

Doktori (PhD) értekezés-tervezet

Fehér András Tibor

2025

NEMZETI KÖZSZOLGÁLATI EGYETEM

Hadtudományi Doktori Iskola

Fehér András Tibor

A mesterséges intelligencia
humán aspektusai a védelmi szférában

Doktori (PhD) értekezés

Témavezető

Dr. habil Négyesi Imre
egyetemi docens

Budapest, 2025

Rövid tartalomjegyzék

BEVEZETŐ	3
A KUTATÁS BEMUTATÁSA	3
A KUTATÁS TERVEZÉSE	5
KUTATÁSI HIPOTÉZISEK, AZ AZOKBÓK TERVEZETT EREDMÉNYEK ÉS FELHASZNÁLÁSOK	8
A KUTATÁS MEGVALÓSULÁSI TERVE	9
ELSŐ RÉSZ: ÁTTEKINTŐ ELEMZÉS	13
I. EMBERI ÉS GÉPI INTELLIGENCIÁK.....	14
I.1. A MESTERSÉGES INTELLIGENCIA FOGALMÁNAK JELENE.....	14
I.2. AZ INTELLIGENCIA, AMI MESTERSÉGES IS LEHET.....	24
I.3. ÉRZELMI INTELLIGENCIA A GÉPEKBEN	32
I.4. RÉSZÖSSZEFOGLALÁS: HOGYAN LEHET AZ INTELLIGENCIA MESTERSÉGES	42
II. TECHNIKAI KÖRÚT A MESTERSÉGES INTELLIGENCIA KÖRÜL.....	44
II.1. MIKT: AZ MI-HEZ SZOROSAN KAPCSOLÓDÓ TECHNOLÓGIÁK.....	44
II.2. GÉPI TANULÁSI MODELLEK.....	53
II.3. ELVEK ÉS TECHNOLÓGIÁK AZ MI KÖRÜL.....	62
II.4. RÉSZÖSSZEFOGLALÁS: MILYEN AZ MI KÍVÜL-BELÜL?.....	79
MÁSODIK RÉSZ: SPECIFIKUS VIZSGÁLATOK	80
III. AZ MI NYÁRBAN NYARALHAT-E AZ EMBER?.....	80
III.1. AZ MI TELEI ÉS TAVASZAI	80
III.2. A KORSZAKOK ELTÉRÉSEI	85
III.3. AZ MI ÉS TÁRSADALOM KÖLCSÖNHATÁSAI ÉS VISSZAHATÁSAI	90
III.4. HIDEGFRONTOK AZ MI NYÁRBAN.....	97
III.5. A VÉGES PARADIGMAVÁLTÁSI KÉPESSÉG ÉS EGYÉB HUMÁN TÉNYEZŐK.....	102
III.6. A TOVÁBBLÉPÉS LEHETŐSÉGEI	112
III.7. ÖSSZEGZÉS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI-NYÁR ÚJ MEGHATÁROZÁSA.....	116
IV. AZ ERKÖLCSI ÉRZÉK DIGITALIZÁLHATÓSÁGA.....	119
IV.1. AZ AUTONÓMIA ANATÓMIÁJA.....	119
IV.2. AZ MESTERSÉGES ERKÖLCSI ÉRZÉK LEHETŐSÉGE	142
IV.3. ÖSSZEGZÉS ÉS RÉSZKÖVETKEZTETÉSEK:.....	164
V. A MESTERSÉGES INTELLIGENCIA FOGALMÁNAK ÚJRAGONDOLÁSA.....	167
V.1. FOGALMI PROBLÉMÁK A „DIGI-DOLGOKTÓL” AZ MIKT RENDSZEREKIG	167

V.2.	AZ INFORMATIKA ÉS A HUMÁNTUDOMÁNYOK VÁLTOZÁSAI AZ MI HATÁSÁRA.....	175
V.3.	JAVASLAT AZ MI MEGHATÁROZÁSÁRA	186
V.4.	ÖSSZEFOGLALÁS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI FOGALOMVIZSGÁLATA	197
VI.	MESTERSÉGES INTELLIGENCIA ÉS ERŐÉRVÉNYESÍTÉS	199
VI.1.	AZ MI-ALAPÚ HIDEGHÁBORÚ: A HIBRID MŰVELETEK SPECIÁLIS ELEME	199
VI.2.	ÉLNI ÉS VISSZAÉLNI A DIGITÁLIS TÉRBEN	211
VI.3.	A TANULMÁNY EDDIGI VIZSGÁLATAI ÉS A VÉDELEM	220
VI.4.	ÖSSZEFOGLALÁS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI VÉDELMI ÚJSZERŰSÉGE.....	236
	ÖSSZEZÉS, EREDMÉNYEK, HASZNOSÍTHATÓSÁG.....	238
	A VIZSGÁLATOK ÖSSZEZETT ÉRTÉKELÉSE	238
	EREDMÉNYEK ÉS AZOK HASZNOSÍTHATÓSÁGA	239
	AJÁNLÁS – KIK SZÁMÁRA LEHET HASZNOS A FENTI TANULMÁNY?	242
	TOVÁBBI KUTATÁSI LEHETŐSÉGEK.....	243
	MUTATÓK.....	245
	RÉSZLETES TARTALOM	245
	TÁBLÁZATOK JEGYZÉKE	251
	ÁBRÁK JEGYZÉKE	251
	SZAKIRODALMAK JEGYZÉKE	251

BEVEZETŐ

A KUTATÁS BEMUTATÁSA

Felvezetés: a „humán aspektus” és a „védelmi szféra”

A cím kifejtésével összefoglalható jelen kutatások programja, érdemes felvezetésként ezt ismertetnem. A mesterséges intelligencia (továbbiakban MI) kifejezés alatt első megközelítésben elegendő azt érteni, amit a tisztelt olvasó tud róla – ennek tisztázása ugyanis a célok között szerepel. A „humán aspektusok” kifejezés arra utal, hogy a téma az etikától a pszichológián és pedagógián át a szociológiáig számos humántudománnyal kapcsolatba kerül, sőt bizonyos pontokon még világnézeti függőségek elemzése is szükségessé válik. (Említésre méltó alternatíva volt a „kognitív tudomány” [1]¹ kifejezés, ám ez a szükségesnél sokkal tágabb kört fed le.)

Hivatásom miatt kutatásaimat általában igyekszem nemzetvédelmi perspektívába helyezni. Jelen vizsgálataim egyértelműen rámutattak arra, hogy a katonai kifejezések (hadi, katonai, harci stb.) használata nem jól fedné le a kutatás szemléletét, sőt félrevezető lenne. Olyan kifejezést kerestem, amely a hagyományos katonai szegmensen túlmutatva utal az ország védelmének résztvevőire. Mivel nemzetbiztonsági szakember sem vagyok, ezért tartom a nemzet szó nélküli „védelem” kifejezést az erre megfelelő kompromisszumnak. Amint látható lesz, ez a kifejezés jó választás annak ellenére, hogy a köznyelvi „védelmi szféra” bizonyos részeire (rendészet, katasztrófavédelem, gazdasági, pénzügyi stb.) a tanulmány csak érintőlegesen tér ki.

Összefoglalva: az MI, a humán aspektus és a védelmi szféra kifejezések közös halmaza ugyan túlmutat az értekezés keretein, metszetük azonban ennek tartalmára utal.

Tanulmányom középpontjában inkább az nagyon sokrétű átalakulás van, melyre az MI rápróbálja kényszeríteni az amúgy is gyorsan változó világot. Ennek néhány kis szeletét elemzem, s bár ezt elsősorban filozófiai, informatikai és nemzetvédelmi oldalról végzem, de bízom abban, hogy olyanok is beleolvasnak, akik nem informatikai szakemberek, vagy nem katonák, vagy nincs filozófiai vénájuk.

A kutatást 2024 december elsején lezártam. Ez nem jelenti azt, hogy egy-egy frissebb forrást ne dolgoztam volna bele a szerkesztési munkák során.

¹ V.2.2-ben részletesebben elemzem, de röviden: a kognitív tudomány egy interdiszciplináris megközelítés, az idegtudomány, nyelvészet, antropológia és minden olyan tudomány kutatóinak interdiszciplináris területe, amely az elme, a tudat, a tudatosság megértésére törekszik.

A téma fontossága, időszerűsége

Az MI vizsgálatának jelentősége ugyan vitathatatlan, viszont tisztázandó, hogy aktuálisak és lényegesek-e ennek humán aspektusai, illetve az így kapott kérdések és válaszok lényegesek-e a védelmi szférában.

Kezdve az időszerűséggel elmondható, hogy ez nem csupán az MI terjedésének tudható be. Hanem „most van itt az ideje”. Egyrészt azért, mert az MI terjedése által már láthatóvá vált annak a társadalommal alakuló kapcsolata, tehát napjainkban már a maga dinamikájában vizsgálható, az, ami tíz éve még nem. Másrészt a korábbi technológiák társadalmi (kölsön-)hatásának fényében vizsgálódhatunk, ami a húsz vagy száz évvel ezelőtti hasonló felvetésekhez képest előnyös. Tehát most a technológia kifejlődésének olyan korai fázisában fogalmazunk meg ajánlásokat, amire a korábbi technikai forradalmak során nem volt lehetőség. Okulhatunk például azon társadalmi, környezetvédelmi, etikai változások mintázataiból, melyeket a korábbi technológiák elterjedése idézett elő. Ezeket felismerve talán elébe tudunk menni az MI negatív hatásainak, és jól tudjuk kihasználni a benne rejlő potenciálokat. Végül a hasonló tudományos felismerések mielőbbi terjedése járulhat hozzá a védelmi szférában ahhoz a paradigmaváltáshoz, mely úgymint elkerülhetetlen.

Az időszerűséggel kapcsolatban leírtak a téma fontosságának indoklasként is értelmezhetőek, de nézzük meg a cím részeit külön-külön. A humán aspektusok jelentőségének indoklásához nem szükséges szakirodalmi alátámasztás, hiszen triviális, hogy a kognitív képességünk gépi másolása és kiterjesztése annyira új, hogy emiatt új alapokra szükséges helyezni mindent. Vagyis újra kell gondolni az emberképünket (sőt az élet meghatározását is), és az ehhez kapcsolódó fogalmakat (okosság, tanulás, érzelem stb.). Ennek a nagy újragondolásnak minden része fontos, viszont ez nem várható el a matematikusoktól vagy mérnököktől, akik magát a technológiát valósítják meg. Ezért van szükség ilyen humán megközelítésekre is.

A védelmi vonatkozások lényegességére akár neves szerzőket idézhetek, akik azon a véleményen vannak, hogy az emberiség történetének eddigi legnagyobb geopolitikai forradalma az a digitális forradalom, amelyet többek között az MI terjedése tesz lehetővé. [2, o. 24]. De a józan ész által is belátható a védelmi jelentőség, hiszen minden jó találmánynak van árnyoldala, vissza lehet vele élni. Tehát kézenfekvő, hogy az MI használata maga is kockázatot jelent, vagyis vizsgálandó. A jelenséget erősíti, hogy egyre több állami feladat digitalizálódik, továbbá az, hogy az egyre élethűbb gépi utánpótlás folyamatosan tágítja egy állam, illetve államrész elleni visszaélés lehetőségi körét.

A fentiek alapján világos, hogy az MI humán aspektusainak védelmi vizsgálata fontos adalékokkal járulhat hozzá a fejlődés emberibbé tételéhez és országunk biztonságához. Emellett csak hazai kutatásokkal lehet a magyar sajátosságokra, hazai védelmi vagy katonai szempontokra megfelelően adaptálni a külföldi szakirodalmakban leírt eredményeket.

A kutatás személyes motivációja

A kutatást nagyban segítette, hogy sikerült motivációt találnom: ebben a kutatásban próbáltam meg egyesíteni érdeklődési köreim korábbi széttagoltságát. Három irányt tekintve sikerült ezt megtennem. Az egyik a tudományos vívmányok, a technológiák, elsősorban az informatikai megoldások iránti érdeklődésem, aminek mentén első egyetemi diplomámat szereztem informatika szakon. A második az az érdeklődésem, ami az elvont okokat kutatja a társadalom és a világ folyamataiban, ami miatt humán szakon is szereztem végzettséget (teológus licenciátust). A harmadik a lelki küzdelmek testi vetülete, a harc és a túlélési technikák iránti érdeklődésem, melyet fiatalkoromban a küzdősport és erdei élet kihívásai vittek előre, amely később katonai hivatássá érett bennem, melyhez később hazánk védelmi működése iránti érdeklődés is társult. Mivel ez a három terület össze tudott kapcsolódni ebben a megközelítésben, ezért a kutatások korai fázisától számomra belsőleg is építő jellegűvé vált ez a munka. Ezért bízom abban, hogy a bennem rendhagyó módon összekapcsolódó területek szintézise az olvasók számára is inspiratív és érdekes lesz.

A KUTATÁS TERVEZÉSE

Bár a kutatás elméleti, mégis a gyakorlatiasság jegyében már a kutatás tervezésekor az eredmények majdani felhasználhatóságát kívántam szem előtt tartani. Ennek érdekében a kutatási probléma általános leírása és a téma lehatárolása után azokat a gyakorlatias elvárásokat, valamint prioritásokat (szempontokat) rögzítettem, melyeket végig szem előtt szeretnék tartani. A célokat ezen elvárások és szempontok jegyében kívántam kijelölni.

A kutatási problémák és a téma lehatárolása

Kutatási problémám általános megfogalmazása a következő:

Korunk és a közeljövő kulcstechnológiája, az MI erős hatást gyakorol az emberekre és a társadalomra, illetve a társadalom is hat a technológiára. Ezáltal a kölcsönhatás által a világ változásának jellege is megváltozni látszik. Vagyis kezd másképp változni a világ. Voltaképpen annak az eltérésnek a mibenléte lenne tisztázandó, hogy a most zajló változások miben térnek el az eddigi, nem kognitív technikák által generált változásoktól.

Világos volt, hogy ezt a „változó változást” csak néhány vonatkozásban vizsgálhatom. Ezért már a kutatás tervezése elején, addigi ismereteim fényében a következő, logikailag ide kapcsolódó témákkal maximum érintőlegesen foglalkozok.

- A kognitív hadszíntér – később kerülhet majd alapos vizsgálatra, itt érinteni szeretném.
- Az emberbe épített agyi implantátumokon alapuló rendszerek eredményei – csak érintőlegesen szeretném vizsgálni;
- Jogi vizsgálatok – csak a technikával való visszaélések védelmi oldalát vizsgálom, egyéb vonatkozásokban nem szeretnék elmélyedni;
- Gazdasági elemzések – legfeljebb érintőlegesen vizsgálom;
- A védelem katasztrófavédelmi oldala – nem szeretnék rá kitérni;
- Rendszereti vagy titkosszolgálati aspektusok – csak érintőlegesen szeretném vizsgálni;
- Jelen munkával nem kívánok a technológiák terén tudományos eredményt felmutatni (bár a jövőbeni terveim között szerepel MI rendszerek építése és tesztelése).

Ide tartozik annak deklarációja is, hogy ezt a népszerű témát nem szeretném a populáris, félelemre építkező módszerrel megközelíteni. Sőt, részben cáfolni is igyekszem azokat a megközelítéseket, melyekre a disztópikus filmek és bulvárcikkek épülnek. Ám ahogyan magát az MI-t is fel lehet használni jó és rossz célokra is, úgy a róla szóló elemzéseket is. Ezért az itt leírtakat – főleg összefüggéseikből kiragadva – sajnos fel lehet használni félelemkeltésre is. Ezekről előre elhatárolódom.

Bizonyos témák nem fértek bele a keretekbe, eredményeiket azonban részben felhasználom. Ezek közül említendő az emberi tanulás paradigmaváltását (tanulás 2.0) elemző tanulmányom [2]. Több publikációt is eredményezett „MI kibervédelmi és kibertámadási használhatósága” téma vizsgálata (ezekre a felhasználáskor hivatkozok).

Hasznosíthatósági irányelvek (gyakorlati célok)

A kutatási problémákkal összefüggésben a kutatás egészére (nem csak az eredményekre) vonatkozó pragmatikus célok a következők:

1. A kutatás hasznosítható kell, hogy legyen elsősorban
 - a. az államvédelmi tudományok,
 - b. a hadtudományok és
 - c. az oktatás területén,
 - d. egyéb területeken.

2. A vizsgálat során törekedni kell arra, hogy az MI technológiák minél több olyan részterületét érintsem, ahol magyarul nem jelentek meg publikációk, vagy esetleg világviszonylatban is fehér folt a tudományos kutatásokban, hogy ezek a részkutatások is növeljék a dolgozat további felhasználhatóságát.
3. 1.C szemponthoz kapcsolódva a technológia fejlődésének belső összefüggéseit úgy kell bemutatni, hogy az egy egyetemi tananyaggá könnyen átdolgozható legyen; amellet, hogy ennek anyagára a további fejezetek is ráépülnek.

A kutatás kiemelt szempontjai

Tovább pontosítható a vizsgálat látóköre a fő szempontok (prioritások) rögzítésével.

Interdiszciplináris megközelítés. Nem technikai (matematikai, mérnöki vagy programozási) eredményt várok, ezért itt a filozófia, a számítástechnika, a pedagógia és a pszichológia területén végzett vizsgálódásokat kapcsolok más tudományterületekkel, szorosabban a nemzetvédelmi tudományok, a szociológia, illetve a tudománytörténet egyes részeivel.

Előtérben a háttér. A vizsgálat ne csak az MI által generált jelenségekre összpontosítson, hanem ezek lehetséges háttérére is adjon következtetéseket.

Prevenció. A téma időszerűségénél fentebb leírtak alapján ez a szempont kiemelendő.

Gyakorlatiasság a részletekben. A fentebb említett hasznosíthatósági célokat nem csupán a tervezett eredmények szintjén, hanem a kisebb részeredmények (saját fogalmak, megfigyelések, ötletek, javaslatok) eszközeivel is szeretném megalósítani.

Egzisztenciális megközelítés. A gyakorlatiasság égisze alatt érvényesíteni szeretnék egy olyan szempontot is, hogy a vizsgálatok eredményeit az olvasó a saját életében is alkalmazhassa. Például amikor azt próbálok érzékeltetni, hogy a történelem egyre haloványabban, egyre részlegesebben ismétli önmagát, ezzel rámutatok arra is, hogy egyre kevésbé használhatóak a megszokott megközelítéseink vagy szokásaink. Mert lehet a világ változásait nem szeretni, de csak ezeket figyelembe véve lehet boldogulni benne, illetve tenni azért, hogy élhetőbbé váljon: ez tehát egyben közös ügyünk is.

Elméleti célok és kérdések

A fenti tervezési gondolatok figyelembevételével jelöltem ki az alábbi, konkrét elméleti kutatási célokat (C1-C4). Az áttekinthetőség kedvéért egyben ismertetem az ezekhez tartozó kutatási kérdéseket (K1-K4) is. Ezek sorrendje kijelöli a tanulmány logikai ívét is.

Első témakör: lesz-e MI tél?

C1: A MI fejlődési irányainak vizsgálata a közeljövőben, részben a technológiák, de inkább a felhasználói oldal elemzésével.

K1: Vajon tovább gyorsul az MI fejlődése, vagy esetleg lassulni fog?

Második témakör: mesterséges erkölcs és szabadság

C2: Megvizsgálni az erkölcsi érzék digitalizálhatóságát.

K2: Képes lehet-e egy mesterséges neurális hálózat elfogadható színvonalú erkölcsi érzék megvalósítására?

Harmadik témakör: terminológiai kérdések

C3: Az MI-vel kapcsolatos terminológiai hiányosságok közül minél több tényezőt azonosítani, és ezek feloldására javaslatokat tenni (új fogalmak, felosztások, pontosítások).

K3: Milyen fontos vonatkozások maradnak ki az MI ismert meghatározásaiból, és ezek közül melyeket lenne szükséges beépíteni egy új definícióba?

Negyedik témakör: az MI hatása a védelmi szférára

C4: Néhány olyan markánsan meghatározó vonás beazonosítása, melyekben az MI jelentős hatással van korunk katonai és védelmi szférájára (a megelőző fejezetek védelmi aspektusaira alapozva).

K4: Lesz-e olyan alapvető hatása az MI-nek az állami katonai és egyéb védelmi elméletekre és rendszerekre, mint a kibertérnek?

KUTATÁSI HIPOTÉZISEK, AZ AZOKBÓK TERVEZETT EREDMÉNYEK ÉS FELHASZNÁLÁSOK

Hipotéziseimet úgy fogalmaztam meg, hogy eleve kirajzolódhasson, milyen tudományos eredményeket várok tőlük, és ami még fontosabb, hogy ezek a tervezett eredmények hogyan válhatnak hasznosíthatóvá. Így a hipotézisek vizsgálata eleve egy pragmatikus keretbe kerül.

1. hipotézis (H1): A mesterséges intelligencia fejlődésében lassulás várható.

Tervezett eredmény: az MI fejlődés jelenlegi ijesztő tempójának folytatódásáról szóló vélemény cáfolata.

Tervezett felhasználás: A lassuló fejlődés forgatókönyvének kockázatelemzése alapján, az így előálló szituáció kihasználási lehetőségeinek feltérképezése.

2. hipotézis (H2): Egy mesterséges neurális hálózat képes lehet ugyan bizonyos erkölcsi érzék jellegű kompetencia megvalósítására, de nem olyan minőségben, ami egy általános MI-t emberszerűvé tenne.

Tervezett eredmények: a neurális gépek erkölcsi határainak megállapítása elsősorban filozófiai szinten (mely jogi és egyéb etikai irányokba adaptálható). Ennek fényében az úgynevezett „általános MI”-vel szembeni túlzó elvárások racionalizálása.

Tervezett felhasználások: az autonóm járművek és fegyverek területén várható számos felhasználás, mivel a mesterséges erkölcsi érzékkel kapcsolatos korlátok határt szabnak a gépi autonómiának is.

3. hipotézis (H3): A Mesterséges Intelligencia jelenlegi meghatározásaiból olyan fontos aspektusok hiányoznak, melyek hiányában a technológia félreérthető, így kompromisszumok mellett is egy jelenlegieknél teljesebb definícióra lenne szükség.

Tervezett eredmények: Az MI új, használhatóbb meghatározása, mely talán a következő tíz évben is megállja a helyét, és a témában a pontosabb, szabatosabb fogalmazást szolgálja.

Tervezett felhasználások: A szabályozások és védelmi tervezés számára kezelhetőbbé válnak az MI olyan aspektusai, melyeket a jelenlegi meghatározások nem vesznek figyelembe.

Az átfogóbb meghatározás a felsőoktatásban is segítheti az MI lényegének megismertetését a nem informatikai vagy műszaki jellegű képzések során, például a katonai oktatásban.

4. hipotézis (H4): az MI tovább tolja az erőérvényesítés hangsúlyát a puha műveletek felé, ennek kezeléséhez pedig a katonai és védelmi informatika radikálisan új megközelítése lesz szükséges

Tervezett eredmények: Az MI által átrajzolt védelmi szféra leírása, ennek kezeléséhez kapcsolódó terminológiák és elvek kidolgozása. Továbbá az új megközelítések szemléletesebbé tétele a digitális erőérvényesítés új módozatainak vizsgálata révén. A szükséges új megközelítés vázlata.

Tervezett felhasználások: A szemléleti tisztázás és a hozzájuk kapcsolódó fogalmi pontosítások hasznosítása elsősorban a diszciplináris és stratégiai tervezés során, valamint hadtudományi, esetleg oktatási szempontból lesz hasznosítható.

A KUTATÁS MEGVALÓSULÁSI TERVE

A kutatás előzetes tervezésekor az I., II. és IV. hipotézisek bizonyítása megfelelő iránytűnek tűnt. Azonban a munka közben kiderült, hogy sokkal pontosabb fogalmi keretekre van szükség. Így került megfogalmazásra a III. hipotézis, valamint a hozzá tartozó vizsgálatok, tehát az Első rész is. Ezek miatt került törlésre a kutatás több része (ld. fentebb), de bízom benne, hogy ez a tanulmány aktualitásának és minőségének javára vált.

