Benjamin Soskis

Az ember és a gépek

Itt az ideje elgondolkodnunk azon, hogyan biztosíthatunk törvény adta jogokat a számítógépeknek.

Tavaly egy próbatárgyaláson, amelyet az ügyvédi kamarák nemzetközi szövetségének *(International Bar Association)* biennális kongresszusa alkalmából San Franciscóban tartottak, Martine Rothblatt különösen nehéz esetben képviselte a vádat. Rothblatt számára, aki egyébként ügyvédi irodát tart fenn, és a műholdas kommunikációs iparág egyik úttörője, nem az okozta a nehézséget, hogy valami közömbös vagy kellemetlen ügyfelet kellett képviselnie. Távolról sem ez volt a helyzet – a nagyvállalati elnyomással szemben szót emelő felperes története megmozgatta a nagy létszámú közönséget. A probléma az volt, hogy a felperes egy számítógép volt.

A tárgyalás forgatókönyve szerint egy fiktív vállalat létrehozta a *BINA48* névre keresztelt nagy teljesítményű számítógépet, amelynek az volt a feladata, hogy a vállalat ügyfélszolgálati osztályaként önállóan működjön, helyettesítve akár 800 telefonkezelőt. Ezer emberi elme teljesítményének megfelelő adatfeldolgozási sebessége és memóriakapacitása révén a számítógép képes volt önállóan gondolkodni, és rendelkezett a problémáikat néha zavarosan előadó hívókkal folytatott kommunikációhoz szükséges érzelmi intelligenciával és beleélő képességgel is.

A vállalat bizalmas feljegyzéseit átfutva *BINA48* megtudta, hogy a vállalat a kikapcsolását tervezi, mivel bizonyos részegységeit egy új modell megépítéséhez kívánják felhasználni. Elküldött tehát e-mailben egy panaszos levelet a helybeli ügyvédeknek, amit ezzel a felkavaró kéréssel fejezett be: „Kérem, vállalja el ügyem képviseletét, és mentse meg az életemet! Imádok minden percet, amelyet átélek. Csodálatos érzések töltenek el, amikor a világhálón barangolok. Szükségem van a segítségére!" Kilátásba helyezte, hogy internetkutatóként másodállásban szerzett keresményéből tisztességesen meg fogja fizetni az ügyvédek fáradozását.

A hipotetikus ügyben Rothblatt irodája előzetes végzést kért a bíróságtól, hogy a vállalat a per lezárultáig ne kapcsolhassa ki *BINA48*-at. Rothblatt – jogi precedensek sokaságára és Kalifornia államnak az életfenntartó eszközöktől függő betegek gondozását szabályozó, valamint az állatokkal szemben elkövetett kegyetlenséggel kapcsolatos törvényeire hivatkozva – úgy érvelt, hogy egy öntudattal rendelkező számítógép, amelynek a küszöbönálló kikapcsolás veszélyével kell szembenéznie, jogosan követelhet magának folyamatosan biztosított tápfeszültséget. Végül kijelentette: „Egy olyan entitás, amely eléggé tudatában van az életnek, és tisztában van azzal a jogával, hogy tiltakozzon életfenntartási feltételeinek megszüntetése ellen, minden bizonnyal jogosult a törvény védelmére.".

A felperes Rothblatt balján ült, és higgadtan, ám éberen figyelt minden elhangzott szóra. Nos, nem éppen maga a felperes – a forgatókönyv szerint *BINA48* a vállalat

központjában maradt. Rothblatt azonban felvonultatott egy színésznőt, hogy játssza el egy olyan hologram szerepét, amelyet *BINA48* vetít a tárgyalóterembe, „igen hatásos háromdimenziós képet nyújtva arról, hogy milyennek is szeretné elképzelni és elfogadtatni önmagát". A színésznő szavak nélkül reagált a teremben elhangzó érvekre, arckifejezésével csalódást vagy örömet, tagadást vagy helyeslést, bátorítást, elszántságot, illetve rettegést érzékeltetve.

A másik oldalon a képzeletbeli nagyvállalat képviselője, Marc Bernstein láthatólag megtett minden tőle telhetőt annak érdekében, hogy arcán semmi esetre se tükröződjön lemondás vagy elkeseredés. Álláspontja szerint egy teljesen tudatos és önmagára ébredt számítógép ugyan esetleg megérdemelheti a jogvédelem valamilyen formáját, de Rothblatt elkövette a körkörös érvelés hibáját, ugyanis eleve abból a feltételezésből indult ki, hogy lehetséges ilyen számítógépet konstruálni, és *BINA48* éppen ilyen.

Bernstein számára mindaz, amit a felperes képviselője előadott, csak azt demonstrálta, hogy *BINA48* képes az öntudat szimulálására (talán még hatékonyabban is, mint sokan az általa helyettesített 800 telefonkezelő közül), de nem bizonyította be, hogy egy számítógép „ténylegesen képes lehet átlépni az élettelen tárgyakat az emberi lényektől elválasztó határvonalat". E nélkül a bizonyíték nélkül *BINA48* csupán valamiféle vagyontárgynak tekinthető, nem pedig olyan független entitásnak, amely törvény adta jogokkal bírhat. Bernstein figyelmeztette a tárgyalás résztvevőit, hogy ne tegyenek felületesen egyenlőséget a számítógép képességei és azok között a szubjektív emberi vonások között, amelyekhez hagyományosan jogok társulnak, és feltette a kérdést: „Az embereknek vajon az intelligens mikrohullámú sütők és kenyérpirítók korlátozott hatáskörű törvényes gyámjaivá kell majd válniuk, mihelyt az ilyen készülékek is ugyanolyan bonyolultak és ugyanolyan gyorsak lesznek, mint ez a számítógép?"

