

Négyesi Imre<sup>✧</sup>

## A mesterséges intelligencia és az etika

DOI 10.17047/HADTUD.2020.30.1.103

A gépi tanuló algoritmusoknak köszönhetően a számítógépek ma már katonai feladatokat is elláthatnak. Ennek megfelelően mindenképpen vizsgálnunk kell a katonai felhasználás során felmerülő etikai kérdéseket és a mesterséges intelligencia egyéb humán vonatkozásait is.

KULCSSZAVAK: mesterséges intelligencia, etika, társadalom, morál

### *The artificial intelligence and the ethics*

*Thanks to machine learning algorithms, computers can now perform military tasks. Accordingly, we must also consider the ethical issues arising in their military use as well as other human aspects of artificial intelligence.*

KEYWORDS: *artificial intelligence, ethics, society, moral*

### **Bevezetés**

A mesterséges intelligencia (a továbbiakban AI)<sup>1</sup> már mindennapjaink elválaszthatatlan részévé vált. Széles körű elterjedése komoly biztonsági és etikai kérdéseket is felvetett. A gépi tanuló algoritmusoknak köszönhetően a számítógépek ma már az autózvezetéstől a robotok irányításán át az automatizált döntéshozásig bármilyen feladatot el tudnak látni, beleértve a katonai feladatokat is. Miután már egyre több kérdésben kezdjük átadni a mesterséges intelligenciáknak az ellenőrzést, sőt, a döntést is, egyre többen sürgetik az olyan biztosítékok beépítését, amelyek garantálják, hogy az AI nem kockáztat emberéleteket. Már több publikációban is foglalkoztam<sup>2</sup> a mesterséges intelligencia katonai alkalmazásának kérdéseivel, de az előzőekben

✧ Nemzeti Közszolgálati Egyetem, Hadtudományi és Honvédtisztképző Kar – National University of Public Service, Faculty of Military Science and Officer Training; e-mail: negyesi.imre@uni-nke.hu; ORCID: 0000-0003-1144-1912

1 AI = Artificial Intelligence.

2 Porkoláb Imre – Négyesi Imre: A mesterséges intelligencia alkalmazási lehetőségeinek kutatása a haderőben. *Honvédségi Szemle*, 147. évfolyam (2019), 5. szám, pp. 3–19.; Négyesi Imre: A mesterséges intelligencia és a hadseregek. *Hadtudomány*, XXIX. évfolyam (2019), 3. szám, pp. 71–79.

leírtaknak is megfelelően, mindenképpen vizsgálnunk kell a katonai felhasználás során felmerülő etikai kérdéseket és a mesterséges intelligencia egyéb humán vonatkozásait is.

### *Az AI felhasználásának általános etikai kérdései*

2019-ben többször beszéltek az AI etikáról, mint korábban bármikor. Több tucat szervezet készített AI etikai irányelveket, mielőtt még a médián keresztül felhívták ezek szükségességére a figyelmüket. Ma már nincs olyan AI-vel kapcsolatos konferencia, ahol a programozás kérdései mellett ne foglalkoznának az etikával kapcsolatos kérdésekkel is. Miután a kifejezetten a hadseregek számára fejlesztett mesterségesintelligencia-alkalmazások mellett a katonai szervezetek használnak egyéb AI-alkalmazásokat is, mindenképpen vizsgálnunk kell a felhasználás etikai kérdéseit a katonai feladatok végrehajtásának vonatkozásában.<sup>3</sup>

Először tekintsük át, milyen általános etikai kérdéseket vetett fel az eddigiek során az AI használata. A főbb kérdések között a szerepelnek:

- Hogyan védjük meg az emberek magánéletét?
- Ha az AI-nak sok adata van szüksége, akkor hogyan biztosíthatók a személyiségi jogok?
- Hogyan dönthet az AI elfogulatlanul?
- Hogyan kerül el az AI a nemek szerinti részlehlajlást az ítéleteiben?
- Hogyan veszi figyelembe az AI a különböző emberek különböző értékrendjét?
- Mit tekinthetünk az AI döntései közül igazságos döntésnek?
- Hogyan bízhatunk a médiában az algoritmikusan létrehozott és elosztott dezinformációval szemben?

Az AI fontosságát jól mutatja, hogy az Amerikai Egyesült Államokban a Fehér Ház 10 alapelvet adott ki az állami ügynökségek számára,<sup>4</sup> amelyeket be kell tartaniuk, amikor új, a magánszektorra vonatkozó AI szabályokat javasolnak. Az alapelveket a Fehér Ház Tudományos és Technológiai Politikai Irodája (OSTP)<sup>5</sup> adta ki az alábbi három fő cél figyelembe vételével:<sup>6</sup>

1. biztosítani a közvélemény részvételét;
2. korlátozni a túlszabályozást;
3. elősegíteni a megbízható, tisztességes, átlátható és biztonságos AI-t.

