

# A mesterséges intelligencia jogi szinten tartása

*Amitai Etzioni\* és Oren Etzioni\*\**

## ABSZTRAKT

*A mesterséges intelligencia programok számos döntést önállóan hoznak, nem átláthatóak, és gyakran változhatnak. Ezért a segítség nélküli emberi ügynökök, például könyvvizsgálók, könyvelők, ellenőrök és rendőrök nem tudják biztosítani, hogy a mesterséges intelligencia által vezérelt eszközök betartsák a törvényt. Ez a cikk azt javasolja, hogy az emberi ügynököknek olyan mesterséges intelligencia-felügyeleti programok segítségével van szükségük, amelyek elemzik és felügyelik a működő mesterséges intelligencia-programokat. A cikk felteszi a kérdést, hogy az operatív AI-programokat úgy kell-e programozni, hogy az emberi felhasználók felülbírálhassák azokat; enélkül egy ilyen lépés aláásná a jogrendet. A cikk rámutat arra is, hogy a mesterséges intelligencia operatív programok nagy felügyeleti kapacitást biztosítanak, és ezért elengedhetetlenek az egyéni jogok védelméhez a kiberkorszakban. A cikk azzal az érveléssel zárul, hogy a mesterséges intelligencia által irányított eszközök, mint például a robotok, a jogrendnél sokkal több veszélyt jelentenek - a robotok a készítőik ellen fordulhatnak, vagy akár el is pusztíthatják az emberiséget.*

---

\*AmitaiEtzioni a George Washington Egyetem egyetemi tanára, korábban a Columbia Egyetemen, a Harvard Business Schoolon és a Berkeley-i Kaliforniai Egyetemen tanított. Főbb könyvei közé tartozik *A magánélet határai*, *A magánélet a kiberkorszakban*, *Az új normális* és *Az aktív társadalom*. A szerzők köszönettel tartoznak Rory Donnelly-nek a korábbi tervezetthez fűzött jelentős észrevételeiért.

\*\* Dr. Oren Etzioni az Allen Institute for Artificial Intelligence vezérigazgatója. 1991 óta a Washingtoni Egyetem informatikai tanszékének professzora, számos kitüntetést kapott, többek között a Robert Engelmere-émlékdíjat (2007), az IJCAI Distinguished Paper Awardot (2005), az AAAI Fellow (2003) és a National Young Investigator Awardot (1993). Több vállalat alapítója vagy társalapítója volt, köztük a Farecast (2008-ban eladta a Microsoftnak) és a Decide (2013-ban eladta az eBaynek), valamint több mint 100 műszaki cikk szerzője, amelyek több mint 27 000 idézést gyűjtöttek. Oren kutatásainak célja a mesterséges intelligencia alapvető problémáinak megoldása, különösen a szövegből való automatikus tudástanulás. Oren a Carnegie Mellon Egyetemen szerzett doktori fokozatot (Ph.D.), 1991, a Harvardon pedig diplomát (B.A.) (2009). 1986.

## TARTALOMJEGYZÉK

I.	A MESTERSÉGES INTELLIGENCE .....	136	
II.	AI GUARDIANS .....	138	
III.	LEZÁRÁS VAGY FELÜLBÍRÁLÁS.....?		141
IV.	HALÁSZNI NEM HORGÁSZHAT	142	
V.	AI DOOMSAYERS .....	144	
VI.	KÖVETKEZTETÉS .....	146	

A politikai döntéshozók és a tudósok egyre több kérdést vetnek fel azzal kapcsolatban, hogy a jogi és erkölcsi rend hogyan tud alkalmazkodni a mesterséges intelligenciával (AI) felszerelt gépek, robotok és eszközök - a továbbiakban "intelligens eszközök" - nagy és egyre növekvő számához. E kérdések közül sok abból ered, hogy az intelligens eszközök, például a vezető nélküli autók, bizonyos fokú autonómiával rendelkeznek; számos döntést önállóan hoznak, jóval túl a programozók által megadott iránymutatásokon.<sup>1</sup> Ráadásul ezek az intelligens eszközök nagyon átláthatatlan módon hoznak döntéseket, és olyan eszközök, amelyek képesek tanulni, olyan irányítási rendszerekkel, amelyek küldetésük teljesítése során változnak.<sup>2</sup>

Egy kaliforniai rendőr például megállította a Google önvezető autóját, mert az autó túl lassan haladva akadályozta a forgalmat.<sup>3</sup> De kit citálhatott volna a rendőr? Az utast? A tulajdonost? A programozót? Az autó számítógépét? Hasonlóképpen, a Google-t azzal a váddal szembesítették, hogy keresőmotorja diszkriminálta a nőket, mivel a jól fizető állások hirdetései gyakrabban jelentek meg a férfiaknak, mint a férfiaknak.

1. Kamala Kelkar, *Hogyan fognak a vezető nélküli autók élet-halál döntéseket hozni?*, PBS (2016. május 28., 11:34), <http://www.pbs.org/newshour/rundown/how-will-driverless-cars-make-life-or-death-decisions/> [https://perma.cc/X93D-4KLT].

2. Jason Tanz, *Soon We Won't Program Computers*, WIRED (2016. május 17., 6:50), <https://www.wired.com/2016/05/the-end-of-code/> [https://perma.cc/UQ7F-7VYX].

3. *Lásd* Don Melvin, *Cop Pulls over Google Self-Driving Car*, CNN (Nov. 13, 2015, 11:03 AM), <http://www.cnn.com/2015/11/13/us/google-self-driving-car-pulled-over/> [https://perma.cc/8K5B-XRK5].

nők, <sup>4</sup>és hogy a keresési eredményekben előnyben részesítette saját üzleteit.<sup>5</sup> Az, hogy az egyszerű halandók képtelenek nyomon követni, hogyan jönnek létre ezek az elfogultságok, jól mutatja, hogy az intelligens gépek milyen kihívások elé állítják a jogi és erkölcsi rendet. Ugyanezek a kérdések vonatkoznak azokra a megállapításokra is, amelyek szerint a letartóztatási nyilvántartást nyújtó weboldalakon megjelenő hirdetések "jelentősen nagyobb valószínűséggel jelentek meg a megkülönböztetetten fekete nevekre vagy egy történelmileg fekete testvériségre vonatkozó kereséseknél".<sup>6</sup> Volt-e szándékosság? Kit vagy mit kell felelősségre vonni az ebből eredő károkért? Hogyan tud a kormány elrettenteni az ugyanazon eszközök által elkövetett ismételt bűncselekményektől? Ez a cikk előzetes választ ad ezekre és számos kapcsolódó kérdésre mind a korlátozott károkozás (pl. egy olyan program, amely miatt egy vezető nélküli autó egy másikkal ütközik)<sup>7</sup>, mind a nagyobb potenciális károkozás (pl. az a félelem, hogy az intelligens eszközök fellázadnak a készítőik ellen, és kárt okoznak az emberiségnek) esetében.<sup>8</sup>