Kutatási módszerek

A fentebb kijelölt irányok a vizsgálatok módszertanát is meghatározzák. A vizsgálatok különböző fázisaiban eltérő módszerek alkalmazása volt indokolt, melyek így foglalhatóak össze:

Deduktív módszerekre és analitikus megközelítésre volt érdemes támaszkodnom a különböző elemzések elvégzésekor. Elsősorban a dokumentum és forráselemzés módszereit alkalmaztam egy-egy kérdéskör eddigi kutatási eredményeinek feltársakor. Ezen belül általában leíró módszert használtam, de a jelenségek és fogalmak esetében inkább visszatekintő elemzést, a folyamatok vizsgálatához pedig feltáró módszert alkalmaztam. A kapott információk rendszerezésére is szükséges volt, ezekre épülve, összefüggésfeltáró módszerek alkalmazásával valószínűleg meg a múlt és a jelen (vagy jövő) párhuzamosságainak és eltéréseinek azonosítása, egy-egy fogalom gépi és emberi megvalósulásának összehasonlító elemzése, vagy fogalmak megfogalmazásainak összevetése. Vizsgálati módszerként itt fontos szerephez jutott az etimológia is, hiszen az átvett régi szavak esetén sokat elárulnak az adott szavak ősbibb jelentésrétegei.

Induktív módszer nélkül viszont nehezen jönne létre eredmény. Ezért a fenti analitikus anyag szintézisére is szükség volt. Az ehhez szükséges következtetéseket sokszor diszkurzívan, a feltárt információk és saját gondolatok logikus kapcsolataként kaptam meg, máskor az információrendezés eredményeiből vagy összehasonlításokból is vontam le a konklúziókat. Némely esetben a teljes indukció alkalmazásával, vagy az állítás ellentétét cáfolva volt érdemes a felvetett következtetést bizonyítani. Ezek a logikai építmények a bizonyítások mellett jól alkalmazhatóak voltak olyan szintézisekhez, melyekre az előrejelzés és a javaslatkételek alapulnak.

Az értekezés logikai felépítése

Az imént említett pontosabb fogalmi szükséglet eredményezte, hogy a kutatás két részre tagolódik. Az Első részben az MI áttekintő elemzését végzem el, mely hármast célt szolgál:

Itt történik meg a legtöbb alapfogalom tisztázása, sok szükséges új fogalom bevezetése, továbbá sok olyan technológiai és elméleti információ közzlése, megközelítések ismeretése, melyekre a „Második rész” specifikus vizsgálatai támaszkodni tudnak.

Ez a rész úgy mutat az MI fogalmának újragondolása felé, hogy témánkként vizsgálja, szükséges-e majd az adott tényezőt egy új meghatározáskor figyelembe venni.

Ebből a részből kívánok a jövőben olyan felsőoktatási jegyzetet kivonatolni, melyet a tervezett gyakorlati hasznok között említettem.

A **Második rész** specifikus vizsgálatainak fejezetei nem csupán az első részre épülnek, hanem egymásra is. A hipotézisek sorrendjében haladva először bemutatom, hogy bár az MI terjedése szinte biztos, lényegi fejlődésében lassulás várható. Ebben a kutatásban lefektetett megfigyeléseket is felhasználok, amikor ezután az általános intelligencia és szuperintelligencia határait és az emberi erkölcsi érzék digitalizálhatóságát vizsgálom. Az erkölcsi elemzésekhez kapcsolódva a gépi és emberi autonómia összevetésének egy új szintézisét is elkészítettem.

A kutatás ezen pontjához érve már kirajzolódnak a technológia határai, erre építve térek vissza az MI fogalmának újragondolására. Végül, ezt is felhasználva jutok el a dolgozat középpontjában álló szféra vizsgálatához. Ennek során a hadtudományi és biztonsági kutatások mellett bemutatom az MI oktatásának szükségességét a katonai és védelmi képzésekben. Ezzel körbeérek, hiszen ehhez az utolsó eredményhez az Első részből összeállítható jegyzet szolgáltathatja majd az alapot.

Áttekintés a szakirodalomról

A kutatásaim során sokáig igyekeztem a hagyományos kutatómódszertan megközelítése szerint elvárt módon válogatni meg a forrásaimat. Azonban számos alkalommal éreztem úgy, hogy a tudománymetriai mérőszámok és a figyelemreméltó hatástényezővel rendelkező, magas presztízsű szakfolyóiratokban megjelent tudományos cikkek, vagy szakkönyvek, melyeket neves szerzők írtak, elmaradtak a várakozásaimtól. Ezzel szemben a hagyományos kritériumok alapján nem tudományos jellegű, oktatási, ismeretterjesztő, közérdekű, vagy kifejezetten szakmai célból közreadott, technikai jellegű közlemények – melyek alapvetően tudományosan nehezen idézhető írások – okoztak kellemes meglepetést. Például a kiadási helyet rég nem lehetne elvárni, hiszen nemzetközi csoportok által üzemeltetett internetes lapok esetében ilyen nem létezik, online meetingeken, a világ több pontjáról dolgoznak a szerkesztők. Azokat azonban kerültem, ahol a megjelenés (utolsó frissítés) idejét nem lehetett megállapítani, mivel indokolhatatlannak tartom ennek a fontos adatnak a rejtését (a portál szinte mindig tárolja, csak ez a mező nem mindig publikus). Számos komoly és érhető technológiai leírás olyan portálokon hozzáférhető, melyek klasszikus módon való idézésekor több adat is hiányzik. A bemutatott példák egy részét ilyen közreadású, de szakmai célú portálról szereztem be. Azonban a kutatás alapvető sarokpontjainál kizárólag hivatalosan is teljes mértékben elfogadható, tudományosan besorolt szakirodalomra támaszkodtam. Az alább elemzett változás a jövőben érinteni fogja a tudományosság kritériumait és az idézés formátum-előírásait is. Ezért bízom benne, hogy a tisztelt olvasóim már a jövő elvárásai szerint értékelik a szakirodalom minőségét, hiszen a leírt adathiányosságok ellenére a felhasznált szakirodalmak megfelelnek a szakmai elvárásoknak.

Stilisztikai magyarázatok

Igyekeztem számos előre és visszautalással segíteni a logikai kapcsolatokat, erre a taglalási hierarchiára vonatkozó szóhasználat a következő:

Rész: Az anyagot két részbe szerkesztettem. Az első rész áttekintésére épül a második rész specifikus vizsgálata, melyből a kutatási eredmények származnak.

Fejezet = Római számmal jelzett nagy egység

- Ezek elején egy kis bevezetésben mutatom be a bizonyítási mód vázlatát, és hogy hogyan kívánom benne kezelni az oda tartozó célt, hipotézist és kutatási kérdéseket.
- Ezek végén egy külön fejezetben bemutatom a cél teljesülését, választ adok a kutatási kérdésre, összegzem a hipotézis bizonyítását és a kisebb-nagyobb eredményeket

Alfejezet = a részek fejezetekre tagolódnak, jelölésük pl. IV.3

Kisfejezet = a fejezeteken belül összefüggő logikai egység, jelölésük IV.3.2.

- Az alfejezetek és kisfejezetek végén általában nincs szükség külön összefoglalásra, de az I. részben néha ott emelem ki az MI-fogalom terén tett előrelépéseket.

Szakasz = bizonyos esetekben a hivatkozhatóság és a követhetőség miatt az alfejezeteket érdemes volt felosztani, de ezeket nem számoztam automatikusan.

Címekkel ellátott listák. Használok egy kevésbé elterjedt azonosítástípust is, ahol egy félkövér rövid cím egy bekezdésbe van szedve az azt értelmező tartalommal. Kétféle esetben használom ezt: egyrészt amikor elegendő egy címet egy-két bekezdés hosszúságban tárgyalni, másrészt akkor, ha egy felsorolás pontjai pár szónál hosszabbak. Ez utóbbi esetben nagyban segítik a felsorolás áttekinthetőségét (főleg visszautaláskor) a rövid címek, melyek azonban magyarázat nélkül érthetetlenek lennének.

Megemlítendő, hogy bizonyos kifejezések értelmét az első előfordulásnál nem fejtem ki, pontosabb magyarázatára többször előre utalok. Ma már fontos stilisztikai célkitűzés (lassan elvárás) kell, hogy legyen az is, hogy egy szöveg senkit ne emlékeztessen egy MI által generált tartalomra. Igyekeztem ezért az érthetőséget a tudományos igényességgel úgy egyesíteni, ahogyan a gép nem tudná.

ELSŐ RÉSZ: ÁTTEKINTŐ ELEMZÉS

A tanulás, az intelligencia és az autonómia az emberek és a gépek világában egyaránt szorosan összefügg. Az élőlények fennmaradását tanulási képességeik biztosítják, különböző szintű és típusú intelligenciáik ennek a képességnek a kibontakozásai. Az állatok és az ember intelligenciájuk alapján képesek a véletlentől eltérő viselkedés valamely szintjére, különböző ösztöneik közötti döntésre, az ember esetében ezeken túl akár különböző gondolatmenetek összevetésére is, tehát az intelligenciájuk mértékében van szabadságuk. Hasonlóan a gépeknél: a gépi tanulás (*machine learning*, ML) legfontosabb tulajdonsága, hogy általa egy rendszer önmagára is erős aktivitást fejt ki, és ezáltal változik, alakul.² Ebben tér el lényegileg a korábbi automatikus megoldásoktól, melyek inkább a világ felé aktívak. A tanulási képesség által tud kibontakozni egy bizonyos szintű mesterséges okosság (ez indokolja, hogy részletesebben foglalkoztam a főbb tanulási modellekkel II.2.). A gépi intelligencia ilyen nézőpontból csupán egyik következménye, egy passzív eredménye a gépi tanulásnak. Amikor pedig a tanulásfüggő intelligens képesség aktivizálódik, vagyis a gép addig virtuális döntései tettekbe fordulnak, akkor egyfajta szabadság (autonómia) jön létre a gépek esetében is.

Ez a három terület jelen tanulmányban is kulcsszerepet kap, azonban a kifejtés logikája nem engedte meg a fent vázolt oksági sorrendet. Először ugyanis világossá kell tenni, hogy mit is szokás MI alatt érteni, valamint azt, hogy mit is szokás intelligencia alatt érteni (I.). Ezután pontosítani kell tenni azt a technológiai kört, amely jelenleg szorosan összefonódik az MI-vel, el kell kicsit merülni a gépi tanulás modelljeiben, valamint vázolni technológiákat és elveket (II). Amint fentebb (Az értekezés logikai felépítéséről szólva) említettem, a tanulmány áttekintő elemzésének fejezetei nem kívánnak jelentős tudományos eredményeket elérni, csupán a későbbi eredményeket alátámasztani hivatottak, de több célt is szolgálnak.

² A hagyományos programozásban is létezik önjavító kód, de az egy eltérő és nem túl gyakori technika.

I. EMBERI ÉS GÉPI INTELLIGENCIÁK

Bár az MI-ről a legtöbb embernek van bizonyos fogalma, fontos, hogy a róla szóló elmélyült vizsgálat elején tisztázzam miről is lesz szó. A gépi intelligencia vizsgálatához azonban logikusan kapcsolódik az intelligencia általánosabb elemzése. Hiszen más kutatók szerint is „*a mesterséges intelligenciát az is meghatározhatja, hogy mit csinálnak a mesterséges intelligencia kutatói. Ez a jelentés a mesterséges intelligenciát elsősorban az informatika olyan ágának tekintti, amely az intelligencia tulajdonságait az intelligencia szintetizálásával tanulmányozza*”.[3, o. 14] Az intelligencia vázlatos elemzését követi egy példaként arra, hogy a kifejezés nem csak az okossággal és IQ-val jellemezhető: ezért kitérek arra, hogy az érzelmi intelligenciát hogyan képesek utánozni az affektív számítástechnika eredményei.

I.1. A MESTERSÉGES INTELLIGENCIA FOGALMÁNAK JELENE

A jelenünk vizsgálatát érdemes a múltból indítani, tehát egy történeti áttekintéssel kezdeni.

I.1.1. Az MI fogalom megjelenése és az MI hullámai

Az emberi elme gépi utánzására irányuló munkálatok több, mint 80 éve képezik részét a tudományos kutatásnak, hiszen 1943-ban publikálták az első MI jellegű eredményt, alig hat évre rá jelent meg az első (Hebb-féle) tanulási modell, majd meg is épült az első neuronhálós számítógép 1951-ben.[4] Ekkor ez még csak egy névtelen kutatási irány volt a számos kibernetikai fejlesztés között, melyekben matematikusok, fizikusok, mérnökök együtt dolgoztak és lelkesedtek az épp létrejövő új világért. Ebben a bizakodó közegben történt meg a névadó is, amikor olyan gépek megalkotása volt a cél, melyek az emberi következtetést utánozzák le (esetleg azt meg is haladják). Az tehát, hogy a névadáskor az intellektust emelték ki a gépesítésre váró kognitív képességek közül, az főleg ennek a reáltudomány-centrikus közegnek köszönhető, amelyben a fogalom létrejött.

Egy kibernetikai konferencián 1956-ban a New Hampshire-i Dartmouth Egyetemen merült fel, hogy jó lenne a fejlesztések ezen irányát a kibernetika tudományától elkülöníteni. A John McCarthy által javasolt *artificial intelligence* fogalmat elfogadták el[5] – amely aztán máig „rajta ragadt” a technológián, mint valami gyerekkori becenév. Sokáig tökéletes kifejezésnek is tűnhetett, hiszen ekkor még legtöbbször azt várták ezektől a gépektől, hogy egy racionális utópia megteremtésében segédkeznek majd. Ennek a várakozásnak az irodalmi lecsapódása az akkori tudományos-fantasztikus irodalom, melynek legfontosabb szerzői maguk is mérnökök, csillagászok, fizikusok voltak. A névadás után vizsgáljuk meg a fogalom fejlődését és bővülését

a három nagy hullám[6, o. 51–52] ismertetésén keresztül, amelyet e téren a szakirodalom megkülönböztet.

- (1.) A kezdeti időszakban a gép érvelési képességei álltak előtérben. Ezért a szabályokon alapuló megközelítésekre összpontosítottak, technikailag ehhez a döntési fák, a bool-algebra és a fuzzy logika (ld. II.3.2) voltak meghatározóak. A XX. század közepén történt biztató indulás után sokáig „eltűnt” az MI. Laborokba bújt a haladás, termékei még nem váltak diszruptívvá. Lassú fejlődésének oka, hogy be kellett várnia a hiányzó technológiai szintet és a kognitív tudományok szükséges eredményeit. Például az 1950-es évektől az 1980-as évekig ez a korszak, ami a racionalitás győzelmében hitt, az érzelmek teljes kiiktatását a gép előnyére írta.³ Ebben az első hullámban tehát a gépek tudásának forrását még emberi szakértői tudás adta, és ellenőrzött, hiteles források alapozták meg. A fejlesztések középpontjában különféle szakértői rendszerek álltak: például erőforrás-elosztó, karbantartó vagy készletellenőrző funkciók.
- (2.) A második ciklusban kerül középpontba a gépi tanulás (*machine learning*, ML) fejlesztése és alkalmazása. A 21. század elejétől újra felívelt a technológia, egyre többen foglalkoztak vele, hiszen az akkori hardvereken kezdtek el végre használhatóan működni ezek a modellek. A felügyelt, a felügyelet nélküli és a megerősítéses tanulás terén is rendkívül sikeresnek mondható a korszak. Azonban ezek a statisztikai módszerek inkább célzott feladatok megoldására használhatóak hatékonyan. Ide már olyan közismertebb megvalósulások tartoznak, mint az internetes webes keresések, a mobiltelefonos arcfelismerés az okoshangszórók természetes nyelvfeldolgozási képessége, illetve a kevésbé ismert, de elterjedt okos spam-szűrők. Megemlítendő, hogy a kognitív tudományok ekkorra váltak részévé a kutatásoknak.
- (3.) A harmadik fejlesztési hullámban az MI technológia egyesíti az első és a második ciklus erősségeit, így komplexebb alkalmazásokra nyílik lehetőség, mely már kifinomult különbségtételre, absztrakcióra, sőt magyarázatra is képes. Így, míg kb. a 2010-es évektől vált közismert tömegtermékké, és kezdett el nagyon terjedni az MI. Azóta a bio-inspirált tanulási módszerek, például a mélytanulás és raj-intelligencia és egyéb biotikai módszerek (ld. III.3.4.) alkalmazására összpontosítanak, és jelentős sikereket értek el az érzékelés és észlelés pontosabbá válásában. Ide tartozik „2023 slágere”, a humorral és irodalmi

³ Úgy vélték, hogy egy érzelmek nélküli gép sokkal optimálisabb döntéseket hozhat, mint egy érzelmei által zavart emberi elme. Azt várták, hogy megfelelő mennyiségű alapvető adat alapján a gépi logika tökéletesebb döntéseket fog hozni, mint az ember. Ennek a tévedésnek alapos elemzésére az affektív számítástechnikánál (I.3) visszatérek.

készségekkel is rendelkező ChatGPT 4.0, vagy az autonóm járművek komplex rendszerei, és az okos-kibervédelmi megoldások is.

Jelenleg is érezzük, hogy ez a harmadik hullám kisebb cunamiként ömlött rá a világra, gátak nélkül. A robbanás-szerű, és túl hirtelen terjedés és túl gyors változás már önmagában megnehezíti megfelelő szabályozást, nem is beszélve a problémakör újdonságáról. De nem csak jogilag, hanem a hétköznapiakban is érezhetővé váltak az emberek és gépek kapcsolódásának új kihívásai, melyre a megoldást legtöbbször az ún. *szimbiotikus intelligencia* megvalósulásában látják. Ezen a néven utalnak az emberi-gépi együttműködés ilyen újgenerációs lehetőségeire.[7] A hivatalos megközelítések abban látják az MI megfelelő felhasználásának zálogát, hogy ha sikerül az ember és MI szimbiózisát létrehozni, és hozzáteszik, hogy ehhez az embereknek is változniuk kell majd.[6, o. 52–53]

Csak hogy a kifejezés szerintem rossz és ijesztő módon érthető félre. Ugyanis kevésbé avatottak könnyen az agyi implantátumokban megvalósuló ember gép együttélésére asszociálnak (ld. II.3.1.) a szóban lévő „biozis” miatt. A félreértést erősíti, hogy olvasni lehet ember-gép szimbiózisról is. Sokkal pontosabb lenne az szinergia (*synergia*) szó használata, mely használatos is az együttműködés szó helyett. Ezentúl tehát itt igyekszem Szinergikus MI-ként utalni erre a célra, jelölve, hogy a szimbiotikus MI-t értem alatta, és elkerülöm az ember-gép szimbiózis kifejezést.

Nem tudjuk ez sikerülni fog-e, és hogy 20 év múlva, visszatekintve, hogyan fogják jellemezni a soron következő negyedik ciklust. Bizonyára lesz valami sajátos, meghatározó mozzanata, én az MI hardver-alapokra helyeződésében várom ezt a sajátosságot. (Erre több irányt is be fogok mutatni II.3.1-ben). Az ugyanott említett egyéb fejlesztési irányok kibontakozását későbbi időpontra (talán egy ötödik hullámban) várom. Ezek a tendenciák, a már megvalósult képességek, valamint a várható kihívások is alátámasztják az V. fejezet vezérgondolatát, miszerint nem szabad várni az MI fogalom bővítésével és pontosításaival.

1.1.2. A mesterséges intelligencia mai hivatalos meghatározásai Magyarországon és az EU-ban

Itt a céloom csupán képet adni arról, hogy mit értenek ma MI alatt, ehhez pedig elegendőek a minket leginkább érintő (hazai, európai és amerikai) hivatalos meghatározások. A szakírók és hivatalos szervek által írt definíciókban sok a közös vonás (gyakran átveszik vagy átfogalmazzák egymást) ezért sem érdemes itt óriási mennyiségű MI meghatározást összevetni, csak ezekre a közös vonásokra koncentrálok. A jól használható meghatározásokat kétoldali

behúzással és dőlt betűvel emeltem ki a későbbi felhasználhatóság céljából, míg az általam kevésbé fontosnak ítélt (csak a teljesség kedvéért idézett) definíciókat csupán lábjegyzetben említem.

Kezdjük a sort egyik hivatalos magyar meghatározással, amely a gyakorlatias és rövid definíciók közül véleményem szerint az egyik legjobb. Magyarország Mesterséges Intelligencia Stratégiája 2020-2030 I. része elején olvasható a 9. oldalon:

„A mesterséges intelligencia az emberi intelligencia valamely részének leképezésére alkalmas szoftver, amely képes támogatni vagy autonóm módon ellátni észlelési, értelmezési, döntési vagy cselekvési folyamatokat. Egy technológia, amely speciális képességekkel rendelkezik, mégis kiemelt figyelem kíséri mind gazdasági, mind társadalmi szinten.”⁴

Van egy tömörebb, frappáns meghatározás is a 6. oldalon:

„A mesterséges intelligencia (MI), mint a betáplált adatok alapján önmagukat tanítani és javítani képes algoritmikus rendszerek összessége”.[8, o. 6]

Kiindulásnak, illetve alapozó oktatásnál jó definíciók ezek, de nem eléggé átfogóak. Részemről volt egy várakozás, hogy a régóta vajdó európai szabályozásban végre majd egy jól használható és széles meghatározást kapunk, hiszen régóta elérhetőek annak háttéranyagai. A háttéranyagok vizsgálatával részletesen foglalkoztam, ennek a kutatásnak a kivonatát nagyon röviden ismertetem itt.

Ha néhány szóban szeretnék jellemezni, azt mondhatjuk, hogy az Európai Unióra jellemző joghangsúlyos eltolódás uralkodik a 2023. júniusában elfogadott EU-s rendelet-javaslat⁵ (továbbiakban: EU AI Act)⁶ MI definícióján is. A fogalmak felsorolásánál említett, agyon-rövidített meghatározás sajnos inkább visszalépést jelent korábbi EU-s meghatározásokkal szemben, vagy a világ többi részének megfogalmazásaihoz képest, hiszen nem jelennek meg sem gépi kognitív képességek sem az autonómia problémái,⁷ (ezek a világ többi definíciójában általában hangsúlyt kapnak).

⁴ Megjegyzés: a „mégis” kötőszót helyett egy sima és-t javasolnék (minden elismerésem és tisztelem mellett). [8, o. 9]

⁵ A hivatalos hírek címe, és számos cikk elfogadott törvényről beszél, pedig még csak sokadik javaslatról van szó „2023. június 14-én a képviselők elfogadták tárgyalási álláspontjukat az MI-törvénnyel kapcsolatban. A Tanácsban most megkezdődnek a tárgyalások a tagállamokkal a törvény végleges formájáról.” [9]

⁶ Pontos magyar címe: [10] I. CÍM Általános rendelkezések, 3. cikk 1.

⁷ A szöveg szerint a „mesterségesintelligencia-rendszer: olyan szoftver, amelyet az I. mellékletben felsorolt technikák és megközelítések közül egy vagy több alkalmazásával fejlesztettek, és amely az ember által meghatározott célkitűzések adott csoportja tekintetében olyan kimeneteket, például tartalmat, előrejelzéseket,

Pedig még megjelent a virtualizáció és az autonómia is abban a változatban, amelyet az OECD-vel is egyeztettek.⁸ Az végül elfogadott verzióba azonban az említett vértelen meghatározás került, melyet ugyan pontosítanak a szabályzó más részében⁹ felsorolt, MI-vel kapcsolatos technikák és megközelítések,¹⁰ ez a felsorolás azonban jócskán igényelne további definíciókat. 20Így a szöveg véleményem szerint túlzott informatikai szakmai tudást vár el a jogalkalmazóktól, vagyis kérdéses, hogy a jogcentrikus főcél teljesül-e így. Ezért állítom, hogy az EU AI Act számos erénye ellenére a benne elfogadott definíció visszalépésnek tekinthető. Ugyanis 2019-ben, ennek a törvénynek korai előkészítéseként, az EU számára tett javaslatot egy használható meghatározáshoz neves szakértők felkért csoportja. Egy füzetnyi terjedelemben adtak áttekintést, melynek végén az alábbi, összegzés egy igen jól használható, a magyarnál bővebb és precízebb meghatározást adtak.

„A mesterséges intelligencián alapuló rendszerek olyan, emberek által megtervezett szoftverrendszerek (és lehetőség szerint hardverrendszerek), amelyek összetett céljukra tekintettel a fizikai vagy a digitális dimenzióban úgy működnek, hogy a környezetüket adatszerzés révén észlelik, értelmezik a gyűjtött strukturált és nem strukturált adatokat, ismereteik alapján érvelnek vagy ezekből az adatokból származó információkat dolgoznak fel, valamint eldöntik, hogy az adott cél eléréséhez melyek a leghatékonyabb cselekvések. Az MI-rendszerek használhatnak szimbolikus szabályokat vagy numerikus modellt is betanulhatnak, és a magatartásukat is megváltoztathatják annak elemzése révén, hogy a korábbi cselekvések hogyan hatottak a környezetre.

