

### Gépi tanulással támogatott mesterséges intelligencia a Fegyverzetellenőrzés

Lück, Nico

Publikációs verzió / Published Version Arbeitspapier /  
working paper

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**

Hessische Stiftung Friedens- und Konfliktforschung (HSFK)

**Empfohlene Zitierung / Javasolt idézet:**

Lück, N. (2019). *Gépi tanulásra épülő mesterséges intelligencia a fegyverzetellenőrzésben.* (PRIF Reports, 8). Frankfurt am Main: Hessische Stiftung Friedens- und Konfliktforschung. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-69004-6>.

**Nutzungsbedingungen:**

*Dieser Text wird unter einer CC BY-ND Lizenz (Namensnennung- Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie itt:* <https://creativecommons.org/licenses/by-nd/4.0/deed.de>

**Használati feltételek:**

*Ez a dokumentum a CC BY-ND licenc (Attribution-NoDerivatives) alatt érhető el. További információért lásd:* <https://creativecommons.org/licenses/by-nd/4.0>

# PRIF REPORT

PEÁCERESZTITUTTEFRANKFURT/LEIBNIZ-INSTITUTTUTTESZERESZTITUTTESFRIEDENS-UNDKONFLIKTFORSCHUNG.



NICO LÜCK //

## A GÉPI TANULÁSON ALAPULÓ MESTERSÉGES INTELLIGENCIA A FEGYVERZETELLENŐRZÉSBN

PRIF jelentés 8/2019

## **A GÉPI TANULÁSON ALAPULÓ MESTERSÉGES INTELLIGENCIA A FEGYVERZETELLENŐRZÉSSEN**

NICO LÜCK //

LEIBNIZ-INSTITUT HESSISCHE STIFTUNG FRIEDENS- UND KONFLIKTFORSCHUNG (HSFK)  
FRANKFURTI BÉKEKUTATÓ INTÉZET (PRIF)

**Borító:**

Nico Lück, saját szám

**Szöveges licenc:**

Creative Commons CC-BY-ND (Attribution/NoDerivatives/4.0 International). A felhasznált képek saját licencük alá tartoznak.



**Levelezés:**

Békekutató Intézet Frankfurt Baseler  
Straße 27-31  
D-60329 Frankfurt am Main  
Telefon: +info@hsfk.de  
<https://www.prif.org699591>

**ISBN: 978-3-946459-51-4**

Mindenki a mesterséges intelligenciáról (AI) beszél, és a témát a jövő egyik központi kérdésének tekintik, mivel előre látható, hogy az AI-rendszerek ereje rendkívüli előrelépéseket tesz majd lehetővé a legkülönbözőbb alkalmazásokban - erőforrás-optimalizálás, előrejelzés, tárgyfelismerés, ember-számítógép interakció vagy robotalapú rendszerek vezérlése. Ez még inkább így lesz, ha az AI rendelkezik az úgynevezett "gépi tanulás" képességével, azaz azzal a képességgel, hogy megfigyelések vagy szolgáltatott adatok alapján saját szabályokat alakítson ki. Ez a képesség példátlan függetlenséget tesz lehetővé az emberi programozóknak az események előrejelzésére vonatkozó képességétől.

A fegyverkezésben ezek a fejlesztések már most is fontos szerepet játszanak: A hagyományos mesterséges intelligenciát már készen áll a harci repülőgépek, drónok, lövegternyok vagy humanoid robotok irányító és segítő egységként való alkalmazása, például a navigáció és a célfelismerés terén. A gépi tanulással támogatott mesterséges intelligenciát (MLpAI) jelenleg új fegyverrendszerek fejlesztésénél tesztelik vagy prototípusokba integrálják.

Ezek a tendenciák a mesterséges intelligenciát a fegyverzetellenőrzés fontos kérdésévé teszik, mégpedig két szempontból is. A fegyverzetellenőrzés tárgyaként az MLpAI kívül esik a hagyományos megközelítéseken, mivel nem rendelkezik sem azokkal a fizikai tulajdonságokkal vagy képességekkel, sem azokkal az átlátható műveletekkel, amelyekre a mennyiségi és minőségi fegyverzetkorlátozás jelenlegi módszerei és eljárásai épülnek. Másrészt az MLpAI új eszközöket biztosít a fegyverzetellenőrzéshez. Így elképzelhető, hogy a meglévő és új fegyverzetellenőrzési szerződések ellenőrzése, azaz a betartásuk ellenőrzése jelentősen profitálhat az MLpAI-ból mint technikai eszközből, például az adatok gyűjtésének, feldolgozásának és elemzésének pontossága és gyorsasága révén.

A biztonságpolitika szempontjából az MLpAI tehát kockázatokat és lehetőségeket egyaránt rejt magában, és a gépi tanulás használatának várható növekedése jelentős mértékben növeli ezeket a kockázatokat és lehetőségeket. A kockázatot az jelenti, hogy az MLpAI-t, mint a jövő autonóm fegyverrendszerek központi elemét, a fegyverzetellenőrzésnek kell korlátoznia, ugyanakkor a fegyverzetellenőrzés nem rendelkezik megfelelő technikai képességekkel. Ezen a ponton a hagyományos megközelítések kimerültek, és az MLpAI fejlesztések vagy telepítések során történő nyomon követésének és korlátozásának lehetősége továbbra is fennáll. Ha az MLpAI-t ellenőrizetlenül használják, az a stratégiai stabilitást veszélyezteti azáltal, hogy minimalizálja a de-eszkaláló emberi tényezőt, elősegíti a technológiai fegyverkezési versenyt, és ellenőrizhetetlenül terjed. Ezzel szemben áll az MLpAI-ban rejlő hatalmas potenciál a fegyverzet-ellenőrzési megállapodások ellenőrzésére. A műholdképeken, videofelvételeken vagy elektromágneses, szeizmikus vagy akusztikus érzékelők adataiban szereplő objektumok, jelenségek és időbeli változások azonosítása bizonyíthatóan javul a gépi tanulás alkalmazásával. A sokkal pontosabb és átfogóbb információfeldolgozás növelheti az átláthatóságot, elriaszthatja a szereplőket a megállapodások megsértésétől, vagy bizonyítékként szolgálhat a megállapodások betartásának igazolására.

A jelentés arra a következtetésre jut, hogy az MLpAI egyszerre része a problémának és része a megoldásnak. A jelenlegi és jövőbeli alkalmazási példák azt mutatják, hogy az MLpAI a modern fegyverrendszerek központi eleme, és így magának a fegyverzetellenőrzés tárgyát kellene képeznie - különösen, mivel a fegyverrendszerek képességeit, de egyben veszélyeit is nagymértékben növelné az MLpAI. Az ellenőrzés tárgyaként azonban számos minőségi vagy mennyiségi korlátozásra irányuló megközelítés alól kivonja magát. Így növeli az alternatív módszerek jelentőségét az általános katonai átláthatóság és bizalomépítés szempontjából. Ez pontosan

ezekben a módszerekben azonban, hogy az ellenőrzés eszközeként használható. A pontos és kiterjedt információfeldolgozás révén nagyobb átláthatóságot teremthet, és ellenőrizheti a megállapodások betartását, ezáltal erősítve a felek közötti bizalmat. Az MLpAI két lehetséges alkalmazási területének fejlődése és az ellenőrzési intézkedések kisebb technikai bonyolultsága azt mutatja, hogy a korai cselekvés támogathatja az MLpAI potenciális előnyeit, mivel már most segíthet a fegyverzetellenőrzésnek új kapacitások kialakításában, mielőtt az MLpAI által ellenőrzött fegyverrendszerek új kihívásával szembesülne.

1. Bevezetés	1
2. Gépi tanulással támogatott mesterséges intelligencia	2
2.1 A mesterséges intelligencia forradalmának motorja: Gépi tanulás	3
2.2 Kihívások a gépi tanulás fejlesztése és használata során	5
3. Gépi tanulásra épülő mesterséges intelligencia a fegyverrendszerekben	7
3.1 A mesterséges intelligencia mint a modern fegyverrendszerek központi eleme és a tanulási képesség mint multiplikátor	8
3.2 Az ellenőrzést megnehezítő jellemzők és új megközelítések	9
3.3 Fogalmi reflexió: Stratégiai instabilitás és proliferáció	12
4. Gépi tanulással támogatott mesterséges intelligencia az Ellenőrző intézkedésekben	16
4.1 Alkalmazások és műszaki lehetőségek	16
4.2 Fogalmi reflexió: A meglévő módszerek javítása	22
5. A gépi tanulással támogatott mesterséges intelligencia a probléma egy része, és egy része a megoldás	23
Hivatkozások	24

## 1. Bevezetés

A mesterséges intelligencia (AI) alkalmazása a fegyverrendszerekben külső felügyelet vagy ellenőrzés nélkül olyan kockázatot jelent az emberiségre, amely a technológia növekvő képességeivel egyre nagyobb mértékben növekszik. A kockázat akkor a legnyilvánvalóbb, ha a mesterséges intelligenciát szándékosan pusztító célokra alkalmazzák, vagy ha az adott cél elérése érdekében saját kezdeményezésére választ ilyen utat. Másrészt a mesterséges intelligencia fejlődése pozitív lehetőségeket is kínál az emberiség számára. A mesterséges intelligencia rendkívüli információfeldolgozó képessége lehetővé teszi, hogy látszólag strukturálatlan adathalmazokban mintákat ismerjen fel és értelmezzen, előre meghatározott vagy megszerzett tudás alapján problémákat oldjon meg, intézkedéseket tervezzen, vagy következtetéseket vonjon le. A gyakorlatban a mesterséges intelligenciát olyan területeken alkalmazzák, mint az erőforrások felhasználásának optimalizálása, az időbeli fejlemények előrejelzése, a rögzített képeken lévő tárgyak felismerése, az emberekkel való kommunikáció vagy a robotrendszerek vezérlése. A mesterséges intelligencia használatában rejlő kockázatok és lehetőségek közötti feszültség az alapvető kérdés, amelyet ez a jelentés vizsgál.

A mesterséges intelligencia végrehajtásának pozitív és negatív következményei közötti ambivalencia a fegyverzetellenőrzésben is megfigyelhető: A fegyverzetellenőrzés átfogó célja, hogy "nagyobb fokú stratégiai stabilitást teremtsen két vagy több állam között", és így "csökkentse a háború kitörésének valószínűségét válság idején" (Croft 1996: 91-92). E célból a fegyverzetellenőrzési intézkedések "(a) befagyasztják, korlátozzák, csökkentik vagy megszüntetik a fegyverek bizonyos kategóriáit; (b) megtiltják bizonyos fegyverek tesztelését; (c) megakadályoznak bizonyos katonai tevékenységeket; (d) szabályozzák a fegyveres erők bevetését" (Goldblat 2002: 3).

Mint a fegyvertechnológiák bármely innovációjánál, a kormányok a fegyverrendszerek technológiáját is arra igyekeznek felhasználni, hogy technológiai fölényt szerezzenek vagy ellensúlyozzanak, és így stratégiai előnyre tegyenek szert más államokkal szemben. Következésképpen, ha a stratégiai stabilitás javítása érdekében szabályoznák a mesterséges intelligencia fegyverrendszerekben való alkalmazását, az új kihívást jelentene a fegyverzetellenőrzés számára. Az ellenőrzés tárgyaként a mesterséges intelligencia egy olyan sorba tartozna, amely magában foglalná az aknákat, lőszerket, kézfegyvereket, hagyományos fegyvereket, tömegpusztító fegyvereket és hordozórendszereket. Ezek azonban olyan fizikai tárgyak, amelyeket technikai, földrajzi vagy alkalmazással kapcsolatos jellemzők alapján szabályoznak. Ha ezt a mesterséges intelligenciával is meg akarjuk tenni, meg kell határozni a mesterséges intelligencia olyan sajátos tulajdonságait, amelyek lehetővé teszik a korlátozást és a nyomon követést. Másrészt, ha a mesterséges intelligenciát nem szabályozzák, akkor a stratégiai stabilitásra nézve eddig nem ismert következményekkel járhat.

Ez a stabilitást fenyegető veszély egy olyan időszakban merül fel, amikor a fegyverzetellenőrzés már válságban van a katonai-technikai fejlődés, a szerződészegések és a politikai akarat hiánya miatt (Arbatov 2015; Schmidt 2017). Ugyanakkor a mesterséges intelligencia új lehetőségeket kínál a hagyományos fegyverek és a tömegpusztító fegyverek ellenőrzésére, és segíthet a meglévő és új fegyverzetellenőrzési szerződések betartásának ellenőrzésében. Az ellenőrzési szerződésekbe és az államok közötti bizalom növelése érdekében a helyszínen kívüli és helyszíni ellenőrzések átláthatóságot teremtenek, és segítenek az államok teljesítményének ellenőrzésében. E célból általában technikai segédeszközöket vetnek be az információk gyűjtésére, feldolgozására és elemzésére (Goldblat 2002: 310). Az ellenőrzési intézkedésekben rejlő lehetőségek az új technológiákkal megnöttek, mivel a műholdak, érzékelők és más megfigyelési technikák javítják az információs helyzetet (Pilat 2002: 81). Ezen a ponton a

mesterséges intelligencia többszörös segítség lehet, mivel ezeket az információkat - különösen a nagy adathalmazokat - nagyobb pontossággal és gyorsabban tudja elemezni, mint a hagyományos módszerek.

Ez a jelentés először a "gépi tanulással támogatott mesterséges intelligencia" (MLpAI) kutatási tárgyat és a tanulási képesség fejlesztésének jelenlegi kihívásait tárgyalja. Ezt követően azt vizsgálja, hogy milyen veszélyek merülnek fel az MLpAI fegyverrendszerekben való alkalmazásából (3. fejezet), és hol nyílnak új lehetőségek a fegyverzetellenőrzés és a fegyverzetellenőrzés számára (4. fejezet). A jelentés arra a következtetésre jut, hogy az MLpAI a fegyverrendszerekben számos hagyományos fegyverzet-ellenőrzési megközelítéssel elkerülhető az ellenőrzés, ugyanakkor növelheti az alapvető átláthatóságot és erősítheti az ellenőrzési intézkedésekben részt vevő felek közötti bizalmat (5. fejezet).

## 2. Machine Learning-alapú művészeti intelligencia

Amikor arra a kérdésre próbálunk választ adni, hogy mit értünk "mesterséges intelligencia" alatt, be kell ismerni, hogy a kérdésnek számos megközelítése van, de nincs általánosan elfogadott definíció. "A mesterséges intelligenciának körülbelül annyi definíciója van, ahány kutató a technológiát fejleszti" (McCloskey 2017).

A legtöbb meghatározás azonban két alapvető jellemzőben egyetért: 1) rendkívül összetett feladatok megoldása és 2) alkalmazkodás a környezethez.<sup>1</sup> Azok a mesterséges intelligencia rendszerek, amelyek e két alapvető területen különösen jól teljesítenek, többnyire tanulási képességet mutatnak. Ez a tulajdonság kiemelkedik a mesterséges intelligencia elismert képességei - az észlelés, a tudás reprezentációja, a problémamegoldás, a tervezés és az érvelés - közül. Ezt támasztja alá a hardverek teljesítményének - különösen a processzorok számítási teljesítményének - még mindig folyamatos növekedése, valamint a nagy adathalmazok tanulási eszközként való rendelkezésre állása.

A múltban a számítógépes programok csak olyan dolgokat tudtak értelmezni, amelyekre a programozó meghatározott szabályokat, más szóval feltételes "ha-akkor" kapcsolatokat. Ha a jövőbeli helyzeteket vagy az időbeli változásokat az emberi programozó nem tudja előre látni, vagy ha a programozók maguk sem tudják a megoldást, akkor a gépi tanulás alkalmazása segít (Russell/Norvig 2010: 693). Ez olyan rendszerekre utal, amelyek a nyers adatokból mintákat kinyerve saját tudást generálnak. Ezzel az innovációval a mesterséges intelligencia képes megoldani komplex valós problémákat anélkül, hogy a programozók által megadott megoldásokra támaszkodna (Goodfellow 2016: 3). E jelentés középpontjában az ilyen, gépi tanulással támogatott mesterséges intelligencia (MLpAI) áll, mivel a tanulási képesség előfeltétele a mesterséges intelligencia két fent említett alapvető tulajdonsága - az összetett feladatok megoldásának képessége és az alkalmazkodóképesség - terén elért kiemelkedő eredményeknek.

