Négyesi Imre

A mesterséges intelligencia és a hadseregek

DOI 10.17047/HADTUD.2019.29.3.71

A cikk megírásának aktualitását a Honvédelmi és Haderõ-fejlesztési Konferencia és a meghirdetett Digital Soldier 2.0 program és egy 2019 elején elérhetõvé vált amerikai dokumentum jelenti. Ez utóbbi dokumentumot az USA Védelmi Minisztériuma adta ki a „Mesterséges Intelligencia Stratégia 2018" címmel, amely hosszú távú program ismét a figyelem központjába irányította a hadseregek mesterséges intelligencia kutatásait.

A honvédelmi miniszter 2018. 12. 13-án, a tárca felsõvezetõinek és a honvédség vezetõinek tartott elõadásában elmondta, hogy a Magyar Honvédség fejlesztésének jogszabályi feltételei adottak.[1] A Honvédelmi és Haderõ-fejlesztési Program többek között tartalmazza a katonák felszerelésének, technikai eszközeinek fejlesztését is. Ennek része a Digitális Katona program[2] is. Ebben a cikkben a magyar tervekhez és fejlesztésekhez kapcsolódóan nézzük meg, hogy az USA Védelmi Minisztériuma milyen lépéseket tett a mesterséges intelligenciának a hadseregben történõ hatékonyabb felhasználása érdekében.

Technológiai és szervezeti elõzmények

A technológia elõrehaladása, az MI megváltoztatta az emberek és gépek között fennálló szerepeket. Eredetileg az emberek a harc hagyományos formáiban vettek részt. Az ipari korszak megjelenésével azonban az emberek felismerték, hogy a gépek nagyban növelhetik háborús harci képességeiket. A hálózatok ezután lehetõvé tették a távfelügyeletet, amely végül kiszolgáltatottnak bizonyult az elektronikus támadásokra, valamint a hosszú jelátviteli távolságok és idõk miatt korlátozott volt a használatuk.

1 1298/2017. (VI.2.) Korm.határozat a Zrínyi 2026 Honvédelmi és Haderõ-fejlesztési Program megvalósításáról.

2 Digital Soldier 2.0

A hadviselés következõ szakasza sokkal ütõképesebb autonóm rendszereket foglalt magában, de mielõtt megengednénk az ilyen gépeknek az emberi harcosok kiegészítését, sokkal nagyobb intelligenciaszintet kell elérniük. Hagyományosan úgy terveztük meg a gépeket, hogy jól meghatározott, nagy mennyiségû és/vagy nagy sebességû feladatokat kezeljenek, és az emberek megszabaduljanak a rutinfeladatoktól.

Az 1950-es és 1960-as években a korai számítógépek már automatizálták az unalmas vagy munkaigényes feladatokat. Ebben a korszakban a tudósok rájöttek, hogy lehetséges az emberi intelligencia szimulálása, és megszületett a mesterséges intelligencia[3] (a továbbiakban: MI) kutatási területe. A feltételezés az volt, hogy az MI lenne az az eszköz, amely lehetõvé teszi a számítógépes problémamegoldását, és olyan funkciók végrehajtását, amelyek rendszerint emberi értelmet igényelnek. Az MI korai kutatásai során kiemelték az ún. kézmûves tudást, és a számítógép-tudósok szakértõi rendszereket állítottak össze, amelyek a szakértõk speciális ismereteit szabályozták azon szabályok szerint, hogy a rendszer alkalmassá váljon különbözõ helyzetekben történõ alkalmazásra. Az ilyen „elsõ hullám” MI-technológiák meglehetõsen sikeresek voltak.

