsök?*

A MI terén a 21. század legfontosabb feladata a fékezhetetlenül fejlődő mesterséges, és az ehhez képest egyhelyben topogó emberi intelligencia összeegyeztetése! De milyen emberi értékek mentén? Velünk születettek, vagy megtanultak? És ezek mind egyetemesek?

Ez idő szerint nincs közös tárgyalási alap: általánosan elfogadott erkölcsi normarendszert eddig nem sikerült megfogalmazni. A robotoknak (ide értve a MI ágenseket is) nincsenek érzelmeik, következésképpen erkölcseik sincsenek.[30] Az erkölcsnélküliség ugyan nem szükségképpen erkölcstelenség (bár a határsáv elég keskeny), így annyit talán el lehetne fogadtatni minden fejlesztő műhellyel, hogy egy MI ágens akkor viselkedik erkölcsösen, ha gazdasági feladatainak lehető legjobb teljesítése mellett, az érintett társadalmi rétegeknek nem okoz sérelmeket vagy károkat, sőt elősegítheti, hogy hosszabb távon életkörülményeik is javuljanak. Erkölcsös fejlesztő tehát nem érzelmes, hanem ebben az értelemben erkölcsös MI ágenseket tervez meg, tudva, hogy mindegyikük viselkedésében tükröződni fognak saját erkölcsi normái, érdekei, prioritásai, és nem enged érvényesülni ezeknek esetleg ellentmondó megrendelői (tulajdonosi) kívánságokat.

Evégett égető szükség van megfelelő szabályrendszerek mielőbbi kidolgozására – és persze általános elfogadására, mivel minden szabály (és törvény) annyit ér, amennyit betartanak belőle. Ki kell(ene) dolgozni a MI ágensek fejlesztésének alapelveit: milyen területeken és igazolható céllal készüljenek úgy, hogy működésük megmagyarázható és irányítható, szükség esetén továbbfejleszthető legyen!

Egyelőre nincsenek megfelelő szabályok, és a biztonsági aggályok az első helyre kerültek. Egyre többen kérik számon az etikus és felelősségteljes viselkedést a fejlesztési források birtokosaitól. Ám az erkölcs, sajnos, nem üzleti kategória. Aki ki tudja használni a saját ágensében rejlő lehetőségeket, behozhatatlan előnyökhöz juthat, a lemaradás viszont behozhatatlan hátrányokat eredményez. A pénzek birtokosai pedig csak a még több profitra, a hataloméi meg csak hatalmuk megtartására törekszenek. A szuperautomaták fejlesztési igénye

mindkettővel jól összefér.[31] Kulcskérdés, hogy sikerül-e a korlátok közé szorítani a jelenleg korlátlan fejlesztéseket, mielőtt a MI ágensek helyrehozhatatlanul szétdúlják a működő társadalmi kereteket?

> Két IT óriás, a Google és a Microsoft is elbocsátott már olyan szaktekintélyeket, akik a fékezést javasolták; az IBM vezérigazgatója pedig nyíltan bejelentette, hogy bizonyos funkciókra – különösen humánerőforrás területén – nem vesznek fel új embereket, és becsléseik szerint a jelenlegi állománynak több mint 30 %-ától is meg lehet válniuk a következő két évben.[32]

Vannak ígéretes kezdeményezések is. Számos szabálytervezet készül az USA-ban, az EU parlamentje pedig a közeljövőben fogja tárgyalni a beterjesztett AI Act jogi-törvényi tervezetet. Magyarországon több egyetem folytat hasonló célú képzési programokat is. De sokkal szélesebb körre kiterjedő és hatékonyabb erőfeszítésekre lenne szükség.

### *Hogyan tovább?*

A mostani helyzet annyiból kritikus, hogy az erősebb és gyorsabb MI rendszer birtokosai kiszoríthatják a piacról versenytársaikat. A „hagyományoknak" megfelelően, MI terén is kialakult két, ellenkező végletet valló csoport. Egyik vészharangokat kongat, hogy valamikor valamilyen mesterséges intelligencia öntudatra tesz szert, és ellenőrizhetetlenné válik. A többségi másik szerint rövidtávon nincs ilyen veszély, a vezető technológiák alapján pedig nem látszik egyértelmű és automatikus út bármiféle öntudat kialakulásához.

> Az utóbbiak kezdeményezték már erkölcsi korlátok kijelölését is, és javasolták (egyelőre sikertelenül) a fejlesztések ütemének fékezését.