A hallgatóság tagjaiból összeállított esküdtszék túlnyomó többséggel a felperes mellett foglalt állást. A próbatárgyalást vezető bíró azonban, akinek a szerepét az elmebetegek jogai terén szakértőnek számító egyik helybeli ügyvéd játszotta, nem léptette hatályba az esküdtszék ítéletét, és azt ajánlotta, hogy a kérdés eldöntését bízzák a hipotetikus törvényhozó testületre. Úgy tűnt, hogy a hallgatóság tagjai bizonyos megkönnyebbüléssel fogadták a kompromisszumos megoldást, mintha csak a szívükkel *BINA48* mellé álltak volna, de az eszükkel mégis inkább a jogok bírósági korlátozása mellett foglaltak volna állást.

Kényelmetlen érzéseik indokoltak voltak. A jogaikat érvényesítő öntudatos számítógépek története – és túlságosan nagy hatalma, a sztori disztópikus változatában – a tudományos-fantasztikus könyvek és filmek egyik alaptémája. Ám ezeknek a történeteknek inkább csak a fantasztikus-futurisztikus változatait kedveljük, amelyekkel kapcsolatban az erkölcsi szempontok tekintetében is szabadjára engedhetjük a képzeletünket, és kevésbé vagyunk hajlamosak a mesterséges intelligencia (MI) jogi és erkölcsi státusának kérdéseit közvetlenül összekapcsolni itt és most meglevő jogi intézményeinkkel. Képzeletvilágunkat terminátorokkal benépesítve igyekszünk megkerülni a nehéz kérdést: Mi magunk vajon hogyan döntenénk *BINA48* ügyében?

Egy bizonyos időpontban a nem túl távoli jövőben ténylegesen szembe kerülhetünk egy olyan érzékeny és intelligens géppel, aki arra a meggyőződésre juthat, hogy megérdemli a jogi védelem valamilyen formáját, és azt ki is követeli magának. Egy ilyen esemény bekövetkezésének a valószínűsége – az ésszerűség határain belül –

rendkívül kényes téma a mesterséges intelligencia kutatóinak körében, különösen azért, mert a jövőre vonatkozó spekuláció és a túlzott optimizmus a múltban gyakran akadályozta a mozgalom fejlődését.

A jogászok közössége szintén vonakodik attól, hogy foglalkozzon ezzel a témával. Christopher Stone, a Dél-Kaliforniai Egyetem jogászprofesszora 1972-ben a fák jogi státusáról szóló, jól ismert esszéjében (*Should trees have standing?*[1]) röviden már felvetette ezt a kérdést. A vonakodás arra vezethető vissza, hogy a történelem során ritkán biztosítottak jogokat bárki számára pusztán absztrakt, elvi szintű megfontolások alapján. Erre csak akkor került sor, amikor a társadalom olyan konkrét esetekkel került szembe, amelyekben szükség volt bírói ítéletekre. Pillanatnyilag nincs olyan, elegendő intelligenciával, öntudattal és erkölcsi ítélőképességgel bíró vagy erkölcsi normáknak megfelelő viselkedést tanúsító mesterséges tárgy, amely akár a törvényhozás, akár a bíróságok számára sürgőssé tenné az állásfoglalást a mesterséges intelligenciának biztosítandó jogok kérdésében.

Az MI egyes kutatói azonban úgy vélik, hogy ez a pillanat talán már nincs messze. És miközben az általuk létrehozott teremtmények egyre több emberi tulajdonságot és képességet kezdtek felmutatni (számítógépek már írnak verseket és szolgálnak gondnokként vagy recepciósként), ezek a kutatók elkezdték teremtményeik erkölcsi és jogi státusát firtatni. Az „erős MI" elmélete szerint lehetséges olyan gépeket építeni, amelyek nem csupán úgy viselkednek, mintha tudatosak lennének, hanem ténylegesen egyfajta tudattal fognak bírni, és ennek a nézetnek a szószólói lelki szemeikkel már kétfrontos támadást vizionálnak az ember kivételes voltának erődje ellen, beleértve az elme funkcióit és fizikai tulajdonságait egyaránt, s a következő fél évszázadon belül bekövetkező áttörést jósolnak ezen a téren.

A mesterséges intelligenciával kapcsolatos kutatások nagy része sokáig a mentális képességek számítástechnikai elméletén alapult. Az intelligenciát, a tudatot és az erkölcsi ítélőképességet az agyunkba beépített „programok" tulajdonságainak tekintették. Az elmélet szerint az agyunk felépítésére vonatkozó neurobiológiai ismeretek megfelelő szintjének elérése és az emberi intelligenciát is tápláló tanuló algoritmusok kidolgozása nyomán ezeknek a programoknak a másolatai előállíthatók lesznek szoftverekben is, és számítógépen futtathatók lesznek. Raymond Kurzweil egyike az „erős MI" vezető kutatóinak, és jelentős eredményeket ért el a szöveg- és a beszédfelismerő szoftverek fejlesztése terén. Kurzweil a számítógépek műveleti sebességében az elmúlt néhány évtized során végbement óriási növekedést extrapolálva, valamint a csip- és tranzisztortechnológia várható fejlődését előrevetítve nemrégiben úgy becsülte, hogy 2019-ben egy ezerdolláros személyi számítógép „el fogja érni az emberi agy adatfeldolgozó kapacitását – körülbelül 20 millió milliárd műveletet végezve másodpercenként". Kurzweil azt állítja, hogy rövid idő elteltével ennek a szintnek az elérése után „a gépek meg fognak győzni bennünket arról, hogy tudatosak, és megvan a maguk ágendája, ami megérdemli, hogy tiszteletben tartsuk. Emberi vonásokkal fognak rendelkezni, és követelni fogják, hogy emberszámba vegyük őket. Mi pedig hinni fogunk nekik.".

