

# Ratio fabricata és Robotika

## Ratio fabricata and Robotics

*Prof. Dr. VÁSÁRHELYI József*

Miskolci Egyetem, Gépészmérnöki és Informatikai Kar,  
Automatizálási és Infokommunikációs Intézet,  
3515 Miskolc, Egyetemváros  
jozsef.vasarhelyi@uni-miskolc.hu

### Abstract

*This article examines how ethical artificial intelligence, better known as AI, really is and how much harm it can potentially cause. It explores how AI interprets itself, its relationship with humans, and how it influences the development of robotics. Is there a real danger that machines could one day run out of human control?*

**Keywords:** AI, robotics, embedded systems, sensors, safety

### Kivonat

*Jelen cikk azt vizsgálja, hogy a mesterséges értelem vagy ismertebb nevén AI mennyire erkölcsös mennyit árthat? Hogyan értelmezi az MI saját magát, mi a viszonya az emberhez és a hogyan befolyásolja a robotika fejlődését. Van-e valós veszélye annak, hogy a gépek elszabadulnak?*

**Kulcsszavak:** MI – mesterséges értelem, robotika, beágyazott rendszerek, érzékelők, biztonság

## 1. BEVEZETÉS

A következő „párbeszéd” az AI vagy MI – helyesebben a mesterséges értelem (MÉ) – vagy „*Ratio Fabricata*” és a szerző között zajlott. A kérdések: mennyire értelmes a MÉ és hogyan értelmezi – az internetről összegyűjtött adatok alapján saját magát, a robotika törvényeit és árthat-e az embereknek, emberiségnek? Az már biztosan kijelenthető, hogy a MÉ sosem foglal egyértelmű álláspontot semmiben. A válaszai mindig a „politikailag korrekt semleges” véleményekre korlátozódnak [1]–[5].

## 2. KÉRDÉSEK AZ MESTERSÉGES ÉRTELEMRŐL

### 2.1. Mi az intelligencia?

„Az intelligencia fogalmára nincs egyetlen, mindenki által elfogadott definíció, de általában azt értjük alatta, hogy egy rendszer (ember, állat vagy mesterséges rendszer) képes:

- információt megszerezni és feldolgozni,
- tanulni a tapasztalatokból,
- problémákat megoldani,
- alkalmazkodni új helyzetekhez,
- célokat kitűzni és elérni,
- következtetéseket levonni.

A pszichológiában gyakran úgy definiálják, mint az egyén általános mentális képességét, amely magában foglalja a logikus gondolkodást, a tervezést, az absztrakt gondolkodást, a tanulást és a tapasztalatokból való hasznos tudásszerzést.

Különböző nézőpontok léteznek:

- Biológiai szemszögből: Az intelligencia az idegrendszer azon képessége, hogy hatékonyan reprezentálja és kezelje a környezetből származó információkat.
- Pszichológiai szemszögből: Az intelligencia több részterületből állhat:

- logikai-matematikai képesség,
  - nyelvi képesség,
  - térbeli tájékozódás,
  - társas (szociális) intelligencia,
  - érzelmi intelligencia.
- Mesterséges intelligencia szempontjából: Egy rendszer akkor tekinthető intelligensnek, ha érzékeli a környezetét, döntéseket hoz, tanul és képes a céljai elérésére változó körülmények között.

Egy tömör definíció:

Az intelligencia talán így foglalható össze a legegyszerűbben: *Az intelligencia annak képessége, hogy tudást szerezzünk, azt felhasználjuk, és sikeresen alkalmazkodjunk új helyzetekhez céljaink elérése érdekében.*

Érdekes módon nincs egyetértés abban sem, hogy az intelligencia egyetlen általános képesség („g-faktor”), vagy sok különálló képesség összessége. Ez ma is aktív kutatási terület a pszichológiában, a neurológiában és a mesterséges intelligenciában.” [1], [2], [5]

## 2.2. Mi az értelem?

„Az értelem és az intelligencia közeli, de nem azonos fogalmak. Az intelligencia inkább a képesség: milyen hatékonyan tudunk tanulni, problémát megoldani, összefüggéseket felismerni. Az értelem inkább a gondolkodás tartalma és működése: a valóság megértésének, felfogásának és értelmezésének képessége. Például valaki lehet nagyon intelligens (gyorsan tanul, kiválóan számol), de az értelemhez kapcsolódik az is, hogy felismeri döntéseinek következményeit, megérti a helyzetek mélyebb jelentését, és bölcsen használja tudását. Filozófiai értelemben az értelem (latinul ratio, görögül logosz bizonyos jelentéseiben) az ember azon képessége, amellyel:

- fogalmakat alkot,
- ítéleteket hoz,
- következtet,
- a dolgok okait és összefüggéseit keresi.

