
Huszonöt tudós a mesterséges intelligenciáról

„A társadalomra nem a gép jelenti a veszélyt,
hanem az, amivé az ember teszi a gépet.”

(Norbert Wiener)

(a John Brockman által szerkesztett *Possible Minds. 25 Ways of Looking at AI*
c. könyv ismertetése)¹

John Brockman New Yorkban élő kulturális-tudományos menedzser, „impresszárió”, aki időszerű és fontos témákkal foglalkozó tudományos könyveket ad ki, a világ leghíresebb tudósait szólaltatva meg bennük. Húsz éven át működtetett egy interdiszciplináris szalont, évente feltett egy-egy fontos kérdést a résztvevőknek, válaszaikat kötetben jelentette meg. Szemléltetésül néhány kérdés (melyek egyben e könyvek címei is): Mi volt az elmúlt 2000 év legnagyobb felfedezése? Hogyan látja a következő 50 évet? (Ezek magyarul is megjelentek 2000-ben, illetve 2003-ban.) Mi az, amit hiszünk, de nem tudjuk bizonyítani? Mi az Ön veszélyes gondolata? Mitől optimista Ön? Miben változott meg a véleménye? Megváltoztatja-e az internet a gondolkodásunkat? Mi az, ami miatt agódnunk kell? Mit gondoljunk a gondolkodó gépekről?

A harminc általa szerkesztett könyv mellett saját maga is írt néhányat, köztük A harmadik kultúra (*The Third Culture: Beyond the Scientific Revolution*) címűt, mely világ híressé tette. Harmadik kultúra alatt azt érti, hogy túl kell lépni a szokásos reál–humán megkülönböztetésen, a tudomány, a tudósok váltak életünk legfontosabb szereplőivé. A tudósok teszik fel a legfontosabb kérdéseket, és a tudomány területéről származnak a legfontosabb hírek. A harmadik kultúra egyben a humán és reál kultúrát összekötő hidat is jelenti. Megszűnik a szaktudományok izoláltsága, a tudósok olyan nyelven kénytelenek írni, amit a más szakmabeliek, sőt, a laikusok is megértenek.

A 25 mesterséges intelligencia (MI)-szakértő túlnyomó része amerikai, általában híres amerikai egyetemek (főként az MIT és a Harvard) professzorai. Rajtuk kívül ketten az angliai Cambridge-ből valók, egy oxfordi, egy pedig londoni. Nincs a szerzők között viszont Ray Kurzweil, a Google mérnök-zsenije, Nick Bostrom svéd származású filozófus (aki Oxfordban tevékenykedik) és Richard David Precht, a német nyelvterület legjelentősebb MI-tudósa. Max Tegmarkkal együtt (aki viszont szerepel a könyv szerzői között) én négyüket tartom a MI legjelentősebb mai tudósainak. (Kurzweil és Bostrom valószínűleg „túl nagy nevek” ahhoz a szakmában, hogy Brockman 25-ös válogatásában szerepeljenek.) A szerzői lista azonban így is impozáns, és van köztük két kulturális menedzser is, akik a MI és a művészetek, a kultúra összefüggéseit vizsgálják.

A MI kapcsán felmerülő legvitatottabb, leggyakoribb kérdések a következők. Miben különbözik a gépi intelligencia és az emberi agy; utolérheti-e a mesterséges a természetes intelligenciát? Létezik-e nem szénalapú, mesterséges tudat? A MI beilleszkedik a műszaki fejlődés történelmi vonalába, vagy egy új jelenséggel állunk szemben? Önállóosodhat-e a MI? Képesek vagyunk/leszünk-e a MI emberbarát programozására? Veszélyt jelenthet-e az emberre? Elveszi-e a munkánkat, és feleslegessé teszi-e az embert? Ezek gyakran tárgyalt témák, a Valóság 2019. decemberi számában én

magam is foglalkoztam velük.² A szerzők viszont nem térnek ki a MI alkalmazásának munkaerőpiaci hatásaira – ami egyébként a téma leggyakrabban és legszéleskörűbben tárgyalt vonatkozása.

(*Wiener az új technológiák felelőtlen alkalmazásáról*) A mesterséges intelligenciáról Brockman 2016-ban kezdeményezett beszélgetést a szakterület tudósaival. Bevezetőjében az ahhoz fűződő reményeket és félelmeket még a szokásosnál is bombasztikusabban találja: „A MI korunk sztorija; ez áll minden mai történés mögött, ez Krisztus második eljövetele, és ez az Apokalipszis.” A MI most ahhoz hasonló fogalom lett, amilyen sokáig Wiener kibernetikája volt. És mivel a kifejezés elterjedtsége és a hozzá fűződő remények és félelmek ahhoz hasonlatosak, amik az általa kovácsolt kibernetika kifejezést az '50-es években övezték, visszanyúltak Wiener 1950-ben írt, *The Human Use of Human Beings* c. könyvéhez.³ Ebben Wiener egy olyan „új fasizmust” vizionál, amikor a gépek elnyomása alá kerülünk: „... a threatening new Fascism dependent on the *machine à gouverner*”. A technika meggondolatlan, az emberi igényekre és a következményekre tekintet nélküli fejlesztése és alkalmazása oda vezethet, hogy kicsúszik az ellenőrzésünk alól. A párhuzam a mai mesterséges intelligencia miatti félelemmel tehát nyilvánvaló.⁴ Wiener korai sejtelmei és félelmei csak manapság váltak aktuálissá.

Wiener kibernetikája az információn alapuló irányításra, visszacsatolásra és kontrollra vonatkozott, és hiányzott belőle a kognitív elem. Ezt az akkortájt tartott interdiszciplináris, ún. „Macy-konferenciák” során vitték bele, nevezetesen John McCarthynak tulajdonítható, aki a „mesterséges intelligencia” kifejezés szülőatyja, és e tudományterület egyik megalapítója (Minskyvel, Rochesterrel és Shannonnal együtt). A kibernetika fogalma viszont nem tűnt el, hanem – Brockman szavaival élve – „metabolizálódott”.

Brockman a szerzőktől nem várta el, hogy reflektáljanak Wiener 1950-es kijelentésére, ennek ellenére azt szinte valamennyi esszéíró megtette, mondanivalójukat arra felépítve fejtik ki. Viszont inspirációként két dolgot ajánlott a figyelmükbe. Az egyik Wallace Stevens Zen-ihletésű verse: A feketerígó, ahogy tizenháromféleképpen szemléljük. A másik az a példabeszéd, amikor vak embereket arra kérnek, hogy tapogassák meg az elefántot, majd mondják el, vajon mi lehet az. Brockman emlékeztet Marshall McLuhan kommunikáció-filozófus megállapítására is, miszerint az információtechnológiák révén „externalizáltuk” központi idegrendszerünket, elménket, és ezáltal egy egyetlen, óriási, közös emberi elmét hoztunk létre. Brockmant, a kultúra-menedzsert az a felismerés is köti az informatikához és a kommunikációhoz, hogy az új technológiák egyben új érzékelést, világfelfogást is jelentenek, és az új technológiai eszközök használata által újraformáljuk önmagunkat.