A gyakorlatban az alapelvek kiadása után a szövetségi ügynökségeknek memorandumot kell benyújtaniuk az OSTP-nek, hogy elmagyarázzák azt, hogy az AI-vel kapcsolatos bármely javasolt szabályozás hogyan felel meg az alapelveknek. Noha az iroda nem rendelkezik felhatalmazással a szabályok megváltoztatására, az eljárás továbbra is biztosíthatja a szükséges nyomást és koordinációt egy bizonyos szabvány

3 Tóth András: Az Amerikai Egyesült Államok Védelmi Minisztériumának felhő koncepciója. *Hírvillám – Signal Badge*, 2018/1. pp. 115–121.

4 <https://raketa.hu/az-usa-kiadta-a-mestersleges-intelligencia-tizparancsolata> (Letöltés ideje: 2020. 02. 15.)

5 OSTP: Office of Science and Technology Policy.

6 <https://www.whitehouse.gov/ostp/> (Letöltés ideje: 2020. 02. 15.)

fenntartása érdekében. R. David Edelman, a MIT technológiai, gazdasági és nemzetbiztonsági projektigazgatója szerint megkísérelnek egy szabályozási előszűrést létrehozni, amelynek segítségével bizonyos szintű minőségellenőrzést vezethetnek be az AI politikába.

A Fehér Ház 10 alapelve a következőképpen hangzik:

1. A nyilvánosság bizalommal forduljon az AI felé. A kormány minden eszközzel elősegíti a megbízható AI-alkalmazások létrehozását.
2. A nyilvánosság részvétele a létrehozásban. A nyilvánosság számára lehetőséget kell biztosítani arra, hogy visszajelzést nyújtson a szabályalkotási folyamat minden szakaszában.
3. Tudományos integritás és információminőség. A politikai döntéseknek minden esetben tudományosan megalapozottnak kell lennie.
4. Kockázatértékelés és -kezelés. Az ügynökségeknek el kell dönteni, milyen szintű kockázatok a még elfogadható kockázatok.
5. Előnyök és költségek. Az ügynökségeknek előzetesen mérlegelni kell valamennyi javasolt szabályozásuk várható társadalmi hatásait.
6. Rugalmasság. Bármelyik megközelítésnek alkalmazkodni kell az AI-alkalmazások gyors változásaihoz és frissítéseikhez.
7. Igazságosság és megkülönböztetésmentesség. Az ügynökségeknek gondoskodniuk kell arról, hogy az AI-rendszerek ne diszkrimináljanak.
8. Közélet és átláthatóság. A nyilvánosság csak akkor bíz az AI-ben, ha tudja, mikor és hogyan kell használni.
9. Biztonság és védelem. Az ügynökségeknek az AI-rendszerek által használt összes adatot biztonságos módon kell kezelniük.
10. Az ügynökségek közötti koordináció. Az ügynökségek koordinációjával biztosítani kell, hogy következtetések és kiszámíthatók legyenek az AI-vel kapcsolatos politikákban.

Az AI etikai kérdések vizsgálatának fontosságát a katonai alkalmazások tekintetében alátámasztja az Amerikai Egyesült Államok Védelmi Minisztériuma<sup>7</sup> által *Summary of the 2018 Department of Defense Artificial Intelligence Strategy Harnessing AI to Advance Our Security and Prosperity*<sup>8</sup> címmel kiadott dokumentum,<sup>9</sup> amelynek egy része külön foglalkozik az AI és az etika összefüggésével. Ebben a dokumentumban is vezető helyen szerepel a katonai etika kérdései között a konzultáció a hazai és a nemzetközi társadalom résztvevőivel (akadémia, magánipar, nemzetközi közösségek stb.). A dokumentum szerint az Amerikai Egyesült Államok Védelmi Minisztériuma arra is törekszik, hogy az AI felhasználásával csökkentsék a nem szándékos és a járulékos károkat a fokozott helyzettudatosság és a fokozott döntéstámogatás révén. A technológia fejlesztésével és annak felhasználásával továbbra is megosztják céljaikat, etikai

7 DoD: Department of Defense.

8 <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf> (Letöltés ideje: 2020. 02. 15.)

