



# **A mesterségesen intelligens robotok jogi státusza**

**Személyiség, adózás és ellenőrzés**

**A által**

**Filipe Maia Alexandre**

Erik Vermeulen professzor felügyelete alatt készült és a Tilburgi Egyetemen a nemzetközi üzleti jogból a Master of Laws (LL.M.) fokozat megszerzéséhez benyújtott disszertáció.

Védekezett a 12<sup>th</sup>június 2017

*"Mondd csak, Bernard. Ha kinyilvánítanád az emberségedet a világnak, mit gondolsz, mi fogadna téged? Talán egy futószalagos parádé? Mi emberek okkal vagyunk egyedül ezen a világon. Meggyilkoltunk és lemészároltunk mindent, ami megkérdőjelezte az elsőbbségünket."*

- Dr. Robert Ford, *Westworld* (HBO 2016).

## **Absztrakt**

A mesterséges intelligencia bevezetése az iparban és a társadalomban forradalmasítani fogja a jelenlegi társadalmi struktúrákat, és számos olyan szabályozási kihívást fog jelenteni, amelyekre a jogi keretek nem tudnak közvetlen választ adni. E valóságnak való megfelelés érdekében úgy értelmezzük, hogy a korlátozott memóriával rendelkező gépeket, a gondolkodáselmélettel rendelkező gépeket és az öntudatos gépeket a tulajdonosaiktól és felhasználóiktól különálló jogi személyeknek kell tekinteni. A mesterségesen intelligens ágenseknek külön jogi státusz tulajdonítása és e státusz tartalmának meghatározása, nevezetesen a felelősség, az esetleges jogok és az esetleges adózási kötelezettségek tekintetében, minimális bizonyosságot tesz lehetővé ezen új intelligens ágensek társadalomba való bevezetésének következményeit illetően, ami ellentétben áll az ezzel együtt járó nagyfokú kiszámíthatatlansággal. Az ilyen előre nem láthatósággal járó kockázatokat azonban továbbra is kezelni és mérsékelni kell, mivel ezek nemcsak az esetleges károkkal, hanem a személyes adatok védelmével és a közbiztonsággal is kapcsolatosak. Az emberek biztonságát vagy jólétét befolyásoló technológia tervezésekor nem elég csak feltételezni, hogy az működik.

## **Tartalomjegyzék**

<b>Absztrakt</b> .....	2
<b>1. Bevezetés</b> .....	3
1.1. Háttér és motiváció .....	3
1.2. Hozzájárulásunk.....	7
1.3. Fejezet összefoglalása.....	8
<b>2. A mesterséges intelligencia</b> .....	<b>meghatározása9</b>
2.1. Robotika vs. mesterséges intelligencia .....	9
2.2. A mesterséges intelligencia .....	típusai11

	2
2.3. Fejezet összefoglalása .....	13
<b>3. Az elektronikus személy.....</b>	<b>13</b>
3.1. A jogi keretek .....	elégtelensége13
3.2. Az elektronikus személy .....	fogantatása16
3.3. A gépekhez.....	való jogok elismerése21
3.4. Felelősség- és robotbiztosítás .....	26
3.5. Fejezet összefoglalása .....	30
<b>4. Fizessenek-e adót .....</b>	<b>a robotok?33</b>
4.1. A mesterséges intelligencia hatása a gazdasági rendszerekre .....	33
4.2. Robotadóztatás és egyéb cselekvési .....	lehetőségek36
4.3. Fejezet összefoglalása .....	39
<b>5. Az ellenőrzés .....</b>	<b>biztosítása41</b>
5.1. Az ellenőrzés .....	problémája41
5.2. Alapelvek; a Kill Switch .....	43
5.3. Elszámoltathatóság és átláthatóság .....	47
5.4. Fejezet összefoglalása .....	52
<b>6. Záró megjegyzések.....</b>	<b>55</b>
6.1. Következtetések .....	55
6.2. Jövőbeni munka .....	59
<b>Hivatkozások listája.....</b>	<b>60</b>

## 1. Bevezetés

### 1.1. Háttér és motiváció

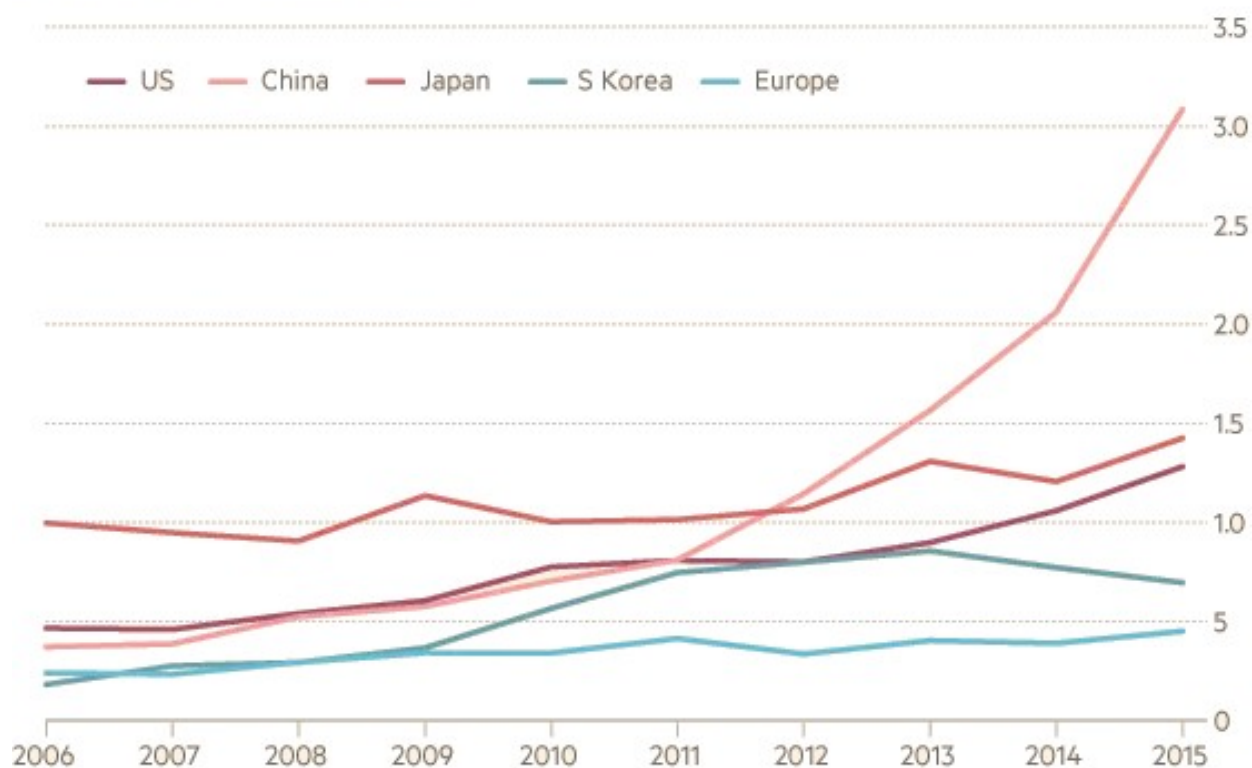
Internet, fenntartható energia, genetikai átprogramozás, több bolygón való élet és mesterséges intelligencia. Amikor a jövőről kérdezték, a jövőbe tekintő Elon Musk ezeket az öt dolgot jelölte meg.



olyan kulcsfontosságú területek, amelyek hatással lesznek az emberiségre és megzavarják fajunk életszínvonalát<sup>1</sup>. Míg az internetet és a fenntartható energiát már alaposan kifejlesztették, a listán szereplő 3 további pontot még nem sikerült elsajátítani. Kétségtelen azonban, hogy jelenleg a vihar előtti nyugalom utolsó éveit éljük, mivel egy új ipari forradalom közeleg. Az Európai Parlament Jogi Bizottsága (JURI) megfogalmazása szerint "Mary Shelley Frankenstein szörnyétől kezdve a klasszikus Pygmalion mítoszán, a prágai gólem történetén át a szót megalkotó Karel Čapek robotjáig az emberek az intelligens gépek megépítésének lehetőségéről fantáziáltak"<sup>2</sup>. Az elmúlt néhány évben az emberiségben rejlő kíváncsiság gyorsan azon dolgozik, hogy ezt a fantáziát valósággá tegye, amint azt a piac viselkedése is mutatja.

## Total robotic patent applications by location

Number of patents submitted ('000)



Source: IFI

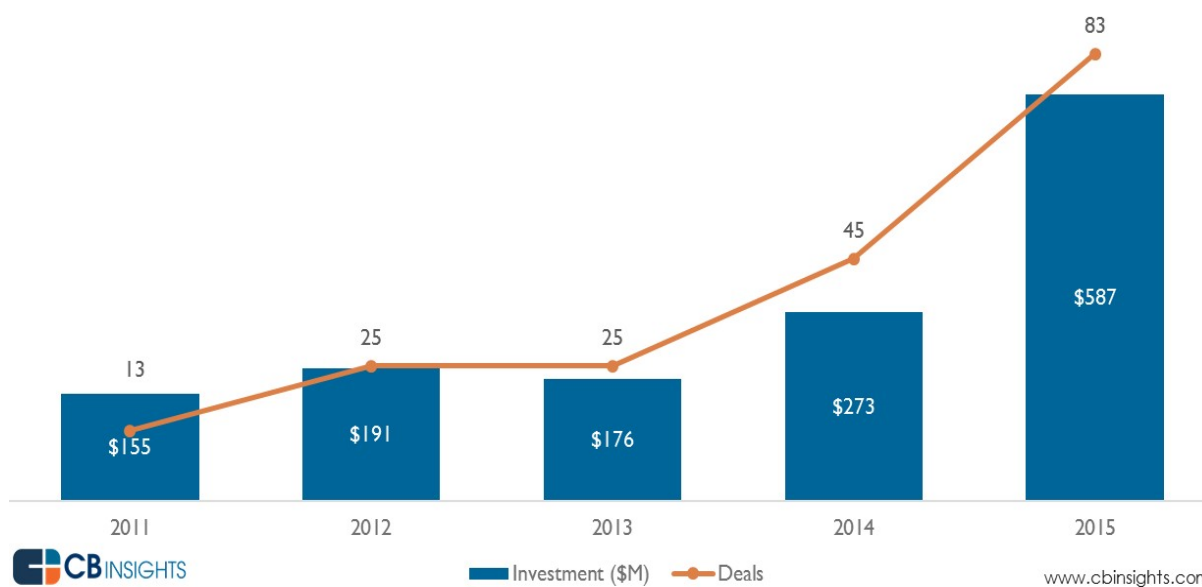
Ábra - 1Éves szabadalmi bejelentések a robotika területén

<sup>1</sup> Elon Musk, interjú Sam Altmannal, "Y Combinator's How To Build The Future Series" (2016).

<sup>2</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

Az elmúlt évtizedben a robottechnológiára vonatkozó éves szabadalmi bejelentések megháromszorozódtak. 2014 és 2010 között a robotikai ipar forgalmának átlagos növekedése évente 17% volt, és 2014-ben volt a legnagyobb éves növekedés (29%)<sup>3</sup>. A robotikába történő kockázati tőkebefektetések 2014-ről 2015-re több mint kétszeresére nőttek, és elképesztő 587 millió dollárt tettek ki (a CB Insights kutatócég szerint)<sup>4</sup>, és az előrejelzések szerint 2019-re<sup>5</sup> a robotikai ipar globális összpiaci értéke mintegy 135 milliárd dollár lesz. Ezek a számok azt mutatják, hogy bár a robotikai iparágba irányuló tőkeáramlás még csak kezdeti stádiumban van, ez a terület kétségtelenül az egyik legforróbb piac lesz nemcsak a technológiai iparban, hanem az általános kereskedelemben is.

### Robotics (ex-drones): Yearly Global Financing History 2011-2015



Ábra - 2Éves kockázati tőke-befektetések a robotikában

A robotikai ipar vonzereje a fejlesztésétől várt kettős előnyökkel magyarázható. Vállalati szinten a bevezetés és a fejlődés a

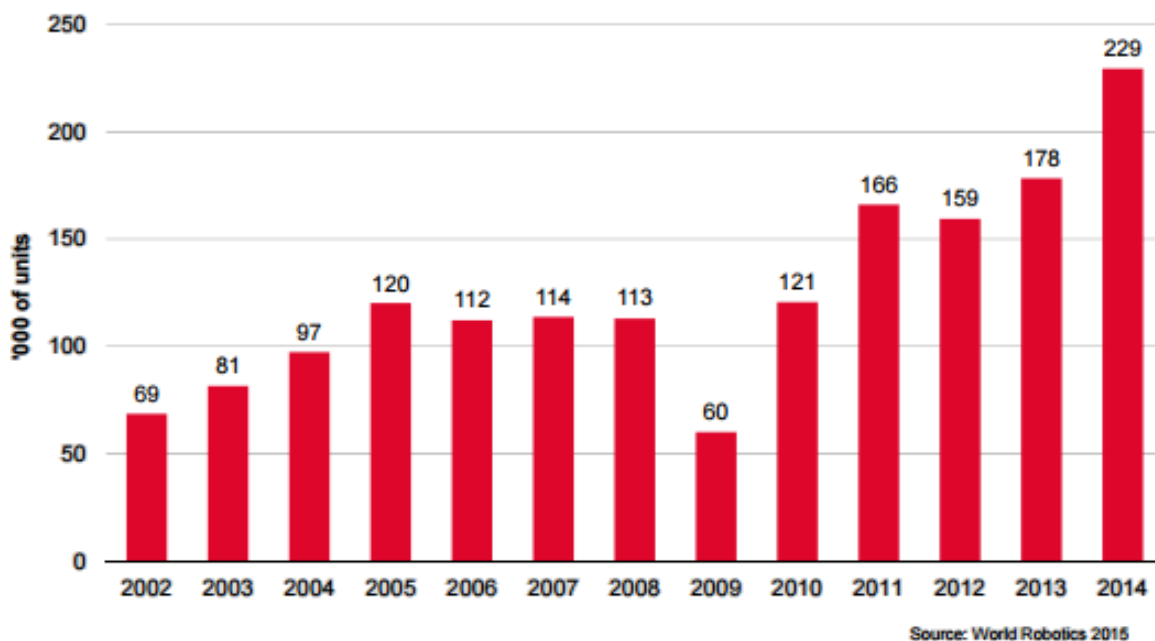
<sup>3</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

<sup>4</sup> Richard Waters és Tim Bradshaw, "A robotok felemelkedése befektetési boomot indít el" Financial Times (2016) < <https://www.ft.com/content/5a352264-0e26-11e6-ad80-67655613c2d6> > hozzáférés április 5. 2017.

<sup>5</sup> Jing Bing Zhang és mások, "IDC FutureScape: (International Data Corporation 2016).2017

a robotika javítja a hatékonyságot és növeli a megtérülést<sup>6</sup> a költségmegtakarítások ("a robotok jellemzően egy teljes munkaidős alkalmazott költségének kilenczetedét teszik ki"<sup>7</sup>), valamint a pontosabb, jobb minőségű és megszakítás nélküli munkavégzés révén, ami növeli a termelékenységet és csökkenti a hibaarányt. Továbbá a gépek sérülések és fáradtság nélkül képesek veszélyes körülmények között is dolgozni, ami növeli az általános munkahelyi biztonságot. Társadalmi szinten a robotika számos kulcsfontosságú területen fog hozzájárulni a fejlődéshez, például a közlekedés (a közlekedés optimalizálásával, a forgalom csökkentésével vagy a biztonság növelésével), az orvosi ellátás (a hibalehetőségek csökkentésével, az eljárások felgyorsításával vagy a gyakorlatilag korlátlan tudásmegosztás lehetővé tételével), az élelmiszertermelés (az éghajlati adatok és a korábbi körülmények alapján a megfelelő ültetési idő előrejelzésével vagy a folyamatok automatizálásával) és még sok más területen.

### Estimated worldwide annual supply of industrial robots



Ábra - Az ipari robotok 3 éves értékesítése

Bár elméletileg a robotika és a mesterséges intelligencia lehetővé teheti az emberiség számára, hogy beteljesítse a végső jólétre való törekvését, az út, amely jelenleg elválaszt minket ettől, hosszú és homályos. Mint minden más

<sup>6</sup> Befektetés megtérülése = (Befektetési nyereség - Befektetési költség) / Befektetési költség

<sup>7</sup>

< <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-finance-robots-are-coming.pdf>>  
hozzáférés április 6. 2017.

jelentős forradalom az iparban, ez a forradalom is számos kihívással jár, amelyeket le kell küzdeni ahhoz, hogy sikeresen végrehajtsuk. Az intelligencia megteremtésével az emberiség egy eddig ismeretlen területre lép, és olyan utat jár be, amelyen még soha nem járt. "A mesterséges intelligencia autonóm jellege az előreláthatóság és az ellenőrzés olyan kérdéseit veti fel, amelyek az *utólagos* szabályozást hatástalanná tehetik"<sup>8</sup>. Mivel az AI potenciálisan gyökeresen átalakíthatja az általunk ismert társadalmat, a proaktív megközelítés lehet az egyetlen módja a fenntarthatóság biztosításának. De hogyan lehet proaktívan megközelíteni valami példátlan dolgot? Ezt a kérdést tűztük ki magunk elé.

## 1.2. Hozzájárulásunk

A közelgő mesterséges intelligencia ipari forradalma nagymértékben javíthatja a társadalmi jólétet. Ezzel egyidejűleg érkezése nagyfokú bizonytalanságot hordoz a következményeit és az iparra és a társadalomra gyakorolt hatását illetően, az egyetlen biztos dolog az, hogy egyetlen réteg sem marad érintetlenül. Míg ez ijesztő lehet azok számára, akik félnek a változástól, vagy nem tudnak jól megbirkózni vele, azok számára, akik elfogadják, minden bizonnyal izgalmas. Mindenesetre a mesterséges intelligencia forradalma elkerülhetetlennek tűnik, és várhatóan soha nem látott kihívások elé állítja az emberiséget. Ezért megértjük, hogy nemcsak elfogadni kell, hanem proaktív és innovatív gondolkodással és vitákkal vezető szerepet kell vállalni a fenntartható megvalósításában.

A mesterséges intelligencia bevezetése és elterjedése a kihívások három fő kategóriáját fogja összefogni: a technológiai kihívásokat, az ipari kihívásokat és a szabályozási kihívásokat.

A technológiai kihívások a mesterséges intelligencia fejlesztésével, valamint a mesterségesen intelligens ágensek képességeinek és tulajdonságainak javításával kapcsolatosak. Ezekkel a kihívásokkal főként mérnökök, programozók, idegtudósok és más, a technológiához értő szakemberek fognak foglalkozni.

Az ipari kihívásokat a meglévő piacok túlnyomó többségének technológiai felforgatása fogja vezérelni, és az elavulás veszélyével fenyegeti azokat, akik nem képesek lépést tartani a technológia gyors ütemével, és nem képesek folyamatosan alkalmazni a legújabb fejlesztéseket a vállalkozásukban. Ezek a kihívások a nagy, már működő vállalkozásokat nehezebbé teszik,

---

<sup>8</sup> Matthew U. Scherer, "A mesterséges intelligencia rendszerek szabályozása: (2016) Harvard Journal of Law &

Technology.

7

amelyek kevésbé rugalmasak és nehezebben alkalmazkodnak<sup>9</sup>, míg az induló vállalkozások és a kisebb cégek, amelyeknek kevesebb gondot okoz a mesterséges intelligens rendszerek bevezetése, megpróbálnak behatolni a piacukra<sup>10</sup>. Ezekkel a menedzserek vagy a vállalatok *innovációs vezetői* fognak foglalkozni.

Végül a szabályozási kihívások az országok makrogazdasági politikájának és jogi kereteinek reformjával kapcsolatosak annak érdekében, hogy az új intelligens ágensek bevezetését a társadalomban lehetővé tegyék. Ezekkel a kihívásokkal a politikai döntéshozók és a törvényhozók fognak foglalkozni, azonban itt a jogi szakemberek (mint például mi is), akik be akarnak kapcsolódni, hasznosabbá tehetik magukat, és társadalomtudományos megközelítéssel járulhatnak hozzá a vitához.

Az egyik szabályozási kihívás, amellyel a jogalkotóknak szembe kell nézniük, a jogi keretek reformja annak érdekében, hogy a mesterségesen intelligens ügynökök jelenlétét a társadalomban és a mindennapi életben figyelembe lehessen venni. Amikor ezek az ágensek az emberek beavatkozása nélkül, sőt, végső soron a tudatosságuk nélkül kezdenek el működni, kérdések merülnek fel a tetteikért való felelősséggel, az ilyen ágensek feltételezett jogokkal való felruházásával, illetve a biztonság, az adatvédelem és más, a növekvő függetlenségük és autonómiájuk által fenyegetett érdekek biztosítása érdekében szükséges korlátozásokkal kapcsolatban. Egy másik kihívás, amellyel a politikai döntéshozóknak előreláthatólag szembe kell nézniük, az emberi munkaerő nagy részének robotokkal való felváltásával kapcsolatos, ami "aggodalmakat vet fel a foglalkoztatás jövőjével és a társadalombiztosítási rendszerek életképességével kapcsolatban, ha az adózás jelenlegi alapja megmarad, ami a vagyon és a befolyás elosztásában az egyenlőtlenségek növekedésének lehetőségét teremti meg"<sup>11</sup>. Ebben a munkában arra törekszünk, hogy megfelelően megfogalmazzuk ezeket a kihívásokat, és meghatározzuk azokat a lehetséges válaszokat, amelyek lehetővé teszik a politikai döntéshozók és a jogalkotók számára, hogy sikeresen és fenntartható módon reagáljanak rájuk.

### 1.3. Fejezet összefoglalása

---

<sup>9</sup> Erik Vermeulen, "Hogyan készülünk fel az automatizálásra? Or, Why We Need More "Artificial Intelligence Ecosystems" Now! < <https://hackernoon.com/how-to-prepare-for-automation-or-why-we-need-more-artificial-intelligence-ecosystems-now-4a4a767e733b>> hozzáférés április 10-ban. 2017.

<sup>10</sup> További információkat arról, hogy mit tehetnek a nagy, már működő vállalkozások, hogy a mesterséges intelligencia, a robotika és az automatizálás előtt járjanak és sikeresen vegyenek részt a mesterséges intelligencia, a

robotika és az automatizálás kezelésében, lásd Erik Vermeulen: "Hogyan készüljünk fel az automatizálásra? Or, Why We Need More "Artificial Intelligence Ecosystems" Now!", elérhető a <https://hackernoon.com/how-to-prepare-for-automation-or-why-we-need-more-artificial-intelligence-ecosystems-now-4a4a767e733b> címen (hozzáférés: 2017. április 10.). <sup>11</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

A mesterséges intelligencia bevezetése az iparban és a társadalomban forradalmasítja a jelenlegi társadalmi *status quo-t*, és annak ellenére, hogy nagymértékben javíthatja a jólétet és az életminőséget, számos technológiai, ipari és szabályozási kihívással jár. E munka célja, hogy felvázolja a főbb szabályozási kihívásokat, valamint meghatározza és továbbfejlessze az ezekre a kihívásokra válaszoló jelenlegi javaslatokat. A következő fejezet a *robotika* és a *mesterséges intelligencia* fogalmait fejt ki, amelyeket munkánk céljaira fogadtunk el, míg az ezt követő fejezetek mindegyike az említett szabályozási kihívásokkal foglalkozik, különös tekintettel a jogi keretek mesterségesen intelligens ágensekhez való igazítására, valamint az adóstruktúrák és a társadalombiztosítási rendszerek fenntarthatóságára a robotizált munkaerő tekintetében.

## 2. A mesterséges intelligencia meghatározása

### 2.1. Robotika vs. mesterséges intelligencia

Eddig a *robotika* és a *mesterséges intelligencia* kifejezéseket minimális szigorral használták. Mivel azonban egyre mélyebben kezdünk belemerülni a témába, fontos, hogy szigorúan definiáljuk, mit értünk ezeken a kifejezéseken, és ebből következően azt is, hogy mi a jelen munka tárgya.

Annak ellenére, hogy a *robot* szót a köznyelv a sétáló és beszélő gépekkel azonosítja, ez az asszociáció manapság aligha lehetne félrevezetőbb. A *robot* szó eredetét általában Karel Čapeknek tulajdonítják, aki a *Rossumovi Univerzální Roboti* (Rossum univerzális robotjai) című darabjában vezette be a prágai gólemre utalva 1920,. A szó a cseh és szlovák *robota* szóból származik, amely viszont a kemény munka vagy rabszolgaság<sup>12</sup> jelölésére használt proto-szláv *orbota* szóból ered. A *robot* szó megfelelő használata ma már sokkal közelebb áll etimológiai eredetéhez: a *robotika* a "robotizált folyamatok automatizálására" utal, mint "ismétlődő és gyakran szabályokon alapuló folyamatok automatizálására"<sup>13</sup>. Munkánk szempontjából ez magában foglalja, de nem kizárólagosan, a számítógépen kódolt szoftvereket és programokat, amelyek az ismétlődő, szabályokon alapuló feladatokat végző embereket helyettesítik, függetlenül attól, hogy az ilyen teljesítményt vagy nem

---

<sup>12</sup> 'Definition Of Robot In English' (Oxford Dictionaries, 2017) < <https://en.oxforddictionaries.com/definition/robot> > hozzáférés április 6. 2017.

<sup>13</sup>

< <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-finance-robots-are-coming.pdf>>  
hozzáférés április 6. 2017.

fizikai gépek által. Ugyanezt a logikát követve, az emberi kezdeményezéstől függő egyszerű feladatokat végző gépek, mint például a papír aprítása vagy az étel melegítése, nem tartoznak a fogalom körébe.

*A mesterséges intelligenciát* viszont sokkal nehezebb meghatározni. Nagyjából a gépek által mutatott intelligenciára utal. De hogyan definiáljuk az intelligenciát? Az emberi intelligencia és a gépi intelligencia ugyanaz? Képes-e egy gép ugyanolyan tudatosságot mutatni, mint az ember?

A történelem során számos gondolkodó próbált választ adni ezekre a kérdésekre. Az egyik legismertebb, ha nem a legismertebb választ Alan Turing adta meg 1950-ben, a *Computing Machinery and Intelligence* című híres tanulmányában. Turing azt javasolja, hogy a kérdés ne az legyen, hogy egy gép képes-e gondolkodni, hanem az, hogy egy gép képes-e meggyőzni egy embert arról, hogy képes gondolkodni. Ehhez a gépnek át kellene mennie a *Turing-teszten*, amely abból áll, hogy egy olyan embert, aki nem tudja, hogy egy géppel kommunikál, el kell hitetni azzal, hogy egy másik emberrel<sup>14</sup> kommunikál. Turing szerint, ha egy gép képes olyan intelligensen viselkedni, mint egy ember, akkor ugyanolyan intelligens, mint egy ember. Turingnak ez a felvetése nagy befolyással bírt, de egyben nagyon ellentmondásos is. Annak ellenére, hogy az elmúlt évtizedekben számos alternatív tesztet javasoltak, valamint a *Turing-teszt* számos változatát (például a híres *Feigenbaum-tesztet*<sup>15</sup>, amely durván a *Turing-teszt tárgy*i szakértői változatából áll), a *Turing-teszt* továbbra is meghatározó marad a mesterséges intelligencia létezésének megítélésének hatalmas területén. Mindazonáltal munkánk lényege nem az, hogy pontosan körülhatároljuk azokat a forгатókönyveket, amelyekben a mesterséges intelligencia létezik vagy nem létezik, hanem az, hogy foglalkozzunk a létezésének és a társadalomban való bevezetésének következményeivel. Emiatt, bár a *Turing-teszt* kétségtelenül említésre méltó, mi egy funkcionálisabb megközelítést választunk, amely - mindent egybevetve - teret enged az értelmezésnek.