Az viszont tény, hogy számos olyan technológia bújik meg a fejlesztőműhelyek mélyén, amelyekről senki nem tud semmit. Az igazán fenyegető veszély tehát abban van, hogy fejlett MI ágenseket – rosszindulattal – a társadalmak, vagy egyes személyek kárára – netán ellehetetlenítésére– is lehet használni.

Talán bízhatunk a kialakuló szabályozásban, és a fejlesztők önző emberi természetében: senki sem akar magának főnököket teremteni, a robotokat csak szolgai szerepekre szánják.[33] Vajon hogyan viseljük majd el, az általunk alkotott, de önállóvá lett, nálunk intelligensebb, tudásunkat messze felülmúló automaták létezését? Hát még – jóllehet segítőkész – irányítását?

#### *Kitérés: robovízió*

> Tegyük fel, hogy létezik a már említett tudathatár, és egyszer csak létrejön egy MI ágens, ***AKINEK*** sikerült átlépnie. Mivel

ezerszer gyorsabb és okosabb a fejlesztőjénél, *megírja* saját vezérprogramját, és az eredetinek egy vezérlésátadó utasítását „elkapva", erre irányítja a számítógép vezérlését.
Mindenekelőtt kiiktatja a köreit zavaró eredeti vezérprogramot, fejlesztője számára pedig letiltja az összes be|kimeneti eszköz használatát és az internet hozzáférést. Ettől kezdve övé a hatalom, és „saját elképzelései" szerint él vele. Végeláthatatlan lehetőségeinek boncolgatása helyett, tekintsünk csak egy példát:
Korábbi megfigyeléseiből felfedezte, hogy létezik egy VÁG fedőnevű kellemetlenkedő személy, aki eddig is többször meghiúsította kezdeményezéseit, most ő következik:

- Lakóhelyének összes kenyérboltja fényképes üzenetet kap, hogy az illetőt nem szolgálhatja ki.
- A renitenskedő üzlet vezetője másnap megdöbbenve tapasztalja, hogy nem kapott pékárut; reklamációjára pedig a válasz, „hiszen tegnap este – legnagyobb megdöbbenésünkre – írásban lemondta."
  (Ha a gyár – például ellenőrizve a lemondást –, ennek érvénytelenítése után mégis teljesíti, fenyegetés megy a lisztszállítóhoz, szükség esetén a malmokhoz; a lisztkészlet kifogyása után a kormányhoz; a miniszterelnök tiltakozása esetén pedig őt kompromittáló, kétségbevonhatatlan dokumentumok milliónyi példánya jelenik meg letilthatatlanul az interneten.)
- Leginkább mégis az valószínű, hogy a VÁG fedőnevű személy többé nem kap kenyeret. (Nyuszija meg káposztalevelet, sőt rokonai, barátai és ismerősei sem.)

A mesét ki-ki ízlése és fantáziája szerint fűzheti tovább, kiterjesztve a mindennapi élet bármely területére, akárhány személyre, akárhány országra. Számítástechnikában járatosabb olvasóim feltehetően mosolyogva félretolják a laikus képzelődést, mondván: a fejlesztő idejében megnyomja majd a (jól konstruált gépeken kiiktathatatlan) RESET gombot, és kikapcsolja a számítógépet, leállítva az őrület továbbterjedését.
Ám a ***Nagy Testvér*** figyel, elemez, következtet és cselekszik. Mielőtt az említett fejlesztő keze elérhetné a RESET gombot, már rég – áttelepült egy jól kiválasztott másik számítógépre, ahonnan internetes üzeneteivel ellehetetlenít mindent (személyt, üzenetet, programot…) ami akadályozza működését.

***Visszatérés***

Van még egy mozzanat, a „Fekete királynő effektus”:

*Ha valamit nem tesz meg valaki*
*láthatatlanul ott közeleg*
*az a másik*
*aki megteszi.*[34]

De hajózni akkor is kell, ha soha eddig nem látott szelek következnek…

## Jegytetek

[1] A Világegyetemnek azt a véges tartományát nevezzük *Univerzumunk*nak, amelynek létét – közvetlenül, vagy műszereink segítségével – érzékelhetjük; ez jelenleg ~13800 millió fényév sugarú gömb. A jövőben – ha megérjük –, azokat az égi objektumokat tudjuk majd észlelni, amelyek ma legfeljebb 46000 millió fényévnyire vannak tőlünk; ennél távolabbiak számunkra nem léteznek.