9 <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/SUMMARY-OF-DOD-AI-STRATEGY.PDF> (Letöltés ideje: 2020. 02. 15.)

irányelveiket és biztonsági eljárásaikat, hogy ösztönözzék a nemzetgazdaság felelős fejlesztését és az AI felhasználását más nemzetek számára.

### *Az AI fejlesztésének kockázatai*

A tudósok eddig arra összpontosítottak, hogy képesek legyenek kifejleszteni egy mesterséges intelligenciát. Mára már az egyik fő kérdéssé az vált, hogy szükséges-e ezt tenniük. Kimondhatjuk, hogy amennyiben az AI-technológiák hatásai többségében inkább negatívak, mint pozitívak lennének, akkor morális felelőssége lenne az ezen a területen dolgozóknak, hogy más területekre helyezték kutatásukat. Bár azok a párhuzamok elég szélsőségesek, amelyek összekötik a belső égésű motorokat a légszennyezéssel vagy a maghasadás felfedezését Csernobillal, mégis igaz, hogy minden tudósnak és mérnöknek szembe kell néznie a munkájukkal kapcsolatos etikai kérdésekkel. A mesterséges intelligencia fejlesztésével kapcsolatosan a következő, általánosnak nevezhető kérdések merülhetnek fel:

- Az emberek elveszíthetik a munkájukat?
- Az embereknek túl sok (vagy túl kevés) szabadságuk marad?
- Az emberek elveszíthetik a személyiségi jogaik egy részét?
- Az AI-rendszerek alkalmazása felveti a felelősségre vonás lehetőségének kérdéseit.
- A mesterséges intelligencia sikere az emberi faj végét jelentheti?

Ha egyenként, részletesebben is megvizsgáljuk ezeket a kérdéseket, akkor érdekes következtetéseket vonhatunk le.

Az emberek elveszíthetik a munkájukat. A modern ipari gazdaság általában véve függővé vált a számítógépektől, sőt néhány különleges AI-programtól is. Az Egyesült Államokban például a hitelkártya-kérelmek elbírálását, a hívások engedélyezését és a visszaélések felderítését napjainkban már mesterséges intelligencia programok végzik. Látszólag ezek a programok több ezer dolgozót tettek munkanélkülivé, de valójában, ha elvonnánk ezeket a mesterséges intelligencia programokat, a többletköltségek miatt nem is léteznének ilyen állások. A mesterséges intelligenciát használó gépesítés mindeddig több állást hozott létre, mint ahányat megszüntetett, és ezek az új állások érdekesebbek, és jobban fizetnek. A mesterséges intelligencia programok fejlődésével kialakult az a nézet, mely szerint az AI alkalmazása javasolt. Ez megszüntette a munkahelyek elvesztésétől való félelmet, amely akkor volt jelentős, amikor még a mesterséges intelligencia célja az embereket helyettesítő „szakértői rendszerek” létrehozása volt.

Arról a kérdésről, hogy az embereknek túl sok (vagy túl kevés) szabadsága marad, Alvin Toffler a *Future Shock*-ban<sup>10</sup> már 1970-ben azt írta: „A századforduló óta a munkahét 50%-kal csökkent. Nem lóg ki a sorból, ha azt jósoljuk, hogy 2000-re ismét felével csökken.” Arthur C. Clarke sci-fi író *The Sentinel* (Az őrszem) című novellájában 1968-ban úgy vélekedett, hogy 2001-ben az emberek talán már a „teljes unalommal néznek szembe, ahol az élet legnagyobb problémája a több száz tv-csatorna közti választás lesz”.

---

10 Alvin Toffler: *Future Shock*. Bantam Books, 1970. ISBN 0-553-27737-5.

Ezekből a jóslatokból lényegében egy valósult meg a tv-csatornák számának növekedése. Azonban, ha a dolgozó emberek egy 24 órában működő, integrált számítógépesített rendszer részévé válnak, az AI alkalmazása miatt, ahhoz, hogy lépést tartsanak, hosszabb munkaidőre kényszerülnek. Tehát egyre növekvő nyomás nehezedik mindenkire, hogy keményebben dolgozzon, ugyanakkor ez jelentős bérnövekedést is eredményez. A mesterséges intelligencia tehát növeli a technológiai innováció ütemét,<sup>11</sup> de egyben tartalmazza annak ígérését is, hogy az automatizált ágensek kicsit átvállalják a teendőket, és egy kis időhöz juttatnak minket.

Arra, hogy az emberek elveszíthetik a személyiségi jogaik egy részét, szintén sok tudós rámutatott. Ennek okát a beszédfelismerési technológia fejlődésében és a lehallgatás elterjedésében, vagyis a mesterséges intelligenciának a tömeges megfigyelés eszközeként történő alkalmazásában látták.