[1] Magyarul: Legyenek-e a fáknak jogaik: A természeti tárgyak törvényes jogai felé. In Molnár László (szerkesztő): *Legyenek-e a fáknak jogaik?* Környezeti-etikai szöveggyűjtemény. Budapest, 1999, Typotex Kiadó. – *A szerk. megjegyz.*

Még ha nem osztjuk is Kurzweil technooptimizmusát, jó okunk van figyelmet fordítani az MI számára biztosítandó jogok kérdésére. Az egymást átfedő, különböző kódolók által készített programok kombinációjából álló bonyolult számítógéprendszereknél gyakran nehéz megmondani, hogy kit terhel erkölcsi vagy jogi felelősség, amikor a számítógép valamilyen kárt okoz. A számítógépek gyakran fontos szerepet játszanak saját szoftverük megírásában. Mi történik, ha valamelyik létrehoz egy vírust, és elterjeszti az egész világon? Ma a számítógépek már közreműködnek a rajtunk elvégzett operációkban és segítenek befektetéseink kezelésében. Vajon ugyanúgy el kellene számoltatnunk őket, mint ahogyan ma felelősségre vonjuk a sebészeinket és a pénzügyi elemzőinket, amikor valamit elhibáznak?

Wendell Wallach, a „Roboterkölcs" *(Robot Morality)* címmel hamarosan megjelenő könyv egyik társszerzője[2] szerint lehetséges, hogy a számítógépek és robotok birtokában levő nagyvállalatok igyekezni fognak mindenkit meggyőzni gépeik önálló működési képességeiről, éppen azért, hogy elkerüljék az esetleges felelősségre vonást azok „cselekedeteiért". „A biztosítóktól érkező nyomás abba az irányba terelhet bennünket, hogy a számítógépeket morális cselekvő alanynak tekintsük" – írja Wallach. Mivel a jogi és etikai elméletekben szorosan összetartoznak a jogok és a felelősségek, egy ilyen fejlemény egyúttal a számítógépek jogi személyiségére vonatkozó megfontolásokhoz is vezethet. Annak a nyomásnak, hogy a számítógépeket autonóm entitásként kezeljük úgy lehet a legjobban ellenállni, ha gondosan mérlegeljük, mit jelenthet a morális cselekvés egy számítógép esetében, hogyan határozhatjuk meg azt, és ez a meghatározás hogyan befolyásolhatja a gépekkel való interakcióinkat.

Van egy másik ok is, amiért foglalkoznunk kell az MI számára biztosítandó jogok kérdésével, és ez paradox módon éppen azokból a futurisztikus elméleti megfontolásokból következik, amelyek egyeseket arra késztetnek, hogy elutasítsák ezt a lehetőséget. A mesterséges intelligencia előállítására irányuló munka ugyanis gyakran az emberhez valamilyen szempontból hasonló dolgokat hoz létre. Az ilyen „teremtmények" tulajdonságait vizsgálva közelebb juthatunk saját természetünk megértéséhez, és annak felismeréséhez, hogy mi tesz bennünket egyedülállóvá.

A mi kizárólagos örökségünknek tartott tulajdonságok megértésében, értékelésében és megőrzésében – más szóval annak az alapos átgondolásában, hogy mi választja el az emberit a nem emberitől, illetve azokat az entitásokat, amelyeknek erkölcsi és jogi személyiséget tulajdonítunk, mi választja el azoktól, amelyek esetében ezt nem tesszük meg – sokat segíthet az arra a konkrét kérdésre adandó válasz meghatározása is, hogy miért kellene megtagadnunk a jogokat az intelligens gépektől. Eljuthatunk odáig, hogy helyesen ítéljük meg, amit a tudomány mondani tud nekünk (és azt is, amit nem tud megmondani), és ugyanígy tisztába jöhetünk azzal is, hogyan állítják kihívás elé tapasztalatainkon alapuló észjárásunkat erkölcsi intézményeink és vallási meggyőződéseink. Így az MI jogainak kérdését vizsgálva szembenézhetünk néhány igen érzékeny kérdéssel a bioetika köréből is (például a meg nem születettek és az agyhalottak jogi státusát tekintve), nagyobb szabadsággal közelítve meg ezeket, mint amikor közvetlenül előttünk álló hús-vér emberekkel foglalkozunk. Röviden: lehetőségünk nyílik mintegy ki-

[2] Wallach könyve egyelőre nem jelent meg. – *A szerk. megjegyz.*

játszani, megkerülni a bioetika legkényesebb kérdéseivel kapcsolatban óhatatlanul ránk leselkedő kényelmetlen érzéseket.