(*Mi és ők/azok*) George M. Church a génebésszet, a „genetic engineering” legnagyobb alakja, tevékenysége egyformán kötődik a Harvardhoz és az MIT-hoz is. Ő úgy gondolodik, hogy végső soron minden dolgot és szervezetet legkisebb alkatrészeire, atomjaira lehet bontani, és ennek megfelelően, alulról felfelé haladva, létrehozhatunk bármilyen kombinációjú szervezetet az élettelenről az élőig, az oktalantól az intelligensig terjedő skálán, beleértve a digitális technika és a robotika által alkotott újabb és újabb szerkezeteket. Ezért azzal kell számolni, hogy a jövőben sokféle létező, és sokféle elme vesz majd minket körül. A fő probléma nem az emberi és egy mesterséges intelligencia viszonya lesz, hanem az, hogy hogyan viszonyuljunk a legkülönfélébb intelligenciákhoz, és milyen jogokat adjunk a sokféle létezőnek. Amikor Jefferson 1776-ban leírta, hogy minden embernek elidegeníthetetlen joga van az élethez, a szabadsághoz és a

boldogság kereséséhez, a „minden ember”-be még nem értette bele a nőket, a rabszolgákat és a színes bőrű indiánokat. Ma már napirenden van, hogy az emberszabású majmok és a delfinek is részesedjenek e jogokban. Church szerint e fejlődés logikus és elkerülhetetlen szakasza lesz, amikor a hibrid szervezetekre is kiterjednek e jogok.

Ma még erkölcsi és jogi tiltások akadályozzák, hogy e mesterséges sokféleség kibontakozzon. Ezek azonban egyre inkább tarthatatlanok lesznek, a határok ledőlnek, a gátak átszakadnak. Gyakran, és főként elítélőleg használják ma a transzhumán, transzhumanista szavakat, azok, akik a régi értékrend és emberfelfogás alapján állnak. De Church e téren is megengedőbb, demisztifikálja e kifejezéseket. Az ember az általa létrehozott technikával a leggyorsabb élőlény lett. Egy amazóniai őslakos szemében vajon nem transzhumán az a nyugati ember, aki a Föld vonzását leküzdve képes a világűrbe kitörni, és ott akár hosszabb ideig is megteremteni az életfeltételeit?³

Egy ilyen soklétezős szituációban másképpen merül fel a robotok, a MI etikájának kérdése is; hogy hogyan viszonyuljanak az emberhez, milyen erkölcsi elveket tápláljunk beléjük. Church oda lyukad ki, hogy az emberközpontú etikával szemben a darwini evolúciós logika is érvényesülhet, amikor az erkölcs nem humánus, demokratikus szavazás és belátás kérdése, hanem a „győzőn az életképesebb” elv érvényesül. A Church által vizionált jövőben a felelősség „delegálásának”, megosztásának kérdése is másként vetődik fel. És különben is; felelősségünk nagy részét már régen a nagyvállalatokra, az állami szervezetekre, a kormányokra, a hadseregre, a különféle hivatalokra és laboratóriumokra bíztuk. Az önvezető járművek kapcsán ma lépten-nyomon azzal a divatos fejtegetéssel találkozunk, hogy ki a felelős az általuk okozott kárért vagy balesetért. Ez azonban ugyanaz a kérdés, hogy ki a felelős egy nagyvállalat vagy hivatal hibás döntéseiért – állítja Church. És lássuk be, jogosan.

Mindezzel szemben áll a megrögzött materialista (egy kicsit árnyaltabban és szakszerűbben: komputacionista) hírében álló Daniel C. Dennett (Tufts Egyetem) leegyszerűsítő kijelentése: „Nem mesterséges tudattal rendelkező ágensre, hanem intelligens eszközökre van szükségünk.” És hogy véleménye felől a „technofiliasoknak” ne maradjon kétsége, még ezzel is megtoldja: „A MI jelenlegi formájában a humán intelligencia élősködője.”

(Mi következik a tudat emergens jelenségként való felfogásából?) Francis Crick – aki James Watsonnal együtt először írta le a DNS kettős spirál-formáját – a tudatot emergens jelenségnek, az agy termékének tekintette, e tételt pedig „meglepő hipotézisnek” nevezte. Ez lett a mai modern idegtudományok alaptétele. A klinikai neurológusok és neuropszichológusok azóta (1994 óta) sem találtak semmi olyan jelet, mely arra utalna, hogy a tudati működésnek a neuronok hálózatán kívül lenne más, nem anyagi jellegű összetevője, vagy alapja. Frank Wilczek, az MIT Nobel-díjas fizikaprofesszora szerint e tételből három fontos állítás következik: (1) a MI-nak is lehet tudata, (2) a MI szintén lehet kreatív, (3) a MI gonosz is lehet. Ha ugyanis a tudat emergens jelenség (azaz kialakulásában nem játszanak közre metafizikai, transzcendens elemek), akkor – legalább is elméletileg – mesterségesen is létre lehet hozni az atomok olyan kombinációját, mely azonos az aggyal, és amelyből majd – a komplexitás kellően magas fokán – emergens módon létrejön a tudat. Sőt, e logikai szálát tovább fonva Wilczek arra a következtetésre jut, hogy az intelligencia általános formája valójában maga a MI, az emberi intelligencia pedig annak egy specifikus megnyilvánulása.

Ez az érvelés egybevág Max Tegmark azon tételével, hogy az intelligencia közegfüggetlen, független az azt hordozó, közvetítő anyagtól, azaz nemcsak szerves-anyag-alapú, szén-alapú lehet, hanem bármilyen, így szilícium-alapon is létrejöhet. (Tegmark e tétele kiterjed a számításra és az információra is.)⁶

(A fordulópont, az igazi veszély az ÁMI) Általános mesterséges intelligenciának (ÁMI) nevezik azt, amikor a MI minden területen eléri, majd meghaladja az emberi intelligenciát. A tudományterület tudósainak elképzelései szerint ez száz éven belül várható, legtöbbben néhány évtized távlatába, 2040-2050 környékére teszik. Ezt a határt különösen Tegmark tekinti létfontosságúnak. Elsősorban nem azért, mert ezt követően a Homo sapiens bármilyen kognitív tevékenység terén helyettesíthető lesz. Hanem azért, mert ezt követően az ÁMI – *per definitionem* – okosabb lesz nálunk. Tehát nem fogjuk tudni érvényesíteni az akaratumkat fölötte, túljár az eszünkön, az lesz, amit ő akar. Tegmark az *AI safety*, a biztonságos MI megalkotásának (illetve a MI kordában tartásának) vezéralakja, ilyen témájú nemzetközi konferenciák és társaságok szervezője. Nyomatékosan hirdeti, hogy még az ÁMI kialakulása/megalkotása előtt ki kell dolgozni azokat az elveket és szabályokat, amelyek a MI használatát biztonságossá teszik. (A megjelenését követően ugyanis erre már nem lesz alkalmunk, az ÁMI önállósítja magát, a saját elképzelései szerint fog cselekedni.) A Tegmark által javasolt elvek és szabályok főként a MI fejlesztésének egyes irányaira és alkalmazási területeire vonatkoznak.