Az AI-rendszerek alkalmazása felveti a felelősségre vonás kérdéseit is. Ez a kérdés a legélesebben az orvostudomány által használt AI-k kapcsán merülhet fel. Az alapvető kérdés az, hogy ha egy orvos a diagnózis felállításakor egy orvosi szakértői rendszer ítéletére hagyatkozik, ki tehető felelőssé a diagnózis hibájáért? Az orvosi diagnózis hibájából lezajlott perek során a bíróságok úgy tartották, hogy az orvosi szakértői rendszerek ugyanazt a szerepet töltik be, mint az orvosi tankönyvek és kézikönyvek, tehát az orvosoknak felelőssége, hogy megértsék a döntések indoklását, és saját megítélésük alapján kell határozniuk, hogy elfogadják-e a rendszer javaslatát. Azonban ha a szakértői rendszerek megbízhatóbban pontosabbak lesznek, mint a diagnosztikusokat felállító emberek, akkor fordított lesz a helyzet, és az orvosok akár jogilag is felelősek lehetnek, ha nem használják fel a rendszer javaslatait.

A legfontosabb kérdés az, hogy a mesterséges intelligencia sikere az emberi faj végét jelentheti-e? Számtalan sci-fi író fantáziáját foglalkoztatta régen is a kérdés, hogy robotok vagy robot-ember kiborgok rendezhetnek-e ámokfutást? Gyakorlatilag bármely technológiában benne rejlik a lehetőség, hogy kárt okozzon rossz kezekben, de a mesterséges intelligencia és a robotika esetében ezek a rossz kezek akár magához a technológiához is tartozhatnak. A fő veszélyt sokan abban látják, hogy a robotok nélkülözhetetlenek lesznek az ember számára, és öntanuló képességükkel olyan magas szintű intelligenciára tehetnek szert, amelyek már meghaladják a természetes emberi intelligenciát.

### *Robotok, erkölcs, etika*

A robotok viszonyát a társadalmi együttélés szabályaihoz, az erkölchöz és az átalakításhoz először egy DARPA kutatáson keresztül nézzük meg.<sup>12</sup> A mesterséges intelligencia fejlődése révén a virtuális- és robotsegédek egyre jobban képesek lesznek a bonyolult feladatok elvégzésére. Ahhoz, hogy ezeket az „intelligens” gépeket biztonságosnak, megbízhatónak, együttműködőnek, esetleg még etikusnak is lehessen

11 Tóth András: A mesterséges intelligencia és az IoT összekapcsolásának lehetőségei. *Hírvillám – Signal Badge*, 2018/1. pp. 122–130.

12 <https://www.darpa.mil/news-events> (Letöltés ideje: 2020. 02. 15.)

tekinteni, a robotoknak képesnek kell lenniük arra, hogy gyorsan felmérjék az adott helyzetet, és alkalmazzák az emberi társadalmi normákat. Az emberi együttélés szabályainak megtanítása a robotoknak természetesen egy új kihívást jelent.

Ennek a kihívásnak a kezelésére a DARPA által finanszírozott kutatók végrehajtottak egy projektet, amelynek célja egy elméleti és formális keret biztosítása volt a normák és a normatív hálózatok számára. Kísérletileg tanulmányozták, hogyan jelenítik meg és aktiválják a normákat az emberi elmék, és megvizsgálták, hogyan lehet megtanítani ezeket a normákat az algoritmusok segítségével a robotoknak. A csoport létrehozott egy kognitív-számítási modellt az emberi normákról, és kifejlesztett egy gépi tanulási algoritmust, amely lehetővé teszi a gépek számára, hogy normákat tanuljanak ismeretlen helyzetekben az emberi adatokra támaszkodva.