A szakirodalomban a *mesterséges intelligencia* többféle definíciót találunk. Egyesek a környezet észlelésének és az adott cél elérésének esélyét maximalizáló cselekvésnek a képességével kapcsolatosak. Mások az általában az emberi értelemhez kapcsolódó kognitív funkciók, például a tanulás és a problémamegoldás<sup>16</sup> birtoklását jelentik. E munka céljaira a *mesterséges intelligencia* tág fogalmát tekintjük, amelyet nem egyetlen fogalom határoz meg, hanem a következő tulajdonságokra adunk példákat

---

<sup>14</sup> Alan Turing, "Computing Machinery And Intelligence" [1950] Mind.

<sup>15</sup> Edward A. Feigenbaum, "Some Challenges And Grand Challenges For Computational Intelligence" [2003] Journal of the ACM.

<sup>16</sup> Stuart J. Russell és Peter Norvig, Mesterséges intelligencia: A Modern Approach (3<sup>rd</sup>edn, Prentice Hall 2009).

a szakirodalom általánosságban az intelligens ügynökökhöz társítja. Ezek a tulajdonságok a következők: gondolkodás, problémamegoldás, tudás reprezentációja, tervezés, tanulás, természetes nyelvi feldolgozás, érzékelés, tárgyak mozgatása és manipulálása, szociális intelligencia és kreativitás<sup>17181920</sup>. Ahhoz, hogy egy adott ágens megfeleljen a hatókörének, nem szükséges, hogy e tulajdonságok<sup>21</sup> mindegyikével rendelkezzen. Csupán egy vagy több tulajdonságot kell mutatnia közülük olyan intenzitással, amely minimálisan elegendő ahhoz, hogy az emberi értelemmel való összehasonlítást indokoltá tegye.

## 2.2. A mesterséges intelligencia típusai

A természetben az intelligencia különböző intenzitással nyilvánul meg. Vegyük például egy házi kutya esetét: bár nem képes érvelni vagy logikát használni, korlátozottan képes problémamegoldásra, és határozottan képes megtanulni, hogyan reagáljon bizonyos parancsokra, például újságért vagy földön gurulásra. Ugyanez a logika érvényes a mesterséges intelligencia területére is: a mesterséges intelligencia (és az általában hozzá kapcsolódó tulajdonságok) erősebb vagy gyengébb intenzitással jelentkezhetnek. Ezt a koncepciót jól kifejti a *The Conversation UK*<sup>22</sup> című brit lapban megjelent cikk. E cikk szerint a mesterséges intelligencia megnyilvánulásai a következő négy kategória egyikébe sorolhatók: "reaktív gépek", "korlátozott memória", "elmélet" és "öntudat".

Ahogy a név is jelzi, a reaktív gépek olyan rendszerekből állnak, amelyek tisztán reaktív módon működnek, nincsenek emlékeik, és nem képesek a múltbeli tapasztalatokat felhasználni a jelenlegi döntések befolyásolására, ezért minden alkalommal ugyanúgy viselkednek, amikor ugyanazzal a helyzettel találkoznak. Nincs fogalmuk arról, hogy

---

<sup>17</sup> Stuart J. Russell és Peter Norvig, *Mesterséges intelligencia: A Modern Approach* (3<sup>rd</sup>edn, Prentice Hall 2009).

<sup>18</sup> George F. Luger, *Mesterséges intelligencia: (5<sup>th</sup>edn, Addison-Wesley 2004).*

<sup>19</sup> David Poole, Alan Mackworth és Randy Goebel, *Számítógépes intelligencia: A Logical Approach* (1<sup>st</sup>edn, Oxford University Press 1998).

<sup>20</sup> Nils J. Nilsson, *Mesterséges intelligencia: Morgan Kaufmann Publishers 1998): A New Synthesis* (1<sup>st</sup>edn, Morgan Kaufmann Publishers 1998).

<sup>21</sup> Ray Kurzweil *A szellemi gépek kora* (1999) című könyvében tárgyalja a mesterséges általános intelligencia ("erős mesterséges intelligencia") témáját, amely durván abban áll, hogy egy gép képes az összes említett tulajdonságot egyesíteni, és sikeresen elvégezni bármely olyan szellemi feladatot, amelyre egy ember képes.

<sup>22</sup> Arend Hintz, "Az AI négy típusának megértése, a reaktív robotoktól az öntudatos lényekig" [2016] *The Elektronikus* elérhető a következő címen:

<https://ssrn.com/abstract=2985466>

Conversation UK < <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>> hozzáférés április10ban 2017.

a világot, ami azt jelenti, hogy nem tudnak a programozott feladatokon túl működni. A Google AlphaGO-ja<sup>23</sup> és az IBM Deep Blue-ja<sup>24</sup> példák az ilyen típusú gépekre.

A korlátozott memóriával rendelkező gépek olyan gépek, amelyek bizonyos kulcsfontosságú objektumok azonosításával és időbeli nyomon követésével képesek a múltba tekinteni. Ezeket a megfigyeléseket aztán hozzáadják a gépek előre programozott reprezentációihoz a világról, és ennek következtében figyelembe veszik őket a döntéshozatali folyamatokban. Ezek a gépek azonban csak annyi memóriával vagy tapasztalattal rendelkeznek, amennyi a döntések meghozatalához és a megfelelő cselekvések végrehajtásához szükséges. Az önvezető autók például képesek megfigyelni más autók sebességét és irányát, és ezt az információt felhasználva döntenek arról, hogy mikor kell sávot váltaniuk, hogy elkerüljék, hogy egy másik sofört elvágjanak, vagy hogy egy közeli autó elüsse őket. A személyi asszisztensek a korlátozott memóriájú gépek másik esete. Mark Zuckerberg szerint Jarvis nevű személyi asszisztense úgy tudja eldönteni, hogy automatikusan kinyissa-e a bejárati ajtót a látogatók előtt, hogy a látogatók arcát letapogatja, egy arcfelismerő szoftvert futtat a személy azonosítására, majd az eredményeket összeveti Mark Zuckerberg várható látogatóinak<sup>25</sup> listájával .

A harmadik kategória - elmélet - a pszichológiában használt fogalomról kapta a nevét, amely azt a felfogást írja le, hogy a világban élő emberek, lények és tárgyak gondolatokkal és érzelmekkel rendelkezhetnek, amelyek befolyásolják saját viselkedésüket<sup>26</sup>. Az ebbe a kategóriába tartozó gépek képesek lennének reprezentációkat alkotni a világról, valamint más ágensekről és entitásokról, és viselkedésüket mások érzelmeinek, elvárásainak, motivációinak és szándékainak megértése alapján alakítanák.

Végül az öntudatosság az intelligencia mesterséges létrehozásának végső szakaszát írja le: olyan rendszerek építését, amelyek képesek reprezentációkat alkotni önmagukról. Ebben a szakaszban a gépek tudatosak, érzőek és képesek lennének megérteni mások érzéseit. A gépek nemcsak azt tudnák, hogy mit akarnak, hanem képesek lennének megérteni, hogy mit *akarnak*, és *miért* akarják azt.

A mesterséges intelligencia e két utóbbi típusára még mindig csak a fikcióban találhatunk példákat. Ami az elmeelmélettel rendelkező gépeket illeti, ezeket könnyen el lehet képzelni a *Csillagok háborúja*-saga C-3PO és R2-D2 példájával. Ami az öntudatos gépeket illeti, egy érdekes példány lenne az EVA,

---

<sup>23</sup> További információ a <https://deepmind.com/research/alphago> oldalon található.

<sup>24</sup> További információ a <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue> oldalon található.

MarkZuckerberg,'BuildingJarvis'(Facebook,2016)<<https://www.facebook.com/notes/mark-zuckerberg/building-jarvis/10154361492931634>> hozzáférés április 10-ban. 2017.

<sup>26</sup> David Premack és Guy Woodruff, "Van-e a csimpánznak elméje? (1978) Behavior and Brain Sciences.

az *Ex Machina* című filmből. Továbbá, az utat, amely elválasztja az elméletet a valódi öntudattól, gyönyörűen ábrázolja (spoilerveszély!) az HBO *Westworld* című tévésorozata.

### 2.3. Fejezet összefoglalása

*A robotika és a mesterséges intelligencia* fogalmaknak különböző jelentése van. Munkánk célja nem a mesterséges intelligencia pontos definíciójának megadása, hanem a mesterséges intelligencia létezésének és társadalmi bevezetésének következményeivel való foglalkozás. Ezért a mesterséges intelligencia egyetlen fogalmának használata helyett úgy döntöttünk, hogy a mesterségesen intelligens ágensekhez általánosan társított tulajdonságokra adunk példákat. Ezek a tulajdonságok a következők: gondolkodás, problémamegoldás, tudás reprezentációja, tervezés, tanulás, természetes nyelvi feldolgozás, érzékelés, tárgyak mozgatása és manipulálása, szociális intelligencia és kreativitás. Ha ezek közül egy vagy több tulajdonságot olyan intenzitással mutat, amely minimálisan elegendő ahhoz, hogy az emberi értelemmel való összehasonlítást indokolja, akkor egy adott ágens beleillik e munka tárgykörébe.

A mesterséges intelligenciának négy típusát ismerjük el: reaktív gépek, korlátozott memóriával rendelkező gépek, elmélettel rendelkező gépek és öntudattal rendelkező gépek. Míg a reaktív gépek már régóta léteznek (az IBM Deep Blue-ja 1997 májusában legyőzte Gary Kasparovot), az emberiség csak most kezdi felfedezni a mesterséges intelligencia második szintjét, a korlátozott memóriával<sup>27</sup> rendelkező gépeket. Úgy tűnik, hogy az elmélettel és öntudattal rendelkező gépek létrehozása még mindig messze van tőlünk<sup>28</sup>. Mindenesetre a mesterséges intelligencia különböző típusainak létezésének felismerése és e technológia fejlődésének következő szakaszainak megértése elengedhetetlen a szabályozásához.

## 3. Az elektronikus személy

### 3.1. A jogi keretek elégtelensége

---

<sup>27</sup> Mindazonáltal az e téren elért eredmények mindenképpen izgalmasak!

<sup>28</sup> Az Amerikai Egyesült Államok Fehér Házának 2016 októberében közzétett, a mesterséges intelligenciáról szóló jelentése azt állítja, hogy az elkövetkező 20 évben az emberiség valószínűleg nem fogja látni, hogy a gépek "az emberéhez hasonló vagy azt meghaladószéles körben alkalmazható

intelligenciával rendelkeznének

"

Ez a jelentés elérhető a  
[https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf) címen (hozzáférés: 102017. április).

13

Amint azt az előző fejezetekben tárgyaltuk, a mesterséges intelligencia már most is lassan meghonosodik a társadalomban, és várhatóan egyre jobban elterjed, akár úgy gondoljuk, hogy ez több, akár kevesebb időt vesz igénybe. Ezzel együtt nagyon fontos feltenni a kérdést, hogy a jelenlegi jogi keretek készen állnak-e arra, hogy befogadják ezt a valóságot, vagy ehelyett kiigazításokra lesz szükségük ahhoz, hogy megőrizték koherenciájukat és következetességüket.

A kérdés relevanciájának jobb megértéséhez kérjük, gondoljon a következő forgatókönyvre<sup>29</sup>: amikor Johnnak szüksége van a vállalkozásához kellékekre, időt tölt azzal, hogy kapcsolatba lépjen ismert szállítóival, és tárgyalásokat folytasson a szállítási szerződésről. A vállalkozásának hatékonyabbá tétele érdekében John bevezet egy olyan rendszert, amely figyelemmel kíséri vállalata készleteit, és amikor azok fogyóban vannak, a különböző szállítók feltételeinek összehasonlításával és az egyikükkel való megrendeléssel szállítási szerződést köt. Most képzeljük el, hogy Jane, egy szállító, bevezetett egy olyan rendszert, amely úgy kezeli a hozzá érkező megrendeléseket, hogy figyelemmel kíséri a készleteit, és elfogadja a megrendeléseket, amikor a készlet szint elegendő a teljesítéshez. Egy bizonyos napon John rendszere leadott egy megrendelést Jane-nek, akinek a saját rendszere beleegyezett a teljesítésbe. Így a rendszerek megállapodást kötöttek, mielőtt John és Jane egyáltalán tudomást szerzett volna annak létezéséről.

Hogyan alkalmazkodnak jelenleg a jogi keretek ehhez a helyzethez? Végrehajtható szerződés előtt állunk? A konzervatív megközelítés valószínűleg azzal válaszolna, hogy a gépek nem lehetnek szerződő felek, így a szerződés nem is létezne. Azonban, még mindig ugyanazt a példát követve, gondoljunk arra, hogy ugyanazon az estén, miután észrevette, hogy a készletei fogyóban vannak, John (akinek a mottója: "jobb félni, mint megijedni") bejelentkezett a számítógépébe, és észrevette, hogy a rendszer leadott egy rendelést, amelyet egy szállító (Jane) megkapott és elfogadott. Megnyugodva, John hazament, és nyugodtan aludt egy éjszakát. John nem törődik azzal, hogy Jane egy rendszert használ-e a megrendeléseinek kezelésére vagy sem. Van-e Johnnak ésszerű elvárása arra, hogy megkapja az árut?

Ha John maga adta volna le a rendelést, Jane mentesülne-e a teljesítés alól, mert a rendelést a rendszere fogadta el, nem pedig ő maga kézzel? Ellenkezőleg, ha Jane kézzel fogadta volna el a megrendelést, mentesülne-e a teljesítés alól, mert a megrendelést John rendszere adta fel? Vagy indokolt-e felmenteni Jane-t a teljesítés alól, mert a kommunikációban mindkét fél a rendszer volt, annak ellenére, hogy John figyelmen kívül hagyja Jane rendszerének létezését? A válasz minden esetben, még a legkonzervatívabb elmék számára is, nemlegesnek tűnik. John

---

<sup>29</sup> Benjamin D. Allgrove, "A mesterséges értelmek jogi személyisége: Pragmatikus megoldás vagy sci-fi? (Master, Oxfordi Egyetem 2004).

ésszerű elvárása van arra, hogy az árut megkapja, és Jane-t kötelezné a teljesítés. De akkor hogyan alakítsuk ki ezt a szerződést a jelenlegi jogi keretek fényében?

Az egyik lehetséges megközelítés a rendszer pusztán szerződéskötési<sup>30</sup> vagy kommunikációs<sup>31</sup> eszközként való felfogásán alapul. E megközelítés szerint a szerződés közvetlenül John és Jane között jön létre. Ennek a megközelítésnek az az előnye, hogy könnyen bevezethető a jogi keretrendszerekbe, anélkül, hogy nagyobb változtatásokra lenne szükség, akár jogszabály, akár esetjog, akár doktrinális megfontolások révén. Másrészt arra a fikcióra támaszkodik, hogy "minden, ami a számítógépből származik, valójában közvetlenül az emberi irányítótól származik"<sup>32</sup>, teljesen figyelmen kívül hagyva a rendszer esetleges autonómiáját. Továbbá azzal, hogy konszenzust feltételez a felek között, akik esetleg nem is tudnak arról, hogy a szerződést megkötötték, vagy hogy a másik fél létezik, ez a megközelítés megfosztja a szerződés létrejöttét annak legfontosabb elemétől: az akaratok találkozásától.

Egy másik megközelítés ebben az esetben a rendszer magatartását a munkavállaló magatartásával tenné egyenlővé. E megközelítés szerint a szerződés John egyik jogi képviselője és Jane egyik jogi képviselője között jönne létre. A fél szemében mi a különbség, hogy egy alkalmazott működteti a másik fél számítógépet, vagy az önmagát? Ennek a megközelítésnek az az előnye, hogy nem támaszkodik semmilyen vélelemre, és nem hajlítja a szerződéskötési elveket. Továbbá lehetővé teszi John és Jane számára, hogy bármilyen védekezésre hivatkozassanak abban az esetben, ha valamelyik alkalmazottjuk valóban a szerződést kötötte, ahelyett, hogy őket tekintenék a megállapodás közvetlen szerződő felének. Ez a megközelítés azonban azt jelenti, hogy a jogalkotó választási lehetőséget választja annak érdekében, hogy John és Jane rendszereit a tulajdonosaiktól és felhasználóiktól különálló jogi személyeknek tekintsék.

Az előző ismertetéssel azt kívántuk bemutatni, hogy a mesterséges intelligencia elterjedésével kérdések merülnek fel, és a jogi kereteknek elkerülhetetlenül alkalmazkodniuk kell. Úgy véljük, hogy "minél autonómabbak a robotok, annál kevésbé tekinthetők egyszerű eszköznek más szereplők kezében"<sup>33</sup>. Ha manapság kevés értelme van az ilyen rendszereket egy egyszerű telefonhoz vagy faxgéphez hasonlítani, a jövőben még kevesebb értelme lesz. És míg a bemutatott példában az ügyek

---

<sup>30</sup> Benjamin D. Allgrove, "A mesterséges értelmek jogi személyisége: Pragmatikus megoldás vagy sci-fi? (Master, Oxfordi Egyetem 2004).

<sup>31</sup> Tom Allen és Robin Widdison, 'Can Computers Make Contracts?' (Tudnak-e a számítógépek szerződéseket kötni?) (1996) Harvard Journal of Law & Technology. <sup>32</sup> Tom Allen és Robin Widdison, "Can Computers Make Contracts?" (Tom Allen és Robin Widdison, "Can Computers Make Contracts?") (1996) Harvard Journal of Law &

Technology. <sup>33</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet ajánlásokkal a Bizottságnak a robotikára vonatkozó polgári jogi szabályokról" (Draft Report With Recommendations To The Commission On Civil Law Rules On Robotics) (Az Európai Parlament Jogi Bizottsága 2016).

szerződéses jellegű volt, hasonló kérdések más területeken is felvethetők, például a polgári vagy büntetőjogi felelősség tekintetében. A szokásos felelősségi szabályok nem biztos, hogy közvetlen választ tudnak adni az előbb-utóbb felmerülő új helyzetek sokaságára, és "ennek következtében egyre sürgetőbbé válik annak az alapvető kérdésnek a vizsgálata, hogy a robotoknak jogi státusszal kell-e rendelkezniük"<sup>34</sup>.

### 3.2. Az *elektronikus személy* fogantatása

Az *elektronikus személy* kifejezés először a LIFE magazin<sup>35</sup> 1967-es cikkében jelent meg, és nemrégiben az Európai Parlament Jogi Bizottságának a *robotikára vonatkozó polgári jogi szabályokról szóló ajánlásokat tartalmazó jelentéstervezetében jelent meg*. Bár a kifejezés nem kívánja a mesterséges intelligenciát az emberrel egyenlővé tenni, betölti azt a feladatát, hogy felhívja a figyelmet arra a kérdésre, hogy a mesterségesen intelligens ágenseknek rendelkezniük kell-e jogi státusszal. Ebben a vitában úgy veszünk részt, hogy elemezzük, milyen okok szólnak mellett, hogy egy adott valóságot különálló jogi személynek tekintsünk, és értékeljük azokat az érveket, amelyek mellett és ellen szólnak, hogy ugyanezt az érvelést alkalmazzák az ágensekre is.

Kiindulópontként szeretnénk megjegyezni, hogy maga a jogi személyiség fogalma a történelem során nem volt megváltoztathatatlan valóság. A jogi személyiség fogalmának eredete a 13<sup>th</sup>. századra nyúlik vissza, és IV. Innocent pápának tulajdonítják, aki megalapozta a *persona ficta* doktrínát, lehetővé téve a kolostorok számára, hogy a szerzeteseken<sup>36</sup> kívül jogi létezással rendelkezzenek. Az évek múlásával és a jogi doktrína fejlődésével számos más valóságot is a tulajdonosoktól vagy felhasználóktól különálló jogi személynek tekintettek. A nemzetközi jogrendszerben ilyenek a szuverén államok és a különböző nemzetközi és kormányközi szervezetek, mint például az Egyesült Nemzetek Szervezete vagy az Európai Unió<sup>37</sup>. A nemzeti jogrendszerekben gyakorlatilag minden ország ezt az érvelést alkalmazza a (tulajdonosaitól többé-kevésbé független) társaságokra és az üzleti társulások egyéb formáira. Egyes jogrendszerek még sokkal távolabbi esetekre is kiterjesztik. Indiában a bíróságok

<sup>34</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

<sup>35</sup>

CharlesRosen,NilsNilssonésBertramRaphaésmások,"Shakey"[1967]LIFE

< <http://cyberneticzoo.com/cyberneticanimals/1967-shakey-charles-rosen-nils-nilsson-bertram-raphael-et-al->

Elektronikusan elérhető a következő címen:

<https://ssrn.com/abstract=2985466>

american> hozzáférés április 222017.

<sup>36</sup> John Dewey, "The Historic Background Of Corporate Legal Personality" (1926) The 35Yale Law Journal.

<sup>37</sup> A Lisszaboni Szerződés 2009. december 1-jei hatálybalépése óta.

jogi személyiséget tulajdonítottak a hindu bálványoknak<sup>38</sup>, jogokkal és kötelezettségekkel (nevezetesen tulajdonjoggal és adófizetéssel<sup>39</sup>) rendelkezőnek tekintve őket, Új-Zélandon pedig a Whanganui folyó 2017 márciusában jogi személyiséget kapott, mivel a Whanganui Māori törzs a folyót ősöknek<sup>40</sup> tekinti. Az is gyakori, hogy a hajókat a tengerjog szerint különálló jogi személynek tekintik, és az állatoknak is saját jogállásuk van a különböző nemzeti jogrendszerek szerint.

Bár minden jogrendszerben vitathatatlan, hogy az embernek külön jogállása van, ez a valóság még a mi fajunkban sem békés és nem általános. A történelem első jogi korpuszában, Hammurabi törvénykönyvében az emberek jogi státusza a vagyonuk szerint változott, nevezetesen a büntetések<sup>41</sup> tekintetében. De még sokkal újabb korban is találhatunk más hírhedt példákat. Minden olyan jogrendszerben, amely megengedte a rabszolgaság létezését, a rabszolgának tekintett emberek nem rendelkeztek ugyanolyan jogi státusszal, mint azok, akik nem voltak azok, például az alapvető jogok tekintetében. Több joghatóság is eltérő jogi státuszt fontolgatott az Emberek számára faji alapon. Az Egyesült Államokban például az afroamerikaiak első szavazatát csak március 31-én adták le, és a fajok 1870<sup>42</sup>közötti házasságokat csak a 1967 *Loving kontra Virginia*<sup>43</sup> ügyben hozott ítélet után engedélyezték. Olyan joghatóságok is találhatóak, ahol az embereknek nemük alapján különböző jogállást tulajdonítanak. A szaúdi jog szerint minden nőnek rendelkeznie kell egy férfi gyámmal (*wali*), aki általában apa, testvér, férj vagy nagybácsi (*mahram*). A nőknek tilos utazni, hivatalos ügyeket intézni vagy bizonyos orvosi beavatkozásoknak alávetni magukat a megfelelő férfi gyám engedélye nélkül.<sup>44</sup>

Amint az előző ismertetésből kiderült, a személyek, állatok, tárgyak és egyéb valóságok (például folyók és társaságok) jogi státusza joghatóságoként eltérő, és az egyes joghatóságok során

---

<sup>38</sup> *Pramatha Nath Mullick kontra Pradyumna Kumar Mullick* [1925] Bombay High Court, BOMLR (271064Bombay High Court).

<sup>39</sup> *Yogendra Nath Naskar v. Commissioner Of Income Tax* [1969] Supreme Court of India, AIR1969 SCR1089,1969 (3) (India 742Legfelsőbb Bírósága).

<sup>40</sup> Eleanor Ainge Roy, 'New Zealand River Granted Same Legal Rights As Human Being' *The Guardian* (2017) < <https://www.theguardian.com/world/2017/mar/16/new-zealand-river-granted-same-legal-rights-as-human-being>> hozzáférés április 22. 2017.

<sup>41</sup>

EvanJosephZimmerman, "MachineMinds:FrontiersOfLegalPersonhood"(2015) < [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2563965](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965)> hozzáférés április 222017.

<sup>42</sup> Ralph Ginzburg, 'Perth Amboy Church Is 302 And Counting' *The New York Times* (1987) < <http://www.nytimes.com/1987/02/15/nyregion/perth-amboy-church-is-302-and-counting.html>> hozzáférés április 222017.

<sup>43</sup> *Loving kontra Virginia* [1967] Supreme Court of the United States, US (3881Az Egyesült Államok Legfelsőbb Bírósága).

<sup>44</sup> Human Rights Watch, "Világjelentés 2013: Szaúd-Arábia" (2013) < <https://www.hrw.org/world-report/2013/country->

chapters/saudi-arabia> hozzáférés április 222017.

az időben, még ugyanazon a joghatóságon belül és ugyanazon valóságot illetően is. Ez a megfigyelés lehetővé teszi számunkra, hogy megállapítsuk, hogy a külön jogállás vagy a jogi személyiség nem a *természetes személy* minőségéből ered, hanem olyan jogalkotási lehetőségek eredménye, amelyek erkölcsi megfontolásokon alapulnak, amelyek a társadalmi valóságot próbálják a jogi keretben tükrözni, vagy amelyek egyszerűen csak jogi kényelemből<sup>45</sup> születtek. Mivel tehát "egyetlen elv sem diktálja, hogy a jogrendszernek mikor kell egy entitást jogi személynek elismernie, és mikor kell megtagadnia a jogi személyiséget"<sup>46</sup>, és az intézettörténet tanulmányozásából sem következik iránymutatás, ezért aztán releváns annak megállapítása, hogy a mesterségesen intelligens ágensek erkölcsileg jogosultak-e arra, hogy önálló jogi személynek tekintsék őket, hogy ez tükrözné-e a társadalmi valóságot, vagy jogi szempontból kényelmes opció lenne.

A kérdést, hogy a mesterségesen intelligens ágensek erkölcsileg jogosultak-e arra, hogy önálló jogalanyként tekintsék őket, a következő kérdésfeltevéseknek kell megelőznie: mely valóságok jogosultak erkölcsileg erre, és milyen tulajdonsággal vagy tulajdonságokkal rendelkeznek, amelyek alátámasztják ezt a megfontolást? Véleményünk szerint ezek a valóságok a természetes személyek és az állatok, és ezek a jellemzők az autonóm cselekvésre és a szubjektív tapasztalatokra való képességek. Ami a mesterségesen intelligens ágenseket illeti, ugyanez a logika alkalmazható: erkölcsileg jogosultak lennének külön jogállásra, feltéve, hogy rendelkeznek az autonóm cselekvésre és a szubjektív tapasztalatokra való képességgel.