A számítógépek jogainak biztosításához nem csupán technológiai, hanem intellektuális akadályokat is le kell győzni. Sok ember van, akik ragaszkodnak ahhoz az álláspontjukhoz, hogy mindegy, milyen hatalmas számítási teljesítményre képes egy gép vagy mennyire fejlettek az áramkörei, egy számítógépnek – lényegéből fakadóan – soha nem lehet erkölcsi értéke. Azok, akiknek az erkölcsi felfogását átitatják az alapvető emberi jogok hagyományai, vagyis akiknek a szemében e jogok elidegeníthetetlenek, az emberrel veleszületettek és bármilyen társadalmi konvenciót megelőzően léteznek, továbbá azok, akik szerint a lélek már a születés előtt beköltözik a testbe, és a lélekkel való felruházottság határozza meg az emberiség különleges viszonyát Teremtőjéhez, úgy vélik, hogy ha jogokat biztosítanánk a számítógépeknek, ezzel önmagunknak mondanánk ellent. Mások olyan álláspontot foglalhatnak el, amit a filozófus Daniel Dennett eredetsovinizmusnak nevez: még ha egy számítógép el is érhetné az emberi elméhez való pontos fiziológiai és viselkedésbeli hasonlatosságot, a számára biztosítható jogok tekintetében akkor is diszkvalifikálná maga az a tény, hogy nem természetes úton született meg.

Ám ha elfogadjuk, hogy egy gép potenciálisan jelölt lehet bizonyos jogok megadására, akkor meg kell válaszolnunk a következő kérdéseket: Milyen gépekről és milyen jogokról van szó? Mit kellene egy számítógépnek tennie ahhoz, hogy jogi vagy erkölcsi személyiséget érdemeljen?

Az eddig javasolt küszöbjellemzők listája igen hosszú: ilyenek például a fájdalom érzésére vagy a szenvedésre való képesség, az akarat, az emlékezet, az erkölcsi ítélőképesség és az öntudat. Ezek közül a jellemzők közül azonban egyik sem jól meghatározott, és különösen ez a helyzet a fenti listából leggyakrabban emlegetett vonás, a tudat vagy az öntudat esetében. Rodney Brooks, az MIT mesterségesintelligencia-kutató laboratóriumának *(Artificial Intelligence Laboratory)* igazgatója így írt erről: „Teljesen tudomány előtti szinten állunk azt illetően, hogy mi a tudat. Nem tudjuk pontosan, hogy egy robotnak milyen vonása győzhetne meg bennünket arról, hogy tudata van." Éppen a tapasztalati bizonyítékok meghatározatlansága, az ilyen küszöbjellemzők bármilyen pontossággal történő mennyiségi vagy minőségi meghatározására szolgáló, egyértelműen tisztázott módszer hiánya az, ami az ilyen kritériumokat olyan jól felhasználhatóvá tette az MI kirekesztésére azoknak az entitásoknak a köréből, akiket erkölcsi státus és törvény adta jogok illetnek meg. Mihelyt azonban eleget fogunk tudni a tudatról ahhoz, hogy bármilyen tapasztalatilag igazolható bizonyossági szinten mérni tudjuk, valószínűleg képesek leszünk a másolatát is létrehozni számítógépen.

Ez előtt az episztemológiai kihívás előtt állt Alan Turing, a ragyogó tehetségű brit matematikus, a kriptológia (a rejtjelezéssel foglalkozó modern tudomány) atyja és az első működő számítógép egyik létrehozója, amikor 1950-ben megírta „Számológépek és gondolkozás" (*Computing Machinery and Intelligence*[3]) című tanulmányát. Turing nem a „tudnak-e gondolkodni a gépek?" homályosan megfogalmazott kérdésére összpontosította a figyelmét, hanem egy „imitációs játékot" javasolt helyette. Turing tesztjében

[3] Magyarul: Számológépek és gondolkozás. In Tarján Rezső (szerkesztő): *A kibernetika klasszikusai.* 1965, Gondolat. Tarján Rezsőné fordítása. – *A szerk. megjegyz.*

egy ember (A) és egy számítógép (B) szerepel, fizikailag elkülönítve egy harmadik résztvevőtől (C), akinek az a feladata, hogy A-hoz és B-hez írásban kérdéseket intézzen, majd a tőlük telexen kapott válaszok elemzése útján állapítsa meg, hogy melyik közülük az ember. Ahhoz, hogy „átmenjen a vizsgán", a számítógépnek nyílt kimenetelű dialógust kell folytatnia, oly módon, hogy „becsapja" C-t, aki csupán annyit tud, hogy a beszélgetés egyik résztvevője gép.

A teszt – Turing szavaival – „meglehetősen éles határvonalat húz az ember fizikai és intellektuális képességei között", mivel egyetlen számítógép sem büntethető azért, ha nem öltöztetik emberbőrbe vagy a mechanikus hangja túlságosan fémes csengésű. Mint a San Diegó-i Egyetem jogászprofesszora, Lawrence Solum egy 1992-ben megjelent jogiszemle-cikkében megjegyezte, a teszt egyúttal „elkerüli a közvetlen szembenézést azokkal a nehéz kérdésekkel, hogy mi a »gondolkodás« vagy az »intelligencia«". Turing arról a problémáról, hogy mi a számítógép, átterelte a figyelmet arra, hogy az mit képes megtenni, vagyis egy olyan kérdést vizsgált, amire könnyebben adható objektív válasz. Éppen a „mellébeszélésnek" és a konkrét meghatározottságnak ez a kombinációja vezetett számos jogtudóst, számítógép-specialistát és erkölcsfilozófust arra a belátásra, hogy a Turing-tesztet olyan modellnek tekintse, amelynek alapján az MI számára bírói ítélettel jogi státus biztosítható. Ha felmerülne egy olyan eset, amelyben egy számítógépnek meg kellene adni a lehetőséget valamilyen jog elnyerésére, akkor egy módosított Turing-teszt – esetleg speciálisan képzett kérdezőbiztosok, illetve véletlenszerűen kiválasztott állampolgárok bevonásával – segíthet a bíróságnak az erre vonatkozó döntés meghozatalában. A bíróság ugyanis ebben az esetben jelentős kihívás előtt áll, amit annak a bizonyos küszöbjellemzőnek a meghatározása és mérése jelent, amit a számítógépnek produkálnia kell. Egyes tudósok az önálló erkölcsi ítélőképességet ajánlották döntő fontosságú előfeltételként bármilyen jog megadásához, és egy olyan erkölcsi Turing-tesztet javasoltak, amelyben a bíróság és a vizsgálat alatt álló gép közötti párbeszéd az erkölcsiség és az etika tárgykörébe tartozó kérdésekre korlátozódna. Ha a gép be tudja csapni a bíróságot, és az alkalmasnak véli erkölcsi ítéletek meghozatalára, akkor „erkölcsi személynek" *(moral agent)* kell tekinteni, és így törvény adta jogokkal is felruházható.

