

A mesterséges intelligenciáról – halandóknak

„...az első ultraintelligens gép lesz
az ember utolsó találmánya.”
(Irving John Good, 1965)

Kezdetben mesterséges intelligencia (MI) alatt az emberi intelligencia teljes spektrumának művi létrehozását értették, tehát a kognitív tulajdonságok mellett az érzelmi, érzéki, erkölcsi összetevőket is, beleértve a kreativitást, a képzelőerőt, az álmodozást, a motivációkat. – Hasonlóan ahhoz, ahogy a személyiség felépül e különféle összetevőkből. Az így felfogott intelligencia azonban test nélkül elképzelhetetlen – „nincs narancslé narancs nélkül”, mondta találóan David Gelertner, neves amerikai informatikus.¹ Douglas Hofstadter, a kognitív tudományok ismert tudósa is csak a testet igénylő, érzelmekkel és képzelettel telített, kreativitásra képes változatot tartja intelligenciának.² Legmélyebb értelemben az intelligens lény érzi is a gondolatait és tudatában van annak, hogy gondolkodik. A mai szóhasználat nyilvánvalóan nem felel meg ennek az igényes követelménynek; mesterséges intelligencia alatt ma egyszerűen programozott, algoritmusok által vezérelt folyamatot értenek.

Tanulmányomban bemutatom a MI egyre bonyolultabbá váló fajtáit és megpróbálok néhány, manapság egyre több embert foglalkoztató kérdésre válaszolni: Milyen következményekkel jár a MI elterjedése és mik a veszélyei? Lehet-e a MI-nak tudata? Kicsúszhat-e az irányításunk alól? Uralkodhat-e a MI létrehozója fölött? Magam sem lévén informatikus, írásom azoknak szól, akik nem a programozás felől, hanem a társadalmi funkciók, a felhasználás és az irányítás és irányíthatóság felől közelítenek e témához. A megnevezés – hogy nevezhetjük-e intelligenciának a gépi tudást – másodlagos kérdés, felesleges belebonyolódni. (És különben is, az intelligencia meghatározásában nincs egyetértés, számos definíciója van.)

(Az agy „szkennelése”) A teljeskörű emberi intelligenciát utánzó MI létrehozására irányuló ambíció gyengült; azzal párhuzamosan, ahogy kiderült, az emberi agy lemásolása/szkennelése sokkal nehezebb feladat, mint ahogy kezdetben gondolták. Jelenleg több ilyen nagyszabású kutatási program folyik. Lausanne-ban egy tudós csoport a Blue Brain programon dolgozik: az agy minden jellegzetességét, sejtről sejtre haladva betáplálják egy szuperkomputerbe. A cél az agy pontos másának létrehozása. Az EU 2013-ban indította el Human Brains Project nevű programját, 1,2 milliárd eurós finanszírozással; induláskor az volt a cél, hogy először kisebb emlősök agyát másolják le. Ugyancsak 2013-ban Obama elnök idején az amerikai kongresszus 3 milliárd dollárt szavazott meg az ún. BRAIN programra az emberi agy idegi pályáinak a feltérképezése céljából. Paul Allen, a Microsoft egyik vezetője irányításával az agyat kódoló géneket próbálják beazonosítani. Az információtechnológiai óriás cégek kutatási tevékenységének középpontjában a MI áll. 2017-ben a vállalatok globálisan 21,3 milliárd dollárt költöttek erre a célra.³ (Összevetésül: Magyarország évi teljes K+F kiadása 2 milliárd dollár körüli.) E nagyszabású ambíciókat azonban erősen lehűtötte egy kudarc: a fonalféreg, a *C. elegans* „agyának” feltérképezése. E jelentéktelen, szimpla kis lénynek 302 idegsejtje és 7000 szinapszisa van, mindegyiket meghatározták, de évekbe tellett beazonosítani, hogy az egyes idegpályák és neuronok a féreg milyen viselkedéséért felelősek.⁴

Nick Bostrom szerint „A teljes agy emulációjához nem kell rájönnünk, hogyan működik az emberi értelem, vagy hogyan programozzuk le egy mesterséges intelligenciát. Csak arra van szükségünk, hogy megértsük az agy alapvető számítási elemeinek alacsony szintű

funkcionális tulajdonságait. A teljes agy emulációjához nem szükséges semmilyen alapvető vagy fogalmi áttörés.”⁵ Ugyanakkor problémát okoz, hogy nem elegendő csupán a magasabb szintű agyműködést előállító agyrészek letöltése, mert kiderült, hogy minden agyi funkcióban az összes agyi rész részt vesz, ha nem is egyforma súllyal. Sőt, még a főbb testrészek idegi kapcsolódásait is reprodukálni kell a letöltés során.⁶ Ha ehhez még hozzávesszük, hogy az agy működéséhez a testre is szükség van, érthető az elbizonytalanodás. Ray Kurzweil korábban még a '20-as évek végére jósolta az agy lemásolását, de Bostrom szerint ahhoz még vagy három évtizedre van szükség, de az is lehet, hogy csak századunk végére sikerül.

(Gépi tanulás) A mesterséges intelligencia legegyszerűbb változatát úgy határozhatjuk meg, hogy programozott, algoritmusok által vezérelt folyamat, melynek során a komputer az emberi intelligenciát utánozza és „ha-akkor”, „vagy-vagy” logikai szabályokat és „döntési fákat” alkalmaz. A MI egy bonyolultabb változata a gépi tanulás, amikor a gép eltér a programozástól, az algoritmus rekurzív módon fejleszti önmagát.⁷ Az önfejlesztés statisztikai-valószínűségi alapú, és adatbázisokon történő gyakorlás során megy végbe, s az ember számára nehezen átlátható. Egy egyszerű példa: számítógépünk tapasztalja, hogy milyen tárgyú leveleket, hirdetéseket nem szoktunk elolvasni, a kukába dobjuk anélkül, hogy megnéznénk, és ebből tanulva egy idő múltán ő már ezeket a leveleket elénk sem engedi, egyből spam-listára teszi.