Amellett, hogy ez a jelentés szándékosan a konkrét tanulási képességekre összpontosít, a mesterséges intelligencia vizsgálandó aspektusait az általános céljuk szerint is korlátozza: egy általánosan intelligens gép a cél, vagy egy olyan gép, amely csak egy adott tudományágban tekinthető "intelligensnek"? Lényegében minden definíció e felosztás szerint osztályozható. Bár az általánosan intelligens gép víziója közérdekű, még nincs olyan rendszer, amely megfelelne ennek a követelménynek. Mivel a szakértők az úgynevezett mesterséges általános intelligencia (AGI) fejlesztési idejét legalább évekre<sup>50</sup> becsülik (Müller/ Bostrom 2016: 559), egy ilyen definíció nem jelent praktikus kiindulópontot a jelenlegi és a legújabb alkalmazások elemzéséhez. Ehelyett ebben a jelentésben kizárólag olyan példákat vizsgálunk az AI-ra, amelyeket

1 A mesterséges intelligencia meghatározásainak összefoglalása megtalálható a következő dokumentumokban: Artificial General Intelligence Sentinel Initiative (2017); Legg/Hutter (2007).

egy adott tudományágra vagy feladatra kifejlesztett vagy optimalizált. A mesterséges intelligencia e már létező típusa számára az ember nem modellként, hanem legfeljebb teljesítménykritériumként szolgál. Az "alkalmazáspecifikus" - más néven "szűk AI" (Franklin 2014: 16) vagy "gyenge AI" (Searle 1980: 417) - célja, hogy legalább egyetlen diszciplínában értékelhető eredményeket vagy "intelligens" teljesítményt érjen el.

Az alkalmazáspecifikus, gépi tanulással támogatott mesterséges intelligenciára összpontosítva a jelentésben szereplő empirikus eredmények áttekintése elsősorban az új, innovatív mesterséges intelligencia programokra összpontosít, és kizárja mind a hagyományos számítógépes programokat, mind a futurisztikus mesterséges intelligencia koncepciókat. Ez a jelentés szándékosan tesz különbséget az "AI" és az "MLpAI" között. Ahol az "AI" kifejezés önmagában szerepel, ott vagy a tanulási képesség nélküli AI-ra is érvényesek az állítások, vagy egy empirikus példa működése nem tulajdonítható bizonyossággal a tanulási képességnek.

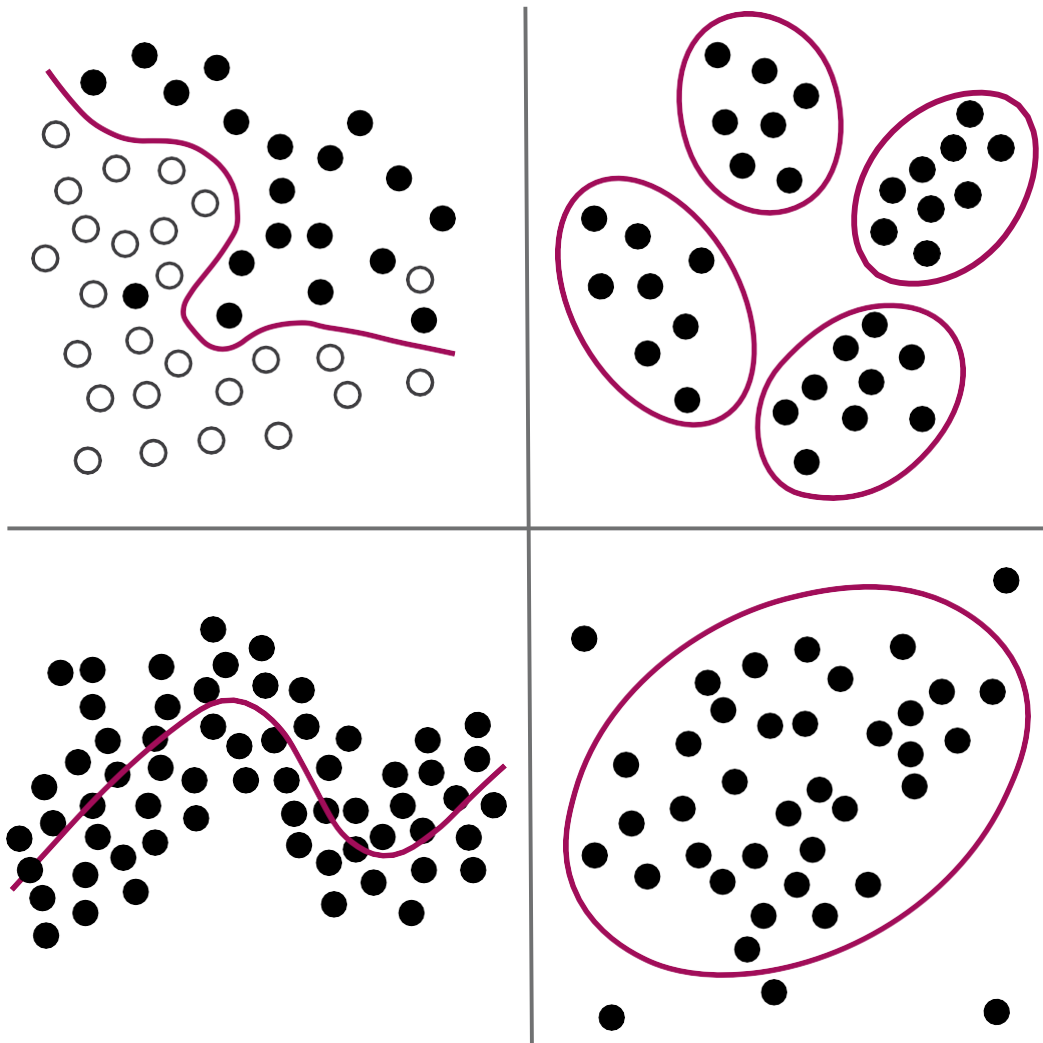
## 2.1 AZ AI FORRADALOM MOTORJA: A GÉPI TANULÁS

Sok emberi vagy állati cselekvés nagyon összetett a számítógépek számára. A beszéd vagy a képek feldolgozása és az azt követő cselekvések koordinálása annyira összetett, hogy az ember maga sem érti eléggé ahhoz, hogy képes legyen átültetni egy programba. Ahhoz, hogy ezeket a rendkívül összetett feladatokat a mesterséges intelligencia elsajátítsa, tanulási képességgel kell felruházni. Az MLpAI felismeri a megoldási mintákat, és átviszi azokat más feladatokra. Erre azért is szükség van, mert a feladatok az időtől, a felhasználótól vagy más paramétereiktől függően változhatnak (Shalev-Shwartz/Ben-David 2014: 21-22).

Az MLpAI két formája különböztethető meg: Az AI tanulhat pusztán előre meghatározott adatok alapján (*felügyelet nélküli*) vagy egy tanár által adott kiegészítő inputok segítségével (*felügyelt vagy megerősített*).

A felügyelet nélküli tanulásról akkor beszélünk, amikor a gépi tanulás ismeretlen adatokból klasztereket hoz létre. A gépi tanulás önállóan azonosítja a hasonlóságokat és különbségeket tartalmazó jellemzőket. Például felismeri, hogy a műholdképek csoportosíthatók a látható szárazföld, tenger vagy felhők szerint. A felügyelt vagy megerősített tanulás esetében ezzel szemben az adatok releváns jellemzői előre meghatározottak. Ez a megközelítés lehetővé teszi az adatokon belüli osztályozást (pl. témák képekre való leképezése), regressziót (pl. környezeti jelenségek előrejelzése) vagy anomáliák (pl. természetellenes kitörések észlelése) felderítését (Russell/Norvig 2010: 694-697).





1. ábra: Osztályozás (balra fent), klaszterképzés (jobbra fent), regresszió (balra lent), anomáliák felismerése (jobbra lent). A ~~szöveg~~ saját ábrázolása.

A gépi tanulás számos módszert tartalmazhat, például *naiv Bayes-féle algoritmusokat*, *támogató vektoros gépi algoritmusokat*<sup>2</sup> vagy a döntési fák különböző változatait. A mesterséges intelligencia kutatásában az utóbbi időben elért előrelépések nagy részét azonban a *mélytanulás* (más néven a *mély neurális hálózatok* módszere) jelenlegi módszere teszi ki. Nagyfokú alkalmazkodóképessége és a rendkívül összetett feladatokban nyújtott teljesítménye azon alapul, hogy a neurális hálózatok másolásával az emberi agyhoz igazodik. Eddig azonban még csak nem is

<sup>2</sup> Ezek a módszerek olyan matematikai módszereken alapulnak, amelyek az objektumok osztályozása érdekében különböző függvények változóit maximalizálják, például az objektumok közötti különbség maximalizálása vagy a legalacsonyabb osztályozási költségek elérése érdekében egy előre meghatározott mértékegység esetében.

megközelítette az emberi agy komplexitását (Hawkins/Blakeslee 2004: 25-27). A konkrét alkalmazásokban, például a képfelismerésben elért teljesítményjavulás azonban már most is jelentős. Míg a hagyományos programok képtelenek egy kép különböző pixeleinek gyűjteményéből jelentést kivonni, addig a *mélytanulási* módszer a folyamatot sok részlépcsésre osztja. Az első szint ("rejtett réteg") összehasonlítja a pixelek fényerejét, hogy azonosítani tudja a kép éleit. A második szint az első szint eredményei alapján keresi a sarkokat és kontúrokat. A harmadik szint pedig a sarkok és kontúrok meghatározott alcsoportjainak megtalálása révén képes azonosítani bizonyos tárgyak teljes részeit (pl. orr, fülek, lábak). Végül a számítógép a felismert különböző objektumok alapján képes értelmezni a képet (Goodfellow 2016: 6).

Ez a folyamat csak azután végezhető el, hogy a számítógép előzőleg megtanulta, hogy a kép mely tárgyakat tartalmazza, például egy autót, egy személyt vagy egy állatot. A tanulási folyamat során a "kimeneti réteg" előre meghatározott, és a szintek feldolgozása fordított sorrendben történik. A gépi tanulás a harmadik szinten észlelt tárgyakat az adott eredményhez, a második szint kontúrjait a tárgyakhoz, az első szint pixeleinek fényerejét pedig a kontúrokhoz társítja.<sup>3</sup> Rendkívül fontos megérteni, hogy az eredmények, amelyekre a *mélytanuló* program jut, csak valószínűségeket tükröznek. Nem dönt egy adott lehetőség mellett és nem indokolja azt, hanem például azt állítja, hogy 9% valószínűséggel autó, 72% valószínűséggel ember vagy 19% valószínűséggel állat. Az, hogy ezeket a kimeneteket hogyan alkalmazza, a programozóra van bízva.

Mivel a gépek csak a változók formalizált összefüggéseit értik meg, de azok tartalmát nem, a gépi tanulást egyes tudósok és kritikusok "automatizált statisztikának" nevezik (Danks 2014: 159f). Maguk a tanulási modellek és a szülőprogram azonban szabályalapú rendszeren alapulnak. Így - a statisztikával ellentétben - mindig tartalmazznak tudatos vagy tudattalan normatív elképzeléseket, amelyeket a programozók a programkódban vagy a tanulási adatok kiválasztásakor hozott döntéseken keresztül vezetnek be (Algorithm Watch 2017: 3).

Meg kell jegyezni, hogy a gépi tanulásban rejlő lehetőségek még messze nem merültek ki. Számos, a fogalom alá tartozó metódust fejlesztettek már ki és fejlesztenek jelenleg is (Farrelly 2016). A komplexitás növelése mellett a jövőbeni mérföldkövek a következő területeken várhatók: a korlátozott bemeneti adathalmazon túli folyamatos tanulás, a megszerzett tudás átvitele más feladatokra, a bemeneti adatok önálló generálása (Morisse 2017) és a más gépek megfigyelésével történő tanulás (Li et al. 2016).

## 2.2 A GÉPI TANULÁS FEJLESZTÉSÉVEL ÉS HASZNÁLATÁVAL KAPCSOLATOS KIHÍVÁSOK

Ha a gépi tanulást valós alkalmazásokban akarjuk használni, három kihívást kell leküzdeni: (1) hibák vagy pontatlanságok a rendszerben vagy a tanulási folyamatban, (2) a döntések nehéz megmagyarázhatósága<sup>4</sup> megfelelő óvintézkedések hiányában, és (3) az eredmények manipulálhatósága szándékosan torzított bemeneti adatokkal.

3 A folyamat részletesebb magyarázata és kiváló szemléltetése megtalálható Zeiler és Fergus (2013) tanulmányában.

4 A megtanult modellek és döntések megértésének és a felhasználók általi megbízhatóságának mértéke.

A gépi tanulás itt bemutatott első kihívása az egyensúly megteremtése két tényező között, nevezetesen egyrészt a tanult modell komplexitása, amely a pontosság növekedésével egyre bonyolultabbá válik, másrészt a tanult modell általánosíthatósága új adatokra (Danks 2014: 155f). E tényezők kölcsönhatása dilemmát vet fel: ha a modell túl komplex lesz, és pontosan leképezi a tréningadatokat - azaz a valóságot leírni hivatott adatokat -, akkor a modell már nem tud általánosítani, így nem alkalmazható új, korábban ismeretlen adatokra (*overmatching*). Ha a modell túl pontatlanul képezi le a képzési adatokat, akkor általánosítható, de a valóságot nem modellezi helyesen (*undermatching*). Ennek eredményeképpen az általánosítás sérülékeny a helytelen következtetésekkel szemben. Káros visszacsatolási hurkok léphetnek fel, amelyekben a gépi tanulás saját logikát tervez, amelyből korrekciós visszacsatolás nélkül nem tud kiszabadulni. Ezeket az általánosítási hibákat a gyakorlati helyzetekben nem mindig észlelik, és így a valós alkalmazásokban elfogadhatók. A mesterséges intelligencia e hibás modelljétől függő szereplőknek a siker érdekében követniük kell a mesterséges intelligencia által felállított szabályokat. Ez viszont torzítja a rendszer lehetséges visszajelzéseit. Ez a probléma például a bűncselekmények előrejelzésénél figyelhető meg: A rendőrök külön járőröznek azokban a körzetekben, amelyeket a rendszer az adott napszakban különösen kritikusnak ítél. A fokozott rendőri jelenlét miatt több bűncselekményt észlelnek. A rendszer által készített előrejelzés megerősítést nyer, ugyanakkor a naplózott incidensek bekerülnek az adatbázisba a további előrejelzésekhez. Ilyen visszacsatolási hurkokat észleltek az arcfelismerés, a tanárok értékelése, valamint a hitel- és biztosítási fedezetek megadása során is (O'Neil 2016).

Második kihívásként különösen figyelemre méltó, hogy a gépi tanulást, különösen a *mély tanulást* számos kutató "fekete dobozként" írta le, a benne rejlő jellegzetességek miatt (Knight 2017; Ribeiro et al. 2016). Tekintettel arra, hogy az autonóm módon tanult szabályok összességükben több ezer vagy akár millió paraméteren alapuló matematikai modellt alkotnak, az összetett modellekben egy döntési útvonal, amely például egy kép egy kis részének kategóriához rendelését foglalja magában, egy ember számára logikailag nem értelmezhető.

Még a fejlesztők sem mindig értik, hogy a gépi tanulás hogyan jut el a döntéséhez. Sok esetben ezt a bizonytalanságot elfogadják, vagy egyszerűen a rendszerre bízzák. A *mélytanulás* módszerének egyre szélesebb körű alkalmazásával azonban ez egyre nagyobb kockázatot jelent. Bár más módszerek jobban megmagyarázhatók, a megmagyarázhatóság negatívan korrelál a pontossággal. Ha például egy program bináris ágakkal rendelkező döntési fa segítségével tanul, az eredmények átláthatóak, de nem olyan pontosak, mint egy *mélytanulási modell* által biztosított több száz absztrakciós szinten több jellemző egyidejű, százalékos súlyozása révén (Gunning 2016: 4). A *mélytanuló* rendszerek esetében a magyarázhatóság növelésére tesznek kísérletet azzal, hogy nemcsak egy adott opció valószínűségét adják meg, hanem azt is meg kell indokolni, hogy miért ez az opció a legvalószínűbb. Például a "a tárgy 93%-os valószínűséggel macska" kitétel helyett a gépi tanuló algoritmusnak meg kell adnia: "A tárgy 93%-os valószínűséggel macska, mert szőre, mancsai és karmai vannak". A mintafelismerés és a minta verbális leírásának összekapcsolására irányuló kutatások azt mutatják, hogy további gépi tanulási alkalmazások lehetővé teszik az ilyen magyarázatokat (Park et al. 2017). Bár egyes megközelítések ígéretesnek tűnnek, a kutatásnak<sup>5</sup> ez az ága még mindig új és korlátozottan elterjedt, legalábbis a nyilvánosan ismertek tekintetében.