Az elmúlt években az MI-kutatások egyik területévé vált az a terület, amely a gépi tanulással kapcsolatban merült fel. Ezen a területen olyan statisztikai és valószínûségi módszereket alkalmaznak a nagy adathalmazokra, általánosított ábrázolásokat hozva létre, amelyek a jövõbeli mintákra is alkalmazhatók lehetnek. Ezek közül a leginkább a mély tanulás (mesterséges) idegrendszeri hálózatai képzelhetõk el, amelyek alkalmasak arra, hogy felhasználják a különbözõ, rendelkezésre álló történeti adatokat. Azonban az adatgyûjtés, címkézés és az adatok azonosításának feladata, amelyre az ilyen „második hullám” MI-technikákat alapozzuk, meglehetõsen költséges és idõigényes.

A Fejlett Védelmi Kutatási Projektek Ügynöksége (Defense Advanced Research Projects Agency – DARPA) egy olyan jövõt tervez, amelyben a gépek többek, mint csupán olyan eszközök, amelyek emberi programozott szabályokat hajtanak végre, vagy az ember által kezelt adatkészletekbõl általánosítanak. A tervek szerint a DARPA által létrehozott gépek inkább kollégaként fognak mûködni, mint eszközként. Ebbõl a célból a DARPA kutatási és fejlesztési terveiben az ember–gép mûködése szimbiózisban célt tûzött ki a gépekkel való együttmûködésre.

A számítástechnikai rendszerek ilyen módon történõ engedélyezése kritikus fontosságú, mivel az érzékelõ, az információ és a kommunikációs rendszerek olyan sebességgel generálnak adatokat, amelyek mellett az emberek lemaradhatnak. Ezeknek a technológiáknak a beépítése a hadviselõkkel együttmûködõ katonai rendszerekbe megkönnyítheti a jobb döntéseket komplex, idõkritikus mûveleti környezetben, lehetõvé teheti a tömeges, hiányos és ellentmondásos információk közös megértését és felhatalmazhatná a pilóta nélküli rendszereket, hogy biztonságosan és nagyfokú autonómiával végezzék el a kritikus feladatokat. (A DARPA jelenlegi befektetéseit egy harmadik MI-hullámra fókuszálja, amelyrõl a késõbbiekben még szó lesz.)

3 A mesterséges intelligencia (Artificial Intelligence) angol rövidítése: AI.

Az USA Védelmi Minisztériuma (Department of Defense – DoD) által kiadott *Mesterséges Intelligencia Stratégia 2018* kiadásának közvetlen elõzménye a Mesterséges Intelligencia Kiválósági Központjának létrehozása volt. A Közös Mesterséges Intelligencia Központ (Joint Artificial Intelligence Center – JAIC) létrehozásának elsõdleges céljaként a hatékonyabb segítségnyújtást határozták meg a hadsereget támogató mesterséges intelligencia alkalmazások megvalósításához.

2018. június 27-én a Pentagon kiadott egy memorandumot, amelyben meghatározta a JAIC konkrét feladatait is. Meghatározták azt is, hogy a feladatokat a Védelmi Minisztérium információs vezetõje (Department of Defense Chief Information Officer – CIO) alárendeltségében kell végrehajtania. (Ez a magas szintû vezetés is jelzi, hogy milyen fontosnak tartják a mesterséges intelligencia katonai alkalmazásának területén folyó kutatásokat.)

A JAIC tevékenységével együttmûködve a hadsereg 2018-ban megalapította a saját MI kutatási tevékenységének támogatására az Army-AI Task Force-t[4] (A-AI TF), amelynek megalakítását egy 2018. október 02-án kelt memorandumban jelentették be. Ennek az alapját a Department of Defense Artificial Intelligence Strategy [1][2] és a Memorandum, Deputy Secretary of Defense, subject: Establishment of the Joint Artificial Intelligence Center [3] dokumentumok jelentették. A létrehozás alapgondolata az elõzõekbõl is adódóan az volt, hogy az új szervezet segítse elõ a jelenlegi technológiák alkalmazásának javítását, szûkítse a meglévõ MI-képességek hiányát.