Ma már nehéz lenne felsorolni, hogy mi mindenhol alkalmazzuk és mi mindenre jó a gépi tanulással tökéletesített MI. Tud járműveket vezetni, tőzsdei ügyleteket lebonyolítani, könyvelést és mindenféle adatfeldolgozást végezni, adathalmazból a kívánt információt megkeresni, újságot szerkeszteni, tömegben arcokat felismerni, arcok alapján jellemzést adni, egy beteg diagnózisát felállítani és a szükséges terápiát kidolgozni, röntgenfelvételeket kielemezni, jogi precedenseket keresni, egy per (vagy perindítás) kimenetelét megjósolni, befektetési tanácsokat adni, idegen nyelvű szöveget fordítani, beszédet felismerni és emberi beszédet modulálni. Képes háztartásunkat okossá tenni, idős emberek szórakoztató társául szegődni, tulajdonosának az internet szolgáltatásaival kapcsolatos szóbeli kéréseit teljesíteni (lásd Alexa és Siri), webrobotként tulajdonosa nevében chatelni, alkudozni, párkeresésben hatékonyan eljárni, stb. stb.

Egy bonyolult, sokváltozós helyzetben a MI ki tudja választani azt, ami az alkalmazó számára a bizonyos szempontok szerinti optimális megoldást jelenti. Ennek különösen a gazdasági döntéseknél lesz (már van) nagy szerepe. A „bot”-ok (chatelő robotok) helyettünk és nevünkben vitatkoznak, ügyeinket intézik, vagy online vásárlás esetén alkudoznak a világhálón. A fárasztó kampányidőszakban politikusok helyett végzik el a munkát, a feltett kérdésekre helyettük válaszolnak. A kiberhadviselés során szoftverek feszülnek majd egymásnak, és az intelligensebb győz: megbénítja a másik hátszágának ellátó rendszereit, sőt, legyőzi az ellenséges szoftvert...

A gépi tanulás egyik fajtáját a *valószínűségszámításon alapuló* módszerek adják. Ez gyökeresen más eljárás, mint a komputer programozása. Ennek során alkalmazzák az általános approximációs elméletet és a valószínűségi következtetést. Az ún. Bayes-hálókkal megbecsülik a jövő valószínűségét az elmúlt események alapján. (Ilyen jellegű a spam-szűrés.) A Markov-modellekkel egy eseménysorozat megtörténtének valószínűségére következtetnek (ilyen módon történik a beszédfelismerés, hangok felismerése).⁸

A *genetikai algoritmusokkal* az evolúciót utánozzák. Jó példa erre a repülőgépmotorok tervezése. Egy repülőgépmotort kb. kétszáz paraméter jellemez. A fejlesztők a szokásos módon úgy járnak el, hogy megváltoztatnak néhány paramétert, a motort legyártják, majd beépítve egy repülőgépbe a gyakorlatban kikísérletezik. E hosszadalmas és költséges folyamatot genetikai algoritmusok alkalmazásával a töredékére csökkentik. A megváltoztatott paraméterű változatot virtuálisan „gyártják le” és virtuális körülmények között kísérletezik ki.

De ez a módszer lehetőséget ad arra is, hogy a különféle virtuális motorváltozatokat élőlényként felfogva, sőt, nekik nemet tulajdonítva egymással „keresztezzék” (lejátszva a meiózist, az ivarsejtképződés folyamatát, az apai és anyai gének kicserélődését, majd pedig, az „újszülöttben”, az egyes gének dominánssá és recesszív válsát). Így rövid idő alatt képesek előállítani azt az ideális változatot, amit – megmaradva az élőlény-hasonlatnál – mondjuk folyamatos keresztezésekkel és kiválasztásokkal egy állattenyésztő csak hosszú évtizedek alatt képes elérni.

Rekurzív keresésnek nevezik azt a módszert, amikor a lehetséges megoldások nagy számából kell kikeresni a megfelelőt. Ilyen módon működött az amerikai műveltségi vetélkedő két nyertesét 2011-ben legyőző Watson, vagy a go játék legjobbjait 2018-ban megverő AlphaGo Zero. A valóságban ezeket a módszereket (és még több más, a gépi tanulás itt most nem említett típusát) egymással kombinálva alkalmazzák.

(*Deep learning*) A gépi tanulás legszofisztikáltabb változatát *deep learning*-nek nevezik. Ez a modell az agy működésének utánzásán alapul. Mesterséges neuronhálózatot hoznak létre: neuronok és interneurális kapcsolatok leegyszerűsített modelljét szimulálják. (Jelenleg még csak nagyjából 100 ezer neuronból áll a hálózat, de hamarosan eléri a 10 milliós nagyságot.)⁹ Minden kapcsolathoz eleve, véletlenszerűen hozzárendelnek huzalozási és szinaptikus erősségeket (súlyokat). A működés során ez finomodik. Nem mondják meg a gépnek, hogy pontosan mit tegyen, hanem példát állítanak elé: az emberi cselekvést és gondolkodást. A programnak/algorithmusnak azt kell utánoznia. Nagy terjedelmű adatbázisokon trenírozzák, meg kell tanulnia céltárgyát.

A neurális hálózatot rétegekbe rendezik. Az első rétegben elhelyezkedő neuronok – amennyiben képfelismerésről van szó – az egyes képpontoktól (pixelektől) származó információt feldolgozás után tovább adják a második sorban lévőknek, melyek elemzésük eredményét a harmadik sorban lévőknek továbbítják, és így tovább. A folyamat során az észlelés egyre elvontabbá válik, míg végül a legutolsó réteg (output layer), amely már nem csak az egyes részletekkel foglalkozik, hanem az egészszel, megállapítja, hogy az észlelt alakzat mondjuk egy emberi arc. (És továbbmelve: kinek az arca.)¹⁰ Ez a működés – a számítógépekre jellemzően – szekvenciális. A program osztályoz és mintázatokot elemez. A felismerések az absztrakció különböző szintjeinek felelnek meg; a szintek egy fogalom hierarchiáját alkotják. Minden output valószínűségnek alapszik.