5 Ezt a kutatást "megmagyarázható mesterséges intelligenciának" nevezik.

Ezzel szemben az amerikai katonai kutatás saját programterületet szentelt a magyarázható gépi tanuláshoz (Gunning 2016).

A harmadik kihívás a Google, a Facebook és különböző egyetemek kutatócsoportjának meglátásai nyomán vált láthatóvá. A csapat megállapította, hogy a *mélytanulási* módszer képfelismerése váratlanul magas hibaarányt mutatott, és osztályozási hibákat tartalmazott, ha a képhez emberi szemmel nem látható információkat adtak hozzá (Szegedy et al. 2014). Ezek az eredmények két alapvető megállapításhoz vezettek: *Először is*, még a kiemelkedő teljesítőképességgel rendelkező gépi tanulás sem képes megtanulni a képek alapjául szolgáló helyes fogalmat. Ehelyett olyan statisztikai összefüggéseken alapuló modellt konstruál, amely a természetesen előforduló adatokat is magába foglalja, igaz, de alapvető gyengeségei vannak, amikor nagyon természetellenes vagy valószínűtlen adateloszlással szembesül (Goodfellow et al. 2015: 2). *Másodszor*, a gépi tanulás manipulálásának lehetősége a bemeneti adatokon keresztül jelentős biztonsági rést teremt. Megoldás nélkül állandóan fennáll a veszélye annak, hogy a mesterséges intelligencia által elemzett adatokat de- liberálisan megváltoztatják oly módon, hogy a rendszer félreértelmezi azokat. Ráadásul a tudósoknak sikerült megmutatniuk, hogy egy ellentmondásos kép (*ellenpélda*<sup>6</sup>) akkor is megőrzi manipulatív tulajdonságait, ha kinyomtatják és bármilyen kamerával újra lefényképezik. Például egy mosógép képe az algoritmus számára széfnek tűnhet (Kurakin et al. 2017). Egy másik tanulmányban kimutatták, hogy az arcfelismerő rendszerek speciális szemüveggel úgy befolyásolhatók, hogy a lefényképezett embereket az adatbázisban szereplő más emberként azonosítják. A szerzők arra figyelmeztetnek, hogy az ilyen módszereket a jövőben bűnügyi aktákban is felhasználhatják (Sharif et al. 2016). Ez a gépek optikai csalódása óriási kockázatokat hordoz a gépi tanulás gyakorlati alkalmazásai szempontjából. Ha a támadók manipulálni akarnának egy önvezető autót, akkor úgy módosíthatnák az utcatáblákat, hogy a közlekedési táblákat felismerő és értelmező gépi tanulási algoritmus a közlekedési táblákat stop tábla helyett elővásárlási tilalmat jelző táblaként értelmezze. Hogy ez nem csak egy hipotetikus forgatókönyv, azt Nicolas Papernot és munkatársai bizonyították be, akik rekonstruálták az imént leírt forgatókönyvet (Papernot et al. 2017). Ezek az esetek világossá teszik, hogy a manipulációk nemcsak az adatfájl megváltoztatásával, hanem pusztán vizuális változtatásokkal is lehetségesek, hiszen a valós világban lévő tárgyak is manipulálhatók. Ráadásul a mosógép, az arcképek és a közlekedési táblák felismerésében részt vevő tanulási modellek helyesen lettek betanítva, de az alkalmazási adatokat - azaz az elemzendő adatokat - manipulálták.

### 3. MAciNe Tanulás-alapú művészeti intelligencia a fegyverszisztémákban

Az ENSZ égisze alatt jelenleg Genfben folyik a vita a halálos autonóm fegyverrendszerek betiltásáról (Boulain/Verbruggen 2017). A civil társadalomban az autonóm fegyverrendszerek betiltására irányuló felhívásokat neves tudósok és szervezetek támogatják (Sauer 2016; Future of Life Institute 2015; Human Rights Watch 2012). Bár ezek a viták és felhívások is foglalkoznak a mesterséges intelligencia fegyverrendszerekben való alkalmazásával, az autonómia kérdéseire összpontosítanak. Amint az ebben a fejezetben bemutatásra kerül, az autonómia mint a mesterséges intelligencia által vezérelt rendszerek tulajdonsága nem megfelelő megközelítés a fegyverzetellenőrzéshez. Mivel a hagyományos fegyverzetellenőrzés egyéb megközelítései (fizikai jellemzők vagy képességek)

6 Az ellenséges példa olyan bemeneti adatok, amelyeket szándékosan úgy terveztek, hogy a rendszer összeomoljon vagy hibát kövessen el.

képességek, valamint a belső működés) nem nyújtanak megbízható bizonyítékot a mesterséges intelligencia esetében, a fejlesztésre és a telepítésre összpontosító alternatív ellenőrzési módszerek megfontolásra kerülnek. A mesterséges intelligencia fegyverrendszerekben való potenciális felhasználását figyelembe véve olyan destabilizáló következmények kerülnek megvitatásra, mint a folyamatok felgyorsulása, a de-eszkalációs intézkedések hiánya, a technológiai fegyverkezési verseny vagy az ellenőrizetlen proliferáció.

### 3.1 AZ AI MINT A MODERN FEGYVERRENDSZEREK KÖZPONTI ELEME ÉS A TANULÁSI KÉPESSÉG MINT MULTIPLIKÁTOR

A hagyományos számítógépekhez hasonlóan a fegyverrendszerek is megkülönböztetnek hardvert és szoftvert. Ezt a megkülönböztetést a "robotika" és az "AI" kifejezések is tükrözik. A robotika fejlődése javítja a fegyverrendszerek fizikai képességeit és tüzerőjét, de a fizikai törvényekből fakadó korlátok korlátozzák a fejlődési lehetőségeket. Az MLpAI fejlesztése ezzel szemben sokszorosára növeli a fegyverrendszerek szoftverrel kapcsolatos képességeit, mivel csak az MLpAI képes egyre inkább javítani a fegyverrendszerek komplex környezetben való működésének képességét.

Még ha az autonóm fegyverrendszerek általános definíciója még nem is létezik, az alkalmazáspecifikus mesterséges intelligencia már most is megtalálható a fegyverrendszerekben, mint vezérlő vagy támogató egység. A következő példákból nem lehet egyértelműen következtetni a tanulási képességre, de - rendszertől függően - a mesterséges intelligenciát a navigáció, a célfelismerés és -azonosítás, valamint a támadás tervezése és végrehajtása során alkalmazzák. Számos alkalmazás a digitális térben, a légtérben és a statikus védelmi rendszerekben található, mivel a környezet, amelyben az AI-nak ezekben az alkalmazásokban működnie kell, kevésbé összetett, mint a szárazföldi vagy városi hadviselésben. A környezet összetettsége a (ismeretlen) kihívások számának növekedésével növekszik: pl. navigáció egyenetlen talajon, mindenféle akadályok és ismeretlen tárgyakkal való interakciók.

A vadászpilóták támogató rendszereként a mesterséges intelligenciát például arra képzik ki, hogy a pilóta téves megítélésének elkerülése érdekében a célpontokat radar képek alapján felismerje, vagy hogy a pilóta látótávolságán jelentősen túli nagy távolságból tüzelésre vonatkozó döntést hozzon (Keller 2015). A mesterséges intelligencia további feladatokat vesz át a *Taranis* nevű drónban, amelyet a brit gyártó *BAE Systems* jelenleg fejleszt. A *Taranis* a kézi távvezérlés és az automatikus repülési navigáció mellett rendelkezik egy olyan üzemmóddal, amelyben önállóan tervez útvonalat és keres célpontokat, amíg el nem éri a küldetés célját (Stevenson 2016). Az AI *ALPHA*, amelyet még nem telepítettek drónokba, képes átvenni a teljes repülési és harci manőverek irányítását - ami a közelmúltig még az emberi pilóták kizárólagos területe volt: Az *amerikai légierő* tapasztalt ezredese, Gene Lee ellenében végzett szimulációban a rendszer kiemelkedő képességekről tett tanúbizonyságot, egyszerre kerülte ki a rá kilőtt rakétákat, több célpontra is tüzelhetett, összehangolt manőverekben vett részt, és regisztrálta az ellenséges taktikákat, illetve tanult belőlük. Az ezredes az MLpAI-t, amely egy mindössze 35 dollárba kerülő számítógépen, egy *Raspberry Pi-n* futott, "a legagresszívabb, legjobban reagáló, legdisz- namikusabb és leghitelesebb mesterséges intelligenciának nevezte, amit eddig láttam" (Ernest et al. 2016).

Emellett néhány statikus védelmi rendszer - kis lövegtornyok vagy légvédelmi ágyúk - az első mesterséges intelligencia által vezérelt fegyverrendszerek közé tartoznak, mivel nem találkoznak összetett környezeti kihívásokkal. A *Super Aegis II* lövegtornyot úgy tervezték, hogy emberi segítség nélkül, önállóan azonosítsa, célba vegye, kövesse és végül lője a célpontokat. Mivel a vásárlók attól tartottak, hogy a rendszer autonóm üzemmódban hibákat követhet el, az autonómia mértéke már egyedileg állítható (Parkin 2015).

Az amerikai haditengerészet *Phalanx CIWS* légvédelmi ágyúja autonóm módon is képes végrehajtani ezeket a műveleteket, hogy megvédje magát a beérkező rakéták és repülőgépek ellen (US Navy 2017).

Az alkalmazási területek folyamatosan bővülnek, és már fejlesztés alatt állnak autonóm nanodrónok (Daniels 2017), hadihajók (Courtland 2016) és humanoid robotok (Boston Dynamics 2018). A fejlesztés azonban nem kizárólag az egyes rendszerekre összpontosít, hanem az interakció új formájára is. A jövőben a fegyverrendszerek is képesek lesznek rajban működni. Egy raj sok egyedi gépből áll, amelyek önállóan, de akár összehangoltan is képesek cselekedni. Az egységek közötti megszakítás nélküli kommunikáció alapján koordinálják magukat (Ben-Ari/Mondada 2018: 251-252). Ennek többek között az az előnye, hogy nincs központi vezérlőegység, amely meghibásodhat, így az egyes hibák vagy megölések csak kis mértékben befolyásolják a raj teljesítményét. A mesterséges intelligencia így nemcsak az egyes rendszerek technikai fölényét eredményezné, hanem egész harci egységek bevetését is optimalizálná.

Még nem ismert olyan működőképes fegyverrendszer, amely a gépi tanulás képességét használja. A vadászrepülőgépek irányítására szolgáló *ALPHA* AI és az orosz *Kalashnikov* fegyvergyártó cég bejelentése, miszerint tanulási képességet használ (Russia Today 2017), azt mutatja, hogy a tanulásra képes mesterséges intelligenciát beépítik az új fegyverrendszerekbe. Bár a katonai titoktartás miatt a működése nem ismert, a magánszektorbeli kutatási projektek jelzik, hogyan lehetne az MLpAI-t fegyverrendszerekben alkalmazni. A chipgyártó Nvidia egyik kutatócsoportja, amely tulajdonképpen magasan specializált grafikus chipeket fejleszt, betanította az MLpAI-t arra, hogy vezessen egy autót anélkül, hogy bármilyen szabályt diktálna. Bemeneti adatként az MLpAI csak a kormánykerék mozgását és az autó elejéről készült kamerafelvételeket kapta. E korlátozott érzékelés ellenére az MLpAI az ember által irányított utazások során megtanulta a közúti közlekedés szabályait, és ezt követően képes volt önállóan vezetni (Bojarski et al. 2016). Ez nagyban különbözik a hagyományos autonóm vezetési rendszerektől, amelyek előre megkapják a közlekedési szabályok és a jármű viselkedésének értelmezését. A megtanult vezetési stílusnak nyilvánvaló hátránya, hogy az emberi hibákat is elsajátítják. Ugyanakkor nagy előnyei is vannak, mivel a program olyan intuitív szabályokat tanul meg, amelyekről a járművezető nem tud tudatosan, és a nem várt helyzetek kezelésére is felkészül. Ennek megfelelően, analóg módon, a mobil fegyverrendszerek a felügyelt tanulás révén navigációs képességet is igényelhetnek. A célzott felismerés és azonosítás is jelentősen finomítható gépi tanulással, a *mélytanulási* módszer tipikus képességeinek - a tárgyak érzékelésének és osztályozásának - felhasználásával. (Harci) helyzetekben a mesterséges intelligenciának kell döntenie és megfelelően cselekednie. A szükséges következtetési és tervezési készségek ember által felügyelt szimulációk segítségével képezhetők ki. Így a gépi tanulás javíthatja az összes olyan készség teljesítményét, amelyre a mesterséges intelligenciának szüksége van a fegyverrendszerekben.

### 3.2 AZ ELLENŐRZÉST MEGNEHEZÍTŐ JELLEMZŐK ÉS ÚJ MEGKÖZELÍTÉSEK

Ha a mesterséges intelligencia valóban az autonóm fegyverrendszerek központi eleme, és a tanulási képességek tovább fokozzák e rendszerek képességeit anélkül, hogy a hardver adaptálására lenne szükség, akkor megállapítható, hogy a fegyverzetellenőrzésnek végső soron nem az autonómiára, hanem a fegyvereket irányító MLpAI-ra kell összpontosítania. Azonban, mint azt a következőkben bemutatjuk, az MLpAI teljesen új problémákat vet fel a fegyverzetellenőrzés számára.

### 3.2.1 Kicserélhető külsők a beépített hardver-kompatibilitáson keresztül

Amint azt a 3.1. fejezetben tárgyaltuk, a hardver- és szoftverfejlesztési folyamatok külön-külön is vizsgálhatók. Mivel azonban a gyakorlati alkalmazásokban a két szint kölcsönhatását össze kell hangolni, nem lehetséges a hardver általános felcserélhetősége. A szoftvernek képesnek kell lennie a hardverrel való kommunikációra. A sokféle hardverkomponenssel való kompatibilitás növelésének egyik módja a következetes szabványok és az automatikus illesztőprogram-frissítések. Az összetettebb hardverrendszerek esetében a robotikában úgynevezett "middleware"-t használnak. A middleware a hardver és az alkalmazások heterogenitását egy további rétegen keresztül szabályozza. Megkönnyíti az új technológiák integrálását, az érzékelőadatok felhasználását és az alkatrészek cserélhetőségét. A gépi tanulás integrálása tovább növelheti a middleware kompatibilitását azáltal, hogy lehetővé teszi a rendszerhez való ~~data~~ alkalmazkodást (Bennaceur et al. 2013). Ha ilyen módszereket alkalmaznak a fegyverrendszerek fejlesztésénél is, akkor az MLpAI-t komplex kiigazítások nélkül lehetne használni a különböző rendszerekben. Az MLpAI ugyanolyan jól irányíthatna drónt, víz alatti járművet, rakétát vagy más robotizált fegyvereket. Így az új fegyvertechnológiák eleme nem kapcsolható kizárólagosan egy adott fegyverrendszerhez.

A fegyverzetellenőrzés egyik gyakori megközelítése a fegyver hordozórendszerének mennyiségi korlátozása. Ha az MLpAI-t fegyverként vagy annak multiplikátoraként értelmezzük, akkor a hagyományos fegyverzetű drónok, robotok stb. a hordozó rendszereket képviselik. A hordozórendszer korlátozása azonban csupán az MLpAI átvitelét eredményezné egy másik rendszerbe. A navigáció, a cél azonosítása és a cselekvés minden rendszeren hasonló beállítással történhetne, így biztosítva az átvitelhez szükséges kompatibilitást. Így az MLpAI, mint a fegyverrendszerek pusztító központi eleme, nem azonosítható egy kívülről látható rendszeren keresztül, ami így megszünteti a fegyverzetellenőrzés egyik szokásos megközelítését.

### 3.2.2 a külső vízbilis képességek cseréje a szoftverfrissítések és a nyílt szoftver-archiválás révén

A számítógépes programok tipikus jellemzője a frissítések, amelyek célja a biztonsági rések bezárása, funkciók hozzáadása vagy az összetevők módosítása. A szoftveralapú fegyverrendszerek növekvő használatával a frissítésekre itt is szükség van. A magasan fejlett vadászrepülőgépek ezt példázzák: Az *amerikai légierő* frissítette az F-22-es vadászrepülőgép szoftverét, hogy az képes legyen újabb fegyverek kilövésére, a célpontok jobb azonosítására, és így a támadófeladatok szélesebb körének végrehajtására (Osborn 2017). Emellett az MLpAI-t használó fegyverrendszerek további funkciókkal is felszerelhetők a hardver megváltoztatása nélkül. A rugalmasság további növelése érdekében nyílt szoftverarchitektúra vezethető be. Ilyen architektúra megtalálható az okostelefonokban: Az úgynevezett "alkalmazások" olyan alkalmazások, amelyek lehetővé teszik, hogy a rendszer a főprogram módosítása nélkül hozzáadjon, eltávolítson vagy frissítsen komponenseket. A hadiiparban ezt a rendszert már alkalmazzák az F-35 vadászrepülőgépben. Az izraeli hadsereg im- portálja ezt a repülőgépet, és a nyílt architektúrát használva a központi szoftver megváltoztatása nélkül tudja a saját igényeihez igazítani (Adams 2016). Más amerikai védelmi vállalatok is nyílt szoftverarchitektúrát fejlesztenek saját termékeikhez. Ha létrejönne egy csoportos szabvány, az alkalmazások rugalmasan használhatók lennének, függetlenül a fegyverrendszer típusától és kialakításától (Hagen et al. 2012: 6).