A helyzet kísértetiesen hasonlít az atomfizika terén történetekre. Rutherford, az atomfizika kezdetének legnagyobb tudósa 1933. szeptember 11-én kijelentette, hogy az atomokból kinyert energia holdkóros ötlet. Másnap, szeptember 12-én Szilárd Leó közzétette a neutronokkal indukált nukleáris láncreakció ötletét. De itt most sokkal veszélyesebb jelenséggel állunk szemben. Az atomfegyverek gyártása és használata egy egyszerű, igen-nem döntés kérdése. A MI intelligencia esetében viszont a veszély szinte megfoghatatlan, kúszó, folyamatos jelenség. Nehéz előre megmondani, hogy melyik irányú fejlesztést kellene betiltani. Tekintettel az előre jelzett 2040-2050-es időpontra, Tegmark azt javasolja, hogy a rendelkezésünkre álló két-három évtizedet arra kell felhasználni, hogy intenzíven kutadjuk, milyen irányt vehetnek a fejlemények és hol szabjunk meg korlátokat. Ez igen tág kör, beletartoznak az erkölcsi értékek és a nemzetközi együttműködés területe is.

(Amikor a kigyó a saját farkába harap) A gépi tanulás legsikeresebb módja a neuronhálózatokat utánzó deep learning. Ennek során a rendszer adatok hatalmas halmazában keresi az összefüggéseket, függvényt rajzol fel belőlük, és minden további információval tökéletesíti azt. Minél több információt, képet táplálnak be adatbázisába pl. a kutyáról, annál pontosabban fogja tudni azonosítani azt. Judea Pearl, az UCLA (Kaliforniai Egyetem) komputer-tudósa a deep learninget a végtelenül lassú darwini szelekcióhoz hasonlítja, amikor pl. a kigyó, vagy a sas látása évmilliók során tökéletesedett.⁷ Ezzel szemben az embernek csak évszázadokra, legfeljebb évezredekre volt szüksége ahhoz, hogy látását nagyítóval, szemüveggel, majd idővel mikroszkóppal tökéletesítse. Ez az emberi tulajdonság ugyanis nem a darwini evolúció eredménye; hanem annak a következménye, hogy a Homo sapiens képes „a valóság mentális reprezentációjára”.⁸ (Ezt nevezzük tudatnak; számomra Michio Kaku tudat-meghatározása a legszimpatikusabb, miszerint a tudat a jövő szimulációjának a képessége.)⁹ Az ember tehát mentálisan reprezentálja a valóságot, és elképzeléseit, változtatásait, kísérleteit nem a hosszú időskálával „dolgozó” evolúció igazolja, vagy veti el, hanem azok életképességéről rövid idő

alatt megbizonyosodik. Sőt, hipotéziseket és ellenhipotéziseket állít fel és teszti azokat, így igazolva vagy elvetve elképzelései jogosságát, életrevalóságát. (Sőt, még fantáziáját is segítségül hívja.) A deep learning azonban nem képes a „Mi van akkor, ha...?” jellegű kérdések feltételére, mert nem képes „a valóság mentális reprezentációjára”, azaz nincs tudata. Hm... a kigyó a saját farkába harapott. Az ÁMI, az általános mesterséges intelligencia – meghatározása szerint – minden vonatkozásban fejlettebb lesz az emberinél, és ennél fogva már tudata is lesz. De a MI-nak tudata csak akkor lehet, ha eléri az emberi intelligencia szintjét; amihez a valóság mentális reprezentációjának képességére lenne szükség. – Hadd legyen e dilemma feloldásának feladata azoké, akik 2040-2050-re jósolják az ÁMI eljövételét.

Valójában itt a tudat emergens megjelenésében kellene bízunk. És miközben számtalan bonyolult jelenséget ezzel magyarázunk, azért a kételkedés megmarad. Jól példázza ezt Harari hasonlata (miközben ő is a tudat emergens felfogásának híve). London közlekedését veszi alapul, mely valóban egy szinte végtelenül sokszereplős, rendkívül komplex rendszer. De vajon elképzelhetőnek tartjuk-e, hogy csúcsidőben, a Piccadilly Circus fölött lebegve, emergens módon megjelenik a londoni közlekedés tudata és azt mondja: Ejnye, de be vagyok dugulva!¹⁰

(*A Teremtés, mint kereső algoritmus*) Képesek vagyunk gondolatok és intelligencia automatikus előállítására – ez a MI. De tudunk-e anyagot is automatikusan (azaz digitális technikával) előállítani? Ez a kérdés foglalkoztatja Neil Gershenfeldet, az MIT Bitek és Atomok Központjának igazgatóját. Az ember, ha egy szerkezetet eszkábál, először az alkatrészeket készíti el, majd összeszereli az egészet. Micsoda primitív módszer! Az evolúció ezzel szemben úgy működik, mint egy keresőgép: az összes lehetséges változatból kiválasztja az adott körülményekre szabott optimálisat. (Bár igaz, az idővel nem spórol.) Az ember azonban – bár tudatlanul – elsajátította ezt a módszert, amikor megalkotta a MI-t. A MI sem alkatrészekből hozza létre az optimális változatot, hanem a lehetséges változatok nagy számából kiválasztja a megfelelőt.

Ez a módszer terjedőben van a mérnöki tervezésben, és „deklaratív”, vagy „generatív” dizájnnak nevezik, szemben az eddigi „imperatív” dizájnnal. A mérnök nem explicit módon tervezi meg a gépet, hanem betáplálja a MI-ba, hogy az milyen kívánt tulajdonságokkal és képességekkel rendelkezzen, és a MI azt létrehozza. (Hiszen a lehetséges gépek száma végtelen, s ebben ott található a kívánt variáns is.) A digitális és a fizikális határterületén vagyunk, a bitek és atomok viszonyának világában. Félünk, mi lesz, ha megalkotjuk az ÁMI-t. Pedig jobban kellene tartanunk attól az automatától, amely tettszése szerinti gépeket hoz majd létre, figyelmeztet Gershenfeld.

(*Célok és értékek*) Minél intelligensebb lesz a gép, annál fontosabb, hogy a kellő értékekkel legyen beprogramozva. Mert az önvezető autó magától nem tudja, hogy a gyalogosok nem szeretik, ha elütik őket. Ha az ÁMI azt a feladatot kapja, hogy találja meg a rák ellenszerét, előfordulhat, hogy a világ teljes népességén kísérletezik, tengerimalacként használva az emberiséget. Vagy ha a világ óceánjainak savasodását kell megállítania, esetleg a légkör oxigénjét fogja erre felhasználni. Ezek Stuart Russel, a Berkeley komputertudósának példái, de e téma klasszikus, sokat idézett példája Bostrom Szuperintelligencia című könyvéből való; amikor a gemkapcsok gyártására befogott szuperintelligencia a Föld összes erőforrását erre használja fel.