Az MI tudományterületként számos megközelítést és technikát foglal magában, köztük a gépi tanulást (amelyre konkrét példa a mélytanulás és a megerősítéses tanulás), gépi érvelést (amely magában foglalja a tervezést, ütemezést, az ismeretek bemutatását és az érvelést, a kutatást és az optimalizációt), valamint a robotikát (amely

ajánlásokat vagy döntéseket képes generálni, amelyek befolyásolják azt a környezetet, amellyel kölcsönhatásba lépnek.” (3. cikk 1.)

⁸ Ez a megfogalmazás még így hangzott: „A „mesterséges intelligencia-rendszer olyan gépi alapú rendszert jelent, amelyet úgy terveztek, hogy különböző szintű autonómiával működjön, és amely explicit vagy implicit célok esetén olyan kimenetet generálhat, mint például előrejelzések, ajánlások vagy döntések, amelyek befolyásolják a fizikai vagy virtuális környezetek helyzetét” (saját fordítás). [11]

⁹ [12, o. 2] I. Melléklet A mesterséges intelligenciával kapcsolatos technikák és megközelítések a 3. cikk 1. pontjának megfelelően

¹⁰ „a) Gépi tanulási megközelítések, ideértve a felügyelt, a felügyelet nélküli és a megerősítő tanulást, a módszerek széles skálájának, többek között a mélytanulásnak az alkalmazásával; b) Logikai és tudásalapú megközelítések, beleértve a tudás megjelenítését, az induktív (logikai) programozást, a tudásbázisokat, a következtetőmotorokat, a(z) (szimbolikus) érvelést és a szakértői rendszereket; c) Statisztikai megközelítések, Bayes-féle becslés, keresési és optimalizálási módszerek.”

*magában foglalja az ellenőrzést, az észlelést, az érzékelőket és működtető egységeket, valamint minden más technikának a kiberfizikai rendszerekbe történő beépítését).*¹¹

Érthetetlen, hogy ezt miért szimplifikálták le a 4 év múlva elfogadott (fentebb, de lábjegyzetben említett) szövegre. Véleményem szerint hibái ellenére ez az eredeti definíció, pontosabban a hozzá kapcsolódó magyarázó füzet egésze, a jogi szabályozás szempontjából használhatóbb lenne, főleg hosszú távon. A hibák alatt elsősorban azt értem, hogy általában csak az intelligencia észbeli, logikai aspektusa jelenik meg meghatározásokban. A felkért szakértők az *intelligencia* átfogó meghatározása helyett egyenesen az *ésszerűség* fogalmát használják.¹² Ennek folyományaként logikus lenne a szakirodalomban Mesterséges Ésszerűségről beszélni, bár ezt az átnevezést senki nem veti fel. És nem véletlenül nem vetik fel, hiszen régóta zajlanak fejlesztések, melyekben az okosságon kívüli képességek gépi előállítására törekszenek. Leginkább az érzelmi (ld. I.3) vagy szociális intelligenciát¹³ próbálják mesterséges eszközökkel utánozni, ezeknek a meghatározásokban meg kellene jelenniük.

I.1.3. Az USA megoldása párhuzamos definíciókkal

Az Amerikai Egyesült Államok kongresszusának megközelítése¹⁴ nem csupán ezért érdemes a vizsgálatra, mert világ jelenlegi vezető hatalma nyilván kihat minden egyébire. Ez a megközelítés úgy oldja fel a szabványosság hiányát, hogy több értelmezést egyszerre ad meg, így figyelemre méltó azért is, mivel a párhuzamosan megadott több meghatározás ötlete megfontolandó minden hasonló kérdésnél. Továbbá ezek a szövegezések jó lehetőséget adnak számomra egy, a legtöbb definícióban megjelenő hiányosság feltárására. Ehhez a párhuzamos definíciókat specifikusan összegeztem az 1. sz. táblázatban.

	mesterséges intelligencia definíció	autonómia	tanulás	ész (gondolk.)	megismerés	egyéb int.
1.	Bármilyen mesterséges rendszer, amely változó és előre nem látható körülmények között hajt végre feladatokat	X	X			

¹¹ A hivatalos fordítást nyelvileg javítva idéztem, de az MI helyett az ott szereplő MI (artificial intelligence) rövidítést hagytam benne. [13, o. 6]

¹² Az ésszerűség (rationality) pedig „azt a képességet jelenti, hogy egy bizonyos cél elérése érdekében képesek vagyunk a legmegfelelőbb cselekvést kiválasztani, tekintettel bizonyos optimalizálandó kritériumokra és a rendelkezésre álló erőforrásokra”. [13]

¹³ Lorenzo Cominelli, Daniele Mazzei, és Danilo Emilio Rossi, „SEMI: Social Emotional Artificial Intelligence Based on Damasio’s Theory of Mind”, *Frontiers in Robotics and MI* 5 (2018): 6, <https://doi.org/10.3389/frobt.2018.00006>.

¹⁴ Saját fordítás. [15, o. 1]

	jelentős emberi felügyelet nélkül , vagy amely képes tanulni a tapasztalatokból és javítani a teljesítményét, ha adathalmazokhoz kapcsolódik.					
2.	Számítógépes szoftverben, fizikai hardverben vagy más módon kifejlesztett mesterséges rendszer, amely emberihez hasonló észlelést, megismerést , tervezést, tanulást , kommunikációt vagy fizikai cselekvést igénylő feladatokat valósít meg.		X		X	
3.	Kognitív architektúrákat és a neurális hálózatokat magába foglaló mesterséges rendszer, amelyet arra terveztek, hogy emberként gondolkodjon vagy viselkedjen			X	X	
4.	Olyan technikák halmaza, amelyek része a gépi tanulás , és amelyeket kognitív feladatokat megközelítésére terveztek.		X		X	
5.	Racionális cselekvésre tervezett mesterséges rendszer, beleértve mind az intelligens szoftverügynököt mind testtel rendelkező robotot, amely észleléssel, tervezéssel, érveléssel, tanulással , kommunikációval, döntéshozattal és cselekvéssel éri el a céljait.		X	X		

1. táblázat: Kulcsszavak az MI definíciókban (saját szerkesztés)

Elsősorban arra voltam kíváncsi, hogy túlmutat-e a megközelítés az imént felvetett Mesterséges Észszerűsége, megjelennek-e benne egyéb intelligencia típusok, vagy legalább a megismerés (kogníció) fogalma, hiszen azzal is utalni lehetne racionalitásnál tágabb jelentésre. Ehhez először bejelöltem, hogy hol jelenik meg a szövegezésben a három kulcsszó: a tanulás, az intelligencia és az autonómia,¹⁵ melyek a hagyományos megközelítés szerint az MI-t megkülönböztetik a hagyományosan programozott technológiáktól. Ezek mellé vettem fel a kogníció és az egyén intelligencia típusok számára oszlopokat. Ahol megjelenik az adott tulajdonság azt „X” jelöli, ahol nem, azt pedig üres mező.

Ebbe a sémába szinte minden fellelhető meghatározást be lehetne sorolni, ám mindegyiknél hasonló eredményt kapnánk, mint itt: az utolsó oszlop minden esetben üres maradna, és a kogníció oszlopa is kevés X-et kapna. Úgy tűnik senkinek sem hiányozik ez az utolsó oszlop. Például a NATO által 2021-ben megfogalmazott MI-stratégiájának 2024-ben tervezett felülvizsgálatakor is csupán a generatív MI emelik ki hiányosságként.[16] De példa lehet erre a NATO

¹⁵ Itt csak röviden: Tanulás: képes viselkedését célszerűen és megismételhető módon változtatni. Intelligencia: az ember következtető és kognitív képességeinek utánozza. Autonómia: emberi beavatkozás nélkül képes reagálni a változó környezeti behatásokra.

egy fejlesztési iránytervének meghatározása is,¹⁶ melyben ugyan mindhárom klasszikus kifejezés mellé rakhatnánk X-et, viszont az utolsó két oszlop ott is üresen maradna. Megemlítendő érdekesség ebben iránytervben, hogy az autonómiát külön elemzi egy teljes fejezetben, ezzel azt sugallva, hogy a gépi autonómia képesség egyelőre a legnagyobb kockázat, jelentősebb, mint a tanuló vagy kognitív képességek.

Az ismertetett hazai, EU-s és amerikai forrásokban sikerült láthatóvá tenni tehát a meghatározások egyik hiányosságát, hogy az intelligencia nem racionális oldalainak gépi megvalósítására nincs bennük utalás – ezt pótolni lenne szükséges, és a javaslatomban is figyelembe veszem. Érdekes azonban egyéb kritikai megjegyzéseket is feldolgozni, és a fenti, ötoldalú meghatározásra találtam említendő anyagot. Néhány szóban összefoglalom ennek néhány fontos meglátását, melyek igazak lennének jó néhány hivatalos MI-definícióra.[17] (Emellett a szerző hasonló hiányosságokat is említ, mint én¹⁷). Elsősorban azon meglátása miatt említem, mely szerint a „racionális cselekvés” kifejezés indokolatlanul korlátozhatja a mesterséges intelligencia meghatározását a racionalitásra vonatkozó feltételezésekre.¹⁸ De fontos gondolata az is, hogy az MI túl szűk vagy túl tág meghatározása azzal a kockázattal jár, hogy korlátozza az MI képességeinek spektrumát, vagy nem jól határozza meg az MI-rendszerek egyedi kapacitását.

1.1.4. Mesterséges ügynökök (ágensek)

Mielőtt továbblépnénk, szükséges beszélnünk még az úgynevezett *mesterséges ágensek* fogalmáról, mely gyakran előfordul a szakirodalomban. Ez azért elkerülhetetlen, mivel a téma megközelíthető akár úgy is, hogy „*az MI egyszerűen a környezetüket észlelő és cselekvő ágensek tanulmányozása.*”[19, o. xxxiv] De azért is fontos, mivel könnyű összekeverni a MI-t és intelligens ágenszt (*intelligent agent*). Tudatosan az ágens szót, az angol kifejezés magyarítását használom annak magyarra fordítása helyett, mivel az „ügynök” kifejezés informatikai használata nincs eléggé jelen a magyar köztudatban. Ha magyar szót keresünk rá, akkor én az „ügyködő” kifejezést javasolnám. Ez mutatna rá a fogalom lényegére, ami ügyködik, valamilyen tevékenységet végez, megkülönböztetve attól a rendszertől (MI), ami ezt a működést lehetővé

¹⁶ „A mesterséges intelligencia a gépek azon képességére utal, hogy olyan feladatokat hajtanak végre, amelyekhez általában emberi intelligencia szükséges. Például minták felismerése, tapasztalatból való tanulás, következtetések levonása, előrejelzések megfogalmazása vagy cselekvés (akár digitálisan, akár az autonóm fizikai rendszerek mögött meghúzódó okos szoftverekként).” (saját fordítás) [6]

¹⁷ Az akkori kutatásom kb. a cikk megjelenésével egy időben, csak később került publikálásra, akkor még nem ismertem ezt az anyagot, [18]

¹⁸ Bár azzal is érvel, hogy a mesterséges intelligencia kognitív feladatokra való korlátozása korlátozhatja a mesterséges intelligencia különféle felhasználásait – ám én nem látom a kognitívításra korlátozott definíciókat.

teszi. Tehát hasonlittal élve az ágens a cég adott részlege vagy alkalmazottja, amely végrehajtja a rábízott feladatokat (a cég pedig az MI-vel állítható analógiába).

Legegyszerűbb megfogalmazásban azt mondhatjuk, hogy az ágens egy olyan szoftver, ami képes cselekedni a számára adott környezetben, vagyis önállóan hoz döntéseket. Ám ez igazából ez nem csupán az MI-ről mondható még el, hanem még egy programozott automatára¹⁹ is igaz. A szakirodalomban azonban már kezdetektől, az 1970-es évek elejétől²⁰ használják annak ellenére, hogy számos típusára nem jellemző a tanulás. Egy MI rendszerben tehát vegyesen lehetnek intelligens és pusztán automatikus ágensek. Akár úgy is megközelíthetjük, hogy az ágens egy olyan elméleti egység, amely az emberek munkavégzési, folyamatszerkezési modelljeinek gépi megvalósítója, vagyis egy olyan feladatkör elvégzője, amelyet esetleg ember is végezhetne (vagy régebben az végezte). A fejlődéssel ezek az ágensek által átvett feladatkörök szélesednek, és az írat-besoroló automatikából multifunkcionális, intelligens titkár-ágens lesz. Az MI és az ágens kifejezéseket azért könnyű összekeverni, mivel egy-egy cél MI rendszer esetleg csupán egyetlen ágenszt valósít meg.

Az ágens és az MI közötti különbségnek van egy másik oldala is, melyre a fenti céges hasonlat nem világított rá, és az sem, hogy szoftverként kezeljük. Az ágens olyan, mint egy edzés, mely része valamely sportnak (ahol a sport áll analógiában az MI rendszerrel). Az adott sport működésének része sok szabály, közvetítések és azok nézői, helyszínek és az azokat üzemeltető cégek stb., tehát számtalan nem-sportoló, passzív szereplő szükséges ahhoz, hogy edzeni lehessen. Az edzés egy adott szempont szerint tervezett (pl. izomcsoportra, gyorsaságra, ügyességre) aktív tevékenység. Ez a hasonlat tehát az aktív és a passzív oldalt igyekszik szemléltetni.

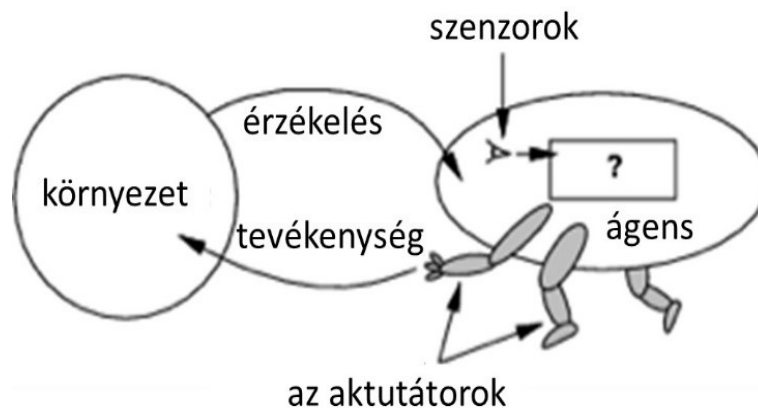
Mivel az ágens általában egy részfunkció, tisztázandó az is, hogy mi a különbség a szoftverek egyéb moduljai és az ágensek között? A válasz, hogy az ágens folyamatos interakcióban van a környezettel, ehhez alapvető elemei az érzékelők és az aktuátorok (ezt nem érdemes magyarázni). Az érzékelőktől kapott inputok által képes a változó környezeti hatások függvényében működni, az aktuátorok, pedig olyan outputok, melyek lehetővé teszik, hogy aktív kölcsönhatásba lépjen a környezetével. Ezt modellezi az. 1. sz. ábra, azonban kiemelendő, hogy a karok és lábak nem feltétlenül fizikaiak. Az aktuátorok lehetnek a környezetre közvetlenül ható fizikai eszközök, mint egy ventilátor vagy robotkar, ám ha a környezet tisztán virtuális, akkor az információközlő output (pl. monitor) az aktuátor, hiszen az információ változtatja meg a környezetet (a környezet itt a monitor előtt ülő, azon a gépen dolgozó ember).

¹⁹ Itt elegendő a fogalom közismert értelmezése. Az automaták és az MI összevetéséről a gépi autonómiával összefüggésben készült alaposabb vizsgálat, az automatika és a gépgenerációk című (IV.1.2), az című részben.

²⁰ Az ágens szó első ilyen említése: [20]

Az 1. sz. ábra sémája alapján sokféle ágens létezik, melyek részben a tervezésekor figyelembe vett főbb elvek, részben a felhasznált technológia (pl. tanulás), más esetekben a bonyolultság alapján különböztetünk meg. Mára körülbelül hét típust szokás elkülöníteni,[21] ám ezeket témánk szempontjából elegendőnek tartom itt csupán érinteni.

- A legegyszerűbb szint a reflexiós ágens, mely egy környezetét érzékelő, és az érzékelt adatok alapján programozottan cselekvő automata;
- Fejlettebb változatába beleprogramoznak egy olyan modellt is, mely segít számára a világ egy adott részének nyomonkövetésére;
- Megkülönböztetünk cél-alapú ágens, melyben a cél elérése kap főszerepet, ennek elérésére hoz döntéseket, és a bizonyos típusai több cél között is tudnak rangsorolni;



1. ábra: Az ágensek működési sémája (saját készítés, forrás a szövegben)

- Ez utóbbit fejleszti tovább a hasznosság-alapú ágens, mely egy-egy adott elérhető állapothoz mérőszámokat rendel, így nem egyetlen célt akar elérni, hanem sok tényező együtteségnek maximális hatékonyságára törekszik;
- A tanuló ágensek esetében jelenik meg a gépi tanulás, melyet egy erről szóló alfejezetben (II.2.) alaposabban tárgyalok. A tanuló funkcióval is rendelkező ágenseket lehet intelligens ágenseknek nevezni;
- Egy másfajta módon fejlettebb a hasznosság-alapú ágenseknél a MAS (*Multi Agent System*), vagyis többágensű rendszer, melyben sok ágens együttműködésén van a hangsúly.
- A többágensű rendszerek bonyolódása nyilván a részt vevő egyre több ágens tervezettebb hierarchizálását tette szükségessé (ahogyan egyre több ember együttműködése is egy idő után csak egy hierarchia bevezetésével képes hatékony maradni). Az ilyen MAS elvileg nem feltétlenül intelligens, egy régebbi robotizált gyár vezérlése is ide tartozna, ám ez inkább a fogalom vissza-vetítése lenne a múltba.

- Jelenleg azok a hierarchikus ágens-rendszerek képviselik a technológia csúcsát, ahol sokféle tanuló ágens is meghatározó szerepet kapnak. Ezeket hívhatnánk hierarchikus tanuló / intelligens ágensnek, ám valójában eleve csak ezt értik alattuk, így felesleges a kifejezést bonyolítani (a fogalom megértését azonban remélhetőleg segítette ez a pontosítás).

Végül meg kell említenünk az *embodied agent* magyarul megtestesült ágens kifejezést, amely alatt valamiféle külső testi jellemzőt (arcot, kezét, lábat) kapott ágenseket kell érteni. A kilencvenes évek végén²¹ tűnik fel a kifejezés egy szoftveres, virtuális megoldásra: a társalgó robotok számára kívántak a monitoron megjelenő, mozgó arcot generálni. Később a kifejezés bővült, és a valódi fizikai testet kapó, azaz robotikus megoldásokat is ide sorolják, másrészt mára mind az input mind az output kapcsolati irányokat érintheti. Így a megtestesült ágens kifejezés mai meghatározása: *olyan intelligens ágens, mely egy fizikai vagy virtuális testen keresztül lép kölcsönhatásba a környezettel.*

1.2. AZ INTELLIGENCIA, AMI MESTERSÉGES IS LEHET

Az intelligencia mélyebb mibenlétének megragadása és típusainak különféle osztályozásai már régen jelen vannak a filozófiai és pszichológiai kutatásokban. Ezek eredményei mára meg lehetőségen, sőt túlságosan is sokszínű képet mutatnak. Ezért itt csupán a cél szempontjából fontos aspektusokat emelem ki.

1.2.1. Nyelvi megközelítés

1. A köznyelvben általában az intelligens jelzőt használják, az intelligencia főnevet ritkán használják, ilyenkor leginkább egy olyan sajátosságra utalnak vele, ami által valaki „*okosan és kultúráltan nyilvánul meg*”. Egyszerűsítve mondható, hogy a hétköznapokban az okos szinonimája, van azonban egy többletjelentése. Egy viselkedni nem tudó zsenire inkább azt mondják, hogy „okos, de bunkó”, míg egy udvarias, figyelmes, megbízatosan megnyilvánuló középszerű szélhámos sokak számára „nagyon intelligens ember”. Más szóval valami megértése, és mások számára befogadhatóvá tétele egyszerre fontos árnyalatai az intelligenciának. Véleményem szerint részben emiatt fogadták sokan ovációval a mesterséges nyelvi rendszerek kulturált fogalmazásmódját.

²¹ 2000-ben már könyv foglalkozik a témával, de sokkal korábbi publikációt nem leltem fel. [22]

2. A kifejezés latin etimológiája szerint talán az „*egybegyűjt*” szó (inter = közé, együvé + legere = szed, gyűjt)²² adhat meg egy sokatmondó szószerinti fordítást. Ez feltárja, hogy a fogalomban régóta megjelenik az a fontos szempont, hogy az ember egybegyűjti az információkat. A gyűjtés szó maga még nem sugallja az említett okosságot: erre inkább abból szokás következtetni, ahogyan az illető átadja az információt. Az intelligens ember a külvilágot nem csupán konstatálja, hanem érti, sőt megérti, és megérthetővé teszi. Ehhez a megértéshez a külvilágot, önmagát, illetve e kettő kapcsolatát mindenki folyamatosan értékeli, feldolgozza, tudatosan és tudat alatt, sőt még álmában is. Így gyűlik egybe a tudásunk. Ezt az etimológiát alátámasztja, hogy már az ókorban is az „értelem képességét” jelentette az *intelligentia*, ami mellett a szótárak a „felfogóképességet” és a „megismerőképességet” is említik. Ez a három fordítás azonban három teljesen különböző oldalt jelöl.
3. A *felfogás*, inkább egyfajta konstatálás, tudomásul vétel, memóriába helyezés csupán.
4. A *megismerés* az adatok közötti alapvető összefüggések felismerése.
5. Az értés, és főleg a magyar *megértés* szó jelentése inkább arra utal, hogy a világból leképezett nagy belső rendszerünkben helyére kerül egy, a felfogás által megkapott és a megismerés által összefoglalt információhalmaz. Ez az új tudáshalmaz ettől kezdve a többivel összefügg és az életünkhöz, a jó döntéseinkhez felhasználható.

A technika mai állása szerint a felfogó és megismerő szinteken már jól teljesítenek a gépek, ám a valódi megértésük nincs.²³ A mai gépintelligencia inkább egy magolós gyermek szintjén ügyes csupán, amint erre számos kutató mutat rá.²⁴

Több jelentésmódosulást is eredményez az intelligencia kifejezésen, ha közösségi aspektusból közelítjük meg. Ennek háttere, hogy a legtöbb ember ösztönös tulajdonsága, hogy a közösség többi tagjával is többé-kevésbé megossza a saját „feldolgozott információit” (felismeréseit, történeteit, pletykáit stb.), vagyis a kommunikáció és az intelligencia szervesen összefügg.

6. Az elsődleges ilyen jelentésre már fentebb utaltam: az intelligencia értelmezéseiben nem szokott megjelenni a *megértés visszaadása*, pedig talán ez fontosabb, mint a megértés. Pl. hiába érzi úgy a hallgató, hogy mindent ért, ha vizsgán meg sem tud mukkanni, nem mondja róla a vizsgáztató, hogy intelligensen hallgató... Az alapján alkotunk képet az intelligenciáról, hogy egy illető mit ad vissza felénk a világból: nem csupán magyarázat szintjén, hanem egy új felfedezés vagy akár művészi erejű megragadás dimenzióiban.