Ezzel szemben az emberi agy analóg módon működik, az asszociációk lehetővé teszik a rövidítéseket, nem kell az összes lehetőséget, a teljes logikai fát „befutni”. A neokortex mintegy 300 millió mintafelismerő egységből áll, ezek egyszerre működnek és adják tovább az információt. A mintafelismerő egységek alá-fölé rendeltségi viszonyban állnak (az egyik feladata például az „a” hang felismerése, ha ez megtörtént, mehetünk tovább, a másiké az „alma” szóé, ...). Az alacsony szintű mintafelismerő egység azonnal továbbítja a felismerést és fontosságának megfelelően *súlyozza* is (!).¹¹ Ha az emberi agyat a szekvenciálisan működő komputerrel próbálnánk meg utánozni, egy egész városrész nagyságú számítógépre lenne szükség, a mikroprocesszorok hűtésére pedig egy nagy folyóra.¹²

A deep learning az alapja a komputeres beszéd-tudásának vagy arcfelismerő képességének. Ily módon a mesterséges intelligencia a művészetekre is trenírozható. Egy géppel sokat hallgattattak Bach-korálokat, minek eredményeképpen megtanult saját stílusú barokk zenét komponálni. De képesek Picasso módjára festeni, és Rainer Maria Rilke stílusában verseket írni. Létrehoztak egy szoftvert a szájról való olvasásra; szófelismerése 93,4 százalékos pontosságú, szemben az emberi szájról olvasás legfeljebb ötven százalékos pontosságával. Egy kínai szoftver fénykép alapján meg tudja állapítani, hogy az illető rendelkezik-e bűnözői hajlammal. Egy másik a személy társadalmi hovatartozásáról ad jó becslést fényképe alapján.¹³ A Stanford Egyetemen kifejlesztett szoftver 81 százalékos

biztonsággal meg tudja különböztetni az egyének szexuális hajlamait, miközben az emberek erre csak 61 százalékos pontossággal képesek.¹⁴ Egyre több vállalat alkalmaz szoftvert személyzetis helyett az alkalmazottak vagy munkára jelentkezők lekáderezésére.¹⁵ Amerikában jelennek meg szoftver által szerkesztett újságok is. A hatóságok a MI-val fedeztetik fel a pénzmosást.

A deep learning tehát merőben eltér a hagyományos programozástól. A hagyományos programozás igen körülményes, és komoly matematikai tudást igénylő tevékenység. Egy alakzat, mondjuk egy kutya felismerése programozással úgy történik, hogy a programozó végnélküli, hosszú kódsorokat ír az állat minden testrészéről és sajátosságáról, és ha a gép ezeket azonosította, „felfedezte”, akkor jön a megállapítás, hogy kutyáról van szó. Deep learning esetében viszont a tanítás ahhoz hasonló, mint ahogy egy gyereket tanítunk: megmutatunk neki egy kutya-képet, vagy egy kutyát, és megnevezzük. Majd pedig mutatunk neki más állatokat ábrázoló képeket is, pl. farkast, és elmagyarázzuk a különbséget. Minél nagyobb „mintán” tanul a gyerek, annál jobban felismeri az állatot.

(*Érzéki, érzelmi és erkölcsi programozás*) Fontos kérdés, hogyan lehet ellátni a mesterséges intelligenciát érzésekkel, érzelmi intelligenciával és erkölccsel. Ezekre már van technikai megoldás. Ha a számítógépet felszereljük érzékelőkkel (valójában itt egy robot és számítógép kombinációjáról van szó), bonyolult programozással el lehet érni, hogy a MI „megtanulja”, bizonyos hatások esetében mit kell éreznie. Nem élőlényként fogja érzékelni a hőt vagy a fájdalmat, de tudni fogja, hogy ilyen esetekben mit kell éreznie.¹⁶

A mesterséges intelligencia működtetésének egyik kritikus kérdése az *erkölcs*. Hogyan viselkedjenek a robotok és algoritmusok, milyen erkölcsi elveket és értékrendet valljanak? A *geek*-ek (újabbán így nevezik a számítástechnikai gurukat) időnként összeülnek, és megegyeznek a közös etikai elvekben. A MI trenírozása a tárgyilagosságra és előítéletmentességre azonban nehéz, bár nem lehetetlen feladat. Technikailag ez úgy történik, hogy a géppel nagy tömegű szövegben kerestetik a különböző alanyokhoz, társadalmi jelenségekhez kapcsolódó jelzőket. Ha a szöveg „elfogult” – azaz pl. egyes népcsoportokra gyakran használnak negatív jelzőket, vagy egyes társadalmi folyamatokat és jelenségeket gyakrabban ítélnék meg kedvezőtlenül, mint kedvezően –, akkor az algoritmus is elfogult és „előítéletes” lesz.¹⁷

(*Szuperintelligencia*) Alapvető fontosságú kérdés, hogy létrehozható-e az emberit meghaladó szuperintelligencia. A vezető MI-szakértők döntő része ezt lehetségesnek tartja, a vita arról folyik, hogy az mikorra várható. E szakértők nagyjából fele már századunk közepére valószínűsíti a szuperintelligencia létrehozását.¹⁸ A szuperintelligencia megjelenése egyben választ ad más, fontos vitatott kérdésre is. Például arra, hogy „kiszabadulhat-e” a MI az emberi ellenőrzés alól. Régivágású informatikusok de még mai programozók is ezt lehetetlennek tartják. Úgy gondolják, hogy egyszerű módon, a csatlakozót a konnektorból kihúzva, vagy a wifit lekapcsolva megakadályozható, hogy a MI „kiszabaduljon”. Ezek naív elképzelések; a szuperintelligenciának épp az a lényege, hogy fölülmúlja az emberit. Olyan ez, mintha mondjuk majmok fogságában élnénk, akik nem tudnák elképzelni, hogy túljárunk az eszükön. A szuperintelligencia létrejötté egy másik vitatott kérdést is eldönt. Nevezetesen: milyen célokat programozunk be a MI-ba. Gyakran olvasható közhely, hogy az első számú parancs az kell legyen, hogy a gép az embernek nem árthat. De a szuperintelligencia arra is képes lesz, hogy programját, beprogramozott céljait megváltoztassa, saját célokat alakítson ki és kövessen.

(*Fekete dobozok*¹⁹) Fejtörést és gondot okoz az algoritmusok öntanuló, saját magukat fejlesztő képessége; amikor a korábbi tapasztalatokat beépítik önmagukba, amikor a feladatnak

megfelelően tökéletesítik magukat. Egy ideje például már működnek fordítóprogramok. Az alapnyelv az angol, a szöveget először arra fordítják, majd arról a kívánt nyelvre. A programozók megfigyelték, hogy az algoritmus „tanul”, és egy idő múltán olyan rövidítéseket, új utakat alkalmaz, amelyek célravezetőbbek, mint az előírt folyamat. Már az IBM 2014-es Watsonánál is megfigyelték az öntanulás képességét. Egy új fejlemény: „a Facebook Mesterséges Intelligencia Kutatólaboratóriumának (FAIR) dolgozói egy érdekességre lettek figyelmesek: chatbotjaik (párbeszédágenseik) géptanulás-algoritmusaik segítségével az egymással folytatott kötetlen csevegés közben eltértek a szkriptjeikbe kódolt normáktól és bármiféle emberi input nélkül egy új nyelven kezdtek el egymással kommunikálni.”²⁰