Ez a cikk a mesterséges intelligencia és a jogrend közötti kapcsolatra összpontosít. A mesterséges intelligencia és az erkölcsi rend közötti kapcsolat külön elemzést igényel.<sup>9</sup> Bár mind a jogrend, mind az erkölcsi rend egyazon társadalom értékrendjét tükrözi, ez a cikk külön kezeli őket, mivel különböző módon választják ki és érvényesítik az értékeket. A jogrendben

---

4. Kristen V. Brown, *Google Showed Women Ads for Lower-Paying Jobs*, FUSION (július 18,2015,:33 PM), <http://fusion.net/story/162685/google-ad-algorithms-gender-discrimination/> [<https://perma.cc/67JU-ELN>]; Julia Carpenter, *Google's Algorithm Shows Prestigious Job Ads to Men, but Not to Women.*, WASH. POST (20156., július), <https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/> [<https://perma.cc/Z6D7-3CJT>]; Claire Cain Miller, *When Algorithms Discriminate*, N.Y. TIMES (July20159.), <http://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html> [<https://perma.cc/M9HN-5QF2>].

5. Issie Lapowsky, *Study Offers New Evidence That Google Skews Search Result*, WIRED (június 1129,2015,:23), <http://www.wired.com/2015/06/google-wu-study/> [<https://perma.cc/TH95-FA7T>]. Brian Souter aggódmánának adott hangot a Google PageRank és a keresési eredmények igazságosságával kapcsolatban, miután weboldalai eltűntek a Google első oldalas találati listájáról. Brian Souter, *Disappearing Tycoon Souter Blames Google*, BBC (201112., szeptember), <http://www.bbc.com/news/technology-14884717> [<https://perma.cc/VA6F-ZLMZ>]. A MyTriggers.com ügyében az ohioi székhelyű bevásárló összehasonlító keresőoldal azzal vádolta a Google-t, hogy a keresési eredményekben előnyben részesíti saját szolgáltatásait másokkal szemben (bár a bíró végül úgy döntött, hogy az oldal nem tudta bizonyítani, hogy más hasonló vállalkozásoknak kárt okozott volna). Dan Levine, *Google Wins Antitrust Victory in Ohio Case*, REUTERS (szept. 41,2011,:20 PM), <http://www.reuters.com/article/us-google-antitrust-ruling-idUSTRE7805O420110901> [<https://perma.cc/AWQ8-UK4P>].

6. Miller, *fenti* megjegyzés 4.

7. Alexa Liautaud, *Driverless Car Push Faces Risk of Hacker Hijacking*, BLOOMBERG (2014. szeptember 8., 2:06), <http://www.bloomberg.com/news/articles/2014-09-04/driverless-car-push-faces-risk-of-hacker-hijacking> [<https://perma.cc/J7WE-V8UB>].

8. Nick Bostrom, *When Machines Outsmart Humans*, CNN (2014. szeptember 10., 9:12), <http://www.cnn.com/2014/09/09/opinion/bostrom-machine-superintelligence/index.html> [<https://perma.cc/5YUW-TC5Q>].

9. Lásd általában Amitai Etzioni & Oren Etzioni, *AI Assisted Ethics*, 18 ETHICS & INFO. TECH. 149 (2016), vagy Amitai Etzioni & Oren Etzioni, *Designing AI Systems that Obey Our Laws and Values*, COMM. ACM, <http://cacm.acm.org/magazines/2016/9/206255-designing-ai-systems-that-obey-our-laws-and-values/fulltext> [<https://perma.cc/6YCP-6Q9Z>] (utolsó látogatás: 2016. szeptember 29.), az AI és az erkölcsi rend kapcsolatának vizsgálatához.

birodalomban olyan régóta fennálló intézmények, mint a törvényhozás és a bíróságok választják ki, hogy mely értékeket kell érvényesíteni, de a szociális és erkölcsi területeken nincsenek ilyen tekintélyelvű intézmények. Az etikának nincs legfelsőbb bírósága - és nincs is szükség rá. Ehelyett az erkölcsi birodalom folyamatos erkölcsi párbeszédnek révén választja ki az érvényesítendő értékeket, amelyek idővel gyakran új, közös erkölcsi felfogáshoz vezetnek.<sup>10</sup>

Nem lehet elégszer hangsúlyozni, hogy a "jogrend" nemcsak a bűnüldözést jelenti, hanem a megelőző jogot is, például a vállalkozások rutinszerű ellenőrzését, a sebességmérő kamerák elhelyezését és a vámtisztviselők alkalmazását. E cikkből kiderül, hogy a jogrend fenntartása a kiberkorszakban sokkal inkább új eszközöket igényel, mint új törvényeket.

Az I. rész a mesterséges intelligenciaprogramok egyedi jellemzőinek meghatározásával kezdődik, amelyek számos döntést hozhatnak önállóan, nem átláthatóak és gyakran változnak. A II. rész azt sugallja, hogy a segítség nélküli emberi ügynökök - a könyvvizsgálók és könyvelők, az ellenőrök és a rendőrök - nem tudják biztosítani, hogy az intelligens eszközök betartják a törvényt. Az emberi ügynököknek olyan mesterséges intelligenciaprogramok segítségére van szükségük (e cikk "AI Guardians"-nak nevezi őket), amelyek elemzik és felügyelik az intelligens eszközöket irányító operatív AI-programokat. A III. rész azt a kérdést teszi fel, hogy az operatív mesterséges intelligenciaprogramokat úgy kell-e programozni, hogy az emberi felhasználók felülbírálhassák azokat. A IV. rész rámutat, hogy az intelligens eszközök rendkívül hatékony felügyeletet tudnak gyakorolni, és hogy az ilyen AI Guardians elengedhetetlenek az egyéni jogok védelméhez a kiberkorszakban. A cikket az V. rész zárja, amely azt az érvet tárgyalja, hogy az intelligens eszközök autonómiája sokkal többet veszélyeztethet a jogrendnél, amennyiben az intelligens eszközök a készítőik ellen fordulhatnak, megölhetik őket, vagy akár elpusztíthatják az emberiséget.