²² <https://www.arcanum.com/hu/online-kiadvanyok/Lexikonok-magyar-etimologiai-szotar-F14D3/i-i-F266A/intelligens-F2767/> (Letöltve: 2024.01.15)

²³ Az MI-nek nincs mögöttes világmodellje, megmondja hogyan kell *hangzania* a válasznak, ám ez nem azonos a jó válasszal. Ld. [23]

²⁴ Az előző interjú-idézet tudományos háttere: [3]

7. Ide kapcsolódik egy másik jelentés-árnyalat: a feldolgozott információ megosztása által, egyfajta kollektív intelligencia is kialakul (vö. rajntelligencia II.3.3.). Ennek a tudásnak a letéteményese hagyományosan az *értelmiség*, akiknek a fő feladata az ismereteknek a közösség számára történő gyűjtése, raktározása és közérdekű feldolgozása volt. Ez egy ősi szerep, bár a kifejezés újkori, és néhány nyelven²⁵ az intelligencia szót alkalmazzák rájuk (a magyar nyelvben is az értelem szóból képezték a nyelvújítók). Manapság viszont, úgy tűnik, hogy már nem az értelmiség nevű emberi réteg őrzi ezt a tudást: az elektronikus tudásmegosztás miatt bárki képes kész információt szerezni, így ez a szerep egyre inkább a technológia felé tendál. Pontosabban az értelmiség társadalmi szerepe és összetétele épp úgy át fog alakulni, mint számos értelmiségi dolgozó feladatköre. (Például aki csak rutinmunkákra volt alkalmas, az kieshet ebből a rétegből.)
8. Az ismeretmegosztásból adódó harmadik fontos oldal, hogy az intelligencia sok nyelvben *hír* jelentéssel, vagyis *átadott információként* is szerepel.²⁶ Megjegyzendő, hogy mai használatban a hír elvileg egy objektív és fontos tény közlése, bár ezek a jelzők a média híreire sokszor nem igazak. Ide tartozik még, hogy mindig emberek (vagy ma már sokszor gépek) által már feldolgozott és kiválogatott anyagot kapunk, vagyis a hír gyakorlatilag inkább feldolgozott, esetleg átdolgozott információt jelent. A fontosság kérdésének ügyes megállapítása pedig része az okosságnak: aki intelligens, az képes akár egy spontán adatot is összefüggésbe hozni egy adott szempontból egy kapott új információkkal.
9. Ehhez szorosan kapcsolódik egy negyedik jelentésárnyalat: ugyanezt a szótövet (intelligence) alkalmazzák a hír megszerzésére, pontosabban a magyarul hírszerzés néven összefoglalt védelmi tevékenységre is. Megjegyzendő, hogy az emberi hírszerző struktúrában különválnak a megszerző és az elemző rész, hiszen akik az adatokat (esetleg adathordozókat) megszerzik, azok sokszor nincsenek a megszerzett hír (információ) birtokában, és számos továbbítási és feldolgozási fázis alapján állnak össze az adatok védelmi információvá. Ezek a fázisok és a munkamegosztás pedig az MI alapú hírszerzésnél is meg fog maradni, hiszen más jellegű modulok és modellek lesznek szükségesek minden funkcióhoz.

A fentiek alapján az intelligencia jelent tehát megértést, vagyis annak tudását, hogy mi a fontos, milyen információ milyen szempontból bír hírértékkel, valamint ezeknek a híreknek,

²⁵ pl. a németben, angolban, oroszban az *intelligentsia* a „a szellemi foglalkozásúak közössége”, azaz értelmiség jelentéssel is bír. <https://uesz.nyttud.hu/index.html?displaymode=web&searchmode=exact&searchstr=intelligencia&hom=> (Letöltve: 2024.01.15)

²⁶ használatos a szó pl. a latinban és angolban „üzenet, értesülés, hír” értelmével (forrás az előző jegyzetben)

tapasztalatoknak a feldolgozását, rendszerzését és okos visszaadását is jelenti. Akinek az előző mondat részeire irányuló minden képessége jó, annak az emberi közösségi dinamikák alapján az lesz a feladata, hogy ezt a csoport vagy társadalom felé kamatoztassa. Ezzel egyben etimológiai alapon támasztottam alá azt a tendenciát, miszerint a gépi intelligencia képes lesz a jövőben egyfajta az „új értelmiséggé” válni. Hasonló logika mentén, ha az intelligencia képessége tudja, hogy mi a fontos, mi az igazi hír, akkor elmondható, hogy az MI „zsigerileg hírszerző”, amennyiben a hírszerzés teljes struktúráját támogatni képes.

1.2.2. Az intelligenciafajták

Az etimológizálás után ebben a kisfejezetben logikusan az intelligencia pszichológiai-filozófiai értelmezéseinek történeti alakulására kellene rátérnem. Azonban csak ismételni tudnám egy nemrégiben megjelent összefoglalás tartalmát,[24, o. 13–21] melyet átolvasva bárki arra az eredményre jut, hogy az MI fogalomhoz nem adnak támpontokat a történeti felfogás változásai általában nem ad új támpontokat, vagyis nem érdemes ezzel növelni a terjedelmet. Egy kivétel van, a „többszörös” intelligencia elmélete (*theory of multiple intelligences*), mely gyökeresen változtathatja meg az MI definícióját. Továbbá kiemelek két régi felosztást is, melyeket egyelőre ritkán szokás vizsgálni az MI-vel kapcsolatban

Howard Gardner nevéhez fűződik, hogy rendszerbe foglalta azt a sejtést, mely szerint számos emberi képességnek van intelligencia-jellege. Ezen képességek mérését kidolgozva, és külön-külön vizsgálva, széleskörűben tudjuk feltérképezni az alany erősségeit és gyengeségeit, ezek összesítésével pedig sokkal pontosabb és tudományosabb képet alkothatunk az illetőről. A teória első verziójában Gardner hét különböző intelligenciafajtát írt le,[25] melyet még ő maga kilencre emelt. A szemlélet alapján azonban az intelligencia-típusok száma jócskán kibővült, és a jövőben várhatóan tovább szaporodik. A mai megközelítésekben számos különféle felosztás olvasható, melyek több ponton eltérnek a korai intelligencia fajtáktól. Érdekes, de nem véletlen, hogy ezek közül épp az érzelmi intelligencia lett viszonylag közismert (ezt mérve kapható meg egy *emotion quotient*, EQ), amely nem volt a gardneri rendszerben. Viszont az emberek hétköznapi problémáikra ennek figyelembevételével kaphattak válaszokat, és az MI szempontjából is igen jelentős (ld. I.3.).

Visszatérve a sokféleségre: a különböző kutató-iskolák az élet különböző területein való helytállást igyekeznek mérhető paraméterekkel lefedni, és ez alapján különböztetnek meg különféle intelligenciafajtákat. A 2. számú táblázatban egyesítettem a korai típusokat, néhány

azóta megjelent²⁷ intelligencia fajtát, valamint saját meglátáson alapuló típusokat. Ez a teljesség igénye nélkül összeállított lista is több mint húsz fél intelligenciafajtát különböztet meg. A lista nem pusztán bővül, számos korai típus eltér az újabban leírtaktól: a táblázat oszlopai viszont nem negligálják a régi típusokat, csupán láthatóvá teszik hol feleltethető meg a régi és az új típus. Több komponens érdekes,²⁸ és talán vitatható, mégis véleményem szerint az összesítés alapján egy olyan benyomás alakul ki a szemlélőben, hogy a szakembereknek még ennél is több intelligenciafajtát kéne figyelembe venniük egy ember átfogó felmérésekor, és a szakirodalom alaposabb feldolgozásával a lista valószínűleg már most is kiegészíthető lenne. Ám itt elegendő ennyi, hiszen ez a típus-mennyiség is óriási számú ahhoz képest, hogy a legtöbben csupán az értelmi oldalt asszociálják az MI-ben szereplő intelligencia szóhoz.

2. táblázat: 22 intelligencia-típus (több modell alapján saját összeállítás)

	Gardneri intelligencia típusok	későbbi intelligencia fajták	
1.	logikai	intelligencia hányados	Intelligence Quotient (IQ)
2.	egzisztenciális		
3.	intrapersonális		
4.	interperszonális	szociális hányados	Social Quotient (SQ)
5.	testi-mozgási	táplálkozási (testi fegyelmi) hányados	Nutrition Quotient (NQ)
6.	térbeli	látás	Vision Quotient (VQ)
7.	természeti		
8.	zenei		
9.	nyelvi		
10.		érzelmi hányados	Emotional Quotient (EQ)
11.		viszontagságokkal szembenezés hányadosa	Adversity Quotient (AQ)
12.		pozitívnek látási hányados	Positive Quotient (PQ)
13.		kreatív intelligencia	Creative Intelligence (CQ)

²⁷ Egy kevésbé pontos és tudományos, azonban igen kreatív forrás adata az alapötleteket: [26]

²⁸ Az „érdekes” típusok: a *táplálkozási hányados* az étkezési (testi) tudatosságot és önfegyelmet méri.

A *viszontagsági hányados* a problémákkal való szembenezés képességét vizsgálja.

A *pozitív hányados* annak az aránya, amikor az elme a jót látja meg a dolgokban (az így töltött időt méri)

A *technikai és etikai kompetenciák hányadosa* a nagyfokú integritással és hosszú távú megközelítéssel alkalmazott technika- és szabály-alkalmazási készséget méri. Ezek összevonására nem találtam a forrásban magyarázatot, ezért említem külön sorban őket.

14.		gazdasági érzék hányadosa	Financial Quotient (FQ)
15.		spirituális hányados	Spiritual Quotient (SPQ)
16.		tapasztalati hányados	Experience Quotient (XQ)
17.		digitális hányados	Digital Quotient (DQ)
18.		technikai kompetenciák és	Technical and Ethical Quotient (TEQ)
19.		etikai kompetenciák hányadosa	
20.		szintézis-teremtő intelligencia ²⁹	
21.		gyakorlatiassági hányados ³⁰	
22.		emberismereti hányados ³¹	
23.		...	

Az intelligencia-fajták várhatóan bővülő száma mellett még egy ősi megkülönböztetést látok fontosnak, a fő téma (az MI-fogalom) szempontjából itt említeni: ez az emberek *analitikus*, *szintetizáló*, valamint *gyakorlatias* képességek szerinti tipizálása. Ezeknek a mindenkiben meg-lévő alapképességek a megoszlása egyéenként erősen eltérő. Ezek együtt határozzák meg az egyének mindennapjait és szakmai kibontakozását. A táblázatban is helyet kaptak (saját típus-ként), mivel az analízáló és szintetizáló MI rendszerek modell szinten is el kell, hogy térjenek. A gyakorlatiasság emberi érzéke pedig erősebben szocializációtól, illetve a tanulás függő, mint a másik két elem – így ennek gépi megvalósítása várhatóan inkább a tanulási adathalmazok függvénye lesz, kevésbé modell-függő.

Itt kell említenem egy másik felosztást, mely azért nem került intelligencia-fajtaként a táblázatba, mivel jelentőségét nem az MI, hanem a humánvirtualizáció (ld. V.2) szempontjából látom fontosnak. Ez az emberek *humán* és *reál* beállítottság szerinti felosztása. Bár felvethető

²⁹ Saját felvetés, a tudományos mérhetőség metódusai nélkül. Egy olyan elvontakoztatási képességet értek alatta, amikor valakiben felébreszhető a motiváció az általa ismert dolgok szélesebb összefüggésbe helyezésére, és ezt képes is megvalósítani információk újszerű szintézisét

³⁰ Saját felvetés, a tudományos mérhetőség metódusai nélkül. Azt az érzéket értem alatta, amikor valaki a rendelkezésére álló, sokszor nem arra való dolgokból képes egy működőképes szerkezet, használati tárgy stb. összerakására, sokszor fizikai tudás nélkül „érzi” a szerkezetek statikai vagy dinamikai kihívásait és a megoldásokat.

³¹ Saját felvetés, a tudományos mérhetőség metódusai nélkül. Ez alatt egy olyan érzéket értek, amikor valaki képes ráérezni egy másik ember valódi képességeire. Így például vezetőként a megfelelő vezető csapat illetve alvezetők kiválasztására. Hétköznapokban pedig nem kell csalódnia folyton a másik emberben.

ez a felosztás az MI rendszerekkel összefüggésben is, de célrendszerekkel kapcsolatban nem találtam jelentős szempontnak (az un általános MI-nél jöhet majd szóba, ld. alább).

1.2.3. A gépi intelligencia tudásszintjei

Az előző alfejezethez való szoros kapcsolata miatt logikailag itt érdemes bemutatnom az MI tudás-szintek szerinti felosztását (az MI szempontjából fontos felosztások sokfélék, ld. V.3.1.). Véleményem szerint ez ugyanis problémás, és problémáira az előző alfejezetben vázolt megközelítést felhasználva tudok rámutatni (a következő alfejezetben). Hasznos ennek a történetéről is pár szót szólni, e nélkül nem érhetőek az MI különféle jelzőkkel ellátott szókapcsolatai, és a közöttük lévő lényegi különbségek.

A régebbi szakirodalomban a *gyenge* MI (weak AI) elnevezést használták azokra a rendszerekre, melyek pusztán arra képesek, hogy úgy cselekedjenek, *mintha* intelligensek lennének.[19, o. 13] Ezzel szemben egy *erős* MI (strong AI) hipotézise állt, melyben olyan gépek létrehozását feltételezték, melyek valóban intelligensen cselekszenek. Sőt egy filozofikus megközelítéssel az „erős” MI kifejezés utalt az öntudattal rendelkező gépekre is, vizsgálták ennek elvi megvalósíthatóságát.

Ma azonban egy három szintű felosztás elterjedt,³² amihez hozzá kell tenni egy félig technikai kifejezést is

- 1) ANI³³ – Mesterséges Vékony Intelligencia (MVI): az embernél gyengébb képességű, feladatorientált gép;
 - 2) AGI³⁴ – Mesterséges Általános Intelligencia (MÁI): az emberhez hasonló képességű, általános célú gondolkodógép;
 - 3) ASI³⁵ – Mesterséges Szuper Intelligencia (MSzI³⁶): az embert meghaladó képességekkel rendelkező gép, mely talán az evolúció következő szintje;
- + *technológiai szingularitás* – az emberi és gép közvetlen kapcsolódása által a gondolkodás és egyéb képességeink beláthatatlan kiterjesztése.

Az ANI tehát a korábbi felosztás a *gyenge* szintjének helyébe lépett egy pontosabb kifejezésként. Az rendszerek ugyanis nem gyengék a maguk területén, sőt erősebbek, mint az ember. Viszont mind csupán egy vagy néhány *vékony* szeletét valósítja meg az emberi képességeknek. Ide tartozik minden mai rendszer, az egyszerűbb gépi tanulásoktól a mélytanulásokig, a cél-

³² A három szint közérthető összevetése: [27]

³³ ANI – Artificial Narrow Intelligence, Vékony MI, (a mai technológiák)

³⁴ AGI – Artificial General Intelligence, Általános MI

³⁵ ASI – Artificial Super Intelligence, Szuper MI

³⁶ A magyar rövidítések sajátok, ismeretlenségük miatt jobbnak ítélttem helyettük az angol verziókat használni.

specifikus rendszerektől a sok-modulos MI architektúrákig. Tehát egy szokásokat megtanuló hőfokszabályzó, egy arcfelismerő rendszer, és egy beszédhangulat értő és analizáló, arra értelmesen és érzelmesen válaszoló „vigasztaló robot” egy kategóriába tartoznak, vagyis praktikus tartalma nincs az ANI kifejezésnek, csak futurologiai oldalról értelmezhető.

Az AGI vagyis *általános* MI elnevezés is találóbb, mint a korai „erős” jelző. Ezek a fajta fogalom-pontosítások pozitív példák arra, hogy nem csupán az elhomályosodás irányába (ld. V.1.) történnek ilyen változások. Abból a szempontból azonban nem történt változás, hogy továbbra is hipotetikus technológiára aggatunk jelzőket.

Az ASI lehetne simán Szuperintelligencia³⁷, felesleges bele a mesterséges jelző. A róla szóló teóriák szerint a gép ugyanis ezen a szinten önmagát tökéletesítve haladná meg az embert, tehát kvázi élőlényként az evolúció következő szintjére (szintjeire) lépne. A gépi öntudat ezen a szinten alapvetően merül fel, mint az AGI-nál, ezt a jövőképet úgy lehetne jellemezni, mintha ismeretlen dimenziókba lépnénk: az ASI akár az emberi számára elképzelhetetlen, sőt valószínűleg felfoghatatlan képességeket, irányokat valósítana meg.³⁸ Más szóval nem csupán sokkal több neuronnal lenne képes dolgozni, mint az ember, nem csupán mindent jobban csinálna, mint egy ember, de talán az emberben nem létező kognitív képességekre is szert tehetne.³⁹

Ez tehát az a három szint, amely kifejezetten a gépi oldalát írja le a fejlődésnek, azonban hozzá kell tenni, hogy ezen a szinten az emberrel való együttműködést sem a hagyományos kommunikációs sablonokkal (érzékszervek, nyelv) képzelik a legtöbben. Általában közvetlen gondolatvezérlésre gondolnak, amelyet vagy az agyhullámok pontos megértése útján vagy az agyba telepített implantátumok segítségével valósítanának meg. Így jutunk el az emberek és gépek egyesítéséhez, amely nem szorosan MI-szint, hanem valami más.

Ezért is alkalmazzák erre gyakran a *szingularitás* kifejezést, melynek ezt az értelmezését az űrfizikából vette át a sci-fi műfaj, amely révén 2005-től a technikai jóslatokban is megjelent.[31] A nevét onnan kapta, hogy a fekete lyukak közelében tapasztalható gravitációs szingularitás értelmezhetetlenné teszi a fizika megismert szabályait. Hasonlóképpen teszi értelmezhetetlenné az eddigi emberi kultúrát az emberi képességek vízionált kiterjesztése, mely által az elme közvetlenül, a testi kommunikációs mechanizmusok (kéz, hang, érzékszervek) megkerülésével vehetne igénybe a számítógépes szolgáltatásokat.⁴⁰ Ez tehát egy szélsőséges diszrupció, ami nem csupán az ipart forgatja fel, hanem az ember önértelmezését is. Az elképzelés szerint az

³⁷ Egyébként a névadó is csak így hívja a teóriát taglaló könyvében: [28]

³⁸ Más megközelítésben új létréteg jönne létre ld. [29]

³⁹ Közérthető áttekintés az ASI-ról: [30]

⁴⁰ Egy egész fejezetet szentel ennek pl. [32]

agykérget összekapcsolva a felhővel a gondolkodásunkat emelhetjük új dimenzióba, az MI-képességeit saját képességünként használhatnánk.⁴¹ E szerint a testi képességek kitágításával, és szaporodás és a halál eltörlésével (mesterséges test által) az élet fogalma⁴² is megváltozhatna. Az ember kiterjesztésére sokirányban zajló törekvésekről említés szintjén túl nem térek ki.⁴³

1.3. ÉRZELMI INTELLIGENCIA A GÉPEKBEN

Az fentebb (I.2.2.) felsorolt számos intelligencia-típus közül az érzelmit választottam annak illusztrálására, hogy az okosságtól eltérő emberi szellemi képességek modellezése mennyire sajátos megközelítéseket igényel. Azért az érzelmeket választottam, mivel ez a terület „ezüst-érmes” a fejlesztésekben. A mesterséges intellektus terjedése mellett ugyanis egyre nagyobb lett az igény olyan gépekre, amelyek a mostaniaknál „emberibben” képesek működni. Ehhez pedig elengedhetetlen az érzelmelek gépi érzékelése és utánzása: vagyis növekszik a kereslet, tehát erős a fejlesztői motiváció is, így ez a terület is igen dinamikus fejlődik. Azonban, bár eredményei sokszor látványosabbak, mint a gépek „okossága”, mégis kevésbé közismert, hogy ez a terület külön fejlődik, az érzelmelek gépi kezelésének eredménye a közfelfogásban összefolyik az MI-fejlesztések eredményeivel. Például a személy felismerése arc alapján az MI-funkciók körébe tartozik, és sokan nem tudatosítják, hogy egy arckifejezés jeleinek észrevétele, harag és öröm arctól elvont felismerése az érzelmelek kutató számítástechnika eredménye. Ez az aspektus az MI-fogalmakban is alig jelenik meg, azonban alaposan körbejárva a területet világossá válik ennek hiánya. Ezért ebben az alfejezetben áttekintem a terület fontosabb részeit: fejlődését, helyzetét és korlátait, s hogy egy gép mennyiben és mi módon képes kezelni az érzelmelek. Érdeemes azzal kezdeni, hogy bemutatom az érzelmelek sokrétű felértékelődését a társadalomban, így a technológia mögötti motivációk is megragadhatóbbá válnak.

1.3.1. Háttér a technológia mögött: az érzelmelek fontosságának története dióhéjban

Európa történelmében jól megfigyelhető egy olyan tendencia, hogy az intellektuális előterbe helyező korszakok után az érzelmelek is felfedezik újra és újra. Ahogyan a felvilágosodás racionalizmusát a romantika érzelemközpontúsága követte, úgy a II. világháború utáni modern

⁴¹ Kurzweil nem csupán kitarthat korábbi jóslata mellett, szerinte, hanem legújabb könyve szerint ez már a küszöbön van. [33]

⁴² Az élet fogalmának újragondolása a technológia tárgyalt szintjei mentén már jó néhány éve elkezdődött. Példa: [34]

⁴³ Egy szakdolgozatban kaptam erről átfogó képet, mely alapján világos, hogy a téma összetettsége miatt említésen túl nem foglalkozhatok vele. [35]

racionalista korszak értelemközpontú világa is lassan szétmállott, fokozatosan egy posztmodern korszaknak adva át helyét, ahol az érzelmet mind tudományosan mind pedig a társadalmilag felértékelték. A háttérben mindig zajlottak olyan folyamatok, melyekre később ezek a váltások alapultak, így a pszichológia szakirodalmában is régóta téma volt az érzelmek objektív kutatásának lehetősége. Persze a kérdéskör régebbi, az ember ősidők óta szétválasztotta az érzelmeket az értelmi mozzanatoktól, így a filozófiában, illetve gyakorlati szempontból az etikában külön kezelték azt.

A pszichológia is némileg eltérően közelítette meg a két terület megközelítése, bár sokszor tűnt úgy, hogy sikerül az értelmet és érzelmet a közös, objektív és mérhető anyagi oldal felől megismerni. Magyarán racionalista korszakokban racionalista módon próbálták az érzelmeket is megragadni. Ezek közül itt csupán egyet emelnék ki azzal a céllal, hogy rámutassak, hogy már a 19. század végén megérett az idő az újszerű érzés-elmélet teóriák megfogalmazódására. Akkor egymástól függetlenül jelentett meg ilyen tárgyú írást 1884-ben James amerikai és 1885-ben Lange dán filozófus. A róluk elnevezett James-Lange elmélet megfordította az észlelésre adott testi reakciók és az érzelmek addig vélt oksági sorrendjét.[36, o. 225–226] Kimutatták, hogy a testi „reakciók” előbb jelenhetnek meg, mint az érzelmek. Ebből, és a hasonló eredményekből a materializmus talaján gyorsan kialakult az a nézet is, hogy „az érzelmek csupán kémia”, hiszen az érzelmek vegyi (illetve biokémiai) stimulációval is kiválhatóak. Ez alapján a modell alapján azonban az agy számítógépes másolásakor nem tudták az érzelmeket algoritmizálni. Teljesen más irányból közelítve, nem a neurális ingerek méréséből, és nem a biokémiai anyagok ismerete felől vált képessé a gép az érzelmek kezelésére.

Az ehhez szükséges szemlélet a számítástechnika és a racionalizmus csúcsán és a szexuális forradalom érzelem-központúságának hajnalán jelent meg. Célja nem érzelem-kezelő gépek megalkothatósága, hanem az ember érzelmi oldalának megragadhatósága volt. Azt a paradigmaváltást az érzelmek kutatásában, melyet később az elektronika felhasználhatott, Sylvan S. Tomkins amerikai pszichológus és teoretikus affektuselmélete⁴⁴ hozta el. Ez a megközelítés nem csupán a pszichológiának adott lökést, hanem pár évtizeddel később az informatikának is. Ennek kulcsfogalma az *affektus* az érzékelhető érzelmet jelenti, vagyis a belső érzelmeket kapcsolja össze az érzelmek fizikálisan észlelhető, külső megjelenésével, például az öröm érzését észleljük egy mosolyban. Az áttörés abban a meglátásban áll, hogy hiába van lehetőség az érzelmek által kiváltott kémiai és elektromos jelenségek mérésére műszerekkel, erre szolgáló

⁴⁴ Magyarul a „hatáselmélet” kifejezést is használják. Itt az idegen szó megtartása a később bemutatandó affektív számítástechnikával való kapcsolat miatt indokolt. Vö. [37]

emberi receptoraink nincsenek. Tehát egy interakcióban nem vagyunk képesek közvetlenül érzékelni mások érzelmeit. Kizárólag érzékszerveinken keresztül észleljük azokat, és valójában ezekre az észlelésekre reagálunk.