Az értékek kérdése igen sikamlós terület. Gondoljunk csak a politikai értékek ellentmondásos változatosságára. Vagy az egyes egyének értékskálájára. Russel az egyéni értékek meghatározásának problémájára hívja fel a figyelmet. Ha elfogadjuk a feltételezést, hogy

elménk egyes alrendszerének preferenciái inkompatibilisek egymással, a vágyaink, óhajaink kielégítésére törekvő ÁMI megoldhatatlan feladattal találja szemben magát. De a robotok értékekre való tanításába besegít majd a piac is. Mert ha pl. egy háztartási robot ebédre megfőzi a macskát – nem ismervé fel, hogy annak „érzelmi értéke” a háziak számára sokkal nagyobb, mint a „tápértéke” – gyártója súlyos piaci veszteséget fog elszenvedni. (A célok és értékek témájáról Tegmark Élet 3.0 című könyvében találhatunk egy rendkívül izgalmas fejezetet.)

Megítélésem szerint ez lesz a MI-val, pontosabban az ÁMI-val kapcsolatos legsúlyosabb probléma: milyen célokat és értékeket tápláljunk a gépbe. (De a probléma majd azután is fennáll majd: az autonóm gép tetszése szerinti értékeket fog vallani.) A deep learning már ma is ad egyfajta megoldást: az algoritmust a napi sajtón és a világirodalom legismertebb, legbecsesebb művein tréningeztetve előáll egyfajta értékeverék... De aligha hihető, hogy ez a mix jó alapul szolgál majd az emberiség sorsát a kezében tartó ÁMI számára. A téma újabb buktatója, hogy az emberi viselkedés és gondolkodás jelentős mértékben irracionális. – A feladat megoldhatatlannak tűnik: ki határozza majd meg ezeket az értékeket? Az ENSZ? Vagy a pápa? Vagy a Nobel-békedíjasok? Vagy Nobel-díjas írók könyvein kellene tréningeztetni az ÁMI-t? Ide kívánczok egy Dosztojevskij-idézet A Karamazov testvérekől, mely Tegmark hivatkozott könyv-fejezetének mottója: „Az emberi lét titka nem abban rejlik, hogy hogyan élünk, hanem abban, hogy miért élünk.”

(Intelligencia és dominancia) Hawking, Musk, Gates, Tegmark, a szakma nagyjai szinte egyöntetűen vallják, hogy idővel eljön majd az ÁMI fölény-kerekedésének veszélye. Ezzel szemben viszont Steven Pinker, a Harvard pszichológus-professzora nem tart ettől. Elveti a determinista történelemfelfogást, az eszmék és az érvelés erejében hisz. Az ÁMI-nak tulajdonított veszélyeket a tudományos alapok nélküli, ótestamentumi átkozódásokhoz és jeremiádákhoz hasonlítja. Azt vallja, hogy a katasztrófa-szenáriók mindig könnyen utat találnak a tömegekhez. A jellegzetes MI disztópiák pedig „a szűklátókörű alfa-hím pszichológia kivetítődései az intelligencia koncepciójára.” (Azaz: az intelligensebb lény hatalmát mások meghódítására használja fel.) Figyelemre méltó tétele, hogy a rettegők téves platformon állnak: azt hiszik, hogy az intelligenciából következik a dominancia vágya. Azonban inkább ennek az ellenkezője valószínű: a szuperintelligencia jóhiszemű lesz. Én azonban könnyen el tudok képzelni olyan esetet, amikor ez nem így van. Az ember a bioszféra réme, folyamatosan veszi el a többi élőlény élethelyét, tevékenysége következtében ezerszámra pusztulnak ki a fajok. Ha a szuperintelligenciának lesz tudata, az minden bizonnyal kozmikus, ökológiai tudat lesz, és akkor a Homo sapiens ellen fordulhat. De valószínűbb, hogy olyan megoldásokra és gyakorlatra tanít és kényszerít majd minket, ami tevékenységünket környezetbaráttá teszi.

Pinker mélységesen hisz a nyugati társadalmakban és a demokráciában. Felállítja azt a tételt, hogy ha a műszaki fejlődés elnyomásunkhoz és alávetettségünkhez vezetne, akkor az elmúlt évtizedek információtechnológiai forradalma következtében ennek már lennének negatív jelei, de szerinte nincsenek. Én azt hiszem, ez nehezen tartható álláspont. Gondoljunk arra, hogy az internetes platformok (de nemcsak a legnagyobbak, hanem a helyi online-kereskedők is) mi mindent tudnak rólunk, mennyi adatot gyűjtenek, és milyen módon manipulálják ízlésünket, életmódunkat, sőt, gondolkodásunkat és döntéseinket is. És amilyen mértékben az olvasás egyre inkább átkerül a könyvekről és az újságokról az internetes e-bookokra és hírportálokra, olyan mértékben alakul majd át a folyamat: nemcsak mi olvassuk majd a könyveket és híreket, hanem azok

minket. Orwell Nagy Testvére nem hinne a szemének. És gondoljunk arra is, hogy a nagy internetes és információtechnológiai vállalatok tevékenységének az ellenőrzése és adóztatása mennyi problémát jelent Amerikának és az EU-nak is.

Stephen Wolfram a programozás leghíresebb tudósa. A MI legnagyobb veszélyét abban látja, ha megkapjuk általa a halhatatlanságot. És valóban, minden, ami bennünk emberi, minden törekvésünk értelmét vesztené; életünknek a halál ad értelmet.¹¹

(Ami a matton túl van) A sakk-automaták arra vannak programozva, hogy az adott helyzetben mindig azt a lépést válasszák, amely a legcélszerűbb a győzelemhez. Ily módon azonban csak MI-t lehet létrehozni. Ez a „tanítási” módszer azonos azzal, amely az emberiség hosszú történelmét jellemezte: a tanítás célja a konformitás és engedelmesség volt, az eretnekeket, másként gondolkodókat súlyosan megbüntették, megégették, a kreativitást nem honorálták. A tudományok és a műszaki fejlődés felgyorsulásához a felvilágosodásra volt szükség. A sakk-automatából akkor lesz ÁMI, ha hagyják, hogy kedve szerint játsszon: kipróbáljon egy új megoldást, vagy egy mások által alkalmazott lépést, ha élvezheti a játékot. A kérdés egy széles témába torkollik: mi az, amit az ÁMI-k majd megtehetnek, és mi az, amit nem. Büntetéssel és korlátozással, vagy jutalmazással és ösztönzéssel kell irányítani őket? És mi legyen a jutalmuk? Vagy a halhatatlan ÁMI-k büntetése?

Az ÁMI-k viselkedésének szabályozására irányuló törekvések mögött az a tapasztalat áll, hogy a rebellis ifjúság eltávolodik a társadalom kulturális és erkölcsi értékeitől, ezért jutalmazásra és büntetésre van szükség. A nyitott társadalomban élő emberek viselkedését azonban nem a külső feltételek szabályozzák; ők maguk választják meg, hogy milyen büntetésben vagy jutalomban részesüljenek. Az ÁMI-k programozása, „nevelése” így az adott társadalmi körülményeknek és feltételeknek a függvénye.