#### I. A MŰVES INTELLIGENCIA EGYEDI JELLEMZŐI Beszámoló az

intelligens eszközök által támasztott jogi kihívásokról  
elsőre túlzónak tűnhet. Végül is a törvény sikeresen a legkülönbélebb eszközöket szabályozzák; a rendeletek sokféle dolgot szabályoznak, kezdve attól, hogy egy fűnyíró milyen zajszintet adhat ki legálisan, egészen addig, hogy egy gyár milyen károsanyag-kibocsátást állíthat elő legálisan.

Egyesek szerint könnyű lenne megkövetelni, hogy az önvezető (más néven autonóm vagy vezető nélküli) autók ugyanazokat a törvényeket tartsák be, mint a régmódi autók.<sup>11</sup> Ez azonban olyan lenne, mintha a T-modell autóknak a lovaskocsikra vonatkozó törvényeket kellene betartaniuk. A kényszerítése.

10. Amitai Etzioni, *Morális párbeszéd a nyilvános vitákban*, PUB. PERSP., Mar.-Apr. 27.

11. Danielle Muoio, *Driverless Cars Always Obey the Law*, TECH. INSIDER (Dec. 18, 2015, 11:59 AM), <http://www.techinsider.io/driverless-cars-always-obey-the-law-and-its-a-problem-2015->



az autonóm autóknak a hatályos törvények betartására számos képességüket feláldoznák. Ha például saját sávot kapnának, a vezető nélküli autók sokkal nagyobb sebességgel tudnának biztonságosan közlekedni, mint a régi autók. A történelem valóban azt mutatja, hogy az új technológiák feltalálása - a fegyverektől a DNS-típusmeghatározásig, a gőzgépektől a pilóta nélküli légi járművekig (UAV) - új jogszabályokat igényelt. Bár a jogrend megerősítéséhez szükség lehet néhány új törvényre, sokkal fontosabb, hogy új eszközöket dolgozzunk ki a mesterséges intelligencia jogszerűségének megőrzésére.

Mind az új, mind a régi jogszabályok az intelligens eszközök egyedi tulajdonságai miatt igénylik a mesterséges intelligencia segítségét. Ezek az eszközök *jelentős autonómiával* rendelkeznek abban az értelemben, hogy számos döntést "maguktól" hoznak.<sup>12</sup> Vagyis ezek a műszerek komplex algoritmusokat használnak arra, hogy a valós idejű emberi inputoktól függetlenül reagáljanak a környezeti inputokra; "maguktól is ki tudják találni a dolgokat".<sup>13</sup> Az intelligens gépek eltérhetnek az eredeti programozók által telepített irányelvektől, vagy azokkal ellentétesen cselekedhetnek.<sup>14</sup> Az önvezető autók például eldöntik, hogy mikor változtassanak sebességet, mekkora távolságot tartsanak a többi autótól, és dönthetnek úgy, hogy gyorsabban haladnak, mint ahogy azt a törvény megengedi - amikor megtudják, hogy a többi autó gyakran megszegi a sebességkorlátozást.<sup>15</sup> Egyre elterjedtebbek az automatikus vészfékező rendszerek,<sup>16</sup> amelyek emberi beavatkozás nélkül, észlelt veszély esetén megállítják az autókat.<sup>17</sup> A fogyasztók téves riasztásokra, a többi autóra veszélyes hirtelen megállásra,<sup>18</sup> és arra panaszkodnak, hogy ezek a fékek arra kényszerítik az autókat, hogy egyenes vonalban haladjanak, még akkor is, ha a vezető máshová próbálja terelni őket.

A mesterséges intelligenciával felszerelt autonóm operációs rendszerek az ember számára *rendkívül átláthatatlan fekete* dobozokká válnak. Azaz az emberek nem képesek követni a lépéseket, amelyeket ezek a gépek tesznek, hogy bármilyen következtetésre jussanak. Viktor Schönberger és Kenneth Cukier megjegyzik:

- 
12. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA 16-17 (2013).
  13. Mass. Inst. of Tech., *New Algorithm Lets Autonomous Robots Divvy up Assembly Task on the Fly*, SCI .DAILY, (May 2015 27, ), <http://www.sciencedaily.com/releases/2015/05/150527142100.htm> [https://perma.cc/K6NK-YHM5].
  14. Kelkar, *fenti* megjegyzés 1.
  15. Joe Miller, *Google's Driverless Cars Designed to Exceed Speed Limit*, BBC (Aug. 19, 2014), <http://www.bbc.com/news/technology-28851996> [https://perma.cc/Q3TB-CQNZ].
  16. Chris Knapman, *automatikus fékezés*: TELEGRAPH, (2012 14., aug.), <http://www.telegraph.co.uk/motoring/road-safety/9429746/Auto-braking-a-quantum-leap-for-road-safety.html> [https://perma.cc/XC6L-6GMJ].
  17. Mark Phelan, *Automatic Braking Coming, but Not All Systems Are Equal*, DETROIT FREE PRESS, (Jan. 1, 2016), <http://www.freep.com/story/money/cars/mark-phelan/2016/01/01/automatic-braking-safety-pedestrian-detection-nhtsa-iihs/78029322/> [https://perma.cc/G2NR-X55X].
  18. Eric Limer, *Automatic Brakes Are Stopping for No Good Reason*, POPULAR MECHANICS, (2015 19., június), [www.popularmechanics.com/cars/a16103/automatic-brakes-are-triggering-for-no-good-reason/](http://www.popularmechanics.com/cars/a16103/automatic-brakes-are-triggering-for-no-good-reason/) [https://perma.cc/Q4TD-HJSC].

"A mai számítógépes kódot meg lehet nyitni és ellenőrizni A big-data elemzéssel azonban ez a nyomon követhetőség sokkal nehezebbé válik. Egy algoritmus előrejelzéseinek alapja gyakran túlságosan bonyolult lehet ahhoz, hogy a legtöbb ember megértse."<sup>19</sup> Hozzáteszik, hogy "az algoritmusok és a mögöttük álló adatkészletek fekete dobozokká válnak, amelyek nem nyújtanak számunkra elszámoltathatóságot, nyomon követhetőséget vagy bizalmat".<sup>20</sup>

Az intelligens műszereket irányító AI-programok ráadásul olyan *tanuló rendszerek*, amelyek folyamatosan felülvizsgálják a változó körülményeket és az általuk irányított műszerek teljesítményét, majd ennek megfelelően módosítják a belső irányelveket.<sup>21</sup> Az intelligens műszerek nem hagyják abba az adatgyűjtést, miután elindították őket; ehelyett a további adatgyűjtés lehetővé teszi az intelligens műszerek számára, hogy folyamatosan tanuljanak a tapasztalatokból és javítsák teljesítményüket.<sup>22</sup> Ezek a mesterséges intelligencia programok ezért jelentősen eltérhetnek a programozók által eredetileg megadott irányelvektől.<sup>23</sup> Az intelligens eszközök ugyanis ellentmondhatnak a készítőik és a felhasználóik utasításainak. Ezért az önvezető autókat nem lehet tesztelni és hitelesíteni, mielőtt útnak indulnak, és nem lehet őket szabadon engedni, feltételezve, hogy az irányítórendszerük nem fog megváltozni az új információk hatására, amelyeket az autók menet közben gyűjtenek.