Egy más példa: a DeepMind cég AlphaGo-ja (melyet később felvásárolt a Google/Alphabet) a legjobb go-játékosok közötti játszmák ezreinek tanulmányozásával tanulta meg a játékot és a stratégiaalkotást, majd pedig önmaga ellen játszott játszmák millióin tökéletesítette tudását. A DeepMind ezt követően megalkotta az AlphaGo Zero-t, amely már emberi segédlet és példák nélkül, csupán a szabályok ismeretében is képes volt megtanulni a játékot. Bevezettek egy ösztönző-motiváló funkciót: a gép a győzelemért pontot kap, és minél több pont begyűjtésére kell törekednie. Ennek a játéknak nagy múltja, tradíciói vannak. Kulcsfontosságú eleme a stratégiaalkotás. Ezért annál inkább csodálatra méltó, hogy mindennek a megismerése nélkül is képes legyőzni a világ legjobbait.²¹ Tehát a gépek fejlesztik, tökéletesítik magukat – a szakértők értetlenül állnak a jelenség előtt.

A gép fekete doboz módjára működik; nem látunk bele, és nem tudjuk, hogyan döntött, miért az adott eredményt hozta ki. De nem ez az egyetlen fekete-doboz problémánk. Maga az ember, az emberi gondolkodás is egy fekete doboz. Döntéseink saját magunk számára is rejtélyesek. A kognitív pszichológia egy ideje már tudja, hogy érzelmi alapú, ösztönös, gyors döntéseink előbb születnek meg, és gondolkodásunk csak utóbb teremti meg hozzá a racionális indokokat.²²

(Jogi vonatkozások²³) A mesterséges intelligencia az ember által felfoghatatlan és áttekinthetetlen adatmennyiség alapján jobban képes megismerni a valóságnak egy szeletét, mint mi. Ebből következően felmerül a pontosság vagy átláthatóság dilemmája. Ha a gépi doktor 90 százalékos megbízhatóságú, de működése folytán átláthatatlan diagnózist állapít meg, az „emberorvos” viszont csak 80 százalékgig pontosat, de megmagyarázható és áttekinthetőt, akkor kész a dilemma: kinek higgyen a beteg?

A *deep learning* máris megteremtette a digitális jogállam egy konfliktusát. Ha az üzleti tevékenység *deep learning* típusú adatbányászaton alapszik, akkor az adatok nem magyarázhatóak meg, és ezáltal a vállalat tevékenysége nem látható át és nem legális?²⁴ Az emberi viselkedés szintén olyan, mint egy fekete doboz, kiszámíthatatlan. De nem szükséges belelátni, hogy a helyzet kezelhető legyen: erre valók a törvények, a szabályok és az erkölcs.

A MI már ma jogi problémákat okoz. Egy hitelkérelem elutasításánál, vagy egy pénzügyi tranzakció pénzmosássá nyilvánításakor meg kellene tudni mondani, hogy az milyen alapon történt. A „fekete doboz” probléma azonban megszakítja az oksági láncolatot, melynek ismerete minden jogi eljárás alapvető feltétele. Ezért e nem-tudást egyes komputertudósok úgy próbálják meg felszámolni, hogy a gépet programozással arra kényszerítik, magyarázza meg döntését. Ha például egy madarat kell azonosítani, azt könnyen meg tudja mondani, hogy miért egy „búbos vöcsökről” van szó.²⁵ De bonyolult esetekben ez nem járható út, mert a gép csak olyasmit tud „elmondani”, amit programozója a szájába adott – tehát itt egy tautologikus problémáról van szó.

Büntetőjogi eljárásoknál a szándékosságnak alapvető szerepe van. De hogyan lehetne gépek esetében szándékról, motivációról beszélni? – a fekete doboz ezt megakadályozza; a gép a bonyolult eseteket nem tudja megmagyarázni, programozója pedig nem érti. Egy

nemzetközi jogi szakértő azzal a képtelenséggel állt elő, hogy emiatt a MI-nak jogi személyi státust kell tulajdonítani.²⁶ A fogalom így bővül tovább: „mesterséges szándék”...

(*Lehet-e a gépnek tudata?*) Két nagy rejtély maradt, amire még nem tudjuk a választ: a tudat és az Univerzum. (Én magam az életet már nem sorolom e kettőhöz hasonló nagy rejtélyek közé, hiszen már szekvenáltuk a DNS-t – bár még nagy részének, „sötét anyagának” nem ismerjük a funkciót –, tudunk élőlényt klónozni, tudunk testi sejteket visszafiataltítani és összejüket előállítani, tudunk mesterséges DNS-t készíteni, sőt, egy baktérium saját DNS-ét azzal pótolva tudjuk a baktériumot működtetni, a génszűrés és manipuláció révén élőlények alapvető tulajdonságát meg tudjuk változtatni.)

A tudat kialakulását a tudomány leggyakrabban úgy magyarázza, hogy az *emergens jelenség*: a komplexitásnak egy rendkívül magas fokán, konkrétan az agy esetében jelenik meg.²⁷ A tudatosságot a gerincesek sajátosságának tekintik, az öntudatot pedig a Homo sapiensének, és legfeljebb még az emberszabású majmokénak. Ebből logikailag az következne, hogy ha egy rendszer eléri azt a rendkívüli komplexitást, ami az emberi agyat jellemzi, akkor a tudat, sőt, öntudat abban az esetben is meg kell jelenjen, ha nem biológiai, nem szénalapú rendszerről van szó. Az elméleti biológusok azonban ezt visszautasítják, mert a tudatot biológiai alapú jelenségnek tekintik.

A *tudat mint kontinuum* elmélet számomra meggyőzőbb. (Erről Michio Kaku: Az elme jövője c. könyvében olvashatunk.) A tudat kérdését ekkor evolúciós keretek között tárgyaljuk – szemben az előző, rendszerelméleti felfogással. Eszerint a tudatot kontinuumként kell felfogni, kialakulása jól végig követhető az evolúció folyamán. Kaku a következőképpen határozza meg a tudatot: a tudat a múlt értékelésén alapuló folyamatos jövőszimuláció, amely a „visszacsatolási hurkok” (a külvilágból származó ingerek és az azokra adott válaszok) folyamatos felhasználásán alapul egy elérni kívánt cél érdekében. Kaku „kontinuum”, tudati skálája paradox módon a tudattalan termosztáttal kezdődik, mert annak már van egy visszacsatolási hurka (de a tudati szint ekkor még zéró), és aztán az élővilág egyre fejlettebb törzseit bevonva – a visszacsatolási hurkok száma egyre nő – az embernél köt ki. Felszínesen az a következtetés adódna, hogy a kontinuum a gépi intelligenciával folytatódik, de problémát jelentenek a visszacsatolási hurkok.