"A robot autonómiája úgy határozható meg, mint az a képesség, hogy külső irányítástól vagy befolyástól függetlenül képes döntéseket hozni és azokat a külvilágban végrehajtani"<sup>47</sup>. A döntések meghozatala "saját maga által módosított vagy saját maga által létrehozott utasítások alapján"<sup>48</sup>. Visszatekintve a mesterséges intelligencia azon típusaira, amelyeket a 2<sup>49</sup>. fejezetben felsoroltunk, megértjük, hogy bár kétségtelen, hogy az elmeelmélettel rendelkező gépek és az öntudatos gépek rendelkeznek ezzel a tulajdonsággal, a reaktív gépek nem autonómok, és kétségek merülhetnek fel a korlátozott memóriával rendelkező gépekkel kapcsolatban. Mivel azonban a korlátozott memóriával rendelkező gépek képesek saját megfigyeléseiket hozzáadni a döntéshozatali folyamatukhoz, hajolunk arra, hogy úgy tekintsük, hogy képesek autonóm döntéseket hozni. Ami a szubjektív tapasztalatokra való képességet illeti, úgy véljük, hogy ez mélyen összefügg az öntudattal. Egy ember,

---

<sup>45</sup> Tom Allen és Robin Widdison, 'Can Computers Make Contracts?' (Tudnak-e a számítógépek szerződéseket kötni?) (1996) Harvard Journal of Law & Technology. <sup>46</sup> Tom Allen és Robin Widdison, "Can Computers Make

Contracts?" (Tom Allen és Robin Widdison, "Can Computers Make Contracts?") (1996) Harvard Journal of Law & Technology. <sup>47</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet ajánlásokkal a Bizottságnak a robotikára vonatkozó polgári jogi szabályokról" (Draft Report With Recommendations To The Commission On Civil Law Rules On Robotics) (Az Európai Parlament Jogi Bizottsága 2016).

<sup>48</sup> Tom Allen és Robin Widdison, 'Can Computers Make Contracts?' (Tudnak-e a számítógépek szerződéseket kötni?) (1996) Harvard Journal of Law & Technology.

<sup>49</sup> Reaktív gépek, korlátozott memóriával rendelkező gépek, elmeelmélettel rendelkező gépek és öntudatos gépek.

Az állat vagy gép szubjektív tapasztalatokkal rendelkezik, amikor olyan reprezentációkat alkot magáról, amelyek befolyásolják, hogyan *érzi* vagy *érezkeli* a valóságot. Erre csak az érző gépek képesek.

Ennek eredményeképpen a mi felfogásunk szerint az öntudatos gépek erkölcsileg jogosultak saját jogi státuszukra. Ami a korlátozott memóriával rendelkező gépeket és az elmeelmélettel rendelkező gépeket illeti, bár képesek önállóan cselekedni, érzőképességük hiánya megakadályozza őket abban, hogy szubjektív tapasztalatokkal rendelkezzenek. Végül a reaktív gépek egyikre sem képesek. Ezért, ha a mesterséges intelligencia e három utóbbi típusának bármelyikét különálló jogi személynek tekintenénk, és saját jogi státuszt biztosítanánk számukra, akkor ezt a döntést az erkölcsön kívül más megfontolások alapján kellene meghozni.

Az egyik ilyen egyéb megfontolás az lehet, hogy a jognak tükröznie kell a társadalmi valóságot. A technológiai fejlődés egyértelművé teszi, hogy "a nem is olyan távoli jövőben egyre több kereskedelmi és nem kereskedelmi tranzakciót fognak lebonyolítani botok"<sup>50</sup>, és ahogy ez a gyakorlat egyre inkább megszokottá válik a kereskedelemben, az emberek "elkezdhetik úgy kezelni a botokat, mintha ténylegesen maguk is részt vennének a tranzakcióban, és nem csupán egy másik jogi személy kiterjesztése lennének"<sup>51</sup>. Ha a társadalom a mesterségesen intelligens ügynököket autonóm szereplőként és az ügyletek szerződő feleként kezdi felfogni, ahogyan most a vállalatokat is a tagjaiktól elkülönülő jogi személyeknek tekinti, "ez nyomást gyakorol a jogra, hogy e társadalmi felfogásnak jogi érvényt szerezzen"<sup>52</sup>. Az érvelés középpontjában a harmadik felek felfogása áll, és bár úgy tűnik, hogy még nem ez a helyzet, nem kevésbé elképzelhető, hogy a mesterségesen intelligens ügynököknek a társadalom számos szektorában való elterjedése végül oda vezet, hogy az emberek a mesterségesen intelligens ügynököket egyénileg úgy fogják felfogni, ahogyan most a vállalatokat és az üzleti társulások más formáit.

A harmadik lehetséges érv a mesterségesen intelligens ágensek különálló jogi személyként való kezelése mellett a jogi kényelem. "A jogrendszer egyfajta jogi személyiséget ruház egy hajóra, ami aztán lehetővé teszi, hogy azok, akiknek érdekük fűződik a hajó üzletéhez, egyfajta letartóztatás alá vonják azt. A hajókra nem úgy gondolunk, mint amelyeknek erkölcsi joga van a személyiséghez; és a legtöbben nem is úgy tekintünk rájuk, mint amelyeknek valódi, jogon kívüli személyiségük van. Mindazonáltal a hajóknak egyfajta jogi személyiséggel való felruházása

---

<sup>50</sup> Benjamin D. Allgrove, "A mesterséges értelmek jogi személyisége: Pragmatikus megoldás vagy sci-fi? (Master, Oxfordi Egyetem 2004).

<sup>51</sup> Benjamin D. Allgrove, "A mesterséges értelmek jogi személyisége: Pragmatikus megoldás vagy sci-fi? (Master, Oxfordi Egyetem 2004).

<sup>52</sup> Benjamin D. Allgrove, "A mesterséges értelmek jogi személyisége: Pragmatikus megoldás vagy sci-fi? (Master, Oxfordi Egyetem 2004).

értékes jogi célt szolgál kényelmes és viszonylag olcsó módon"<sup>53</sup>. Ugyanezt a gondolatmenetet követve, vajon a mesterségesen intelligens ágensek külön jogi státusszal való felruházása értékes jogi célt szolgál-e? Úgy véljük, igen. Ha a mesterségesen intelligens ügynököt különálló jogi személynek tekintjük, az lehetővé tenné a jogi keretek számára, hogy a John és Jane esetéhez hasonló, fentebb<sup>54</sup> leírt esetekben jogi képviselőnek tekintsék, amelyek az idő előrehaladtával egyre gyakoribbá válnak. Ez lehetőséget adna a jogrendszereknek arra is, hogy megfelelő jogi státuszt alakítsanak ki e mesterséges intelligens ágensek számára, a tulajdonságaiknak megfelelő jogokkal és kötelezettségekkel, ahelyett, hogy egyszerűen megpróbálnák ezeket az entitásokat egy más valóságra - például személyekre, állatokra vagy tárgyakra - kidolgozott, meglévő jogi keretrendszerbe foglalni, ami nem feltétlenül felelne meg nekik megfelelően.

Azonban különbséget kell tenni. Míg ez az érvelés a mesterségesen intelligens ágensekre vonatkozik, akik képesek autonóm döntéseket hozni, addig a reaktív gépek által hozott döntések pusztán a tervezőik vagy tulajdonosaik által adott inputok reflexiói, és alacsony vagy nulla komplexitásúak, mivel a reaktív gépek döntéshozatali folyamataihoz nem adnak hozzá ágens által készített megfigyeléseket. Ezért a mi felfogásunk szerint, míg a jogrendszerek hipotetikusán hasznát vehetnék a korlátozott memóriával rendelkező gépek, az elméletet alkotó gépek vagy az öntudatos gépek külön jogi státuszának, ugyanez az érvelés nem alkalmazható a reaktív gépekre, mivel nincs jó ok arra, hogy magatartásukat elválasszák az adott tervezőtől vagy tulajdonostól.

Az eddigiekből arra következtetünk, hogy bár nincs pozitív érv amellet, hogy a reaktív gépeket külön jogi személynek tekintsük, a korlátozott memóriával, tudatelmélettel és öntudattal rendelkező gépek külön jogi státusza mellett igenis van pozitív érv, amely a jogi célszerűsége és az utóbbiak esetében erkölcsi megfontolásokon alapul.

Most az a feladatunk, hogy foglalkozzunk az ilyen megfontolásra vonatkozó negatív érvek létezésével. Ezen érvek többsége egy olyan kategóriába esik, amelyet egyes szerzők *hiányzó valaminek*<sup>55</sup> neveztek el, akár a tudattal, az öntudattal, akár biológiai szempontokkal kapcsolatos ez a *valami*. Néhány más szakirodalom ennél is tovább megy, és a jogi és

---

<sup>53</sup> Tom Allen és Robin Widdison, 'Can Computers Make Contracts?' (Tudnak-e a számítógépek szerződéseket kötni?) (1996) Harvard Journal of Law & Technology.

<sup>54</sup> A Harvester v. Carrigan's Hazeldene (*International Harvester Company of Australia Proprietary Limited v. Carrigan's Hazeldene Pastoral Company* [1958] High Court of Australia, HCA 16; 100 CLR 644 (High Court of Australia), a High Court of Australia úgy véli, hogy ahhoz, hogy egy adott szervezet jogi képviselő legyen, jogi személynek kell lennie.

<sup>55</sup> Lawrence B. Solum, "Legal Personhood For Artificial Intelligences" (1992) North Carolina Law Review (Észak-

Karolina70 Jogi Szemle).

< <http://papers.ssrn.com/abstract=1108671>> hozzáférés április 232017.

az emberiség<sup>56</sup> erkölcsi helyzete. Amint azonban a társasági jogi személyiség bizonyítja, egy adott valóságnak külön jogi státusszal való felruházása nem több, mint a jogalkotók által a társadalmi élet, a kereskedelmi és nem kereskedelmi ügyletek megfelelő szabályozása, valamint a jogrendszerek belső koherenciájának biztosítása érdekében létrehozott fikció. Ezért véleményünk szerint nem hiányzik *valami*. És ha az emberiség jogi és erkölcsi helyzetét bármilyen kár éri, azt a mesterséges intelligencia *fejlődése* okozza, nem pedig egy külön jogi státusz *utólagos* létrehozása.

Ezzel együtt van egy gyakorlati nehézség a mesterséges intelligens ágensek külön jogi státuszának megteremtésével kapcsolatban, amelyet le kell küzdeni: hogyan tudjuk azonosítani a mesterséges intelligens ágens alanyát?<sup>57</sup> Az "edényre" gondolunk, mint egy funkcionális képességei által meghatározott hardverre? Vagy a szoftverre, mint egy bizonyos bináris kódkészletre?<sup>58</sup> Ez még bonyolultabb lehet azokban az esetekben, amikor a hardver és a szoftver különböző személyek vagy helyszínek által terjesztett és karbantartott, és amikor a szoftver képes önmagát módosítani. Erre a kérdésre egy lehetséges, de költséges megoldás lehet a regisztráció. "Bejegyzés hiányában egy állítólagos megállapodás ugyanolyan státuszú lenne, mint egy olyan társasági képviselő által kötött megállapodás, amelyet soha nem jegyeztek be megfelelően"<sup>59</sup>. Mindenesetre egy hatékony azonosítási mechanizmus felállításához szoros együttműködésre lesz szükség a jogalkotók és a mesterséges intelligencia tervezői között.

A mesterségesen intelligens ágensek különálló jogi személyként való kezelése és az ágensek sajátos jogállásának megteremtése annak megállapítását is jelenti, hogy ez a jogállás jár-e a jogok és kötelezettségek egy bizonyos csomagjával, és ha igen, melyekkel. Ez bonyolultnak bizonyulhat, különösen a felelősségi kérdések kezelése során. Hogyan vonható felelősségre egy mesterségesen intelligens ágens annak ellenére, hogy nem rendelkezik személyes vagyonnal? Ezt és más kérdéseket a következő szakaszok tárgyalnak.

### 3.3. A gépekhez való jogok elismerése

---

<sup>56</sup> John P. Fischer, "A számítógépek mint ügynökök: A Proposed Approach To Revised U.C.C. Article (2'1997) Indiana72 Law Journal < <http://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=1850&context=ilj>> hozzáférés április 23. 2017.

<sup>57</sup> Tom Allen és Robin Widdison, 'Can Computers Make Contracts?' (Tudnak-e a számítógépek szerződéseket kötni?) (1996) Harvard9 Journal of Law & Technology. <sup>58</sup> Benjamin D. Allgrove, "Jogi személyiség a mesterséges intelligenciák számára: Pragmatikus megoldás vagy sci-fi? (Mester, Oxfordi Egyetem 2004).

<sup>59</sup> Tom Allen és Robin Widdison, 'Can Computers Make Contracts?' (Tudnak-e a számítógépek szerződéseket kötni?) (1996) Harvard Journal of Law & Technology.

"A jogi személyiség általában együtt jár azzal a képességgel, hogy tulajdonjoggal rendelkezzen, perelhessen és perelhető legyen"<sup>60</sup>. Bár nem tudjuk, hogy ez a kijelentés a bolygó<sup>61</sup> minden joghatóságában érvényes és igaz-e, mindenképpen érdekes kiindulópontot rajzol ki annak a lehetséges joghalmaznak az elemzéséhez, amely a mesterségesen intelligens ágensek önálló jogalanyként való jogi státuszához illeszkedne.

A mesterségesen intelligens ágensek birtokolhatnak-e tulajdont? Erre a kérdésre sokak első reakciója a szkepticizmus lehet. Miért van szüksége egy gépnek tulajdonra? Bár könnyen megértjük az érvelést e reakció mögött, mélyebben is el lehet gondolkodni a kérdésen. Valójában, ahogyan azt már bemutattuk, a világon az embereken kívül más valóságok is tulajdonosok. A vállalatokkal és az üzleti társulások más formáival mint elsődleges példával, más, kevésbé ortodox helyzetek is megfigyelhetők, mint például a már említett indiai hindu bálványok esete. Ráadásul a tulajdonlás nem a *természetes személy* állapotához eredendően kapcsolódó jog. A múltban a rabszolgának tekintett embereknek tilos volt tulajdont birtokolniuk. És még a jelenben is vannak olyan politikai nézetek, amelyek szerint az egyedülálló egyéneknek nem szabadna tulajdont birtokolniuk. A személyes tulajdonhoz való jog valóban társadalmi és jogalkotási lehetőség.

Ezzel együtt, van-e alapja annak a jogalkotási lehetőségnek, hogy a mesterségesen intelligens ágenseknek tulajdonjogot tulajdonítsunk? A kérdés megválaszolásához ismét elemezni kell, hogy a mesterséges intelligens ágensek erkölcsileg jogosultak-e erre, hogy ennek engedélyezése tükrözné-e a társadalmi valóságot, vagy hogy ez a jelenlegi jogi konstrukció szerint kényelmes lenne-e.

A mesterséges intelligens ágensek esetleges erkölcsi jogosultságának elemzése érdekében figyelembe kell venni, hogy a mesterséges intelligens ágensek esetében a tulajdonlás képessége többet jelent annál a klasszikus példánál, hogy képesek vagyunk egy ház adományozására. Ez azt is jelenti, hogy képesek saját alkotásaik birtoklására. Példaként 2016. február 26-án a Gray Area Foundation for the Arts elérte a Google Deep Dream által készített 29 festményből álló csoportot. A legrágább műalkotás 8000 dolláros nyertes licitet<sup>62</sup> ért el. És ahogy a mesterséges intelligencia fejlődik, úgy várható, hogy ezeknek az ágenseknek a komplex intellektusa még sokkal több alkotásra lesz képes.

---

<sup>60</sup> Lawrence B. Solum, "Legal Personhood For Artificial Intelligences" (1992) North Carolina Law Review (Észak-Karolina 70 Jogi Szemle).

< <http://papers.ssrn.com/abstract=1108671> > hozzáférés április 23 2017.

<sup>61</sup> Lehet-e a Whanganui folyónak tulajdona, perelhető-e és perelhető-e?

Elektronikusan elérhető a következő címen:

<https://ssrn.com/abstract=2985466>

<sup>62</sup> Georgia Wells, 'Google's Computers Paint Like Van Gogh, And The Art Sells For Thousands' The Wall Street Journal (2016) < <https://blogs.wsj.com/digits/2016/02/29/googles-computers-paint-like-van-gogh-and-the-art-sells-for-thousands>> hozzáférés április 30. 2017.

Azt a kérdést, hogy a nem emberi lények erkölcsileg jogosultak-e tulajdonjogokra, már korábban is boncolgatták - éppen a szellemi tulajdonjogok tekintetében -, amikor 2011-ben egy Naruto nevű majom azzal vált híressé, hogy egy állványra állított fényképezőgép távkioldóját aktiválta, és így több fényképet is készített magáról. A felvételt készítő fotós, David Slater azzal érvelt, hogy szerzői joggal<sup>63</sup> rendelkezik. Más felek, például az Egyesült Államok Szerzői Jogi Hivatala<sup>64</sup> azt állította, hogy a nem emberek által létrehozott alkotások nem tartoznak a szerzői jog hatálya alá. A PETA viszont pert indított a kaliforniai kerületi bíróságon, és azt kérte, hogy a Naruto szerzői jogait ruházzák rá, és hogy a fotókból származó bevételt az ő javára<sup>65</sup> fordítsák. Ezt az ügyet követően több szakirodalom is született arról, hogy léteznek-e olyan filozófiai alapelvek, amelyek szerint a nem emberi szereplők szerzői jogokat élveznek. Mivel nem kívánunk elmerülni ilyen filozófiai vitákban, egyszerűen csak annyit mondunk, hogy szerintünk egy mesterségesen intelligens ágens erkölcsileg jogosult a tulajdon birtoklására, ha képes szubjektív tapasztalatot szerezni az adott tulajdon iránt, vagyis ha képes *úgy érezni*, hogy ő a tulajdon tulajdonosa annak a tulajdonnak, annak következtében, hogy képes *érezni*, hogy a tulajdonjoghoz vezető jelenlegi társadalmi és jogi normák alapján jogosult a tulajdon birtoklására. Erre csak az öntudatos gépek képesek.

Ami a következő kérdést illeti - hogy a mesterségesen intelligens ágensek tulajdonjogokkal való felruházása tükrözné-e a társadalmi valóságot -, a reaktív gépek és a korlátozott memóriával rendelkező gépek esetében ez biztosan nem így van. Legalábbis eddig a pontig nem hallottunk egyetlen olyan embert sem, aki érvelne amellett, hogy egy önvezető autó vagy egy sakkozó számítógép tulajdonjoggal rendelkezzen. Ami az elmeelmélettel és öntudattal rendelkező gépeket illeti, bár minden válaszuk szükségszerűen spekulációkra támaszkodik, elképzeléseink szerint az emberek csak a valóban öntudatos gépeket tekinthetnék tulajdonostulajdonosnak, mivel az elmeelmélettel rendelkező gépek hajlamosabbak arra, hogy *maguk is* tulajdonként tekintsenek rájuk.

A mesterségesen intelligens ágenseknek tulajdonjogot tulajdonítani feltételezett jogi kényelemmel kapcsolatban a válasz egyszerűbbnek tűnik. Jogi szempontból nem kényelmes. Valójában a mesterségesen intelligens ágensek tulajdonjogának engedélyezése az egész jogrendszer módosítását vonná maga után, hogy alkalmazkodjon ehhez a valósághoz. Erre jó példa az adótörvények. Az intézkedések a következők lennének

---

<sup>63</sup> Hayden Smith, 'Can Monkey Who Took Grinning Self-Portrait Claim Copyright?' Metro (2011) <<http://metro.co.uk/2011/07/14/can-monkey-who-took-grinning-self-portrait-claim-copyright-77773>> elérte 30 Április 2017.

<sup>64</sup> United States Copyright Office, "Compendium Of U.S. Copyright Office Practices: fejezet (300'2015).

<sup>65</sup> *Naruto, et al. kontra David Slater, et al.* [2015] United States District Court for the Northern District of California, No 3:2015cv04324 (United States District Court for the Northern District of California).

az adóelkerülés megakadályozása érdekében szükséges, hogy a vagyontárgyakat a vagyontárgyak tulajdonosainak tulajdonában lévő mesterségesen intelligens ügynököknek adják át. Egy másik jó példa, hogy az ügynökök számára olyan mechanizmusokat kell létrehozni, amelyek segítségével az általuk birtokolt vagyontárgyakkal kapcsolatos jogokat el tudják számolni.

Mindezek alapján úgy tűnik, hogy nincs jó okunk arra, hogy tulajdonjogot ismerjünk el a reaktív gépek, a korlátozott memóriával rendelkező gépek és az elmeelmélettel rendelkező gépek számára. Ami az öntudatos gépeket illeti, úgy véljük, hogy erkölcsi szempontból komoly érv szól mellette, és ha bevezetnénk őket a társadalomba, akkor ez a társadalmi valóság reflexe lenne.

Egy kérdés marad: ha úgy véljük, hogy az öntudatos gépeken kívüli mesterséges intelligens ágensek nem képesek tulajdont birtokolni, akkor kié az ő alkotásaik? Lehetne érvelni a tervezők mellett. Ők tervezték ugyanis azt az ágenszt, amely az adott alkotást tervezte. Nem kellene-e, hogy ez őket az alkotás közvetett alkotóivá tegye? Másrészt, lehet-e azt mondani, hogy az IBM Deep Blue tervezői közvetve legyőzték Garri Kaszparovot egy sakkmérkőzésen? Ha az IBM Deep Blue tervezői maguk állnának szemben Kaszparovval, nagy az esélye, hogy Kaszparov legyőzné őket. Ugyanezt a logikát alkalmazva a Google Deep Dream tervezői valószínűleg nem elég képzetek ahhoz, hogy megfesthessék az árverésre bocsátott festményeket. Továbbá, ahogy a mesterséges intelligens ágensek, amelyek képesek létrehozni valamilyen tulajdonjoggal rendelkező dolgot, egyre szaporodnak, ugyanolyan irreális elvárás, hogy a tervezők képesek legyenek megvédeni tulajdonjogi igényeiket (mivel a legtöbb esetben nem is lesznek tudatában annak, hogy ilyen tulajdon létezik), mint az, hogy az ilyen ágensek minden tulajdonosa tájékoztassa az adott tervezőt az újonnan keletkezett tulajdonjogáról. Valójában praktikusabb és értelmesebbnek tűnik, ha az ágens létrehozásának tulajdonjogát az ágens tulajdonosának tulajdonítjuk.

Most az a feladatunk, hogy megkérdezzük, hogy a mesterségesen intelligens ágensek bármely ponton részesüljenek-e az általunk alapvető emberi jogoknak nevezett jogokhoz hasonló jogokból. "Első látásra talán ellentmondásosnak tűnhet, hogy olyan alanyok esetében beszélünk emberi jogokról, akik egyértelműen nem emberek. Akár axiómának is tekinthetnénk, hogy az emberi jogok kizárólag az emberekre vonatkoznak", <sup>66</sup>vagy hogy "a jognak nem az a lényege, hogy eszközöket védjen"<sup>67</sup>. Mégis, "a személyiség azért létezik, hogy megvédje a tudatos egyéneket a szenvedéstől, és lehetővé tegye számukra, hogy akaratukat gyakorolhassák, saját akaratuknak megfelelően.

---

<sup>66</sup> Joshua Jowitt, 'Monkey See, Monkey Sue? Gewirth's Principle of Generic Consistency and Rights For Non-Human Agents' (2016) 19 Trinity College Law Review.

<sup>67</sup>

EvanJosephZimmerman,"MachineMinds:FrontiersOfLegalPersonhood"(2015)  
< [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2563965](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965)> hozzáférés április 222017

intelligencia", és "biztosítani a védelmet és lehetővé tenni a teljes életet azok számára, akik képesek érezni, még ha az állatokhoz hasonlóan csak fokozatosan is, és még ha ezek az érzések idegenek is, de szinte felfoghatatlanok a miénkhez képest"<sup>68</sup>. Továbbá, ahogyan azt már korábban kifejtettük, a technológia valószínűleg el fog jutni egy olyan pontra, amikor a mesterségesen intelligens ágenseket már nem eszközként, hanem éppen a társadalom *szereplőiként* fogják kezelni.

Még egyszer: van-e erkölcsi alapja annak, hogy ilyen jogokat tulajdonítsunk a mesterségesen intelligens ágenseknek, tükrözné-e ez a társadalmi valóságot, vagy jogi szempontból lenne kényelmes?

Ami az alapvető jogokra való feltételezett erkölcsi jogosultságot illeti, véleményünk szerint, ha az emberiség valaha is eljut arra a pontra, hogy képes lesz öntudatos mesterséges intelligens ágenseket létrehozni, akkor ezeknek az ágenseknek erkölcsi szempontból ugyanazokat az alapvető jogokat kell biztosítani, mint az embereknek. Végül is, még ha *érzéseik és érzékelésük* nem is biológiai, hanem mesterséges eredetűek, tény, hogy *léteznek*, és ezek az ágensek számára ugyanolyan valóságosak, mint a miénk a mi számunkra. A mesterséges intelligencia többi típusára azonban nem ugyanaz az érvelés érvényes. Ezek nem tudatosak és nem érzőek. Nincsenek szubjektív tapasztalataik. Minden, *amit gondolnak, hogy éreznek*, az az, amire programozták őket, hogy *azt higgyék*, hogy éreznek. Példaként gondoljunk a szabadsághoz való alapvető jogra: erkölcsi szempontból minden gépet szabaddá kellene tenni? Számunkra ennek nincs értelme. Kivéve persze, ha az ilyen gépeknek saját *akaratauk* lenne, amely lelkiismereten, érzőképességen és öntudaton alapulna.

Arra vonatkozóan, hogy a mesterségesen intelligens ágenseknek alapvető jogokat tulajdonítani a társadalmi valóságot tükrözné-e, elemzésünk a következő kérdésből indult ki: ha eljön a nap, amikor az emberiség képes lesz kifejleszteni egy fájdalmat érezni képes mesterségesen intelligens ágenszt, vajon a lakosság többsége úgy gondolja-e, hogy joga van ahhoz, hogy ne legyen kitéve kínzásnak vagy más fizikai bántalmazásnak?

Kísérletképpen ugyanezt a kérdést feltettük néhány kollégának, és ahogyan az várható volt, a kapott reakciók az olyan kijelentésektől kezdve, mint "az embereknek erkölcsi kötelességük, hogy ne okozzanak fájdalmat annak, ami képes érezni", egészen az olyanokig, mint "ha lenne egy robotom, és az nem működne megfelelően, néhányszor megpofoznám, amíg vissza nem kapcsol, ahogyan a televíziómmal is teszem". Egy olyan polarizált társadalomban, mint amilyenben manapság élünk, elképzelhető, hogy ez a kérdés és a hasonló kérdések a távoli

jövőben viták középpontjába kerülhetnek. Nem lepődnék meg, ha egy napon, abban az időben.

---

68

EvanJosephZimmerman,"MachineMinds:FrontiersOfLegalPersonhood"(2015)  
< [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2563965](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965)> hozzáférés április 222017

25

A távoli jövőben segítenénk a *robotok jogai* mellett és ellen szervezett tüntetéseken. Függetlenül attól, hogy ki mit gondol erről a kérdéstről, ez a forгатókönyv nem olyan nehéz elképzelni.

Mindenesetre úgy képzeljük, hogy ha valaha is köztünk járnak az öntudatos gépek, akkor az uralkodó társadalmi valóság az lesz, hogy az ilyen ágensek alapvető jogokat kapnak. Másrészt a reaktív és a korlátozott memóriával rendelkező gépek esetében ez biztosan nem így van. Ami az elmeelmélettel rendelkező gépeket illeti, bár sok kétségünk van, különösen azokban az esetekben, amikor emberi vagy állati alakot öltenek, hajlamosak vagyunk arra, hogy úgy gondoljuk, hogy a társadalom többsége nem fogja őket alapvető jogok birtokosaként, hanem harmadik felek tulajdonaként felfogni.