Ghanadan Reza, a DARPA programmenedzsere azt mondta, hogy *„Jelenleg nem tudjuk, hogyan lehet az értelmes normafeldolgozást beilleszteni a hatékony számítási architektúrákba”*, és hozzátette, hogy a társadalmi és etikai normáknak számos olyan tulajdonsága van, amelyek egyedülálló kihívást jelentenek. Az emberek a normákat rendkívül gyorsan aktiválják, vagyis az emberek nagyon gyorsan észlelik a normák megsértését! A normák értelmezése azonban az emberi szubjektumtól is függ, és ez rendkívül megnehezíti az emberi normák gépi tanulásának vizsgálatát. *„Ahhoz, hogy egy robot etikus is legyen, képesnek kell lennie számos olyan norma megtanulására, reprezentálására, aktiválására és alkalmazására, amelyeket egy adott társadalomban az emberek várnak, hogy egymásnak engedelmeskedjenek”* – mondta Ghanadan. Ez a feladat sokkal bonyolultabbá válik, mint az AI-rendszerek szabályainak megtanítása egyszerűbb feladatokra, például képek címkézésére, spam észlelésére vagy az emberek adóbevallásuk során történő vezetésére. Azáltal, hogy keretet biztosít ilyen komplex algoritmusok kifejlesztéséhez és teszteléséhez, az új kutatás hamarabb elhozhatja azt a napot, amikor a gépek az emberi viselkedés legjobb módszereit követik. Összefoglalva Ghanadan szavait, kimondhatjuk, hogy ha ugyanolyan közel kerülünk egymáshoz a jövőbeni robotokkal, ahogyan azt ma elképzeljük, akkor ezeknek az aszisztenseknek ugyanazokat a normákat kell betartaniuk, mint nekünk.

### *Erkölcstelen kompetens robotok kutatása*

A kutatások természetesen nem csak a DARPA-nál folynak. A tudósok érdeklődését egyre inkább felkelti, hogy milyen módon tehető egy robot „érző emberi lényé”. Néhány kérdés, amelyre a kutatások folyamán összpontosítottak:

- Hogyan lehetne beépíteni egy robotba az emberi normarendszerek néhány alapvető tulajdonságát?
- Milyen alapvető erkölcsi szókinccset kell „megtanítani” a robotnak, amely lehetővé tenné a mesterséges ágensek számára az erkölcsi események megismerését és az azokkal kapcsolatos kommunikációt?
- Hogyan használhatja fel a robot a megszerzett információkat saját cselekedeteinek irányításához?
- Hogyan ítélik meg az emberek az erkölcsi döntéseket meghozó robotokat;?
- Milyen különbségeket találunk az ember és a robot között a döntések során betartott erkölcsi normák tekintetében?

Az erkölcsi normák tulajdonságaival kapcsolatosan, valamint az ember és a robot döntéseinek vizsgálatáról számtalan kutatási eredményt közzétettek. Sok tanulmány foglalkozott azzal a kérdéssel is, hogy az emberek miként viszonyulnak az erkölcsi döntéseket meghozó robotokhoz. A tapasztalatokat összegezve arra a következtetésre jutottak, hogy az emberek kétharmadának nincs gondja, hogy a robotokat vagy az AI-t erkölcsi döntéshozókként kezelik. Kimutatták azt is, hogy amikor az emberek erkölcsileg értékelik a mesterséges ágenseket, ugyanazokat a normákat alkalmazzák, mint az emberek esetében. Érdekes eredmény az is, hogy az emberek néha kevésbé hibáztatják az embereket (vagyis önmagukat) olyan esetekben, amikor a döntés erkölcsileg indokolható. Igaz ez akkor is, ha a robot ugyanerre a döntésre jutott a tényleges probléma megoldásánál. Például az emberek kevésbé hibáztatják az embereket abban az erkölcsi döntést igénylő helyzetben, amikor az emberek nem hajlandóak feláldozni egy egyént négy ember megmentése érdekében. Ez egyértelműen abból adódik, hogy az emberek megértik és a saját gondolataikban szimulálják, milyen nehéz lenne egy ilyen áldozatot hozó döntés meghozatala. Ez a fajta „gondolati szimuláció” és megértés nem érhető el egy robot műveleteinek megfontolásakor és kiértékelésekor.

A jövőbeni robot-csapatokban a sikeres interakciók támogatásához szükség van arra, hogy a robot „emberi bizalmat” érezzen csapattársa iránt. Ez a projekt abból indult ki, hogy két vagy több partner közötti együttes fellépés megköveteli a viselkedés összehangolását, amely azonban többek között az erkölcsi normákra is támaszkodik. Azok a partnerek, akik betartják ezeket a normákat megbízhatónak minősülnek. Az emberek, akik a robotokkal közösen feladatot látnak el, ennek megfelelően arra számítanak, hogy „robot-partnereik” ismerik és követik a megfelelő normákat. Ha a robotok megsértik ezeket a normákat, magyarázatot kell adniuk, és igazolniuk kell viselkedésüket az emberek számára érthető módon. (Ez a projekt tehát összekapcsolódik a robot viselkedésének magyarázatával kapcsolatos projektekkel.)