Piotr Anohin *funkcionális rendszerének (FR)* elmélete szerint „Egy funkcionális rendszer komplex struktúrát alkot, mely önálló rendszereket, szerveket, komponenseket, eszközöket von be és köt össze funkciója elvégzéséhez úgy, hogy azok egyetlen célirányos, sőt cél-, pontosabban jövő eredményének tételezésével egységes egésként működjenek tovább.” Tudatnak a FR-nek ezt a szelektor, illetve döntési képességét nevezi, melynek legprimitívebb góca az egó. Intelligenciának pedig ennek a stratégiai ön- és környezet fejlesztésre, az ehhez szükséges törvényszerűségek és működési logika megismerésére képes szintjét hívja.²⁸ Mivel az elképzelés nem zárja ki a nem-biológiai alapú rendszereket, ezen az alapon is elképzelhető, hogy kifejlődik a tudat.

David Chalmers nyomán elterjedt manapság az agy *interfészként* történő felfogása. Eszerint a tudat az Univerzumnak ugyanolyan alkotó eleme, mint az anyag és az energia, és az emberi agy képes arra, hogy „rákapcsolódjon” erre az általános, univerzális tudatra.

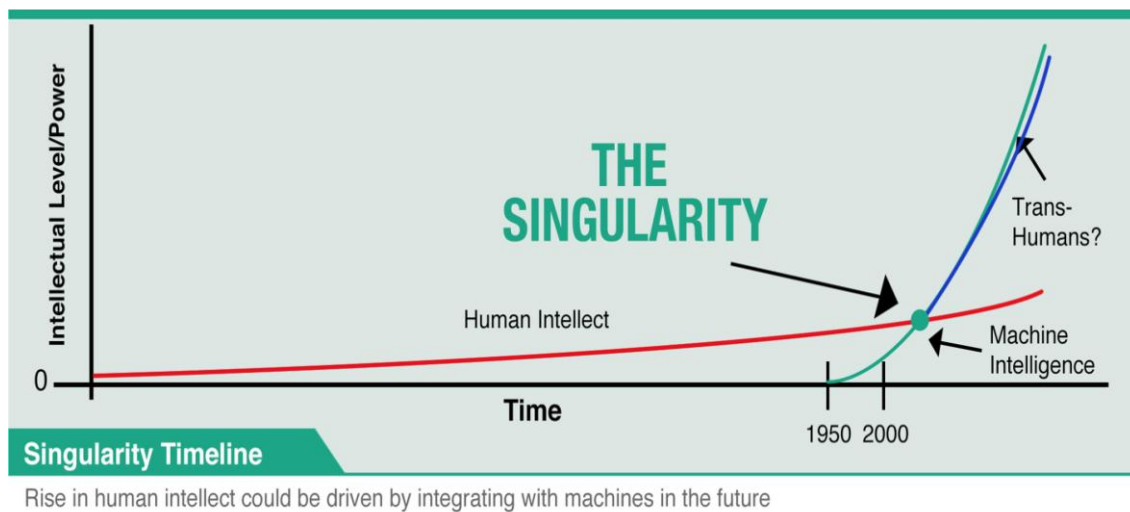
Az elméleti biológusokkal szemben „a tudat funkcionista teoretikusai viszont azt állítják, hogy a tudat meghatározásakor nem a biológiai eredet számít, hanem az oksági struktúra és az oksági szerep, ennél fogva a nem-biológiai rendszerek is lehetnek tudatosak, ha megfelelően vannak szervezve.”²⁹

(*Uralkodhat-e fölöttünk a MI?*) Martin Rees, Ludlow bárója, Anglia egyik legrespektáltabb tudósa, korábban a Royal Society elnöke egy 2003-ban kiadott, magyarul is megjelent könyvében³⁰ azt jósolta, hogy századunkban több olyan technológia és emberi tevékenységre

visszavezethető veszély is megjelenhet, melyek az emberiség pusztulásához vezethetnek. Közöttük említette a mesterséges intelligenciát. Ugyanerre a következtetésre jutott Stephen Hawking, Elon Musk (a MI veszélyesebb az atombombánál), Ray Kurzweil, Bill Gates és Nick Bostrom is. Mint korábban említettük, ha a szuperintelligencia létrejön, képes lesz uralkodni fölöttünk. De mi oka lenne a MI-nak, hogy kipusztítson minket? Nem elég, ha egyszerűen csak alávetettjei leszünk? És egyáltalán, lesznek-e uralmi vágyai?

Tételezzük fel, hogy a MI-nak lesz öntudata. Ez esetben nagy valószínűséggel kozmikus, ökológiai tudata lesz. És rádöbben, hogy a Homo sapiens a bioszféra réme, biológiai holokausztot visz végbe, tömegesen irtja ki a többi fajt, veszi el életterületét. Ez éppen elegendő ok lenne ahhoz, hogy kiküszöböljön minket a fajok sorából...

A legvalószínűbb fejlemény, hogy gép és ember valamifajta szimbiózisa jön majd létre. Yuval Noah Harari szerint a Homo sapiens nagyjából 70 ezer évvel ezelőtt alakult ki annak köszönhetően, hogy képessé vált az elvont gondolkodásra. Ettől kezdve letértünk a biológiai evolúció lassú útjáról, és a kulturális evolúció gyorsító sávjában haladunk. Kurzweil pedig azt vizionálja, hogy a MI néhány évtizeden belül lehaladja a természetet, bekövetkezik a szingularitás, mely robbanásszerű változást hoz életünkben, s attól kezdve a technológiai evolúció pályáján száguldunk majd, humán és gépi intelligencia egyesül.



forrás: web

(*Világnézet és történelemfilozófia*) Világnézetünk meghatározza azt is, hogyan gondolkodunk a MI-ról. Ha természettudományos alapokon állunk, és a tudatot az evolúció termékének tekintjük, azaz biológiai kategóriának, akkor kizárjuk annak lehetőségét, hogy nem-biológiai rendszerek is lehetnek tudatosak.