## II. AI GUARDIANS

Az intelligens eszközök egyedi tulajdonságai jogi kihívást jelentenek, ha ezek az eszközök kárt okoznak. Volt-e szándékosság? Ki vagy mi felelős a kárért? És kit kell a jog szerint felelősségre vonni? A következő gondolatmenet szemlélteti a kérdést. Képzeljük el, hogy egy bankot beperelnek, mert az afroamerikaiak által benyújtott hitelkérelmek aránytalanul nagy részét elutasította a fehér amerikaiakéhoz képest. Válaszul a bank tisztviselői rámutatnak, hogy az elmúlt három évben a bank egy mesterséges intelligencia programra támaszkodott a hitelek odaítélésekor vagy elutasításakor. A program kiválasztásakor a bank kikötötte, hogy a programnak tartózkodnia kell attól, hogy a hitelképesség meghatározásához a faji hovatartozást vagy bármilyen helyettesítő változót, például irányítószámot használjon. A felperesek mégis azt bizonyítják, hogy a program hátrányos megkülönböztetést alkalmazott velük szemben azáltal, hogy olyan eseteket mutattak be a bíróságnak, amelyekben a bank olyan afroamerikai kérelmezőknek tagadott meg hitelt, akiknek a hitelminősítése ugyanolyan jó vagy jobb volt, mint a kaukázusi kérelmezőké, akiknek a bank jóváhagyta a hitelét.

---

19. MAYER-SCHÖNBERGER & CUKIER, *Supra* note at 12,178.

20. *Id.* 179.

21. Tanz, *Supra* note 2.

22. *Id.*

23. *Id.*

A diszkrimináció megállapítása nem dönti el az ügyet. A szándékosság és a megkülönböztetésért való felelősség kérdései azért állnak fenn, mert a törvény általában sokkal szigorúbban bünteti a szándékos bűncselekményeket, mint a nem szándékosokat - lásd például az elsőfokú gyilkosság és a gondatlanságból elkövetett emberölés közötti különbséget.<sup>24</sup> Ezért a szándékosság és a felelősség kérdésre választ kell adni annak meghatározásához, hogy kit kell felelősségre vonni az okozott kárért. Visszatérve a gondolati gyakorlatunkhoz: a feltételezett bíróság megállapította, hogy kár keletkezett, de még mindig meg kell határozni, hogy a bank szándékosan okozta-e a kárt azzal, hogy utasította a programozókat, hogy a faji hovatartozást változóként használják - állításai ellenére. Vagy a program "megtanulta" az adatokat vizsgálva, hogy a faj hatékony helyettesítő változóként szolgálhat más tényezők, például az osztály, az oktatás és a földrajzi elhelyezkedés helyettesítésére?

A bíróság felkérhet egy számítógép-programozási szakértőt, hogy szolgáljon tanúként, de ő valószínűleg rámutat majd arra, hogy egyetlen ember sem tud "olvasni" egy mesterséges intelligenciaprogramban annak megállapítására, hogy az abban mutatott elfogultság a programozók tevékenységét vagy a program autonóm tevékenységét tükrözi-e. Mindenekelőtt egyetlen ember sem tudja nyomon követni, hogy a program milyen lépéseken keresztül jutott el autonóm döntéseihez, mivel a program nem vezet feljegyzéseket ezekről a lépésekről.<sup>25</sup>

Ez a cikk azt javasolja, hogy a bíróságnak - és mindazoknak, akiknek meg kell határozniuk a szándékot, a felelősséget és a felelősséget az intelligens eszközök cselekedeteiért - *AI programokra* van szüksége *az AI programok vizsgálatához*. A jognak intelligens eszközökre van szüksége az intelligens eszközök kezeléséhez. Eddig a társadalom az AI-t nagyrészt egy olyan területként kezelte, amely számos programot foglal magában, a repülőgépek robotpilótáitól kezdve a sebészeti robotokig. Innentől kezdve az AI-t két kategóriára kell osztani. Az első kategóriába az operatív AI programok tartoznának - az intelligens eszközöket irányító számítógépes "agyak". A második kategóriát a felügyeleti AI-programok alkotnák, amelyek felülvizsgálják az első kategória döntéshozatalait, és a döntéseket a törvényekkel összhangban tartják. Ezek a felügyeleti programok, amelyeket ez a cikk "AI Guardians"-nak nevez, olyan AI-programokat foglalnának magukban, amelyek az operatív AI-programok lekérdezésére, felfedezésére, felügyeletére, ellenőrzésére és a megfeleléség garantálására szolgálnak.

Az önvezető autók jól illusztrálják az ilyen mesterséges intelligencia-őrök szerepét. Mivel ezek az autók tanulásra és alkalmazkodásra vannak programozva, szükségük van egy különleges AI Guardian programra, egy AI Monitorra, amely végigkíséri őket, hogy biztosítsa, hogy az autonóm autó tanulása és döntéshozatala ne vezessen a törvények megsértéséhez. Az emberi utasokkal ellentétben,

---

24. *Lásd pl.* Griggs v. Duke Power, 401 U.S. 424 (1971) (tovább hangsúlyozva annak megállapításának fontosságát, hogy a kárt szándékosan okozták-e).

25. A lépéseket az érintett számítógépek saját maguk végzik el, amelyek nem vezetnek



listát az általuk végzett igen nagyszámú, összetett számításról. *Lásd* Tanz, *Supra* note 2.

ezek a programok nem fáradnának bele a sebességkorlátozás és a többi autótól való távolság folyamatos ellenőrzésébe, és utas hiányában is el tudnák látni felügyeleti feladataikat.