### *Etikai szabályozók*

A következőkben nézzük meg, hogy milyen szabályozók készültek el az eddigi évek során a mesterségesintelligencia-alkalmazások és azok etikus felhasználásának területén. Az látszik, hogy egyre több szinten, egyre több szervezetnél kezdtek vizsgálni a mesterséges intelligencia etikai, erkölcsi, társadalmi szempontjaival kapcsolatban.

Először nézzük meg, mit mond az Európai Bizottság független magas szintű mesterséges intelligencia szakértői csoportja<sup>13</sup> az „Etikai útmutatás a megbízható AI számára” (*Ethics Guidelines for Trustworthy AI*) című dokumentumában.<sup>14</sup> A dokumentum első tervezetét 2018. december 18-án adták ki nyílt konzultációra, és több mint 500 közreműködőtől kaptak visszajelzést. A dokumentum az alapvető jogokon alapuló megközelítés alapján az első fejezetében meghatározta az etikai alapelveket

13 Independent High-level Expert Group on Artificial Intelligence set up by the European Commission (AI HLEG). <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

14 <https://www.eismd.eu/about/> (Letöltés ideje: 2020. 02. 15.)

és azokkal összefüggő értékeket, amelyeket tiszteletben kell tartani az AI-rendszerek fejlesztése, telepítése és használata során. A legfontosabb elvnek tartják az emberi autonómia tiszteletben tartását, a méltányosságot és a megmagyarázhatóságot. Fontosnak tartják, hogy felhívják a figyelmet a gyerekekre, a fogyatékosággal élőkre, a történelmileg hátrányos helyzetűekre és az olyan helyzetben levőkre, akiknél fennállhat a hátrányos megkülönböztetés lehetősége. A dokumentum rámutat az AI használatának társadalmi kockázatára, amelynek figyelembe vételével szükségesnek tartja a kockázat mértékével arányos intézkedések kiadását is.

Az Etikai útmutatás külön pontokban felsorolja a megvalósítás kulcsfontosságúnak ítélt alapelveit is. Ezek az alapelvek az AI-rendszerek teljes életciklusára vonatkoznak:

- teljes körű emberi felügyelet;
- műszaki megbízhatóság és felügyelet;
- a magánélet és a személyes adatok sérthetlensége;
- átláthatóság;
- a környezeti és társadalmi jólét szem előtt tartása;
- elszámoltathatóság.

A legnagyobb technikai cégek, mint a Google, a Facebook, a Microsoft, de még az Amazon is gyakorlatilag rákényszerült, hogy foglalkozzon az AI etikai kérdéseivel. Ennek oka, hogy növekvő ellenállásba ütköztek a képzett munkaerők piacán, amióta a szoftverfejlesztők egyre hangosabban adnak hangot az elégedetlenségüknek a cégek szerintük etikátlan együttműködései kapcsán. A tiltakozások egyik formája az volt, hogy néhány technikai dolgozó nyílt leveleket írt alá, amelyekben ellenzik többek között a Google Project Maven-t<sup>15</sup> és az amerikai katonasággal kötött szerződést. A Microsoft nem ért egyet az USA bevándorlási és vámügyi protokolljával, valamint az Amazon arcfelismerési technológiájának a bűnüldöző szervezetek történő eladásával. A cégek technikai alkalmazottai mellett már olyanok is csatlakoztak a kampányhoz, akik még nem köteleződtek el egyik nagy cég felé sem. Ők a nagy cégek fejedelmének küldött e-mailekben fejtik ki ellenérzésüket. A konkrét munkavállalók esetében a *Tech Workers Coalition* (Technológiai Munkások Szövetsége) nevű munkajogi szervezet biztosít platformot azoknak a fejlesztőknek, akik anyacégeik működését etikátlannak tartják, és nem kívánnak részt venni az emberi jogi aggodalmakat kiváltó projektekből.

Az nyilvánvaló, hogy a szóbeli tiltakozás önmagában nem elég, mert sok szervezet AI etikai irányelve továbbra is homályos és nehéz végrehajtani. Kevés vállalat mutat be kézzelfogható változásokat az AI termékek és szolgáltatások értékelésének és jóváhagyásának módjában. A Google például létrehozott egy AI etikai tanácsot, amely azonban nem rendelkezik tényleges vétőjoggal a megkérdőjelezhető projektek felett, és csak pár taggal jött létre, akiknek a bevonása is vitát váltott ki. De 2019-ben eredményeket is sikerült elérni. Számos város – köztük San Francisco és Oakland, Kalifornia és Somerville (Massachusetts) – betiltotta az arcfelismerés nyilvános használatát, és a javasolt szövetségi törvények hamarosan megtilthatják azt

---

15 Az Egyesült Államok Védelmi Minisztériumának mesterséges intelligencia projektje.



az amerikai állami épületekben is. A technológiai óriások, például a Microsoft, a Google és a Salesforce alkalmazottai szintén felhívták a figyelmet arra, hogy vállalkozások AI-t használnak a migránsok nyomon követésére és a drónok megfigyelésére.