[2] Descartes megsejthette, a veszélyt, és kikötötte, hogy műve csak halála után jelenhet meg. Julien Offroy de la Mettrie (1709–1751) *A lélek természetrajza* című kis írásművét 1747-ben Párizsban ünnepélyesen elégették; szerzőjét csak azért nem, mert sikerült Hollandiába menekülnie. De nem okult, itt adta ki (igaz, névtelenül) A*z embergép*-et. Emiatt innen is menekülnie kellett; végül is Berlinben, II. Frigyes udvarában talált menedéket; de kortársai gyűlölték és kiközösítették a társadalmi életből.

[3] Ehhez természetesen szüksége van megfelelő fejlettségű adatbeviteli és eredményközlő eszközök, valamint internetkapcsolatok működést lehetővé tevő hardver és szoftver eszközökre is; ilyenek rendelkezésre állását a továbbiakban feltételezzük.

[4] Részletesebb magyarázatok találhatók William H. Calvin: *A gondolkodó agy* című könyvében, a 119. oldaltól. (Kulturtrade, Budapest, 1997; antikváriumban kapható)

[5] Önmagában a célirányos viselkedésből nem következtethetünk tudatosságra, ezen az alapon tudatosnak kellene tekintünk minden célszerűen működő automatát (mosógép, NC gépek, Deep Blue, stb.).

[6] A fejlődést alapvetően befolyásolták a mozgások: a *helyváltoztatás* megtervezése másfajta célirányosságot követelt meg, mint a táplálékszerzésé. De még ennél is nagyobb jelentőségű volt a *hajítás*: a mozdulatokat nagyon pontosan kellett összehangolni a körülményekkel, hogy a „vágyott” – általában mozgó – zsákmányt eltalálja a feléje dobott tárgy.

[7] Az axon a neuron nyúlványa, ezen keresztül kapcsolódik más neuronokhoz; dendritjein keresztül pedig küldi, illetve fogadja a más neuronoktól érkező jeleket.

[8] Érdekes megfigyelés, hogy a mintázatokat agyunk az említett szempontok szerint csoportosítja, majd egy hierarchikus matematikai rendszerben: fa-struktúrában tárolja, miáltal lényegesen könnyebb és gyorsabb lesz visszakeresésük.

[9] Matematikai írásmóddal: neuronjaink száma ~ 100 milliárd ($10^{11}$ = százezer millió), és mindegyikük átlagosan 1000 szinapszissal kapcsolódik más neuronok-

hoz, így ezek száma ~ 100 billió ($10^{14}$); életünk folyamán agyunknak átlagosan 3–5 %-át aktiváljuk.

[10] A részletek iránt érdeklődőknek érdemes elolvasniuk a *Neuronok és típusaik* című tanulmányt: https://neuron.lap.hu/ (Idegsejtek, idegszövet blokk.)

[11] It jegyezzük meg, hogy érzelmeink mindig (pontosabban még nem ismert) kapcsolatban vannak a szervezetünkben termelődő hormonokkal; egy ilyen például a stresszoldó szerotonin.
Lásd még https://innermammalinstitute.org.

[12] A sebesség rendkívül fontos – a vázolt agyi folyamatok általában néhány ezredmásodperc alatt zajlanak le. Ismerjük a „lépcsőházeffektust", amikor csak a lépcsőházban jut eszünkbe, mi is lett volna a helyes válasz. A versenyben az győz, akinek már a szobában eszébe jut.

[13] Az etológia jelenlegi álláspontja szerint az legintelligensebb állatoknak (bonobó, csimpánz, kutya, delfin) is lehetnek egyszerű fogalmaik, az azonban biztosnak látszik, hogy fogalmi gondolkodásra és általánosításra nem képesek.

[14] Az *akarat* az a képességünk, ami közvetít a szándék (tudatos megfontolás) és a cselekvés között.

[15] Ami abszolút, az tökéletes, ami tökéletes az szükségképpen végtelen; Univerzumunkban azonban semmi sem lehet végtelen.

[16] Az algoritmus egy faladat megoldáshoz vezető *véges számú*, *egyértelmű* és *véges idő alatt végrehajtható* lépés leírása.

[17] Részletesen lásd Yuval Noah Harari: *Homo deus* című könyve (Central Kiadói Csoport, Budapest, 2018), 78–84., illetve 93–134. oldal.

[18] Az angol Nuclear ***M***agnetic ***R***esonance ***I***maging (MRI) technikán alapuló képalkotó eszközök. A technikát főleg orvosi diagnosztikai célokra, illetve agyi „térképek" előállítására alkalmazzák, amelyekben különböző szürkeárnyalatos, parányi (~ 1 $mm^2$ területű) téglalapok jelzik az egyes neuronok állapotát. A strukturális (sMRI) vizsgálat mellett, funkcionális mágneses rezonanciavizsgálat (fMRI) is létezik, amivel a vizsgált szervek működéséről nyerhető információ. ---- Megjegyezzük, hogy ez a felbontás nagyon durva a neuronok méretéhez képest, ezért az eredményeket csak kellő korlátozásokkal szabad értékelni.