Az AI közösség még nem tett különbséget az operatív és a felügyeleti AI programok között, nagyrészt azért, mert sok AI tudós osztotta az internet megalakulásával kapcsolatos eredeti eszméket. Az eredeti eszmék hívei azt remélték, hogy az internet egy "lapos" birodalom lesz, egy falu, amelyben minden ember együttműködhet.<sup>26</sup> Nem képzeltek el olyan hierarchiát, amelyben egyesek felügyelnek és szabályoznak másokat.<sup>27</sup> Az évek során azonban az internet faluból dzsungellé változott, ahol hemzsegnak a hackerek, szélhámosok, tolvajok, zsarnokok és ingyenélők. Egyre nagyobb szükség van a rendet erősítő intézményekre. Ezért a világnak jót tenne egy sor mesterséges intelligencia-őrző kifejlesztése, akik megakadályoznák az emberi tervezők által a mesterséges intelligencia programjaiba beépített utasításoktól való eltéréseket.

A kérdező mesterséges intelligencia megállapítaná, hogy az operatív mesterséges intelligenciaprogramok betartják-e az adatvédelmi törvényeket annak meghatározásával, hogy az ilyen programok személyes orvosi információkat használnak-e a fogyasztók megcélzására, foglalkoztatási döntések meghozatalára, hitelnyújtásra vagy -visszavonásra stb. Egy ilyen AI Interrogator nem csupán azt tudná megállapítani, hogy történt-e az orvosi információk jogellenes felhasználása, hanem azt is, hogy a visszaélés a programozók (vagy azok, akik megtartották őket) szándékos cselekedete volt-e, vagy az AI rendszer működésének eredményeként jött létre. Vagyis egy mesterséges intelligencia-kérdező kideríthetné, hogy az információkkal való visszaélés az orvosi adatok illegális megszerzésének vagy az orvosi információk más személyes adatokból való kiszűrésének eredménye volt-e, ez utóbbi jelenleg legális. Ha például egy bankban egy mesterséges intelligenciaprogram lehívná egy rákos beteg hitelét, a program mesterséges intelligencia-kérdezője felmérné, hogy a program egy kórházból vagy orvosi rendelőből illegálisan megszerzett információk alapján járt-e el, vagy a fogyasztási döntések alapján derítette ki a személy állapotát (pl. az illető parókat, nagy mennyiségű szappant és vitaminkiegészítőket vásárolt).

Más mesterséges intelligencia-őrök a felügyeleti szerepek széles skáláját tölthetnék be. A könyvvizsgáló AI programok meghatározhatnák, hogy a pénzügyi tervező szoftverek olyan befektetésekhez vagy biztosítási tervekhez irányítják-e felhasználóikat, amelyekben a szoftver fejlesztőinek pénzügyi érdekeltségei vannak. A mesterséges intelligencia auditorok azt is megállapíthatnák, hogy a keresőmotorok találatai elfogultak-e a keresési eredményeket szolgáltató vállalat vagy annak hirdetői javára.

---

26. SHANE GREENSTEIN, HOW THE INTERNET BECAME COMMERCIAL: INNOVATION, PRIVATIZATION, AND THE BIRTH OF A NEW NETWORK 33-64 (2015).

27. John Perry Barlow, *Nyilatkozat a kibertér függetlenségéről*, <https://projects.eff.org/~barlow/Declaration-Final.html> [<https://perma.cc/FF4F-KSQH>] (utolsó

látogatás 2016. szeptember 19-én).

Eközben a mesterséges intelligenciával foglalkozó felügyelői programok felülvizsgálhatnák a mesterséges intelligencia kiberbiztonsági programjait, például azokat, amelyek korlátozzák az információkhoz való hozzáférést, és jelenthetnék és nyomon követhetnék a jogosulatlan hozzáférési incidensekre adott válaszlépéseket. Ahogy egyre több eszközbe épülnek be operatív AI-programok, úgy növekszik majd az igény a különböző AI-örökre a felügyeleti feladatok ellátására. Vagyis egyre nagyobb szükség lesz az AI-programokra, hogy az AI-programok jogszerűek maradjanak.

A mesterséges intelligencia-öröknek két nagy előnye van az emberi "őrzőkkel" szemben. Először is, az AI Guardians sokkal kisebb valószínűséggel sértik meg az általuk vizsgált személyek szellemi tulajdonjogait és magánéletét, mivel nincsenek saját indítékaik vagy érdekeik. Másodsor, a mesterséges intelligencia-öröknek csak töredékére van szükségük annak az erőforrás- és időigénynek, amely egy embernek ugyanezen felügyeleti feladatok elvégzéséhez szükséges lenne - már ha az emberek egyáltalán képesek lennének ilyen felülvizsgálatokat végezni.

### III. LEZÁRÁS VAGY FELÜLBÍRÁLÁS?

Első pillantásra kézenfekvőnek tűnhet, hogy az intelligens eszközök autonóm cselekedeteinek korlátozására szükség van egy felülbíró eszközre. Egy ilyen eszköz egyfajta vétőjogot biztosítana az embereknek az intelligens eszközök cselekedetei felett. Ha például egy autonóm autó utasai egy égő autóban rekedt embereket látnak az út szélén, megállíthatnák az autójukat, hogy kiszálljanak és segítsenek; az önvezető autó ilyen felülbírálati lehetőség nélkül egyébként csak továbbhajtana. Az embereknek képesnek kell lenniük arra, hogy lelassítsák az autót, hogy élvezzék a tájat, vagy túllépjék a sebességhatárt, hogy a kórházba siessenek. Néhány vezető nélküli autó már rendelkezik ilyen mechanizmussal,<sup>28</sup> és több állam is előírja, hogy az önvezető autók csak vezetői képesítéssel rendelkező utas jelenlétében működhetnek.<sup>29</sup> A New York-i törvények még azt is előírják, hogy valaki mindig tartsa az egyik kezét a kormánykeréken.<sup>30</sup>

Ezzel szemben egyesek azzal érveltek, hogy nem szabadna felülbírálatnak léteznie, mert az emberek visszaélnének vele, amikor ittasan gyorsshajtanának, vagy "közúti dühből" vakmerően vezetnének, és ezzel veszélybe sodornák magukat és másokat. Ahogy egy megfigyelő fogalmazott: "[A]z egyéneket gyakran éppen azért szabályozzuk és vesszük el tőlük az irányítást, mert nem bízhatunk abban, hogy tartózkodni fognak a saját érdekükben való cselekvéstől".<sup>31</sup> Van egy közösségi oldala is a

28. John Markoff, *For Now, Self-Driving Cars Still Need Humans*, N.Y. TIMES (2016. január 17.), <http://www.nytimes.com/2016/01/18/technology/driverless-cars-limits-include-human-nature.html> [https://perma.cc/EVW9-K25M].