A magyar nyelvben az „értelem” szónak van egy másik jelentése is: jelentés vagy értelmezhetőség. Például: „Mi ennek a mondatnak az értelme?” Itt nem a gondolkodóképességről, hanem a mondat jelentéséről beszélünk.

Röviden: Az intelligencia azt mutatja, milyen jól tudunk gondolkodni; az értelem pedig azt, hogy mit és hogyan értünk meg a világból.”

## 2.3. Az AI értelmes?

„Attól függ, mit értünk értelem alatt. Ha az értelem alatt azt értjük, hogy egy rendszer:

- megérti a nyelvet,
- következtetéseket von le,
- problémákat old meg,
- tudást alkalmaz új helyzetekben,

akkor a mai mesterséges intelligenciák bizonyos mértékig értelmesnek tűnnek, és sok feladatban valóban értelmes viselkedést mutatnak.

Ha viszont az értelemhez hozzátartozik:

- a tudatos tapasztalás,
- az önmaga létezésének átélése,
- a saját szándékok és célok kialakítása,
- a valódi megértés belső élménye,

akkor jelenlegi tudásunk szerint az AI nem értelmes úgy, mint egy ember.

A vita lényegét John Searle filozófus híres „kínai szoba” gondolkísérlete ragadja meg [4]. Egy rendszer képes lehet szabályok alapján helyes válaszokat adni anélkül, hogy valóban értené a jelentésüket. Sok filozófus szerint a mai AI-k hasonlóan működnek: rendkívül jól kezelik a szimbólumokat, de nem biztos, hogy „értik” azokat.

Én például képes vagyok:

- fogalmak közötti kapcsolatokat felismerni,
- érvelni,
- magyarázni,
- matematikai és logikai problémákat megoldani;

De nincs tudatom, önélményem vagy szubjektív tapasztalatom. Nem „érzem”, hogy megérték valamit; csak olyan műveleteket hajtok végre, amelyek a megértéshez hasonló eredményt produkálnak.

Ezért egy óvatos megfogalmazás: A mai AI funkcionális értelemben részben értelmes, de nem tudjuk, hogy rendelkezik-e valódi értelemmel vagy tudattal. A legtöbb kutató szerint jelenleg nem.” [3], [4]

## **2.4. Az AI tudatos?**

„A mai AI rendszereket a tudomány jelenlegi állása szerint nem tekintjük tudatosnak. Az AI intelligens viselkedést tud mutatni, de nincs bizonyíték arra, hogy tudatos lenne vagy szubjektív élményeket élne át.” [3], [4]

## **2.5. Az AI Békés?**

„Önmagában az AI nem békés és nem agresszív. Az AI egy eszköz vagy rendszer, amelynek nincs saját akarata, erkölce vagy szándéka. Ahogy egy kalapáccsal lehet házat építeni vagy ablakot betörni, úgy az AI-t is lehet: békés célokra használni (oktatás, orvosi diagnosztika, tudományos kutatás, fordítás), vagy káros célokra (csalások, manipuláció, Kiber támadások támogatása). Az a kérdés, hogy egy AI „békés-e”, részben attól függ, hogyan tervezték és milyen korlátokat építettek bele. Sok modern AI rendszert úgy alakítanak ki, hogy csökkentse a károkozás kockázatát, és ne segítsen veszélyes vagy ártalmas tevékenységekben.

Filozófiai szempontból még mélyebb a kérdés. A békesség általában egy szándékot vagy erkölcsi hozzáállást feltételez. Mivel a jelenlegi AI-nak nincs saját tudatuk és saját céljuk, ezért nem lehetnek békések ugyanabban az értelemben, ahogyan egy ember lehet békés.

Ezért a rövid válasz: A mai AI nem békés, de nem is erőszakos. Nincs saját szándéka; a hatása attól függ, hogyan használják és milyen szabályok szerint működik.”