A Kongresszusi Kutatási Szolgálat (Congressional Research Service – CRS) 2019 elején frissítette a Mesterséges Intelligencia és Nemzetbiztonság (Artificial Intelligence and National Security) elnevezésû összefoglalóját, amelybõl további részletek derülnek ki az amerikai kutatási programokról.

Mesterséges intelligencia a gyakorlatban

Természetesen már JAIC megalakítása elõtt is megindultak a mesterséges intelligenciakutatások. Ezek egyik eredménye, a Project Maven már segítette az ISIS elleni harcot Irakban és Szíriában.

A Projekt Maven-t, más néven Algorithmic Warfare Cross-Functional Team (AWCFT), 2017 áprilisában indítottak el azzal a céllal, hogy még 2017-ben a rendszert harci övezetekbe telepítsék. A katonai tervek szerint a Project Maven gépi tanulási algoritmusait használták volna fel a rögzített anyagok rendezésére azért, hogy potenciálisan segítséget nyújtsanak a katonai elemzõk és a hírszerzõk számára az ellenséges célok rangsorolásában.

Jelenleg folyamatban vannak a számítógépes látás algoritmusainak finomításai, amelyeket kisebb, alacsonyan repülõ drónokkal és nagy magasságú autonóm repülõgépekkel fejlesztettek ki. (Ilyen volt például a Global Hawk, amely 60 000 méteres magasságból képes megfigyelni a harci övezeteket.)

4 Teljes nevén Army Artificial Intelligence Task Force in Support of the Department of Defense Joint Artificial Intelligence Center (magyarul: Hadsereg Mesterséges Intelligencia Munkacsoport A Védelmi Minisztérium Közös Mesterséges Intelligencia Központjának Támogatása Érdekében).

A Project Maven célja az volt, hogy olyan szoftveres algoritmusokat dolgozzon ki, amelyek megkülönböztethetik a járműveket, az embereket és az autókat, valamint nyomon követhetik az érdeklődésre számot tartó objektumokat. A rendszer iterációja[5] kiemelheti a digitális térképen lévő elemeket, a gyakran ismétlődő hibák csökkentése az új objektumok nyomon követése érdekében.

A fejlesztést a Google-al együtt hajtotta végre a DoD egészen addig, amíg több mint 3000 Google-alkalmazott aláírt egy petíciót, amelyben tiltakoztak az ellen, hogy a vállalat részt vesz egy amerikai védelmi minisztérium mesterséges intelligencia (MI) projektjében. A Sundar Pichai vezérigazgatónak címzett nyílt levélben a Google alkalmazottai aggodalmukat fejezték ki azzal kapcsolatban, hogy az amerikai hadsereg hadrendbe állíthatja az MI-t és alkalmazhatja a technológiát a drónnal végrehajtott, és egyéb halálos támadások finomítására. A nyílt levél úgy kezdődött: „*Úgy véljük, hogy a Google nem vehet részt a háborús tevékenységben*", majd elmagyarázta, hogy a Google részvétele a Project Maven-ben a márkanév, és a nyilvánosság bizalmának sérelméhez vezethet. A Project Maven-t azonban csak PILOT jelleggel alkalmazták, ezért csak egy – bár úttörő – lépcsőfoknak tekinthető a „harctéri" alkalmazásban.[6]

A következő említésre méltó szervezet az SRI International, amely már díjat is kapott az amerikai hadsereg integrált vizuális bővítési rendszerének támogatásában kifejtett tevékenységért. A SRI International a díjat a Digitális Éjszakai Látványkamerák Amerikai Hadseregben Történő Alkalmazása (Integrated Visual Augmentation System – IVAS) című programjának támogatására kapta. [4] A díjat a System of Systems Consortium Inc. (SOSSEC) adta ki a Night Vision és az Electronic Sensor Directorate (NVESD) nevében.