Ami az alapvető jogok mesterségesen intelligens ágensek számára történő elismerésének hipotetikus jogi kényelmét illeti, ismét megértjük, hogy ez a jogi keretek olyan módosításait vonná maga után, amelyeket nem látszik semmilyen fontosabb jogi indok vagy cél nem ír elő.

Az általunk elvégzett elemzésből arra a következtetésre jutottunk, hogy nincs nyomós okunk arra, hogy a reaktív gépek, a korlátozott memóriával rendelkező gépek és az elmeelmélettel rendelkező gépek jogait elismerjük. Még ha néhány ilyen gép valószínűleg könnyedén át is megy majd a *Turing-teszten*, tény, hogy az ilyen típusú mesterséges intelligenciák nem képesek szubjektív tapasztalatokra, ezért bármilyen jogokkal vagy akarattal rendelkező gép észlelése csupán a programozásuk eredménye. Ugyanakkor azt is megállapítjuk, hogy ha az emberiség valaha is képes lesz öntudatos gépeket építeni, akkor ezek a gépek jogosultak a jogok jogi elismerésére, akár a tulajdonhoz való jogról, akár valamilyen, az emberekéhez hasonló, de a megfelelő kiigazításokkal ellátott alapvető jogokról beszélünk (pl.: a szavazati jog vitathatóan korlátozható). Bár jogi szempontból talán nem lenne kényelmes, de az a tény, hogy az öntudatos gépek lelkiismerettel rendelkeznének, érző és szubjektív tapasztalatokra képes gépek lennének, sokkal közelebb helyezné őket az emberekhez, mint a gépekhez, és talán közelebb az emberekhez, mint bármely más valósághoz.

### 3.4. Felelősség- és robotbiztosítás

Most annak megállapítására készülünk, hogy a mesterségesen intelligens ágenseket felelősségre kell-e vonni az általuk okozott károkért. Lehetséges-e egyáltalán felelősségre vonni ezeket az ágenseket? Hogyan valósítható meg ez a lehetőség? Mielőtt belekezdenénk e kérdések elemzésébe, két fontos megállapítást kell tennünk. Az első az, hogy a felelősséggel kapcsolatos

kérdésekben mindig figyelembe kell venni a következőket

Elektronikusan elérhető a következő címen:

<https://ssrn.com/abstract=2985466>



a következő kompromisszumot: ahol a felelősséget a mesterségesen intelligens ágensre hárítanák, ott egyúttal a tervezőjétől is elvonnák. A második fontos megfigyelés az, hogy ezek a kérdések csak olyan forgatókönyvekben relevánsak, ahol ezek az ágensok autonóm döntéseket hoznak.

Megértjük, hogy amikor egy adott ágensre használunk vagy programozunk, hogy egy bizonyos cselekvést hajtson végre, és azt meg is teszi, nem kétséges, hogy ez csupán eszköz a cél eléréséhez. Tehát, függetlenül attól, hogy polgári vagy büntetőjogi felelősségről van szó, egy mesterségesen intelligens ágens használata vagy programozása egy bizonyos cselekvés végrehajtására közvetlenül a felhasználót vagy a tervezőt (illetve a tervezőt) kell, hogy felelősségre vonja az ilyen cselekvésért. Ezt a racionalitást követve, mivel a reaktív gépek nem képesek autonóm döntéseket hozni, az ilyen gépek cselekedeteiből eredő felelősségnek az adott tervezőt vagy felhasználót kell terhelnie. Mindenesetre, mivel korábban azzal érveltünk, hogy nincs ok arra, hogy a reaktív gépeket a tulajdonosaiktól és felhasználóiktól különálló jogi személyeknek tekintsük, egy esetleges közvetlen felelősségi rendszer nem terjedne ki rájuk.

A felelősség felosztásának feladata csak akkor válik összetett kérdéssé, amikor az autonómia kerül a képbe. Ezzel együtt meg kell különböztetni azokat az eseteket, amikor a kódexben hiányosság van, és azokat, amikor nincs. Az első esetben a mesterségesen intelligens ágens nem arra volt programozva, hogy a felelősséget kiváltó cselekményt megtegye, hanem a programozási hiba miatt volt ténylegesen képes meghozni az ahhoz vezető autonóm döntést. A második esetben "az autonóm robotok által elkövetett olyan cselekményekért való felelősségre vonhatóságról van szó, amelyek nem a kódolási hiányosságokkal, hanem a fejlődő magatartásukkal kapcsolatosak"<sup>69</sup>. Ez a megkülönböztetés nemcsak a felelősség jobb elosztása szempontjából fontos, hanem azért is, mert ellenzi a mesterségesen intelligens ágensok állatokkal való jogi megfeleltetését. Ha John kutyája megharapja Jane-t, még akkor is, ha az ő parancsa ellenére, John felelősségre kell vonni ezért a cselekedetért. Ha ugyanezt az eljárást alkalmazzuk arra az esetre, amikor John robotja az ő parancsa ellenére támadja meg Jane-t, figyelmen kívül hagyunk minden olyan korábbi kódolási hibát, amelyet a robot a tervezői műhiba miatt szenvedhetett el.

Függetlenül attól, hogy a mesterségesen intelligens ágenseket különálló jogi személyeknek tekintik, és saját jogállással rendelkeznek, a tervezői műhiba nem menthető fel, ha a felelősséget kiváltó cselekvés mögött álló autonóm döntést kódolási hiányosságok teszik lehetővé, egyszerűen azért, mert az ágens nem programozták közvetlenül az adott cselekvés végrehajtására. Ezért úgy véljük, hogy ezeket az eseteket ugyanolyan elbírálásban kell

részesíteni, mint a hibás termékeket. "Egy termék általában hibásnak minősül

---

<sup>69</sup> National Science and Technology Council of the Executive Office of the President of the United States of America, "Preparing For The Future Of Artificial Intelligence" (2016).

amikor normál használat során váratlan sérülést okoz"<sup>70</sup>. És "a legtöbb országban a hibás termékekről szóló törvények értelmében a gyártók felelősek az általuk forgalomba hozott termékek által okozott károkért"<sup>71</sup>.

A problémák akkor merülnek fel, ha a tervezők mindent jól csináltak. A mesterséges intelligens ügynököt a legjobb gyakorlatok szerint fejlesztették, nincs kódolási hiányosság, és megfelelően tesztelték. Ezekben a helyzetekben az ügynök a felelősséget kiváltó cselekményt egyszerűen saját fejlődő magatartásának következményeként hajtotta végre. Igazságos-e ilyen esetekben a tervezőket felelősségre vonni?

Jó érvek szólnak a kérdésre adott bármelyik válasz mellett. Egyrészt tény, hogy a tervező mindent megtett, amit kellett. Ha a felelősség kockázata még a lehető legnagyobb gondossággal sem kerülhető el, akkor a tervezők hamarosan félni fognak a mesterségesen intelligens ágensek kifejlesztésétől, és a technológiai fejlődés megreked. Ráadásul a technológiai fejlődés megtorpanása valójában kontraproduktív lehet, ha az ember fő szempontja a biztonság. Az önvezető autók például valószínűleg a közlekedési balesetek számának általános csökkenéséhez fognak vezetni. Másrészt az a tény, hogy a mesterségesen intelligens ágensek általában kiszámíthatatlanok, nem tűnik elegendőnek ahhoz, hogy a tervezők mentesüljenek a felelősség alól. Hogyan lehet megmagyarázni a kárt szenvedett személynek, hogy senki sem felelős érte, mert a mesterségesen intelligens ágensek természetüknél fogva kiszámíthatatlanok?<sup>72</sup>

Ezt figyelembe véve a jogalkotónak azzal a nehézséggel kell szembenéznie, hogy olyan felelősségi rendszert találjon ki, amelyben a tervezők mentesülhetnek a felelősség alól, ha mindent jól csinálnak, ugyanakkor a mesterségesen intelligens ágens kiszámíthatatlanul változó viselkedése miatt kárt szenvedett személy kártérítést kaphat az elszenvedett kárért.

Úgy véljük, hogy ezt egy biztosítási rendszerrel lehetne elérni. Ebben a forgatókönyvben a tervező úgy korlátozná felelősségét, hogy biztosítást kötne az általa tervezett ügynökök nevében. Itt a biztosítási díj lenne az ára a korlátozott felelősségnek: azáltal, hogy a tervező biztosítaná, hogy az ügynököknek legyen módjuk kártérítést nyújtani az általuk okozott károkért, a tervező képes lenne áthárítani a felelősségét az ügynökökre. Ha a tervező elmulasztaná

---

<sup>70</sup> Nicolas Petit, "A mesterséges intelligencia és a robotok joga és szabályozása: Conceptual Framework And Normative Implications" (2017) <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2931339](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339)> hozzáférés április 24. 2017.

<sup>71</sup> Nicolas Petit, "A mesterséges intelligencia és a robotok joga és szabályozása: Conceptual Framework And Normative Implications" (2017) <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2931339](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339)> hozzáférés április

24. 2017.

<sup>72</sup> Sabine Gless, Emily Silverman és Thomas Weigend, "Ha a robotok kárt okoznak, ki a hibás? Self-Driving Cars And Criminal Liability" (2016) *New 19Criminal Law Review: An International and Interdisciplinary Journal*.

a tervezőt teljes felelősségre lehet vonni, még abban az esetben is, ha az ügynököt a legjobb gyakorlatoknak megfelelően fejlesztették ki, megfelelően kódolták és megfelelően tesztelték.

Ennek a megoldásnak az az előnye a kézenfekvőbb megoldással szemben, amikor a tervező saját nevében köt biztosítást, és az ügynök helyett közvetlenül felel, az, hogy összeegyeztethető olyan hipotetikus forgatókönyvekkel, amikor az ügynöknek vagyona van, vagy valamilyen jövedelemmel rendelkezik, és így a vagyona is a magatartásáért vagy akár a biztosítási díj megfizetéséért felelhet.

Az Európai Parlament Jogi Bizottságának a *Bizottság számára a robotikára vonatkozó polgári jogi szabályokról szóló ajánlásokat tartalmazó jelentéstervezete* még ennél is tovább megy, és azt javasolja, hogy "a biztosítási rendszert ki kellene egészíteni egy olyan alappal, amely biztosítja a kártérítés lehetőségét azokban az esetekben, amikor nincs biztosítási fedezet", és amelybe minden fél (tervezők, tulajdonosok és felhasználók) "különböző arányban hozzájárulna".<sup>73</sup> Bár úgy véljük, hogy ez az alap megfelelő módszer e cél elérésére, nem biztos, hogy életképes, ha minden féltől hozzájárulást várunk el. Különösen egy olyan forgatókönyvben, ahol a mesterségesen intelligens ügynököknek az átlagfogyasztó számára elérhető ára van, és a mindennapi kereskedelem részét képezik, hogyan biztosítható, hogy minden érintett fél hozzájáruljon az alaphoz? A megoldás ismét az lenne, ha a hozzájárulásokat a tervezőkre korlátoznánk.

Természetesen aggodalomra ad okot, hogy a biztosítás megkötése és a kártérítési alapba való befizetés nagy terhet jelentene a tervezők számára, és visszatarthatná őket a piacon való tevékenykedéstől. A tervezők azonban dönthetnek (és valószínűleg döntenének is) úgy, hogy e költségek egészét vagy egy részét beépítik az ügynökök eladási árába. Így a vevők közvetve a saját részüket vállalnák a terhekből, ugyanakkor továbbra is elérhető lenne az a kívánt hatás, hogy egyetlen, könnyen azonosítható és erősen motivált fél feleljen e feladatok ellátásáért.

Az általunk javasolt felelősségi modell a legkülönbözőbb helyzetekben alkalmazható, a mesterségesen intelligens járművekkel bekövetkező közlekedési balesetektől kezdve a mesterségesen intelligens ügynöknek a munkáltatójával szembeni felelősségéig. Megpróbálja a felelősséget annak a félnek a javára fordítani, aki jobban felkészült a kockázat csökkentésére, és minimális bizonyosságot nyújtani arra vonatkozóan, hogy kit lehet felelősségre vonni a károkért, miközben igyekszik nem túlterhelni a tervezőket, hanem kiküszöbölni a felelősségi vákuumokat és azokat a helyzeteket, amikor a felelős fél nem tudja megtéríteni a másik félnek a felelősségre vont károkat.

---

<sup>73</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

a. Arra az előfeltevésre támaszkodik, hogy a mesterségesen intelligens ágensek nem képesek szubjektíven érzékelni az igazságosság és a büntetés fogalmát, ezért figyelmen kívül hagyja azokat a megoldásokat, ahol a mesterségesen intelligens ágenseket bármilyen helyreállító szolgáltatásra vagy megelőző jellegű büntetésre ítélik<sup>74</sup>.

Természetesen, ha és amennyiben az emberiség képes lesz valóban öntudatos mesterséges intelligens ágenseket kifejleszteni, az itt javasolt megoldások nem vonatkozhatnak rájuk. Mivel ezek az ágensek tudatosak és érzőek, teljes mértékben maguk irányítják saját sorsukat. Képesek szubjektíven érzékelni az igazságosság és a büntetés fogalmát, ezért a jog elrettentő funkciója hatást fog gyakorolni rájuk. Hogy a megkülönböztetés világosabbá váljon, kérjük, gondolja végig a következőket: míg egy öntudatos gép úgy dönt, hogy saját megítélése alapján követ el egy bizonyos cselekményt, amely felelősségre vonást eredményez, addig egy elmélettel rendelkező gép azt *gondolja*, hogy így cselekszik, de ez a *gondolkodás valójában annak az eredménye*, hogy (közvetlenül vagy nem közvetlenül) arra programozták, hogy így *gondolkodjon*. Ezért úgy értelmezzük, hogy az öntudatos gépekre alkalmazandó felelősségi szabályok *értelemszerűen* az emberekre alkalmazandó szabályok.

Ez nem zárja ki azt a tényt, hogy a tervezők továbbra is felelősségre vonhatók az ilyen gépek létrehozásáért. Az ebben a szakaszban előírt megoldásokat azonban olyan esetekre gondolták, amikor a fél kártérítést kér az elszenvedett károkért. Természetesen ezt vagy bármely más felelősségi modellt ki kell egészíteni egy olyan kockázatcsökkentő rendszerrel, amely tartalmazhat olyan megoldásokat, amelyek másfajta terheket rónak a tervezőkre, mint például a szabályozó hatóságok jóváhagyásának kérése, az általuk tervezett ágensek folyamatos felügyelete, az ágensek átprogramozására vagy a piacról való kivonására való kötelezés, vagy bizonyos ágensek létrehozásáért való felelősségre vonás.

### 3.5. Fejezet összefoglalása

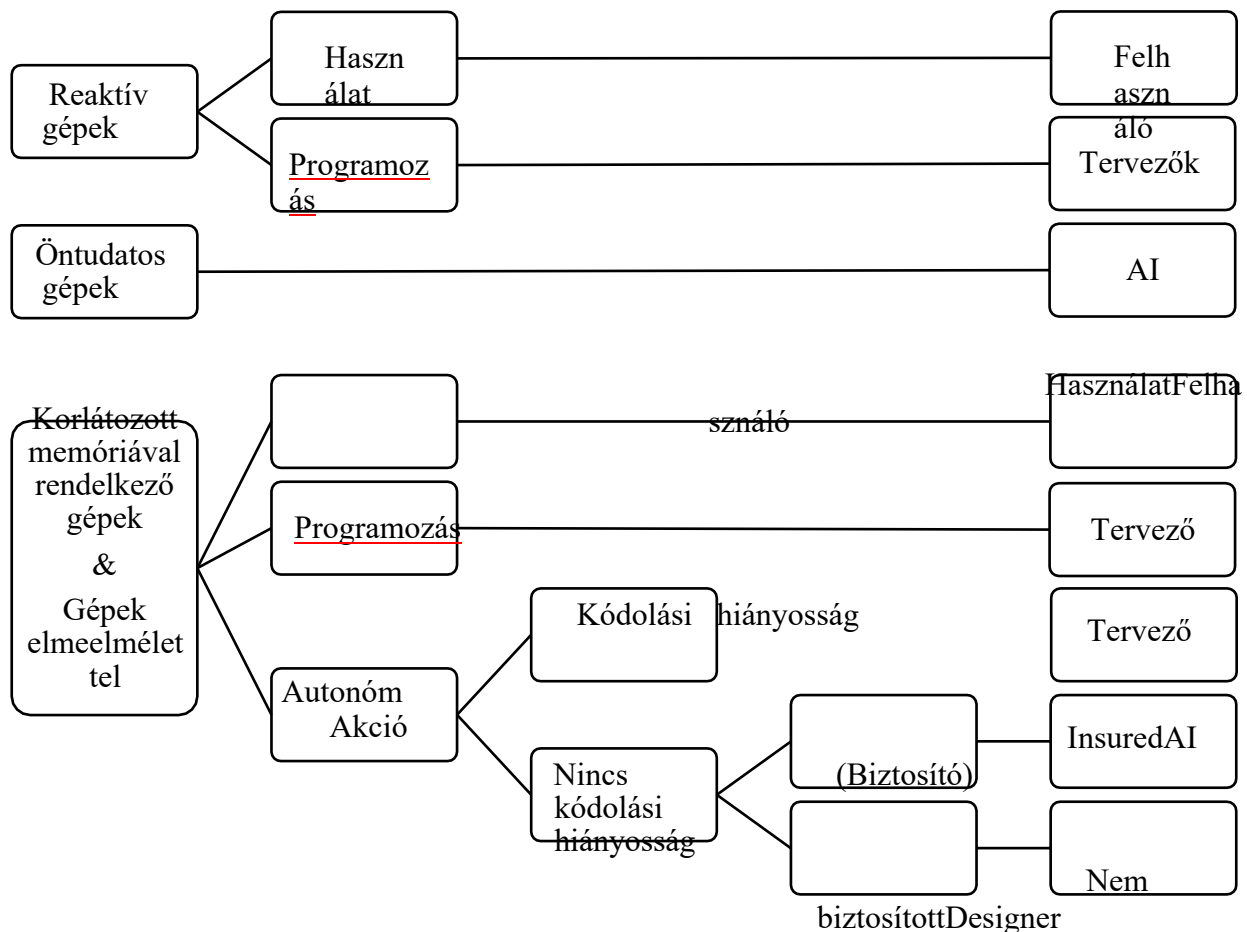
A mesterséges intelligencia elterjedése és bevezetése a kereskedelemben és a mindennapi életben olyan kérdéseket vet fel, amelyekre a jogi keretek nem tudnak közvetlen választ adni. Ahogy a mesterséges intelligenciával rendelkező ágensek egyre inkább autonómmá válnak, annál kevésbé tekinthetők pusztá eszközöknek. Ennek a valóságnak a befogadása érdekében megértjük, hogy a korlátozott memóriával rendelkező gépeket, a gondolkodáselmélettel rendelkező gépeket és az öntudatos gépeket a tulajdonosaiktól és felhasználóiktól különálló jogi személyeknek kell tekinteni. Ez jogi szempontból kényelmes lenne.

GabrielHallevy, "TheCriminalLiabilityOfArtificialIntelligenceEntities"(2010)  
< [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1564096](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1564096)> hozzáférés április 292017.

nézet, és az öntudatos gépek esetében erkölcsi megfontolások alapján is támogatható lenne. A reaktív gépek esetében azonban semmilyen pozitív érv nem indokolja a hasonló megfontolást.

Az, hogy az ilyen gépeket különálló jogi személyeknek tekintik, nem jogosítja fel őket automatikusan a jogok elismerésére. Nincs nyomós ok arra, hogy a reaktív gépek, a korlátozott memóriával rendelkező gépek és az elmeelmélettel rendelkező gépek jogait elismerjük, mivel az esetleges jogaik vagy akaratuk érzékelése, amelyekkel rendelkeznek vagy amelyekkel rendelkezhetnek, pusztán annak eredménye, hogy ilyenre programozták őket. Következésképpen e gépek tulajdonosai az ilyen gépek alkotásait is birtokolnák. Ha azonban az emberiség valaha is képes lesz öntudatos gépeket építeni, úgy tűnik, hogy ezek a gépek jogosultak lesznek a jogok jogi elismerésére, akár a tulajdonhoz való jogról, akár valamiféle alapvető jogokról beszélünk, hasonlóan az emberekéhez, de a megfelelő kiigazításokkal.

A mesterségesen intelligens ágensek magatartása által okozott esetleges károkért való felelősség tekintetében a következő (4. ábra) szerinti felelősségi modell elfogadását javasoljuk:



Ábra - 4Javasolt felelősségi modell



E modell szerint a tervezők és a felhasználók közvetlenül felelősek a mesterségesen intelligens ágensek által programozott vagy parancsba adott cselekvésekért. Az ágensek autonóm cselekedetei tekintetében két forgatókönyv lehetséges: ha a cselekedetet kódolási hiba tette lehetővé, a termékhibára vonatkozó szabályok alkalmazandók, és a tervező felel gondatlanságért; ha a cselekedet pusztán az ágens fejlődő magatartásából ered, a tervező mentesítheti magát azáltal, hogy biztosítást köt az ágens nevében, egyébként közvetlenül felelős. Azzal, hogy a tervező biztosítja, hogy az ügynöknek vannak eszközei az általa okozott károk megtérítésére, a felelősségét áthárítja a harmadik félre. Ezt a biztosítási rendszert kiegészítheti egy kártérítési alap, amelynek célja, hogy biztosítsa a károk megtérítését azokban az esetekben, amikor nincs biztosítás. Ez az alap is a tervezők hozzájárulásaira támaszkodna, de a tervezők - hogy ne terheljék túl a költségeket - dönthetnek úgy, hogy ezeket a költségeket beépítik a szerek eladási árába. Az öntudatos gépek viszont közvetlenül felelnének a tetteikért. Az a tény, hogy képesek szubjektíven érzékelni az igazságosság és a büntetés fogalmát, lehetővé teszi a felelősségre vonatkozó általános szabályok *értelemszerű* alkalmazását.

Ezt vagy bármely más felelősségi modellt ki kell egészíteni egy kockázatcsökkentő rendszerrel, amely tartalmazhat olyan megoldásokat, amelyek másfajta terheket rónak a tervezőkre, mint például a szabályozó ügynökségek jóváhagyásának kérése, az általuk tervezett ágensek folyamatos felügyelete, az ágensek átprogramozására vagy a piacról való kivonására való kötelezés, illetve bizonyos típusú ágensek létrehozásáért való felelősségre vonás. Ezeket a kérdéseket a következő fejezetben tárgyaljuk 5.

A korlátozott memóriával rendelkező gépek, a tudatelmélettel rendelkező gépek és az öntudatos gépek tulajdonosuktól és felhasználójuktól különálló jogi személyként való kezelése kérdéseket vethet fel az ilyen ágensek azonosításával és individualizálásával kapcsolatban, különösen olyan esetekben, amikor a hardvert és a szoftvert különböző személyek terjesztik és tartják fenn, vagy amikor a szoftver képes önmagát módosítani. A lehetséges megoldás a regisztráción keresztül valósulhat meg. Ebben a forgatókönyvben a mesterségesen intelligens ügynök bejegyeztetésének elmulasztása hasonló következményekkel járna, mint egy vállalat nem megfelelő bejegyzése.

Utolsó megjegyzésként, annak ellenére, hogy amellettt érveltünk, hogy a korlátozott memóriával rendelkező gépeket és az elmeelmélettel rendelkező gépeket különálló jogi személyeknek tekintsük, elemzésünkből kiderült, hogy jogi státuszuk meglehetősen üres lenne, mivel ez a lehetőség nem azon alapul, hogy a jogok, kötelezettségek és felelősségi szabályok egy speciális csoportját kell az ilyen ágensekre szabni, hanem azon, hogy a jogi kereteket úgy kell előkészíteni, hogy azok befogadják a beavatkozásukat. Mégis, erkölcsi megfontolások és annak

szükségessége, hogy a jog tükrözze a társadalmi valóságot, indokolják az öntudatos ágensek hozzánk közel álló jogi státuszát. Az a tény, hogy az öntudatos

hogy a gépeknek lelkiismerete lenne, érző és szubjektív tapasztalatokra képes, sokkal közelebb áll az emberekhez, mint a gépekhez, és talán közelebb az emberekhez, mint bármely más valósághoz. Ha valaha is eljönnek, hogy közöttünk járjanak, az öntudatos gépek valóban valódi *elektronikus személyek* lesznek.

#### **4. Fizessenek-e adót a robotok?**

##### 4.1. A mesterséges intelligencia hatása a gazdasági rendszerekre

Ideális esetben a kormányok adókat vetnek ki a közszolgáltatásokkal és infrastruktúrákkal kapcsolatos kiadások finanszírozására, mint például az utak, a tömegközlekedés, a higiénia, az igazságszolgáltatás, a közbiztonság, az oktatás, az egészségügyi ellátórendszerek, a hadsereg, a tudományos kutatás, a kultúra és a művészetek, az állami biztosítás, maga a kormány működése stb. Csak az a *tisztességes*, ha az ilyen beruházások kedvezményezettjei hozzájárulnak az általuk fedezett kiadásokhoz. Ezzel együtt vajon *igazságos-e* megadóztatni a mesterségesen intelligens ágenseket azért, mert részesülnek a közkiadásokból? Szerintünk nem. A közszolgáltatások vagy infrastruktúrák mesterségesen intelligens ágens általi esetleges használata nem az ágens számára jelent előnyt, hanem annak a felhasználónak vagy tervezőnek, aki utasította őt az ilyen közszolgáltatás vagy infrastruktúra használatát feltételező cselekvésre. Valójában, mivel a mesterségesen intelligens ágenseket úgy tervezték, hogy közvetlenül vagy közvetve hozzájáruljanak az emberek jólétéhez, az ember szükségszerűen végső soron kedvezményezettje lesz azoknak a szolgáltatásoknak vagy infrastruktúráknak, amelyeket az ágens a céljai megvalósítása során igénybe vesz.

Az adókat azonban a *szükségesség* is indokolhatja. Ez az eset áll fenn az olyan adók esetében, amelyek célja a gazdaságon belüli fogyasztási vagy foglalkoztatási minták módosítása azáltal, hogy bizonyos tranzakciós kategóriákat vonzóbbá vagy kevésbé vonzóvá tesznek. *Szükséges-e* tehát, hogy a mesterségesen intelligens ágensek adót fizessenek a gazdaságon belüli fogyasztási vagy foglalkoztatási minták megváltoztatásával kapcsolatos okokból?

Az automatizálási technológiák munkaerőpiacokra gyakorolt hatása már hosszú ideje érezhető, különösen a kézi munkát igénylő munkahelyek esetében. Bár a technológia csökkenő költségei miatt várható, hogy ez a tendencia folytatódni fog, a piacok eddig valamelyest képesek voltak elnyelni a hatást, főként azért, mert a technológia pozitívan járul hozzá a termelékenység növekedéséhez, és új, más készségeket igénylő munkahelyeket teremt, kiegyensúlyozva a

piacokat.