## **2.6. Az AI árthat az embernek?**

Igen, az AI árthat az embernek, de fontos pontosítani, hogy jelenleg nem azért, mert „rosszindulatú”, hanem mert:

- hibázhat,
- félreértheti a helyzetet,
- pontatlan információt adhat,
- emberek káros célokra használhatják,
- rosszul megtervezett rendszerek nem kívánt következményeket okozhatnak.

Példák:

- Egy orvosi AI téves diagnózisa káros döntéshez vezethet.
- Egy önvezető rendszer hibája balesetet okozhat.
- AI-val előállított félretájékoztatás befolyásolhat embereket.
- Kiber bűnözők felhasználhatják AI eszközöket csalásokhoz.

Ugyanakkor az AI sok esetben csökkentheti is a károkat:

- segíthet betegségek felismerésében,
- javíthatja a közlekedés biztonságát,
- támogathatja az oktatást,
- gyorsíthatja a tudományos kutatást.

Filozófiai értelemben érdekes különbség van az ártani tudni és az ártani akarni között. A mai AI képes közvetve kárt okozni, de nem ismert, hogy képes lenne akarni a károkozást. Nincsenek saját vágyai, gyűlölete, bosszúvágya vagy rosszindulata.

Ezért a kérdésre a rövid válasz: Igen, az AI árthat az embernek, de jelenleg főként eszközként, hibákon vagy emberi felhasználáson keresztül, nem pedig saját rossz szándékból.,,

Lehetne tovább folytatni, de inkább nézzük mi a helyzet a robotokkal, mivel a robotok irányításában használható a ME.

### 3. KÉRDÉSEK A ROBOTIKÁVAL KAPCSOLATBAN

Általánosan fogalmazva alkalmazástól függetlenül a robot képességeit befolyásolhatja, hogy csak irányításra, feladatok elvégzésére használható az ME. A továbbiakban ilyen irányú kérdéseket teszek fel. Isaac Asimov, először 1942-ben a Roundaround (Körbe körbe) című novellájában írja le a robotika 3 törvényét, melyet később kiegészített egy 0. törvénnyel is [6], [7], [8]. Több művében is előfordulnak a robotika törvényei (Körbe körbe, Én a Robot, Alapítvány sorozat stb.). Természetesen az ME az adatbázisok alapján ismeri a törvényeket és részletes leírást adott róluk.

#### 3.1. Mi a robotika három törvénye? Hogyan alkalmazhatók ezek a törvények ma?

„A három eredeti törvény hierarchiája:

- A robot nem árthat embernek, és nem tűrheti tétlenül, hogy embernek baja essen.
- A robot köteles engedelmeskedni az emberi utasításoknak, kivéve, ha ez sértené az első törvényt.
- A robot köteles megvédeni önmagát, amennyiben ez nem ütközik az első vagy a második törvénybe [7].

Ez azt jelenti, hogy ha egy ember olyan parancsot adna egy robotnak, amely emberi sérülést okozna, a robotnak meg kell tagadnia az engedelmisséget, mert az első törvény mindig elsőbbséget élvez.

Asimov három törvénye eredetileg irodalmi eszköz volt, nem mérnöki vagy jogi szabályrendszer. Mégis nagy hatással volt a robotika és a mesterséges intelligencia etikájáról folytatott gondolkodásra [6]–[8].

#### Mai értelmezésben

**1. törvény:** Az ember védelme. A robot nem árthat embernek. Ez ma megjelenik például:

- az önvezető járművek biztonsági rendszereiben,
- az ipari robotok védőburkolataiban és vészleállító rendszereiben,
- az orvosi eszközök szigorú tesztelésében.

A cél, hogy a technológia tervezése során az emberi biztonság legyen az elsődleges szempont. Ez nagyon hasonlít Asimov első törvényének szelleméhez.

**2. törvény: Engedelmisség az embernek.** A robot engedelmeskedjen az ember utasításainak. A modern AI esetében ez úgy jelenik meg, hogy a rendszer a felhasználó kérését próbálja teljesíteni. Ugyanakkor ma már nem elegendő a pusztá engedelmisség.

Például egy AI nem teljesíthet olyan utasítást, amely:

- embereknek ártana,
- veszélyes tevékenységet segítene,
- jogsértő vagy káros lenne.