Az érdekelteltek közötti párbeszéd elősegítésére megalakult az egyre több érdekelt felet képviselő fórum, az *AI4People's*,<sup>16</sup> amely összehozhatja az innováció új alkalmazásai társadalmi hatásának kialakításában érdekelt összes szereplőt, beleértve az Európai Bizottságot, az Európai Parlamentet,<sup>17</sup> a civil társadalmi szervezeteket, az ipart és a médiát. A hároméves ütemtervvel 2018 februárjában elindított *AI4People* célja egy közös nyilvános tér létrehozása az alapelvek, politikák és gyakorlatok meghatározására, amelyek alapján a „jó AI társadalom” felépülhet. Ahhoz, hogy ez sikerrel járjon, meg kell állapodniuk abban, hogy miként lehetne legjobban ápolni az emberi méltóságot, elősegíteni az emberi fejlődést és gondoskodni a jobb világról. Ez nem csak a jogi elfogadhatóság kérdése, hanem etikai preferencia kérdése is. A *The AI4People's Ethical Framework for a Good AI Society* dokumentum<sup>18</sup> a bevezetés utáni 2. szakasz ismerteti az AI által kínált alapvető lehetőségeket az emberi méltóság és az emberi fejlődés előmozdítására, valamint azok megfelelő kockázatait. A 3. szakasz áttekintő képet nyújt a szervezetek előnyeiről, ha etikai megközelítést alkalmaznak az AI-k létrehozásakor. A 4. szakasz megfogalmazza az AI öt etikai elvét, a meglévő elemzésekre építve, amelyek alátámasztják az AI etikai elfogadását a társadalomban. Végül, az 5. szakasz 20 ajánlást kínál a „Jó AI Társadalom Európában” kialakítása céljából.

A 4. szakaszban a szerzők szándékának megfelelően nem új alapelvek kerültek megfogalmazásra. Az *AI4People* célja pontosan az volt, hogy párbeszéd alakuljon ki az érdekelteltek között, és ezt segítené a kiadott dokumentum is. Az 5. szakasz ennek megfelelően az AI társadalom ajánlásait ismerteti. Két részből áll: egy preambulumból és 20 akciópontból. A cselekvési pontok négy típusa létezik: felmérés, fejlesztés, ösztönzés és támogatás. Néhány ajánlást közvetlenül a nemzeti vagy európai döntéshozók tehetnek meg, adott esetben az érdekelt felekkel együttműködve. A szerzők határozott szándéka, hogy az AI-alapelvek beépüljenek az AI alapértelmezett gyakorlatába. Kiemelten fontosnak tartják, hogy az AI megmagyarázható legyen, mivel a megmagyarázhatóság kritikus eszköz a nyilvánosság technológiába vetett bizalmának és megértésének felépítésében. Ennek megfelelően a 20 akciópont a következőképpen alakult:<sup>19</sup>

1. Megvizsgálandó a meglévő intézmények, például a nemzeti polgári bíróságok azon képessége, hogy kijavítsák az AI rendszerek által elkövetett hibákat vagy károkat.

16 Egy több érdekelt felet képviselő fórum.

17 <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence> (Letöltés ideje: 2020. 02. 15.)

18 <https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf> (Letöltés ideje: 2020. 02. 15.)

19 <https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf> (Letöltés ideje: 2020. 02. 15.)