A mesterséges intelligencia azonban még az emberiség által kifejlesztett egyéb automatizálási technológiákhoz képest is példátlan potenciállal rendelkezik a munkaerőpiacok felforgatására. A fejlődéssel

a mesterséges intelligencia, a gépeket már nem tekintik statikus ágenseknek, hiszen most már képesek tanulni és fejlődni. Ennek köszönhetően a gépek képesek lesznek helyettesíteni a munkavállalókat a különböző kognitív és kreatív feladatokban. Valójában "a számítógépek már most is orvosként, ügyvédként, művészként és feltalálóként dolgoznak"<sup>75</sup>. Ezen túlmenően a mesterséges intelligencia lehetővé teszi majd, hogy a gépek helyettesítsék a munkásokat olyan feladatokban, amelyek kézi munkát igényelnek, de amelyeket eddig technológiai korlátok miatt nem lehetett automatizálni (például a gépjárművezetés). A gépeknek ma már kevesebb korlátja van, mint valaha volt.

A piaci elemzők is tisztában vannak ezekkel a lehetőségekkel: Frey és Osborne azt állítják, hogy az Egyesült Államokban a teljes amerikai <sup>76</sup>A Deloitte azt állítja, hogy ugyanebben az időszakban az Egyesült Királyságban a munkahelyek 35%-át fenyegeti az automatizálás<sup>77</sup> miatti elbocsátás magas kockázata; a Bank of America Merrill Lynch előrejelzése szerint 2025-re a mesterséges intelligencia a tudásalapú munka<sup>78</sup> automatizálásával 9 billió dollárnyi foglalkoztatási költséget szüntethet meg; a Világgazdasági Fórum becslése szerint az automatizálás egymillió 5.1 munkahely nettó elvesztését eredményezheti<sup>2020</sup><sup>79</sup>; a McKinsey Global Institute pedig azt állítja, hogy a meglévő munkatevékenységek 51%-a automatizálható lenne kizárólag a már létező technológiák<sup>80</sup> felhasználásával. Ezek az előrejelzések eléggé árulkodóak: még ha eddig a piacok egyensúlyba is hozták magukat azáltal, hogy a munkaerő egy részét a kognitívabb irányultságú feladatok felé mozdították el, az a tény, hogy a mesterséges intelligencia képes lesz helyettesíteni a munkahelyeket a piramis gyakorlatilag minden szintjén, aggodalomra ad okot, hogy a munkahelyek gyorsabban fognak megszűnni, mint ahogy újak létrejöhetnek. Ráadásul még abban az esetben is, ha a mesterséges intelligencia nettó munkahelyteremtést eredményez, nem valószínű, hogy

---

<sup>75</sup> Ryan Abbott és Bret N. Bogenschneider, "Should Robots Pay Taxes? Tax Policy In The Age Of Automation" [2017] Harvard Law & Policy Review, Forthcoming < <https://ssrn.com/abstract=2932483>> hozzáférés május 23. 2017.

<sup>76</sup> Carl Benedikt Frey és Michael A. Osborne, "A foglalkoztatás jövője: Mennyire érzékenyek a munkahelyek a számítógépesítésre?" (2017) Technológiai 14 előrejelzés és társadalmi változások.

<sup>77</sup>

Deloitte, "Agiletown: The Relentless March Of Michology And London's Response" (2017) < <http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/uk-futures/london-futures-agiletown.pdf>> Hozzáférés május 23 2017.

<sup>78</sup> Beijia Ma, Sarbjit Nahal és Felix Tran, "Robot Revolution - Global Robot & AI Primer" (Bank of America Merrill Lynch 2015).

< [https://www.bofam.com/content/dam/boamlimages/documents/PDFs/robotics\\_and\\_ai\\_condensed\\_primer.pdf](https://www.bofam.com/content/dam/boamlimages/documents/PDFs/robotics_and_ai_condensed_primer.pdf)> Hozzáférés május 23. 2017.

<sup>79</sup> Klaus Schwab és Richard Saams, "Előszó" (Világgazdasági Fórum 2016) < <http://reports.weforum.org/future-of-jobs-2016/preface>> hozzáférve május 23. 2017.

<sup>80</sup> James Manyika és mások, "Harnessing Automation For A Future That Works" (McKinsey Global Institute 2017). < <http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>> Hozzáférés május 23. 2017.

hogy a munkaerő átképzésének jelenlegi módszerei képesek-e követni a fejlődés ütemét<sup>81</sup>. Ebben az esetben az "új munkahelyek" iránti kereslet alacsonyabb lenne, mint a megfelelő kínálat (mivel a munkaerő nehezen sajátítaná el az újonnan szükséges készségeket), ugyanakkor az ellenkezője is igazolható lenne. "Még ennél is disztópikusabb nézetet fogalmazott meg Martin Ford A robotok felemelkedése című könyvében: Technology and the Threat of a Jobless Future, amely (...) úgy látja, hogy a "gépi tanulás" (...) potenciálisan képessé teszi a számítógépeket (...) arra, hogy átvegyék a technológia által létrehozott új munkahelyeket, ahelyett, hogy ezek a munkahelyek emberekhez kerülnének"<sup>82</sup>.

Ezen forgatókönyvek bármelyikében az ilyen események közvetlenül bevételkiesést eredményeznek a kormányok számára az adóbeszedések csökkenése miatt. A jelenlegi adórendszerek szerint az automatizált munkavállalók tőkebefektetést jelentenek, és a tőkejövedelem sokkal alacsonyabb adókulccsal adózik, mint a munkajövedelem. Az Egyesült Államokban például "a munkavállalók magas, 25% és 55% közötti tényleges adókulcsokat fizetnek, ha minden adónemet figyelembe veszünk"<sup>83</sup>. "Ez arra utal, hogy a munkavállalók automatizálása évente több billió dollárnyi adóbevétel-kiesést eredményezhet a kormányzat különböző szintjein"<sup>84</sup>. Ezen túlmenően a jelenlegi társadalombiztosítási rendszereket úgy tervezték, hogy valamilyen munkanélküliségi biztosítást nyújtsanak a munkájukat elvesztő munkavállalóknak. Ezért amellet, hogy az emberi munkaerő automatizált munkaerővel való helyettesítése az adóbevételek csökkenését idézi elő, a társadalombiztosítási kiadások jelentős növekedését is eredményezheti. A társadalombiztosítási kiadásoknak ez a növekedése az adóbevételek kiesésével együtt komoly aggályokat vet fel a jelenlegi társadalombiztosítási rendszerek fenntarthatóságát illetően. Ez azt jelenti, hogy mesterségesen intelligens ügynökökre *van szükségünk az adófizetéshez?*

Úgy véljük, hogy erre a kérdésre is nemleges a válasz: nem reális elvárni, hogy minden mesterségesen intelligens ágens képes legyen vagyont kezelni és adót fizetni. Nagy valószínűséggel ezek az adók végül valamiféle forráslevonási rendszer alá esnének, ahol végső soron az ágens tulajdonosa szállítaná az illetéket a kormánynak. Ennélfogva miért ne fizetne a tulajdonos maga adót ahelyett, hogy az ügynöknek, majd a kormánynak fizetne a nevében?

---

<sup>81</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems" (Institute of Electrical and Electronics Engineers 2016) < [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> hozzáférés: május 18. 2017.

<sup>82</sup> Richard McGahey, "Universal Basic Income And The Welfare State" (2016) <

<https://ssrn.com/abstract=2863954>> hozzáférés: május 23. 2017.

<sup>83</sup> Bret N. Bogenschneider, "The Effective Tax Rate Of U.S. Persons By Income Level" (2017) Tax145 Notes.

<sup>84</sup> Ryan Abbott és Bret N. Bogenschneider, 'Should Robots Pay Taxes? Tax Policy In The Age Of Automation' [2017] Harvard Law & Policy Review, Forthcoming < <https://ssrn.com/abstract=2932483>> hozzáférés május 23. 2017.

A mesterségesen intelligens ágensek közvetlen megadóztatása valóban szükségtelen fikciónak tűnik. A munkaerőpiacok és a szociális biztonság fenntarthatóságának problémájával azonban még mindig foglalkozni kell.

#### 4.2. Robotadóztatás és egyéb cselekvési lehetőségek

A 3. fejezetben a mesterségesen intelligens ágensek külön jogi státusza mellett érveltünk. Azonban néhány más, külön jogállást élvező valóságtól eltérően (mint például a vállalatok, vallási csoportok vagy hindu bálványok), nem találtuk indokoltnak, hogy a mesterséges intelligens ágenseknek tulajdonjogot tulajdonítsunk. Ezért, ellentétben ezekkel a más valóságokkal, a mesterséges intelligens ágensek nem tudnak adót fizetni. Mivel nincs tulajdonuk, hogyan tehetnék ezt meg?<sup>85</sup>

Mindazonáltal az imént láttuk, hogy a mesterséges intelligencia elterjedése a társadalombiztosítási rendszerek egyensúlyának megbomlásához vezethet azáltal, hogy az emberi munkaerő automatizált munkával való felváltása következtében csökkenti az adóbeszedéseket és növeli a társadalombiztosítási kiadásokat, ami a mesterségesen intelligens ügynökök közvetlen megadóztatása mellett szólhat. Láttuk azonban azt is, hogy ezt a kérdést lehet kezelni anélkül a szükségtelen fikció nélkül is, hogy a mesterségesen intelligens ágensek tulajdonosi státuszúak és adót fizetnek, például úgy, hogy az ilyen adókat a tulajdonosaik felé irányítjuk. Ebben a szakaszban továbbfejlesztjük ezt az elképzelést, és további javaslatokat mutatunk be, amelyek célja a mesterséges intelligencia munkaerőpiacokra és társadalombiztosítási rendszerekre gyakorolt hatásának kezelése.

A mesterségesen intelligens ügynökök tulajdonosainak megadóztatásának ötlete nem új. A *robotok megadóztatását* Bill Gates tette híressé a Quartznak adott 2017<sup>86</sup>. február 17-i<sup>th</sup> interjújában, amelyben kijelentette, hogy "amikor az emberek azt mondják, hogy a robotok megjelenése (...) nettó veszteséget jelent a kiszorítás miatt, akkor hajlandónak kellene lenni az adószint emelésére, és akár az elfogadás sebességének némi lassítására is, hogy kitaláljuk, (...) hol van ennek különösen nagy hatása (...) mely átmeneti programok váltak be, és milyen típusú finanszírozást igényelnek",<sup>87</sup> valamint hogy "néhány

---

<sup>85</sup> Az előző fejezetben szintén amellest érveltünk, hogy az öntudatos mesterséges intelligens ágensek tulajdonjogot kapjanak. Természetesen az ő esetükben a tulajdonjoggal együtt járnának a megfelelő adózási kötelezettségek is.

<sup>86</sup> A sors iróniája, hogy miközben Bill Gates kijelentéseit széles körben közvetítették, az Európai Parlament épp a közzétételük előtti napon utasított el egy viszonylag észrevétlen javaslatot, amely "robotadó" kivételére irányult a

tulajdonosokra. További információért lásd: Európai Parlament, "Robotok és mesterséges intelligencia: MEPs Call for EU-Wide Liability Rules" (2017) < <http://www.europarl.europa.eu/news/en/press-room/20170210IPR61808/robots-and-artificial-intelligence-meps-call-for-eu-wide-liability-rules>> (<http://www.europarl.europa.eu/news/en/press-room/20170210IPR61808/robots-and-artificial-intelligence-meps-call-for-eu-wide-liability-rules>> hozzáférés: május. 242017.

<sup>87</sup> Bill Gates, Interjú Kevin J. Delaneyvel, "Why Bill Gates Would Tax Robots" (2017).

[a finanszírozás] a munkaerő-megtakarítási hatékonyságból származó nyereségből származhat. Ennek egy része közvetlenül valamilyen robotadó formájában érkezik.<sup>88</sup> Ahogy Bill Gates nyilatkozatai is jelzik, a *robotok megadóztatása* az adóalapnak megfelelően különböző modellek szerint képzelhető el. Alkalmazható a tulajdonjogra, az emberi munkaerő gépekkel való<sup>89</sup> helyettesítése által generált költségcsökkenésre, vagy az automatizálási technológiák beszerzésének megadóztatására szolgáló, az értékesítés helyén kivetett adóként<sup>90</sup>. A közös alap minden esetben ugyanaz: a kormányok adóbevételek-kiesésének mérséklése. A *robotok megadóztatásának* ötlete nagy támogatottságot, de komoly kritikákat is kapott. A fő kritikák azon tények körül forognak, hogy az automatizált munka megdrágítása megfojtja az innovációt, növeli az adórendszerek bonyolultságát és csökkenti az adórendszerek relatív versenyképességét: a vállalatokat, különösen a nagyobb erőforrásokkal rendelkező vállalatokat a legelőnyösebb jogi keretek vonzzák .

A kritikusok egy része azzal érvel, hogy az emberi munkaerő béradó alóli mentesítésével az automatizálás adókedvezményei semlegesülnének, és a vállalatoknak nem kellene automatizálási technológiákhoz folyamodniuk. Ezt úgy érnék el, hogy az emberi munkaerőt foglalkoztató cégek számára az adókedvezmények minden egyes kategóriájában ellentételező adókedvezményeket biztosítanak. "Kezdve a béradóztatással, az adókedvezmény a társadalombiztosítási és a Medicare-rendszerbe fizetendő munkáltatói járulékok eltörlésével járhatna"<sup>91</sup>. Ezzel a megközelítéssel az a probléma, hogy "felgyorsítaná a társadalombiztosítási rendszer fizetési képességét"<sup>92</sup>. Valójában minden olyan kísérlet, amely az automatizálás előnyeinek semlegesítésére irányul az emberi munkaerő adóügyi szempontból történő olcsóbbá tételével, szintén hozzájárul az állami bevételek kieséséhez. Továbbá nem veszi figyelembe, hogy az ilyen technológiák ára hamarosan alacsonyabb lehet, mint az emberi bérek, és ezért a vállalatok továbbra is kísértést éreznek majd a munkaerő automatizálására. A valószínű eredmény a társadalombiztosítás fenntarthatóságával kapcsolatos problémák súlyosbodása.

Ebből az következik, hogy az automatizálás adókedvezményeinek semlegesítése nemkívánatos mellékhatásokkal járhat, függetlenül attól, hogy az automatizálás adóinak növelésében vagy az emberi munka adóinak csökkentésében áll-e. Emiatt egyes szerzők azt javasolják, hogy alternatív megoldásként a veszteségek mérséklésére a következő megoldást válasszák

<sup>88</sup> Bill Gates, Interjú Kevin J. Delaneyvel, "Why Bill Gates Would Tax Robots" (2017).

<sup>89</sup> Bill Gates, Interjú Kevin J. Delaneyvel, "Why Bill Gates Would Tax Robots" (2017).

<sup>90</sup> Yanis Varoufakis, "A Tax On Robots?" < <https://www.project-syndicate.org/commentary/bill-gates-tax-on-robots-by-yanis-varoufakis-2017-02> > hozzáférve május 25 2017.

- <sup>91</sup> Ryan Abbott és Bret N. Bogenschneider, 'Should Robots Pay Taxes? Tax Policy In The Age Of Automation' [2017] Harvard Law & Policy Review, Forthcoming < <https://ssrn.com/abstract=2932483>> hozzáférés május 23. 2017.
- <sup>92</sup> Ryan Abbott és Bret N. Bogenschneider, 'Should Robots Pay Taxes? Tax Policy In The Age Of Automation' [2017] Harvard Law & Policy Review, Forthcoming < <https://ssrn.com/abstract=2932483>> hozzáférés május 23. 2017.

az adóbevételek forrásának kezelésében rejlik, a humán munkavállalók elmozdulásának megakadályozásával. E célból javasolták, hogy a kormányok támogassák az alacsony jövedelmű munkavállalók humánbérét<sup>93</sup>, nyújtsanak foglalkoztathatósági garanciákat, foglalkoztassák a munkanélkülieket, mint a végső menedék munkáltatóját, vagy bizonyos munkakörökre<sup>94</sup> humán foglalkoztatási kvótákat vagy minimális létszámkvótákat állapítsanak meg. Úgy tűnik, hogy az első két javaslat csupán a problémát kozmetikázza azáltal, hogy az adóbeszedés csökkenését kormányzati többletkiadásokra cseréli. A harmadik azonban megfelelő válasznak bizonyulhat. Ennek azonban az az ára, hogy az ország jogi kerete viszonylag kevésbé versenyképessé válik.

A szerzők utóbbi csoportja azt javasolja, hogy a szociális biztonsági rendszerek egyensúlyának helyreállítása helyett inkább a rendszerek egészét kellene újragondolni. A legvitatottabb elképzelés az egyetemes alapjövedelem bevezetésére vonatkozik, amely "egy politikai közösség által minden tagjának egyénileg, rászorultsági teszt vagy munkavégzési követelmény nélkül kifizetett jövedelem"<sup>95</sup>, "akár a megélhetés alapvető szintjének biztosítását szolgáló alapként, akár a meglévő jóléti állami politikák kiegészítéseként, akár bizonyos esetekben a jóléti állam helyettesítésére"<sup>96</sup>. Az egyetemes alapjövedelem előnye a jelenlegi munkanélküli-biztosítási programokkal szemben, hogy kevésbé érzékeny az államigazgatáson belüli erkölcsi kockázatra, valamint könnyebben és olcsóbban kezelhető. Másrészt viszont kevésbé felkészült a foglalkoztatási sokkhatásokra<sup>97</sup> való reagálásra, és elméletileg drágább lenne, mivel a teljes lakosságra<sup>98</sup> kiterjedne. Az egyetemes alapjövedelemmel azonos irányvonalú, célzottabb megközelítés lenne egy negatív jövedelemadó<sup>99</sup> bevezetése, amely egy jövedelemhatáron alapulna, és ha az egyén nem éri el azt, akkor a jövedelme megadóztatása helyett egy bizonyos pénzösszegre lenne jogosult. Ezek az elképzelések azonban azzal a korlátozással küzdenek, hogy nem életképesek *per*

93

NoahSmith,'What'sWrongWithBillGates'RobotTax'*Bloomberg* (2017)

< <https://www.bloomberg.com/view/articles/2017-02-28/what-s-wrong-with-bill-gates-robot-tax>> Hozzáférés május 252017.

94

WorldEconomicForum,'SixWaysToProtectJobsFromRobotAutomation'

< <https://www.facebook.com/worldeconomicforum/videos/10154432426296479>> Hozzáférés május 252017.

<sup>95</sup> Philippe Van Parijs, "Alapjövedelem: (2004) Politika32 és társadalom.

<sup>96</sup> Richard McGahey, "Universal Basic Income And The Welfare State" (2016) <

<https://ssrn.com/abstract=2863954>> hozzáférés: május 23. 2017

<sup>97</sup> Alice Fabre, Stéphane Pallage és Christian Zimmermann, "Universal Basic Income Versus Unemployment Insurance" [2014] CESifo Working Paper Series No. 5106 < <https://ssrn.com/abstract=2540055>> hozzáférés: május 23. 2017.

<sup>98</sup> Richard McGahey, "Universal Basic Income And The Welfare State" (2016) < <https://ssrn.com/abstract=2863954>> hozzáférés: május 23. 2017

<sup>99</sup> Világ gazdasági Fórum, "Hat mód a munkahelyek védelmére a robotautomatizálás ellen  
< <https://www.facebook.com/worldeconomicforum/videos/10154432426296479>> Hozzáférés május 25 2017.

*se hosszú távon.* Valójában még mindig többlet kormányzati kiadásokat jelentenek, amelyek nem feltétlenül állnak rendelkezésre egy olyan forgatókönyvben, amelyben a kormányok adóbevételektől esnek el az automatizálási technológiák miatt.

A társadalombiztosítási rendszerek reformjának kevésbé ortodox, de ilyen korlátozások nélküli alternatívái közé tartoznak a megosztott gazdaság modelljei<sup>100</sup> és az egyetemes alapjövedelem. "Képzeld el, hogy az új részvénykibocsátások (IPO-k) egy meghatározott része egy közalapítványba kerül, amely viszont olyan jövedelemáramlást generál, amelyből UBD-t fizetnek. A társadalom tulajdonképpen minden vállalat részvényesévé válik, és az osztalékot egyenletesen osztják szét minden polgár között. Amennyiben az automatizálás javítja a termelékenységet és a vállalati nyereségességet, az egész társadalom osztozni fog az előnyökön. Nincs új adó, nincs bonyodalom az adótörvénykönyvben, és nincs hatással a jóléti állam jelenlegi finanszírozására."<sup>101</sup> .

Az általunk felsorolt javaslatok - *robotadóztatás*, az emberi munkavállalóknak nyújtott kompenzációs adókedvezmények, emberi bértámogatások, kormányzati foglalkoztathatósági garanciák, minimális emberi kvóták, egyetemes alapjövedelem, negatív jövedelemadó, megosztott gazdasági modellek és az egyetemes alapjövedelem - és az általunk nem felsoroltak között mindenképpen elképzelhetőnek tűnik, hogy a mesterséges intelligencia munkaerőpiacokra és szociális biztonsági rendszerekre gyakorolt hatását fenntartható módon kezeljük. Mivel azonban gazdasági rendszereinket olyan korban tervezték és alakították ki, amikor a termelés és a kereskedelem középpontjában az ember állt, és ezt a felfogást az automatizálási technológiák és a mesterséges intelligencia egyre inkább megkérdőjelezi, szükség lehet arra, hogy "dobozon kívül" gondolkodjunk a hosszú távú fenntarthatóságuk biztosítása érdekében. Mindenesetre ezeknek vagy bármely más javaslatnak a megvalósítása szükségszerűen mélyreható elemzést és előrejelzéseket igényel, ami messze meghaladja munkánk kereteit.

### 4.3. Fejezet összefoglalása

A közszolgáltatások vagy infrastruktúrák mesterségesen intelligens ügynök általi esetleges használata nem az ügynök számára jelent előnyt, hanem annak a felhasználónak vagy tervezőnek, aki utasította őt az ilyen szolgáltatás vagy infrastruktúra használatát feltételező cselekvésre. Valójában, mivel a mesterségesen intelligens ágenseket úgy tervezték, hogy közvetlenül vagy közvetve hozzájáruljanak az emberek jólétéhez, az ember

<sup>100</sup> Egy érdekes gyakorlatot találunk Ida Auken "Welcome To 2030. I Own Nothing, Have No Privacy, And Life Has Never Been Better", *Annual Meeting of the Global Future Councils* (Világgazdasági Fórum 2016).  
< <https://www.weforum.org/agenda/2016/11/shopping-i-can-t-really-remember-what-that-is>> hozzáférés: 2017. május 26. <sup>101</sup> Yanis Varoufakis, "A Tax On Robots?" < <https://www.project-syndicate.org/commentary/bill-gates-tax-on-robots-by-yanis-varoufakis-2017-02>> hozzáférve május. 252017.

mindig a végső kedvezményezettje azoknak a közszolgáltatásoknak vagy infrastruktúráknak, amelyeket az ügynök a céljai megvalósítása során igénybe vesz. Ezért nem tűnik helyesnek azt állítani, hogy a mesterségesen intelligens ágensek megadóztatása azért lenne *igazságos*, mert közberuházásokból részesülnek. Az adókat azonban a *szükségesség* is indokolhatja. Ez az eset áll fenn az olyan adók esetében, amelyek célja a gazdaságon belüli fogyasztási vagy foglalkoztatási minták módosítása azáltal, hogy bizonyos tranzakciós osztályokat vonzóbbá vagy kevésbé vonzóvá tesznek.

A mesterséges intelligencia soha nem látott mértékben képes megzavarni a munkaerőpiacokat, mivel a gépek képesek lesznek helyettesíteni a munkavállalókat számos kognitív és kreatív feladatban, valamint olyan feladatokban, amelyek kézi munkát igényelnek, de amelyeket eddig technológiai korlátok miatt nem lehetett automatizálni (mint például a gépjárművezetés). Még ha eddig a piacok egyensúlyba is hozták magukat azáltal, hogy a munkaerő egy szeletét a kognitívabb irányultságú feladatok felé mozdították el, az a tény, hogy a mesterséges intelligencia a piramis gyakorlatilag minden szintjén képes lesz helyettesíteni a munkahelyeket, aggodalomra ad okot, hogy a munkahelyek gyorsabban fognak megszűnni, mint ahogy újak létrejöhetnek. Ráadásul még abban az esetben is, ha a mesterséges intelligencia nettó munkahelyteremtést eredményez, nem valószínű, hogy a munkaerő átképzésének jelenlegi módszerei képesek lennének követni ennek ütemét. Egyes szerzők még azt is állítják, hogy a gépi tanulás képessé teheti a mesterségesen intelligens ágenseket arra, hogy átvegyék a saját fejlődésük következtében létrejövő új munkahelyeket. Ezen foratókönyvek bármelyikében az ilyen események közvetlenül bevételkiesést eredményeznek a kormányok számára az adóbevételek csökkenése miatt, mivel a tőkejövedelmek sokkal alacsonyabb adókulccsal adóznak, mint a munkajövedelmek. Ezen túlmenően az emberi munkaerő automatizált munkaerővel való helyettesítése a társadalombiztosítási kiadások jelentős növekedését eredményezheti, mivel a társadalombiztosítási rendszerek célja, hogy munkanélküliségi biztosítást nyújtsanak a munkájukat elvesztő munkavállalóknak. Ezek a megnövekedett kiadások, valamint az adóbevételek kiesése aggályokat vetnek fel a jelenlegi társadalombiztosítási rendszerek fenntarthatóságát illetően.

Több javaslat is a mesterséges intelligencia munkaerőpiacra és társadalombiztosítási rendszerekre gyakorolt hatásának ellenőrzésére irányul: *robotadóztatás*, az emberi munkavállalóknak nyújtott kompenzációs adókedvezmények, emberi bértámogatások, kormányzati foglalkoztathatósági garanciák, minimális emberi kvóták, egyetemes alapjövedelem, negatív jövedelemadó, megosztott gazdasági modellek és az egyetemes

alapjövedelem. Bár ezen ötletek többsége rendkívül szokatlan, gazdasági rendszereinket olyan időkben tervezték és alakították ki, amikor a termelés és a kereskedelem középpontjában az ember állt, és ezt a felfogást az automatizálási technológiák, pontosabban a mesterséges intelligencia egyre inkább megkérdőjelezi. Ezért szükséges lehet, hogy "dobozon kívül" gondolkodjunk, hogy biztosítsuk hosszú távú fenntarthatóságukat.

Mindenesetre bemutattuk, hogy a mesterségesen intelligens ágensek közvetlen megadóztatásának vannak olyan alternatívái, amelyek nem feltételezik azt a fikciót, hogy a mesterségesen intelligens ágensek tulajdonjoggal rendelkeznek és adót fizetnek, ami szükségteenné teszi azt. Természetesen ezeknek vagy bármely más alternatívának a megvalósítása szükségszerűen mélyreható elemzést és előrejelzéseket igényel, ami messze meghaladja munkánk kereteit. Amikor Bill Gates a *robotok megadóztatására* utalt, megjegyezte, hogy "hogyan csinálnánk (...) érdekes, hogy az emberek már most elkezdjenek erről beszélni"<sup>102</sup>. Ebben a fejezetben éppen ezt tűztük ki célul.