A program támogatása érdekében az SRI egy nagy fényérzékenységű Complementary Metal-Oxide Semiconductor (CMOS) képérzékelőt tervez és integrálja az eszközt egy kisméretű, kis súlyú és nagy teljesítményű, optimalizált kameramo- dulba. Colin Earle, az Imaging Systems társigazgatója a programról azt mondta, hogy az „… *IVAS program hatalmas lehetőséget kínál az SRI International számára, hogy bemutassa a szilárd állapotú, alacsony fényszintű képalkotási technológiát egy alacsony SWAP kameramodulban, amely fokozhatja a katonák helyzetének pontos meghatározását*". Majd hozzátette, hogy az „… *SRI folyamatosan fejlődött az éjjellátó CMOS képérzékelők alacsony fényszintű teljesítményében, és örülünk annak, hogy az IVAS program már beépíti a negyedik generációs NV-CMOS képeit*".

Az integrált vizuális bővítési rendszert (IVAS) úgy alakították ki, hogy az egyes katonákra fej-, test- és fegyvertechnológiákat alkalmazzon. Ez egy olyan platform, amelyet a hadsereg használhat arra, hogy csökkentse a személyi veszteségeket. (A rendszer magában foglalja a csapat szintű harcképzési képességeket is és az ismétlések lehetőségével a begyakoroltatást is.)

Hagyományosan a nappali és az éjjellátó képalkotó rendszerek szükséges képerősítőkkel és nappali kamerákkal rendelkeznek. Az SRI új NV-CMOS képérzékelő technológiája azonban arra szolgál, hogy a fényes napfénytől a csillagfényig terjedő

5 Az *iteráció* valójában egy függvény ismételt végrehajtását jelenti az előző függvényértéken.

6 https://gizmodo.com

teljes fényviszonyok között képeket készítsen. Ez a lehetõség a hadseregek számára is felértékeli a technológiát. Technikailag ez úgy néz ki, hogy az SRI NV-CMOS képérzékelõi a nagy fényérzékenységet lehetõvé teszik, amely analóg képerõsítõ csövével közelít a digitális CMOS képfeldolgozó chipekhez költségében, teljesítményében, szilárdságában, rugalmasságában. Az NV-CMOS multi-megapixeleket biztosít alacsony képzajú video-képsebességgel, nagy láthatósággal a látható és közeli infravörös sávokon.

A hagyományos képerõsítõkkel ellentétben az NV-CMOS képérzékelõ digitális jelet ad ki, amely ideális a videó felvételéhez, vagy megosztásához, valamint a hõképekkel való fúzióhoz. Az eredmény a méret és a súly jelentõs csökkenése, ideális korlátozott küldetésekhez (például UAV-khoz) és mobil mûveletekhez. Az SRI mozgásadaptív zajcsökkentõ feldolgozása tovább növeli az érzékenységet és csökkenti a kép elmosódását. A szélsõségesen gyenge fényviszonyok között rögzített képeken a mozgó célpontok követése nagy kihívást jelent.

A DARPA több mint öt évtizede vezetõ szerepet tölt be a kutatás és fejlesztésben (K + F), amely elõsegítette a szabályalapú és statisztikai-tanulás alapú MI-technológiák fejlõdését és alkalmazását. [5] Napjainkban a DARPA továbbra is vezeti az innovációt az MI-kutatásban, mivel a kutatás-fejlesztési programok széles választékát finanszírozza, kezdve az alapkutatástól a fejlett technológia fejlesztéséig. A DARPA úgy véli, hogy a „Harmadik Hullám" (vagy harmadik generációs) MI-technológiák kifejlesztése és alkalmazása során ez a jövõ. Megvalósíthatók olyan rendszerek, amelyek generatív kontextusos és magyarázó modellek révén képesek új ismereteket szerezni és létrehozni.

A fejlesztések elõsegítésére a DARPA 2018 szeptemberében több mint 2 milliárd dolláros többéves beruházást jelentett be. A DARPA a fejlesztési kampánya az „MI Next Campaign" néven fut, és a következõ három pilléren nyugszik:

– AI Exploration.
– Ongoing AI Programs.
– AI Colloquium.