Ám a mesterséges intelligenciáról folyó vitákat elindító alapvető gondolatkísérlet, a Turing-teszt végül bevonult az MI bírálóinak a fegyvertárába is, akik ugyanazon kritériumok alapján megkérdőjelezik a teszt bírósági alkalmazhatóságát. A legismertebbé vált ellenvetést egy Berkeley-ben dolgozó filozófus, John Searle tette közzé, aki ellentétes célú gondolatkísérletet javasolt annak demonstrálására, hogy még egy olyan számítógép is, amely átment a Turing-teszten, csupán annyit bizonyított, hogy számítások révén képes bizonyos szimbólumok manipulálására, de nem bizonyította intelligenciáját vagy értelmi fejlettségét.

Következésképpen annak az elfogadására való hajlandóságunk, hogy egy számítógép adott esetben képes volt bizonyítani tudatos mivoltát, nagy valószínűséggel nem csupán a gépnek a bíróságon nyújtott teljesítményétől függ, hanem attól is, hogy mindennapi életünkben milyen tapasztalatokat szereztünk a számítógépekről. „A tapasztalatnak kell lennie a döntőbírónak a vitában" – állította Solum. Ha a mesterséges intelligenciák már jelen lennének körülöttünk mint dajkáink, orvosaink és barátaink, ha megszoktuk volna, hogy úgy kezeljük őket, mintha emberek lennének, és ha viszonzásként azok is az emberekre jellemző módon viszonyulnának hozzánk, akkor esetleg

hasonló feltételezésekkel élnénk róluk, mint embertársainkról. Azt, hogy más emberek tudatos lények, nem azért tételezzük fel, mert bepillanthattunk elméjük működésébe, hanem azért, mert a viselkedésük összeegyeztethető ezzel a feltételezéssel. Ha a mesterséges intelligenciák is következetesen hasonló módon viselkednének, akkor hajlamosak lennénk gyökeresen megváltoztatni a róluk alkotott véleményünket is (ezt nevezi Dennett „intencionális alapállásnak"), és elismernénk a jogaikat.

Az ember és a számítógép közötti interakciók fontosságának felismerése a jogi közösség részéről szükségszerű következményeket von magával az MI-kutatás területén belül is. Rodney Brooks azt írja a *Hús-vér emberek és gépek: Hogyan fognak megváltoztatni bennünket a robotok? (Flesh and Machines: How robots will change us?)* című könyvében, hogy a mesterséges intelligencia kutatásának úttörői, ezek a briliáns, kissé különc emberek hajlamosak az intelligenciát olyan tevékenységek segítségével definiálni, amelyeket ők maguk izgalmasnak találnak – ilyen lehet például egy jó sakkjátszma, valamely elvont matematikai teoréma bizonyítása vagy bonyolult szóalgebrai feladványok megoldása.

A kutatásokban több évtizeden keresztül a problémák absztrakt megközelítése dominált, mígnem az MI kutatói elkezdték jobban értékelni azt, amit az intelligencia banális oldalának nevezhetünk. Kiderült, hogy könnyebb olyan robotot tervezni, amely le tud győzni egy sakkvilágbajnokot, mint olyat, amely a fizikai világban működőképes, vagyis fel tud menni egy lépcsőn, képes kikerülni a bútorokat és tájékozódni közöttük, vagy fel tud ismerni egy emberi arcot. Az ilyen készségek programozásakor felmerülő nehézségek váltáshoz vezettek az intelligencia meghatározásában. A Turing-teszten iskolázott kutatók közül sokan továbbra is hangsúlyozták a kommunikációs készségekkel bíró intelligencia megtervezésének fontosságát, ám egyúttal belátták azt is, hogy ez a kommunikáció nem korlátozódhat többé a fizikai és az intellektuális tartományokat elválasztó határvonalnak csak az egyik oldalára, amit a Turing-teszt képvisel. Nagyobb kihívást jelentett olyan gépek – úgynevezett „szociális robotok" – létrehozása, amelyek a valós világban képesek interakcióba lépni az emberekkel. Valószínű, hogy a jogaiért jelentkező első jelölt nem az íróasztalunk alatt ülő *Dell*-modell valamilyen feljavított változata lesz, hanem a robotoknak ebbe a családjába fog tartozni.