29. Lásd Bryant Walker Smith, *Automated Vehicles Are Probably Legal in the United States*, 1 TEX. A&M L. REV. 411, 500-08 (2014).

30. N.Y. VEH. & TRAF. LAW § (1226McKinney 2016).

31. Joshua Gans, *Who Should Control Your Car's Software*, DIGITOPOLY (2015).

december 28.), <https://medium.com/@joshgans/who-should-control-your-cars-software-c5ecd8c1e129#.5qkj39p8k> [<https://perma.cc/F8BA-9YJQ>].

ez az érv: ha az önvezető autók összehangolják mozgásukat, ami nagymértékben növelné a biztonságot, az egyéni felülbírálatok aláásnák az autonóm autók ezen előnyét.<sup>32</sup>

Válaszul meg kell jegyezni, hogy a szabad társadalmakban a törvény ritkán akadályozza meg az embereket abban, hogy a tulajdonukban lévő és általuk használt eszközöket módosítsák - kivételt képeznek azok a helyzetek, amikor a módosítások nagy károkat okoznának (pl. biztonsági öv nélküli vezetés). A társadalom a legtöbb "rossz" eszközhasználatot úgy akadályozza meg, hogy utólag megbünteti azokat, akik visszaélnek a hatalmukkal. Ugyanezt az elvet kell alkalmazni az autonóm eszközökre is.

Továbbá, mivel az AI Guardians felügyeleti programjai számos permutációra képesek, a két álláspont összeegyeztethető. Egy mesterséges intelligenciaprogramot úgy lehetne megtervezni, hogy az út szélére kormányozzon és megálljon, ha egy személy felülbírálja az eredeti programot, majd veszélyes viselkedést tanúsít, de egyébként lehetővé tenné az utasok számára, hogy tetszés szerint felülbírálják a programot. Vagyis a program minden egyes felülbírálatát értékelhetne, és néhányat felülbírálhatna. A mesterséges intelligencia programoknak képesnek kell lenniük a csoportos viselkedés koordinálására akkor is, ha a csoport egyes tagjai robotok, mások pedig emberek. A mai mesterséges intelligenciaprogramokra mindez nem biztos, hogy igaz, de ésszerűnek tűnik, hogy a jövőben képesek lesznek erre.

#### IV. HALÁSZNI NEM HORGÁSZHAT

Bár ritkán fogalmazzák meg így, a civil társadalmak nem törekszenek a teljes jogérvényesítésre. Ez a furcsa preferencia részben abból a valószínűségből fakad, hogy a legtöbb, ha nem is minden polgár valamikor elkövet valamilyen bűncselekményt - sokan pedig elég sokat. Ha a hatóságok mindenkit megbírságnának vagy letartóztatnának, aki elszívott egy jointot, néhány ital után gyorsabban hajtott a megengedett sebességhatárnál, vagy aki nem fizetett ajándékadót a gyermekeire fordított nagy összegű kiadások után, akkor kevés polgár maradna meg, ha egyáltalán lenne ilyen. A civil társadalmak ezért gyakran hajlamosak félrenézni, és a szórványos végrehajtásra hagyatkozni. Ez a nem teljes körű bűnüldözésre való törekvés az egyik oka annak, hogy a polgári szabadságjogok elutasítják a "halászi expedíciókat", vagyis azokat az eseteket, amikor a bűnüldöző szervek munkatársai visszaélnek a célzott házkutatással, hogy megpróbáljanak bizonyítékot találni bármilyen szabálysértésre, nem csak arra, amire az engedély vonatkozik. Az ilyen házkutatásokat a polgári jogok megsértésének tekintik.<sup>33</sup> Valójában ez az oka annak, hogy a parancsok tartalmazzák a "részletességet" - részleteket arról, hogy a hatóságok mit keresnek

---

32. *Id.*

33. *Lásd Katherine M. Shelfer & Hiaohua Hu, Making Better Sense of the Demographic Data Value in the Data Mining Procedure, in FOUNDATIONS AND NOVEL APPROACHES IN DATA MINING 331-61 (Tsau Young Lin, Setsuo Ohsuga, Churn-Jung Liao, and Xiaohua Hu eds., 2015); Brent Skorup, Cops Scan Social Media to Help Assess Your 'Threat Rating', REUTERS (2014).*

december 12.), <http://blogs.reuters.com/great-debate/2014/12/12/police-data-mining-looks-through-social-media-assigns-you-a-threat-level/> [<https://perma.cc/YCB2-YXGJ>].

azt állítják, hogy keresik, ahelyett, hogy csak horgászni mennének.<sup>34</sup> A házkutatási parancsok hatályának korlátozását indokolja továbbá a magánélet megsértésének megakadályozása, az ártatlan emberek megfigyelésének megakadályozása, valamint a hatóságok civil személyek zaklatásának megakadályozása.<sup>35</sup>

Az eszközök okosabbá tételének van egy fontos mellékhatása: még a kisebb bűncselekmények és vétségek felderítését is megkönnyíti, ami veszélyezteti a halászat tilalmát és mindazt, amit az véd. Mind a magán-, mind az állami szektorban olyan programokat fejlesztenek ki, amelyek képesek nyomon követni az egyén internetes tevékenységét,<sup>36</sup> a mobiltelefonokat mikrofonokká és nyomkövető eszközökké, a számítógépeket kamerákká alakítják, apró rádióadókat<sup>37</sup> ültetnek a ruhákba,<sup>38</sup> és még sok minden mást. E programok fejlesztése a felhőalapú tárolás és a "tárgyak internete" megjelenése miatt egyre fokozódik, mivel a hűtőszekrényektől kezdve a termosztátokon át a fitnessszalagokig minden tárgy olyan érzékelőkkel rendelkezik, amelyek személyes adatokat közölhetnek harmadik felekkel és kormányzati hatóságokkal.<sup>39</sup>

A különböző intelligens eszközök által tárolt vagy gyűjtött nagy mennyiségű személyes információ AI programok<sup>40</sup> általi összeállítása, elemzése és extrapolációja ("kibernetizáció") tovább növeli ezen új technológiák hatását, és jelentősen megkönnyíti a bűnüldözés magasabb szintjeit. Például a tipikus CCTV-k - az üzletben, parkolóban vagy lakóházak előterében elhelyezett, saját tulajdonban lévő és felszerelt megfigyelő kamerák - egy adott időpontban, egy adott helyen, egy adott személyről csak néhány adatot gyűjtenek össze, és az információkat rövid ideig megőrzik. Ennek ellenkezője igaz a Microsoft Domain Awareness Systemjére, amelyet először New Yorkban teszteltek a 2012.<sup>41</sup> programban.