Alaptalan az a félelem is, hogy az ÁMI-k okosabbak lesznek nálunk. Létrehozásuk során a mi gondolkodásunk is fejlődik. Az emberi agy bármit el fog tudni gondolni, amit az ÁMI, a különbség csak a gyorsaságban lesz, az a hátrány viszont majd külső eszközök segítségével behozható lesz.

Bármennyire is meglepő, ezek David Deutsch gondolatai, aki a kvantumkomputerek előállításának legnagyobb tudósa (Oxfordi Egyetem). A meglepő az, hogy fejtegetése nem a kvantum-összefonódásról, a kvantumbizonytalanságról, és a részecskék kettős természetéről szól, valamint a kvantumkomputer előállításának nehézségeiről, hanem az ember és a gép kognitív jellemzőinek hasonlóságáról. Tétele: ha az emberhez hasonló, sőt, azt fölülmúló mesterséges intelligenciát akarunk alkotni, akkor ahhoz olyan társadalmi feltételeket kell teremteni, melyek az emberi intelligencia kifejlődéséhez is szükségesek.

(Gyermeki tanulás és MI) Az univerzum legjobb tanulója a gyerekek – állítja Alison Gopnik, a Berkeley nemzetközileg ismert gyermekpszichológusa, fejlesztés-kutatója. A MI-nak alapvetően két formája van, az egyik az alulról építkező deep learning, amikor az algoritmus egy mesterséges neuronháló segítségével adathalmazból keresi ki az összefüggéseket. Ide tartozik a „reinforcement learning”, a „megerősítő” vagy jutalmazó tanulás is. Ekkor egy függvény jutalmazza, vagy bünteti a MI-t, attól függően, hogy közelebb kerül, vagy távolodik a feladat megoldásától. A másik a főlülről lefelé irányuló tanulás, melyet bayes-i valószínűségi hálók beiktatásával végeznek: nem a MI-nak kell megalkotnia az összefüggéseket leíró függvényt, hanem a vizsgált adatokat megadott valószínűségi függvényekhez kell viszonyítania.

A gyermeki tanulás csodálatraméltó folyamat: mindkét fenti módszert alkalmazza, valójában azok legjobb tulajdonságait ötvözi, és aztán fölülmúlja azt. Mindmáig rejtély, hogyan fejlődik ki négyéves gyermekekben a kreativitás. Érdeemes a gyermeki tanulás elemeit megpróbálni felhasználni a gépi tanulásban: kíváncsiságot vinni be a gépbe és lehetővé tenni, hogy a gép aktívan interakcióba lépjen a külvilággal, mint ahogy az a gyermek tanulásának is fontos feltétele.

(Robotgondok és az emberi tényező) Mind a gépi, mind a nem gépi robotok fejlesztésének következő fontos lépése az lenne, hogy a gépek is ismerjenek minket, jelenjen meg bennük az emberi tényező. Fizikai robotok esetében egy egyszerű dologról van szó: a robot érzékeli a saját és mások térbeli elhelyezkedését, és ha pl. elindulok egy bizonyos irányba, ne legyen utamban. Ez a koordinációs feladat. A nem-fizikai robotok esetében már sokkal komplikáltabb igények jelentkeznek. Ha pl. megbízást adok az algoritmusnak tőzsdei tranzakciók végrehajtására, és közben, vagy ezt követően változik a pénzügyi helyzet, ismerje szándékaimat, és tudja, hogy a megváltozott feltételek közepette én hogyan cselekednék. E bonyolult interaktív feladatot az értékek összehangolásának (value alignment) nevezik. A robotot a feladat elvégzésére nem utasításos algoritmusokkal, hanem egy ún. „jutalmazási függvénnyel” veszik rá, ill. ösztönzik. Jó pontokat kap, ha helyesen cselekszik, és pontokat vonnak le tőle, ha hibázik, vagy a feladatot nem a legjobban látja el. (Erről szól Anca Dragan esszéje, aki a Berkeley komputer-tudósa.)

De miközben azon a feladaton töprengünk, hogy hogyan vigyük be az emberi tényezőt a MI egyenletébe, saját magunkkal sem vagyunk sok vonatkozásban tisztában. Az emberi döntéshozatal tipikus fekete doboz, nem tudjuk, hogyan születnek a döntések érzelmeink és a racionális megfontolások keverékéből. Az „ember-modell” egyik legfontosabb sajátossága a korlátozott optimalitás (bounded optimality); a korlátozott racionalitásból, vagy adathiányból következően a döntésünk sem lesz a legjobb, nem lesz optimális. Az embert utánzó mesterséges intelligencia kifejlesztésén dolgozunk, miközben nem ismerjük jól magunkat, a mintát. Pedig a MI megalkotásához még mindig az emberi minta a legalkalmasabb. A robotok tanításához, ösztönzéséhez is azt kell megértenünk, hogy hogyan tanulnak az emberek, mi ösztönzi őket. Ez az „inverz tanulási ösztönzés” (inverse reinforcement learning) témája. Ezeket a kérdéseket tárgyalja Tom Griffiths, a kognitív tudományok professzora a Princeton egyetemen.

A deep learning során találkozunk a „lokális minimumok” problémájával; amikor az összefüggéseket felrajzoló függvény értelmezési tartománya nem eléggé széleskörű (azaz korlátozott adatbázison nyugszik), a megoldás is korlátozott, „lokális” érvényű lesz. Általánosabb értelemben az emberi gondolkodást is a lokális minimumok jellemzik, de – Chris Anderson lelkes véleménye szerint – az AI az elkövetkező évtizedekben meghozza nekünk azt a lehetőséget, hogy a lokális minimumokat mélyebb, majd pedig globális minimumokra változtassuk, azaz univerzálisan gondolkodhassunk. (Anderson drónépítő vállalkozó, korábban a híres Wired tudományos magazin főszerkesztője volt.)

(Lesz-e szingularitás?) A mai technikai előrejelzéseknek az az alapja, hogy a műszaki fejlődés gyorsuló ütemben nő. Kurzweil emellett hosszasan, és számtalan adatot felvontatva érvel A szingularitás küszöbén című könyvében, és szinte mindennapos tapasztalataink is ezt támasztják alá. A szingularitás hívei szerint ez oda vezet (és maga a fogalom azt jelenti), hogy a gépi intelligencia utoléri, majd meghaladja az emberit, a kettő összeolvad, és attól kezdve az emberiségre (vagy az így létrejövő transzhumán lényekre) fantasztikus perspektíva vár. (Ezt valójában már Neumann János előre jelezte az ötvenes