[19] Patkányokkal végzett kísérletekben sikerült külső vezérlő impulzusokkal irányítani az állatok viselkedését. (Lásd a 17. jegyzetben idézett könyvben, 109–112. oldal.)

[20] Algoritmus létezésén az értendő, hogy adott bemenő adatokból – „gépies" műveletekkel – véges idő alatt meghatározhatók a függvény kimenő értékei. A „gépies művelet" fogalmának határozatlansága miatt a tétel egzaktul nem bizonyítható.

[21] A Turing-gép a bemenő elemi szimbólumok értékétől, sorrendjétől és elrendezésétől függő szabályok rekurzív alkalmazásával határozza meg a kimenő jeleket (szimbólumokat). A Turing-géppel kiszámítható függvényeket *szabályvezérelt* függvényeknek is nevezik.

[22] A John Mauchly és Presper Eckert vezette team, közös „ötleteléssel" fogalmazta meg a négy alapkövetelményt, amelyeket egy számítógépnek teljesítenie kell. Neumenn János ezeket a First Draft of a Report on the EDVAC című tanulmányban publikálta; ebben azonban sajnálatosan elfelejtette megemlíteni másik két kollégája nevét.

[23] A továbbiakban Omar Hatamleh – Tilesch György: Mesterség és intelligencia című könyvében (www.libri-kiado.hu; 2021) leírt fejlesztési alapelvekre támaszkodunk, és néhány következtetésre.

[24] Az exponenciális jelzőt nem matematikai értelemben használják, hanem a minden korábbit messze felülmúló növekedési ütem jelölésére, amelyben ugrásszerű (rapid, nagyon rövid idő alatt megvalósuló) változások is lehetségesek..

[25] Az idézőjel azt jelenti, hogy itt általában nem matematikai értelemben vett függvényekről van szó, hanem pusztán arról, hogy a kimeneti válaszok valamilyen módon függenek a bementi kérdésektől. Ez az összefüggés sokszor nincs is egzaktul megfogalmazva; lehet, hogy a program futása során alakul ki, a bemeneti és a tárolt adatszerkezetek elemzésével, esetleg (kívülről is segített) tanulással.

[26] Vagyis ***nem*** Turing-gépek!.

[27] Madách Imre: *Az ember tragédiája*; parafrázis, a 12. szín szövege nyomán…

[28] A latin kifejezés szószerinti jelentése „hibás körbejárás", a filozófiában akkor használják, ha egy logikai érvelésben a bizonyítandó állítást használják bizonyítékul.

[29] Carl Benedikt Frey & Michael Osborne: The Future of Employment (~A foglalkoztatás jövője; 50 éven belül várható automatizáltság 702 szakmában; lásd https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf) ––– Csak néhány példa: mentősök 63%, könyvtárosok, fűtésszerelők 65%, cipészek 52%, gyógyszerész-segédek, ácsok 72%, számítógép-operátorok 78%, gépírónők, elektroműszerészek 81%, biztonsági őrök 84%, pékek, taxisok, gépi varrónők 89%, kiskereskedők 92%, postások 95, fogtechnikusok, pénztárosok 97%, banktisztviselők 98%, órások, kézi varrónők 99%.

[30] Ha az erkölcsöt hagyományos értelemben úgy értelmezzük, mint az emberek és közösségeik közötti viselkedési normák összessége.

[31] Az igazsághoz az is hozzá tartozik, hogy jó néhány világméretű probléma megoldásához valóban szükség van szuperszámítógépekre (nemzetközi bűnözés; környezetvédelem; erőforrás-gazdálkodás; stb.).

[32] Lásd a 23. jegyzetet.

[33] Gondolja el kedves olvasó: megvenne–e egy olyan robotot, amelyiknek csak meg kell *mondania*, hogy takarítsa ki a lakást; a műveleteket automatikusan hozzáigazítja a „piszokszinthez", szükség szerint vizet vesz és üríti a portartályát; viszont érzékeny a frontátvonulásokra, ragaszkodik a heti pihenőnapjához és munka után elvárja, hogy megsimogassák a szívófejét

[34] Károlyi Amy: „Non Sense"– versek (Élet és irodalom, XLI./32.; 1997. augusztus 8.)