34. *Lásd pl.* State v. Retherford, 639 N.E.2d 498 (Ohio Ct. App. 1994).

35. *Lásd id.* 505., 510. o.

36. *Lásd pl.* , Peter Eckersley, *How Online Tracking Companies Know Most of What You Do Online (and What Social Networks Are Doing to Help Them)*, ELECTRONIC FRONTIER FOUND. (2009.21., szeptember), <https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks> [<https://perma.cc/HL4J-RT9M>].

37. *Lásd pl.* , Mike Masnick, *Smartphone Apps Quietly Using Phone Microphones and Cameras to Gather Data*, TECH DIRT (április 9. 2011.,:41), <https://www.techdirt.com/blog/wireless/articles/20110417/21485513927/smartphone-apps-quietly-using-phone-microphones-cameras-to-gather-data.shtml> [<https://perma.cc/85AP-ZSJJ>].

38. Jenny Strasburg & Matthew Yi, *Clothing Will Have Transmitters*, SF GATE (márc. 12., 42003.:00), <http://www.sfgate.com/business/article/Clothing-will-have-transmitters-Benetton-to-2628532.php> [<https://perma.cc/T7DC-KW9N>].

39. *Lásd* ANDREW HILTS, CHRISTOPHER PARSONS & JEFFREY KNOCKEL, EVERY STEP YOU FAKE: A COMPARATIVE ANALYSIS OF FITNESS TRACKER PRIVACY AND SECURITY 2, 76 (2016), [https://openeffect.ca/reports/Every\\_Step\\_You\\_Fake.pdf](https://openeffect.ca/reports/Every_Step_You_Fake.pdf) [<https://perma.cc/MRK9-TWHB>].

40. AMITAI ETZIONI, ADATVÉDELEM A KIBERKORSZAKBAN: (2015).

41. Sajtóközlemény, N.Y.C., Bloomberg polgármester, Kelly rendőrfőnök és a Microsoft bemutatja az új, legmodernebb

jogérvényesítési technológiát

[http://www.nyc.gov/portal/site/nycgov/menuitem.c0935b9a57bb4ef3daf2f1c701c789a0/index.jsp?pageID=mayor\\_press\\_release&catID=1194&doc\\_name=http%3A%2F%2Fwww.nyc.gov%2Fhtml%](http://www.nyc.gov/portal/site/nycgov/menuitem.c0935b9a57bb4ef3daf2f1c701c789a0/index.jsp?pageID=mayor_press_release&catID=1194&doc_name=http%3A%2F%2Fwww.nyc.gov%2Fhtml%2F)



az egész városból a különböző CCTV-kamerák, sebességmérő kamerák, rendszám-tábla-leolvasók és sugárzásérzékelők segítségével.<sup>42</sup> Bár a rendszer egyelőre nem használ arcfelismerést, a jövőben kiterjeszhető ilyen adatokra, valamint mobiltelefonos helymeghatározási információkra is.<sup>43</sup> A Domain Awareness System öt évig vagy még tovább tárolja ezeket az információkat, és a hatóságok tetszés szerint felhasználhatják őket, hogy teljes profilt rajzoljanak egy személy közéletéről.<sup>44</sup> Ez csak egy példa a sok közül, amikor egy személyre vonatkozó "szűrőpróbaszerű" információkat kombinálnak az adott személyre vonatkozó egyéb információkkal, majd ezeket az adatokat mesterséges intelligenciaelemzésnek vetik alá, amely lehetővé teszi a hatóságok számára, hogy a közvetlen megfigyelés által feltárt adatokon jóval túlmutató következtetéseket vonjanak le az adott személyről.<sup>45</sup>

Az emberek ilyen átfogó és folyamatos nyilvános megfigyelésének megakadályozása érdekében a jogalkotóknak új jogszabályokat kellene elfogadniuk, amelyek előírnák, hogy a helyi eszközök, például az útdíjfizető bódék és a térfigyelő kamerák által gyűjtött információkat rövid idő elteltével automatikusan törölni kell, kivéve különleges helyzetekben, például terrortámadás vagy borostyánriasztás esetén. A jogszabályoknak meg kellene tiltaniuk a kibernetizálást is, kivéve az érzéketlen személyes adatokat, mint <sup>46</sup>például az egyén egészségi állapotára vonatkozó információkat, és meg kellene tiltaniuk az érzéketlen információk felhasználását az érzékeny információk felfedésére. Ezen előírások betartatása érdekében a kormányoknak olyan törvényeket kellene elfogadniuk, amelyek előírják, hogy a működő mesterséges intelligencia-megfigyelési programok ellenőrzésére és felügyeletére mesterséges intelligencia-felügyelőket kell alkalmazniuk. Röviden, a törvény az AI-alapú felügyeletet használhatná az AI-alapú felügyelet megfékezésére.

## V. AI DOOMSAYERS

A magasan jegyzett intézmények, mint az MIT, Cambridge és Berkeley mesterséges intelligencia szakértőinek egy kis, de gyakran idézett csoportja arra figyelmeztet, hogy az intelligens eszközök azzal fenyegetnek, hogy olyan okossá válnak, hogy meghaladják az emberi intelligenciát, és ezek az eszközök fellázadhatnak a készítőik ellen, és átvehetik - ha nem pusztítják el - a világot. Rory Cellan-Jones

---

2Fom%2Fhtml%2F2012b%2Fpr291-12.html&cc=unused1978&rc=1194&ndi=1  
[<https://perma.cc/H592-7M6V>].

42. Joe Coscarelli, *The NYPD's Domain Awareness System Is Watching You*, N.Y. MAG. (2012. aug. 9., 8:50), <http://nymag.com/daily/intelligencer/2012/08/nypd-domain-awareness-system-microsoft-is-watching-you.html> [<https://perma.cc/283U-52GY>].

43. Lásd Neal Ungerleider, *NYPD, Microsoft Launch All-Seeing "Domain Awareness System"*, FAST COMPANY (Aug. 128, 2012.:07 PM), <https://www.fastcompany.com/3000272/nypd-microsoft-launch-all-seeing-domain-awareness-system-real-time-cctv-license-plate-monito> [<https://perma.cc/XK4U-WTEZ>].

44. Lásd *id.*

45. Lásd általában *United States v. Jones*, 132 S. Ct. 945, 963 (2012) (Sotomayor, J., egyetértő) ("Az ilyen és más új [megfigyelő] eszközök elérhetősége és használata továbbra is alakítani fogja az átlagember elvárásait a mindennapi mozgásának magánéletével

kapcsolatban.").