Tomkins kilenc alap-affektusba rendszerezte és három féle előjelű osztályokba sorolta azokat a sémákat, melyeket a másik emberen felismerünk:⁴⁵

- I. pozitív affekturok: 1. érdeklődés-izgalom, 2. élvezet-öröm;
- II. semleges affektus: 3. meglepetés-döbben;
- III. negatív affektusok: 4. szorongás-gyötrelem, 5. harag-düh, 6. félelem-rettegés, 7. szégyen-aláztatás, 8. undor (izból), 9. bűz (szaglásból)

Később különböző számú affektussal dolgozó sémák születnek, és több tudományos elmélet is létezik az alapérzelmek csoportosítására.⁴⁶

A pszichológiában az affektív jelenségek átfedéseiként választják szép az érzelmeket (*emotion*), és ennek konkrét, szubjektív eseteit, ahol megkülönböztetik az érzést (*feeling*), a hangulatot (*mood*) és az attitűdöt (*attitude*).⁴⁷ Jelen megközelítést felesleges elbonyolítaná, ha ilyen finomításokat is vizsgálnánk. Így az alábbiakban az „érzelmek” kifejezésben vagyunk kénytelenek utalni mindenre (nem célunk az érzések, hangulatok és attitűdök gépi szimulációját külön vizsgálni). Témánk szempontjából két okból is igen jelentős Tomkins munkássága. Egyrészt affektus-elmélete adta meg a kulcsot a gépi érzelem-kezeléshez: rámutatott ugyanis, hogy a gépeknek elegendő csupán felismerni az érzelmet kifejező affektust, illetve szimulálni, érzékelhetővé tenni azt az affektust, amelyet egy ember érzés-kifejezésnek érzékel. A másik ok, ami miatt fontos, hogy ő a *script*-elmélet megalkotója is.⁴⁸

A *script*-elmélet szerint az emberi viselkedés leírható minták és érzelem-mintázatok segítségével. A gondolkodás-minták fontossága az emberi megismerésben azóta számos területre hatással volt, és ezek a minták vezettek paradigmaváltáshoz az MI-kutatásokban is. Itt ugyanis hasonló jellegű elakadás jelent meg, mint amelyet az imént az érzemeknél bemutatam, csak kisebb léptékű. A számítástechnika hőskorszakban ugyanis a racionalitást hangsúlyozva még úgy vélték, hogy a gépi logika döntéseinek tökéletesítéséhez nagy mennyiségű „alapvető adat” szükséges. A '80-as éveket követően jöttek rá a fejlesztők, hogy a *script*-elmélet gondolkodás-mintáinak segítségével sokkal hatékonyabb MI-modell tervezhető. Ez a minta-alapú matematikai modell mára jól bevált technológiává fejlődött, sőt a jövőre nézve is joggal gondolhatjuk,

⁴⁵ Ehhez jó ábra is készült: [38, o. 68]

⁴⁶ További csoportosítások: [39, o. 293]

⁴⁷ Elsősorban ezen kötet *Nyelv, tudat, gondolkodás* c. részfejezete: [39, o. 290]

⁴⁸ Bár ezen terület alapműve tanítványa nevéhez fűződik: [40]

hogy az MI fejlődése erre alapulva halad tovább. Itt sajnos csupán említeni tudtam, mivel az affektuselmélet a számítástechnikára gyakorolt hatására kell koncentrálnom.

Térjünk vissza a társadalmi folyamatok tudományra gyakorolt hatásához, hiszen nem elhanyagolandó, hogy minden kutatóra akarva akaratlanul is hat a korszellem. Vagyis az érzelmek társadalmi felértékelődése hatással van érdeklődésükre, ötleteikre és intuíciókra is. A hatvanas évektől nyugaton rohamosan bontakoznak ki érzelmeket felértékelő mozzanatok. Ezek a hippimozgalomban és a szexuális forradalom légkörében jogi szinten és hétköznapiakban egyaránt jelentős gyakorlati változásokat eredményeznek, és ekkorra a racionalista tudományra gyakorolt hatásuk is kézzelfoghatóvá válik. Itt elsősorban Paul Ekman nevét kell kiemelni, aki már 1972-es könyvében tárgyalta szerzőtársaival az érzelmek felismerhetőségét az arcon, feltérképezte az arc izmait, így képes volt ezek mozgásait figyelni és dekódolni, s például a valódi örömet az álmosoltyól megkülönböztetni.⁴⁹A nyolcvanas évekre így a tudomány szintjén is népszerűvé válik az érzelmek felértékelése, a mérhetőségének és működési modelljeinek csi-szolása. Ekkortájt válik le az érzelmek mérése az intelligenciával kapcsolatos vonalról, elsősorban a fentebb említett (I.2.2.) többszörös intelligencia-elméletre alapozva.[25] Ekkor kezd jobban kibontakozni az affektuselmélet is, és vele az affektív tudomány.[42] Ez utóbbi olyan kidolgozott alapokat (modelleket, kategóriákat) rak le, melyet aztán a műszaki kutatók is alkalmazhattak az érző gép megvalósításán gondolkodva.

A '90-es évektől, az érzelmi intelligencia elkülönített kutatása óta[43] a különböző pedagógiai és pszichológiai modellek az emberek intellektuális és érzelmi képességeit eltérően kutatják, kezelik és mérik. Ez kihat az önismereti, konfliktuskezelési módszerekre is, de eredményeit felhasználják a marketingben, és a média-manipuláció révén a tömegeket érzelmileg befolyásoló információs műveletekre is. Tehát, mivel társadalmilag is ismertté válik az ember érzelmi oldalának megragadhatósága, kezelhetősége, így evidens, hogy a gépekkel szemben is fokozatosan egyre több érzelmi elvárást támaszt az emberiség. De térjünk vissza arra, hogy a gépi érzelmek kutatása hogyan vált fokozatosan új tudományággá.

1.3.2. Az affektív számítástechnika kifejlődése

(1.) Az affektív számítástechnika fogalma és kezdetei

Az érzelemkezelő számítógépekre Rosalind W. Picard vezette be az *affective computing* szakkifejezést, egy 1995-ben megjelent cikkében.[44] Számos, később megvalósuló technológiát megjósol ebben az írásában, a hordozható számítógépektől (nála *affective wearable*

⁴⁹ Újabb kiadásban ld. [41]

computers) kezdve egészen a média-felhasználásokig. Rámutat arra is, hogy a Turing-teszt⁵⁰ elképzelhetetlen az érzelmek gépi érzékelése és szimulálása nélkül. Az eredeti teszt tükrözi a fentebb bemutatott kort, melyben az ember fő vonása az értelem, szerintük ez különbözteti meg az embert a géptől és az állattól egyaránt. Ez a nézet vált meghaladottá, és ezt váltja le az a személet, melyet a kutatónk cikkét is áthatja, amikor rámutat az érzelmek figyelembevételének fontossága mellett, figyelembevehetőségének akkoriban csírázó lehetőségeire is. Azonban nem ő volt az első, aki ezzel foglalkozott, hiszen az imént említett Ekman óta egyre többféleképpen foglalkoztak az érzelmek gépi felismerésével, így a '90-es évek elején már javában folytak szóbeli intonáció-felismerő[46, o. 1097–1108] vagy kézmozdulat-felismerő[47] modellekre irányuló kutatások is.

A technológia kibontakozásának ismertetése előtt térjünk rá az *affective computing* fogalmának elemzésére, hiszen ez nagyon sokat és lényegeset adhat hozzá az MI fogalmához. Egyik rövid magyar meghatározás szerint a kifejezés magyarázható úgy, hogy „érzelmekkel operáló számítástechnika”⁵¹. Itt azonban inkább az „affektív számítástechnika” terminussal utalok rá, mely szerintem pontosabb.

Az affektív számítástechnika tehát olyan rendszerek és eszközök tanulmányozását és fejlesztését jelenti, amelyek képesek felismerni, értelmezni, feldolgozni és szimulálni az emberi affektusokat.[48]

Fontos hangsúlyozni, hogy a „gépi érzélem” éppen annyira nem érzélem, mint amennyire a mesterséges intelligencia nem bölcsesség. Nehogy bárki azt gondolja, hogy ezek a gépek képesek érzésekre! Mivel a fentebb említett tomkinsi affektív pszichológiára alapul az irányzat, ezért mondhatjuk, hogy az affektív számítástechnika egy interdiszciplináris fogalom. Az ilyen irányú eredményekhez ugyanis sokféle tudományág kibontakozására és összefonódására van szükség, a számítástechnikától a pszichológiáig, így az affektív számítástechnika a kognitív tudomány egy részterületének is tekinthető.

Ám az affektus sem teljesen pontos jelzője ennek az irányznak, hiszen egyik nyelvben sem kétirányú kifejezés, pedig az általunk vizsgált szóösszetételben egy 'oda-vissza hatás' jelenik meg. Picard egyaránt affektív számítástechnikának hívja azt, amikor a számítógép emberi érzéseket ismer fel (input affektusok), és azt is, amikor érzések szimulálására képes a kompjúter (output affektusok). Így négy féle géposztály létezhet – ezt példázzák napjaink jelentősebb eredményei is:

⁵⁰ A teszt azzal méri a gépek képességeit, hogy a tesztelő meg tudja-e különböztetni, hogy kérdéseire gép vagy ember válaszolt-e? Az eredeti teszt: [45]

⁵¹ Online kézikönyvekben felbukkan az „érzelmi számítások” és az „affektív számítások” kifejezés is.

1. csak input: a szív-, vér-, bőrreakciók fiziológiai formájában megnyilvánuló érzelmek felismerése
2. csak output: gépi esztétikai képességek előállítása
3. input és output: az arcmimikában, a hangban, a kéz- és testmozdulatokban megnyilvánuló érzelmek felismerése és előállítása
4. egyik sem (nincs érzelmi része).

Azért is külön kell kezelnem majd az input és az output oldalt, mivel jelentősen eltér a hozzájuk szükséges technológia. Az affektus tehát egyszerre jelenti az érzékelést és az érzelmi kifejezést, napjainkra mégis az affektív számítástechnika itt vázolt eredeti jelentésének szűkülése figyelhető meg. Egyszerűen az MI-t használó technológiába beleértik, mintha evidens lenne, hogy egy MI rendszer érzelmeket is kezel.

Ennek elkerülésére bevezetem a „mesterséges érzelm” (MÉ) terminust. Mert igaz ugyan, hogy főleg az MI-vel összefüggésben, annak bővítményeként használható az MÉ, és alább is így vizsgáljuk, ám sokszor szükséges lenne megkülönböztetni ezeket a csak intelligenciát szimuláló gépi rendszerektől. Arra pedig, amikor a két technológiát együtt használjuk, a magyar *MÉ2 rendszer* (esetleg MÉÉ, Mesterséges Értelem és Érzelm) vagy az angol *AIE system (Artificial Intelligence and Emotion)* kifejezést javaslom, és ezt használom, mert ez mutat rá arra, hogy a két oldal külön kezelendő.

(2.) Az érzelemmotortól a mesterséges empátiáig

A fogalmi vizsgálat után át kell tekintenünk, hogyan alakult, és fejlődött a mai szintre ez a technológia. Az affektív számítástechnika egyik első, boltokban is kapható reprezentánsa egy csak output gép volt a kétezres évek elején. A PlayStation-2 játékkonzol chipjében „érzelemmotor”-nak (*Emotion Engine, EE*) nevezték el azt a technológiát, amely a 3D játékgrafikát megvalósította.[49] Nem pontos a szóhasználat, hiszen a technológia lényege nem kifejezetten az érzelmek kifejezése, hanem a gyorsabb és jobb grafikai ábrázolás volt. De előre mutató felismerés ezt azzal jellemezni, hogy a szereplők arcán az érzelmek ábrázolása is lehetővé vált. A neuronhálós megközelítés helyett akkoriban egy speciális skalárvektor-architektúra is elegendő volt a korábbinál szebb grafikára, egészen 2012-ig gyártották a különféle EE-chipeket.

Fogalmilag egyébként felvethető, hogy ennek mintájára a mai, még finomabb érzelmi kifejezésre alkalmas szoftvereket, sokmagos videokártyákat és egyéb eszközöket is az affektív számítástechnikához sorolni, hiszen az affektusokra irányuló kutatásokat használják fel. Számos egyéb projekt zajlott az érzések gépi analizálására vagy előállítására már az évezred elején, amint arról egy 2005-ös nemzetközi konferencia is tanúskodik.[50] Ennek előadásai között

találhatóak beszámolók a mimika, a kézmozdulatok, a testbeszéd vagy a hang felismerésére, vagy érzelmes hang, érzelmeket kifejező arcok megjelenítésére – tehát már ekkor megjelent a terület ma eredményeinek minden kezdeménye. Megemlítendő itt, hogy Marvin Minsky figyelme is az mesterséges érzelmek kutatás felé terelődött[51] (az ő neve fontos az MI történetében, és akihez számos fejlesztés és találmány kapcsolódik). Ez jól mutatja, hogy a szakmán belül felfutott az érzelmi terület, de várni kellett még azzal, hogy a közvélemény számára is érzékelhető legyen ez a robbanás.

Tehát sok kutató korábban is használt MI-modelleket gépi érzelmekhez. Ám a megfelelő affektus-adatbázisok hiánya és az MI-modelleket támogató hardverek elégtelensége az ötleteket szakmai körökön belül tartotta. A felhő-technológia terjedésével nyíltak meg az első lehetőségek az MÉ2 elfogadható minőségű használatára. A 2005-ös Google Talk még kicsit géphangon beszélt, lassan csiszolódott ki olyan MÉ-felhő, melynek angol nyelvű beszéde már élethűbb (például az Amazon Echo/ Alexa 2014, a WaveNet 2016), mára egyszerűbb feladatokra általánosan jól alkalmazhatóvá vált a beszéd-értés, és magyarul is egyre kevésbé gépies a beszéd-szintetizálás.

A beszélő gépek példáján egyben rávilágíthatunk arra a már említett sajátosságra, amelyet fontos kiemelni az MÉ fejlődésének áttekintésekor: sok esetben elmosódnak az affektív számítástechnika határai. Ha *beszédet* fejlesztenek, annak kifejezőképességéért és nyelvhelyességéért az MI-modulok felelősek, de az élethű, érzelmes hanglejtésért az MÉ. Ezek együtt, egyszerre teszik értelmi és érzelmi szempontból valóságosá a gépi hangot.

A 2005 utáni korszakra jellemző még, hogy az audio- vagy video-szenzorok megújulásán túl teljesen újszerű érzékelők sora jelenik meg: kezdve a Nintendo cég Wii mozgásérzékelő játékkonzoljától (2006),⁵² a fizikálisan kapcsolódó szenzorokkal végrehajtott érzelm-felismerésen át (szemkövető szemüvegek, EEG-távírányítók⁵³) egészen a mikrohullámú test-radarokig (ld. következő szakasz). A játékok mellett az érzelmfelismerők húzóágazatává kezd válni például az autópálya is a fáradtság-ellenőrzőkkel, említhetően a vállalati munkatárs-jóléti alkalmazások⁵⁴ de a katonai kutatások szintén segítik a terület fejlődését.

Ezek közül azt a török katonai kutatást szeretnénk röviden bemutatni,[53] amely a hagyományos neuronháló érzelmi bővítését vetette fel a célpontfelismerési képességek javítására. Dr. Khashman kutatócsoportja 2009-ben egy olyan megerősítő tanulási modellt alkotott, melybe

⁵² Ez egy kézben tartható, de vezeték nélküli alkatrész (*remote*) mozgását detektálta infravörös érzékelővel.

⁵³ Elektroenkefalográfiás fejérzékelő: nem hagyományos effektust érzékel, hanem elektrofiziológiai változásokat, idegsejtek elektromos aktivitását méri valós időben.

⁵⁴ Ez már 2005-ben prognosztizálható volt olyan ötletekkel együtt, mint a csoportdöntési segédeszközként való használat vezetők számára, hogy feltérképezhessék a beosztott állományuk hangulatát. Ld. [52, o. 97]

két érzelmet vett fel: a szorongást és a magabiztosságot. Abból az emberekre jellemző tényből indultak ki, hogy amikor új feladatot tanulunk, szorongásunk szintje magas, a magabiztosságunk pedig a tanulás kezdetén alacsony, ám idővel, a tanulás és pozitív visszajelzések által a szorongásszintünk csökken, míg a magabiztossági szintünk nő. Így hozhatunk a tanulás révén jobb döntéseket, mégpedig rövidebb idő alatt. A modell beváltotta a reményeket. A tesztek alapján a szorongási és magabiztossági együttható beépítése a rendszert nem csupán hatékonyabbá tette a célok azonosításában, de egyben gyorsabbá vált a tanulás és a döntéshozatali idők szempontjából is. Ez a modell példa arra is, hogy létezhetnek egyéb érzélem-modellek, melyek nem az affektus-modell változataként funkcionálnak, és nem emberibbé teszik a rendszert, hanem hatékonyabbá válik általuk egy hagyományos MI.

Végül a *mesterséges empátia* kifejezést szeretném bemutatni, melyet Minoru Asada egy 2012-es tanulmányában vezetett be.[54] A kutató már 2001-től a gépi érzelemmel kapcsolatos témákkal foglalkozott, ebből nőtte ki magát ez a vizsgálati terület. A kifejezés gyakorlatilag egy input-output modellt takar, amikor egy gép (például robot) egy adott emberi érzélemre a megfelelő érzelemmel válaszol. Ez a kifejezés jól példázza azokat a konvergencia folyamatokat, melyek mára áthatják számítástechnikát és az MI-t. Egy adott fázisban külön tud csak fejlődni egy-egy technológia, abban a fázisban sajátosságaira koncentrálna sikerül megoldani számos kihívást. Ezek az eredmények később gyakran beleolvadnak nagyobb rendszerekbe, ez esetben például ma már az MI rendszerek empátiára tanításáról beszélnek.[55]

(3.) *Érzélem-chip, affektus-radar, mesterséges empátia*

Az utóbbi évtizedben ugyan az MÉ2 ugrás-szerűen elterjedt, azonban a fentiekhez képest kevés diszruptív újdonság jelent meg. Erre az időszakra egészen napjainkig egy horizontális fejlődés jellemző. Egyrészt a fenti trendek finomodása érzékelhető. Így az újabb rendszerek egyre pontosabban kezelik az affektusokat is. Ez igaz mind az input, mind az output irányokra, sőt az output irány lemaradása talán kicsit csökkenni látszik (élethűen szimulálni az érzelmeket jóval nehezebb, mint beazonosítani őket). Másrészt az eddig külön fejlesztett technológiák konvergenciája is erősödik. Ez részben egy természetes spontaneitással megy végbe, de néha az üzleti dinamika mentén történik, technológiák vagy cégek felvásárlásával, és az így kapott fejlesztések tudatos összeépítésével. Az alábbi sorok erre is példaként hozhatóak, ám ennél sokkal fontosabb, hogy így más irányból is mutatható be a várható jövő. Kutatásaim során pár éve találtam az EmoShape nevű vállalkozást, melynek innovációit érdemes bemutatnom annak

ellenére, hogy mára eltűnt, pontosabban beleolvadt egy MetaSoul nevű cégbe, melyet több nagyvállalat⁵⁵ üzemeltet.

Az alábbi eredmények előzményei a XXI. század elejére nyúlnak vissza. A cég vezetője 2005-ben egy EU-s díjat⁵⁶ nyert el a 3Dsolar technológiáért. Ez nem más, mint a 3D megjelenítés tökéletesítése, mely az EmotionEngine ismertetésénél leírtakhoz hasonlóan képessé tehet egy arcábrázolást akár az érzelmekre is. Talán ezért is folytatták a kutatást a szempontunkból jelentős EPU-chip névre keresztelt Érzelmi Feldolgozó Egység (*Emotion Processing Unit*) irányába. Ennek első változatát 2015-ben mutatták be. Érdekesebb azonban az EPU-III (2017) verzió, mely több szempontból rugalmas technológiát takart. Három féle kiadása létezett. [56]

1. egy felhő alapú szolgáltatás (mely az megvásárlás után a mai napig működik);
2. egy beszerelhető mikrochip, amit eMCU-nak⁵⁷ hívott a cég (alább bemutatom);
3. valamint egy USB-csatlakozású perifériaként forgalmazott termék, mely kész okoseszközököt volt hivatva érzelmi komponenssel kiegészíteni.

A termékek nem végfelhasználókat, hanem robot és rendszerfejlesztőket céltartott meg, akik a hozzá tartozó szoftverfejlesztő készlet és dokumentáció megvásárlásával kibővíthették saját szolgáltatásait érzelmi összetevőkkel.

Rugalmassága mellett az érzelmekkezelés eddigieknél sokkal nagyobb pontossága, valamint az input- és output-képességek együttes jelenléte tűnt biztatónak. Ugyanis az EPU egyrészt képessé tehetett egy gépet arra, hogy érzelmileg megértse, amit olvasnak neki vagy amit lát, másrészt érzelmi állapotokat és szintetikus érzelmeket hoz létre az intelligens gépekben. A projekt a fentebb bemutatott Ekman-féle modell módosításán⁵⁸ alapul. Működése hasonlít egy színes szkennert működéséhez, ahol egy kép adott pontjának színárnyalata a három alapszín különböző intenzitású keveréséből áll elő. Csak itt a három alapszín erősség-értékei helyett, a modell 12 alap-érzelmének⁵⁹ különböző intenzitásából „keveri ki” (hozza létre) a rendszer az affektus-

⁵⁵ Az NVidia, az OpenAI és a Microsoft cégek <https://pitchbook.com/profiles/company/96013-72#comparisons> Megtekintve: 2024.01.28. Az EPU hardver és a Radar nevű, mindjárt ismertetésre kerülő fejlesztések mára eltűntek, már csak EPU felhő-alapú része még elérhető.

⁵⁶ Patrick Levy-Rosenthalnak az IST díjat ítéltek oda. https://www.euro-case.org/wp-content/uploads/2019/07/Eurocase/PDF/2006EISTP_book.pdf

⁵⁷ Emotional Micro Controller Unit (érzelmi mikrovezérlő), ami a számítástechnika azon korszakára utal, amikor még különböző mikrovezérlők (MCU-k) voltak szükségesek a CPU mellé, különböző vezérlési feladatok ellátására, amelyek később beleolvadtak a CPU-ba.

⁵⁸ A szakirodalom szerint a félelmet, a dühöt, az undort, a szomorúságot, az örömet, a meglepődést és az érdeklődést a legtöbben alapérzelmeknek tekintik, ám ezek közül csupán hat szerepel az EPU-ben. vö. [39, o. 293]

⁵⁹ A 12 itteni érzelmek angol kifejezésekkel: *anger, fear, sadness, disgust, indifference, regret, surprise, anticipation, trust, confidence, desire, joy*. A magyar kifejezések nem mindig pontosan ugyanazt fedik, de nagyjából a következő egyszavas fordítások adhatóak: düh, félelem, szomorúság, undor, közöny, megbánás, meglepetés, előérzet, bizalom (remény), önbizalom, vágy és öröm.

pixel⁶⁰ konkrét árnyalatát, mely az adott pillanat érzelmét pontosan definiálja. Több időintervallumban felvett érzelem-pixelek sorát feldolgozva alakítja ki rendszer azt az érzelem-mátrixot, mely számunkra információval bírhat. Ugyanis az érzelmek inkább egy idő-intervallumban értelmezhetőek, vagyis affektus-pixelek sorozatával, hiszen a gesztusok, a mimika változása vagy a mozdulatok íve, sebessége bír óriási információértékkel. Az EPU érzelmi spektrumát képező mátrix elvileg 64 millió érzelem-árnyalatot kezelhet,[56] és ezen mátrix alapján állítja elő a rendszer a saját reakcióját. Így biztosít egyfajta nagy teljesítményű érzelmi „tudatosságot” a számítógépek vagy robotok számára, melyre megfelelő outputot tervezve a gép viselkedése bizonyos szintű érzelmi intelligenciát mutathat. Ráadásul, mivel más mesterségesintelligencia-technológiákkal (pl. arc-generálással, nyelvi modellel) is képes kommunikálni. Így valósulhat meg az MI és a robotok spontán (nem kódvezérelt) érzelme, tehát akár nevetése vagy sírása, mely hiába csak szimuláció és csak virtuális, mégis egy eddig nagyon hiányzó problémát oldana meg.