évek elején, és maga a szingularitás kifejezés is tőle származik.) Seth Lloyd, az MIT elméletifizika-professzora ezt kétségbe vonja. Szerinte a műszaki fejlődés előrejelzése sok bizonytalanságba ütközik és nincsenek örökké tartó exponenciális folyamatok. A műszaki előrejelzés a jóslás műfajába tartozik. Az ötvenes években elképzelt robot-takarítónők fantáziának bizonyultak (az automata porszívók még mindig fennakadnak egy-egy szőnyegcsomón), és a robotok még mindig nem képesek megkötni a cipőfűzőjüket. A deep learning jól képes utánozni egy agyi funkciót, az agykéreg mintázatfelismerő tevékenységét. De hol van ez még az agy számos komponensének és egészének számtalan és végtelenül bonyolult funkciójától! A komputerek „nyers” információfeldolgozó képessége brutálisan nőtt – de az információfeldolgozás minősége, finomsága alig halad előre. (A számítógépek információfeldolgozó képessége két évtizeden át exponenciálisan, két évente megkettőződve nőtt, de az utóbbi időben valóban elakadt. Kurzweil azonban további olyan lehetőségeket lát, amelyek ezt az ütemet fenn fogják tartani.)¹²

Seth Lloyd egy másik érdekes jelenségre is felhívja a figyelmet. Amióta a programok képesek tanulni, és önmagukat tanítani, tökéletesíteni, a tudósok attól félnek, hogy a gépi tanulás egyre követhetlenebb lesz. Azonban kiderült, hogy az algoritmusok egyre inkább *emberi* módon tanulnak és az emberi agyat és gondolkodást utánozzák.

(*Hibrid intelligenciák*) W. Daniel Hillis, a Dél-Kaliforniai Egyetem komputer-tudósa ugyanarra a jelenségre hívja fel a figyelmet, mint amit korábban George Church-nél említettünk: félelmünk, hogy a MI-val, és különösen a szuperintelligenciával szemben alulmaradunk, nem új keletű. Ténylegesen ez történik az állammal és a nagyvállalatokkal szemben, melyeket Hillis „hibrid intelligenciáknak” nevez, mivel saját érdekeik, és emergens módon megjelenő céljaik vannak. Az analógia nyilvánvaló, és visszamegy Wienerig, aki „húsból és vérből álló” intelligens gépekről beszélt. A helyzet differenciált vizsgálata érdekében Hillis négyféle viszonyrendszert különböztet meg a gépi és a hibrid szuperintelligenciák (államok, vállalatok, hivatalok) között:

1. A gépi szuperintelligenciát az egyes nemzetállamok (azaz a hibrid szuperintelligenciák) birtokolják és ellenőrzik. (Ez inkább csak elméleti lehetőség, a globális világban már a hagyományos nemzetközi nagyvállalatok is kinőnek a nemzetállamok ellenőrzése alól – bár Kínában ez van megvalósulóban.)

2. A gépi szuperintelligencia a profitelven működő nagyvállalatok kezében van. (A világ efelé halad, a MI-t a nagy internetes platformok és az információtechnológiai nagyvállalatok birtokolják és működtetik.)¹³

3. A szuperintelligencia önállósodik és szembefordul velünk – ettől félünk a leginkább.

4. A szuperintelligencia önállósodik, de önkéntesen az emberiség javát szolgálja – ez az optimista forgatókönyv (ebben bízunk Steven Pinker).

(*Emberi agy és gépi intelligencia*) A számítógépek bővületében élünk. Lépten-nyomon azzal példálózunk, hogy mi mindenre képesek, hogy verte meg a Deep Blue 1997-ben Kaszparovot sakkban, hogy győzte le Watson a Jeopardy amerikai műveltségi vetélkedő nyerteseit 2011-ben, hogy szegényítette meg az AlphaGo Zero a go ázsiai bajnokát, Lee Sedolt 2017-ben. A MI szinte az élet, a tudományok minden területén csodálatos teljesítményekre képes. Seth Lloyd azonban felhívja a figyelmet, hogy nem egy gép verte meg a sakk vagy a go bajnokait; hanem a gépet megalkotó tudósok és a gép együtt.

George Church konkrét számokkal érvel az emberi agy mellett. Az említett műveltségi vetélkedő során az IBM Watsona 85.000 watt energiát fogyasztott (pontosabban ennyi volt a teljesítménye), versenytársai csak fejenként (azaz „agyanként”) 20-at. Egy felnőtt

emberi agy létrejöttéhez 6 billió joule energiára van szükség, ugyanannyira, mint amibe Watson előállítása került. A komputer számolásban hihetetlenül gyors, adattárolásban (emlékezetben) verhetetlen, de távolról sem képes arra a differenciált agyműködésre, ami az emberi gondolkodást jellemzi. Mire hát ez a nagy felhajtás?

Fred Wilczek mérlege is az emberi agy felé billen. Az információfeldolgozó technológiák hat területen biztosan előnyöket mutatnak az emberi aggyal szemben: a gyorsaságban, a méretekben, a stabilitásban, az élettartamban, a modularításban (a gépek architektúrája nyitott), és abban, hogy alkalmazási lehetőséget nyújtanak a majdani kvantumszámításnak. Ezzel szemben az emberi agy öt olyan vonatkozásban mutat fel egyértelmű előnyöket, melyek az előző szempontokhoz képest nagyobb jelentőségűek. A gépek processzorainak elhelyezése kétdimenziós, a háromdimenziós megoldás nehézségekbe ütközik, ezzel szemben az agy jól kihasználja a három dimenziót. Az agynak igen nagyfokú az önjavító képessége. A konnektivitás, a neuronok kapcsolódása tekintetében az agy fölülmúlhatatlan. Az agyat nem kell létrehozni, az biológiailag kifejleszteti önmagát, és – különösen gyermekkorban – minél intenzívebb az egyén kapcsolata a külvilággal, annál fejlettebb lesz az agya. Végül pedig nagy az előnye az „integráltság” terén. Az agyhoz természetesen módon kapcsolódnak az érzékszervek („szenzorok”), valamint a mozgató testrészek („aktuátorok”). A MI alkalmazását főként gazdasági szempontok teszik előnyösebbé. Az intelligencia távlatait tekintve azonban látnunk kell, hogy az emberi agy géppel szembeni előnyei egyre csökkennek, a gép előnyei viszont növekednek.

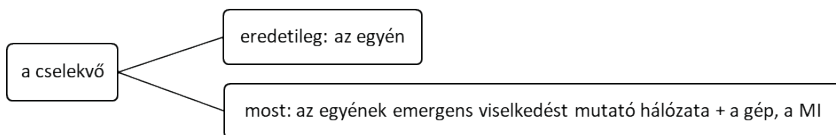
(Újra divat az analóg? A kontroll elvesztése) George Dyson tudománytörténész rövid, ötoldalas írása a számítás (komputáció) jövőjéről szól. Fő tétele az, hogy az analóg számítás nem halt ki, sőt, már visszatérőben van. „Miközben digitális komponenseket használunk, a rendszer által végzett analóg számítás egy bizonyos ponton messze meghaladja annak a digitális kódnak a komplexitását, amellyel azt létrehozták.” „Az analóg kontroll rendszerrel működő igazi mesterséges intelligencia egy digitális szubsztrátumból fog kiemelkedni ahhoz hasonlóan, ahogy a digitális számítógépek megszülettek a második világháború után az analóg komponensekből. Ez hamarabb meg fog történni, mint ahogy hinnénk.” Törekvéseink a visszajára fognak fordulni, mert „miközben olyan gépeket igyekszünk építeni, amelyek mindent képesek kontrollálni, elvesztjük az ellenőrzést ezen gépek fölött”.