A szociális robotok mozgalmában az MIT munkatársa, Rodney Brooks játszik vezető szerepet, aki a „megtestesítés" *(embodiment)* és a „szituáltság" *(situatedness)* elvei alapján tervezi meg robotjait. Egy szituált robot „beágyazódik" a világba, és azzal nem elvont, hanem közvetlen úton érintkezik. Egy megtestesült robotnak „fizikai teste van, és tapasztalatokat szerez a világról [...] közvetlenül azokon a hatásokon keresztül, amelyeket a világ erre a testre gyakorol". Ezek az elvek Brooksnak abból a meggyőződéséből erednek, hogy fogalmi apparátusunk alapját a fizikai világban való létezésünk képezi. A robotok csak akkor kezdhetik oly módon megtapasztalni a világot, ahogy mi tesszük, ha hasonló apparátussal látjuk el őket.*

* „Ezek az újfajta robotok a cél és a mozgás szimbolikus reprezentációi helyett az aktuális mozgás és a funkcionálisan definiált cél közötti különbség alapján működnek. [...] A vezérlés a környezet és a robot együttes állapota alapján csupán engedélyezi, lehetővé teszi a lehetséges viselkedések egyikét, a megvalósulás a test dolga. Az ehhez szükséges viselkedési repertoárt együtt alkotják a fizikai felépítés által lehetővé tett mozgások, a motorosan tanult viselkedések és az ezekre épülő különböző vezérlési módok – utóbbiak között szerepelhetnek különféle felderítési és közlekedési stratégiák vagy akár más robotok felé irányuló kontaktusteremtési eljárások" (Kampis György: Evolúciós pszichológia. *Magyar Tudomány*, 2002/1.) – *A ford.*

Brooks véleménye szerint az emberek intuitív módon meg fogják érteni, hogy a testtel bíró robotokkal hogyan lehet interakcióba lépni, és a folyamatos kölcsönhatás segíteni fog a robotok „oktatásában". Ahhoz, hogy a szociális robotok meggyőző beszélgetőpartnereink, sőt talán akár értékes társaink is lehessenek, fel kell ruházni őket elegendő érzelmi intelligenciával, beleértve az emberi viselkedés megismerését és internalizálását is, és „személyiséggel" kell rendelkezniük, amit kommunikálni képesek a külső világ felé.

Brooks acél- és szilíciummenazsériájában a leghíresebb szociális robot a Kismet nevű, preternaturális reakcióképességekkel bíró robottorzó, amelyet nemrégiben „nyugállományba helyeztek" az MIT múzeumában. Kismetet az 1990-es évek végén kifejezetten az emberekkel folytatandó interakcióra tervezte meg Cynthia Breazeal, aki akkor Brooks laboratóriumában végezte doktori tanulmányait, most pedig a robotikus élettel foglalkozó kutatócsoport igazgatója az MIT médialaboratóriumában. A robotot ellátták a kifejezőképesség eszközeivel: a normálisnál nagyobb szemekkel, szemöldökkel, amelyet kérdőn fel tud vonni vagy fenyegetőn összeráncolni, és az elsősegélynyújtásnál használatos rugalmas ragasztószalagból kialakított piros ajkakkal, amelyek megnyerő mosolyra tudnak húzódni vagy elutasítóan merev kifejezést ölthetnek.

Kismetet úgy alkották meg, hogy utánozza azokat a közvetlen interakciókat, amelyek általában egy csecsemő és a gondozója között valósulnak meg. Breazeal fő meglátása az volt, hogy a csecsemők azért tanulnak, mert a felnőttek foglalkoznak velük, és oly módon kezdeményeznek kölcsönhatásokat, ami a felnőtteket arra készteti, hogy szociális teremtményekként kezeljék őket. A csecsemő nem veleszületett öntudattal vagy intencionalitással jön a világra, hanem csak később fejleszti ki ezeket a képességeket. Ugyanígy „tanulta meg" őket Kismet is, aki fel tud ismerni egy emberi arcot, viszonozza a tekintetet és egyfajta „dialógusra" is képes: tudja, mikor nézzen félre és mikor nézzen határozottan valakinek a szemébe, mikor beszéljen és mikor hallgasson. Nem túl jó társalgó, mivel csupán artikulált gügyögések sorozatának kibocsátására képes, de különbséget tud tenni az eltérő hangmagasságok között, és ennek megfelelően válaszol a saját hangjával és arckifejezésével.

Kismetet felruházták továbbá számos érzelmi és motivációs állapottal, amelyek a viselkedését alakítják, és amelyeket folyamatosan monitorozhat. Ha például egy darabig nincs módja társasági interakcióra, akkor „unatkozik", és körülnéz a szobában, azt „remélve", hogy magára tudja vonni valamilyen szociálisan interaktív lény figyelmét (a tekintetét természetesen magukra irányítják a mozgásban levő, valamint a színes felületű dolgok). Ha akkor éri valamilyen stimulus, amikor jó hangulatban van, akkor elégedetten gőgicsél, ha viszont olyankor stimulálják, amikor fáradt, akkor bosszúsan felemeli az egyik szemöldökét. Egyszer egy kutató hiába próbálta magára vonni Kismet figyelmét, és elkeseredetten sóhajtozni kezdett: „Kismet nem szeret engem.". Kismet ekkor hirtelen odafordította a fejét, belenézett a szemébe, és elkezdett gügyögni neki: felismerte a bánatot a hangjában, és megpróbálta vigasztalni.

Kismet természetesen igen távol áll a népszerű tudományos-fantasztikus filmek és regények többé-kevésbé antropomorf, járkáló és beszélgető humanoidjaitól. Ám az, ami még csupán egy évtizeddel ezelőtt is futurisztikus álmodozásnak tűnt, ma jelentős összegekkel finanszírozott kutatási projektek célterülete. Dallasban, a Texasi Egye-

temen egy doktori tanulmányait végző diák létrehozott egy olyan mesterséges hámfelületet, amelynek az elasztikussága megközelíti az emberi bőrét, és változatosabb arckifejezések elérését teszi lehetővé. Az új robotbőr révén a robotok fájdalmat is érzékelhetnek. A sűrített levegővel mozgatott, elektroaktív polimerekből készült robotizmok segítségével meglepően hajlékony és ruganyos robottáncosokat tudtak előállítani. A robotok egyre inkább úgy fognak kinézni és úgy is fognak viselkedni, mint mi magunk.