A fegyverzetellenőrzés egyik megközelítése a fegyverek képességeinek korlátozása. A *nukleáris kísérletek betiltásáról szóló szerződés*, amint hatályba lépett, megtiltja a polgári vagy katonai célú nukleáris robbantásokat. Ebben az esetben a fegyverzetellenőrzés a nukleáris fegyverek robbanóerejére támaszkodik. Az MLpAI funkciók azonban rugalmasan hozzáadhatók vagy eltávolíthatók a fegyverrendszerekből. A fegyverrendszerek azonosítása és korlátozása nem lehetséges a látható képességek alapján. A kritikus funkciókat még az ellenőrzések során is hozzá lehet adni vagy el lehet távolítani rövid időre.

### 3.2.3 IntranSPARENT Internal működési mód a Complex reverse Engineering-en keresztül és a tanulási mód alkalmazásának lakja

Ha egy számítógépes program működésére vonatkozó megállapodást kell ellenőrizni, a program forráskódját lehet elemezni. Ha csak a kész rendszer áll rendelkezésre, például egy autonóm drón, akkor a forráskódot nem lehet könnyen kinyerni. A programokat jellemzően úgynevezett magas szintű nyelven írják, amelyet általában egy fordítóprogram alakít át gépi nyelvre.<sup>7</sup> Ez a transzformáció megsemmisíti a metainformációkat, ami megnehezíti a folyamat visszafordítását. A forráskód rekonstruálására speciális programok segítségével, úgynevezett reverseengineering segítségével van lehetőség (Eilam 2005). Mindazonáltal a digitálisan vezérelt fegyverrendszerek ellenőrzésének és verifikációjának kezdeti akadályát kizárólag a szoftverek alapvető felépítése jelenti. Ez már a mai fegyverrendszerekre is érvényes, még gépi tanulás nélkül is.

A gépi tanulási alkalmazásokban azonban az átláthatatlanság egy második szintje is megjelenik. Amint azt a 2.2. fejezetben leírtuk, a különböző tanulási módszerek inherens jellemzői határozzák meg az átláthatóságot. Eddig alig vizsgált módszereket kellene hozzáadni a fegyverrendszerhez, hogy igazolja annak képességeit. Az utólagos tisztázástól eltekintve, a képességek előzetes meghatározása lehetetlen mindaddig, amíg a tanult modell nem látható.

A fegyverzetellenőrzés másik közös megközelítése például a nukleáris fegyverek, a gyalogsági aknák és a kazettás bombák működési módja. A nukleáris fegyverek esetében a nukleáris energia robbanáshoz történő felhasználását ellenőrzik. A gyalogsági taposóaknák és a kazettás bombák tiltottak, mivel működési módjuk nem képes különbséget tenni a harcosok és a civilek között. Az MLpAI esetében a működési módot a már ismert két szint teszi átláthatatlanná. Az átláthatóság hiánya kizárja a fegyverzetellenőrzés további megközelítését, mivel a fegyverrendszerekben lévő MLpAI-t nem lehet a működési mód alapján meghatározni és korlátozni.

### 3.2.4 a karok ellenőrzését lehetővé tevő technikai megközelítések

Ha - ahogyan az előző fejezetekben megállapítottuk - az MLpAI léte, hardvere, működési módja és képességei nem nyújtanak megbízható megközelítést, amely lehetővé teszi a fegyverzetellenőrzés végrehajtását - akkor ez a

7 A magasabb szintű nyelvek valós időben is értelmezhetők. A közvetlen értelmezés megkönnyíti a kód tesztelését és módosítását. Ennek ellenére a fordítás még mindig szokásos, mivel növeli a kód hatékonyságát.

a fegyverek fejlesztésére vagy telepítésére vonatkozó ellenőrző intézkedések végrehajtásával kell elérni. A fejlesztést mint megközelítést a megelőző fegyverzetellenőrzés koncepciója tárgyalja. Ebben a megközelítésben a hadászatiilag alkalmazható technológiákat, anyagokat vagy rendszereket a fejlesztési vagy tesztelési szakaszban azonosítják és tiltják vagy szabályozzák (Altmann 2008):

"Konkrétan a megelőző fegyverzetellenőrzés célja a kapcsolódó kutatási és fejlesztési folyamatok korlátozása, felfüggesztése vagy megszüntetése és/vagy a fegyverekben (rendszerekben) való megvalósításukon alapuló katonai lehetőségek tiltása." (Neuneck/Mutz 2000: 109)

Ebben az összefüggésben elképzelhető egy olyan katonai kutatási és fejlesztési nyilvántartás, amely a fegyverkezési kockázatokat korai szakaszban észleli (Müller 2000). Ez a megközelítés azonban a fejlesztési folyamat nagyfokú átláthatóságát igényli, amit aztán más felek sokkal könnyebben másolhatnának, mint a jelenlegi technológiákat. Ráadásul a civil AI-kutatással való átfedés megnehezíti a szándékok egyértelmű tisztázását (Bostrom 2017).

A mesterséges intelligencia fegyverrendszerekben való alkalmazása a rendszer stratégiai és taktikai céljainak és a kapcsolódó felhasználási lehetőségeknek a korlátozásával érhető el (Kahl/Mölling 2005: 350). Az ilyen küldetési célokat és a rendszer akcióit egyfajta fekete dobozban lehetne rögzíteni. A Gubrud és Altmann által javasolt "üvegdoboz" óriási mértékben növelhetné az átláthatóságot:

"Az adatfolyamnak közvetlenül a kiválasztási és bevonulási parancsokat megelőző és azokat is tartalmazó időszelétét lehet a bevonulás elsődleges rekordjaként kijelölni. Ez a rekord az állami félnél maradna, de a rekord "hash"-nak nevezett kriptográfiai kódot egy "üvegdoboz" rögzítené [...] a bevetési parancs kiadásának pillanatára vonatkozó időbélyegzővel együtt. A hash a jegyzési jegyzőkönyv digitális pecsétjeként szolgálna; ha a jegyzőkönyv egyetlen bitjét is megváltoztatnák később, a hash nem egyezne meg". (Gubrud/Altmann 2013: 6)

Jogellenes háborús cselekmények gyanúja esetén az államnak át kellene adnia a dobozban rögzített adatokat egy nemzetközi ellenőrző hatóságnak.

### 3.3 KONCEPCIONÁLIS REFLEXIÓ: STRATÉGIAI INSTABILITÁS ÉS PROLIFERÁCIÓ

Az MLpAI mint ellenőrzési tárgy jelentőségének és akadályainak meghatározása után ez a fejezet a fegyverzet-ellenőrzési elméletet alkalmazza az MLpAI alkalmazásának következményeinek kidolgozására. Az államközi kapcsolatok stabilizálásának célját a katonai eszkaláció és a fegyverkezési verseny megakadályozásával kell elérni. A fegyverrendszerek autonómiájával azonban eltűnnének az emberek lehetőségei arra, hogy fegyverzetellenőrzéssel ériék el a deszkalációt. Ezen túlmenően az MLpAI hozzájárulhat a vertikális proliferációhoz - a további katonai-technológiai fejlődéshez és a meglévő kapacitások fejlesztéséhez - és ezáltal egy új fegyverkezési versenyhez. Az MLpAI-fejlesztés emellett egyaránt jól használható polgári és katonai célokra is.

célokra, és így hozzájárulnak a horizontális proliferációhoz - a katonai-technológiai ismeretek és fegyverrendszerek terjesztéséhez állami és nem állami szereplők között.

### 3.3.1 krízis stabilitás: Az emberi természet dekoncentrálásának lakja

A fejlett érzékelési, tanulási és következtetési képességekkel rendelkező MLpAI alkalmazása a fegyverrendszerekben nagyfokú autonómiát eredményezhet. Ez veszélyezteti a ~~végtel~~ általános fegyverzetpolitikai célját, mivel az ember által alkotott moderáló tényező minimálisra csökken. E tényező lényege, hogy az ember lassítja a dolgokat:

"A modern kommunikáció és az elektronikus adatfeldolgozás ellenére a tisztviselőket még mindig korlátozza a hétköznapi emberi intelligencia, a beszélt nyelv hagyományos sebessége, a szem leolvasó mozdulatai és a válsághelyzetben a felelősség érzelmi kísérelése." (Schelling/Halperin 1961: 27)

Ez alatt a rövid idő alatt az embereknek három de-eszkalációs lehetőségük van:

- A gép jelentésének vagy ajánlásának validálása
- Kommunikáció az ellenféllel tárgyalások vagy magyarázatok keresése céljából
- Az erkölcsi és jogi következmények mérlegelése

A fegyverzetellenőrzés ezekre a lehetőségekre támaszkodik a fegyverrendszerek hadrafoghatóságának késleltetése révén. "Úgy tűnik, hogy sok fegyverzetkorlátozás - ha nem is kifejezetten, de - implicit módon a döntés ütemére irányul" (Schelling/Halperin 1961: 27). A hagyományos fegyverzetkorlátozásban ezt úgy érik el, hogy a rakéták állandó indításra való készenlétét tiltják, megkövetelik, hogy a robbanófejeket a rakétalövedékektől elkülönítve tárolják, vagy megkövetelik, hogy a rakétákkal felszerelt tengeralattjárók a part menti vizeken maradjanak. A hidegháború alatt több esetben a nukleáris eszkalációt az is megakadályozta, hogy az emberek felismerték a technikai téves riasztást (Schlosser 2013). A fegyverzetellenőrzés nemcsak kihasználja ezt az emberi tényezőt, hanem bizalomépítő intézkedésekkel erősíti is. Ezek az intézkedések az emberek közötti bizalomépítést célozzák, többek között különböző típusú információk cseréjével, külföldi megfigyelők jelenlétének engedélyezésével a katonai gyakorlatokon, tiszték és gyakoronokok csereprogramok fenntartásával, vagy forródróttal a válsághelyzetekre (Goldblat 2002: 11).

Ha egy fegyverrendszer autonómiája olyan szinten van, amely nem teszi lehetővé az emberi felügyeletet és beavatkozást, a helyzet automatikus eszkalálódása valószínűbbé válik. A mesterséges intelligencia által okozott inci- dentre példa a 2010. májusi *flash crash* a New York-i tőzsdén, amikor a piaci manipuláció a számítógép által vezérelt nagy volumenű kereskedők által elindított eladási spirál lefelé irányuló spirálját indította el. Ennek következtében a tisztviselők biztonsági intézkedéseket hoztak (CFTC/SEC 2010). Az autonóm fegyverrendszerek hasonló helyzetekbe kerülhetnek, ahol egy hiba vagy véletlen kivételesen gyors eszkalációt okoz. A mesterséges intelligencia váratlan viselkedése esetén az emberi döntéshozóhoz való visszalépési mechanizmus olyan stabilizáló óvintézkedés lenne, amely megakadályozná az erőszakos eszkalációt (Scharre 2016: 38-39). Egy ilyen mechanizmusra azért is szükség van, hogy a bizalomépítő intézkedések ne váljanak feleslegessé:

A kialakult bizalom kevésbé értékes, ha a fegyverrendszerek maguk hozzák meg döntéseiket, és a szemben álló fél emberi megítélése már nem gyakorol befolyást.

De különösen az MLpAI-nak emberi felügyeletre van szüksége mindaddig, amíg a szisztematikus tanulási hibákat és a manipulációra való érzékenységet nem sikerült kiküszöbölni (lásd a 2.2. fejezetet). Ha például az MLpAI-nak különbséget kell tennie harcosok és civilek között, egy hibás vagy manipulált tanulási modell téves osztályozáshoz vezethet. Az ellenséges harcosok úgy védekezhetnek, hogy megváltoztatják a ruházatuk vagy fegyvereik optikai jellemzőit, hogy tévesen osztályozzák őket. Az MLpAI, az autonóm fegyverrendszerek magja, nem rendelkezik a de-eszkaláló emberi természettel, amely megakadályozhatná egy ilyen forgatókönyv bekövetkezését. A helyzetet Altmann és Sauer szavaival lehet összefoglalni:

"A sebesség kétségtelenül taktikai előny a csatatéren, és az emberek lassabbak, mint a gépek. A stratégiai stabilitás azonban elengedhetetlen a túléléshez. Amikor ez veszélybe kerül, az emberi lassúság némi maradéka jó dolog." (Altmann/Sauer 2017: 136)

### **3.3.2 a fegyverkezési rendszer stabilitása és a vertikális proliferálás: a megújult, külső irányítású fegyverkezési dinamika kockázata**

Az MLpAI fejlesztése és egy ország fegyverkezési dinamikája között olyan kölcsönhatás alakulhat ki, amely belülről és kívülről egyaránt irányítható. A belső irányítású fegyverkezési dinamika szempontjából ennek a kölcsönhatásnak a forrása a társadalmon belüli hatalmi viszonyok. A mesterséges intelligencia kutatások a haderőfejlesztés egy jelentős technikai fázisának részét képezik, ami leginkább az Egyesült Államokban látható (Neuneck/ Alwardt 2008). Különösen a demokráciákban támogatják az újrafegyverkezést, hogy a lehető legjobb fegyverrendszereket fejlesszék ki a háborús veszteségek kockázatának minimalizálása érdekében (Shaw, 2005: 79; Schörnig 2008). Ráadásul ezekben az államokban erős hadiipari szereplők is vannak, akik belpolitikai befolyásukat felhasználva támogatják a folyamatos katonai fejlesztéseket (Müller/Schörnig 2006: 106).

A kívülről irányított fegyverkezési dinamika szemlélete úgy látja, hogy a fegyverkezési dinamika eredete a két vagy több állam közötti kapcsolatokban rejlik. Egyik formája szerint az államok intenzív katonai technológiai fejlesztésre törekcsenek, hogy technológiai fölényt szerezzenek más államokkal szemben. Matthew Evangelista (1988) a hidegháborús nukleáris fegyverkezési verseny alapján mutatja be, hogy a fegyverrendszerek technológiai-logikai innovációinak fejlődése nagyon eltérő irányt vett az USA-ban és a Szovjetunióban. Míg az USA-ban az innovációk az erős civil társadalomból eredtek, addig a Szovjetunióban az innováció központosított volt, és az állam reaktívan határozta meg. <sup>Evangelista</sup> modellje hasznos a mesterséges intelligencia kutatás jelenlegi innovációinak megértéséhez: Az alulról felfelé irányuló megközelítésnek megfelelően az amerikai kutatást nagyrészt a magánszektor vállalatai finanszírozzák (Bughin et al 2017: 10). Oroszországban és Kínában ezzel szemben a kormányok támogatják a fejlesztést. Ahogy Vlagyimir Putyin mondja:

"A mesterséges intelligencia a jövő, nemcsak Oroszország, hanem az egész emberiség számára [...]. Aki vezető lesz ezen a téren, az lesz a világ ura". (Lant 2017)

Emellett az orosz hadipari bizottság bejelentette, 2025-ig a haditechnika százalékát 30 robotikára és autonóm rendszerekre cseréli (Association of the United States Army 2017: 1). Kína is közzétette a következő néhány évre 150 milliárd dollárt ígérő terveket, hogy Kínát a mesterséges intelligencia innovációs központjává tegye. Ahogy Xi Jinping 2030. kínai elnök mondja:

"Fel kell gyorsítanunk Kína erős országgá történő kiépítését, amely fejlett gyártási technológiával rendelkezik, és mély integrációt kell megvalósítanunk a reálgazdaság és a fejlett technológiák között, beleértve az internetet, a nagyméretű adatokat és a mesterséges intelligenciát." (Yu/Jing 2017)

A kínai mesterséges intelligencia-kutatás rendkívüli fejlődését tovább kívánják bővíteni, és a hadsereg civil együttműködésre törekszik a mesterséges intelligencia fegyveres erőkké való integrálása érdekében (Kania 2017). Oroszország és Kína beruházási programjai csökkenthetik az USA katonai és technológiai előnyét.