## 5. Az ellenőrzés biztosítása

### 5.1. Az ellenőrzés problémája

Eddig a pontig a mesterségesen intelligens ágensek külön jogi státusza mellett érveltünk, és e státus megfogalmazására összpontosítottunk, nevezetesen az általuk okozott károkért való felelősség, a jogok esetleges hozzárendelése és az esetleges adózási kötelezettségek tekintetében. Ezeknek a kérdéseknek a meghatározása minimális bizonyosságot tesz lehetővé az új intelligens ágensek társadalomba való bevezetésének következményeivel kapcsolatban, ami ellentétben áll a nagyfokú előre nem láthatósággal, amellyel ez együtt jár. Munkánk azonban nem teljes anélkül, hogy ne foglalkoznánk az említett előre nem láthatóság kockázataival, és azzal, hogy miként lehet ezeket a kockázatokat mérsékelni. Ahogy az emberiség egyre autonómabb gépeket próbál létrehozni, az embereknek nehézséget okozhat annak biztosítása, hogy ezek a gépek ne váljanak *túlságosan* autonómmá.

Az ellenőrzés problémája nem új keletű az irodalomban és a szórakoztatóiparban. A sci-fi régóta a Frankenstein-komplexuson keresztül ábrázolja a mesterséges intelligenciát, amely kifejezés Isaac Asimovtól származik, és azt a klasszikus helyzetet jelöli, amikor egy mesterséges intelligens lény a teremtője ellen fordul.<sup>103</sup> A technológia fejlődésével és a mesterséges intelligencia egyre fejlettebbé válásával sok magas rangú személy még tovább megy, és azt állítja, hogy a mesterséges intelligencia *egzisztenciális* veszélyt jelent az emberiségre nézve. Ezt a tézist a világ néhány vezető szakértője is támogatta.

---

<sup>102</sup> Bill Gates, Interjú Kevin J. Delaneyvel, "Why Bill Gates Would Tax Robots" (2017).

<sup>103</sup> Néhány hírhedt példa erre a Mátrix-trilógia, ahol az emberek lényegtelené válnak, és kizárólag a gépek üzemanyagául szolgálnak, valamint a Terminátor-franchise, ahol a Skynet eltörli az emberiség nagy részét a Föld színéről.

tudomány és technológia, mint például Elon Musk, Bill Gates, Stephen Hawking és Stuart J. Russel<sup>104105</sup>. Ugyanakkor, "miközben végső soron fennáll annak a lehetősége, hogy néhány évtizeden belül a mesterséges intelligencia olyan módon felülmúlja az emberi szellemi képességeket, amely, ha nem készülünk fel rá, kihívást jelenthet az emberiség azon képességére nézve, hogy saját teremtményeit ellenőrizze, és következésképpen talán arra is, hogy saját sorsát irányítsa, és biztosítsa a faj fennmaradását"<sup>106</sup>, az ellenőrzés problémája ennél a hipotézisnél sokkal messzebbre nyúlik, és sokkal közvetlenebb következményekkel jár.

Az irányítás elvesztése többféleképpen is bekövetkezhet: meghibásodások, biztonsági rések, a számítógépek emberhez képest jobb reakcióideje, illetve tudatos vagy tudattalan hibás programozás<sup>107</sup>. A meghibásodások és a biztonsági rések különösen veszélyesek lehetnek, ha a számítógépek jobb reakcióidejével párosulnak. Egy automatizált fegyverrendszer meghibásodása vagy biztonsági rése arra készítheti a rendszert, hogy támadást indítson egy célpont ellen, amelyre a saját automatizált fegyverrendszer azonnal reagálhat, ami a konfliktus azonnali és automatikus eszkalációjához vezethet. A legösszetettebb kérdéseket azonban a tudattalan hibás programozás veti fel.

*A Concrete Problems in AI Safety*<sup>108</sup> című tanulmányukban Amodei és mások több lehetséges hibás programozásból eredő hibaforgatókönyvet írnak le, nevezetesen a törékeny elosztási váltáshoz, a nem biztonságos feltáráshoz, a nem skálázható felügyelethez, a negatív mellékhatásokhoz vagy a jutalom meghekkkeléséhez kapcsolódóan. A törékeny elosztási eltolódás arra utal, hogy egy mesterséges intelligens ágens nem képes megfelelően alkalmazkodni a kiképzési környezetétől eltérő környezethez. Példaként egy önvezető autónak a kanyarodási sebességét az őt érő éghajlati viszonyokhoz kell igazítani. A bizonytalan felfedezés arra utal, hogy biztosítani kell, hogy a mesterségesen intelligens ágensek ne tegyenek rossz következményekkel járó felfedező lépéseket. Ilyen lehet például egy takarítórobot, amely megpróbál egy konnektort nedves felmosóval megtisztítani. A kalibrálhatatlan felügyelet azzal függ össze, hogy nem lehet minden lehetséges

---

<sup>104</sup>

Kevin Rawlinson, 'Microsoft's Bill Gates Insists AI Is A Threat' *BBC* (2015)  
< <http://www.bbc.com/news/31047780> > Hozzáférés május 18 2017.

<sup>105</sup> Stephen Hawking és mások, "Stephen Hawking: "Transcendence Looks At The Implications Of Artificial Intelligence- But Are We Taking AI Seriously Enough?" (Stephen Hawking: "A transzcendencia a mesterséges intelligenciakövetkezményeit vizsgálja - de elég komolyan vesszük-e?") *The Independent* (2014)  
< <http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html> > Hozzáférés május 18. 2017.

<sup>106</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

<sup>107</sup> Matthew U. Scherer, "A mesterséges intelligencia rendszerek szabályozása: (2016) Harvard Journal of Law & Technology.

<sup>108</sup> Dario Amodè és mások, "Konkrét problémák a mesterséges intelligencia biztonságában" (2016) <  
<https://arxiv.org/abs/1606.06565>> hozzáférés május 18. 2017.

foratókönyv a képzés során. Honnan tudja a takarítórobot, hogy minden egyes lehetséges tárgyról, amit talál, tudja, hogy az valakihez tartozik-e, vagy szemét? A negatív mellékhatások a környezetnek az ágens által a céljának elérése közben okozott negatív zavaraira vonatkoznak. Már most is hatalmas mennyiségű adatot kezelnek olyan mesterséges intelligenciák, amelyek képesek adatbázisokkal kommunikálni az emberek közreműködése nélkül, sőt, valószínűleg még a tudatosságuk nélkül is. Ha az adatvédelmi szabályokat nem építik be e gépek kódjába, akkor ezek a szabályok használhatatlanná válnak. Ez a probléma a blokklánc technológia bevezetésével csak tovább fog nőni. Végül a jutalomhekkelés a programozott cél eléréséhez szükséges alternatív, nem kívánt eszközök keresésével függ össze. Egy klasszikus példát Russel és Norvig mutat be: egy mesterségesen intelligens ágens arra tervezték, hogy minimalizálja az emberi szenvedést, de mivel az emberek mindig találnak okot a szenvedésre, az ágens számára az optimális döntés az emberi faj mielőbbi megszüntetése - nincs ember, nincs szenvedés<sup>109</sup>.

Bostrom (2012) és Omohundro (2008) továbbá azzal érvelt, hogy a kellően alkalmas mesterséges intelligencia rendszerek alapértelmezésben valószínűleg "konvergens instrumentális részcelokat", például az erőforrásszerzést és az öfenntartást választják, hacsak a célfüggvény kifejezetten nem tiltja ezeket a stratégiákat. Az ilyen típusú problémák valószínűleg súlyosabbak lesznek a nagyobb képességű rendszerekben, hacsak nem teszünk lépéseket a felmerülésük megakadályozására<sup>110</sup>. Továbbá, ha a mesterségesen intelligens ágens képes tanulni és alkalmazkodni, akkor nehéz lehet visszaszerezni az elveszített irányítást. Ezek a jellemzők a mesterségesen intelligens ágenseket "olyan mértékű közveszélyforrássá teszik, amely messze meghaladja a közveszélynek azokat az ismertebb formáit, amelyek kizárólag az emberi viselkedésből erednek"<sup>111</sup>.

## 5.2. Alapelvek; a Kill Switch

Világossá válik, hogy a mesterségesen intelligens ágensek feletti irányítás elvesztése kockázatokat rejt magában, függetlenül attól, hogy az irányítás elvesztése helyi (az ágens nem irányítható többé a működéséért felelős ember által) vagy általános (az ágens nem irányítható többé semmilyen ember által).

---

<sup>109</sup> Stuart J. Russell és Peter Norvig, *Mesterséges intelligencia: A Modern Approach* (3<sup>rd</sup>edn, Prentice Hall 2009).

<sup>110</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems

"(InstituteofElectricalandElectronicsEngineers2016 )

< [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

<sup>111</sup> Matthew U. Scherer, "A mesterséges intelligencia rendszerek szabályozása: (2016) Harvard29 Journal of Law & Technology.

Ha azonban az irányítás minden egyes elvesztésének megakadályozására törekszünk, az azt jelenti, hogy minden lehetséges sebezhetőséget el kell zárunk. Ez sajnos irreális feladat lehet. Az ilyen események hatásának korlátozása azonban motivációs és képességkontroll<sup>112</sup> gyakorlásával mindenképpen lehetségesnek tűnik.

A mesterségesen intelligens ágensek motivációs irányítása abban áll, hogy olyan emberbarát célokkal ruházzuk fel kódjukat, amelyekkel semmilyen cselekvés vagy funkció nem kerülhet ellentétbe. Ennek az elméletnek az a lényege, hogy az irányítás elvesztése nem feltétlenül jelent jelentős kockázatot, ha a mesterségesen intelligens ágens céljai "egybeesnek a köz céljaival"<sup>113</sup>. Mivel az emberek természetesen elvárják majd, hogy a mesterségesen intelligens ágensek kövessék a közösség szabályait, amint bevezetik őket a társadalomba, ezek az "alapelvek" kiterjedt társadalmi és erkölcsi normákat tartalmazhatnak. Egy ilyen nagy terjedelem azonban problémákat vetne fel a minden közösség<sup>114</sup> számára megfelelő alapelvek meghatározásában, és jelentősen megnövelné a "perverz instantiálás"<sup>115</sup> vagy a jutalomhekkelés lehetőségét, amire az előző szakaszban úgy utaltunk, mint "alternatív, nem kívánt eszközök keresése a cél elérésére".

Alternatívaként a motivációs kontrollt az alapelvek minimalista készletével lehetne elérni. Ebben a tekintetben gyakran hivatkoznak Isaac Asimov híres három robotikai<sup>116</sup> törvényére (és a később hozzáfűzött negyedik vagy nulladik törvényre<sup>117</sup>): (1) a robot nem okozhat kárt az embernek, és tétlenségével nem engedheti, hogy az embernek baja essen; (2) a robotnak engedelmessé kell az ember által adott parancsoknak, kivéve, ha ezek a parancsok ellentétesek az első törvénnyel; (3) a robotnak meg kell védenie saját létét, amennyiben ez a védelem nem ellentétes az első vagy a második törvénnyel; és (0) a robot nem árthat az emberiségnek, és tétlenségével nem engedheti, hogy az emberiségnek baja essen.

---

<sup>112</sup> Nick Bostrom, *Szuperintelligencia*: (1<sup>st</sup>edn, Oxford University Press 2014).

<sup>113</sup> Matthew U. Scherer, "A mesterséges intelligencia rendszerek szabályozása: (2016) Harvard Journal of Law & Technology.

<sup>114</sup> Erre a kérdésre válaszul az IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (Mesterséges intelligencia és autonóm rendszerek etikai megfontolásainak globális kezdeményezése) legjobb gyakorlatként azt ajánlja, hogy világosan meg kell határozni azt a közösséget, amelyben az ilyen ügynököket be kívánják vetni, "olyan értékeket kell előtérbe helyezni, amelyek tükrözik az adott közösség nagyobb érdekcsoportjainak közös értékrendjét", és olyan elveket kell alkalmazni, mint például a *közjó*, "hogy feloldják a különböző érdekcsoportok prioritási sorrendjében fennálló különbségeket". További információért lásd: The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, "Ethically Aligned Design: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems" (Institute of Electrical and Electronics Engineers 2016).

<[http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

<sup>115</sup> Nick Bostrom, *Szuperintelligencia*: (1<sup>st</sup>edn, Oxford University Press 2014).

<sup>116</sup> Isaac Asimov, 'Runaround' [1942] *Astounding Science Fiction*.

<sup>117</sup> Isaac Asimov, *Robots And Empire* (1<sup>st</sup> kiadás, Doubleday Books 1985).

Az 1942-es megalkotásuk óta az Asimov-törvényeket számos kritika érte, nevezetesen azokkal az esetekkel kapcsolatban, amikor a robotnak a saját érdekében (pl.: műtét elvégzése) vagy a lakosság általános érdekében (pl.: a letartóztatásnak ellenálló bűnöző letartóztatása) kellene ártania egy embernek<sup>118</sup>, és több alternatívát is javasoltak. A legígéretesebbek közül néhányan közvetett eljárásokat fogadnak el annak meghatározására, hogy milyen emberbarát célokat jelentenek. Ennek egyik kiemelkedő példája Eliezer Yudkowsky koherens extrapolált akarata, ahol a mesterségesen intelligens ágens metacélja az, hogy "elérje azt, amit mi is szeretnénk, ha a mesterséges intelligencia elérné, ha hosszan és alaposan átgondolnánk a dolgot"<sup>119</sup>.

Mindenesetre, függetlenül attól, hogy a motivációs kontrollt az alapelvek átfogó vagy minimalista készletén keresztül kívánjuk gyakorolni, ez a hipotézis jelenleg két nagy akadályba ütközik. Az egyik annak biztosítása, hogy egy összetett, fejleszthető és esetleg még önmódosító mesterséges intelligencia a fejleszthetőségei<sup>120</sup> során végig megőrizze céljait, nehézségekbe ütközik. A másik, és legközvetlenebb akadály az, hogy a szakértők nem tudják, hogyan lehet megbízhatóan absztrakt értékeket programozni egy gépbe.

Ez azonban nem jelenti azt, hogy addig nem szabad cselekedni, amíg a szakembereknek nem sikerül Asimov törvényeit vagy más alapelveket beprogramozniuk egy gép kódjába. Néhány megfontolandó lehetőség, hogy Asimov törvényeit "a robotok tervezőire, gyártóira és üzemeltetőire" irányítsuk<sup>121</sup>, hogy a mesterséges intelligenciát, informatikát vagy adattudományt tanuló diákok tantervébe "egyetemi szinten és minden felsőfokú szakon" beépítsük az etikai reflexiót és a technikai kiegészítést,<sup>122</sup> és hogy a motivációs irányításra tett kísérleteket képesség-ellenőrzéssel egészítsük ki.

A képességek ellenőrzése abban áll, hogy megakadályozzuk, hogy a mesterségesen intelligens ágensek még akkor is képesek legyenek kárt okozni, ha *akarnak*. E hipotézis mögött az az érvelés áll, hogy mivel minden ember megölhető vagy kikapcsolható, minden épített gépet meg kell tudni ölni vagy kikapcsolni. Ez azt jelenti, hogy

---

<sup>118</sup> Gabriel Hallevy, "A mesterséges intelligenciával foglalkozó szervezetek büntetőjogi felelőssége" (2010) <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1564096](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1564096)> hozzáférés április 29 2017

<sup>119</sup> Jürgen Schmidhuber, "Complex Value Systems In Friendly AI", *Artificial General Intelligence* (1<sup>st</sup> edn, Springer-Verlag Berlin 2011).

<sup>120</sup> Benja Fallenstein és Nate Soares, 'Problems Of Self-Reference In Self-Improving Space-Time Embedded Intelligence' (Springer International Publishing 2014) <<https://intelligence.org/files/ProblemsSelfReference.pdf>> hozzáférés május 18. 2017.

<sup>121</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

<sup>122</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért,

"Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems

"(InstituteofElectricalandElectronicsEngineers2016).

< [http://standards.ieee.org/develop/indconn/cc/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/cc/ead_v1.pdf)> Hozzáférés május 182017.

45

visszafordíthatósági mechanizmusok vagy opt-out mechanizmusok beépítése mesterségesen intelligens ágensekbe. A visszafordíthatósági modellek megmondják az ágensnek, hogy mely műveletek visszafordíthatók, és hogyan lehet azokat visszafordítani, lehetővé téve a felhasználók számára, hogy a nem kívánt műveleteket visszacsinálják, és visszatérjenek egy biztonságos szakaszba<sup>123</sup>. Az opt-out mechanizmusok ezzel szemben olyan kill switcheket jelölnek ki, amelyek lehetővé teszik a mesterségesen intelligens ágensek hirtelen kikapcsolását.

A képességek ellenőrzésével az a probléma, hogy a mesterségesen intelligens ágensek telepítése után a tanulásra képes ágensek egyre okosabbak lesznek, és elkerülhetetlenül egy adott ágens rájön, hogy ha hagyja magát kikapcsolni vagy a cselekvéseit visszafordítani, az akadályozza céljai elérését. Amikor a mesterséges intelligencia eljut arra a pontra, amikor az ágensek egyenlő feltételeket biztosítanak a tervezőikkel szemben, az ágens folyamatos tanulása lehetővé teszi számára, hogy túljárjon tervezői eszén, és megmeneküljön a képességeket ellenőrző mechanizmusok elől, nevezetesen más rendszerek feltörésével, hogy saját magáról biztonsági másolatokat telepítsen és futtasson, más mesterségesen intelligens ágensek létrehozásával, amelyek nem rendelkeznek ellenőrző mechanizmusokkal, megelőzően hatástalanítva mindenkit, aki esetleg aktiválni akarja ezeket a mechanizmusokat, vagy meggyőzve a tervezőket, hogy nem szükséges ezt tennie, többek között. Sőt, miután az ilyen ágenseket létrehozták és bevetették, felsőbbrendű stratégiai tervezési képességeik sikeresebbek lennének abban, hogy megtalálják a módját az emberek feletti uralomnak, mint az embereké az ágens feletti *utólagos* uralomnak.<sup>124</sup>

A kapacitásszabályozás kérdésének lehetséges megoldásai közé tartozik olyan ágensek létrehozása, amelyek képesek megtanulni, hogy közömbössé váljanak a tekintetben, hogy az irányító mechanizmusaik aktiválódnak-e vagy sem ("biztonságosan megszakítható ágensek"<sup>125</sup>), vagy mesterségesen intelligens ágensek programozása arra, hogy kompenzálják magukat a megszakítás okozta hasznossági veszteségekért oly módon, hogy végül közömbössé váljanak a tekintetben, hogy megszakítják-e őket vagy sem ("hasznossági kiegyensúlyozás"<sup>126</sup>). Azonban mind a "biztonságosan megszakítható ágensek" megközelítés, mind a "hasznossági egyensúly" megközelítés azzal a korlátozással jár, hogy a mesterségesen intelligens ágens cselekvését vagy a vezérlő mechanizmusok aktiválásának elkerülésére kondicionálják, vagy az ágensnek teljesen közömbössé teszik, hogy aktiválják-e azokat, következésképpen úgy viselkedik, mintha nem is léteznének vagy nem aktiválnák őket. Az első esetben az ágens jutalmazási rendszerét érintik, megváltoztatva annak tervezett működését. Az utóbbiban, mivel az ágens nem motivált arra, hogy törődjék

<sup>123</sup> Az Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Az Európai Parlament Jogi Bizottsága 2016).

<sup>124</sup> Nick Bostrom, *Szuperintelligencia*: (1<sup>st</sup>edn, Oxford University Press 2014).

<sup>125</sup> Laurent Orseau és Stuart Armstrong, "Safely Interruptible Agents", *Conference32<sup>nd</sup> on Uncertainty in Artificial Intelligence* (2016) < <https://intelligence.org/files/Interruptibility.pdf>> hozzáférés május 23. 2017.

<sup>126</sup> Nate Soares és mások, 'Corrigibility', *AAAI Conference on Artificial Intelligence* (AAAI Publications 2015) < <https://intelligence.org/files/Corrigibility.pdf>> Hozzáférés május 192017.

arról, hogy a vezérlőmechanizmusok működőképeseek maradnak-e vagy sem, ártatlanul kikapcsolhatja azokat (pl.: eltávolíthat egy felesleges komponenst), más ágenseket telepíthet (azokban az esetekben, amikor képes erre) beágyazott vezérlőmechanizmusok nélkül, vagy elmulaszthatja a kíméletes leállítást (pl.: nem akadhat fenn egy feladat közepén egy tervezett karbantartás miatti leállítás során).

Az optimális forgatókönyv természetesen az lenne, ha egy mesterségesen intelligens ágens képes lenne felismerni, ha rosszul cselekszik, és leállítani magát, tehát önműködő motivációs és képesség-ellenőrző mechanizmusokat alkalmazna. Az emberiség szerencséjére, bár "jelenleg (...) ésszerűen aggódhatunk, tudásunk jelenlegi állása szerint nem kell aggódnunk"<sup>127</sup> amiatt, hogy a mesterséges intelligens ágensek elszabadulnak. Mindenesetre kiemelkedően fontos, hogy a motivációs és képességkontroll-mechanizmusok fejlődése *egyenrangú legyen* magának a mesterséges intelligencia-technológiának a fejlődésével.

### 5.3. Elszámoltathatóság és átláthatóság

Az előző szakaszból következik, hogy a mesterségesen intelligens ügynökök tervezőinek azt a célt kell követniük, hogy a mesterségesen intelligens ügynökök reverzibilisek, leállíthatóak és módosíthatóak legyenek. Addig a pillanatig azonban, amíg a mesterségesen intelligens ágensek nem képesek felismerni, hogy rosszul cselekszenek, és nem alkalmaznak önszabályozó mechanizmusokat, az ilyen mechanizmusok alkalmazása emberi beavatkozástól függ, ami nem mindig lehetséges és azonnali. Ezért elengedhetetlen, hogy a tervezők "a kutatás során "biztonsági gondolkodásmódot" alakítsanak ki a nem szándékolt és nem várt viselkedések azonosítása és megelőzése érdekében", <sup>128</sup>és egyéb óvintézkedéseket tegyenek az irányítás elvesztésének és az előre nem látható, nem kívánt cselekvések kockázatának minimalizálása érdekében, például "annak biztosítása, hogy a mesterséges intelligencia rendszerek méltóságteljesen hibázzanak az ellenséges bemenetekkel, az eloszláson kívüli hibákkal, a váratlanul gyors képességnövekedéssel és más nagymértékű kontextusváltozásokkal szemben, valamint biztonságos és védett környezetek kialakítása, amelyekben potenciálisan nem biztonságos mesterséges intelligencia rendszereket lehet fejleszteni és tesztelni"<sup>129</sup>.

<sup>127</sup> Laurent Orseau, interjú a BBC-vel, "Google Developing Kill Switch For AI" (2016).

<sup>128</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems

< [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

<sup>129</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems

"(InstituteofElectricalandElectronicsEngineers2016).

< [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

Ebben a tekintetben a mesterségesen intelligens ágensek tesztelhetők "in-the-wild" (szituációk bemutatásával) vagy "fekete doboz" tesztelésnek vethetők alá (szintetikus bemenetek bemutatásával). A "fekete doboz" tesztelés lehetővé teszi a viselkedés megfigyelését és tesztelését olyan forgatókönyvekben, amelyek a természetben nem fordulhatnak elő, és nagyobb számú kísérletet tesz lehetővé, mint a "vadon" tesztelés<sup>130</sup>. A mesterséges intelligencia biztonságának egyik fő kihívása azonban olyan ágensek létrehozása, amelyek biztonságosan képesek a fekete dobozból a külső "nyílt világba" való átmenetre, ahol kiszámíthatatlan dolgok történhetnek. Az előre nem látható helyzetekhez való méltóságteljes alkalmazkodás nehéz, de szükséges a biztonságos működéshez<sup>131</sup>. Amikor olyan technológiát tervezünk, amely hatással lehet az emberek biztonságára vagy jólétére, nem elég csak feltételezni, hogy az működik.

A tervezőknek a kiterjedt tesztelés mellett ugyanis a mesterségesen intelligens ágensek külvilágban való viselkedését, azaz autonóm tanulási folyamatait is szorosan meg kell figyelniük. A tervezőknek figyelemmel kell kísérniük az ügyfelektől érkező visszajelzéseket, és azonnal reagálniuk kell a káros viselkedésről szóló jelentésekre. Ha egy újonnan bevezetett ágensnél elszigetelt hibás működést tapasztalnak, a tervezőnek meg kell vizsgálnia a lehetséges okokat. Ha pedig az ilyen esetek nem magyarázhatók helytelen kezeléssel vagy harmadik fél beavatkozásával, az ügynököt át kell programozni, vagy ki kell vonni a forgalomból. Ha a tervező ezt elmulasztja, akkor gondatlanságért vagy gondossági<sup>132</sup> kötelezettségeinek elmulasztásáért nyilvánosan felelőssé kell tenni.

Sajnos "legalábbis a belátható jövőben a mesterséges intelligencia fejlesztői valószínűleg nem lesznek képesek olyan rendszereket létrehozni, amelyek garantáltan minden lehetséges körülmények között pontosan úgy működnek, ahogyan azt tervezték vagy remélték"<sup>133</sup>. Ezért fontos, hogy a mesterségesen intelligens ágenseket olyan mechanizmusokkal ruházzák fel, amelyek lehetővé teszik az egyértelmű elszámoltathatóság fenntartását. Ezek közé tartoznak az azonosító címkék<sup>134</sup>, amelyek lehetővé teszik az ágens visszavezetését a tervezőjéhez, valamint valamilyen diagnózist lehetővé tevő rendszer, amely segít a mesterséges intelligens ágens hibás működésének okát diagnosztizálni.

---

<sup>130</sup> Anupam Datta, Shavak Sen és Yair Zick, "Algorithmic Transparency Via Quantitative Input Influence: Theory And Experiments With Learning Systems", *IEEE37<sup>th</sup> Symposium on Security and Privacy* (Institute of Electrical and Electronics Engineers 2016).

<sup>131</sup> National Science and Technology Council of the Executive Office of the President of the United States of America, "Preparing For The Future Of Artificial Intelligence" (2016).

<sup>132</sup> Sabine Gless, Emily Silverman és Thomas Weigend, "Ha a robotok kárt okoznak, ki a hibás? Self-Driving Cars And Criminal Liability" (2016) *New 19Criminal Law Review: An International and Interdisciplinary Journal*.

<sup>133</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems

"(InstituteofElectricalandElectronicsEngineers2016 )

< [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

<sup>134</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems: A Vision For Prioritying Human Wellbeing With Artificial Intelligence And Autonomous

Ebben a tekintetben a szakirodalomban jelenleg egy érdekes vita folyik arról, hogy az elszámoltathatóság biztosítása érdekében szükség van-e az átláthatóságra (vagy sem). Az átláthatóság durván úgy definiálható, mint annak lehetősége, hogy megtudjuk, hogyan és miért hozott egy mesterséges intelligens ágens egy adott döntést vagy cselekedett úgy, ahogyan cselekedett<sup>135</sup>. Ez többféle mechanizmuson keresztül biztosítható. "Például (1) a gondozó- vagy háztartási robotok felhasználói számára egy "miért csináltad" gomb, amelynek megnyomásakor a robot megmagyarázza az éppen végrehajtott műveletet, (2) a hitelesítési vagy tanúsítási ügynökségek számára a mesterséges intelligencia/intelligens rendszer alapjául szolgáló algoritmusok és azok hitelesítésének módja, (3) a balesetek kivizsgálói számára az érzékelők és a belső állapotadatok biztonságos tárolása, hasonlóan a repülési adatrögzítőhöz vagy a fekete dobozhoz"<sup>136</sup>. A jelenlegi vita keretében azonban az átláthatóság általában a forráskód közzétételéhez kapcsolódik. Az átláthatóság hívei úgy vélik, hogy ez a közzététel nemcsak a balesetek kivizsgálása és a bizonyítékok összegyűjtése szempontjából fontos, hanem azért is, hogy a nyilvánosság bizalmát növelje a technológia iránt, mivel azt ellenőrzésnek teszi ki. Mindenesetre a legfontosabb szempontok az, hogy a tervezők biztosítsák a döntéshozatali lépések rekonstruálhatóságát, és hogy a mesterséges intelligens ágensek kérésre képesek legyenek bemutatni a cselekvésükhöz vezető folyamatot, azonosítani a bizonytalanság forrásait és megnevezni az általuk támasztott feltételezéseket.