2. Mérjék fel, mely feladatok és döntéshozatali funkció nem adható át az AI-rendszereknek.
3. Megvizsgálandó, hogy a jelenlegi szabályozás kellően megalapozott-e etikailag ahhoz, hogy olyan jogi keretet biztosítson, amely képes lépést tartani a technológiai fejleményekkel.
4. Hozzanak létre egy keretet a társadalmilag jelentős döntéseket meghozó AI-rendszerek megmagyarázhatóságának fokozására.
5. Dolgozzanak ki megfelelő jogi eljárásokat az igazságszolgáltatási rendszer informatikai infrastruktúrájának fejlesztésére, az algoritmikus döntések bírósági ellenőrzésének lehetővé tétele érdekében.
6. Fejlesztetni kell az AI-rendszerek ellenőrzési mechanizmusait a nem kívánt következmények elkerülésére (pl. tisztességtelen elfogultság azonosítása).
7. Fejlesszenek ki jogorvoslati eljárást vagy mechanizmust az AI által okozott tévedések vagy sérelmek orvoslására, kompenzálására.
8. Fejlesztetni kell az AI-termékek és -szolgáltatások megbízhatósági mutatóit.
9. Új, a közjólét védelméért felelős EU felügyeleti ügynökség fejlesztése az AI-termékek, -szoftverek, -rendszerek vagy -szolgáltatások tudományos értékelése és felügyelete útján.
10. Az AI európai megfigyelőközpontjának kialakítása.
11. Jogi eszközök és szerződéses sablonok kidolgozása az ember és gép közötti együttműködés megteremtésére.
12. Ösztönözni kell pénzügyileg is az EU szintjén az AI-technológiák fejlesztését és használatát. Cél a társadalmi szempontból előnyösebb AI-k létrehozása.
13. A fenntartható, megnövekedett és koherens európai kutatási erőfeszítések ösztönzése pénzügyi szempontból is.
14. A multidiszciplináris és ágazatközi együttműködés és eszmecsere ösztönzése a technológia, a társadalmi kérdések, a jogi tanulmányok és az etika kérdéseiben.
15. A jogszabályok rendszeres felülvizsgálata annak tesztelésére, hogy milyen mértékben ösztönzik a társadalmilag pozitív innovációt.
16. Az AI-rendszerek empirikus tesztelése és fejlesztése.
17. Segíteni kell az AI és annak alkalmazásai közvélemény általi felfogására és megértésére, valamint a nyilvános konzultációs mechanizmusok végrehajtására irányuló kutatások pénzügyi támogatását.
18. Az AI-vel kapcsolatos politikák és szabályok kidolgozása.
19. Támogassa az AI-vel kapcsolatos szakmákra vonatkozó önszabályozó magatartási kódexek kidolgozását, etikai kötelezettségekkel.
20. Támogatni kell a vállalati igazgatóságok azon képességét, hogy vállalják a felelősséget a vállalatok AI-technológiáinak etikai következményeiért.

## Összefoglalás, következtetések

A leírtakból látszik, hogy Európa és a világ egésze olyan technológia megjelenésével szembesül, amely az emberi élet szempontjából sok előnnyel járhat, de úgy tűnik, hogy komoly veszélyeket is jelenthet. Az idézett dokumentumok és a leírt ajánlások mindegyike arra törekszik, hogy az irányítást az etikai és társadalmi szempontból előnyösebb eredmények irányába mozdítsa el az AI-technológiák fejlesztése, tervezése és bevezetése során. Az AI alapvető társadalmi lehetőségeinek és kockázatainak, valamint az elfogadott etikai alapelveknek a felismerése alapján az együttműködés szellemében megszülethetnek a konkrét és konstruktív válaszok az AI által felvetett legsürgetőbb társadalmi kihívásokra, amely válaszokat figyelembe kell venni a katonai célú AI-k létrehozása során is.

## FELHASZNÁLT IRODALOM

- <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>  
(Letöltés ideje: 2020. 02. 15.)
- <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>
- <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>  
(Letöltés ideje: 2020. 02. 15.)
- <https://raketa.hu/az-usa-kiadta-a-mesterseges-intelligencia-tizparancsolatot> (Letöltés ideje: 2020. 02. 15.)
- <https://www.darpa.mil/news-events> (Letöltés ideje: 2020. 02. 15.)
- <https://www.eismd.eu/about/> (Letöltés ideje: 2020. 02. 15.)
- <https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf> (Letöltés ideje: 2020. 02. 15.)
- <https://www.whitehouse.gov/ostp/> (Letöltés ideje: 2020. 02. 15.)
- Négyesi Imre: A mesterséges intelligencia és a hadseregek.  
*Hadtudomány*, XXIX. évfolyam (2019), 3. szám, pp. 71–79. DOI 10.17047/HADTUD.2019.29.3.71
- Porkoláb Imre – Négyesi Imre: A mesterséges intelligencia alkalmazási lehetőségeinek kutatása a haderőben.  
*Honvédségi Szemle*, 147. évfolyam (2019), 5. szám, pp. 3–19.
- Tóth András: A mesterséges intelligencia és az IoT összekapcsolásának lehetőségei,  
*Hírvillám – Signal Badge*, 2018/1. pp. 122–130.
- Tóth András: Az Amerikai Egyesült Államok Védelmi Minisztériumának felhő koncepciója,  
*Hírvillám – Signal Badge*, 2018/1. pp. 115–121.