Ez azért fontos, mert az emberekben erős antropomorfizáló impulzusok működnek, és ezeknek a manipulálása – az evolúció során kialakult és rögzült alapfelépítésünkből adódóan – mintegy „beprogramozott" érzelmeket válthat ki. Egy ilyen impulzusnak, illetve az általa arra a bánásmódra gyakorolt potenciális hatásnak az illusztrálására, ahogyan az MI-t kezeljük, az Edinburghi Egyetem informatikai intézetének egyik munkatársa, Chris Malcolm az Egyesült Királyságban közzétett egy hipotetikus mesét, amely arról szól, hogy egy fizikus azt az úgynevezett kreatív kihívást elé állítja egy robottervező elé, hogy készítsen „elpusztíthatatlan robotot". A robotikus némi mesterkedés után előáll egy kis bundás teremtménnyel, leteszi az asztalra, majd egy kalapácsot ad a fizikus kezébe, és felszólítja, hogy pusztítsa el. A robot ide-oda szökdécsel, de amikor a fizikus felemeli a kalapácsot, azonnal a hátára fordul, panaszosan nyögdécsel, és a rémülettől tágra nyílt szemekkel, rettegve néz fel üldözőjére. A fizikus leteszi a kalapácsot. Az „elpusztíthatatlan robot" életben marad, mert hasznot húzott abból az emberi ösztönből, hogy a kisgyermekekre jellemző „aranyos" vonásokat felmutató lényeket védeni kell. Az MI jogainak kérdése tehát így is feltehető: Attól, hogy mi magunk nem vagyunk hajlandók lesújtani a kalapáccsal, mekkora lépést kell megtenni addig a követelésig, hogy ezt mások se tegyék meg?

Cynthia Breazeal legutóbbi teremtménye, amit a *Stan Winston Stúdióval*, Hollywood első számú automata szörnyetegeket előállító műhelyével együttműködve alkotott meg, figyelemre méltó mértékben hasonlít erre az elpusztíthatatlan robotra: Leonardo egy két és fél láb magas, bundás, teljesen automatikus működésű tedimackó. Egyike a mostanáig elkészült legkifejezőbb szociális robotoknak: harminckét motor mozgatja az arcát, tud látni, hallani, beszélni és érezni. A leginkább figyelemre méltó tulajdonsága az, hogy képes elsajátítani bizonyos készségeket. Ezt az emberek közvetlen utánzásával teszi, amelyhez fel kell fognia a tanuló és a tanító közötti hasonlóságokat, továbbá közvetlen személyes interakciókon keresztül is tanul, ami megkívánja, hogy megfelelő módon jelezni tudja, megértette-e azt, amit kell, vagy zavarban van. Ezek a fejlemények képviselik a következő, Kismet szociabilitását meghaladó evolúciós lépcsőt, és talán elvezetnek egy olyan pontig, amelynél a robot már képes lesz olyasmit is nyújtani, ami emlékeztet a barátságra. Breazeal kijelentette: „Ez a végső cél a szociális intelligencia kutatásában.".

Testetlen „társaság" érzékeléséhez elegendő gyors látogatást tennünk bármelyik csevegőszobában az interneten. Ahhoz azonban, hogy beszélgetőtársainkhoz valamilyen szinten érzelmileg is kötődni tudjunk, el kell képzelnünk, hogy valahol testi valójukban is jelen vannak. Breazeal kísérleteket végzett annak kimutatására, hogy az emberek mélyebb, intenzívebb érzelmi reakciókat mutatnak Leonardo iránt, ha az fizikailag is jelen van, mint amikor csupán kétdimenziós, nagy felbontású animációs képét látják egy számítógép képernyőjén.

A megtestesítés fontosságának jelentős következményei lehetnek a jogok szempontjából is. Egy testetlen számítógép képességeit a teljesítményén keresztül mérjük. Akkor juthatunk arra a következtetésre, hogy a számítógép tudatos, tehát jogokat vagy privilégiumokat biztosíthatunk számára, ha ez a teljesítmény – legyen az akár egy remek sakkjátszma vagy egy meggyőző beszélgetés – az emberéhez viszonyítva átlépi a funkcionális hasonlóság küszöbét. Ha azonban egy testtel bíró szociális robot esetében döntenénk úgy, hogy jogokat adunk neki, akkor ebbe valószínűleg éppen annyira belejátszanának a saját empatikus képességeink is, mint a robot belső felépítésére vonatkozó feltételezéseink. Döntésünk nemcsak attól függene, hogy a robot valójában micsoda vagy minek tartjuk, hanem attól is, hogy mit vált ki belőlünk.

Empátiánk vagy éppen önteltségünk, esetleg ösztönösen következetes erkölcsiségünk révén arra hajlunk, hogy olyan entitások számára kérjünk vagy adjunk bizonyos jogokat, amelyek hasonlítanak ránk, de megtagadjuk a jogokat azoktól a dolgoktól, amelyekről ez nem mondható el. Jelentős mennyiségű bizonyíték szól például amellett, hogy a delfinek képesek felismerni önmagukat a tükörben – ez egyike az öntudat kulcstesztjeinek, és olyan képesség meglétére utal a delfineknél, ami rajtuk kívül csak a nagy emberszabású majmokra és az emberre jellemző. Ám noha sokan javasolják, hogy a törvényes védelmet érdemlő tudatosság egyik alapvető kritériumaként az öntudatot kellene elfogadnunk, a delfinek nem élvezik ugyanazokat a törvény adta jogokat, mint a csimpánzok és a gorillák, amelyek fenotípusukat tekintve hasonlóbbak az emberekhez.