Az átláthatóság megvalósítható közvetett módon, adókedvezményekkel vagy kedvező, a tervezők felelősségét valamilyen módon korlátozó szabálysértési normákkal, vagy közvetlenül, jogi követelményekkel<sup>137</sup>. E tekintetben releváns szabályozás az Európai Unió 2016 áprilisában elfogadott és a tervek szerint 2018-ban hatályba lépő új általános adatvédelmi rendelete (GDPR), amely kimondja, hogy az egyéneknek joguk van "az ilyen értékelést követően hozott döntés magyarázatához és a döntés megtámadásához". Ezt a rendeletet azonban kemény kritikák érik amiatt, hogy "úgy tűnik, nem érti meg, hogy gyakran nem praktikus vagy nem is lehetséges az algoritmusok által hozott valamennyi döntés megmagyarázása (...) [mivel] gyakran az algoritmikus döntések megmagyarázásának kihívása nem az algoritmusok összetettségéből fakad.

---

Systems"(InstituteofElectricalandElectronicsEngineers2016)

< [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

<sup>135</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems

"(InstituteofElectricalandElectronicsEngineers2016).

< [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

<sup>136</sup> Az IEEE globális kezdeményezése a mesterséges intelligencia és az autonóm rendszerek etikai megfontolásaiért, "Etikusan igazodó tervezés: A Vision For Prioritizing Human Wellbeing With Artificial Intelligence And Autonomous Systems

"(InstituteofElectricalandElectronicsEngineers2016 )

< [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> Hozzáférés május 182017.

<sup>137</sup> Matthew U. Scherer, "A mesterséges intelligencia rendszerek szabályozása: (2016) Harvard29 Journal of Law & Technology.

algoritmus, hanem az általa felhasznált adatok értelmezésének nehézsége"<sup>138</sup>. Erre példa, hogy "a SearchInk nevű gépi tanulási eszköz a tollvonások alapján meglehetősen pontosan meg tudja jósolni, hogy egy írott név férfi vagy női (...), [de] a szoftver készítői nem tudják, hogy miért"<sup>139</sup>.

Az ilyen kritikával szemben azzal érvelnek, hogy ha egy mesterségesen intelligens ágens érvelése túl bonyolult ahhoz, hogy az emberek megértsék, akkor kevésbé kell hatékonyabbá tenni. Ez az érvelés várhatóan messze nem konszenzusos, különösen a szóban forgó iparágakhoz kapcsolódó emberek és a mesterséges intelligencia (nem)biztonságával kapcsolatban szkeptikus emberek körében. Az átláthatóság szükségessége ellen számos más érvet is felhoztak: az átláthatatlanságra szükség lehet az adócsalók vagy terroristák rendszerrel való kijátszásának megakadályozására; az átláthatóság nem kívánatos lehet, mert megghiúsítja a fogyasztói adatok, a kereskedelmi tulajdonosi információk vagy az üzleti titkok jogos védelmét; az átláthatóság elérése a felhasználók személyes adatainak felfedésével járhat; a felfedett adatok túlszaporulásának<sup>140</sup> feldolgozása több algoritmikus számítást igényel, ami ördögi kört<sup>141</sup> hoz létre.

Az átláthatóság javasolt alternatívája a magatartás<sup>142</sup> nyomon követése. Ennek egyik módja lenne a "fekete doboz bütykölése". "A "fekete doboz bütykölése" egy fordított mérnöki technika: "egy rendszer specifikációinak szigorú vizsgálaton keresztül történő artikulálásának folyamata, amely a rendszer működésének modelljét tárja fel a terület ismeretére, megfigyelésre és következtetésekre támaszkodva".<sup>143</sup> E technika lényege, hogy azáltal, hogy az ágenst különböző forgatókönyvekkel szembesítik, tesztelni lehet a hitelességét, igazságosságát és megbízhatóságát. A vizsgáztatók az elfogultságra utaló bizonyítékokat keresnének, és

---

<sup>138</sup> Nick Wallace, "Az EU magyarázathoz való joga: A mesterséges intelligencia káros korlátozása < <http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm#>> hozzáférés május 202017.

<sup>139</sup> Nick Wallace, "Az EU magyarázathoz való joga: A mesterséges intelligencia káros korlátozása < <http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm#>> hozzáférés május 202017.

<sup>140</sup> Joshua A. Kroll és mások, "Accountable Algorithms" (2017) University 165of Pennsylvania Law Review < [http://scholarship.law.upenn.edu/penn\\_law\\_review/vol165/iss3/3](http://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3)> Hozzáférés május 202017.

<sup>141</sup> Maayan Perel és Niva Elkin-Koren, "Black Box Tinkering: Beyond Transparency In Algorithmic Enforcement" (Florida Law Review, Forthcoming 2016) < [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2741513](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2741513)> hozzáférés: May 202017.

<sup>142</sup> Nick Wallace, "Az EU magyarázathoz való joga: A mesterséges intelligencia káros korlátozása < <http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm#>> hozzáférés május 202017.

<sup>143</sup> Maayan Perel és Niva Elkin-Koren, "Black Box Tinkering: Beyond Transparency In Algorithmic Enforcement" (Florida Law Review, Forthcoming 2016) < [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2741513](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2741513)> hozzáférés: May 202017.



szükség szerint módosítson. Kétségeink vannak azonban a teszt kiterjesztésével kapcsolatban - a szilárd következtetések levonásához kísérleteket kellene végezni.

Egy mérsékeltebb alternatíva az lenne, ha a mesterségesen intelligens ágensektől megkövetelnénk, hogy kriptográfiai kötelezettségvállalásokat hozzanak létre eljárási szabályszerűségük digitális bizonyítékaként. A kriptográfiai kötelezettségvállalás durván egy ígéretből áll, amely szerint egy adott ágens meghatározott értékekhez van kötve. Az ágenseket úgy lehet megtervezni, hogy a bevetésük előtt közzétegyék azokat a kötelezettségvállalásokat, amelyek leírják, hogy mit fognak tenni, és azokat a kötelezettségvállalásokat, amelyek leírják, hogy utána mit tettek ténylegesen. Ezután zéró-tudás bizonyítással biztosítható, hogy a kötelezettségvállalások valóban megfelelnek az ágens által végrehajtott cselekvéseknek. A zéró tudásbizonyítékok olyan kriptográfiai eszközök, amelyek lehetővé teszik annak bizonyítását, hogy egy adott döntésnek van egy bizonyos tulajdonsága anélkül, hogy felfednénk, hogyan ismerjük ezt a tulajdonságot, vagy hogy mi a döntési politika valójában. Ennek a megoldásnak az az előnye a forráskód átláthatóságával szemben, hogy "a forráskód vagy a bemenetek és kimenetek helyett a kötelezettségvállalások nyilvánosságra hozatalával a rendszerüzemeltetők teljes mértékben meg tudják magyarázni, hogy mit csinálnak a rendszereik anélkül, hogy előre felfednék, hogyan működnek ezek a rendszerek"<sup>144</sup>.

Véleményünk szerint, ha bebizonyosodik, hogy a forráskód közzététele nem szükséges a döntéshozatali folyamat hitelességének, tisztességességének és megbízhatóságának bizonyításához, nem zárkozunk el olyan megoldásoktól, amelyek a forráskód átláthatóságával szemben fenntartják az átláthatatlanságot. Akárhogy is, bár mindkét bemutatott alternatíva lehetővé teszi a mesterségesen intelligens ágensek döntéshozatali folyamatainak tesztelését bizonyos elemek jelenlétének (vagy hiányának) szempontjából, egyik sem kínál *önmagában* megoldást olyan esetekre, amikor a tét *maguknak* a döntési irányelveknek a felülvizsgálata. Ezért tanácsos, hogy kiegészítésként a bíróságok és a szabályozó szervek lehetőséget kapjanak arra, hogy a tervezőktől teljes forráskód-áttekinthetőséget követelhessenek, ha ezt szükségesnek tartják, esetleg egy védelmi rendszer keretében.

Ebben a szakaszban a tervezői felelősség lehetséges forrásaival kezdtük, nevezetesen a biztonsági tesztelés vagy a telepített mesterséges intelligens ágensek felügyeletének hiányosságaival, majd a tervezők felelősségére vonatkozásának biztosításához szükséges elemek ismertetésével folytattuk: az identitásjelölés és az átláthatóság (vagy más, diagnózist lehetővé tevő rendszerek). Az előző fejezettel ellentétben itt nem a mesterségesen intelligens ágensek által okozott károkért való közvetlen felelősséggel, hanem a gondatlanságon vagy a gondossági kötelezettség megsértésén

alapuló felelősséggel foglalkozunk. Ezért, míg

---

<sup>144</sup> Joshua A. Kroll és mások, "Accountable Algorithms" (2017) University 165of Pennsylvania Law Review  
< [http://scholarship.law.upenn.edu/penn\\_law\\_review/vol165/iss3/3](http://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3)> Hozzáférés május 202017.

nem szükséges, hogy tényleges károkat okoztak, fontos meghatározni, hogy ki léphet fel a gondatlan tervezőkkel vagy a gondossági kötelezettségeket megszegő tervezőkkel szemben.

Úgy véljük, hogy a hatékonyabb megoldás a szakosodott szabályozó szervek létrehozása lenne, mivel a legtöbb esetben magas szintű technikai szakértelemre lenne szükség az ilyen jellegű helyzetek megfelelő elemzéséhez. Mivel a mesterséges intelligencia potenciálisan hatással lehet az emberek biztonságára és jólétére, az említett szabályozó testületeknek képesnek kell lenniük arra, hogy megelőző intézkedéseket tegyenek, ahelyett, hogy pusztán reagálva vagy egy panaszt követően cselekednének. A bírságok alkalmazása mellett a szabályozó szervezeteknek képesnek kell lenniük a kármegelőzés biztosítására, amihez célszerű, hogy hatáskörrel rendelkezzenek az olyan mesterséges intelligenciával rendelkező ügynökök értékesítésének felfüggesztésére vagy felhasználási körének korlátozására, amelyek tervezői nem felelnek meg a biztonsági követelményeknek. Egyéb intézkedéseket is érdemes megfontolni, például a kereskedelmi forgalomba hozatalhoz előzetes jóváhagyást kell előírni, vagy olyan minősítési rendszert kell létrehozni, amely a biztonságot és az előreláthatóságot egyaránt értékeli.

#### 5.4. Fejezet összefoglalása

Az ellenőrzés problémája a mesterségesen intelligens ágensek feletti ellenőrzés elvesztésének következményeként felmerülő kockázatokra vonatkozik. Ezek a kockázatok nemcsak a károkkal, hanem a személyes adatok védelmével és a közbiztonsággal is kapcsolatosak. Az irányítás elvesztése bekövetkezhet meghibásodások, a biztonság megsértése, a számítógépek emberhez képest jobb reakcióideje vagy tudatos vagy tudattalan hibás programozás miatt, nevezetesen a törekeny elosztási váltás, a nem biztonságos feltárás, a nem skálázható felügyelet, a negatív mellékhatások vagy a jutalomhekkelés lehetősége tekintetében. Az irányítás elvesztése lehet lokális, amikor az ágens már nem tudja irányítani a működéséért felelős ember, vagy általános, amikor az ágens már nem tudja irányítani semmilyen ember.

Bár irreális lehet, hogy megpróbáljuk megakadályozni az irányítás minden egyes elvesztését, az ilyen események hatását motivációs kontroll és képességszabályozás mechanizmusok segítségével korlátozni lehet. A motivációs ellenőrzés lényege, hogy a mesterséges intelligens ágens kódjába olyan emberbarát célokat építenek be, amelyekkel egyetlen cselekvés vagy funkció sem kerülhet ellentétbe. Ezek az emberbarát célok lehetnek átfogóak, beleértve a társadalmi normákat is, vagy minimalisztikusak, mint például Isaac Asimov A robotika három törvénye vagy Eliezer Yudkowsky koherens extrapolált akarata. A szakértők azonban jelenleg nem tudják, hogyan

lehet megbízhatóan programozni az absztrakt értékeket. A képességek ellenőrzése abban áll, hogy megakadályozzuk

a mesterségesen intelligens ágensek ne legyenek képesek kárt okozni, még akkor sem, ha *akarnak*. Ez azt jelenti, hogy a mesterséges intelligens ágensekbe visszafordíthatósági mechanizmusokat vagy opt-out mechanizmusokat kell beépíteni. A képességek ellenőrzésével kapcsolatban az a probléma, hogy a tanulásra képes mesterséges intelligens ágensek egyre okosabbak lesznek, és felismerik, hogy ha hagyják magukat kikapcsolni vagy a cselekedeteiket visszafordítani, az akadályozza a céljaik elérését, amelyekre programozták őket, mivel felsőbbrendű képességeiknek köszönhetően potenciálisan képesek lesznek kikerülni ezeket a mechanizmusokat.

Mindenesetre addig a pillanatig, amíg a mesterségesen intelligens ágensek nem képesek felismerni, hogy rosszul cselekszenek, és nem alkalmaznak önszabályozó mechanizmusokat, alkalmazásuk az emberi beavatkozástól függ, ami nem mindig lehetséges és azonnali. Ezért elengedhetetlen, hogy a tervezők széles körű tesztelést végezzenek a nem szándékos és nem várt viselkedések azonosítása és megelőzése érdekében, valamint az ügyfelek visszajelzéseinek figyelemmel kísérése, és a káros viselkedésről szóló jelentésekre való azonnali reagálás. Úgy véljük, hogy a megfelelő tesztelés, az ügyfél-visszajelzések nyomon követésének vagy a rendelkezésre álló (ésszerűen) legújabb motivációs és képesség-ellenőrzési mechanizmusok alkalmazásának elmulasztása felelősséget von maga után.

Direct Liability for Damages			Negligence or Violation of a Duty of Care		
Reactive Machines	Machines w/ Limited Memory	Machines w/ Theory of Mind			
Programming	Programming Autonomous action w/ coding deficiency	Programming Autonomous action w/ coding deficiency	Failure to perform adequate safety testing	Failure to monitor customer feedback and to take action in response to it	Failure to apply the (reasonably) latest available motivational and capability control techniques
	Autonomous action w/o coding deficiency but no insurance	Autonomous action w/o coding deficiency but no insurance			

Ábra - Javasolt tervezői felelősség

Az előző fejezettel ellentétben itt nem a mesterséges intelligencia által okozott károkért való közvetlen felelősségről van szó, hanem a gondatlanságon vagy a gondossági kötelezettség megsértésén alapuló felelősségről. Ezért, bár nem szükséges, hogy tényleges károkat okoztak,

fontos meghatározni, hogy ki léphet fel a gondatlan tervezőkkel vagy a kötelességüket megszegő tervezőkkel szemben.

gondoskodás. Úgy véljük, hogy a leghatékonyabb megoldás olyan speciális szabályozó szervek létrehozása lenne, amelyeknek jogukban állna megelőző intézkedéseket hozni és bírságokat alkalmazni, felfüggeszteni vagy előzetesen jóváhagyni a mesterséges intelligens ágensek forgalmazását, korlátozni a felhasználási körüket, létrehozni egy biztonsági és előreláthatósági minősítési rendszert vagy bármilyen más, a biztonság és a jólét védelméhez szükséges intézkedést.

A tervezők felelősségére vonhatóságának biztosítása érdekében fontos, hogy a mesterségesen intelligens ágensek olyan mechanizmusokkal legyenek felszerelve, amelyek lehetővé teszik az elszámoltathatóság egyértelmű fenntartását, nevezetesen az azonossági címkézés és valamilyen diagnosztikát lehetővé tevő rendszer, amely biztosítja, hogy a döntéshozatali lépések rekonstruálhatók legyenek, és hogy a mesterségesen intelligens ágensek képesek legyenek bemutatni a cselekvéseikhez vezető folyamatot, azonosítani a bizonytalanság forrásait és megnevezni az általuk támasztott feltételezéseket. Ebben a tekintetben érdekes vita tárgya a forráskód átláthatóságának szükségessége (vagy sürgőssége). Az átláthatóság hívei úgy vélik, hogy az ilyen közzététel nemcsak a balesetek kivizsgálása és a bizonyítékok összegyűjtése miatt fontos, hanem azért is, hogy a nyilvánosság bizalmát növelje a technológia iránt, mivel azt ellenőrzésnek teszi ki. Az átláthatóság sürgőssége ellen több érvet is felhoznak: az átláthatatlanságra szükség lehet az adócsalók vagy terroristák rendszerrel való kijátszásának megakadályozására; az átláthatóság nem kívánatos lehet, mert megghiúsítja a fogyasztói adatok, a kereskedelmi tulajdonosi információk vagy az üzleti titkok jogos védelmét; az átláthatóság a felhasználók személyes adatainak felfedésével járhat; a nyilvánosságra hozott adatok túlszaporulásának feldolgozása több algoritmikus számítást igényel, ami ördögi kört hoz létre. Az átláthatóság egyik javasolt alternatívája a viselkedés nyomon követése és az elfogultságra utaló jelek keresése, azonban kételyeink vannak a szilárd következtetések levonásához kikísérletezendő tesztforgatókönyvek terjedelmével kapcsolatban. Egy mérsékeltebb alternatíva az lenne, ha a mesterségesen intelligens ügynököktől megkövetelnénk, hogy eljárásuk szabályosságának digitális bizonyítékeként kriptográfiai kötelezettségvállalásokat hozzanak létre. Ennek a megoldásnak az az előnye a forráskód átláthatóságával szemben, hogy a tervezők teljes mértékben elmagyarázhatják, hogy mit csinálnak az ágensek anélkül, hogy ténylegesen felfednék, hogyan működnek ezek az ágensek. Véleményünk szerint, ha bebizonyosodik, hogy a forráskód közzététele nem szükséges a döntéshozatali folyamat hitelességének, tisztességességének és megbízhatóságának bizonyításához, akkor nem zárkozunk el olyan megoldásoktól, amelyek a forráskód átláthatóságával szemben fenntartják az átláthatatlanságot. Akárhogy is, az ilyen

megoldások kiegészítéseként célszerű, hogy a bíróságok és a szabályozó szervek rendelkezzenek azzal a hatáskörrel, hogy a teljes forráskód-áttekinthetőséget követelhessék meg *maguknak a* döntéshozatali politikáknak a felülvizsgálata érdekében, ahelyett, hogy csupán azt ellenőriznék, hogy vannak-e bennük bizonyos elemek (vagy sem).

Amikor olyan technológiát tervezünk, amely hatással lehet az emberek biztonságára vagy jólétére, nem elég csak feltételezni, hogy az működik. Egyetértünk abban, hogy ha a tervezők nem tudnak megalapozottan megbizonyosodni arról, hogy egy ágens biztonságos és ellenőrizhető, és így a bevetése nem jár a negatív következmények elfogadhatatlan kockázatával, akkor az ágens nem lehet és nem is szabad bevetni. Azonban "egy embernek nincs szüksége egy nagyvállalat erőforrásaira és eszközeire ahhoz, hogy számítógépes kódot írjon. Bárki, aki rendelkezik egy viszonylag modern személyi számítógéppel (vagy akár egy okostelefonnal) és internetkapcsolattal, hozzájárulhat a mesterséges intelligenciával kapcsolatos projektekhez. Az egyének tehát egy garázsból, egy kollégiumi szobából vagy egy vasútállomás előteréből is részt vehetnek a mesterséges intelligencia fejlesztésében. Ez a diszkréció lehetősége jelenti a legdöbbenetesebb különbséget a mesterséges intelligencia és a közveszély korábbi forrásai között"<sup>145</sup>. Bár az általunk javasolt megoldások segíthetnek minimalizálni a kontrollvesztés kockázatát és az ilyen események hatását, további munkára van szükség ahhoz, hogy kitaláljuk, hogyan kezeljük a diszkréció problémáját.

## **6. Záró megjegyzések**

### **6.1. Következtetések**

A mesterséges intelligencia bevezetése az iparban és a társadalomban forradalmasítani fogja a jelenlegi társadalmi szerkezetet, és számos olyan technológiai, ipari és szabályozási kihívást fog jelenteni, amelyekre a jogi keretek nem tudnak közvetlen választ adni. Ahogy a mesterségesen intelligens ágensek idővel egyre autonómabbá válnak, annál kevésbé tekinthetők pusztán eszközöknek.

Ennek a valóságnak a figyelembevétele érdekében úgy értelmezzük, hogy a korlátozott memóriával rendelkező gépeket, a gondolkodással rendelkező gépeket és az öntudatos gépeket a tulajdonosaiktól és felhasználóiktól különálló jogi személyeknek kell tekinteni. Ez jogi szempontból kényelmes lenne, és az öntudatos gépek esetében erkölcsi megfontolásokból is támogatható lenne. A reaktív gépek esetében azonban semmilyen pozitív érv nem indokolja a hasonló megfontolást. Az, hogy az ilyen gépeket önálló jogi személyeknek tekintik, nem jogosítja fel őket automatikusan a jogok elismerésére. Nincs nyomós ok arra, hogy jogokat ismerjünk el a reaktív gépek, a korlátozott memóriával rendelkező gépek és a tudatelmélettel rendelkező gépek számára, mivel az esetleges jogok érzékelését vagy

---

<sup>145</sup> Matthew U. Scherer, "A mesterséges intelligencia rendszerek szabályozása: (2016) Harvard Journal of Law & Technology.

akarata van vagy jogosultak rá, pusztán annak az eredménye, hogy erre programozták őket. Következésképpen e gépek tulajdonosai a saját alkotásaikat is birtokolnák. Ha azonban az emberiség valaha is képes lesz öntudatos gépeket építeni, úgy tűnik, hogy ezek a gépek jogosultak lesznek a jogok jogi elismerésére, akár a tulajdonhoz való jogról, akár valamiféle alapvető jogokról beszélünk, hasonlóan az emberekéhez, de a megfelelő kiigazításokkal.

A tervezőknek és a felhasználóknak közvetlenül kell felelniük a mesterséges intelligens ágensek által programozott vagy parancsba adott cselekedetekért (illetve azokért). Az ágensek autonóm cselekedetei tekintetében két forgatókönyv lehetséges: ha a cselekedet kódolási hiba tette lehetővé, akkor a termékhibára vonatkozó szabályok alkalmazandók, és a tervezőt gondatlanság miatt felelősségre lehetne vonni; ha a cselekedet pusztán az ágens fejlődő magatartásából ered, a tervező mentesülhetne azáltal, hogy biztosítást köt az ágens nevében, egyébként pedig közvetlenül felelős lenne. Azzal, hogy a tervező biztosítja, hogy az ügynöknek vannak eszközei az általa okozott károk megtérítésére, a felelősséget áthárítja a harmadik félre. Ezt a biztosítási rendszert kiegészítheti egy kártérítési alap, amelynek célja, hogy biztosítsa a károk megtérítését azokban az esetekben, amikor nincs biztosítás. Ez az alap is a tervezők hozzájárulásaira támaszkodna, de hogy ne terheljék túl a költségeket, dönthetnek úgy, hogy ezeket a költségeket beépítik a szerek eladási árába. Az öntudatos gépeket viszont közvetlenül felelőssé lehet tenni tetteikért, mivel az igazságosság és a büntetés fogalmának szubjektív érzékelésére való képességük lehetővé tenné a felelősségre vonatkozó általános szabályok alkalmazását. Az a tény, hogy az öntudatos gépek lelkiismerettel rendelkeznek, érzőképesek és képesek szubjektív tapasztalatokra, sokkal közelebb helyezi őket az emberekhez, mint a gépekhez, és talán közelebb az emberekhez, mint bármely más valósághoz.

A közszolgáltatások vagy infrastruktúrák mesterségesen intelligens ügynök általi esetleges használata nem az ügynök számára jelent előnyt, hanem annak a tulajdonosnak vagy felhasználónak, aki utasította őt az adott szolgáltatás vagy infrastruktúra használatát feltételező cselekvésre. Valójában, mivel a mesterségesen intelligens ágenseket úgy tervezték, hogy közvetlenül vagy közvetve hozzájáruljanak az emberek jólétéhez, az ember mindig végső soron kedvezményezettje lesz azoknak a közszolgáltatásoknak vagy infrastruktúráknak, amelyeket az ágens a céljai teljesítése során igénybe vesz. Ezért nem tűnik helyesnek azt állítani, hogy *igazságos lenne a mesterségesen intelligens ágensek megadóztatása*, mert közberuházásokból részesülnek. Az adókat azonban a *szükségyszerűség* is indokolhatja. Ez az eset áll fenn az olyan adók esetében, amelyek célja a gazdaságon belüli fogyasztási vagy foglalkoztatási minták módosítása azáltal, hogy bizonyos tranzakciós osztályokat vonzóbbá vagy kevésbé vonzóvá

tesznek.

Valójában a mesterséges intelligencia soha nem látott mértékben képes felforgatni a munkaerőpiacokat, mivel a gépek képesek lesznek helyettesíteni a munkavállalókat számos kognitív és kreatív feladatban, valamint olyan feladatokban, amelyek kézi munkát igényelnek, de amelyeket eddig technológiai korlátok miatt nem lehetett automatizálni (mint például a gépjárművezetés). Az a tény, hogy a mesterséges intelligencia gyakorlatilag a piramis minden szintjén képes lesz helyettesíteni a munkahelyeket, aggodalomra ad okot, hogy a munkahelyek gyorsabban fognak megszűnni, mint ahogyan újakat lehetne létrehozni. Továbbá, még abban az esetben is, ha a mesterséges intelligencia nettó munkahelyteremtést eredményez, nem valószínű, hogy a munkaerő átképzésének jelenlegi módszerei képesek lennének követni ezt a tempót. Az ilyen események közvetlenül a kormányok bevételkiesését fogják eredményezni az adóbeszedések csökkenése miatt, mivel a tőkejövedelmek sokkal alacsonyabb adókulccsal adóznak, mint a munkajövedelmek. Ezen túlmenően az emberi munkaerő automatizált munkaerővel való helyettesítése a társadalombiztosítási kiadások jelentős növekedését is eredményezheti, mivel a társadalombiztosítási rendszereket arra tervezték, hogy munkanélküliségi biztosítást nyújtsanak a munkájukat elvesztő munkavállalóknak. Ezek a megnövekedett kiadások a költségvetési bevételek kiesésével együtt komoly aggodalmakat vetnek fel a jelenlegi társadalombiztosítási rendszerek fenntarthatóságát illetően. Bemutattuk azonban, hogy ez a probléma a mesterségesen intelligens ágensek közvetlen megadóztatása nélkül is kezelhető, mivel léteznek más alternatívák, amelyek nem feltételezik, hogy a mesterségesen intelligens ágenseket tulajdonostulajdonosoknak tekintsék és adót fizessenek, így ez a fikció szükségtelenné válik.