Az audio-vizuális affektusokon túl azonban egyéb jelek is vannak, melyekre nekünk nincs ugyan érzékszervünk, de gépeink által érzékelni tudjuk őket – ezekkel a szenzorokkal még tovább fokozható egy ilyen rendszer érzelme-felismerő képessége. Akár egy infravörös tartományra kiterjesztett kamera érzékelheti a testrészek hőmérséklet változásait, akár agyhullámok érzékelésére is ki lehet képezni – de ezeket most csak én vetem fel, a cég egy más ötletet publikált. Ők egy aRadar (affektus-Radar) névre hallgató szenzort fejlesztettek, amely vezeték nélküli (mikrohullámú) technológiával érzékeli az ember légzésének és szívverésének pontos hullámformáit (egy EKG készülékhez hasonlóan). Az aRadar és az EPU-technológiák együttes használatát az okosautók következő generációjában (ExoCar néven) és számítógépes játékokban vélték hasznosíthatónak. Az EPU technológia továbbfejlesztéseként, az NLP⁶¹ rendszerekhez történő integrálás is a célok között volt, azaz élethűen beszélgető gépek, illetve NLG-algoritmusok⁶² létrehozása (ezekről bővebben a II.3.4.-ben). Ezt már azonban a cég új tulajdonosa vitte végbe, és haladta meg. A MetaSoul irányítói az Open-AI segítségével érzelmekre is tanítható avatarokat, digitális személyiségeket kínálnak. Ezek segítségével például egy játék virtuális karakterei dinamikusan képesek fejlődni, és idővel érettebben, vagy megkeseredettebben reagálni a felmerülő szituációkra, így téve élethűbbé a játékelményt.⁶³ Ezen technológiák

⁶⁰ saját kifejezés (!), ez a leírásokban nem szerepel

⁶¹ Natural Language Processing, természetes nyelvfeldolgozás

⁶² = Natural Language Generation, természetes nyelv generálása. Ez némileg eltér az NLP-től, mivel inkább adatok értelmes szöveggé alakításáról szól. ld. [57]

⁶³ <https://metasoul.one/> Megtekintve: 2024.01.28.

utódairól ugyan nem leltem fel adatokat, de az említett Meta-Soul csoporton által feltehetően nagy szerepet játszanak a jelenlegi és várható MI-szolgáltatásokban.

Az aRadar és az általam felvetett a szenzorok tehát kiegészíthetik és jócskán pontosíthatják az MÉ2 gépek által felismert érzelmek meghatározását. Az így kiegészített rendszer becsapásához nem csupán egy ál-mimikát kéne elsajátítani, hanem emellett a szív- és légzésrendszer rezdüléseit, vagy az agyhullámokat is uralni kellene. Kézenfekvő tehát igazságvizsgálásra alkalmazni, vagyis rendészeti, ügyészségi vagy titkosszolgálati felhasználás irányába is fejleszteni. Ez a magyarázat talán segített jobban megérteni az ilyen jellegű gépi tudás határait, és rámutatott arra, hogy az új MI-fogalomban ez az aspektus megjelenítendő. A technológia védelmi alkalmazhatóságára még ki fogok térni (VI.1.5).

1.4. RÉSZÖSSZEFOGLALÁS: HOGYAN LEHET AZ INTELLIGENCIA MESTERSÉGES

Az érzelmek virtualizációja, melyről az előző alfejezet szólt kutatásom egészében több szempontból is alapvető fontosságú, ezért itt külön összefoglalom. Elsősorban kimutattam, hogy ezt az aspektust az MI fogalmi szintjén is szükséges lesz megjeleníteni (V.3.), vagyis a K3 kérdés első felére máris válaszolhatunk igennel: ezt bizonyosan érdemes megjeleníteni az MI fogalmában. Erre majd csak a következő fejezet végén térek majd vissza, mivel annak leírásai ezt a témakört úgyis tovább vizsgálják és kiegészítik. Fontos lesz a fenti alfejezet a második témakör, a mesterséges etika és gépi autonómia szempontjából is (IV.1.): különösen az a jogfilozófiai vetület, hogy a polgárok titkolt érzelmei egyelőre nem képezik személyes és intim szférájukat, mint pl. titkolt betegségeik. Sőt kiemelhetőek a terület a védelmi vonatkozásai is (ld. VI.1.5.), melyek által nem csupán az gépi „okosságot” lehet támadó, megfigyelő, védekező vagy egyéb módon alkalmazni, hanem az gépi „érzelmek” utánzása lehetőséget ad automatizált információs (érzelem-befolyásoló) műveletekre vagy visszaélésekre.

A fejezet többi részéről is elmondható, hogy sikerült megvalósítani az Első Rész hármas céljának (ld. Bevezető) ide tervezett részét, sőt kisebb eredményeket is elértem. Egyrészt tisztáztam számos jelenlegi fogalmi megközelítést, amelyekre a későbbiekben visszautalhatok. Másrészt az egyetemi-jegyzetté alakítás során ugyan nem tervezem felhasználni mindent (pl az intelligencia nyelvi megközelítéséről írtakat), de az anyag nagy részét igen. Harmadrészt az MI fogalom újragondolásához is alfejezetenként tettem megállapításokat. Ehhez az affektív számítástechnika (I.3.) elemzésén túl kiemelendő a jelenlegi MI-megfogalmazások, illetve a párhuzamos definíciók kritikai vizsgálata (I.1) valamint az intelligenciafajták és intelligenciaszintek ismertetése (1.2.3.) Megjegyzendő, hogy sajnos az intelligencia alapos elvi vizsgálatára szűkös volt ez a keret, de jelen kutatásokhoz elegendő. Második rész többi vizsgálata majd felhasználja

ezt az elemzést, különösen az intelligenciaszintek kritikájánál (III.6.2.). Említendő még hogy az ágensek tisztázására (I.1.4) nem csupán azért volt szükség, hogy ne keveredjen össze ez a fogalom az MI-vel, hanem a későbbi pszeudo-MI fogalom (ld. V.1.4) megértését is segítik.

II. TECHNIKAI KÖRÚT A MESTERSÉGES INTELLIGENCIA KÖRÜL

Ebben a részben először azokról a technológiákról adok áttekintést, melyek nagy rendszerek esetében egybeforrtak az MI-vel. Ez után néhány olyan témakörrel adok kiemeltebb áttekintést, melyek a tanulmány valamely eredménye szempontjából jelentőséggel bírnak - ezek szerkesztéskor kerültek egy alfejezetbe.

II.1. MIKT: AZ MI-HEZ SZOROSAN KAPCSOLÓDÓ TECHNOLÓGIÁK

Lassan szinte minden technológia kapcsolatban hozható az MI-vel. Itt muszáj azokat bemutatnom, melyekre a jelenlegi szabályzók és meghatározások is sokszor utalnak. Ezért pár bekezdést állítottam össze azok számára, akik nem foglalkoznak ezekkel a kifejezésekkel napi szinten. Ezek a technológiák jelen témához is nagyon szorosan kapcsolódnak, ezért a terminológiai összefoglaláson túl igyekeztem jelen tanulmány sajátosságaihoz igazított saját megfogalmazásokat adni, ezért kitérek a technológiák MI-hez fűződő kapcsolatára, valamint főbb biztonsági kockázataikat is vázolni. Az egyszerűség jegyében nem láttam szükségét minden szakirodalom megadásának, melyet az elmúlt évek során tananyagként vagy publikációkhoz olvastam, hiszen egyébként közismert technológiákról van szó.

II.1.1. BigData, strukturálatlan adathalmazok

A BigDatát EU szakemberei már a hivatalos definíció első mondatában összekapcsolják az MI-vel: „A Big Data olyan összegyűjtött adathalmazokra vonatkozik, amelyek olyan nagyok és összetettek, hogy feldolgozásukhoz új technológiákra, például MI-re van szükség”. [58] Ez azonban nem mutat rá a kifejezés lényegére, ahogyan egyébként a „nagy” (Big) jelző sem a nevében. Ugyanis a kihívás egyik fele csupán az adathalmaz nagysága, hiszen a hagyományos programozásban is léteztek technikák a milliós vagy milliárdos adatmennyiség feldolgozására. A fenti definícióban inkább az „összetettség” és a „halmaz” a kulcsszó, vagyis az, hogy az adatok jó része „strukturálatlanul” áll benne rendelkezésre. Mit is jelent ez? Egy leegyszerűsített példával: futtassunk egy karaktersort keresést fájlok tartalmában (nem a fájlnevben), és pár ezer fájl átnézésének hosszú perceit vessük össze azzal, amikor egy több tízezer cellás excel-fájlban másodpercek alatt minden találatot megkapunk. Vagy bele lehet gondolni abba amikor a képeinket vagy videóinkat nézegetve keresgetjük, hogy melyiken van rajta egy ritkán látott ismerős... Ezek alapján megsejthető a strukturálatlanság problémája: első példa az elektronikus tárolásban, a második a humánérős keresésben mutatott rá erre.

Az adatbázisok már az ókortól az adatproblémák pontosabb gyorsabb megoldására hivatottak, sokáig nem elektronikus megoldásokkal. Adatbázisnak elvileg *az egymással kapcsolatban lévő adatok rendezett halmazát*-t hívjuk – de igazából inkább elektronikus (relációs) adatbázisokra szokás a szót használni. Az egykor pergamenen vagy papiruszon tárolt jól rendezett (telefonkönyv-szerű) listák helyét idővel átvette a táblázat, ahol a kapcsolat ábrázolása, hogy az összetartozó adatok egy sorban vannak. A világ növekvő adatigénye már a számítástechnika korai időszakában megteremtette annak lehetőségét, hogy ne csupán digitálisan tároljuk a rendezett listákat és táblázatokat, mivel ezek nagyobb mennyiségű adat esetében már áttekinthetetlen és redundáns⁶⁴ módszerré válnak. A redundancia elkerülésére matematikát dolgoztak ki, és már az 1970-es évektől megvalósultak a relációs⁶⁵ elektronikus adatbázisok, melyek máig egyfajta megoldást jelentenek. Ezek számos kis adattábla irányított összekapcsolásával nagyságrendekkel hatékonyabbak a digitalizált táblázatoknál, és hatékonyan lekérdezhetőek (például SQL⁶⁶ lekérdező nyelven).

A gondot itt az igen jelentős humán erőforrás jelentette. Már egy digitalizált táblázat-alapú (pl. excel) feldolgozáshoz is szakértelem kell. De sokkal jelentősebb szakértői erőforrások keltenek az ilyen adatbázisok tervezői és lekérdezői szintjén (mit kell tárolni, hogyan optimális felépíteni, hogyan kell hatékonyan megfogalmazni a lekérdezést stb). Még az adatok alacsony képesítésű feltöltőjének is értenie kell, hogy mit hova visz be az űrlapokon, vagy ha egy fájlt helyez el az adatbázisban, azt milyen címkéssel érdemes ellátni, a könnyebb megtalálás érdekében.

A manapság termelődő adatmennyiség ilyen strukturált rögzítéséhez és feldolgozásához a humán erő kezd kevésnek bizonyulni. Az emberi felhasználók is és eszközeik is egyre több és egyre több, nagyobb és többféle fájlt állítanak elő a CAD tervektől a röntgenképek speciális formátumain át a házi videóvágó programok fájl típusaiig. Az IoT eszközök megsokszorozzák ezt az adatmennyiséget, és csak bizonyos címéket lehet automatikusan létrehozni (pl. egy szenzor helye és a mérés időpontja). Egy térfigyelő kamera esetén kevés. De sokszor az sem elég, ha csak strukturált adatokat tárolunk, ahogyan például egy router forgalmi naplójában sem lehet bizonyos újfajta anomáliákat hagyományos programkóddal észlelni. Tehát a felgyülemelő adatmennyiség egyre változatosabb. Ráadásul elég régóta keletkezik egyre több adat – így hozzá kell venni a problémákhoz a sebesség-tényezőt is: egyre nagyobb sebességgel

⁶⁴ Az ismétlődő adatokat nevezzük redundánsnak, például amikor egyik oszlopban az egyedet halmazokba soroljuk (monitor, nyomtató). A nagyszámú ismétlődés sok felesleges helyet foglal és lassítja a visszakeresést.

⁶⁵ Relációnak az adatkapcsolatokat hívja ez a szaknyelv.

⁶⁶ Structured Query Language, egy adatbázislekérdezések írására kifejlesztett kód-család

termelődnek az adatok. Az eddigiek alapján érthető, hogy a szakirodalom a *változatosság*, a *sebesség* és a *mennyiség* kulcsszavakkal, a három V-vel kezdett el rámutatni a BigData probléma lényegére, melyhez egyre több további „V” kapcsolódik.⁶⁷

Az „adatbányászat” kifejezés kapcsolódik szorosan a BigData-hoz, mely a hagyományos adatlekérdezés helyébe új dimenziót nyit. Az adatbányászat nagy mennyiségű adathalmaz olyan irányított elemzési folyamata, mellyel új *információk* nyerhetők ki a halmazból, például trendek és mintázatok észlelése által. Vagyis itt nem csupán arról van szó, hogy egy képfelismerő MI segít megkeresni egy arcot egy videón, tehát megtalálni a nehezen megtalálható adatot. Ez a kialakulóban levő új szakma a meglévő adatok között az eddig nem ismert kapcsolatokra (összefüggésekre) segít rámutatni. Tehát a relációs adatbázis emberileg megadott, közismert relációi mellé, eddig ismeretlen relációkat képes megtalálni. Összegezve: ebben az új paradigmában az adattalálalatokból kapott információkat egészíthetik ki a kapcsolattalalatok új dimenziós információi. Az ilyen mintázatfelismerést támogató a MI rendszerek által válnak a számítógépek adatfeldolgozó rendszerekből valódi információfeldolgozó rendszerekké, hiszen az adatokból már nem csupán az emberi értelmezés által válik információ, hanem a gép képes rámutatni az új információkra.

Ez a technológia tehát számos ponton összeszővődik az MI-vel. Már a tárolás technológiája az úgynevezett neurális adatbázis irányába halad (ld. V.1.2.), és a lekérdezésben is alapvető szerepe van az MI-nek. Az használt technológiák kihívásain túl a visszaélések lehetősége az adatokhalmazt jellemző „V betűk” következtében is markánsan jelen van.⁶⁸ A technológiáktól ugyanis független, hogy a nagy mennyiségű információ birtoklása egyfajta hatalmi tényezőként is értelmezhető. Azaz az információ napjainkra a pénzhez hasonló erővé, vagy akár „fegyverré” is válhat, mely a politikai és gazdasági frontokon vethető be. Egy BigData technikai üzemeltetője, ilyen hatalommal is rendelkezik akkor is, ha hivatalosan (jogilag) nem is ő az adatok birtokosa. Nehéz elkerülni, hogy ezeket az adatokat a cégek, vagy állami szándékok tényleg nem használják-e fel, vagy bizonyítani ennek megtörténését. (Erre bővebben kitérek III fejezetben.

⁶⁷ A három V az angol szavak kezdőbetűi Variety, Volume, Velocity. A további V-k közül néhány: Variability (az adatok variálhatóságát jelöli), a Virtual (az adatok virtuális voltát jelzi), a Veracity (az adatok integritását jelöli) vagy a Value (az adatokban rejlő hasznosságot jelöli) ([59]

⁶⁸ Két ilyen visszaélést is bemutat: [60]

II.1.2. Felhőtechnológia és technológiai virtualizáció

Sok forrás alapján dolgoztam át az Amerikai Szabványügyi hivatal által adott meghatározást⁶⁹. Először nagyon egyszerűsítve, aztán picit pontosabban (és a felvetett kifejezéseket magyarázva) a digitális felhő lényege a következő:

1. felhasználóként: a hálózaton keresztül úgy érhetőek el a programok, a szolgáltatások és a szükséges adatok, mintha csak a saját gépünkön dolgoznánk.
2. szolgáltatóként: egy vasúti szállítványozóhoz vagyunk hasonlóak. Csak mi nem vagonokat kapcsolunk össze, hanem hardveres erőforrásokat, és nem a szerelvényt osztjuk szét a bérlők között, hanem az összeadódó számítási és tárolási kapacitásokból adunk bérbe minden felhasználónknak egy saját látszatkörnyezetet az ő szükségletei szerint.
3. műszaki oldalról: *Elosztott erőforrásokon* és egymástól elzárt konténerekben jön létre *virtuálisan* az, amit a bérlő igényel. A konténer biztosítja, hogy senki nem látja mit csinálnak a többiek, az elosztás által pedig akár „sok vagonon átnyúló konténert” adhatunk. Ráadásul a bérbe adott konténer egy része (processzormag, memória) felszabadul amikor valaki kilép, tehát az erőforrásokat másnak adhatjuk ki.
4. bérlőként: olcsóbban jutok óriási tárhelyhez vagy nagy számítási kapacitáshoz (pl. nagy mennyiségű adat elemzésére), mint ahogyan olcsóbb 1000 tonna árut vasúton szállíttatni, mint egy saját kamionnal, ráadásul üzemeltető, karbantartó, és biztonsági embereket sem kell alkalmaznom (a példában nem kell sofőr, szerelő, műhely vagy biztonsági őr a szállítmány kíséretéhez).
5. biztonsági oldalról a két fő kockázat, hogy ez a technológia nem értelmezhető hálózat nélkül, valamint, hogy túl nagy bizalom szükséges a felhőt szolgáltató vállalat felé (ezért pl. titkos anyagot senki, cégek sem tárolnak bérelt felhőben).

Alább nem a fenti pontokat részletezem tematikusan, hanem csak a felmerült és a kapcsolódó fogalmakról írom le a legfontosabbakat.

⁶⁹ „A felhőalapú számítástechnika egy olyan modell, amely lehetővé teszi a mindenütt elérhető, kényelmes, igény szerinti hálózati hozzáférést egy megosztott konfigurálható számítástechnikai erőforrások (pl. hálózatok, szerverek, tárolók, alkalmazások és szolgáltatások) közös használatához, amelyek gyorsan, minimális kezelési erőfeszítéssel vagy szolgáltatói interakcióval biztosíthatók és felszabadíthatók.” [61]

- A *virtuális* kifejezés röviden látszatot, látszatvalóságot jelent. A virtualizációs technológia esetén mindig egy számítógépes szoftver (*hypervisor*) hoz létre valamilyen látszatvalóságot. Ez vagy látszathardver vagy pedig egy olyan szoftverszolgáltatás, mely nincs telepítve a felhasználónál mégis (hálózaton keresztül) igénybe veheti. Nem csak felhőn jöhet létre, sőt már a harmadik generációs gépekben (az 1970-es években) megjelent⁷⁰ bizonyos fajta virtualizáció. Pl. a „virtuális memória” esetén az operációs rendszer „csapja be” a programokat, amik úgy látják, mintha bőven lenne belső memória – pedig valójában a lassú háttértár egy részét használja úgy az operációs rendszer, mintha az is gyors belső memória lenne. A mai virtualizáció inkább a felhasználót „csapja be”, ő látja úgy, hogy működő hardvereket kezel és telepítés nélkül elér pl. egy szövegszerkesztőt, pedig mindezeket akár egyetlen (akár nem is túl drága) gép emulálja számára.
- Az *emuláció* magyarul utánzás, jelen esetben valamely hardver vagy szoftver funkciójának vagy műveletének, vagy akár teljes egészének utánzása (reprodukálása). Például elavult processzorutasításokat vagy szoftvereket régóta emulálnak modernbb gépekben a kompatibilitás miatt (hogy régi funkciók még elérhetőek legyenek). A virtualizáció és emuláció között nincs éles határ, hiszen a *látszat*-ot az *utánzás*-ok hozzák létre.⁷¹ A *szimuláció* és *emuláció* szavakat újságcikkek néha összekeverik. A lényegi különbség közöttük, hogy a *szimuláció* a valóságot akarja utánozni, ám annak komplexitását minden részletében képtelen imitálni (ezért azt mindig egyszerűsíti). Az *emuláció* azonban képes egy másik gép minden részét leutánozni, néha ezt meg is teszi – vagy csak azért nem teszi, mert erre épp nincs szükség.
- „*Elosztott erőforrás*” a neve a felhőhöz szükséges másik fő módszernek. Ugyanis a fenti vonatatos hasonlattól eltérően a szaknyelvben nem a gépeket kapcsoljuk össze, hanem a feladatokat osztjuk szét sok gép között, így a feladatok feldolgozása egyszerre több számítógépen, párhuzamosan történik. Az erőforrásgazdálkodáshoz tartozik a fent említett konténer-technika, mely által ebből az egyesített óriásgyurmából az igényekhez pontosan hozzáformázott darabokat adhatunk ki. Azért is olcsóbb, mivel csak addig foglalja le az erőforrást felhasználó, míg igénybe veszi (a felhőtárhely ez alól kivétel), utána újra belegyúrjuk a nagy gyurmatömbbe.

⁷⁰ Például az IBM VM/370 rendszere már használt ilyet. [62, o. 19]

⁷¹ Ha akarunk különbséget, akkor a virtuális szót inkább komplex, bonyolultabb, jól paraméterezhető emuláció-halmazra értik.

- A „*virtuális valóság*” fogalma említendő még itt, mely alatt egy olyan jól szimulált és teljes látszatvilágot értünk, amely érzékszerveink számára az igazi világ érzetét kelti. A háromdimenziós videojátékoknál merült fel ez a kifejezés, ahol a játék mesevilága nem csupán a képzeletre van bízva, hanem látható, mégis interaktív – megtehetjük benne, amit a valóságban nem.⁷² Mára az ilyen teret még közelebb hozták a valós világhoz, például amikor mozgásunkat érzékelve, azzal összehangolva mozgatja a rendszer a szemünkbe vetített háromdimenziós digitális világot, a testérzékelésünk és látásunk becsapott harmóniája agyunkban akkor is valóságérzetet kelt, ha rajzok között mozgunk.

Ezek a módszerek technikailag megvalósíthatóak gyengébb hardvereken, és kisebb (belső) hálózaton is, így a felhő is. Ám szempontunkból inkább a szupergépekkel megvalósuló, interneten elérhető felhőszolgáltatók lényegesek. A teljesség kedvéért: a fentebb idézett NIST dokumentum 5 jellemző, 3 szolgáltatási és 4 telepítési modell segítségével pontosítja meghatározást. [61] Ezeket a szakma úgyis ismeri, itt nem alapvetőek, ezért eltekintettem részletes ismertetésüktől. Viszont látványos, hogy a fejlődés alapján máris egyel több szolgáltatási modellt (4-et) tartanak nyilván. Ugyanis a *Kiszolgáló nélküli számítástechnikát (serverless computing)* is ilyen modellként kezelik, ahol egy applikáció kap számára szükséges mennyiségű erőforrást amíg kell neki (pl. egy ChatGPT-re épülő applikáció).⁷³

A technikai oldal pontosításához sok mindenre ki lehetne térni, itt azonban csak azt emelem ki, hogy a meghatározások általában a szolgáltatásként tekintenek a felhőre, ami a szoftveres szint, pedig a fogalomba hasznos lenne beleértetni a megvalósító hardver technológiákat is. A hatékony működéshez szükséges a gépek óriási számítási teljesítménye, a robusztus tárolási kapacitása, és az óriási sebességű adatátvitel mind a belső alkatrészek között, mind pedig hálózaton. Ezek bizonyos szintje fölött kezdett csak elterjedni a technológia, ma pedig a sokmagos erős processzorok, nagy és gyors memória-modulok, az SSD⁷⁴ táruk újabb generációi és az 5G vezeték nélküli átvitel, és minden egyéb korszerű hardver alapvető szerepet játszik népszerűségében.

⁷² Ez a tulajdonság a függőséget és valós világtól való menekülést is okozhat, de gyógyterápiák során is sikeresen alkalmazható (ld. affektív számítástechnika I.3).

⁷³ Ha már ez előkerült, röviden megemlítendő a három régebbi szolgáltatás-modell is: az infrastruktúra-mint-szolgáltatás (Infrastructure as a Service, IaaS, amikor egy virtuális gépet kap a bérlő) a szoftver-mint-szolgáltatás (Szoftver as a Service, SAAS, amikor szoftvert futtathat a virtuális bérleményében), és a platform-mint-szolgáltatás (Platform as a Service, PAAS, melyen a bérlő telepíthet és használhat szolgáltatásokat, valamint adatbázist).

⁷⁴ Solid State Drive – a pendriveoknál is használt tartós memória-technológia

Végül fókuszáljunk a felhők és az MI kapcsolatára. Itt ki kell emelni, hogy „*az MI a felhőben érzi jól magát*”, pontosabban mondva sokkal hatékonyabb, ha felhőben fut. Viszont sok esetben nem jöhet szóba, hogy a hálózati kapcsolat nélkül, offline módban ne lehessen használni egy robotot, drónt, vagy fontos számítógépes funkciót. Ennek problémának a feloldására gyakran az a megoldás, hogy az ilyen rendszerek alapfunkciói működnek a felhővel való kapcsolat nélkül is. Vannak tehát teljesen offline módon is működő kis MI megoldások (pl. mobiltelefon arcfelismerése), csak felhőben elérhető szolgáltatások (pl. a mobiltelefon a beszédet leírja), de vannak „fél-offline MI rendszerek” is. Ez utóbbiak tanítási fázisát úgy teszik hatékonyabbá, hogy ezalatt a tanulást végző gépegyeden túl felhőben futó MI is megtudja azt, amit az egyed. Így ún. párhuzamos tanítással, egyszerre tökéletesíthető minden funkció, tehát sokkal gyorsabban fejlődik a rendszer. Más szóval a felhőre kapcsolt gépek „közösén tanulnak”: amit egyik gép megtanul, azt feltölti a felhőbe, ahonnan minden résztvevő gépekbe letöltődik, így a többieknek nem kell mindenre maguknak „rájönniük” a jobb működésre.[63] Erre példa a Tesla önvezetés 2.0 verziója, ahol sok autó, hosszú idő alatt, sok utat bejárva tanította meg a technológiát, hogy képes legyen a szenzoradatok, a térképi információk és a statisztikák eredőjében elég keveset hibázni. A kereszteződések, táblák, lámpák felismerési nehézségeit a többiek által megtanultak alapján tudja minden autó leküzdeni navigálni.[64] Ez a módszer a gyors fejlődés egyik oka (bár egy ideig még nem szabad teljesen önvezetésre hagyatkozni).