A DARPA kutatásainak kulcsfontosságú területei közé tartoznak a kritikus DoD üzleti folyamatok automatizálása, mint például:

– a biztonsági ellenõrzés vagy a szoftverrendszerek akkreditálása az operatív telepítéshez;
– az MI-rendszerek robusztusságának és megbízhatóságának javítása;
– a gépi tanulás és az MI-technológiák biztonságának és rugalmasságának növelése;
– a teljesítmény növelése, az adathiány és a teljesítményhiány csökkentése;
– a következõ generációs MI-algoritmusok és MI-alkalmazások.

Ezek a területek összhangban állnak a DARPA mesterséges intelligencia kutatási programjával (AIE)[7], amely az ügynökség szélesebb körû mesterséges intelligencia befektetési stratégiájának kulcsfontosságú eleme. [6] Ennek megfelelõen a DARPA erõteljes képességeket teremt a DoD számára az alábbi területeken való részvételével:

7 AIE – Artificial Intelligence Exploration.

– *Új képességek kifejlesztése*
  Az MI-technológiákat rutinszerûen alkalmazzák, hogy támogassák a DARPA K + F projektjeit, beleértve a több mint 60 létezõ programot, mint például a DARPA Microsystems Technology Office (MTO) által bejelentett Elekctronics Resurgence Initiative[8] (ERI) és más, a kifinomult számítógépes támadások valós idejû elemzéséhez kapcsolódó programokat.[9] Az ERI foglalkozik az elektronikai biztonság és a magánélet védelmével kapcsolatos kérdésekkel, a differenciált elektronikai hozzáférések kérdéseivel és maximálni akarja a nemzeti védelmi alkalmazásokra gyakorolt hatását. Kutatási területei közé tartoznak a manipulált képek felderítése, az emberi nyelvtechnológiákhoz, az automatikus célfelismeréshez, az orvos-biológiai elõrelépésekhez és a protézis végtagok ellenõrzéséhez kapcsolódó programok.
– *Robusztus MI*
  Az MI-technológiák nagy értéket képviseltek olyan változatos területeken, mint az ûralapú képelemzés, a számítógépes figyelmeztetés, az ellátási lánc logisztikája és a mikrobiológiai rendszerek elemzése. Ugyanakkor az MI-technológiák meghibásodási módjai rosszul érthetõek, ezért a DARPA arra törekszik, hogy foglalkozzon ezzel a hiányossággal. A DARPA sikere elengedhetetlen ahhoz, hogy a minisztérium az MI technológiákat alkalmazza, különösen a harcászati szinten, ahol megbízható teljesítményre van szükség a személyi állomány megóvása érdekében.
– *Az ellentétes MI*
  A legerõsebb MI-eszközök ma a gépi tanulás (ML) rendszerei. A tanulórendszereket könnyen meg lehet semmisíteni az olyan bemenetek megváltoztatásával, amelyek soha nem csapnának be egy embert. Az ilyen rendszerek képzéséhez használt adatok megsérülhetnek és a szoftver maga is ki van téve a számítógépes támadásoknak. Ezeket a területeket és még többet kell kezelni, amivel több MI-képes rendszer mûködtethetõ.
– *Nagy teljesítményû MI*
  A számítógép teljesítményének növekedése az elmúlt évtizedben lehetõvé tette a gépi tanulás sikerességét, nagy adatállományokkal és szoftverkönyvtárakkal kombinálva. Az adatközpont és a taktikai telepítések lehetõvé tétele érdekében nagyobb teljesítmény érhetõ el alacsonyabb elektromos teljesítmény felhasználásával. A DARPA az MI-algoritmusok analóg feldolgozását 1000-szeres gyorsítással és 1000-szeres energiahatékonysággal mutatta be a legmodernebb digitális processzorokon, és az MI-specifikus hardverterveket kutatja. A DARPA is kutatja a gépi tanulás jelenlegi hatástalanságát azáltal, hogy olyan módszereket vizsgál, amelyek jelentõsen csökkentik a címkézett képzési adatok követelményeit.
– *A következõ generációs MI*
  Az arcfelismerést és a vezetõ nélküli jármûveket lehetõvé tevõ gépi tanulási algoritmusokat több mint 20 éve találták fel. A DARPA vezetõ szerepet tölt be a kutatásban az MI-algoritmusok következõ generációjának kifejlesztésében, amely az eszközöket problémamegoldó partnerekké alakítja át.