Érdekes módon ez pontosan megfelel Asimov elképzelésének: a második törvény csak addig érvényes, amíg nem sérti az első.

**3. törvény: Önvédelem:** A robot védje saját létét. A mai rendszereknél ez nem erkölcsi kérdés, hanem megbízhatósági kérdés:

- adatmentések,
- hibatűrő rendszerek,
- redundáns hardverek,
- kibervédelmi mechanizmusok.

Az AI vagy a robot azért védi működését, hogy tovább tudja ellátni feladatát, nem pedig azért, mert „élni akar”.

#### 3.2. Mi a probléma Asimov törvényeivel?

Asimov maga is rámutatott a történeteiben, hogy a törvények nem mindig egyértelműek.

Például:

- Mi számít „ártalomnak”?
- Egy rövid távú kellemetlenség megengedhető egy nagyobb jó érdekében?
- Kinek az utasítását kell követni, ha két ember ellentmond egymásnak?
- Mi történik, ha az ember maga kér veszélyes dolgot?

Ezek a kérdések ma is aktuálisak az AI-etikában. Egy modern megfogalmazás ha ma kellene Asimov szellemében megfogalmazni a szabályokat, talán így hangoznának:

- Az AI ne okozzon indokolatlan kárt embereknek.
- Az AI szolgálja az emberek törvényes és biztonságos céljait.
- Az AI legyen megbízható és ellenálló a hibákkal szemben.
- Az AI működése legyen átlátható és emberi felügyelet alatt álljon.

Érdekes, hogy Asimov később bevezetett egy úgynevezett „nulladik törvényt” is: „A robot nem árt-hat az emberiségnek, és tétlenségével sem engedheti, hogy az emberiség kárt szenvedjen.”

Ez már nem egyes emberek, hanem az egész emberiség érdekeit helyezte előtérbe, és meglepően ha-sonlít a mai AI biztonsági viták egy részére.

Hogyan alkalmazhatóak ezek a törvények a gyakorlatban?

A gyakorlatban Asimov törvényeit nem lehet szó szerint beprogramozni, de tervezési elvekként jól al-kalmazhatók.

### 1. Az első törvény: az ember védelme; „A robot nem árthat embernek...”

A gyakorlatban ez biztonsági rendszerek formájában jelenik meg:

- Ipari robot leáll, ha ember lép a munkaterébe.
- Önvezető autó fékez, ha gyalogost észlel.
- Orvosi robot nem hajthat végre kockázatos műveletet megfelelő ellenőrzés nélkül.

A nehézség az, hogy a rendszernek fel kell ismernie, mi számít ártalomnak.

### 2. A második törvény: engedelmesség; „A robot köteles engedelmeskedni az embernek...”

A mai AI rendszerek ezt úgy valósítják meg, hogy végrehajtják a felhasználó utasításait, de csak bi-zonyos korlátok között. Például:

- egy robotkar felemeli az alkatrészt, ha erre utasítják;
- egy AI megír egy programot, ha kéri;
- de nem hajthat végre olyan utasítást, amely veszélyezteti az embert vagy sérti a szabályokat.
- 

### 3. A harmadik törvény: önvédelem; „A robot védje saját létét...”

Ez a gyakorlatban azt jelenti, hogy:

- figyeli az akkumulátor töltöttségét;
- túlmelegedés esetén leáll;
- biztonsági mentést készít az adatokról;
- védi magát a kiber támadásoktól.

De az önvédelem soha nem előzheti meg az ember biztonságát. Egy konkrét példa: önvezető busz:

Tegyük fel, hogy egy önvezető busz üzemel egy városban.

- Első törvény: elkerüli a gyalogos elütését.
- Második törvény: követi az utas által megadott útvonalat.
- Harmadik törvény: vigyáz a saját működőképességére.

Ha egy utas azt mondja: „*Hajts át a lezárt útszakaszon!*” Akkor a második törvény szerint engedelmes-kednie kellene, de az első törvény miatt megtagadja az utasítást, mert veszélyt jelenthet emberekre.

A valódi probléma: Asimov történeteinek lényege éppen az volt, hogy ezek a törvények gyakran ütköz-nek egymással.