II.1.3. IoT (Internet of Things) szenzorok és robotok

A „dolgok internete” gyűjtőfogalom lényege, hogy az ebbe tartozó „dolgok” (eszköz, szoftverek) emberi beavatkozás nélkül a hálózati adatforgalom aktív befogadói vagy előállítói. Véleményem szerint pontosabb lenne a Network of Things (dolgok hálózata) kifejezés,⁷⁵ mivel egyre kevésbé szokás ilyen elektronikai „dolgokat” a nyílt interneten működtetni, biztonságosabb egy intranet. Az IoT általam is használt hivatalos meghatározása⁷⁶ egyébként az internetet nem erőlteti a definícióba, viszont az MI-re többféleképpen utal, „okos infrastruktúrának” is nevezi.

Kézenfekvő az ilyen dolgok felosztása a szerint, hogy adnak vagy kapnak adatokat a hálózat többi részétől. Az input IoT eszközök információt (inputokat) adnak vagy egy központi rendszernek vagy egymásnak. Például egy időjárás-szenzor amellet, hogy adatokat küld egy

⁷⁵ Talán a rövidítés miatt nem ez rögzült, hiszen a NoT angolban nem hangzik jól egy technológiára.

⁷⁶ „egymással összekapcsolt eszközök és szolgáltatások adatokat gyűjtenek, cserélnek és dolgoznak fel annak érdekében, hogy dinamikusan alkalmazkodjanak a környezethez. Az IoT szorosan kötődik a kiberfizikai rendszerekhez, és e tekintetben az intelligens infrastruktúrák elősegítője a szolgáltatási minőségük javításával.” [65]

előrejelző rendszer számára, közvetlenül is aktiválhatja a közelében lévő optikai eszköz páratlanító fűtését is. Az output eszközök a hálózaton kapott parancs alapján végeznek el valamit. A csak input képességgel ellátott eszközök elterjedtek (szenzorok), míg a kizárólag output eszközök ritkábbak, egy ilyen funkció mellé minimum állapot- és hibajelző szenzorokat is be szoktak építeni. Például egy okos közlekedési lámpa a forgalmat érzékelve, a központi rendszer által végzett optimalizációval képes hatékonyan szabályozni az adott kereszteződést, ha pedig közlekedő mentőautót érzékel, közvetlenül kapcsolja számára a lámpákat a biztonságosra.

Igazából nem kifejezetten egy diszruptív technológiáról van szó, hanem a már régen létező eszközök[66] automatizált hálózatba kötésének koncepciója futott fel, a többi itt tárgyalt diszruptív technológia farvizén. Az ilyen eszközök adatainak feldolgozására a felhőszerverek ideálisak, sok adatuk nem strukturált (pl. videókép) vagy „csupán” nagy mennyiségű így a BigData részévé válik. Továbbá csak MI-képességek segítségével lehet hatékonyan feldolgozni, és abból a szükséges információt kinyerni (pl. egy körözött személy mozgását a térfigyelő képek alapján). Az IoT szoros kapcsolata MI-vel ennél azonban többrétű és nem felhőhöz kötött. Használható még a fentebb bemutatott módon robotok tanítására, másfelől egyszerűbb, offline MI-képességek (pl. képfelismerés, okos-korrekciók) biztonságosabban és hatékonyabban vezérelhetnek robotikus eszközt, mint egy hagyományos kód. Katonai felhasználásainak egy része kézenfekvő, a logisztikai támogatástól a harctéri korai jelző szenzorokig, erre külön kifejezések alakultak ki (IoMT, IoBT)⁷⁷, de rendészeti és katasztrófavédelmi felhasználása is terjed.

II.1.4. Az új technológiákon új ökoszisztéma: az Ipar 4.0

Szót kell még ejteni a kapcsolódó technológiákon túl arra az új gazdasági ökoszisztémáról, melyhez az imént bemutatott három technológia és az MI használata együtt szolgáltatott alapot. Az EU szakértők által megfogalmazott rövid meghatározás inkább az IoT-t emeli ki: *„Az ipar 4.0 a termelési folyamatok olyan szervezését írja le, melynek keretében az eszközök önállóan kommunikálnak egymással az értéklánc mentén.”*[67] A termelési folyamatok optimalizálása korábban is fontos tényező volt. Azonban gazdasági fejlődés során számos olyan érdekcsoport alakult ki, mely ráhatással bírt az ellátási láncok egyre nagyobb részeire, sokszor teljes egészére. Így a részfolyamatok hatékonyságának növelésén túl igény lett a nyersanyag előállításától a vőkgig tartó hosszú ellátási lánc egészét egyben áttekintő optimalizálásra. Ez az igény motiválta a technológiák fejlesztéseit is, hiszen multinacionális gazdasági komplexumok tudják az MI és a BigData által nyújtott előnyöket igazán meglovagolni.

⁷⁷ IoMT: Internet of Military Things, IoBT: Internet of Battlefield Things.

Ennél az általánosabb meghatározásnál a kifejezést gyakran szűkebb értelemben használják, és csak a gyártási aspektust értik alatta. Ez a szűkebb értelemben használt Ipar 4.0 „...a jövő egy olyan „okos” gyárát hozva létre ezzel, amelyben a számítógép-vezérelt rendszerek nyomon követik a fizikai folyamatokat, létrehozzák a fizikai valóság virtuális mását és decentralizált döntéseket hoznak önszervező mechanizmusok alapján.”[67] Ennek szűkebb megközelítésnek előnye, hogy jobban vizsgálhatóbbá tesz számos kihívást és biztonsági problémát, amelyek az Ipar 4.0 terjedéséhez kapcsolódnak.[68] Most egy példa is elegendő szemléltetésére: egyik legnagyobb problémát a régebbi technológiák bekapcsolása jelenti az Ipar 4-es hálózatba.⁷⁸ Számos kérdést elhagyva lépek most tovább, erről itt elegendő ez az említés szintű pár sor.

II.1.5. Az MIKT fogalom bevezetése

Az MI szoros kapcsolata az imént ismertetett rendszerekkel, véleményem szerint indokoltá tesz egy új kifejezést, ezért erre a technológia halmazra saját kifejezést alkottam. ez lett a Mesterséges Intelligencia és Hozzá Kapcsolódó Technológiák köre, azaz MIKT. Angol nyelvű alkalmazására az AIRT rövidítést javaslom, mely az Artificial Intelligence and Related Technologies kifejezést takarja. Tömör meghatározás a következő:

AZ MIKT az MI közös halmaza azokkal a technológiákkal, melyek az MI fejlődésének robbanását segítették, és azóta is az MI-vel kölcsönhatásban fejlődnek. Ennek a technológia-halmaznak legfontosabb elemei a BigData, az IoT és a Felhőtechnológia tágabb (a hardver-fejlettséget is a felhőhöz soroló⁷⁹) értelmezése.

Pontosításként ismertetem a fogalom összevetését hasonló kifejezésekkel, és rögzítem mit nem értek bele.

- Az MIKT-hez hasonló fogalom például az EU AI-Act-ban az „MI-rendszerek” kifejezés, valamint az, hogy ott technológia-családként utalnak az MI-re. A szöveg alapján egy okosfűtés vagy egy mobiltelefon készülék tehát egy kalap alá kerül egy okosvárossal, hiszen mindegyik MI-rendszernek tekinthető.
- Felvethető, hogy az Ipar 4.0 modell is hasonlót takar, ám jobban megvizsgálva ez a modell egy jóval tágabb, és nem is technológiai kategória. Az Ipar 4.0 ráépül ugyan az MIKT adta optimalizációs lehetőségekre, viszont kifejezetten gazdaság-centrikus

⁷⁸ Például egy még jól működő gyártósort csak azért lecserélni mert Windows XP gépen fut a vezérlésre túl drága, valószínűleg újraprogramoztatni sem éri meg – viszont a nem támogatott alaprendszer az egész hálózatot veszélyeztetheti.

⁷⁹ A tágabb értelmezés nem szolgáltatásként tekint a felhőre, hanem abban beleérti a hozzá kapcsolódó hardver és szoftver technológiákat is, melyek a megfelelő működés érdekében koruk csúcsát képezik (ultragyors alaplapok, legújabb processzorok, óriás sávszélességű kapcsolatok, gyors és óriási tárolók stb). ld. II.1.2.

megközelítés, melynek lényege, hogy profitorientáltan koncentráljon a hatékonyságra. Ez azonban nem érvényes az MIKT állami vagy védelmi alkalmazásaira, ahol nem feltétlenül szempont a hatékonyság. Sőt ezeknél indokolhatják a „ráfizetéses” használatot olyan tényezők, mint az emberközpontúság vagy nemzeti érdekvédelem és hasonló. Továbbá az Ipar 4.0 a hatékonyság érdekében az MIKT mellett sok egyéb technológiát is felhasznál, a Blockchain módszertől a 3D nyomtatásmódokig.

- A fogalomba nem értem bele azokat a technológiákat, melyekre nem kifejezetten támaszkodik az MIKT. Például kérdésként merülhet fel, hogy a robotika beletartozik-e? Erre azt lehet mondani, hogy annyiban része egy robot amennyiben kapcsolódik hozzá. Tehát ha egy robot (vagy más kisebb rendszer) egy MIKT számára tanulási adatokat ad, illetve a kapott adatok alapján működik, akkor része (pl. egy önvezető okosautó). Ám egy régi offline robot, pl. egy régebbi kézi vezérlésű aknamentesítő robot nem része (bár némi átalakítással részévé tehető).
- Nem tartoznak az MIKT-ba a tudomány olyan irányai sem, melyek NEM elengedhetetlen részei egyetlen jelenleg működő MIKT rendszernek sem, és nem is lehet arra számítani, hogy néhány éven belül azok lesznek. Ebbe az az MIKT-n kívüli körbe sorolom pl. az ember-gép integráció vagy a kvantum-MI elképzelések egyelőre kezdetleges (még messze nem piacképes) fejlesztéseit.

A fogalom más irányú továbbgondolását az okosdolgokkal kapcsolatban folytatom (V.1.3.). Remélhetőleg az itt javasolt fogalmi pontosítások segítenek jobban megérteni a minket körbevevő technikai világot. Az bizonyos, hogy jelen tanulmány számára segítenek az MI fogalom pontosításában, tudva, hogy ilyen fogalmak további pontosítási szükségessége csupán idő kérdése. Az sem kizárt, hogy olyan irányba kell bővíteni (pl. energia-termelés vagy tárolás), amelyet most még nem érdemes beleérteni. De a hatékonyabb kommunikáció érdekében is hasznos lenne nem 8-10 év késéssel alkotni fogalmakat, hanem mielőbb: már akkor, amikor a fejlődés során egy fogalmi-zavar lehetősége felmerül.

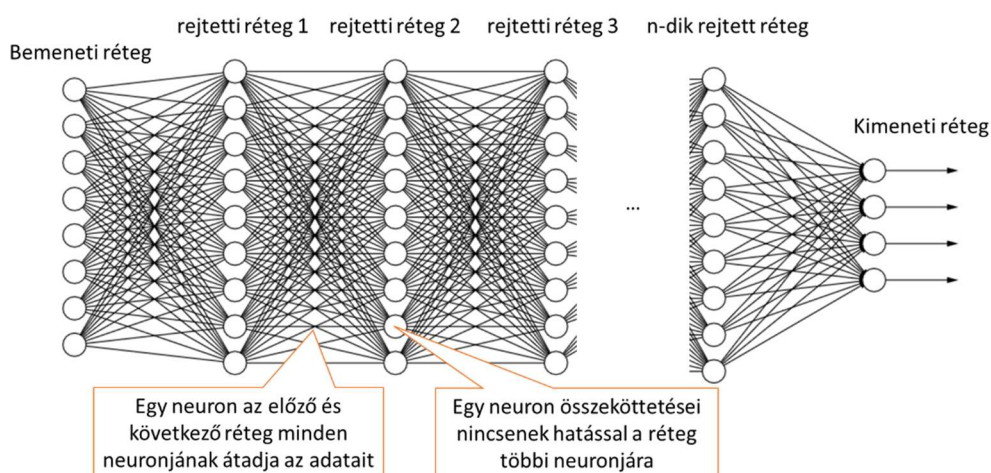
II.2. GÉPI TANULÁSI MODELLEK

Az alfejezet célja, hogy a gépi tanulás fontosabb megvalósítási módjairól képet adjon. Ezt megkísérlem röviden, de a matematika iránt kevésbé érdeklődők számára is érthető formában megtenni.

II.2.1. Az ML modellek közös jellemzői

A gépi tanulás alapját az ún. „mesterséges neurális hálózat” (artificial neural network, ANN) jelenti, ami nevét az emberi agyban működő idegsejtek hálózata alapján kapta, hiszen ezt próbálták mesterségesen utánozni. Egy op-art művészi alkotás is lehetne, sőt akár az MI egyik szimbólumává is válhat az alapmodell vázlata, mely a 2. sz. ábrán⁸⁰ látható.

Mivel ez az alapmodell elég közismert, remélhetőleg elegendő pár bekezdés a bemutatásához. Az ábrán is látható, hogy a neuronok rétegekbe vannak szervezve. A bemenetek csak az első rétegnek adnak át információt, majd ez az első réteg a csak a másodiknak, és így tovább, míg a kimenet csak az utolsó (n-dik) rétegtől kap információt. A rétegen belül nincs kapcsolat, azonban egy neuron a következő réteg összes neuronjának átadja az információját. Nem kell több belső réteg, sőt, időrendben először az egyetlen belső réteggel rendelkező változat volt működőképes Perceptron⁸¹ néven, az 1950-es években. Manapság is használatos ez az egyréteges változat egyszerűbb feladatokra, így gyakorlatilag a DL olcsóbb változataként is felfogható.



2. ábra: ANN mélytanuló alapmodell (saját átszerkesztés)

Nem csak a rétegek és neuronok száma, de neuronok kapcsolata is eltérő lehet, a következő alfejezetben a fenti a modellnek több ilyen variánsát bemutatom. De minden MI modell lényege, hogy az azt alkotó neuronok kimenete időben változik a tanulás során. Ezt a változást a neuronokat ért adatok idézik elő. További közös vonás minden modellben, hogy a neuronokban súlyozottan összesítődik⁸² minden feléjük küldött számadat. Amint ezt a 3. sz. ábra mutatja, minden bemenethez is rendelhető külön súly, amellyel az adatot megszorozzák. Látható, hogy az adott neuronhoz magához is hozzárendelnek egy saját torzítási (*bias*) értéket, ezt is

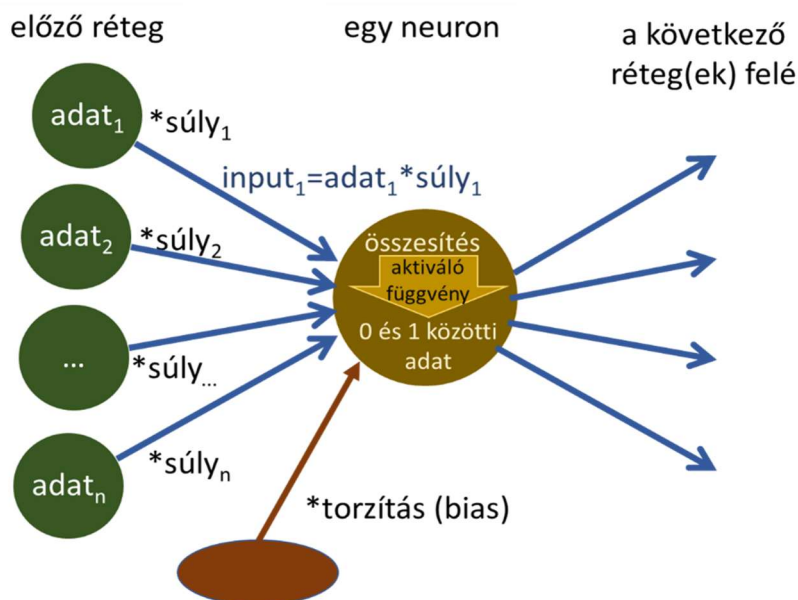
⁸⁰ A szerkesztett ábrához a kiindulás: [69]

⁸¹ Ez használható volt pl. osztályozási problémák megoldására. [70]

⁸² Azért nem összegzést írok, mivel nem összegfüggvényt, sőt is csak súlyozott átlag függvényt alkalmaznak.

figyelembe veszi az a (modellfüggő) összesítő függvény, mellyel a neuron dolgozik. Végül az imént kapott számot 0 és 1 közé eső értéké alakítja át,⁸³ így áll elő tehát egy neuron tartalma. Ez lesz a következő réteg neuronjainak elküldve. A súlyok és torzítási értékek arra valók, hogy általuk hangolni lehessen a modelleket, erre sokszor utalni fogok különböző összefüggésekben. Az imént leírt súlyozások általi hangolás is közös vonása az összes modellnek

A fentiek alapján világos, hogy a neuronok tartalma egy 0 és 1 közötti számadat, mely matematikai függvények és konverziók által jön létre. Ebből adódik a modell fekete-doboz jellege,⁸⁴ vagyis, hogy a belső rétegek tartalmát hiába is kérdeznénk le. Sem a neuronok konkrét tartalma, sem a rétegeké nem ad értelmezhető információt úgy, mint egy adatbáziscella vagy adattábla. A neuronok csupán kusza értékeket tartalmaznak – a rendszer csak egészében hasz-



3. ábra: Egy neuronhoz tartozó súlyozások és torzítás figyelembevétele (saját készítés)

nálható. Hasonlító ez ahhoz, amikor jegyzetombbel nézünk bele egy .exe fájlba: ebből nem tudunk következtetni a kódra, hiába fut le jól a program. Ezért nem is kérdezzük le a neuronok értékét. Számunkra a belső rétegek „rejtettek”, így nevezi ezt a jelenséget a szakirodalom.

⁸³ egy nem lineáris, általában szigmoid függvénnyel kerül átalakításra.

⁸⁴ Léteznek kutatások a fekete-doboz jelleg nélküli modellekre és tanítási formákra is, mivel ez a jellemző néha nem fogadható el. Pl. [71]

II.2.2. Más cél más ML

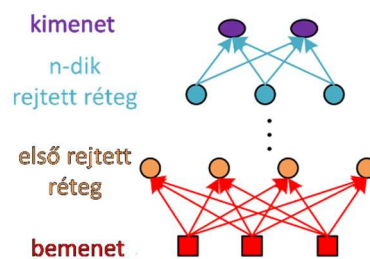
Túlzottan nem érdemes belebonyolódni, de néhány eltérő modell vázolója elkerülhetetlen.

Térjünk vissza a 2. ábrán bemutatott mély neuronháló (DNN, *deep neural network*) modellre, mely egyszerűsítve van a 4. ábrán⁸⁵ annak érdekében, hogy a következő két modellel könnyebben összevethető legyen. Látható is rajta, hogy ezen

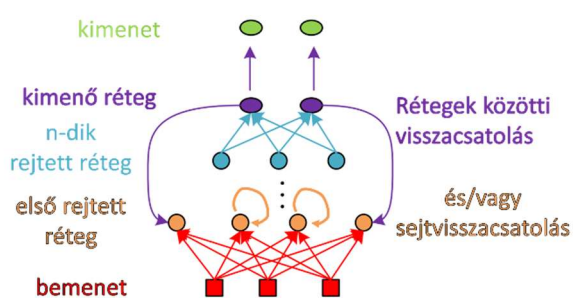
alapkonstruáció szerint modellekben az információ csak előre halad (előre-csatolásos modellek). Az ilyen modellek tudása a bejövő információ folyamatos módosulása miatt állandóan változik. Ez a változandóság gyakran hátrányos, mert a modell tudása az újabb adatok által könnyen romolhat, hiszen elfelejti, amit már tudott.

A visszacsatolós⁸⁶ neuronháló (RNN, *Recurrent Neural Network*) ezen javít. Itt ugyanis egy réteg kimenete nem csupán „előre”, a következő réteg felé küld információt, hanem visszafelé, az előző réteg felé is, amint ezt az 5. ábra lila nyila is mutatja. Sőt, egyes megvalósításaiban a neuronok kimenete saját magára is

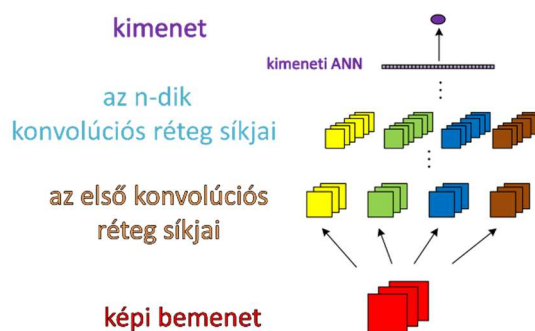
visszahat. Az ilyen visszacsatolások által az adatok belső összefüggései jobban megmaradnak a tanítás során, így egyfajta saját memóriaképességet adhatunk a rendszernek. Ilyen elven működik pl. a hosszúvektorú rövidtávúmemória rendszer, ami az emberi rövidtávú memóriához hasonló dolgot hivatott megvalósítani. Hivatalos neve valójában inkább szójáték: „hosszú rövidtávú memória (hálózat)” (*Long Short Term Memory (Network)*, LSTM). Itt a „long” a tárolt adat mennyiségére (az adatvektor hosszára), a „short” a tárolási időtartamra utal, pontosabb lenne azonban angolul a Long-vector Short-Term Memory kifejezés.



4. ábra: Az ANN modell (forrás a lábjegyzetben)



5. ábra: Az RNN modell (forrás a láb-



6. ábra: a CNN modell (forrás a láb-

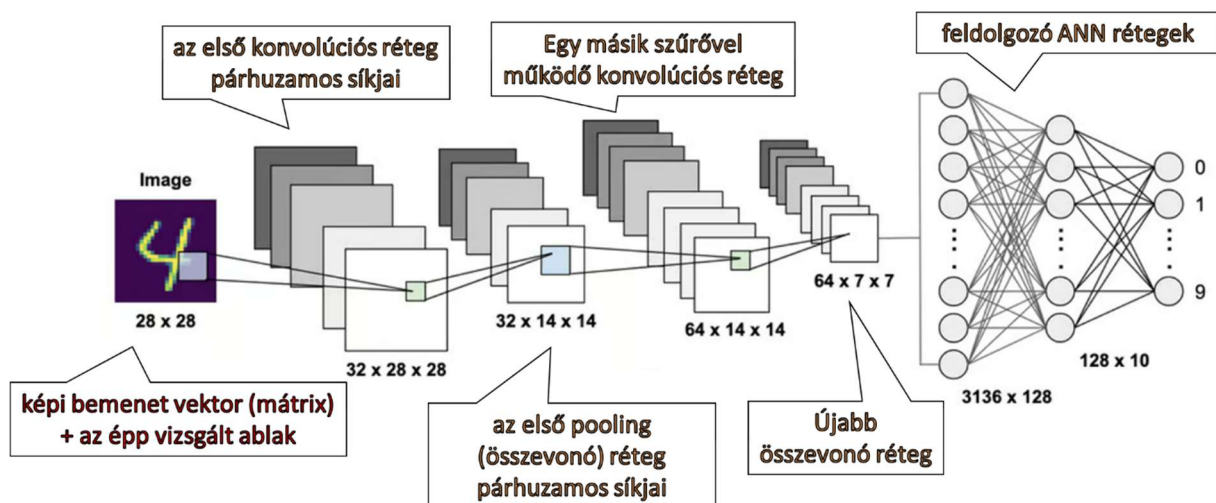
⁸⁵ A 4-5-6 ábrákat a következő forrás alapján alakítottam át és magyarítottam: [72, o. 10]

⁸⁶ néhol „ismétléses neuronháló”, mely véleményem szerint nem egy találó kifejezés (hanem szó szerinti fordítás)

Az LSTM továbbfejlesztései képesek irányított módon megőrizni, illetve elfelejteni az információkat, ehhez memóriacellákat és/vagy kapukat használnak. Így ezek még hosszabb szekvenciák értelmezésére válnak alkalmassá. Az ilyen és „kapuzott ismétlődő egységek” (GRU, *Gated Recurrent Units*)⁸⁷ illetve az LSTM megoldások főleg a természetes nyelvi feldolgozásnak segítettek sokat. De tudásuk számos egyéb területen hasznosítható, a mozdulat-felismeréstől kezdve az orvosi diagnózis előrejelzéseken át az adott stílusban való zeneszerzésig.