8 Elektronikus újraélesztési kezdeményezés.

9 A DARPA már megkezdte az ERI II. Fázisát is.

Az MI-kutatás új irányai, eredményei

A DARPA egyik új kutatási iránya az *MI-rendszerek tanítása a dinamikus környezetekhez való alkalmazkodás érdekében* címet viseli. Az új program célja olyan mesterséges intelligencia (MI) létrehozása, amely önállóan felismeri és reagál a valós körülmények változásaira. A jelenlegi MI-rendszerek kiemelkednek a merev szabályok által meghatározott keretek között végrehajtható feladatok közül, mint például a Go and Chess játékok és a világszínvonalú emberi játékosokat meghaladó képességekkel rendelkeznek.

Az MI-rendszerek azonban nem igazán alkalmasak arra, hogy alkalmazkodjanak a folyamatosan változó körülményekhez, amelyekkel a csapatok általában a valós világban szembesülnek (például az ellenfél meglepetésére való reagálástól a változó idõjárásig, az ismeretlen terepen való mûködésig). Ahhoz, hogy az MI-rendszerek hatékonyan együttmûködjenek az emberekkel a katonai alkalmazások széles skáláján, az intelligens gépeknek zárt világban kell megoldaniuk a korlátozott határokon belül a folyékony és új helyzetek által jellemzett nyílt világ kihívásait.

Az új kihívásokra való reagálásként a DARPA bejelentette a *Mesterséges intelligencia és tanulás a nyílt világ újdonságaira* (Science of Artificial Intelligence and Learning for Open-world Novelty – SAIL-ON) címet viselõ programját. [7] A SAIL-ON olyan kutatási alapokat és általános mérnöki technikákat és algoritmusokat kíván kutatni és fejleszteni, amelyek szükségesek ahhoz, hogy a világban elõforduló, a különbözõ területeken megjelenõ új helyzetekben megfelelõen és hatékonyan mûködjenek. A program célja, hogy tudományos alapelveket dolgozzon ki az újdonságok számszerûsítésére és jellemzésére a nyílt világban, létrehozza az újdonságokra reagáló MI-rendszereket, és bemutassa ezeket a rendszereket egy kiválasztott alkalmazási területen.

A kutatási irányokhoz az alapot az szolgáltatja, hogy a meglévõ MI-rendszerek hatástalanná válnak, és nem tudnak alkalmazkodni, ha valami jelentõs és váratlan történik. Ellentétben az emberekkel, akik felismerik az új élményeket és megfelelõen módosítják viselkedésüket, a gépek továbbra is elavult technikákat alkalmaznak, tehát szükséges a gépek „továbbképzése". Elegendõ adatot adva, a gépek képesek statisztikai érvelésre (például képek felismerésére, arcfelismerésre stb.). Egy régebbi példa: a 2000-es évek elején az vezetõ nélküli autókba a DARPA MI-beépítése volt, ami az autonóm jármûvek jelenlegi forradalmához vezetett. A hatalmas mennyiségû adatnak köszönhetõen, amelyek több tízmillió mérföldrõl gyûjtöttek össze ritka eseményeket, az vezetõ nélküli technológia megvalósítható lett. A rendelkezésre álló adatok azonban általánosan jól meghatározott környezetekre vonatkoztak, ismert útszabályokkal.