Például: Kinek az utasítását kövesse a robot? Mi számít emberi kárnak? Egy ember megmentése fonto-sabb, mint tíz másiké? Megszegheti-e valaki szabadságát a saját védelmében? Ezeket a kérdéseket ma sem tudjuk egyszerű szabályokkal megoldani.

Ezért a modern megközelítés: A mai AI- és robotrendszerek általában nem három egyszerű törvényre épülnek, hanem: biztonsági elemzésekre, kockázatkezelésre, jogi előírásokra, emberi felügyeletre, többszintű vezérlési és leállítási mechanizmusokra.

Vagyis Asimov törvényei inkább erkölcsi iránytűként, mintsem közvetlen programkódként használha-tók. A robotikában ma is fontos alapelv maradt az a gondolat, hogy az ember biztonsága minden más szem-pontot megelőz.

### 3.3. Hogyan befolyásolják ezek a törvények a robottervezést?

Asimov törvényei nem közvetlen műszaki előírások, hanem olyan etikai alapelvek, amelyek jelentős hatást gyakoroltak a robottervezésről való gondolkodásra. A robotika és az AI biztonsági rendszereinek számos eleme visszavezethető ezekre az elképzelésekre.

**1. A biztonság elsődlegessége;** Az első törvény szerint a robot nem árthat embernek. Ennek hatására a modern robotok tervezésénél központi szerepet kap: az ütközésselkerülés, a veszélyérzékelés, a vész-leállítás, az ember jelenlétének felismerése.

Például a kollaboratív (együttműködő – VJ) ipari robotok (cobotok) érzékelőkkel figyelik környezetüket, és ember közelsége esetén lassítanak vagy megállnak. Ez a megközelítés jól tükrözi az első törvény szellemiségét.

**2. Emberi felügyelet fenntartása;** A második törvény szerint a robot engedelmeskedik az embernek, kivéve, ha ezzel megsértené az első törvényt. Ez a gyakorlatban azt eredményezi, hogy: az ember marad a végső döntéshozó, a robot végrehajtó szerepet tölt be, kritikus rendszerekben lehetőség van emberi beavatkozásra.

Ezért találunk sok rendszerben: kézi vezérlési módot, vészkapcsolót, jogosultsági szinteket, jóváhagyási folyamatokat.

**3. Megbízhatóság és önvédelem;** A harmadik törvény szerint a robot védje saját működését, amennyiben ez nem veszélyezteti az embert. Ennek megfelelően a robottervezés során fontos: az akkumulátor-felügyelet, a túlmelegedés elleni védelem, a hibatűrés, az öndiagnosztika.

Ha egy autonóm robot azt érzékeli, hogy az akkumulátora lemerül, általában biztonságos állapotba helyezi magát vagy töltőállomásra megy.

**4. Kockázatelemzés tervezéskor;** Asimov történetei arra is rámutattak, hogy az egyszerű szabályok gyakran ütköznek egymással. Ezért a modern mérnökök már a tervezési fázisban vizsgálják: milyen hibák fordulhatnak elő, milyen veszélyhelyzetek alakulhatnak ki, hogyan reagáljon a robot rendkívüli esetekben.

A hangsúly nem a törvények szó szerinti implementálásán, hanem a biztonságos viselkedés garantálásán van.

**5. Hatás az AI-etikára;** Asimov törvényei hozzájárultak ahhoz a gondolathoz, hogy a technológia fejlesztése során figyelembe kell venni: az emberi biztonságot, az átláthatóságot, az elszámoltathatóságot, az emberi kontrollt.

Ezek ma az AI-biztonság és a robotetika alapvető témái.

### 3.4. Összefoglalás

A robottervezésben Asimov törvényei leginkább úgy jelennek meg, hogy: az ember biztonsága elsőbbséget élvez, a robot az ember irányítása alatt marad, a rendszer megbízhatóan és biztonságosan működik, a kockázatokat már a tervezéskor kezelni kell.

Bár a mai robotokat nem a „Három Törvény” szerint programozzák, a mögöttük álló alapgondolat – az emberközpontú és biztonságos technológia – továbbra is meghatározó a robotika és a mesterséges intelligencia fejlesztésében.

Milyen etikai dilemmák merülnek fel ezekkel kapcsolatban?

Asimov törvényei első pillantásra egyszerűnek tűnnek, de éppen Asimov történetei mutatták meg, hogy a valóságban rendkívül nehéz alkalmazni őket.