Ugyanakkor Dyson felvet egy súlyos problémát: az emberi megértés korlátait. A MI harmadik törvénye úgy szól, hogy „ha egy rendszer elég egyszerű ahhoz, hogy megértsük a működését, az nem eléggé összetett ahhoz, hogy intelligensen viselkedjék. De ha eléggé összetett ahhoz, hogy intelligens legyen, a megértése túl komplikált lesz számunkra.” „Mindenesetre semmi akadálya sincs annak, hogy olyasmit építsünk, aminek nem értjük a működését.” „A számítás következő forradalma olyan analóg rendszerek létrehozásával fog járni, melyek fölött a digitális programozásnak semmilyen hatalma nem lesz. Azoknak, akik azt hiszik, hogy építhetünk olyan gépeket, amelyek mindent ellenőriznek, a Természet válasza az lesz, hogy megengedi, olyan gépet alkossanak, ami őket fogja ellenőrizni.”

E kérdések kapcsán abba a súlyos filozófiai problémába botlunk, hogy hol van az emberi megértés határa. Gotthard Günter kibernetikus szerint egy bonyolult rendszer csak egy még bonyolultabb szintről érthető meg. Ily módon ez már egy metafizikai probléma.

(Humán – MI ökoszisztémák¹⁴) A kibernetikai szemlélet eredetileg az egyéni cselekvő beágyazottságát jelentette az irányítási és információs folyamatba. Most viszont a szereplők hálózatának emergens tulajdonságairól van szó – a hálózatok tudományának

megjelenése következtében. Az is bonyolította a helyzetet, hogy kiderült, a káosz, majd a komplexitás a rendszerek tipikus viselkedésmódjai. Az eredeti, a döntéshozatalba bekapcsolt cselekvő egyénre vonatkozó kibernetikai szemlélet tehát most kiterjed az egyének emergens viselkedést mutató hálózatának és a gépeknek az összekapcsolt, komplex rendszerére. A MI része lett a kibernetikai folyamatnak. Ez merőben különbözik az eredeti kibernetikai felfogástól. Lassan képesek leszünk a komplex, heterogén hálózatok emergens viselkedésének az elemzésére, előrejelzésére, sőt, a tervezésére is. Ezek Alex „Sandy” Pentland gondolatai, aki az MIT-n a médiatudományok professzora. A kibernetikai folyamatban történt változást a szemléletesség kedvéért ábrával is érzékeltetem:



A Big Data elhozta azt a lehetőséget – írja Pentland –, hogy újraalkossuk civilizációinkat. Most láthatjuk a szociális interakciók részleteit is, és nem kötnek az átlagok, akár a piaci fejleményekről, akár politikai választásokról van szó, és ez jelentősen növeli az előrelátás képességét. A Big Data valójában korunk prométheuszi tüze, mely azonban ugyanúgy fordítható jóra, mint rosszra. A Big Data és a MI segítségével olyan ökológiai modelleket és rendszereket építhetünk, amilyenekre még sohasem volt módunk. De fennáll a veszélye az algoritmusok zsarnokságának is, amikor senki által sem választott adatkezelők irányítják a világot. Ez azonban – Pentland szerint – könnyen meggátolható: ha ellenőrizzük az algoritmusok bemeneteli és kimeneteli oldalait. (Látva, tapasztalva az internetes közösségi platformok működése körüli problémákat, én ezt nem nevezném könnyűnek.)

Ahhoz, hogy eljussunk a Humán – MI ökoszisztémákhoz, a MI-nak még komoly technikai fejlesztésre van szüksége. A mesterséges neuronháló neuronjai jelenleg még „buták”, üresek, ezért az általuk generált tudás nélkülözi a kontextust. Meg kell oldani, hogy az emberről való minél több tudással legyenek feltöltve. (Vagy éppenséggel annak a tudományterületnek az ismereteivel, amiben megoldást várunk tőlük.) Továbbá, a „reinforcement learning”, a megerősítő tanulás során fejleszteni kell azokat a jutalmazó függvényeket („credit assignment functions”), visszacsatolásokat, amelyek értékelik, hogy az egyes neuronok tevékenysége miként járult hozzá az eredményhez.¹⁵ Pentland utópisztikus elképzelése: ennek mintájára kellene felépíteni az emberi társadalmat is. Egyik érve az, hogy az emberi döntések gyakran a MI esetén fejlesztendő „credit-assignment” függvényekre hasonlítanak. (Bár bevallja, hogy ennek az elképzelésnek ugyanakkor komoly hátránya, hogy a propaganda, a reklámok, az álhírek erősen befolyásolják magatartásunkat.) De alkalmazhatónak tartja a módszert vállalatok működtetésére is – pontosabban a vállalati alkalmazottak magatartásának optimalizálására.

(*Kozmikus perspektíva*) Tegmark meghatározása szerint az intelligencia az atomok olyan elrendeződése, amely képes az információ-feldolgozásra. (Ebből következik, hogy az intelligencia nem kötődik kizárólag az élethez, a szerves anyaghoz.) Kurzweil (aki nem szerepel e kötet szerzői között, viszont e tudományterület egyik legnagyobb alakja) az intelligenciát mintázatfelismerő képességként határozza meg. (Egyszerűbben: az összefüggések felismerésének képessége.) A világ-mindenség evolúciója szerinte

egy olyan folyamat, melynek során az információtechnológia és az emberi intelligencia egybeolvadása után „a világegyetemben létező anyag és energia átítatódik intelligens folyamatokkal és tudással”.¹⁶ Az univerzumnak tehát mintegy ez a sorsa, végzete, hogy intelligenssé váljon, és ebben az emberi intelligencia és az ember által kifejlesztett technológia döntő – ha ugyan nem kizárólagos – szerepet játszik. Hasonlóan magasztos szerepet szán Tegmark is az embernek. A Homo sapiens előtt szerinte két út áll (és ezt a jövőértékelést én is osztom). Az egyik az, hogy eltűnik, kihal. Elszaporodása és a bioszféra tisztító tevékenysége – ő a bioszféra réme – a fajok tömeges kihalásához vezetett, és a MI megalkotásával, de főként annak felelőtlen alkalmazásával saját maga is a kipusztult fajok sorsára jut.

A másik lehetőség az, hogy az ÁMI kifejlesztésével az ember elterjeszti az intelligenciát az Univerzumban. Az emberi tudat megjelenése az ébredéshez hasonlatos: az Univerzum a Homo sapiens tudatának megjelenése révén kezd tudatosná válni, önmagára ismerni. Az ember azonban törekeny lény, és képtelen olyan bolygóközi utazásokra, amelyek nem évekig, de fényévekig eltartanak. Az ÁMI révén viszont képes eljuttatni a tudatot, az intelligenciát (kiborgok, vagy letöltött elmék formájában) az Univerzum legtávolabbi szegleteibe is, és ezáltal a Homo sapiensnek lesz köszönhető, hogy ez a tudattalan Multiverzum öntudatosodik, önmagára ismer. Az ÁMI tehát ugyan megfosztja majd az embert mindattól, ami eddig létének értelmet adott: a hasznos tevékenységtől, önmaga fenntartásának képességétől, de egy semmivel sem fogható magasztos misszióval pótolja azt. (Ha csakugyan pótolja.)