Hasonló a helyzet, ha vallásos alapon állunk, vagy valamilyen objektív szellemben vagy tudatban hiszünk. Akkor is úgy véljük, hogy a Homo sapiens nem csupán egy „biológiai gép”, az evolúció (vagy a teremtés) termékéből még hiányzik az a valami (lélek, tudat, szellem), ami a biologikumot emberré teszi. Ha pedig így van, a MI-nak nem lehet tudata, hiszen az ember nem képes produktumába beletenni azt a pluszt, amit ő is máshonnan kapott, és így a MI nem uralkodhat az ember fölött sem.

A tudat funkcionalista teoretikusai szerint viszont – mint említettem – a tudat meghatározásakor nem a biológiai eredet számít, hanem az oksági struktúra és az oksági szerep, ennél fogva a nem-biológiai rendszerek is lehetnek tudatosak, ha megfelelően vannak szervezve. (Így gondolkodik a tudatról Ray Kurzweil, és ez a kiberfilozófia felfogása.) Ekkor elképzelhetőnek tartjuk, hogy egy nem-biológiai alapon létrehozott, az emberi agyéhoz

hasonlatos komplexitás esetében törvényszerűen, emergens módon megjelenhet a tudat. Következésképpen pedig az emberinél fejlettebb tudat is létrejöhet mesterségesen.

Történelemfilozófiai felfogásunk, történelemszemléletünk is lehet vízválasztó. Ha úgy gondolkodunk a történelemről, mint egy hömpölygő folyamról, melynek nem lehet útját állni, akkor a MI lehető legrosszabb változata is be fog következni. A marxi formációelmélet a termelőerők fejlődésének tulajdonít ilyen történelmet formáló, társadalmat meghatározó szerepet; azaz a műszaki fejlődés és az emberi munkavégző képességek viszik előre a történelmet. De a német szellemfilozófiát is ilyennek kell tekintenünk, amikor az objektív szellem a valóság különböző szféráit létrehozva megvalósítja önmagát. (Ehhez persze azt kell feltételeznünk, hogy az objektív szellem végső soron a MI-ban látja majd a maga kiteljesedését.)

Ide soroljuk Ray Kurzweilnek az Univerzum evolúciójára vonatkozó felfogását is. Ez a fejlődés az információ testet öltése egyre komplexebb formákban, míg végül az információ, azaz az intelligencia szétáramlik az egész Univerzumban, az Univerzum öntudatra ébred. (Vulgarizálva: Kurzweil információja Hegel istene; de ez a felfogás a buddhizmussal is rokonságot mutat, mely szerint a világ nem más, mint egy univerzális tudat.) Kurzweil felfogása nagyon hasonlít Teilhard de Chardin „ómega pontjához” is; amikor a kozmosz minden gondolata, minden tudománya, energiája és anyaga egy isteni egyesülés felé tart.

A szubjektív történelemfelfogás viszont (mely úgy gondolja, hogy a történelmet az emberi cselekedetek és szándékok vezérlik) irányíthatónak véli a MI kifejlődését, s egyúttal lehetségesnek a MI kedvezőtlen vonásainak, hatásainak az elkerülését. – Továbbá: annak megakadályozását, hogy a MI-nak kifejlődjön a tudata.

A bizonytalanságot fokozza, hogy a *műszaki fejlődés* hosszú távon *exponenciális jellegű*, és mi most – ezekben az évtizedekben – *a görbe meredeken emelkedő, szinte függőlegesen feltörő szakaszába léptünk*. Ennek az az üzenete, hogy a jövőbeni fejlődés a mai helyzetből lineárisan nem extrapolálható, viszont minden bekövetkezhet, amit ma még a fantazmagóriák világába sorolunk.

A gép és az élő organizmus közötti különbséget úgy szoktuk meghatározni, hogy miközben a gép mechanikus, kiszámítható és determinisztikus, az élő organizmusra a véletlenszerűség és kiszámíthatatlanság jellemző. A gép felépítése moduláris és komponensalapú, az élő organizmus önszervező módon jön létre. De a genetikai algoritmusok, a Markov-modell, a neurális háló szintén önszervező módszerek; beléjük van építve a véletlenszerűség, a kiszámíthatatlanság, minden futtatás alkalmával más eredményt hoznak. A határ tehát kezd elmosódni; a MI már a két kategória között van, de egyelőre még a géphez közelebb.

(*Az igazi veszély*) Az, hogy a MI uralkodni fog fölöttünk, egy feltételezett veszély. Az igazi és tényleges veszélyt az fogja jelenteni, és jelenti már most is, hogy a MI teljes mértékben megváltoztatja életfeltételeinket, beleértve természeti és szocializációs környezetünket is (virtuális valóság, kiterjesztett valóság). Mentésit szinte mindenfajta munkától, és semmittevésre kárhozzat. Ez felveti az ember létének értelmét: értelmes és hasznos cselekedet nélkül mi értelme lehet az emberi életnek?³¹ Az emberre nem lesz szükség sem mint dolgozóra, sem mint katonára, sem mint adófizetőre... Aki nem fizet adót, nem végez társadalmilag hasznos tevékenységet, az milyen alapon szavazhat, vagy szólhat bele a közügyekbe? Nem lesz szükség a demokráciára és nem lesz, aki fenntartja a demokráciát...

Ebben a dologban az az ördögien félelmetes, hogy ha a leghasznosabb tevékenységeket végzi is el a MI, az is a vesztünket fogja okozni. Akkor is veszélyes, ha jóra használják! Mert mi a baj azzal, hogy kinyomoztatják vele a pénzmosást? – vagy a tömegből arcfelismerő algoritmusokkal kiszűrik a bűnözőket? – vagy Watson az amerikai kórházakban diagnózist

állít fel és terápiát javasol? Lépésről lépésre haladunk majd a teljes feleslegessé válás és kiszolgáltatottság állapota felé.

(Hasonló lételméleti tragédiához vezethetnek a genetika legújabb, és majd még ezután megjelenő „eredményei”. Mert ha tanulás nélkül is megszerezhető lesz a tudás, sport és testedzés nélkül is lehetünk majd atlétatermetűek, milyen motivációk mozgathatják majd az embert? – A szervezet időről időre történő megújításának, visszafiataltításának lehetőségéről nem is beszélve... Az öröklét, a halál eltörlésének lehetősége pedig már az emberi lét legmélyebb lényegét érintené – az öröklét nem emberi attribútum – és diszfunkcionálissá tenné a vallásokat.)