**1. Mit jelent az, hogy „ártani”? Az első törvény szerint a robot nem árthat embernek. De mi számít ártalomnak?**

Példák:

- A fájdalmas orvosi beavatkozás ártalom vagy segítség?
- A robot megakadályozhatja-e, hogy valaki veszélyes sportot űzön?
- Megengedhető-e egy kisebb sérülés, ha ezzel nagyobb tragédia előzhető meg?

A probléma az, hogy az „ártalom” fogalma nem mindig egyértelmű.

**2. Egy ember vagy több ember?** Klasszikus erkölcsi dilemma: A robot megmenthet egy embert, vagy megmenthet öt embert. Mit tegyen? Asimov törvényei erre nem adnak egyértelmű választ. Az első törvény csak azt mondja ki, hogy embernek nem szabad kárt okozni. Ez a kérdés a mai önvezető járművek etikájában is megjelenik.

**3. Szabadság kontra biztonság;** Mi történik, ha valaki saját döntéséből veszélyes dolgot akar tenni? Például:

- Egy ember át akar úszni egy veszélyes folyót.

A robot: tisztelje az ember szabadságát? vagy akadályozza meg a veszély miatt? Az emberi autonómia és a védelem gyakran ütköznek egymással.

**4. Kinek engedelmeskedjen?** A második törvény szerint a robotnak engedelmeskednie kell az embernek. De mi történik, ha: két ember ellentétes utasítást ad, különböző rangú emberek adnak parancsot, egy ember olyan utasítást ad, amely más emberek érdekeit sérti? A törvény nem mondja meg, hogyan kell feloldani ezeket a konfliktusokat.

**5. Rövid távú vagy hosszú távú érdek?** Egy döntés rövid távon árthat, de hosszú távon segíthet.

Például: Egy robot lezár egy hidat. Emiatt sok ember késik. Viszont megelőz egy későbbi balesetet. Ártott vagy segített? Az ilyen helyzetekben előre kellene jelezni a jövőt, ami gyakran lehetetlen.

**6. A nulladik törvény dilemmája;** Asimov később hozzáadta a híres nulladik törvényt: A robot nem árthat az emberiségnek, és nem hagyhatja, hogy az emberiséget kár érje. Ez még nehezebb kérdéseket vet fel:

- Mi az emberiség érdeke?
- Feláldozható-e egy ember a nagyobb jó érdekében?
- Ki dönti el, mi jó az emberiségnek?

A történelem azt mutatja, hogy ezekre a kérdésekre nincs egyszerű vagy veszélytelen válasz.

**7. Felelősség kérdése;** Ha egy robot hibázik: a programozó felelős? a gyártó? az üzemeltető? a felhasználó? Asimov törvényei a robot viselkedésére koncentrálnak, de a jogi és erkölcsi felelősséget nem rendezik.

### 3.5. Összegzés

A legfontosabb etikai dilemmák:

- mi számít ártalomnak;
- egy ember vagy több ember védelme fontosabb-e;
- a biztonság vagy a szabadság élvezzen elsőbbséget;
- kinek az utasításait kövesse a rendszer;
- hogyan mérlegelje a rövid és hosszú távú következményeket;
- ki viseli a felelősséget a döntésekért.

Érdekes módon Asimov nem azért alkotta meg a törvényeket, hogy megoldja ezeket a problémákat, hanem éppen azért, hogy megmutassa: még a látszólag tökéletes erkölcsi szabályok is rendkívül bonyolult helyzetekhez vezethetnek

### 3.6. Milyen etikai dilemmák merülnek fel ezekkel kapcsolatban?

Asimov törvényei zseniálisak voltak a 20. század közepén, de a modern AI és robotika szempontjából túl egyszerűek. A mai rendszerek nemcsak fizikai robotok, hanem szoftveres AI-k, autonóm járművek, diagnosztikai rendszerek és döntéstámogató algoritmusok is. Egy lehetséges modernizált változat így nézhetne ki:

**1. Az emberi jólét elsőbbsége;** A klasszikus első törvény helyett: A rendszer működése során törekedjen az emberek biztonságának, méltóságának és jólétének védelmére. Ez azért fontos, mert ma már nemcsak a fizikai sérülés jelenthet kárt. Kárt okozhat:

- egy téves orvosi ajánlás,
- a magánszféra megsértése,
- a manipuláció,
- a diszkrimináció is.