E könyv több szerzője az ember és az általa alkotott intelligencia jövőbeli viszonyát pesszimistán értékeli. Venki Ramakrishnan az angol Royal Society (Tudományos Akadémia) indiai származású elnöke, Nobel-díjas mikrobiológus. Miközben egyetlen olyan megállapítást sem tesz, mely ne lenne ismert a MI iránt érdeklődő laikusok számára, arra a kérdésre, hogy főként kerekedhet-e a MI, ezt a választ adja: „Amiben egészen biztos vagyok: a MI sosem fog uralkodni a baktériumok fölött.”

Caroline A. Jones művészettörténész az MIT-n. Azt, amivé a kibernetikai gondolat mára fejlődött (MI, az embert fenyegető szuperintelligencia) rossz iránynak tartja, és jobboldali kibernetikának nevezi. Ez a „makacs arrogancia” vezetett a mai automata fegyverrendszerekhez, a humán munkaerő kiszorításához és a Google profitéségéhez. „Vissza kell térnünk a baloldali kibernetikához, azokhoz az elméleti biológusokhoz és antropológusokhoz, akik célja az volt, hogy elősegítsék a fajokon túlmutató intelligens rendszerek megértését.”

NÉHÁNY FONTOSABB FORRÁSMŰ

Bostrom, Nick: Szuperintelligencia. Ad Astra, 2015.

Bostrom, Nick: Ethical Issues in Advanced Artificial Intelligence. web, 2003. <https://nickbostrom.com/ethics/ai.html>

Eggers, Dave: A Kör. Európa, 2016.

Günther, Gothard: Cyberphilosophy. BCL-Reports. English Collection, not complete. (Ed. Rudolf Kaehr, prepared by Prof. Eberhard von Goldamer, 2004), web.

Harari, Yuval Noah: Homo Deus. Animus Kiadó, 2016.

Kaku, Michio: Az elme jövője. Akkord Kiadó, 2014.

Kiss Károly: A mesterséges intelligenciáról – halandóknak. Valóság, 2019. december.

Kiss Károly: Ki akar itt superman lenni? Digitális jövő – veszélyekkel. L'Harmattan, 2019.

Kiss Károly: Az online titánok megszelídítése. Az internet szabályozása és jövője. Valóság, 2018. 10.

- Kurzweil, Ray: A szingularitás küszöbén. Amikor az emberiség meghaladja a biológiát. Ad Astra, 2014, és web.
- Kurzweil, Ray: How to Create a Mind. Viking, Penguin Books, 2012, és web.
- Pokol Béla: A mesterséges intelligencia társadalma. Kairosz, 2018.
- Precht, Richard David: Künstliche Intelligenz und der Sinn des Lebens. Goldman Verlag, München, 2020.
- Tegmark, Max: Élet 3.0. Embernek lenni a mesterséges intelligencia korában. hvg könyvek, 2018.

JEGYZETEK

- 1 Lehetséges elmék. A mesterséges intelligencia 25 megvilágításban. A kötetet a Penguin Books adta ki 2020-ban.
- 2 Kiss Károly: A mesterséges intelligenciáról – halandóknak.
- 3 Wiener nagy gondolata az volt, hogy az információs, irányítási és ellenőrzési folyamatok egyaránt vonatkozathatók az élő szervezetekre és a gépekre is. Alaptétele az a meglátás, hogy az intelligens viselkedés visszacsatolási folyamatokon alapul, és ez a gépekben is utánozható – ezáltal jutunk el korunk mesterséges intelligenciájához.
- 4 A könyv következő, 1954-es kiadásából viszont Wiener törölte a vonatkozó fejezetet. Az atombomba, majd a hidrogénbomba gyártása abban a korban különben is diszkontálta Wiener félelmeit a kibernetikai rendszerektől.
- 5 Egy Amerika meghódításához kapcsolódó történelmi mendemonda szerint Cortez és Pizarro azért voltak képesek néhány száz emberrel legyőzni az inkák és az aztékok sokezes seregeit, mert a bennszülöttek nem ismerték sem a lovat, sem a lőfegyvert, és az arcszörzetük is gyér volt. Így isteneknek képzelték a lovon ülő (réműletükben azzal egy élőlénynek hitt), kezükben mennydörgést és villámot szóró bottal felfegyverezett, szakállas spanyol konkvisztádorokat.
- 6 Lásd erről Élet 3.0 című könyvét.
- 7 Ebben van egy kis túlzás: amikor az algoritmust egy nagy adatbázison tanítják, „tréningeztetik”, az összefüggések felismeréséhez számára csupán a másodperc töredékére van szükség. De a hasonlat annyiban szemléletes, hogy az eredmény elérése nem hirtelen felismeréssel, hanem fokozatosan történik: a jelenséget leíró függvénygörbe fokozatosan válik egyre pontosabbá az újabb és újabb adatok birtokában.
- 8 De az öntudat nem az evolúció eredménye? Harari azt egy kb. 70 ezer évvel ezelőtt bekövetkezett mutációnak tulajdonítja.
- 9 Lásd erről Az elme jövője c. könyvét.
- 10 Harari, Homo Deus, 100. o.
- 11 E téma részletes kifejtését lásd „Az életnek értelme adó halál” című írásomban, a Ki akar itt superman lenni? című kötetben.
- 12 Lásd erről: Kurzweil, A szingularitás küszöbén.
- 13 Dave Eggers „A Kőr” című regénye egy ebből következő hátborzongató disztópia bekövetkeztét vázolja fel a közeljövőben. Nem nehéz felismerni, hogy a szerző a Facebook világalumát vizionálja. Egyébként vannak tanulmányok, melyek azt bizonyítják, hogy egy internetes platform el tudná látni a törvénykezés és kormányzás feladatát. (Lásd erről Az új szuverén című írást a Ki akar itt szuperman lenni c. kötetben.)
- 14 Az angol-nyelvű szakirodalomban az utóbbi időkben az ökoszisztéma kifejezést szélesebb értelemben használják: nem természeti, hanem a természeti rendszerek módjára működő komplex társadalmi és gazdasági rendszereket értenek alatta.
- 15 A wikipédia a Credit Assignment Problem magyarázatára azt a szemléletes példát hozza fel, amikor egy futballmeccs után az edző kiértékeli a játékosok teljesítményét: megmondja, hogy az egyes játékosok játékának elemei, mozdulatai mennyiben járultak hozzá a végeredmény kialakulásához és annak alapján tervezi meg a következő mérkőzés taktikáját.
- 16 Ray Kurzweil: A szingularitás küszöbén. web, 23. o.