„Nem lenne célszerû megpróbálni egy hasonló adatbázist készíteni több millió km^2-re a katonai földi rendszerek számára, amelyek a terepen dolgoznak, ellenséges környezetben, és állandóan új feltételek mellett, magas tétekkel szembesülnek. Nem is beszélve a levegõben és a tengeren dolgozó autonóm katonai rendszerekrõl"[10] – mondta Ted amerikai szenátor.[11] Ha azonban sikerül, a SAIL-ON, akkor egy MI-rendszer vezérelné, hogyan kell tanulni és megfelelõen reagálni anélkül, hogy nagy adatállományon át kellene képeznie.

10 https://www.darpa.mil/news-events/2019-02-14.

11 Defense Sciences Office (DSO) Program Manager.

A program arra törekszik, hogy megteremtse a technikai alapot, amely felhatalmazza a gépeket, függetlenül a tartománytól, hogy maguktól figyeljék meg a helyzetet, reagáljanak arra, amit megfigyelnek, döntsék el a legjobb cselekvési módot, majd cselekedjenek.

Pontokba szedve az MI tevékenységeit, ez a folyamat így nézne ki:

– Az elsõ dolog, amit az MI-rendszernek meg kell tennie, hogy felismerje a környezetet.
– A második dolog, amit meg kell tennie: a környezet megváltoztatásának jellemzése.
– A harmadik dolog, amit meg kell tennie: megfelelõ választ kell adnia.
– A negyedik dolog, amint megtanulja az alkalmazkodást az, hogy frissítse a környezet modelljét.

A DARPA következõ megemlítendõ programjának célja, hogy új generációs védelmet alakítson ki a gépi tanulási algoritmusok megtévesztésének megakadályozására. [8] Ma a gépi tanulás (mashine learning – ML) az emberiséget sokféle alkalmazásban (a rendkívül hatékony gyártástól, az orvostudománytól és a tömeges információelemzéstõl az önálló vezetésig, és azon túl) képes kiszolgálni. Ha azonban helytelenül alkalmazzák, rosszul használják, akkor az ML nagy veszélyeket hordoz. „*Az elmúlt évtizedben a kutatók arra összpontosítottak, hogy megvalósítsák a valós feladatok elvégzésére és hatékonyabbá tételére alkalmas gyakorlati ML-eket*” – mondta dr. Hava Siegelmann, a DARPA Információs Innovációs Iroda (I2O)[12] programmenedzsere.[13] De véleménye szerint kevés figyelmet szenteltek az ML-platformokra jellemzõ sebezhetõségeknek, különösen azokra, amelyek a rendszerek megváltoztatását, sérülését vagy megtévesztését célozták.

Ahhoz, hogy ezt az akut biztonsági kihívást megelõzzék, a DARPA létrehozta az MI Szilárdságának Biztosítása A Megtévesztés Ellen (Guaranteeing AI Robustness against Deception – GARD) elnevezésû programot. A GARD célja, hogy új generációs védelmet alakítson ki az ML-modellekkel szembeni ellentmondásos, megté- vesztõ támadások ellen.

A jelenlegi védelmi erõfeszítéseket úgy tervezték, hogy kivédjék a meghatározott, elõre meghatározott, ellentmondásos támadásokat, és a tesztelés során kitértek a tervezési paramétereken kívüli támadásokra is. A GARD másképp igyekszik megközelíteni az ML védelmet, olyan széles körû védelmet fejlesztve, amelyek az adott forgatókönyvben szereplõ számos lehetséges támadást kezelik. Siegelmann azt is kimondta, hogy „*Szükség van az ML védelemre, mivel a technológia egyre inkább beépül néhány kritikus infrastruktúránkba. A GARD program célja, hogy megakadályozza a közeljövõben elõforduló káoszt, amikor a támadási módszerek már rombolóbb szintre értek. Biztosítanunk kell, hogy az ML biztonságos és megtéveszthetetlen legyen!*”

A GARD új válaszai, ennek megfelelõen, három fõ célra összpontosítanak:

– a megvédhetõ ML elméleti alapjainak és az ezeken alapuló új védelmi mechanizmusok lexikonjának kidolgozása;

12 DARPA’s Information Innovation Office.

13 https://www.darpa.mil/news-events/2019-02-06.