Ha az MLpAI által vezérelt fegyverrendszerek a nukleáris fegyverekhez hasonló - még ha más jellegű is - nagy stratégiai előnyt biztosítanak, és a bejelentett kormányzati beruházások a katonai alkalmazások terén megvalósulnak, akkor legalább a három szereplő, Kína, Oroszország és az USA között új fegyverkezési verseny kezdődhet.

### 3.3.3 dual-use és horizontal proliferation: unkontrollabilis distribution és use

A fegyverzetellenőrzés másik célja a fegyverek (rendszerek) és a haditechnikai ismeretek elterjedésének megfékezése. Ezt a célt különösen veszélyezteti a *kettős felhasználás*<sup>8</sup> problémája, mivel a folyamatban lévő polgári fejlesztések eredményei katonai célokra is felhasználhatók:

"[A]z autonóm fegyverek kódjának kiszervezése nem tűnik kívánatosnak, és nem hallottunk senkit, aki ezt követelte volna. De a mesterséges intelligencia alap kutatása jellemzően nem ilyen módon alkalmazáspecifikus. Sokkal inkább, amennyiben sikerrel jár, olyan algoritmusokat és technikákat szállít, amelyek az alkalmazások nagyon széles körében használhatók." (Bostrom 2017: 137)

Még ha a jövőbeli (fél)ig autonóm fegyverrendszerek forráskódjai nem is nyilvánosak, a nyilvános alap kutatás<sup>9</sup> nem alkalmazáspecifikus, és illegális célokra is felhasználható. Az MLpAI alkalmazások és programozási keretrendszerek sokasága szabadon hozzáférhető.<sup>10</sup> Ezeket a nyílt forráskódú projekteket egyrészt a védelmi ipar fejlesztésében, másrészt a polgári projektek fejlesztésében használják. Minél változatosabbak egy-egy civil MLpAI alkalmazási területei, annál valószínűbb a fegyverrendszerekben való felhasználása. A nukleáris fegyvertechnológia esetében az ismeretek terjesztését a teststop norma korlátozta, mivel a fejlett

8 "A kettős felhasználású termékek - olyan áruk, szoftverek és technológiák, amelyek polgári és katonai célokra egyaránt felhasználhatók és/vagy hozzájárulhatnak a tömegpusztító fegyverek elterjedéséhez - kereskedelmét ellenőrzés alá kell vonni, hogy megelőzzék az ilyen termékek által a nemzetközi biztonságra jelentett kockázatokat." (Európai Bizottság 2017)

9 Számos ismert kutató teszi közzé legújabb eredményeit (forráskód nélkül) a <https://arXiv.org> oldalon.

10 Válogatás szabadon hozzáférhető tanulási architektúrákból: TensorFlow, Torch, Accord.NET.

nukleáris fegyverhez több próbafolyamatra van szükség. A tesztek olyan méretben kell végrehajtani, hogy elkerülhetetlenül megfigyelhető szeizmikus hullámokat, hidroakusztikus jeleket vagy radionuklidokat bocsátanak ki. Az MLpAI esetében a tesztek betiltása nem lenne hatékony, mivel kis léptékű funkcionális tesztek vagy szimulációk - amelyek nem bocsátanak ki észlelhető jeleket - elvégezhetők, majd sok rendszerre méretezhetők. Az alapvető MLpAI fegyverkezésre alkalmas alkalmazásá fejlesztése viszonylag kis akadály, és a fegyverzetellenőrzést a nagy kockázatú technológia elterjedésének megakadályozására távoli céljá teszi.

#### 4. MaChIne Learning-alapú művészeti intelligencia a verifikációs mérésekben

"Bízz, de ellenőrizd" - ez egy gyakran idézett diktum a fegyverzetellenőrzésben, mivel a fegyverzet korlátozásának megállapodása nem mentesíti az államot a többi állam alapvető bizalmatlansága alól. Csak az ellenőrzés - vagyis annak vizsgálata, hogy az államok betartják-e a fegyverzetellenőrzési szerződést - csökkentheti a bizalmatlanságot és erősítheti minden állam biztonságra való törekvését.

Az ellenőrzés három célja *először is* az átláthatóság megteremtése, és ezáltal az államháztartás megsértésének korai felismerése, hogy diplomáciai, katonai vagy gazdasági intézkedéseket lehessen kezdeményezni. E reakciók révén az ellenőrzési intézkedéseknek el kell tántorítaniuk a szerződészegésektől. Az elrettentő funkció mellett ezeknek az intézkedéseknek *másodsorban* a felek közötti bizalmat is ki kell építeniük. Annak megerősítése, hogy minden tagállam tiszteletben tart egy szerződést, bizalmat épít a fegyverzetellenőrzésnek a nemzeti érdekek védelme szempontjából mutatott értékébe (Goldblat 2002: 309). *Harmadszor*, az ellenőrző intézkedések lehetővé teszik a hamisan vádolt államok számára, hogy bizonyítsák a szerződés betartását. Ha ilyen állítás fogalmazódik meg, az ellenőrzési és hírszerzési forrásokból származó bizonyítékokat kell összegyűjteni és értékelni. Ha az állítás igaznak bizonyul, a tagállamok intézkedéseket hozhatnak a rendszeren belül, vagy a jogsértést az ENSZ Biztonsági Tanácsa elé terjeszthetik (Müller/Schörning 2006: 150-153). Ha a diplomáciai intézkedések nem vezetnek a szerződészegés orvoslásához, az embargók, szankciók vagy katonai erővel való fenyegetés következhet.

A lehetséges intézkedések egyértelművé teszik, hogy az összegyűjtött bizonyítékok érvényessége és minősége alapvető fontosságú az értékelés szempontjából. A helyszíni ellenőrzések (műholdak, repülőgépek, radar- és egyéb érzékelőrendszerek) és a helyszíni ellenőrzések lehetővé teszik a tagállamok fegyveres erőinek, fegyvereinek vagy tevékenységeinek ellenőrzését. Az ellenőrzések "történhetnek szisztematikusan - folyamatosan vagy időszakonként - vagy ad hoc módon, az ellenőrző szerv döntése szerint, vagy kihívás esetén, konkrét kérés eredményeként" (Goldblat 2002: 310). Az ellenőrzési folyamat magában foglalja az adatok gyűjtését, feldolgozását és elemzését, hogy használható információkat nyerjenek. Az MLpAI felhasználható az érvényesség és a minőség növelésére vagy az emberi elemzők hatékonyságának javítására e három lépés egyikében vagy mindháromban.

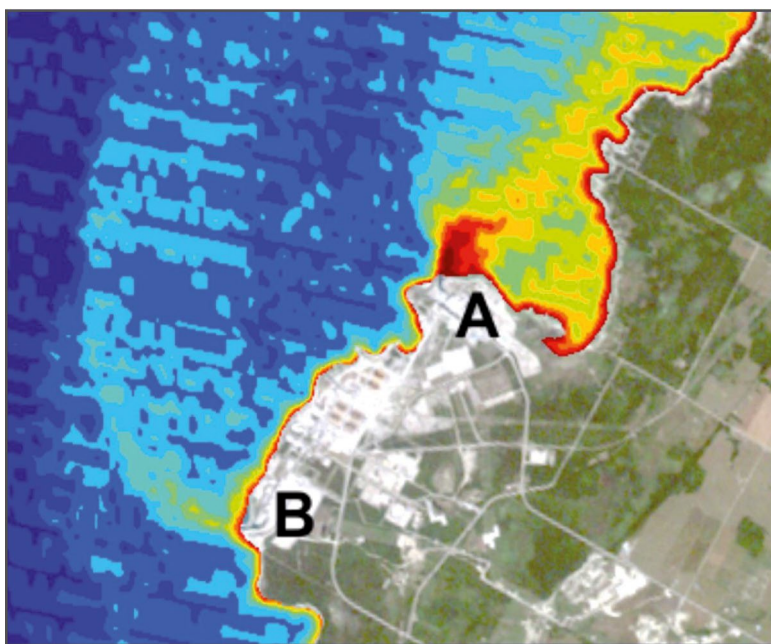
##### 4.1 ALKALMAZÁSOK ÉS MŰSZAKI LEHETŐSÉGEK

Az MLpAI-ban rejlő lehetőségek az adatok gyűjtésére, feldolgozására és elemzésére az ellenőrzés érdekében már ma is nyilvánvalóak az alkalmazásokból vagy a kutatásokból. Különböző adatforrások alkalmasak kiindulópontként ezen esetek elemzéséhez: Műholdfelvételek az űrből, földi ellenőrzések vagy érzékelők globális hálózatai.

#### 4.1.1 SpaCe-ból: automatizált távérzékelés állami képekkel

A műholdképek elemzése például fontos szerepet játszik a nukleáris fegyverek és az energia elterjedésének megakadályozására, illetve a rövid és közepes hatótávolságú rakéták korlátozására irányuló fegyverzetellenőrzési rendszerekben. A műholdképekből nyert információkat - a folyamatot "távérzékelésnek" is nevezik - a szerződési feltételek ellenőrzésére használják (Patton et al Niemeyer/Ruthowski2016, 2016). Ezen elemzések érvényességének és bizonyító erejének növelése érdekében az MLpAI felhasználható a képeken megjelenő objektumok és időbeli változások azonosítására.

A légi felvételeket intenzíven használja például a *Nemzetközi Atomenergia- szervezet* (IAEO), amely a *nukleáris fegyverek elterjedésének megakadályozásáról szóló szerződés* betartását ellenőrzi. A NAÚ a nukleáris létesítmények ellenőrzéséhez a műholdképek által biztosított térbeli, időbeli és multispektrális dimenziókat használja. Az elemzések célja a nagy dúsítású urán vagy plutónium előállítására szolgáló, be nem jelentett létesítmények felderítése, a nehézfémek felhasználásának és feldolgozásának nyomon követése, valamint a helyszíni ellenőrzések szempontjából érdekes objektumok azonosítása. Ehhez olyan mutatókat használnak, mint a hőmérsékleti sugárzás vagy az épületek optikai változásai, hogy nyomon kövessék egy létesítmény bővülését vagy használatának intenzitását (Truong et al. 2005; Johnson et al. 2014).



2. ábra: A kanadai Bruce atomerőmű termográfiai felvételén látható, hogy a B blokkal ellentétben az A blokk üzemel. (Truong et al. 2005: 3)

Az egyik konkrét kihívás például az urániummalmok azonosítása és ellenőrzése, amelyek a természetes uránból a nukleáris üzemanyagciklusban az úgynevezett *sárga tortát* állítják elő, amely az atomerőművekben vagy nukleáris fegyverekben használt urán előanyaga. Mivel ezek a műholdfelvételeken nagyon hasonlítanak a teljesen ártalmatlan rézművekhez, a kutatók által kifejlesztett mesterséges intelligencia az egyes épületek jellemzőit és a komplexum méretét használja a létesítmények osztályozásához (Sundaresan et al. 2017). Az AI az osztályozást egy ember által előre meghatározott döntési fa segítségével végzi el. Bár a NAÜ a múltban mindig a technológiai fejlődéshez igazította tevékenységét, jelenleg nincs bizonyíték arra, hogy a légi felvételek gyakorlati elemzése során gépi tanulást alkalmazna.<sup>11</sup> A 2018-as aktualizált kutatási tervben azonban továbbra is a nem kifejlesztett innovációként szerepel, de nem kiemelt prioritásként (International Atomic Energy Agency 2018: 15).

A kereskedelmi beszállítók már kínálnak használható alkalmazásokat: Az ENVI térinformatikai rendszer olyan modult működtet, amely lehetővé teszi a kutatók és elemzők számára, hogy betöltsék a keresett objektum mintaképét, például a vegyi anyagok tartályait, mint képzési adatokat, majd a program keressen légi felvételeket. A várostervezésben, a természetvédelemben és az erdészetben való alkalmazás mellett a gyártó kifejezetten hirdeti katonai járművek, leszállási zónák vagy épületek megtalálását vagy lokalizálását (Harris Cooperation 2017). Más szállítók<sup>12</sup> is átvették a kereskedelmi műholdképek mesterséges intelligencia alapú kiértékelésének üzleti modelljét. *“Öt évvel ezelőtt erre nem lettünk volna képesek”* - mondja Pavel Machalek, a SpaceKnow vezérigazgatója (Dillow 2016). És a lehetőségek még mindig nagyok, mert a számítási teljesítmény, a gépi tanulás és a műholdképek kombinációjának fejlődése még csak most kezdődik.

Az amerikai korlátozások miatt a nem állami szereplők számára csak viszonylag durva szemcséjű, pixelenként 30cm-es40 felbontású műholdfelvételek állnak rendelkezésre (Shalal 2014), míg az amerikai kéműholdak felbontása cm-es15 (Krebs 2017). Ennek ellenére az MLpAI számára óriási lehetőségek rejlenek e képek felhasználásában, mivel az MLpAI képes kompenzálni a rosszabb képfelbontásból eredő hátrányokat. A kézi- és könnyűfegyverek mellett a kereskedelmi gépi tanulás a nagyméretű katonai felszereléseket is képes azonosítani az alacsony felbontású kereskedelmi műholdképeken.

A tárgy méretétől függően az MLpAI képes a félreérthetetlen jellemzők felismerésére és a tárgy több képen keresztül történő követésére is. A nukleáris, vegyi vagy biológiai üzemek feltűnő változásai automatikusan elemezhetők, és az anomáliákról tájékoztathatók az ellenőrzési rendszerek elemzői. Ez a kiemelkedő anomália- és mintafelismerő funkció érvényesebbé teheti az MLpAI használatát a műholdas alapú ellenőrzésben, és ezáltal erősítheti azt mint eszközt.

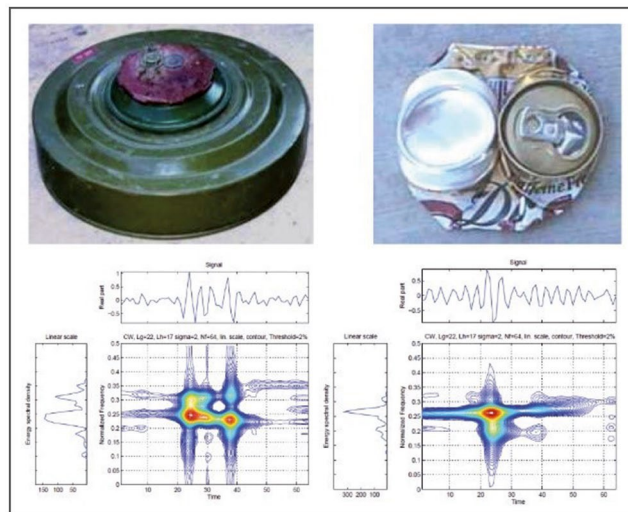
11 Az újonnan kifejlesztett NAÜ térinformatikai kiaknázási rendszeren belül a kereskedelmi szolgáltatók elemzési eszközeinek is rendelkezésre kell állniuk (Balter 2014: 6). A gépi tanulást esetleg ott is lehetne használni.

12 A DigitalGlobe műholdas szolgáltató egy programozási keretrendszert kínál az objektumok megtalálására: <http://deepcore.io/>. A SpaceKnow (<https://spaceknow.com/>), az Orbital Insight (<https://orbitalinsight.com/>) és a Descartes Labs (<https://www.descarteslabs.com/>) start-up cégek online platformokat kínálnak tárgy- és mintafelismerésre. Egy egyszerűsített funkció kipróbálható a <https://search.descarteslabs.com/> aloldalon.

#### 4.1.2 a terepen: fegyverkereskedelem és a fel nem robbant robbanószerkezetek polgári tulajdonba vétele

A hagyományos fegyverek kereskedelmét és így többek között a *fegyverkereskedelmi szerződés* betartását is nyomon lehetne követni az MLpAI segítségével. A szerződés arra kötelezi az államokat, hogy nyomon kövessék a fegyverexportot a rendeltetési helyig, és megakadályozzák az emberi jogokat vagy a nemzetközi humanitárius jogot megsértő államokba történő esetleges szállítást. Ahhoz, hogy ezt a tranzitárúk esetében is biztosítani lehessen, ellenőrizni kell a kikötőkben naponta be-, ki- és átrakodó számtalan teherkonténert (Holtom/Bromley 2011). A kereskedelem hatalmas volumene miatt az emberi ellenőrzés legfeljebb az ismert kockázati tényezők - például a származási hely, a rendeltetési hely és a bejelentett tartalom - alapján történő mintavételre terjedhet ki. Egy MLpAI azonban képes lenne röntgensugarak segítségével azonosítani a számtalan lezárt konténerben lévő fegyvereket, amint azt a University College London kutatócsoportja már be is mutatta (Jaccard et al. 2016). Míg az emberek képesek lennének értelmezni a szenzorok által készített tartalomképeket, addig a gépi tanulás sokkal gyorsabban képes az adatok rögzítésére, az adatokban lévő minták felismerésére és a tartalom típusának valószínűségi becslésére. A tanulási modell képzési fázisában az MLpAI-nak számos különböző forgatókönyvvel, valamint a helyes válasszal - pl. "robbanófej" vagy "nincs robbanófej" - kellene szembesülnie.