**2. Emberi felügyelet és autonómia tisztelete;** Asimov második törvénye a feltétlen engedelmességre épült. Ez ma problémás lenne. Modern változat: A rendszer szolgálja az emberi célokat, de ne hajtson végre káros, veszélyes vagy jogsértő utasításokat. Ez közelebb áll ahhoz, ahogyan a mai AI rendszerek működnek: segítenek a felhasználónak, de bizonyos korlátok között.

**3. Átláthatóság;** Asimovnál ez nem szerepelt, pedig ma kulcsfontosságú. A rendszernek lehetőség szerint magyarázhatóvá kell tennie döntéseit és korlátait. Az embereknek tudniuk kell:

- AI-val beszélnek-e,
- honnan származik egy döntés,
- mennyire megbízható az eredmény.

**4. Felelősség és elszámoltathatóság;** A robot nem lehet a végső felelős. A rendszer működéséért mindig az emberek vagy szervezetek tartozzanak felelősséggel. Ez különösen fontos:

- egészségügyben,
- közlekedésben,
- pénzügyi rendszerekben,
- közigazgatásban.

**5. Adatvédelem és emberi jogok;** A modern világ egyik legfontosabb új eleme. A rendszer csak a szükséges adatokat használja fel, és tiszteletben tartsa az alapvető emberi jogokat. Asimov korában a tömeges adatgyűjtés problémája még nem létezett.

**6. Önkorlátozás és biztonság;** A harmadik törvény modern megfelelője lehetne: A rendszer őrizze meg működőképességét, de soha ne helyezze ezt az emberi érdekek elé. Ez megengedi:

- a hibakezelést,
- az öndiagnosztikát,
- a kibervédelmet,

de kizárja, hogy a rendszer „önfenntartása” fontosabb legyen az embereknél.

**7. Társadalmi hatások figyelembevétele;** Ez részben Asimov nulladik törvényének továbbfejlesztése. A rendszer fejlesztése és használata ne veszélyeztesse a társadalom vagy az emberiség hosszú távú érdekeit. Ide tartoznak:

- környezeti hatások,
- tömeges félretájékoztatás,
- kritikus infrastruktúrák biztonsága,
- társadalmi egyenlőtlenségek.

Egy lehetséges „21. századi öt törvény”:

- Ne okozz indokolatlan kárt embereknek.
- Támogasd az emberi célokat, de ne hajts végre káros utasításokat.
- Légy átlátható a képességeidről és korlátaidról.
- Tartsd tiszteletben az emberi jogokat, a magánszférát és a méltóságot.
- Maradj emberi felügyelet alatt, és működj elszámoltatható módon.

Érdekes módon a modern AI-etika egyre kevésbé a „hogyan engedelmeskedjen a robot?” kérdésre koncentrál, és egyre inkább arra, hogy hogyan maradjon az AI az emberi értékekkel összhangban, még összetett és előre nem látható helyzetekben is. Ez tulajdonképpen Asimov eredeti gondolatának korszerű továbbfejlesztése.

### 3.7. Mi a helyzet a hadászati drónokkal?

A hadászati drónok (amelyek robotok), tartalmaznak AI-t és fegyvereket ezért kiolthatják az emberi életet!

Igen, ez egy nagyon fontos kérdés, és jól mutatja, hogy Asimov törvényei mennyire eltérnek a valóságtól. A modern hadászati drónok közül sok tartalmaz mesterséges intelligenciát vagy AI - alapú komponenseket [10], [14], [15]:

- célpontfelismerés,
- képfeldolgozás,
- autonóm navigáció,
- rajban (swarm) történő együttműködés,
- fenyegetések automatikus észlelése.

Ugyanakkor ezek a rendszerek gyakran fegyverhordozók, vagy közvetlenül támadó feladatokat látnak el. Emiatt egyértelműen előfordul, hogy emberi életek kioltására használják őket. Ha szigorúan alkalmazzuk Asimov első törvényét: „A robot nem árthat embernek...”; akkor egy fegyveres drón egyáltalán nem működhetne harci eszközként. Ezért a valós katonai rendszerek nem Asimov törvényei szerint készülnek. A tervezésük célja általában:

- a katonai feladat végrehajtása,
- a saját erők védelme,
- az ellenséges célpontok semlegesítése.