– a védett rendszerek létrehozása és tesztelése a különbözõ beállítások között;
– egy új próbaplattform építése az ML-védettség jellemzésére a fenyegetettségi forgatókönyvekhez viszonyítva.

Ezen egymástól függõ programelemek révén a GARD célja, hogy megtévesztõ rezisztens ML-technológiákat hozzon létre, amelyek szigorú kritériumokkal rendel- keznek az értékelésekhez. A GARD számos kutatási irányt figyel a potenciális véde- lemre, beleértve a biológiát is. A GARD a jelenlegi igények kezelésére fog törekedni, de a jövõbeli kihívásokat is szem elõtt tartja. A tervek szerint a program kezdetben a legmodernebb képalapú ML-re összpontosít, majd továbblép a video, az audio és a bonyolultabb rendszerekre. Arra is törekszik, hogy az életciklusa során elõrejelzé-sekre, döntésekre és alkalmazkodásra képes ML-t kezelje.

Összefoglalás, következtetések

Összességében tehát a mesterséges intelligenciával kapcsolatos védelmi kutatások ismét az érdeklõdés központjába kerültek. A Magyar Honvédség tekintetében ezt tükrözi a Honvédelmi és Haderõ fejlesztési Konferencia megtartása és a meghirde- tett Digital Soldier 2.0 program. Továbbá az is, hogy a megalakult Mesterséges Intelli- gencia Koalíció egyik alapító tagja a Honvédelmi Minisztérium.

A NATO-országok kutatási irányait egy 2019 elején elérhetõvé vált amerikai dokumentummal jellemezhetjük. Ez utóbbi dokumentumot az USA Védelmi Minisz- tériuma adta ki a *Mesterséges Intelligencia Stratégia 2018* címmel és ezzel hosszú idõre kijelölte a védelmi kutatások jellemzõ irányait.

FELHASZNÁLT IRODALOM

[1] Summary of the 2018 National Defense Strategy of The United States of America; https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf (Letöltés: 2019. 06. 20.)

[2] Summary of the 2018 Department of Defense Artificial Intelligence Strategy; https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF) (Letöltés: 2019. 06. 20.)

[3] Memorandum, Deputy Secretary of Defense, subject: Establishment of the Joint Artificial Intelligence Center; https://admin.govexec.com/media/establishment_of_the_joint_artificial_intelligence _center_osd008412-18_r....pdf (Letöltés: 2019. 06. 20.)

[4] SRI International Awarded Contract to Support U. S. Army Integrated Visual augmentation System (Menlo Park, 2019. április 15.); https://www.sri.com/newsroom/press-releases/sri-international-awarded-contract-support -us-army-integrated-visual (Letöltés: 2019. 06. 20.)

[5] AI Next Campaign; https://www.darpa.mil/work-with-us/ai-next-campaign (Letöltés: 2019. 06. 20.) [6] Accelerating the Exploration of Promising Artificial Intelligence Concepts; https://www.darpa.mil/news-events/2018-07-20a (Letöltés: 2019. 06. 20.)

[7] Mr. Ted Senator: Science of Artificial Intelligence and Learning for Open-world Novelty (SAIL-ON); https://www.darpa.mil/program/science-of-artificial-intelligence-and-learning-for -open-world-novelty (Letöltés: 2019. 06. 20.)

[8] Defending Against Adversarial Artificial Intelligence; https://www.darpa.mil/news-events/2019-02-06 (Letöltés: 2019. 06. 20.)