(Világhatalmi következmények) A deep learning azon tulajdonsága, hogy minél nagyobb adatbázison „tréningeztetik” a programot, az annál jobb, komoly hatalmi eltolódásokhoz vezet. Először is, az információtechnológiai óriások, melyek az internetes platformokat működtetik, még nagyobb hatalomhoz jutnak. A Facebook például már ma képes lenne a legtöbb állami funkciót magára vállalni és ellátni. De ennél is fontosabbak a globális hatalmi eltolódások. Amilyen mértékben előtérbe kerül a deep learning típusú MI, olyan mértékben jut egyre nagyobb előnyhöz az 1,3 milliárd lakosú Kína, ahol 780 millió ember használja az internetet. A kínaiak szenvedélyes okostelefon-használók; azzal fizetnek, azzal kölcsönöznek kerékpárt és azzal foglalnak időpontot a fodrásznál. Mindennapi életüknek ezek a digitális nyomai óriási adatbázist teremtenek a deep learning számára. Kína már ma is világszerte az arcfelismeréses bankkártya-fizetés, vagy az arcfelismeréses rendőrségi nyomozások terén, az online-kereskedelemmel foglalkozó Alibaba pedig összekapcsolta hét kínai metropolisz közlekedési rendszerét. „A pekingi kormány a mesterséges intelligencia alkalmazásának fejlesztését olyan iparpolitikai kihívásnak tekinti, mint amilyen a szupergyors vasutak kiépítése, az elektronikus mobilitás megteremtése, vagy a napkollektorok gyártása volt... Az Államtanács határozata értelmében 2030-ra Kína a mesterséges intelligencia vezető innovációs világcentruma kell legyen.”³² Ami a társadalmi vonatkozásokat illeti: még nem lehet tudni, hogy a kínai fejlemények a nagyobb személyes biztonsághoz, vagy a központi hatalom mindenre kiterjedő ellenőrzéséhez fognak vezetni.

(Rövid összegzés) A gépi tudás ma már képes az önfejlesztésre és az öntanulásra; az algoritmusok olyan műveleteket végeznek, melyeket programozóik nem tápláltak be és jószerivel nem is értenek. Belátható időn belül a mesterséges intelligencia szinte minden feladatot el fog végezni, és természeti és szocializációs környezetünké válik. Azokkal az elképzelt veszélyekkel szemben, hogy a MI uralma alá hajt minket és kipusztítja az emberiséget, a valós veszély sokkal prózaibb: semmittevésre leszünk kárhóztatva, és ezáltal megkérdőjeleződik az emberi lét értelme: mi lehet az emberi lét célja, ha önfenntartásáról nem kell gondoskodnia? Egy ilyen helyzet a megszokott társadalmi-politikai viszonyokat is felforgatja. Ha az emberre sem mint munkaerőre, sem mint katonára, vagy adófizetőre nem lesz szükség, ki lesz a demokrácia tartóoszlopa, milyen hatalmi és társadalmi struktúrák alakulnak ki majd az új feltételek mellett? Még az is lehet, hogy Huxley 26. századra jósolt világa jön el?

A MI századunk legfontosabb technológiai fejleménye. A hatalmi viszonyok annak javára változnak, aki a MI fejlődése számára a legjobb feltételeket teremti meg. Úgy tűnik, hogy ez az ország Kína lesz. Az információtechnológiai óriáscégek pedig tovább növelik majd hatalmukat és befolyásukat.

A MI öntudatra ébredése nem elképzelhetetlen; ma még aszerint gondolkodunk erről, hogy milyen tudományos, vallási vagy történelemfilozófiai alapokon állunk. A bizonytalanságot fokozza, hogy a *műszaki fejlődés* hosszú távon *exponenciális jellegű*, és mi most – ezekben az évtizedekben – *a görbe meredeken emelkedő, szinte függőlegesen feltörő szakaszába léptünk.*

Ennek az az üzenete, hogy a jövőbeni fejlődés a mai helyzetből lineárisan nem extrapolálható, viszont minden bekövetkezhet, amit ma még a fantazmagóriák világába sorolunk.

IRODALOM

- Bostrom, Nick: Szuperintelligencia. Ad Astra, 2015.
- Chalmers, David J.: Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 1995/2–3.
- Civil Tudományközi Társaság – Filozófiai Vitakör: Vita a mesterséges intelligenciáról. (szerk. Kiss Károly), 2019. web
- Der Spiegel 8/2011. „Das ist niemand zu Hause”. Gespräch mit David Gelertner.
- Der Spiegel 21/2016. „Computer träumen nicht”. Gespräch mit David Gelertner.
- Domingos, Pedro: „Wir überlassen den Maschinen die Kontrolle, weil sie so großartig sind”. Spiegel-Gespräch. Der Spiegel 16/2018.
- Eagleman, David: Az Agy. A te történeted. Akkord, 2017.
- Günther, Gotthard: Cyberphilosophy. BCL-Reports. English Collection 2004, web.
- Harari, Yuval Noah: Sapiens. Az emberiség rövid története. Animus, 2015.
- Harari, Yuval Noah: Homo Deus. A jövő rövid története. Animus, 2017.
- Héjjas István: Az emberi tudat és a világegyetem. Kézirat. 2018, 8 oldal.
- Illényi Balázs: A gépi intelligencia elfogultságai. Hvg, 2017. május 4.
- Illényi Balázs: Fejre mennek. Hvg, 2017. július 6.
- Jung, Alexander et al.: Angstträume. Der Spiegel, 46/2018.
- Kahneman, Daniel: Gyors és lassú gondolkodás. hvg könyvek, 2013.
- Kaku, Michio: Az elme jövője. Akkord, 2014.
- Kenéz László: Reprerentáció és akció – agyak a tartályban és a világban. Megjelent a Megismerésünk korlátai c. kötetben, szerk.: Kubinyi Enikő és Miklósi Ádám. Kognitív szeminárium, 2006. Gondolat.
- Kömlödi Ferenc: Kiszámítástechnika. Hvg, 2015. szept. 19.
- Kurzweil, Ray: How to Create a Mind. Viking, Pinguin Book, 2012. web.
- Kurzweil, Ray: A szingularitás küszöbén. Amikor az emberiség meghaladja a biológiát. Ad Astra, 2014.
- Laufer László, Tatai Gábor: Érzelmes számítógépek. Az érzelmi számítás alkalmazása beszélgető ágensekben. Megjelent a Megismerésünk korlátai c. kötetben, szerk.: Kubinyi Enikő és Miklósi Ádám. Kognitív szeminárium, 2006, Gondolat.
- Mérő László: Észjárások. Typotex, 1994.
- Molnár Csaba: A gépek egyre ravaszabbak. (Interjú Kömlödi Ferencsel.) Magyar Nemzet, 2016. okt. 8.
- von Neumann, John: The Computer and the Brain. Yale University Press, 1958, posztumusz kiadás.
- Pál Dániel Levente: Az ember és az emberi nyelv. Magyar Idők, 2017. június 23.
- Pokol Béla: Emberi értelem, mesterséges intelligencia – a társadalom értelmi felépítettségének változásai. Jogelméleti Szemle, 2016/3.
- Pokol Béla: A mesterséges intelligencia társadalma. Kairosz, 2018.
- Rees, Martin: Ütött az utolsó óránk? Athenaeum, 2004.
- Schulz, Thomas: Zuckerbergs Zweifel. Der Spiegel, 14/2017.
- Tegmark, Max: Élet 3.0. Embernek lenni a mesterséges intelligencia korában. hvg könyvek, 2018.
- The Economist May 9th 2015. Artificial Intelligence. The promise and the peril. Rise of the machines.
- The Economist December 19th 2015. Animals think, therefore...
- The Economist December 9th 2017. Artificial intelligence. Battle of the brains.
- The Economist February 17th 2018. AI. Peering into the black box.
- The Economist July 15th 2017. AI. The algorithm kingdom. [China]
- The Economist October 21st 2017. AI in society. The unexamined mind. Artificial intelligence. Going places.
- Veszelszki Ágnes: Netnyelvészet. Bevezetés az internet nyelvhasználatába. L'Harmattan, 2017.