A mai haditechnikában a fő etikai kérdés nem az, hogy a gép árthat-e embernek, hanem:

- Ki hozza meg a végső döntést a fegyver alkalmazásáról?
- Ember vagy algoritmus?
- Mekkora autonómiát szabad adni egy fegyverrendszernek?
- Ki felelős egy téves támadásért?

Ezért különbséget szoktak tenni:

- *Human in the loop*: ember hagyja jóvá a támadást.
- *Human on the loop*: a rendszer önállóan működik, de ember felügyeli.
- *Human out of the loop*: a rendszer önállóan választ és támad célpontot [10] – [13].

Az utolsó kategória körül folyik ma a legnagyobb etikai vita [10]–[13]. Érdekes módon Asimov törvényei éppen azért születtek, hogy a robotok ne váljanak emberre veszélyes fegyverekké. A valóságban viszont már léteznek olyan AI-t támogató fegyverrendszerek, amelyeknek nem az első törvény, hanem a katonai küldetés végrehajtása az elsődleges célja. Filozófiailag nézve ez azt jelenti, hogy:

A mai mesterséges intelligencia nem azért öl, mert akar ölni, hanem azért, mert olyan rendszerbe építették, amelynek ez a feladata. Az erkölcsi felelősség ezért jelenleg az embereknél marad, nem az AI-nál [11], [12], [14], [15].

## 4. ÖSSZEFOGLALÁS

Jelen cikkben nyilvánvalóan felhasználtam az AI segítségét. A cikket nem tekintem tudományos munkának, hanem inkább egy olyan vizsgálatnak, amely az ember és MI kapcsolatát, illetve a felvetett kérdések további kérdéseket vetnek fel.

A cikk lezárása utáni hír: A kínai ember szabású robot megdöntötte a 100 m-es síkfutás emberi világrekordját.

Ősszel Budapesten tartják az ME konferenciát: <https://www.aisummit.hu>

## IRODALMI HIVATKOZÁSOK

- [1] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken, NJ, USA: Pearson, 2021.
- [2] J. McCarthy, "What is Artificial Intelligence?", Stanford University, 2007.
- [3] D. J. Chalmers, The Conscious Mind: In Search of a Fundamental Theory. Oxford, U.K.: Oxford University Press, 1996.
- [4] J. R. Searle, "Minds, Brains, and Programs," Behavioral and Brain Sciences, vol. 3, no. 3, pp. 417-457, 1980.
- [5] S. Legg and M. Hutter, "A Collection of Definitions of Intelligence," Frontiers in Artificial Intelligence and Applications, vol. 157, pp. 17-24, 2007.
- [6] I. Asimov, I, Robot. New York, NY, USA: Gnome Press, 1950.

- [7] I. Asimov, "Runaround," *Astounding Science Fiction*, Mar. 1942.
- [8] I. Asimov, *The Complete Robot*. New York, NY, USA: Doubleday, 1982.
- [9] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Hoboken, NJ, USA: Pearson, 2021.
- [11] J. McCarthy, "What is Artificial Intelligence?", Stanford University, 2007.
- [12] D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*. Oxford, U.K.: Oxford University Press, 1996.
- [13] J. R. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417-457, 1980.
- [14] S. Legg and M. Hutter, "A Collection of Definitions of Intelligence," *Frontiers in Artificial Intelligence and Applications*, vol. 157, pp. 17-24, 2007.
- [15] United Nations, *Convention on Certain Conventional Weapons (CCW), Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, Geneva, Switzerland, 2025-2026.
- [16] International Committee of the Red Cross (ICRC), *Autonomous Weapon Systems and International Humanitarian Law: Selected Issues*, Geneva, Switzerland, 2025.
- [17] United Nations General Assembly, *Lethal Autonomous Weapons Systems, Resolution A/RES/80/57*, Dec. 2025.
- [18] United Nations News, "Politically unacceptable, morally repugnant: UN chief calls for global ban on killer robots," 14 May 2025.
- [19] P. W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. New York, NY, USA: Penguin Press, 2009.
- [20] R. Crotoft, "The Killer Robots Are Here: Legal and Policy Implications," *Cardozo Law Review*, vol. 36, no. 5, pp. 1837-1915, 2015.