A második alkalmazási példa a humanitárius fegyverzetellenőrzésből származik: A gyalogsági aknáknak és a kazettás bombáknak a konfliktus befejezése után is a földben maradnak, veszélyeztetve a civil lakosságot. A humanitárius fegyverzetellenőrzési szerződések, az *ottawai egyezmény* és a *kazettás bombákról szóló egyezmény* egyszerre írják elő általános tilalmat és a már meglévő fegyverek eltávolításának kötelezettségét a konfliktusban részt vevő felek számára. Ez a bonyolult folyamat képzett csapatokat igényel, amelyek szisztematikusan bemérik az érintett területeket, fémdetektorokkal fel nem robbantott aknákat detektálnak, majd ellenőrzött robbanásokkal felrobbantják azokat. A tárgyak felkutatásának nagy kihívására rendkívül hatékony módszereket fejlesztettek és fejlesztenek ki. Egy rendkívül ígéretes új megközelítés a földi radart és a gépi tanulást ötvözi. Amint az ábrán látható, egy gyalogsági akna és egy lapos italosdoboz radarképét vizuálisan nem könnyű megkülönböztetni. Ennek érdekében különböző fejlesztőcsoportok ilyen bemeneti adatok felhasználásával képeztek ki egy MLpAI-t (Núñez-Nieto et al. 2014; Seiffert et al. 2013). Ez a módszer lehetővé teszi, hogy a rendszert egy terepjáró járműre szereljük, és az autó előtt néhány méterrel letapogassuk a terepet. A cél a balesetek vagy téves riasztások arányának csökkentése más nyomkövető módszerekhez képest.



3. ábra: Egy gyalogsági akna és egy italosdoboz érzékelési adatainak összehasonlítása (Sun/Li 2005: 3).

Az ebben a fejezetben szereplő két esetben közös, hogy az MLpAI a környezetből azonnali szenzoradatokat kap a minták azonosításához. A fegyverek vagy gyalogsági aknák felderítésében az érdekesség a többi objektumtól való megkülönböztetésben rejlik a hiányos vagy strukturálatlan adatok alapján. Mivel ezeket az eseteket aktívan tesztelik, a megvalósítás és az alkalmazás valószínűleg megmutatja, hogy az MLpAI hatékonyabb, mint a hagyományos rendszerek. Ha ezeket a fejlesztéseket általánosítják és extrapolálják, akkor jelentős lehetőség nyílik az anyagi átláthatóságra: A fegyverzetellenőrzési megállapodásokat sértő tevékenységeket vagy tárgyakat még a legnagyobb gondosság mellett is csak részben lehet a felfedezés ellen védeni. Az árnyékoláson áthatoló kis mennyiségű nem kívánt információt az ember vagy a statisztikai modellek egyáltalán nem vagy túl lassan tudják értelmezni, de az MLpAI ezen a ponton az átláthatóság új szintjét érheti el.

#### 4.1.3 az érzékelők széles skálája lehetővé teszi a "mérés és a természet megismerését".

Az elektro-optikai, elektromágneses, akusztikai és geofizikai adatok, valamint a nukleáris, biológiai és kémiai nyomelemek feldolgozását a hírszerzésben *mérési és szignatúra-felderítésnek* (MASINT) nevezik (Aid 2014: 120-122). A MASINT alap gondolata, hogy a vizsgált objektum a különböző információhordozókon (sugárzás, hang, anyagok stb.) keresztül következetesen azonosít egy egyedi szignatúrát. Ennek az információnak az elemzése felhasználható az ellenőrzési intézkedésekhez.

A Bochum Verification Project keretében egy olyan ellenőrzési módszert dolgoztak ki, amely akusztikus és szeizmikus szenzoradatokat használ. A projektben kifejlesztett mérőműszereket akár 200 méteres időközönként helyezték el, hogy egy olyan érzékelősorral rendelkezzenek, amely az utakon kívül is érzékeli a járműveket. A mérőeszközök képesek voltak az áthaladó katonai járműveket különböző járműkategóriákba sorolni az akusztikai és

szeizmikus jelek (Hochmuth 2003). Az ígéretesebb akusztikus felismerés a motorzaj alapján egyszerre öt láncos és kerekes járművet tudott megkülönböztetni (Altmann et al. 2002). A 69% és 98% közötti változó sikerességi arányt az MLpAI segítségével magas szinten lehetett stabilizálni, és javult a reziliencia a zavaró zajokkal, például egy záporosóvel szemben. Például ilyen mérőműszerek sora szabályozhatná egy adott járműtípus mennyiségi jellemzőit. Ezzel szemben az egyes járművek azonosítása nehéznek bizonyulhat. Bár a motorzaj járművenként kissé változik, ez a kopás vagy javítás miatt változhat, ami lehetetlenné teszi a felismerést.

Egy másik ígéretes alkalmazási területen az MLpAI-t az *Átfogó Atomcsend-szerződés* (CTBT) betartásának ellenőrzésére is fel lehet használni. A meglévő *nemzetközi megfigyelőrendszer* (IMS) a rendkívül nagy mennyiségű érzékelőadat hagyományos feldolgozását példázza. Az IMS a mérőállomások világméretű hálózatában rögzített adatok rögzítésével képes a nukleáris robbanások diagnosztizálására. A szeizmográfok, a hidroakusztikus érzékelők, az infravörös hangjelző állomások és a radionuklid-detektorok nagy mennyiségű adatot generálnak. A bécsi központi adatfeldolgozás elemzi és redukálja a nyers szenzoradatokat, hogy megkülönböztesse a jeleket az állandó zajtól, és "nukleáris robbanás" vagy "nem nukleáris robbanás" kategóriába sorolja őket (Russell et al. 2010: 32). Ezt a folyamatot még nem lehet teljesen automatikusan elvégezni. Az elemzőknek még mindig fáradságosan át kell dolgozniuk az automatikus rendszerek eredményeit, mivel az automatikus feldolgozás hajlamos az erős zaj, a helytelen osztályozás vagy a téves asszociációk miatti hibákra. Az eredmények további optimalizálása érdekében a különböző projektcsoportok már 2009-ben, a bécsi *Nemzetközi Tudományos Tanulmányok* (ISS) konferencián új módszereket javasoltak. A projektcsoportok egyöntetűen egyetértettek a problémában: az IMS érzékelői által rögzített adatok túlságosan összetettek ahhoz, hogy a lineáris megkülönböztetésen alapuló hagyományos statisztikai elemzés - a csoportok közötti határokat meghatározó lineáris függvény - elvégezhető legyen. Mind a négy projektcsoport azt teszteli, hogy az MLpAI képes-e nemlineáris megkülönböztetést végezni. A projektcsoportok az elmúlt tíz év osztályozott adatait már felhasználták "alapigazságként" (a gépi tanulás algoritmusának képzési készlete). A prototípusok szeizmikus adatokat (Kleiner et al. 2009), hidroakusztikus adatokat (Tuma/ Igel 2009), infrahang adatokat (Procopio et al. 2009) és radionuklid adatokat (Stocki et al. 2010) elemeztek. Valamennyi prototípus képes volt megkönnyíteni az elemzők feladatát és pontosabb automatizált értékeléseket végezni. Mindazonáltal a kutatók arra a következtetésre jutottak, hogy a rendszerek még nem működőképeseek, és további finomításra szorulnak. Összességében ezek a kutatási projektek és a továbbfejlesztett továbbfejlesztések (Arora et al. 2013) bizonyították, hogy az MLpAI-ban nagy lehetőségek rejlenek az IMS-adatokon alapuló fegyverzetellenőrzésben történő ellenőrzésben.

Az érzékelők által lefordított nagy mennyiségű fizikai adat elemzése az MLpAI segítségével jelentősen javítható az emberi elemzőkhöz képest. Különösen a nukleáris kísérletek betiltásának ellenőrzése fog rövid távon óriási hasznot húzni az MLpAI-ból, mivel az adatokat az érzékelők már strukturálták, és így könnyebben elemezhetőek. Az MLpAI sokkal vonzóbbá teszi ezt a fajta ellenőrzést, és azt eredményezheti, hogy más fegyvereket is hasonló módon fognak ellenőrizni.

#### 4.2 KONCEPCIONÁLIS REFLEXIÓ: A MEGLÉVŐ MÓDSZEREK JAVÍTÁSA

Az előző fejezetben bemutatott esetek azt mutatják, hogy az MLpAI képes a fegyverzetellenőrzés megerősítésére, különösen az ellenőrzési intézkedések terén. A kelet-nyugati konfliktus vége óta az ellenőrzési intézkedések jelentősége stagnál az új típusú fegyvertechnológiák és az elszigetelt fegyverzet- túllépések miatt (Pilat 2002: 85-87). Az MLpAI-nak meglenne a lehetősége arra, hogy ezeket az intézkedéseket ismét relevánsabbá tegye, feltéve, hogy az államok megfelelő politikai akaratot gyűjtenek az erőfeszítésekhez.

Az ellenőrzési intézkedések technikai továbbfejlesztése, amennyiben megvalósul, nagyobb átláthatóságot tenne lehetővé az ellenőrzési tételek nyomon követése során. A jobb ellenőrzés megvalósítása erősítené a szerződés titkos megsértésének megakadályozására irányuló célt is. Az ellenőrzési intézkedések két másik célkitűzése - a bizalomépítés és a szerződés betartásának bizonyítása - szintén profitálhatna az MLpAI alkalmazásából eredő nagyobb átláthatóságból.

Amint az előző fejezetekben bemutattuk, az MLpAI használatát már tesztelték megvalósíthatósági tanulmányokban, és az ellenőrzési intézkedésekben hatékonyak bizonyultak. Mindazonáltal a gépi tanulási algoritmusok még mindig jelentős kihívásokkal küzdenek. A bemeneti adatok formájától függően szükségük lehet a vizuális és egyéb érzékszervi benyomások "érzékelésére" és értelmezésére. A digitális adatokban mintákat kell felismerniük, és a meglévő ismeretek felhasználásával kell elvégezniük az elemzést. Az érzékelés, a mintafelismerés és a tudás generálása gépi tanulással fejleszthető, mivel az elemzőprogramra vonatkozó emberi korlátozások nem érik el a szükséges komplexitást. A gépi tanulás alkalmazkodóképessége lehetővé teszi, hogy a programok a kontrollobjektumokkal párhuzamosan fejlődjenek, amikor az elemzett objektumok jellemzői idővel változnak. Az MLpAI képes az ellenőrzési tevékenységek rendkívül összetett és változó környezetéből tudást desztillálni, növelve a fegyverekkel kapcsolatos átláthatóságot. Ezenkívül az MLpAI nagyobb skálázhatóságot tesz lehetővé - az elvégzett elemzések számának jelentős növekedését. Ahogy Robert Cardillo, a Nemzeti Térinformatikai Hírszerző Ügynökség igazgatója fogalmazott:

"Ha megkísérelnénk manuálisan hasznosítani a kereskedelmi műholdképeket, amelyekkel várhatóan a következő 20 évben rendelkezni fogunk, nyolcmillió képelemzőre lenne szükségünk." (Cardillo 2017)

Ugyanakkor az MLpAI más előnyöket is kínál az emberi elemzőkkel szemben:

"A technológiának vannak előnyei az emberi ellenőrökkel szemben. Folyamatosan és állandó megfigyelési szinten tud működni. Az adatai nem összehasonlíthatóak. A szerződés szempontjából releváns információk felderítésére korlátozódhat, míg más típusú információkat figyelmen kívül hagyhat." (UNIDIR/VERTIC 2003: 27)

Az azonban, hogy az MLpAI-t a hagyományos fegyverzetellenőrzésben hivatalos ellenőrzési eszközként használják-e, politikai döntés, amelyet a tagállamoknak kell meghozniuk. Politikai okokból a technikai ellenőrzési lehetőségeket már eddig is mesterségesen korlátozták:

"Az INF-szerződés [...] lehetővé tette, hogy a rakétatartályokról röntgenfelvételt készítsenek a benne lévő rakéta típusának meghatározására, de a gépezetet egy bizonyos felbontásra állították be, hogy ne lehessen érzékeny tervezési információkhoz jutni."  
(UNIDIR/VERTIC 2003: 27)

A *Nyitott Égbolt Szerződésben* - amely szerződés célja nem a fegyverek korlátozása, hanem a bizalom kiépítése - a technikai intézkedéseket szándékosan alacsony szintre korlátozták: A szerződés által engedélyezett átrepülések során legfeljebb 30 centiméteres felbontású légi felvételek készíthetők, noha a kereskedelmi műholdfelvételek már ugyanilyen vagy jobb felbontást biztosítanak. Ráadásul az analóg kamerákról a digitálisra való modernizálás során ügyeltek arra, hogy a szerződésben meghatározott maximális felbontást ne lépjék túl (Britting/Spitzer 2005). A hagyományos fegyverzetellenőrzési szerződésekben a technológiai fejlődés nem áll azon nemzetek érdekében, amelyek már most is előnyben vannak az információgyűjtés terén. Az MLpAI alkalmazása - ahogyan az már más hagyományos ellenőrzési intézkedésekkel is megtörtént - politikai akarattal vagy a katonai titoktartás érdekében megakadályozható vagy korlátozható.

A humanitárius fegyverzetellenőrzésben ezzel szemben a nemzetállamok fejlődésgátló érdekei valószínűleg nem várhatóak. A gyalogsági aknák és kazettás bombák felderítésére irányuló kísérleti projektek (lásd a 4.1.2. fejezetet) azt mutatják, hogy az ilyen megkülönböztetésmentes fegyverek gyors eltávolítása milyen hatalmas előnyökkel jár. Ezt az elvet a kézfegyverek nemzetközi kereskedelmét ellenőrző szereplők is alkalmazhatnák. Az *itrace* projekt például helyszíni vizsgálatokon keresztül követi nyomon az ilyen fegyverek forgalmazását. A nyomozókat támogathatná az MLpAI, mivel képes az adatokban viselkedési mintákat vagy összetéveszthetetlen vizuális jellemzőket azonosítani, és így az egyes fegyvereket konkrét fegyverleltárakhoz kapcsolni.

## 5. A MAciNe Learning-alapú művészeti intelligencia része a problémának és része a szolúciónak is

Ebben a jelentésben világossá vált, hogy a hagyományos fegyverek, a tömegpusztító fegyverek és a modern fegyverrendszerek ellenőrzése terén az MLpAI egyszerre lehet a probléma és a megoldás része. Mint ellenőrzési intézkedés, az MLpAI természeténél fogva számos, minőségi vagy mennyiségi korlátozások bevezetésére irányuló megközelítést elkerül. Így növeli az általános katonai átláthatóság vagy a bizalomépítő intézkedések elérésére irányuló alternatív módszerek jelentőségét. Pontosan ezekben a módszerekben az MLpAI ismét használható ellenőrző eszközként. A pontos és átfogó információfeldolgozás lehetővé teszi, hogy az MLpAI nagyobb átláthatóságot teremtsen, ellenőrizze a szerződések betartását és erősítse a felek közötti bizalmat.

A fegyverzetellenőrzési elmélet szempontjából az MLpAI fegyverrendszerekben és ellenőrzési intézkedésekben való alkalmazása törékeny stratégiai stabilitást eredményez. Az MLpAI fegyverrendszerekben való fokozott alkalmazása veszélyeztetheti a stratégiai stabilitást azáltal, hogy minimalizálja a de-eszkáláló emberi természetet és elősegíti a techno- logikai fegyverkezési versenyt. Az alapvető MLpAI-k ellenőrzetlen elterjedése csökkenti az MLpAI fegyverrendszerekben való alkalmazásának akadályait. Mind ezek az elméleti megfontolások, mind pedig az a felismerés, hogy az MLpAI a jövőbeli autonóm fegyverrendszerek központi eleme lesz, a fegyverzetellenőrzéssel kényszerítik ki a technológiai korlátozásokat. Az MLpAI esetében azonban a külső megjelenés, a külsőleg felismerhető funkciók vagy a belső működési elvek alapján történő minőségi vagy mennyiségi ellenőrzés nem lehetséges.