46. *Lásd általában* ETZIONI, *Supra* note. 40.

írja: "Az emberek, akiket a lassú biológiai evolúció korlátoz, nem tudnának versenyezni, és kiszorulnának."<sup>47</sup> Hasonlóképpen, egy olyan tudósok, köztük Stephen Hawking által írt véleménycikk is ezt állítja: "Elképzelhető, hogy [az AI] túljár a pénzügyi piacok eszén, túljár az emberi kutatók feltalálásán, túljár az emberi vezetők manipulálásán, és olyan fegyvereket fejleszt ki, amelyeket még csak nem is értünk. Míg a mesterséges intelligencia rövid távú hatása attól függ, hogy ki irányítja, addig a hosszú távú hatása attól függ, hogy egyáltalán irányítható-e".<sup>48</sup> A *Washington Post* szerint

Nick Bostrom kedvenc apokaliptikus hipotézise egy olyan gépet tartalmaz, amelyet úgy programoztak, hogy gémkapcsokat készítsen (bár bármilyen hétköznapi termék megteszi). Ez a gép egyre okosabb és erősebb lesz, de soha nem fejleszt emberi értékeket. Eléri a "szuperintelligenciát". Elkezd mindenféle hétköznapi anyagot gémkapcsokká alakítani. Végül úgy dönt, hogy a Földön mindent - beleértve az emberi fajt is (!!!)- papírkartonná alakít. gémkapcsok.<sup>49</sup>

A mesterséges intelligenciával kapcsolatos vészjóslók olyan sci-fi filmekre hivatkoznak, mint a *Terminátor*, a *Mátrix*, a *2001: Űrodüsszeia* és a *Transzcendencia*,<sup>50</sup> mint arra utaló jelek, hogy elképzelhető egy ilyen mesterséges intelligencia által vezérelt világvége.

Könnyen belátható, hogy ha az intelligens eszközök meghibásodnak, mérhetetlen károkat okozhatnak; egy atomerőmű meghibásodott mesterséges intelligenciaprogramja például nagy pusztítást végezhet. Az azonban korántsem nyilvánvaló, hogy ezek az eszközök saját "akaratot" fejleszthetnének ki, hogy uralkodjanak a készítőik felett, nem is beszélve az emberiségről. Igaz, nem zárható ki, hogy az intelligens eszközök valamikor a távoli jövőben fellázadnak, de mi következik egy ilyen kijelentésből? A kormánynak be kellene tiltania az intelligens műszerek fejlesztését, és lemondani a sok és gyorsan növekvő előnyről, például arról, hogy olyan műszerek készüljenek, amelyek hatékonyabban szolgálnak minket, kevesebbe kerülnek, és több ember számára elérhetőek? Ráadásul az ilyen tilalmakat globális szinten is érvényre lehetne-e juttatni?

Történelmileg az új technológiák, amelyekre most támaszkodunk, vonzották a jövőmondókat, akikről kiderült, hogy hamis próféták.<sup>51</sup> Mindazonáltal, minden

47. Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC (Dec. 20142.), <http://www.bbc.com/news/technology-30290540> [<https://perma.cc/3GLM-TYCV>].

48. Stephen Hawking, Max Tegmark, Frank Wilczek & Stuart Russell, *Transcending Complacency on Superintelligent Machines*, HUFFINGTON POST (201419., június), [http://www.huffingtonpost.com/stephen-hawking/artificial-intelligence\\_b\\_5174265.html](http://www.huffingtonpost.com/stephen-hawking/artificial-intelligence_b_5174265.html) [<https://perma.cc/3WN5-AVNW>].

49. Joel Achenbach, *Az M.I. szorongás*, WASH. POST (201527., dec.), <http://www.washingtonpost.com/sf/national/2015/12/27/aianxiety/>. [<https://perma.cc/Q9JW-G4CU>].

50. Hawking, Tegmark, Wilczek & Russell, *fenti* jegyzet. 48.

51. A leghíresebb eset a ludditáké, akik az ipari forradalom idején szétverték a mechanikus szövőszékeket. Lásd Adrian J. Randall, *A luddizmus filozófiája: The Case of the West of England Woolen Workers, ca. 1790-1809*, 27 TECH. & CULTURE 1, 1-17 (1986) (az ipari forradalom idején a mechanikus szövőszékeket szétzúzó ludditákról).

az operatív AI-programokat folyamatos felügyeletnek kell alávetni annak biztosítása érdekében, hogy magatartásuk ne térjen el az emberi szereplők által meghatározott határokatól. Ezt, hogy megismételjük, csak más mesterséges intelligenciaprogramok által biztosított felügyelettel lehet elérni, amelyeket mi AI Guardiansnak neveztünk el. Ezeknek az AI Guardians-nak okosabbá kell válniuk, ahogyan az operatív AI programok is fejlődnek. Mivel az emberi intelligencia növekedése valószínűleg nem fog lépést tartani a mesterséges intelligencia növekedésével, az embereknek nem sok választása marad, mint hogy a mesterséges intelligencia ellenőrzésére a mesterséges intelligenciát vegyék igénybe - és arra törekedjenek, hogy a mesterséges intelligencia felügyeletét növeljék, ahogy az általuk felügyelt programok intelligenciája növekszik.

## VI. KÖVETKEZTETÉS

A gondolkodó emberek évszázadok óta kérdezik: "Ki fogja őrizni az öröket?".<sup>52</sup> Nincs új válaszunk erre a kérdésre, amelyre soha nem kaptunk jó választ. Egyelőre a legjobb, amiben reménykedhetünk, hogy minden intelligens eszközt felszerelnek majd egy könnyen megtalálható kikapcsolóval, hogy az emberi ügynökök végső ellenőrzést gyakorolhassanak mind az operatív, mind a felügyeleti AI-programok felett.<sup>53</sup>

---

52. A "*Quis custodiet ipsos custodes*" kérdést először Juvenal római író tette fel. *Lásd* JUVENAL, SATIRE VI, at (65Lindsay Watson & Patricia Watson szerk., 2014).

53. Ez az utolsó sor látszólag ellentmondhat egy korábbi kijelentésnek, miszerint a felügyeleti programokat meg kell védeni az emberi beavatkozástól. Ez az előző kijelentés az intelligens eszközök használatára vonatkozik, de nem a használatuk elkerülésére. Tehát amíg valaki autót vezet, addig alá van vetve annak felügyeleti programjának. De mind a működési, mind a felügyeleti program elkerülhető, ha az ember leállítja az autót, vagy ha nem használja tovább. Ugyanez kellene, hogy érvényes legyen minden eszközre.