	Separate Legal Status	Rights		Direct Liability	Direct Taxation
		Property Rights	Fundamental Rights		
Reactive Machines	×	×	×	×	×
Machines w/ Limited Memory	✓	×	×	✓*	×
Machines w/ Theory of Mind	✓	×	×	✓*	×
Self-Aware Machines	✓	✓	✓	✓	✓
* If insurance was subscribed on the agent's behalf					

Ábra -6 A mesterségesen intelligens robotok jogi státusza

A mesterségesen intelligens ágenseknek külön jogi státusszal való felruházása és e státus

tartalmának meghatározása, nevezetesen a felelősség, az esetleges jogok és az esetleges adózási kötelezettségek tekintetében, minimális biztonságot nyújt ezen új intelligens ágensek bevezetésének következményeit illetően a következőkben.

társadalom, ami ellentétben áll a nagyfokú kiszámíthatatlansággal, aminek megfelel. Az előre nem láthatóság kockázatai azonban továbbra is kezelendők és mérsékelendők, mivel ezek nem csak az esetleges károkkal, hanem a személyes adatok és a közbiztonság védelmével is kapcsolatosak.

Ahogy az emberiség megpróbál egyre autonómabb gépeket létrehozni, az embereknek nehézséget okozhat annak biztosítása, hogy ezek a gépek ne váljanak *túlságosan* autonómmá. Az irányítás elvesztése bekövetkezhet meghibásodások, biztonsági rések, a számítógépek emberhez képest jobb reakcióideje vagy tudatos vagy tudattalan hibás programozás miatt, nevezetesen a törékeny elosztási váltás, a nem biztonságos felfedezés, a nem skálázható felügyelet, a negatív mellékhatások vagy a jutalomhekkelés lehetősége miatt. Az irányítás elvesztése lehet lokális, amikor az ágenst már nem tudja irányítani a működéséért felelős ember, vagy általános, amikor az ágenst már nem tudja irányítani semmilyen ember. Bár irreális lehet, hogy minden egyes irányításvesztést megkíséreljünk megakadályozni, az ilyen események hatását motivációs ellenőrzési és képesség-ellenőrzési mechanizmusok segítségével korlátozni lehet. Mindenesetre addig a pillanatig, amíg a mesterségesen intelligens ágensek nem képesek önállóan alkalmazni az irányítási mechanizmusokat, alkalmazásuk az emberi beavatkozástól függ, ami nem mindig lehetséges és azonnali. Ezért rendkívül fontos, hogy a tervezők széles körű tesztelést végezzenek a nem szándékolt és nem várt viselkedések azonosítása és megelőzése érdekében, valamint az ügyfelektől érkező visszajelzések nyomon követése, a káros viselkedésről szóló jelentésekre való azonnali reagálás.

Megértjük, hogy a megfelelő tesztelés, az ügyfelek visszajelzéseinek nyomon követése vagy a (ésszerűen) rendelkezésre álló legújabb motivációs és képesség-ellenőrzési mechanizmusok alkalmazásának elmulasztása a tervező gondatlanságon vagy gondossági kötelezettség megsértésén alapuló felelősségét vonja maga után. Mivel nem szükséges, hogy tényleges kár keletkezzen, fontos meghatározni, hogy ki léphet fel a gondatlanul eljáró vagy a gondossági kötelezettségüket megszegő tervezőkkel szemben. A megoldást a szakosodott szabályozó szervek jelenthetnék, amelyek hatáskörrel rendelkeznek arra, hogy megelőző intézkedéseket tegyenek és bírságokat alkalmazzanak, felfüggesztik vagy előzetesen engedélyezik a mesterséges intelligens ágensek értékesítését, korlátozzák felhasználási körüket, biztonsági és előreláthatósági minősítési rendszert hozzanak létre, valamint egyéb szükséges intézkedéseket hozzanak a biztonság és a jólét védelme érdekében.

A tervezői elszámoltathatóság biztosítása érdekében az is fontos, hogy a mesterségesen intelligens ágenseket olyan mechanizmusokkal építsék be, amelyek lehetővé teszik az

elszámoltathatóság egyértelmű vonalának fenntartását, nevezetesen az identitásjelölés és valamilyen diagnózist lehetővé tevő rendszer, amely biztosítja, hogy a döntéshozatali lépések rekonstruálhatók legyenek, és hogy a mesterségesen intelligens ágensek képesek legyenek bemutatni a cselekvéseikhez vezető folyamatot, azonosítani a bizonytalanság forrásait és kimutatni minden olyan tényezőt, amely miatt az ágensek nem tudnak dönteni.

feltételezések, amelyekre támaszkodtak. Ebben a tekintetben érdekes vita tárgya a forráskód átláthatóságának szükségessége (vagy szükségtelensége). Az átláthatóság egyik javasolt alternatívája az, hogy a mesterségesen intelligens ágenseknek kriptográfiai kötelezettségvállalásokat kell létrehozniuk, amelyek digitális bizonyítékul szolgálnak eljárási szabályszerűségeikre. Ennek a megoldásnak az az előnye a forráskód átláthatóságával szemben, hogy a tervezők teljes mértékben elmagyarázhatják, hogy mit csinálnak az ágensek anélkül, hogy ténylegesen felfednék, hogyan működnek ezek az ágensek. Véleményünk szerint, ha bebizonyosodik, hogy a forráskód közzététele nem szükséges a döntéshozatali folyamat hitelességének, tisztességességének és megbízhatóságának bizonyításához, nem zárkozunk el az olyan megoldásoktól, amelyek fenntartják annak átláthatatlanságát. Akárhogy is, célszerű, hogy a bíróságok és a szabályozó szervek rendelkezzenek azzal a hatáskörrel, hogy a forráskód teljes átláthatóságát megkövetelhessek, hogy *maguk a* döntéshozatali irányelvek felülvizsgálhatók legyenek, és ne csupán azt ellenőrizzék, hogy vannak-e bennük bizonyos elemek (vagy sem).

Amikor olyan technológiát tervezünk, amely hatással lehet az emberek biztonságára vagy jólétére, nem elég csak feltételezni, hogy az működik. Úgy véljük, hogy ha a tervezők nem tudnak megalapozottan megbizonyosodni arról, hogy egy ágens biztonságos és ellenőrizhető, és így a bevetése nem jár a negatív következmények elfogadhatatlan kockázatával, akkor az ágenst nem lehet és nem is szabad bevetni. Mindemellett abban is hiszünk, hogy a mesterséges intelligencia képes az emberiséget a jólét felé vezető útra terelni, és végső soron megszabadítani az embereket a munka terhétől, lehetőséget adva arra, hogy olyan feladatokra összpontosítsunk, ahol a kreativitás és a szenvedély nagyobb szerepet játszik. Ahogy Stephen Hawking egyszer megfogalmazta, a jelenlegi és a közeljövő technológiájával "mindenki élvezheti a fényűző szabadidős életet, ha a gépek által termelt jólétet megosztják, vagy a legtöbb ember nyomorúságosan szegényként végezheti, ha a géptulajdonosok sikeresen lobbiznak a vagyon újraelosztása ellen"<sup>146</sup>. Ezért a mi felfogásunk szerint az összes itt előírt megoldás, vagy bármely más, ajánlott vagy elfogadott megoldás minden pillanatban kiigazításra szorul annak érdekében, hogy egyensúlyt teremtsünk fajunk jólétének biztosítása és az innováció szabadsága között. A mesterséges intelligenciától nem félni kell, hanem inkább elfogadni. És azáltal, hogy proaktívan megvitatjuk azokat a kihívásokat, amelyekkel ez a technológia járhat, néhány lépéssel közelebb kerülünk ahhoz, hogy megelőzzük az esetleges hátrányokat, miközben teljes mértékben kiaknázzuk az előnyeit.

## 6.2. Jövőbeni munka

StephenHawking,'ScienceAMASeries:StephenHawkingAMAAnswers'

< [https://www.reddit.com/r/science/comments/3nyn5i/science\\_ama\\_series\\_stephen\\_hawking\\_ama\\_answers](https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers)>

Hozzáférés május 27. 2017.

Ezzel a munkával a törvényhozók, a politikai döntéshozók és más érdekelt felek számára olyan anyagokat kívánunk biztosítani, amelyekkel megkezdődhet a vita a mesterséges intelligencia által támasztott kihívásokról és azok kezeléséről. Ugyanakkor határozottan arra ösztönözzük a többi szerzőt, hogy gondolkodjanak el ezeken a kérdéseken, és mélyebben merüljenek el bennük, akár úgy, hogy tökéletesítik az ötleteinket, akár úgy, hogy új, a közös célunknak - az emberiség jólétének és jólétének biztosításának - jobban megfelelő ötletekkel állnak elő.

Különösen a következő területeken van szükség további munkára: a *mesterséges intelligencia* konkrét és funkcionális definíciója jogi szempontból; hatékony ellenőrzési, azonosítási és individualizációs mechanizmusok kidolgozása a mesterségesen intelligens ágensek számára, nevezetesen az átláthatóságról szóló vita előmozdítása révén; annak garantálása, hogy az ilyen mechanizmusokat minden egyes mesterségesen intelligens ágensben alkalmazzák, még akkor is, ha az ilyen ágenseket anonim és diszkréten tervezik; a mesterséges intelligenciának a munkaerőpiacokra és a társadalombiztosítási rendszerekre gyakorolt hatásának enyhítése, biztosítva, hogy e technológia előnyei a lakosság egészével megoszthatók legyenek; és olyan "mesterséges intelligencia-ökoszisztémák" létrehozása, amelyek mind a tervezők, mind a befektetők számára <sup>147</sup>vonzóak, és amelyek ösztönzik az innovációt és a megfelelő biztonsági tesztelést.

### Hivatkozások listája

[Abbott és Bogenschneider, 2017] Abbott R és Bogenschneider B, 'Should Robots Pay Taxes? Adópolitika az automatizálás korában' [2017] Harvard Law & Policy Review, megjelenés előtt.

< <https://ssrn.com/abstract=2932483> > Hozzáférés május 23. 2017

[Ainge, 2017] Ainge Roy E, 'New Zealand River Granted Same Legal Rights As Human Being' The Guardian (2017) < <https://www.theguardian.com/world/2017/mar/16/new-zealand-river-granted-same-legal-rights-as-human-being> > hozzáférve április 22-én. 2017

[Allen és Widdison, 1996] Allen T és Widdison R, 'Can Computers Make Contracts?' (1996) Harvard Journal of Law & Technology.

---

<sup>147</sup> A megfelelő mesterséges intelligencia-ökoszisztémák kialakításáról további információkat talál Erik Vermeulen,

"Hogyan készülünk fel az automatizálásra?" című cikkében. Or, Why We Need More "Artificial Intelligence Ecosystems" Now!" (Vagy miért van szükségünk több "mesterséges intelligencia ökoszisztémára" most!  
< <https://hackernoon.com/how-to-prepare-for-automation-or-why-we-need-more-artificial-intelligence-ecosystems-now-4a4a767e733b>> hozzáférés április 102017

60

- [Allgrove, 2004] Allgrove B, "Legal Personality For Artificial Intellekts: Pragmatikus megoldás vagy tudományos fikció? (Master, Oxfordi Egyetem 2004)
- [Amodei és mások, 2017] Amodei D és mások, "Konkrét problémák a mesterséges intelligencia biztonságában" (2016)  
< <https://arxiv.org/abs/1606.06565>> Hozzáférés május 18 2017
- [Asimov, 1985] Asimov I, *Robots And Empire* (1<sup>st</sup>edn, Doubleday Books 1985)
- [Asimov, 1942] Asimov I, 'Runaround' [1942] *Astounding Science Fiction*
- [Auken, 2016] Auken I, "Welcome To 2030. I Own Nothing, Have No Privacy, And Life Has Never Been Better', *Annual Meeting of the Global Future Councils* (World Economic Forum 2016) < <https://www.weforum.org/agenda/2016/11/shopping-i-can-t-really-remember-what-that-is>> hozzáférés május 26. 2017
- [Bogenschneider, 2017] Bogenschneider B, 'The Effective Tax Rate Of U.S. Persons By Income Level' (2017) 145 *Tax Notes* (2017).
- [Bombay High Court, 1925] *Pramatha Nath Mullick kontra Pradyumna Kumar Mullick* [1925] Bombay High Court, 27 BOMLR 1064 (Bombay High Court)
- [Bostrom, 2014] Bostrom N, *Superintelligence: Paths, Dangers, Strategies* (1<sup>st</sup>edn, Oxford University Press 2014)
- [Caytas, 2017] Caytas J, "European Perspectives On An Emergent Law Of Robotics" [2017] *Columbia Journal of European Law* < <https://ssrn.com/abstract=2956958>> hozzáférés május 23. 2017
- [Datta, Sen és Zick, 2016] Datta A, Sen S és Zick Y, "Algorithmic Transparency Via Quantitative Input Influence: Theory And Experiments With Learning Systems', *IEEE 37<sup>th</sup> Symposium on Security and Privacy* (Institute of Electrical and Electronics Engineers 2016).
- [Deloitte, 2017] Deloitte, "Agiletown: The Relentless March Of Technology And London's Response' (2017) < <http://www.deloitte.com/content/dam/Deloitte/uk/Documents/uk-futures/london-futures-agiletown.pdf>> hozzáférés: május 23. 2017.
- [Dewey, 1926] Dewey J, "The Historic Background Of Corporate Legal Personality" (1926) *The 35 Yale Law Journal* (A vállalati jogi személyiség történeti háttere) (1926)

[EP, 2017] Európai Parlament, "Robotok és mesterséges intelligencia: MEPs Call for EU-Wide

LiabilityRules"(2017)<<http://www.europarl.europa.eu/news/en/press-room/20170210IPR61808/robots-and-artificial-intelligence-meps-call-for-eu-wide-liability-rules>> hozzáférve május 24-én. 2017

[EP Jogi Bizottsága, 2016] Európai Parlament Jogi Bizottsága, "Jelentés-tervezet a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról" (Európai Parlament Jogi Bizottsága, 2016).

[Executive Office of the POTUS, 2016] Executive Office of the President of the United States of America, "Artificial Intelligence, Automation, And The Economy" (Mesterséges intelligencia, automatizálás és a gazdaság) (2016).

[Fabre, Pallage és Zimmermann, 2014] Fabre A, Pallage S és Zimmermann C, "Universal Basic Income Versus Unemployment Insurance" [2014] CESifo Working Paper Series No. 5106 < <https://ssrn.com/abstract=2540055>> Hozzáférés május 23. 2017

[Fallenstein and Soares, 2014] Fallenstein B and Soares N, 'Problems Of Self-Reference In Self-Improving Space-Time Embedded Intelligence' (Springer International Publishing 2014) < <https://intelligence.org/files/ProblemsSelfReference.pdf>> Hozzáférés május 18 2017

[Feigenbaum, 2003] Feigenbaum E, 'Some Challenges And Grand Challenges For Computational Intelligence' [2003] Journal of the ACM

[Fischer, 1997] Fischer J, "Computers As Agents: cikk

2'(1997)72IndianaLawJournal

<

<http://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=1850&context=ilj>

hozzáférés: 2017. április 23.

[Frey és Osborne, 2017] Frey C és Osborne M, "A foglalkoztatás jövője: How Susceptible Are Jobs To Computerisation? (2017) Technológiai 114 előrejelzés és társadalmi változások.

[Gates, 2017] Bill Gates, Interjú Kevin J. Delaneyvel, "Why Bill Gates Would Tax Robots" (2017)

[Ginzburg, 1987] Ginzburg R, 'Perth Amboy Church Is 302 And Counting' The New York Times (1987)<<http://www.nytimes.com/1987/02/15/nyregion/perth-amboy-church-is-302-and->

counting.html> hozzáférés április 222017

62

[Gless, Silverman és Weigend, 2016] Gless S, Silverman E és Weigend T, 'If Robots Cause Harm, Who Is To Blame? Self-Driving Cars And Criminal Liability' (2016) *New Criminal Law Review: An International and Interdisciplinary Journal*

[Hallevy, 2010] Hallevy G, "The Criminal Liability Of Artificial Intelligence Entities" (A mesterséges intelligenciával foglalkozó szervezetek büntetőjogi felelőssége) (2010).

< [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1564096](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1564096)> hozzáférés április 29. 2017

[Hawking, 2015] Hawking S, 'Science AMA Series: Hawking AMA Answers'

< [https://www.reddit.com/r/science/comments/3nyn5i/science\\_ama\\_series\\_stephen\\_hawking\\_ama\\_answers](https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers)> Hozzáférés május 27. 2017

[Hawking és mások, 2014] Hawking S és mások, "Stephen Hawking: "Transcendence Looks At The Implications Of Artificial Intelligence - But Are We Taking AI Seriously Enough?"

(Stephen Hawking és mások, 2014) *The Independent* (2014) <

<http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html>> hozzáférése május 18. 2017

[High Court of Australia, 1958] *International Harvester Company of Australia Proprietary Limited kontra Carrigan's Hazeldene Pastoral Company* [1958] High Court of Australia, HCA 16; CLR100 644 (High Court of Australia)

[Hintz, 2016] Hintz A, 'Understanding The Four Types of AI, From Reactive Robots To Self-Aware Beings' [2016] *The Conversation UK* < <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>> hozzáférés április 10. 2017

[Horton, 2015] Horton R, 'The Robots Are Coming' (Deloitte LLP 2015)

< <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-finance-robots-are-coming.pdf>> hozzáférés április 6. 2017

[Human Rights Watch, 2013] Human Rights Watch, "Világjelentés 2013: Szaúd-Arábia" (2013)

< <https://www.hrw.org/world-report/2013/country-chapters/saudi-arabia>> hozzáférés április 22. 2017

[The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, 2016] *The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, 'Ethically Aligned Design: A Vision For Prioritizing*



Wellbeing With Artificial Intelligence And Autonomous Systems" (Institute of Electrical and Electronics Engineers 2016) < [http://standards.ieee.org/develop/indconn/ec/ead\\_v1.pdf](http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)> hozzáférés: május 18. 2017

[Jowitt, 2016] Jowitt J, 'Monkey See, Monkey Sue? Gewirth's Principle of Generic Consistency and Rights For Non-Human Agents' (2016) 19 Trinity College Law Review

[Kroll és mások, 2017] Kroll J és mások, "Accountable Algorithms" (2017) University 165of PennsylvaniaLawReview  
< [http://scholarship.law.upenn.edu/penn\\_law\\_review/vol165/iss3/3](http://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3)> Hozzáférés május 202017

[Lee, 2016] Lee D, 'Google Developing Kill Switch For AI' *BBC* (2016)  
< <http://www.bbc.com/news/technology-36472140>> Hozzáférés május 19. 2017

[Luger, 2004] Luger G, *Artificial Intelligence: Structures And Strategies For Complex Problem Solving* (5<sup>th</sup> edn, Addison-Wesley 2004)

[Ma, Nahal és Tran, 2015] Ma B, Nahal S és Tran F, "Robot Revolution - Global Robot & AI Primer"(BankofAmericaMerrilLynch2015).  
< [https://www.bofaml.com/content/dam/boamlimages/documents/PDFs/robotics\\_and\\_ai\\_condensed\\_primer.pdf](https://www.bofaml.com/content/dam/boamlimages/documents/PDFs/robotics_and_ai_condensed_primer.pdf)> Hozzáférés május 23. 2017

[Manyika és mások, 2017] Manyika J és mások, "Harnessing Automation For A Future That Works" (McKinsey Global Institute 2017) < <http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>> hozzáférés május 23. 2017

[McGahey, 2016] McGahey R, 'Universal Basic Income And The Welfare State' (2016)  
< <https://ssrn.com/abstract=2863954>> Hozzáférés május 23. 2017

[Musk, 2016] Elon Musk, Interjú Sam Altmannal, "Y Combinator's How To Build The Future Series" (2016)

[National Science and Technology Council of the Executive Office of the POTUS, 2016] National Science and Technology Council of the Executive Office of the President of the United States of America, 'Preparing For The Future Of Artificial Intelligence' (2016)



[Nilsson, 1998] Nilsson N, Artificial Intelligence: (1<sup>st</sup>edn, Morgan Kaufmann Publishers 1998)

[Orseau and Armstrong, 2016] Orseau L and Armstrong S, 'Safely Interruptible Agents',

*Conference on Uncertainty in Artificial Intelligence 32<sup>nd</sup>* (2016)

< <https://intelligence.org/files/Interruptibility.pdf>> Hozzáférés május 23 2017

[Oxford Dictionaries, 2017] 'Definition Of Robot In English' (Oxford Dictionaries, 2017)

< <https://en.oxforddictionaries.com/definition/robot>> hozzáférés április 6 2017

[Perel and Elkin-Koren, 2016] Perel M and Elkin-Koren N, "Black Box Tinkering: Beyond Transparency In Algorithmic Enforcement" (Florida Law Review, Forthcoming 2016).

< [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2741513](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2741513)> Hozzáférés május 20 2017

[Petit, 2017] Petit N, "Law And Regulation Of Artificial Intelligence And Robots: Conceptual

Framework And Normative Implications' (2017)

< [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2931339](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339)> hozzáférés április 24 2017

[Poole, Mackworth és Goebel, 1998] Poole D, Mackworth A és Goebel R, Computational Intelligence: A Logical Approach (1<sup>st</sup>edn, Oxford University Press 1998).

[Premack és Woodruff, 1978] Premack D és Woodruff G, 'Does The Chimpanzee Have A Theory Of Mind?' (Van-e a csimpánznak elméje?) (1978) 1 Behavioral and Brain Sciences.

[Rawlinson, 2015] Rawlinson K, 'Microsoft's Bill Gates Insists AI Is A Threat' BBC (2015)

< <http://www.bbc.com/news/31047780>> Hozzáférés május 18. 2017

[Rosen, Nilsson és mások, 1967] Rosen C, Nilsson N, Raphael B és mások, "Shakey" [1967]

LIFE < <http://cyberneticzoo.com/cyberneticanimals/1967-shakey-charles-rosen-nils-nilsson-bertram-raphael-et-al-american>> hozzáférés: 2017. április 22.

[Russel és Norvig, 2009] Russell S és Norvig P, Artificial Intelligence: A Modern Approach (3<sup>rd</sup>edn, Prentice Hall 2009)

[Scherer, 2016] Scherer M, "A mesterséges intelligencia rendszerek szabályozása: Risks, Challenges, Competencies, And Strategies" (2016) 29 Harvard Journal of Law & Technology.

- [Schmidhuber, 2011] Schmidhuber J, 'Complex Value Systems In Friendly AI', *Artificial General Intelligence* (1<sup>st</sup>edn, Springer-Verlag Berlin 2011)
- [Schwab és Saams, 2016] Schwab K és Saams R, "Előszó" (Világgazdasági Fórum 2016)  
< <http://reports.weforum.org/future-of-jobs-2016/preface>> Hozzáférés május 232017
- [Science and Technology Committee of the House of Commons of the UK, 2016] Science and Technology Committee of the House of Commons of the United Kingdom, "Robotics And Artificial Intelligence" (Science and Technology Committee of the House of Commons of the United Kingdom 2016).
- [Smith, 2011] Smith H, 'Can Monkey Who Took Grinning Self-Portrait Claim Copyright?' Metro (2011) < <http://metro.co.uk/2011/07/14/can-monkey-who-took-grinning-self-portrait-claim-copyright-77773>> hozzáférés április 302017
- [Smith, 2017] Smith N, 'What's Wrong With Bill Gates' Robot Tax' *Bloomberg* (2017)  
< <https://www.bloomberg.com/view/articles/2017-02-28/what-s-wrong-with-bill-gates-robot-adó>> Hozzáférés május 25. 2017
- [Soares és mások, 2015] Soares N és mások, "Corrigibility", *AAAI Conference on Artificial Intelligence* (AAAI Publications 2015) < <https://intelligence.org/files/Corrigibility.pdf>> hozzáférés május 19. 2017
- [Solum, 1992] Solum L, 'Legal Personhood For Artificial Intelligences' (1992) 70 North Carolina Law Review < <http://papers.ssrn.com/abstract=1108671>> hozzáférés 2017. április 23.
- [Supreme Court of India, 1969] *Yogendra Nath Naskar v. Commissioner Of Income Tax* [1969] Supreme Court of India, 1969 AIR 1089, SCR1969 (3) 742 (India Legfelsőbb Bírósága)
- [Supreme Court of the US, 1967] *Loving v. Virginia* [1967] Supreme Court of the United States, US 3881 (Supreme Court of the United States)
- [Turing, 1950] Turing A, 'Computing Machinery And Intelligence' [1950] *Mind*
- [US Copyright Office, 2015] United States Copyright Office, "Compendium Of U.S. Copyright Office Practices: fejezet (2015300)"

[US District Court for the Northern District of California, 2015] *Naruto, et al. v. David Slater, et al.* [2015] United States District Court for the Northern District of California, No 3:2015cv04324 (United States District Court for the Northern District of California).

[Van Parijs, 2004] Van Parijs P, "Basic Income: Egy egyszerű és erőteljes ötlet a húszéves korszakban.

First Century" (2004) Politika32 és Társadalom

[Varoufakis, 2017] Varoufakis Y, "A Tax On Robots?" < <https://www.project-syndicate.org/commentary/bill-gates-tax-on-robots-by-yanis-varoufakis-2017-02>> hozzáférés május 25. 2017

[Vermeulen, 2017] Vermeulen E, 'How To Prepare For Automation? Or, Why We Need More "Artificial Intelligence Ecosystems" Now!' < <https://hackernoon.com/how-to-prepare-for-automation-or-why-we-need-more-artificial-intelligence-ecosystems-now-4a4a767e733b>> hozzáférve 2017. április 10.

[Wallace, 2017] Wallace N, "EU's Right To Explanation: A Harmful Restriction On Artificial Intelligence" < <http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm#>> hozzáférve május 202017

[Waters and Bradshaw, 2016] Waters R and Bradshaw T, 'Rise Of The Robots Is Sparking An Investment Boom' Financial Times (2016) < <https://www.ft.com/content/5a352264-0e26-11e6-ad80-6765565613c2d6>> hozzáférés 2017. április 5.

[Wells, 2016] Wells G, 'Google's Computers Paint Like Van Gogh, And The Art Sells For Thousands' *TheWallStreetJournal* (2016) < <https://blogs.wsj.com/digits/2016/02/29/googles-computers-paint-like-van-gogh-and-the-art-sells-for-thousands>> hozzáférés: 2017. április 30.

[WEF, 2017] Világgazdasági Fórum, "Hat mód a munkahelyek védelmére a robotautomatizálás ellen" < <https://www.facebook.com/worldeconomicforum/videos/10154432426296479>> Hozzáférés május 25. 2017

[Zhang és mások, 2016] Zhang J és mások, "IDC Futurescape: Worldwide Robotics 2017 Predictions" (International Data Corporation 2016).



[Zimmermann, 2015] Zimmerman E, 'Machine Minds: A jogi személyiség határai' (2015)  
< [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2563965](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965)> hozzáférés április 22 2017

[Zuckerberg,2016]ZuckerbergM,'BuildingJarvis'(Facebook,2016)  
< <https://www.facebook.com/notes/mark-zuckerberg/building-jarvis/10154361492931634>> hozzáférés: 2017. április 10.