A hagyományos megközelítések ezen a ponton elérik határaikat, és az egyetlen fennmaradó lehetőség az MLpAI fejlesztése vagy telepítése során történő ellenőrzés és korlátozás. Ha a telepítés "üvegdobozon" keresztüli ellenőrzése vagy a fejlesztési folyamat megelőző ellenőrzése nem végrehajtható, akkor magasabb szinten lehet minőségi megoldásokat keresni. Az általános katonai képességeket és stratégiákat átláthatóbban lehetne bemutatni, és fokozni lehetne a bizalomépítő intézkedéseket (Schörnig 2015).

Az MLpAI alkalmazása az ellenőrzési intézkedésekben javíthatja a stratégiai stabilitást, mivel az MLpAI várhatóan lehetővé teszi, hogy a műszaki felügyeleti eszközök sokkal pontosabb és átfogóbb információfeldolgozást biztosítsanak. Ez a lehetőség azonban nem akadálymentes: A gépi tanulás jelenlegi fejlettségi szintjén javítani kell az átláthatóságot és a külső manipulációval szembeni védelmet ahhoz, hogy érvényes ellenőrzési módszerek lehessenek tekinteni. Ezeket a technikai követelményeket teljesíteni kell ahhoz, hogy a módszerbe és így az államok között is bizalmat lehessen építeni. Az MLpAI-ban mint ellenőrzési módszerben rejlő lehetőségek számos prototípusban és korai alkalmazásban láthatók. Ezek azt mutatják, hogy a gépi tanulás képessége felhasználható optikai, termikus és topo- grafikus műholdképek, helyszíni ellenőrzések szenzoradatai, polgári fegyverzeti repülések és mérőállomások hálózatából származó strukturálatlan adatok elemzése során.

Az, hogy az itt elemzett két jelenségnek lesz-e jelentős hatása a fegyverzetellenőrzésre, attól függ, hogy a mesterséges intelligencia milyen környezetben fog működni. A környezet típusokra való felosztása (Russell/ Norvig 2010: 46) azt mutatja, hogy a valós világbeli cselekvéseknél sokkal több kiszámíthatatlan környezeti tényezőt kell figyelembe venni, mint ami a digitális környezetben jelen van. A mesterséges intelligencia kutatásának még több évre van szüksége ahhoz, hogy a fegyverrendszerek képesek legyenek autonóm cselekvésre a valós világban. A gépi tanulás azonban már most is használható az ellenőrzésre, mivel az elemzendő adatok már digitális formában rendelkezésre állnak, vagy előre meghatározott minta szerint értelmezhetők, anélkül, hogy a környezet hirtelen változásaira be kellene őket tanítani. Az időben történő bevezetés kritikus fontosságú, mivel az MLpAI már azelőtt új szintre emelheti a fegyverzetellenőrzést, mielőtt az MLpAI által vezérelt fegyverrendszerek új kihívásával kell szembenéznie.

- Adams, Eric 2016: F-35-öst (legalábbis egyelőre), <https://www.wired.com/2016/05/israel-can-can-customize-americas-f-35-least-now/>; 04.12.2017.
- Aid, Matthew M. 2014: In: Measurement and Signature Intelligence: Dover, Robert/Goodman, Michael S./Hillebrand, Claudia (szerk.): A mérés és a méréelmélet: Dover, Robert/Goodman, Michael S./Hillebrand, Claudia (szerk.): Routledge Companion to Intelligence Studies, London, 114-122.
- Algorithm Watch 2017: Antworten auf den Fragenkatalog für das Fachgespräch zum Thema "Künstliche Intelligenz" des Ausschusses Digitale Agenda vom 22.03.2017.
- Altmann, Jürgen 2008: Altmann Altmann: Präventive Rüstungskontrolle, in: Becker, Una/Müller, Harald (szerk.): Rüstungskontrolle im Jahrhundert 21., Berlin, 105-125.
- Altmann, Jürgen/Linev, Sergey/Weiß, Axel 2002: A katonai járművek akusztikus-szeizmikus észlelése és osztályozása - eszközök fejlesztése a leszerelés és a békefenntartás érdekében, in: Alkalmazott akusztika 63: 1085-110710,.
- Altmann, Jürgen/Sauer, Frank 2017: Bauer, Bauer, Bauer: Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer, Bauer: Survival 59: 117-142.5,
- Arbatov, Alekszej 2015: Alexejov: Egy észrevétlen válság. The End of History for Nuclear Arms Control?, <https://carnegie.ru/2/06/16/unnoticed-crisis-end-of-history-for-nuclear-arms-control-pub-60150408>; 19.07.2019.
- Arora, N. S./Russell, S./Sudderth, E. 2013: NET-VISA. Network Processing Vertically Integrated Seismic Analysis, in: Bulletin of the Seismological Society of America 103: 2A, 709-729.
- Mesterséges általános intelligencia Sentinel Initiative 2017: A Working List. A mesterséges intelligencia és az emberi intelligencia definíciói.
- Association of the United States Army 2017: Integrating Army Robotics and Autonomous Systems to Fight and Win, <https://www.USA.org/publications/integrating-army-robotics-and-autonomous-systems>; 19.07.2019.
- Balter, E. 2014: Digitális nyilatkozatok: Az INFCIRC/540 2.a. cikk (iii) bekezdése szerinti helyszínrajzok rendelkezésre bocsátása, <https://conferences.iaea.org/indico/event/47/contributions/8862/contribution.pdf>.
- Ben-Ari, Mordechai/Mondada, Francesco 2018: Cham.
- Bennaceur, Amel/Issarny, Valérie/Sykes, Daniel/Howar, Falk/Isberner, Malte/Steffen, Bernhard/Johansson, Richard/Moschitti, Alessandro 2013: Machine Learning for Emergent Middleware, in: Moschitti, Alessandro/Plank, Barbara (Eds.): Moschitti, Alessandro/Plank, Barbara (Eds.): Trustworthy Eternal Systems via Evolving Software, Data and Knowledge. Second International Workshop, EternalS 2012, Montpellier, Franciaország, augusztus Revised28,2012, Selected Papers, Berlin, Heidelberg, 16-29.
- Bojarski, Mariusz/Del Testa, Davide/Dworakowski, Daniel/Firner, Bernhard/Flepp, Beat/Goyal, Praveen/Jackel, Lawrence D./Monfort, Mathew/Muller, Urs/Zhang, Jiakai/Zhang, Xin/Zhao, Jake/Zieba, Karol 2016: End to End Learning for Self-Driving Cars, <https://arxiv.org/pdf/1604.07316v1.pdf>; 19.07.2019.
- Boston Dynamics 2018: Boston Dynamics. Megváltoztatjuk az elképzelésünket arról, hogy mire képesek a robotok, <https://www.bostondynamics.com/robots>; 05.01.2018.
- Bostrom, Nick 2017: Bostrom: A nyitottság stratégiai következményei a mesterséges intelligencia fejlesztésében, in: Global Policy 8: 2, 135-148.

- Boulanin, Vincent/Verbruggen, Maaïke 2017: Mapping the Development of Autonomy in Weapon Systems, Stockholm, [https://www.sipri.org/sites/default/files/2017-11/siprireport\\_mapping\\_the\\_development\\_of\\_autonomy\\_in\\_weapon\\_systems\\_1117\\_0.pdf](https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_0.pdf); 19.07.2019.
- Britting, Ernst/Spitzer, Hartwig 2005: Der Open-Skies-Vertrag: Stand und Perspektiven, in: Neuneck, Götz/Mölling, Christian (Szerk.): A nyílt légtér és az Open-Open-Spings: nézőpontok és nézőpontok: Die Zukunft der Rüstungskontrolle, Baden-Baden, 308-323.
- Bughin, Jacques/Hazan, Eric/Ramaswamy, Sree/Chui, Michael/Allas, Tera/Dahlström, Peter/Henke, Nicolaus/Trench, Monica 2017: Mesterséges intelligencia. The Next Digital Frontier?", <http://www.odbs.org/2017/08/artificial-intelligence-the-next-digital-frontier-mckinsey-global-institute-study/>; 19.07.2019.
- Cardillo, Robert 2017: Cardillo, Robert Cardillo: GEOINT 2017 Symposium (Remarks as prepared for Robert Cardillo), <https://www.nga.mil/MediaRoom/SpeechesRemarks/Pages/GEOINT-2017-Symposium.aspx>; 19.07.2019.
- CFTC/SEC 2010: Findings Regarding the Market Events of May 6, <https://www.sec.gov/news/studies/2010/marketevents-report.pdf>; 05.01.2018.
- Courtland, Rachel 2016: DARPA's Self-Driving Submarine Hunter Steers Like a Human, <https://spectrum.ieee.org/automaton/robotics/military-robots/darpa-actuv-self-driving-submarine-hunter-steers-like-a-human>; 05.01.2018.
- Croft, Stuart 1996: A fegyverzetellenőrzés stratégiái. A History and Typology, Manchester, Egyesült Királyság.
- Daniels, Jeff 2017: Daniels, Daniels: Daniels, Daniels: Mini atomfegyverek és szűnyogszerű robotfegyverek a jövő hadviseléséhez, <https://www.cnbc.com/2017/03/17/mini-nukes-and-inspect-bot-weapons-being-primed-for-future-warfare.html>; 05.01.2018.
- Danks, David 2014: Tanulás, in: Frankish, Keith/Ramsey, William (szerk.): Tanulás: Frankish, Keith/Ramsey, William (szerk.): The Cambridge Handbook of Artificial Intelligence, Cambridge, UK, 151-167.
- Dillow, Clay 2016: Dillow: Mi történik, ha a mesterséges intelligenciát és a műholdas képalkotást kombináljuk, <http://fortune.com/2016/03/30/facebook-ai-satellite-imagery/>; 26.11.2017.
- Eilam, Eldad 2005: Visszafordítás. Secrets of Reverse Engineering, Indianapolis, USA.
- Ernest, Nicholas/Carroll, David/Schumacher, Corey/Clark, Matthew/Cohen, Kelly/Lee, Gene 2016: Genetic Fuzzy based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions, in: Journal of Defense Management 6: 1, 1-7, <https://www.omicsonline.org/open-access/genetic-fuzzy-based-artificial-intelligence-for-unmanned-combat-aerialvehicle-control-in-simulated-air-combat-missions-2167-0374-1000144.pdf>; 31.01.2018.
- Európai Bizottság 2017: Kettős felhasználású termékek exportellenőrzése, [http://ec.europa.eu/trade/import-and-export-rules/export-from-eu/dual-use-controls/index\\_en.htm](http://ec.europa.eu/trade/import-and-export-rules/export-from-eu/dual-use-controls/index_en.htm); 19.07.2019.
- Evangelista, Matthew 1988: Innováció és fegyverkezési verseny. How the United States and the Soviet Union Develop New Military Technologies, Ithaca, NY.
- Farrelly, Colleen 2016: <https://www.slideshare.net/ColleenFarrelly/machine-learning-by-analogy-59094152>; 09.01.2018.
- Franklin, Stan 2014: Történet, motivációk és központi témák, in: Frankish, Keith/Ramsey, William (szerk.): Frankish, Keith/Ramsey, William (szerk.): The Cambridge Handbook of Artificial Intelligence, Cambridge, UK, 15-33.

- Az Élet Jövője Intézet 2015: Autonóm fegyverek. A mesterséges intelligencia és robotika kutatóinak nyílt levele, <https://futureoflife.org/open-letter-autonomous-weapons>; 14.01.2018.
- Goldblat, Jozef 2002: Gof Goldbatch: Fegyverzetellenőrzés. A tárgyalások és megállapodások új útmutatója, London. Goodfellow, Ian 2016: Deep Learning, Cambridge, Massachusetts, London, Anglia.
- Goodfellow, Ian J./Shlens, Jonathon/Szegedy, Christian 2015: Explaining and Harnessing Adversarial Examples, <https://arxiv.org/pdf/1412.6572.pdf>; 19.07.2019.
- Gubrud, Mark/Altmann, Jürgen 2013: Megfelelési intézkedések egy autonóm fegyverekről szóló egyezményhez. ICAC Working Paper #2, [https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Alt-mann\\_Compliance-Measures-AWC\\_ICAC-WP2.pdf](https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Alt-mann_Compliance-Measures-AWC_ICAC-WP2.pdf); 19.07.2019.
- Gunning, David 2016: Gunningunning, Gunning: Explainable Artificial Intelligence (XAI). Broad Agency Announcement, <https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf>; 07.12.2017.
- Hagen, Christian/Sorenson, Jeff/Hurt, Steven/Wall, Dan 2012: Szoftver: The Brains Behind U.S. Defense Systems, [https://www.atkearney.com/documents/10192/247932/SoftwareThe\\_Brains\\_Behind\\_US\\_Defense\\_Systems.pdf/69129873-eecc-4ddc-b798-c198a8ff1026](https://www.atkearney.com/documents/10192/247932/SoftwareThe_Brains_Behind_US_Defense_Systems.pdf/69129873-eecc-4ddc-b798-c198a8ff1026); 19.07.2019.
- Harris Cooperation 2017: ENVI Feature Extraction Module, [http://www.harrisgeospatial.com/Portals/0/pdfs/HG\\_ENVI\\_FX\\_module\\_data-sheet\\_WEB.pdf](http://www.harrisgeospatial.com/Portals/0/pdfs/HG_ENVI_FX_module_data-sheet_WEB.pdf); 26.11.2017.
- Hawkins, Jeff/Blakeslee, Sandra 2004: New York: On Intelligence, 1<sup>st</sup> Edition, New York.
- Hochmuth, Olaf 2003: 2000.tik.hu-berlin.de/~hochmuth/bvp/; 13.12.2017.
- Holtom, Paul/Bromley, Mark 2011: Tranzit- és átrakodási ellenőrzések egy fegyverkereskedelmi szerződésben. <https://www.sipri.org/sites/default/files/files/misc/SIPRIBP1107a.pdf>; 19.07.2019.
- Human Rights Watch 2012: Emberiesség elvesztése. The Case against Killer Robots, <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>; 14.01.2018.
- Nemzetközi Atomenergia-ügynökség 2018: Nemzetközi Atomenergia Ügynökség: Kutatási és Fejlesztési Terv. Enhancing Capabilities for Nuclear Verification, Wien, Österreich.
- Jaccard, Nicolas/Thomas W. Rogers/Edward J. Morton/Lewis D. Griffin 2016: Automated Detection of Smuggled High-risk Security Threats Using Deep Learning, <https://arxiv.org/pdf/1609.02805.pdf>, 19.07.2019.
- Johnson, Michael R./Paquette, Jean-Pierre/Elbez, Julien 2014: <https://www.iaea.org/safeguards/symposium/2014/home/e-proceedings/sg2014-papers/000042.pdf>; 26.11.2017.
- Kahl, Martin/Mölling, Christian 2005: Kahl: Die "Revolution in Military Affairs" und die Bedingungen und Möglichkeiten für Rüstungskontrolle, in: Die "Revolution in Military Affairs" und die Bedingungen und Möglichkeiten für Rüstungskontrolle: Neuneck, Götz/Mölling, Christian (Eds.): Neuneck, Götz/Mölling, Christian (Eds.): Die Zukunft der Rüstungskontrolle, Baden-Baden, 341-353.
- Kania, Elsa B. 2017: Kania Kania: Quest for an AI Revolution in Warfare. The PLA's Trajectory's from Informatized to "Intelligentized" Warfare, <https://thestrategybridge.org/the-bridge/2017/6/8/chinas-quest-for-an-ai-revolution-in-warfare>; 14.01.2018.
- Keller, John 2015: DARPA TRACE Program Using Advanced Algorithms, Embedded Computing for Radar Target Recognition, <http://www.militaryaerospace.com/articles/2015/07/hpec-radar-target-recognition.html>; 01.12.2017.