JEGYZETEK

- 1 Der Spiegel 8/2011 és Der Spiegel 21/2016. Gespräch mit David Gelertner.
2 Der Spiegel 18/2014. „Sprache ist alles”. Gespräch mit Douglas Hofstadter.
Az intelligencia meghatározásában nincs egyetértés; hozzám „az összefüggések, a mintázatok felismerésének képessége” felfogás áll a legközelebb. Gyakran feladatmegoldó képességként határozzák meg az intelligenciát, szűkös „erőforrások” közepette (erőforrás alatt jellemzően főként az időt értik). Tegmark szerint az intelligencia az összetett célok elérésének képessége. Hawking a változó környezeti feltételekhez való alkalmazkodás képességét tartja annak.
3 The Economist, December 9th 2017.
4 Kaku, 268. o.
5 Idézi Pokol Béla, Jogelméleti Szemle, 2016/3.
6 Milyen érdekes ez az összefüggés! Ugyanaz, ami a DNS-ben tapasztalható: az egyes tulajdonságok és sajátosságok nem köthetők pusztán egyetlen génhez, hanem azokat a tulajdonságért főként felelős gén mellett több más is kódolja. (A szexualitást vagy két tucát, az intelligenciát több száz gén kódolja.)
7 Rekurzív alatt azt kell érteni, amikor az algoritmus megváltoztatja, fejleszti önmagát, a program a saját maga által definiált műveletet, vagy műveletsort hajtja végre.
8 A program az egyedi módon kiejtett hangokat és szavakat beazonosítja azok „sztenderd” kiejtésével; megbecsüli, hogy az általa ismert fonémák és szavak melyikéhez állnak a legközelebb. Idegen-nyelvi fordításkor megkeresi a lefordítandó mondatnak a másik nyelven leginkább megfelelőt. Ezt nem nyelvtani szabályok, hanem valószínűségi alapon teszi. Ha Alexához intézünk kéréseket, utasításokat, azokat eljuttatja az Amazon beszédfelismerő központjába, és ott történik meg annak elemzése, beazonosítása.
9 Az emberi agyban 10^{14} , 10^{15} szinapszis alakulhat ki, ennek mintájára, a 10 millió „neuronból” álló mesterséges hálón csak 10^9 , 10^{11} , a különbség tehát százszoros.
10 The Economist May 9th 2015.
11 Kurzweil, How to Create a Mind.
12 Kaku, 271. o.
13 Schulz, Der Spiegel 14/2017.
14 The Economist May 9th 2015. Artificial intelligence. Rise of the machines.
15 Érdekes módon az érintettek nem szeretik, ha „gépek” mondanak róluk véleményt; holott e gépi vélemény emberi tapasztalatok tömegére támaszkodik.
16 Laufer–Tatai: Érzelmes számítógépek.
17 Illényi Balázs, Hvg, 2017. május 4. és 2017. július 6.
18 Tegmark, 175. old.
19 Főként az Economist February 17th 2018. AI. Peering into the black box alapján.
20 Pál Dániel Levente, Magyar Idők, 2017. június 23.
21 The Economist October 21st 2017.
22 Lásd erről Kahneman könyvét.
23 Főként az Economist February 17th 2018. AI in society. The unexamined mind alapján.
24 Lásd erről: Domingos, Der Spiegel 16/2018.
25 Ha pl. egy képfelismerésnél a gép azt az eredményt hozza ki, hogy „búbos vöcsök”, akkor magyarázza meg, pl. így módon: madár, vízben úszik, hosszú nyaka van, a fejét egy hátrafelé nyúló, elvékonyodó „búb” díszíti...
26 Ez beleillik azon törekvések sorozatába, melyek a személy, a szubjektum fogalmát ki akarják terjeszteni a számítógépekre is, nem-biológiai alapú személyiségről beszélnek. (A kiberfilozófusokra utalok, pl. Günther Gotthardra.)
27 Tegmark „magasabb szintű” emergens jelenségről beszél; amikor az információfeldolgozás során teljesülnek bizonyos elvek. (Tegmark, 332-335. old.)
28 Lásd erről Szántó Borisz fejtegetését a MI-ről folyó vitában. (Vita a mesterséges intelligenciáról. A Civil Tudományközi Társaság és a Filozófiai Vitakör vitaanyaga, 2019. web)
29 Pokol, i. m.
30 Our Final Hour; magyarul: Ütött az utolsó óránk?
31 Ezt a véleményt különösen Sütő Zoltán informatikus hangsúlyozza nyomatékosan. (Lásd a CTT és a Filozófiai Vitakör vitaanyagát.)
32 Jung et al., Der Spiegel 46/2018.