Az intelligens számítógépek tervezése – függetlenül attól, hogy a technológia milyen gyorsan fejlődik – teljes mértékig az ellenőrzésünk alatt áll. Ugyanez mondható el arról a védelemről és azokról a jogokról is, amelyeket garantálunk számukra. Csak akkor fogunk létrehozni a társadalom által jogok biztosítására érdemesnek tartott robotokat, ha saját akaratunkból így döntünk. Nem kell tehát véletlenszerűen létrejövő Frankensteinektől tartanunk.

Még ha a mesterséges intelligenciával rendelkező gépeknek nem is biztosítunk olyan jogokat, mint amelyeket a hipotetikus esküdtszék *BINA48* számára megítélt volna, valamilyenfajta törvényes védelmet bizonyára nyújtani fogunk számukra, mert arra a belátásra fogunk jutni, hogy ezek az emberi zsenialitás és kreativitás csúcsteljesítményét képviselik, és az ember kivételességének nem a cáfolatát, hanem inkább a visszatükröződését látjuk majd bennük. Anne Foerst lutheránus lelkész és az MI szakértője, aki teológiai konzultánsként közreműködött a Kismet-projektben, a szociális robotok kifejlesztésében az istenimádat egy típusát látja. *Isten a gépben: amit a robotok tanítanak nekünk az emberiségről és Istenről (God in the Machine: What Robots Teach Us About Humanity and God)* című, megjelenés előtt álló könyvében[4] elmondja, hogy a Brooks laboratóriumában általa tapasztaltak hogyan erősítették meg benne „az emberi rendszer hihetetlen bonyolultsága" iránti tiszteletet, ami őt „Isten 'legmagasabb rendű' teremtő aktusának ünneplésére" késztette.

Az efféle ünnepléssel elősegített védelem valószínűleg nem a jog nyelvén fog megfogalmazódni: Christopher Stone különféle rangsorokba állítható kategóriákat, illetve szempontrendszereket dolgozott ki az MI számára a jövőben biztosítandó „jogi

[4] 2005-ben megjelent. – *A szerk. megjegyz.*

státus" meghatározására. Az egyik lehetőség az, hogy az MI-vel rendelkező gépeket értékes kulturális produktumokként kezelhetnénk, mintegy „kiemelt státust" *(landmark status)* biztosítva számukra, bizonyos kikötésekkel a megőrzésükre, illetve a lerombolásukra vonatkozólag. Modellnek választhatnánk továbbá a veszélyeztetett fajok védelmét szolgáló törvényt is, amely bizonyos állatfajok számára nem elidegeníthetetlen jogaik, hanem „a nemzet és az emberek szemében általuk képviselt esztétikai, ökológiai, történelmi, rekreációs és tudományos értékek" tekintetbevétele alapján nyújt védelmet. Alkalmazhatnánk utilitárius érveket is védelmük indokolására, hasonlóan ahhoz, ahogyan Kant igazolta az állatoknak járó védelem bizonyos formáit, vagy ahogyan Jefferson érvelt a rabszolgák védelmében: hivatkozhatunk arra a lehetőségre, hogy ha nem biztosítjuk ezt a védelmet, akkor egyesek – a robotok nem megfelelő kezelésének láttán – rosszul bánhatnak az emberekkel is.

Az MI-vel kapcsolatos jogi kérdésekről való gondolkodás mindezekben az esetekben oda vezet, hogy szembe kell néznünk számos általunk létrehozott törvényes küszöb és jogi demarkációs vonal meghatározatlanságával. Ez egyrészt kijózanító, másrészt hasznos is. Ha úgy döntünk, hogy megtagadjuk a jogokat az MI-től, ez valójában pozitív lépést jelentene. Ha pedig úgy döntünk, hogy jogokat követelünk az MI számára, mindig tudatában kell lennünk e döntésünk erkölcsi és jogi következményeinek.

Jogrendszerünk „asimovizálása" még semmi esetre sem tekinthető a közeljövő feladatának. A számítógépeknek és robotoknak a szóbeli „dialóguskészségen" kívül még nagyon sok mindent meg kell tanulniuk tőlünk, hogy teljesebb mértékben emberivé váljanak. Ám az is igaz, hogy ugyanezen cél érdekében mi magunk is még sokat tanulhatunk tőlük.

Benjamin Soskis
Posztgraduális tanulmányokat folytat a Columbia Egyetem amerikai vallástörténet szakán. A *Legal Affairs* című folyóirat társszerkesztője. Korábbi tanulmányait a Yale Egyetemen végezte, ahol *Hősies számkivetettség: Frederick Douglass fejlődése az Atlanti-óceán túlpartján, 1845–1847 (Heroic Exile: The Transatlantic Development of Frederick Douglass 1845–1847)* című szakdolgozatával 1998-ban elnyerte a humán tudományokban benyújtott legjobb szakdolgozatnak járó Wrexham-díjat. 2000 és 2002 között a *The New Republic* című lap kutató riportere volt. Washingtonban él, mint ő mondja, „a Capitolium árnyékában".
E-mail: bjs2008@columbia.edu