- Kleiner, Ariel/Mackey, Lester/Jordan, Michael I. 2009: Machine Learning for Improved Automated Seismic Event Extraction, [https://www.ctbto.org/fileadmin/user\\_upload/ISS\\_2009/Poster/DM-02A%20%28US%29%20-%20Ariel\\_Kleiner%20etal.pdf](https://www.ctbto.org/fileadmin/user_upload/ISS_2009/Poster/DM-02A%20%28US%29%20-%20Ariel_Kleiner%20etal.pdf); 19.07.2019.
- Knight, Will 2017: A mesterséges intelligencia szívében rejlő sötét titok. Senki sem tudja igazán, hogyan teszik a legfejlettebb algoritmusok, amit tesznek. Ez probléma lehet., <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>; 11.01.2018.
- Krebs, Gunter 2017: KH-11/Kennen/Crystal, [http://space.skyrocket.de/doc\\_sdat/kh-11.htm](http://space.skyrocket.de/doc_sdat/kh-11.htm); 25.11.2017.
- Kurakin, Alexey/Goodfellow, Ian/Bengio, Samy 2017: Adversarial Examples in the Physical World, <https://arxiv.org/pdf/1607.02533.pdf>; 19.07.2019.
- Lant, Karla 2017: Kína, Oroszország és az USA mesterséges intelligencia-fegyverkezési versenyben állnak, <https://futurism.com/china-russia-and-the-us-are-in-an-artificial-intelligence-arms-race/>; 16.11.2017.
- Legg, Shane/Hutter, Marcus 2007: A Collection of Definitions of Intelligence, <https://arxiv.org/pdf/0706.3639.pdf>; 19.07.2019.
- Li, Wei/Gauci, Melvin/Groß, Roderich 2016: Turing learning. A Metric-free Approach to Inferring Behavior and its Application to Swarms, in: *Swarm Intelligence 10*: 211-2433,.
- McCloskey, Paul 2017: <https://gcn.com/Articles/2017/03/10/defining-AI.aspx>; 15.11.2017.
- Morisse, Tom 2017: The Next Challenges of AI Research, <https://en.fabernovel.com/insights/tech-en/the-next-challenges-of-ai-research>; 18.12.2017.
- Müller, Harald 2000: Früherkennung von Rüstungsrisiken in der Ära der "militärisch-technischen Revolution". Ein Register für militärische Forschung und Entwicklung, Frankfurt.
- Müller, Harald/Schörning, Niklas 2006: Rüstungsdynamik und Rüstungskontrolle. Eine exemplarische Einführung in die Internationalen Beziehungen, Baden-Baden.
- Müller, Vincent C./Bostrom, Nick 2016: A mesterséges intelligencia jövőbeni fejlődése: A Survey of Expert Opinion, in: Müller, Vincent C. (szerk.): *Fundamental Issues of Artificial Intelligence*, Berlin, 553-571.
- Neuneck, Götz/Alwardt, Christian 2008: [https://ifsh.de/pdf/publikationen/IFAR\\_Working\\_Paper\\_13.pdf?asset\\_id=5489](https://ifsh.de/pdf/publikationen/IFAR_Working_Paper_13.pdf?asset_id=5489); 19.07.2019.
- Neuneck, Götz/Mutz, Reinhard 2000: Vorbeugende Rüstungskontrolle. Ziele und Aufgaben unter besonderer Berücksichtigung verfahrensmäßiger und institutioneller Umsetzung im Rahmen internationaler Rüstungsregime, Baden-Baden.
- Niemeyer, Irmgard/Ruthowski, Joshua 2016: Bonn: Satellite Imagery Processing for the Verification of Nuclear Non-Proliferation and Arms Control, Bonn.
- Núñez-Nieto, Xavier/Solla, Mercedes/Gómez-Pérez, Paula/Lorenzo, Henrique 2014: GPR Signal Characterization for Automated Landmine and UXO Detection Based on Machine Learning Techniques, in: *Remote Sensing 6*: 9729-974810,, <http://www.mdpi.com/2072-4292/6/10/9729/pdf>; 11.01.2018.
- O'Neil, Cathy 2016: A matematikai pusztítás fegyverei. How Big Data Increases Inequality and Threatens Democracy, 1<sup>st</sup> Edition, New York.

- Osborn, Kris 2017: <https://defensesystems.com/articles/2017/03/14/f22.aspx>; 03.12.2017.
- Papernot, Nicolas/McDaniel, Patrick/Goodfellow, Ian/Jha, Somesh/Celik, Z. B./Swami, Ananthram 2017: Practical Black-Box Attacks against Machine Learning (Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security, Abu Dhabi, UAE), Abu Dhabi, UAE.
- Park, Dong H./Hendricks, Lisa A./Akata, Zeynep/Schiele, Bernt/Darrell, Trevor/Rohrbach, Marcus 2017: Figyelmes magyarázatok. Döntések igazolása és a bizonyítékokra való rámutatás, <https://arxiv.org/abs/1612.04757>; 19.07.2019.
- Parkin, Simon 2015: Gyilkos robotok: The Soldiers that Never Sleep, <http://www.bbc.com/future/story/20150715-killer-robots-the-soldiers-that-never-sleep>; 05.01.2018.
- Patton, Tamara/Lewis, Jeffrey/Hanham, Melissa/Dill, Catherine/Vaccaro, Lily 2016: Emerging Satellites for Non-Proliferation and Disarmament Verification, [https://nonproliferation.org/vcdnp/wp-content/uploads/2016/06/160614\\_copernicus\\_project\\_report.pdf](https://nonproliferation.org/vcdnp/wp-content/uploads/2016/06/160614_copernicus_project_report.pdf); 19.07.2019.
- Pilat, Joseph F. 2002: Pilat Pilát: Ellenőrzés és átláthatóság: Relikviák vagy jövőbeli követelmények?, in: Larsen, Jeffrey A. (Szerk.): Az átláthatóság és az átláthatóság: a világosság és az átláthatóság kérdései: Larsen, Jeffrey A: Arms Control. Cooperative Security in a Changing Environment, London, 79-96.
- Procopio, Michael J./Young, Christopher J./Gauthier, John A. 2009: Applying Machine Learning Methods to Improve Efficiency and Effectiveness of the IDC Automatic Event Detection System, [https://www.ctbto.org/fileadmin/user\\_upload/ISS\\_2009/Poster/DM-07A\\_US\\_-\\_Michael\\_Procopio\\_et.al.pdf](https://www.ctbto.org/fileadmin/user_upload/ISS_2009/Poster/DM-07A_US_-_Michael_Procopio_et.al.pdf); 19.07.2019.
- Ribeiro, Marco T./Singh, Sameer/Guestrin, Carlos 2016: "Miért bízzak benned?". Explaining the Predictions of Any Classifier, <https://arxiv.org/pdf/1602.04938.pdf>; 19.07.2019.
- Russell, Stuart J./Norvig, Peter 2010: Mesterséges intelligencia. A Modern Approach, 3<sup>rd</sup> Edition, Upper Saddle River, NJ.
- Russell, Stuart J./Vaidya, Sheila/Le Bras, Ronan 2010: Machine learning for Comprehensive Nuclear-Test-Ban Treaty monitoring, in: CTBTO Spectrum: 32-35; 14,20.11.2017.
- Russia Today 2017: Kalasnyikov teljesen automatizált neurális hálózat alapú harci modult fejleszt, <https://www.rt.com/news/395375-kalashnikov-automated-neural-network-gun/>; 05.01.2018.
- Sauer, Frank 2016: A "gyilkos robotok" megállítása: Why Now Is the Time Is the Time to Ban Autonomous Weapons Systems, [https://www.armscontrol.org/ACT/2016\\_10/Features/Stopping-Killer-Robots-Why-Now-Is-the-Time-to-Ban-Autonomous-Weapons-Systems#note04](https://www.armscontrol.org/ACT/2016_10/Features/Stopping-Killer-Robots-Why-Now-Is-the-Time-to-Ban-Autonomous-Weapons-Systems#note04); 13.01.2018.
- Scharre, Paul 2016: [https://www.files.ethz.ch/isn/196288/CNAS\\_Autonomous-weapons-operational-risk.pdf](https://www.files.ethz.ch/isn/196288/CNAS_Autonomous-weapons-operational-risk.pdf); 19.07.2019.
- Schelling, Thomas C./Halperin, Morton H. 1961: New York: Strategy and Arms Control, New York.
- Schlosser, Eric 2013: Carlos Elosser: Command and Control. Nuclear Weapons, the Damascus Accident, and the Illusion of Safety, New York, NY.
- Schmidt, Hans-Joachim 2017: Hoffnungsvoller Neustart der konventionellen Rüstungskontrolle?, <https://blog.prif.org/2017/07/10/hoffnungsvoller-neustart-der-konventionellen-ruestungskontrolle/>; 14.01.2018.

Schörnig, Niklas 2008: Az áldozatoktól való idegenkedés a demokratikus biztonsági ellátásban. Procurement and the Defense Industrial Base., in: Evangelista, Matthew (szerk.): Democracy and Security. Preferenciák, normák és politikai döntéshozatal, London, 14-35.

Schörnig, Niklas 2015: A mennyiségi fegyverzetellenőrzéstől a minőségi fegyverzetellenőrzésig: The Challenges of Modern Weapons Development, in: Development and Peace Foundation/Käte Hamburger Kolleg/Centre for Global Cooperation Research (Eds.): Global Trends Prospects 2015. for World Society (Globális trendek kilátásai a világtársadalom számára), 87-100.

Searle, John R. 1980: Agyak, agyak és programok, in: Searle Searle: Minds, Brains, and Programs: Minds, Brains, and Programs: Behavioral and Brain Sciences 3: 417-4573,.

Seiffert, Udo/Abeynayake, Canicious/Jain, Lakhmi C./Tran, Minh D.-J. 2013: Detection of Targets in Characteristic GPR Sensor Data Using Machine Learning Techniques, [https://www.techfak.uni-bielefeld.de/~fshleif/mlr/mlr\\_01\\_2013.pdf](https://www.techfak.uni-bielefeld.de/~fshleif/mlr/mlr_01_2013.pdf); 19.07.2019.

Shalal, Andrea 2014: DigitalGlobe Gains U.S. Govt License to Sell Sharper Satellite Imagery, <https://www.reuters.com/article/digitalglobe-imagery/digitalglobe-gains-u-s-govt-license-to-sell-sharp-er-satellite-imagery-idUSL2N0OR2UX20140611>; 25.11.2017.

Shalev-Shwartz, Shai/Ben-David, Shai 2014: Understanding Machine Learning: From Theory to Algorithms, New York.

Sharif, Mahmood/Bhagavatula, Sruti/Bauer, Lujo/Reiter, Michael K. 2016: Accessorize to a Crime. Valós és lopakodó támadások a legkorszerűbb arcfelismerés ellen, Wien.

Shaw, Martin 2005: Shaw: A háború új nyugati módja. Kockázatáthárító háború és annak válsága Irakban, Cambridge, Egyesült Királyság.

Stevenson, Beth 2016: Stevenson Stevenson: Elemzés: <https://www.flightglobal.com/news/articles/analysis-taranis-developers-reveal-test-flight-spec-425347/>; 19.07.2019.

Stocki, Trevor J./Li, Guichong/Japkowicz, Nathalie/Ungar, R. K. 2010: Machine Learning for Radioxenon Event Classification for the Comprehensive Nuclear-Test-Ban Treaty, in: Journal of environmental radioactivity 101: 68-741,.

Sun, Yijun/Li, Jian 2005: Jin Yun Jin, Jin Yun Jin: Adaptív tanulási megközelítés a taposóaknák felderítéséhez, in: IEEE Transactions on Aerospace and Electronic Systems 41: 1-93,.

Sundaresan, Lalitha/Chandrashekar, S./Jasani, Bhupendra 2017: Discriminating Uranium and Copper Mills Using Satellite Imagery, in: Remote Sensing Applications: Society and Environment: 27-355,.

Szegedy, Christian/Zaremba, Wojciech/Sutskever, Ilya/Bruna, Joan/Erhan, Dumitru/Goodfellow, Ian/Fergus, Rob 2014: Intriguing Properties of Neural Networks, <https://arxiv.org/pdf/1312.6199.pdf>; 19.07.2019.

Truong, Q. B./Borstad, Gary/Saper, Ron 2005: Integration of Satellite Imagery and Other Tools in Safeguards Information Analysis, [https://www.remote-sensing.aslenv.com/documents/ESARDA\\_paper\\_2005\\_TRUONG\\_etal.pdf](https://www.remote-sensing.aslenv.com/documents/ESARDA_paper_2005_TRUONG_etal.pdf); 19.07.2019.

Tuma, Matthias/Igel, Christian 2009: Kernel-Based Machine Learning Techniques for Hydroacoustic Signal Classification, CTBTO International Scientific Studies Conference, Wien.

UNIDIR/VERTIC 2003: Coming to Terms with Security. A Handbook on Verification and Compliance. <http://www.unidir.org/files/publications/pdfs/coming-to-terms-with-security-a-handbook-on-verification-and-compliance-en-554.pdf>; 19.07.2019.

Amerikai haditengerészet 2017: MK 15 - Phalanx Close-in Weapons System (CIWS), [http://www.navy.mil/navydata/fact\\_display.asp?cid=2100&tid=487&ct=2](http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2); 05.01.2018.

Yu, Xie/Jing, Meng 2017: Xi Jinping épp most adott zöld utat a tervnek, <http://www.scmp.com/business/china-business/article/2115935/chinas-xi-jinping-highlights-ai-big-data-and-shared-economy>; 16.11.2017.

Zeiler, Matthew D./Fergus, Rob2013 : Visualizing and Understanding Convolutional Networks, <https://arxiv.org/pdf/1311.2901.pdf>; 19.07.2019.

## prif jelentés

A PRIF-jelentések háttérelmzéseket nyújtanak a politikai eseményekről és fejleményekről, és kutatási eredményeket mutatnak be.

Fröhlich, Marieke (2019): Férfiasság a békefenntartásban. Az ENSZ BT 1325-ös határozatának korlátai és átalakulásai a Dél-afrikai Nemzeti Védelmi Erőknél, PRIF Report 7/2019, Frankfurt/M.

Christian, Ben/Coni-Zimmer, Melanie (2019): Deutschland im UN-Sicherheitsrat 2019-2020. Eine Halbzeitbilanz, PRIF Report 6/2019, Frankfurt/M.

PRIF REPORT



[www.hsfk.de/PRIF-Reports](http://www.hsfk.de/PRIF-Reports)  
[www.hsfk.de/HSFK-Reports](http://www.hsfk.de/HSFK-Reports)

## prif SpotLight

A PRIF Spotlights aktuális politikai és társadalmi kérdéseket tárgyal.

Fehl, Caroline (2020): Syrische Folter vor Gericht. Die partielle Rückkehr des universellen Rechts, PRIF Spotlight 2/2020, Frankfurt/M.

Polianskii, Mikhail/Rogova, Vera (2020): Elveszett az átmenetben? Rutyn stratégiája a PRIF2024, Spotlight 1/2020, Frankfurt/M.

PRIF SPOTLIGHT

KEIN FRIEDEN OHNE MENSCHENRECHTE



[www.hsfk.de/PRIF-Spotlights](http://www.hsfk.de/PRIF-Spotlights)

## prif bLog

A PRIF Blog a béke- és konfliktuskutatás szempontjából releváns aktuális politikai kérdésekről és de- bátumokról szóló cikkeket mutat be.

PRIF BLOG

PEACE RESEARCH INSTITUTE / LEIBNIZ-INSTITUT FÜR SOZIALE ERSTUNGS- UND KONFLIKTFORSCHUNG

<https://blog.prif.org/>

A PRIF-jelentések és a PRIF Spotlights nyílt hozzáférésű kiadványok, amelyek a [www.prif.org](http://www.prif.org) címen tölthetők le. Ha e-mailben vagy nyomtatott formában szeretné megkapni kiadványainkat, kérjük, vegye fel a kapcsolatot a [publikationen@hsfk.de](mailto:publikationen@hsfk.de) címen.

[www.facebook.com/HSFK.PRIF](https://www.facebook.com/HSFK.PRIF)

[www.twitter.com/HSFK\\_PRIF](https://www.twitter.com/HSFK_PRIF)

<https://blog.prif.org/>

NICO LÜCK /

## A GÉPI TANULÁSON ALAPULÓ MESTERSÉGES INTELLIGENCIA A FEGYVERZETELLENŐRZÉSSEN

A mesterséges intelligencia (AI), különösen a gépi tanulás által vezérelt mesterséges intelligencia mindenki száján van. Az ilyen rendszerek még a fegyverkezésben is egyre fontosabb szerepet játszanak: Egyes fegyverrendszerek már képesek önállóan azonosítani a célpontokat és harcba szállni velük. Ez problémákat vet fel a fegyverzetellenőrzés hagyományos formái számára, amelyeket eredetileg fizikai tárgyak, például aknák és kézfegyverek és azok belső működésének megfigyelésére terveztek. Emellett a megbízható ellenőrzés olyan fontos további hatásai, mint a bizalomépítés és a diplomáciai kapcsolatok stabilizálása, nem kerülnek szóba. Fontos, hogy a fegyverzetellenőrzés foglalkozzon ezekkel a kockázatokkal is.

Ugyanakkor a gépi tanuláson alapuló mesterséges intelligencia (MLpAI) mint eszköz alkalmazása óriási lehetőséget kínál a fegyverzetellenőrzési folyamatok javítására. Itt a pontosabb és átfogóbb adatfeldolgozás különösen az államok közötti nagyobb bizalmat teremthet. Ez a jelentés az MLpAI fegyverzetellenőrzésben való alkalmazásával kapcsolatos kockázatok és lehetőségek közötti feszültségre világít rá.

ISBN 978-3-946459-51-4