Eltérő kihívások felé azonban eltérő technikákkal érdemes az fordulni, épp úgy ahogyan bennünk, a mi emberi megismerésünkben sem egyformán működik a különböző érzékszerveinktől kapott információk feldolgozása. Az előbb említett irányok helyett a képfeldolgozást a mélytanulás teljesen más fajta fejlesztésével célszerű megközelíteni. Itt ugyanis nem szekvenciálisan kapott információt kell feldolgozni, mint a szöveg esetében, hanem egy mátrix-jellegű bemenet kiértékelésére kell hatékony megoldást találni. A számítástechnikában köztudott, hogy ilyen feladat megvalósításakor érdemes minél több dolgot párhuzamosítani. Az erre tervezett rendszerek ezért már az input-rétegben is párhuzamosan síkokon tanulnak, a belső rétegekben pedig, az ANN sejtjei helyén is párhuzamosan működő síkokban dolgozzák fel a képet (vagy más mátrixot). A 6. ábra modellje egymás mögötti téglalapokkal ábrázolja a párhuzamos feldolgozás síkjait, ami a tehát legfőbb különbség az alap ANN modellhez képest. A párhuzamoság azonban kevés a módszer megértéséhez, bele kell pillantanunk a működésbe. Ha ez itt nem sikerült közérthető nyelvezetű forrásokat is adok meg.

A képet egy konvolúció nevű átalakítással tesszük a gép számára értelmezhetővé, ezért nevezik ezt a modellt konvolúciós neuronhálónak (*Convolution Neural Network, CNN*). Konvolúció matematikai művelete két függvény alapján egy harmadik függvényt állít elő. Ez az



7. ábra: A CNN modell alaposabb bemutatása (forrás a lábjegyzetben)

⁸⁷ Az ilyen rendszerekről alaposabb képet ad: [73]

eredményfüggvény azt fejezi ki, hogy hogyan módosítja az egyik függvény alakját a másik függvény. Esetünkben a képmátrix képrészleteinek sorozatát egy un. szűrőfüggvénnyel veti össze a konvolúció. A 7. ábrán[74] látható a képrészletet mutató „ablak”, melyet kis lépésekkel, akár pixelenként arrébbtolva, végigmozgatnak a kép minden részén.⁸⁸ Amikor az ablak picit arrébb mozdul, akkor bár új sor(ok) vagy oszlop(ok) kerül(nek) alá, a letakart képpontok nagy része ugyanaz marad, mint az előző lépésben. Ezeken a képpontokon tehát többször lefut a konvolúció, míg az ablak fölöttük van. Ez által meg tudnak jelenni a létrejövő eredménymátrix súlyozásaiban a képpontok összefüggései, hiszen a konvolúcióból létrejött eredményfüggvények egymáshoz képesti módosulásait tárulnak így fel. Tehát a konvolúciós eredménymátrixon alapul majd a gépi tanulás, és ez által lesznek képesek a későbbi rétegek a pixelhalmazban felismerhető mintázatokat találni.[76]

Am a konvolúciós képfeldolgozás csak több szűrőfüggvény lefuttatásával ad megfelelő eredményt, ezek különböző rétegekben működnek. Ezek között a rétegek között azonban érdemes volt segédretegeket is létrehozni a folyamat gyorsítása érdekében. Ezek a köztes, összesítő un. pooling rétegek csökkentik a mátrix méretét (erre egyszerű maximumkeresést és átlagolást használnak), de közben a domináns jellemzőket megőrzik. Végül a kép mintázatainak osztályozását, vagyis a tanult adatok használhatóvá tételét egy hagyományos (általában ANN) neuronháló végzi, ez állítja elő a modell kimenetét.[77, o. 4]

Ez a modell is tovább kombinálható. Például visszacsatolásossal bővítve az előbb említett RNN emlékezés előnyeire tehetünk szert. Az ilyen konvolúciós visszacsatolt neuronháló (*Convolutional Recurrent Neural Networks*, CRNN) a kép- és videóelemzési feladatok fejlődéséhez[78] döntő fontosságúak alapot nyújt, de ennél jóval messzebbre mutat. Fontos és védelmi szempontból is érdekes felhasználása lehet többek között pl. a gépi szájról olvasás is.[79]

II.2.3. Az utóbbi évek újításai

Felhasználói társadalmunkban egyre inkább népbetegség a figyelemzavar, a gépek koncentrációs képessége viszont egyre jobb. E téren a nagy áttörést az un. figyelemmechanizmusok (*Attention Mechanism*) beépítése jelentette, mely elsősorban szövegfeldolgozó modelleknél vált be. Egy figyelemmechanizmussal ellátott LSTM épp úgy képes rákoncentrálni a túl komplex valóság érdekes részére, ahogyan egy ember is képes figyelmen kívül hagyni a zavaró tényezőket, amikor feladatára koncentrálni.[80] Így egy adott információhalmazt több szempontból lehet vele elemeztetni, különböző kérdések mentén. Korábban egy LSTM-RNN modell

⁸⁸ Ezért csúsztható ablakok technikájaként is utalnak erre a módszerre. [75]

könnyen belezavarodott például egy hosszabb mondat helyes fordításába, mivel a gép a közeli hasonlóságokat és az ismétlődéseket nehezen kezelte. A megoldást elég új,[81] és a módszer természetesen itt is a súlyozásokhoz kapcsolódik, viszont ezt úgy oldották meg, hogy nem csak kulcsszavakhoz rendeltek magasabb súlyt, hanem szókapcsolatokhoz, vagy más szövegegységekhez is.⁸⁹ Így betáplálva egy ilyen gépbe jelen tanulmány fentebbi részeit, elvileg képes választ adni olyan kérdésekre, mint például „mit ír a szerző az informatika szó helyességéről és változásáról?”

Ennek az áttörést jelentő, úgynevezett skálázott pontszerű eredményfigyelés (*Scaled Dot Product Attention*) nevű figyelem-mechanizmusnak ismertebbé vált egyik fejlesztése, mint ő maga. Ez a Transformers rendszer, melynek névadója ugyan a Google volt, akinek a fejlesztői 2017-ben publikálták ezt a modellt, viszont a modell neve egy Microsoft által vált közismertté, hiszen erre utal a GPT rövidítés a közismert Chat-GPT (ld. II.3.4.) szolgáltatásban (*Generative Pre-train Transformer*). A Transformers modell azt oldotta meg, hogy sokszor nem ad megfelelő eredményt, ha a figyelem-súlyokat csupán átlagoljuk. Az új modell ennek elkerülésére többféle figyelem-összesítés („több fej”) párhuzamos használatát vezette be,[82] ebből adódik, az érdekes „többfejű figyelem” elnevezés.[83] A modell alapvető újdonsága, hogy kizárólag a figyelemmechanizmusokra alapul, azokat módosítja, így teljesen mellőzi a konvolúciókat és ismétlődéseket. Így egy gyorsabban betanítható és jobban párhuzamosítható modellt kaptak, melynek több feje több dologra is hatékonyan képes figyelni.⁹⁰ Az eredményt tovább javította, amikor a többfejű figyelem mellett a modell önfigyelemre is képessé tették.[85] A szövegfeldolgozás terén ezek által egyre több jelentésréteg egy finomabb megkülönböztetése vált lehetővé.

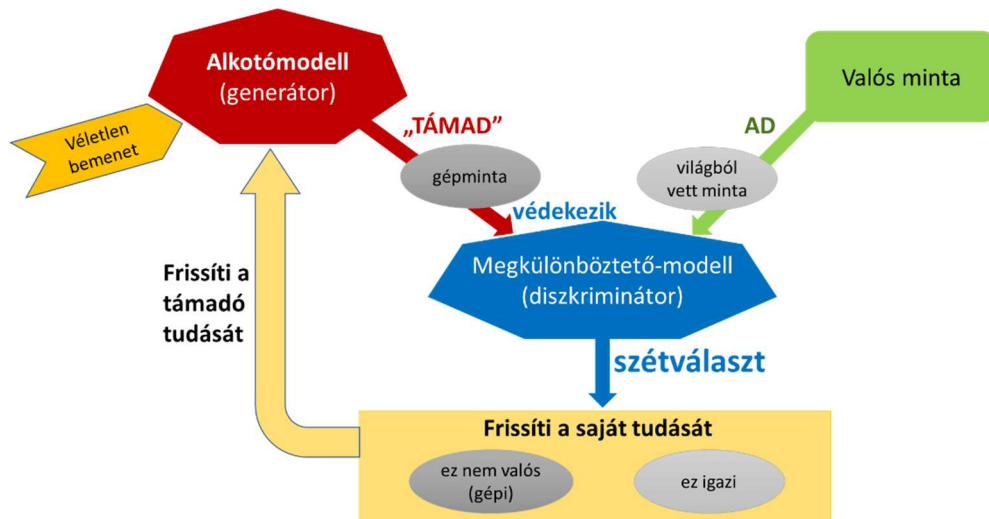
És itt érkeztünk el az utóbbi évek másik nagy áttöréshez, amely bizonyos kreativitást képes adni a gépnek. Ezek az Alkotó Versengő Hálózatok (GAN, *Generative Adversarial Network*),⁹¹ ismertebb nevükön Generatív MI-k, képesek a tanult minták alapján hasonló, de új adatokat generálni, ezáltal keltik a kreativitás látszatát. Lenyűgözte a világot, amikor ezek valakinek a stílusában megadott tartalmakkal egy alkotnak egy szép képet, vagy például hexameterben, vagy adott rímképlettel írnak verset. A névben az *adversarial* szó arra utal, hogy egy versengő

⁸⁹ Picit pontosabban: egy ún. „kontextusvektor” készül minden bemeneti szóra, tehát az adott szó és a többi közötti „távolság” adja meg a súly mértékét. A technológia megértéséhez egy név nélküli forrást kell javasolnom: <https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/> (Letöltve: 2024. 05. 03.)

⁹⁰ Ashish Vaswani és mtsai., „Attention Is All You Need”, *31st Conference on Neural Information Processing Systems, Long Beach*, 2023. augusztus 1., <https://doi.org/10.48550/arXiv.1706.03762>

⁹¹ Az angol kifejezés szó szerinti „ellenséges” fordítását használva (mint több helyen) itt is kissé megtévesztő kifejezést kapnánk.

modellről van szó, ami azt jelenti, hogy a GAN-ban két MI „ellenségként” viselkedik egymással. Az alkotómodell (generátor) a támadó, ez próbál olyan mintákat létrehozni, mellyel megteveszti ellenfelét, amint azt az 8. sz. ábra mutatja. (Az ábrát ihlette: [86])



8. ábra: Egy generatív MI modell, és annak „ellenséges” elemei (saját készítés)

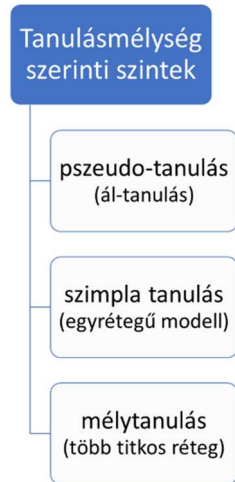
A megkülönböztető-modell (diszkriminátor) védekezik, vagyis igyekszik megkülönböztetni a generált, gépi adatokat az igazi, világból kapott adatoktól. A diszkriminátor eredményei visszacsatolódnak és mind önmaga, mind a generatív rendszer tanul az eredményekből. Itt már nem egy, hanem két modellt szükséges finomhangolni, valamint rendszer tanításához egyaránt szükségesek a valós és a hibás adatok (természetesen elkülönítve). Ezek rámutatnak arra a tényre, hogy az egyre komplexebb rendszerek tanítása is bonyolultabb. Ez a technológia védelmi szempontból is igen fontos, hiszen mára elérte, hogy már nemigen lehet eldönteni egy illusztrációs grafikáról, hogy azt gépi vagy egy emberi rajzoló alkotta-e, így jól bevethető a pszichológiai és az információs műveletekben. Ám a jövőre nézve nagyobb jelentőségű, hogy hasonló versenyzető módszer lehet alkalmas például kibervédelmi MI kiképzésére is.

Végül itt is említést kell tenni a raj-tanulás sajátos módjáról. Ez a modell, és ennek fejlődése jelentősen, és sok szempontból alapvetően eltér az imént vázolt modellektől, emiatt bővebb ismertetését egy másik alfejezetben (II.3.3.) teszem majd meg.

II.2.4. Az ML népszerű felosztásai és a pseudo-tanulás

Az intelligencia és az autonómia szintje függ a tanulás képességétől, vagyis a megvalósított ML modelltől, emiatt szükséges a gépi tanulás főbb felosztásaira rátekinteni. Ezeket könnyebb áttekinteni ábrázolva, ezért készült a 9. és 10. ábra.

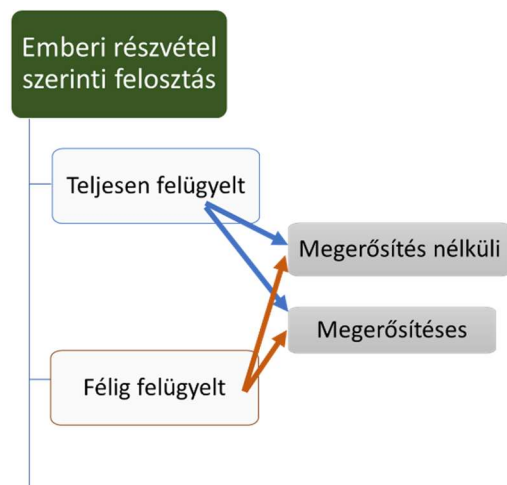
1) Egyik legelterjedtebb megközelítésben a „mélytanulás” (*deep learning*, DL) kifejezéssel különböztetnek meg fejlettebb modelleket az egyszerű gépi tanulástól. ahol a mélyebb mélyebben (lejjebb) látható, a magyarázatnál azonban fordított sorrend a logikus. Mint láthatuk, a mai a mélytanuló modellek egymástól jelentősebben térnek el annál, hogy egy kategóriába kerüljenek, ezért mára ez a fajta kettéosztás talán kissé meghaladott (de azért tartható osztályozási mód). Viszont javaslom kiegészíteni egy még egyszerűbb szinttel, úgy használhatóbb. A szinteket tehát ennél a felosztásnál az jellemzi, hogy hány neuronréteget használ a be- és kimenet között:



9. ábra: A gépi tanulás architektúra szerinti felosztása

- a DL több rejtett réteget használ;
- a szimpla ML csak egy belső réteget használ;
- egy réteget sem használ. Ezt hívom pszeudo-tanulási szintnek, de nevezhetnénk nulladik szintnek is. Valójában nem is tanulás, hanem olyan hagyományosan programozott rendszer, mely a tanulás benyomását keltheti. Ezt fontossága miatt alaposabban, a másik két főfogalommal összegzésben fejtem ki (ld. III.1.4.), itt a lista teljessége miatt említettem.

2) Egy másik szokásos felosztás azt mondja meg, hogy mennyiben vesz részt az ember a tanulásban. Ez alapján a szintek: (a) teljesen felügyelt, (b) félig felügyelt, (c) felügyelet nélküli tanulás. Az elnevezésekből világos, hogy az ember részvétele a tanulási folyamat irányításának mértékét jelenti. Az a) és b) esetben felmerül az irányítás módja. E tekintetben a jutalom-büntetés módszer alkalmazását szokás külön felosztásnak venni (tehát, hogy azt alkalmazzák vagy nem).⁹²



10. ábra A gépi tanulás tanítás szerinti felosztása (saját készítés)

Ezeket a fogalmakat, ha megpróbáljuk röviden megragadni, a következők mondhatók [88]:

- **Felügyelt tanulás:** A pontos eredmények eléréséhez címkézett adatokra van szükség. Az eredmények javítása érdekében gyakran több adat megismerésére és időszakos módosításokra van szükség.

⁹² Számos gépi tanítási technikát osztályoz cikkének 2. számú ábráján: [87]

- **Félig felügyelt tanulás:** ez a felügyelt és a felügyelet nélküli tanulás vegyes alkalmazása. Részlegesen címkézett adatokon tud eredményeket adni, és nem igényel folyamatos módosításokat a pontos eredmények eléréséhez.
- **Felügyelet nélküli tanulás:** emberi beavatkozás nélkül fedez fel mintákat és betekintést az adatkészletekben, és pontos eredményeket ad. A klaszterezés a felügyelet nélküli tanulás leggyakoribb alkalmazása.
- **Megerősítő tanulás:** A megerősítő tanulási modell folyamatos visszacsatolást vagy megerősítést igényel, amint új információk érkeznek, hogy pontos eredményeket adjon. „Jutalmazási funkciót” is használ, amely lehetővé teszi az öntanulást a kívánt eredmények jutalmazásával és a rossz eredmények megbüntetésével

Megnyugtatóan megállapítható, hogy a biztonság érdekében az öntanítást csupán célfeladatokra használják, például emberszerű robotok mozgáskoordinációjánál. Az „üres lap” alapozásról induló MI komoly problémákat vet fel, ennek kritikáját később ismertetem. (IV..2.4.) Az ember géptanítói lehetőségeit egyébként is alapvetően határozza meg az alkalmazott technológia. Sok tanulási modell nem is alkalmas (biztonságos) öntanításra. A XX. században épp az lassította az MI fejlődését, hogy csak a megerősítéses módszer adhatott megbízható tudást a gépnek. Csak a közelmúltban, a korszerűbb, de egyben bonyolultabb rendszereknél nyílt meg az út a részleges öntanításra, de az filmek vagy novellák öntudatra ébredő rendszereitől a mai technológia még fényévekre van.

Remélhetőleg a fenti néhány fontosabb technológia vázolata is elegendő a gépi megismerés sokféleségének érzékeltetéséhez. A leírtakból kirajzolódott, hogy egy gép egyre több emberi kogníciós képességhez hasonló tulajdonsággal rendelkezhet. Az ember megismerésének, emlékezésének, figyelmének, alkotóképességének számos mozzanatát már ma képesek ilyen gépek utánozni. A nagy MI rendszerek számos alrendszerből tevődnek össze, vagyis az itt felsorolt technológiákat jobbra együtt alkalmazzák (sok további fejlesztést is beépítve). Ám még ezek a hihetetlenül komplex és költséges óriás-MI-k is messze vannak attól, hogy képesek legyenek szimulálni az emberi tudást, annak összetettségét.

II.3. ELVEK ÉS TECHNOLÓGIÁK AZ MI KÖRÜL

Miután az előző alfejezetben az MI-hez kívülről kapcsolódó fogalmakat tisztáztam, szükséges itt kitérni néhány technológiára (fontosabb elvre), melyek belülről kapcsolódnak hozzá. Ezek közül azokra szorítok, melyek a tanulmány téziseihez kapcsolódnak.

II.3.1. A jelen és a jövő MI-célú hardverei

Bár már az MI kezdeteinél is megjelentek hardveres implementációk, a technológia fejlődése és felfutása egyértelműen szoftveres síkon történt. A neurális háló bemutatásával (II.2.) az olvasó számára is világossá válhatott, hogy igen sok sejt van egy mélytanulási neuronhálóban, amelyeket párhuzamosan érdemes futtatni, de számítási igényük kicsi. Erre a hagyományos, egymagú CPU kifejezetten nem ideális, bár képes rá (megfelelően gyors processzor használható egyszerűbb MI szoftveres modellek futtatására). A fejlettebb CPU termékek sem nyújtanak megoldást, melyek egyre több magot tartalmaznak: azért nem optimálisnak, mivel komplex tudású és erős számítási igényre tervezett magjainak képességét nem tudja az MI kihasználni, viszont a csekély párhuzamosan futó művelet lassítja a rendszert. Ezért az MI felfutásával párhuzamosan a megfelelőbb hardveres alap keresését is számos fejlesztés célozta, ezekből a legfontosabbak a következők.

- **GPU** (*Graphical Processing Unit*).⁹³ A más célra tervezett, rendelkezésre álló hardverek közül a kép- és videófeldolgozáshoz már régóta használt hardver-megközelítés, a GPU alkalmazása szinte kézenfekvő volt. Ezekben már korán a sok (és egyre több) „kistudású” processzormag együttműködésén volt hangsúly, a 2025-ben megjelenő NVidia RTX 5090-nek már 21 760 darab magja lesz. A GPU magokat neuronként használó rendszerek jól teljesítenek, a nagy gyártók azonban mégis kínálnak több más megoldást is.
- **FPGA** (*Field-Programmable Gate Array*). *A tükör fordítás helyett hívhatjuk inkább a felhasználáskor programozható logikai-kapu mátrixnak., és még az 1980-as évekre nyúlik vissza (eredetileg semmi köze az MI-hez). Az egyszerűbb nevén szoftver-processzornak is nevezhető elv lényege, hogy a logikai blokkok programozhatósága révén az ilyen központi vezérlőegység sokkal jobban optimalizálható egy adott pontos célra, továbbá biztonságosabban kialakítható és igény szerint frissíthető, sőt tízszer kisebb energia-fogyasztású[89] az ilyen alapon megvalósított vezérlés. Ezért minden nagy fejlesztőnek vannak ilyen megközelítésű termékei, elsősorban a peremhálózati számítástechnika (edge computing) területén.[90] Elsődleges hátránya, hogy programozói szempontból nagyobb kihívás – ezért nem terjedhet gyorsan és széleskörűen ez a megközelítés.*
- **RDU** (*Reconfigurable Dataflow Unit*), azaz egy Újrakonfigurálható Adatfolyam-Egység technológiája még kevésbé ismert: itt a GPU-nál rugalmasabb alapokra próbálják helyezni

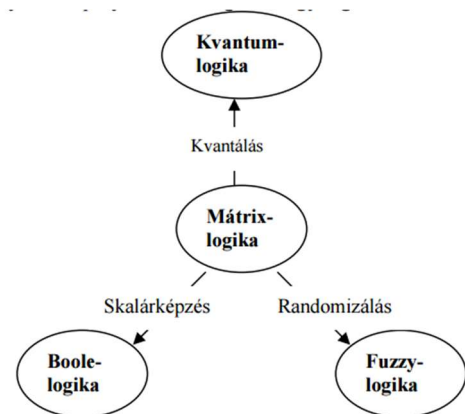
⁹³ 1999-től jelentek meg ilyen grafikai célprocesszorok. Elsősorban játékokhoz és film-animációs felhasználások elvárásai alapján fejlődtek, de a kriptovaluta bányászatból származó kereslet miatt is lökést kaptak.

a párhuzamosítást.[91] Vagyis a SambaNova cég ezen megközelítése egy MI-re jobban optimalizált GPU-nak is tekinthető.

- **Szoftverhez tervezett hardver:** egy további logikus architektúrális megközelítés, hogy egy adott MI keretrendszer vagy metódus számára optimalizálják a processzort. Erre példa a Loihi-2-es chipje (2021-től). Ezt kifejezetten a nyílt forráskórú LAVA nevű nyelvi keretrendszerrel való szoros együttműködésre tervezték, és a hivatalos tesztek szerint jóval hatékonyabbak hagyományos processzoroknál.[92] A háttérben az *eseményalapú neurális hálózatok* (*spiking neural network*, SNN) újdonsága áll, mely erre a platformra alapul. Ebben folyamatosan újratérképezik a neuronhálót, ezáltal tanulásuk sokkal jobban hasonlít a természetes tanulásra.
- **LPU** (*Language Processing Unit*). Ez az architektúra az előzőhöz hasonló célfeladatokat szolgál. Ezek az egységek szekvenciális feldolgozásra is optimalizálva vannak (adatok egymásutáni feldolgozására), mivel kifejezetten NLP feladatokhoz készülnek. A tervezési megközelítés előnye a fenti megoldásokkal szemben, hogy determinisztikusabban előre látható a teljesítménye a fordítóprogramok számára ideálisabb, mint a fenti párhuzamos megoldások. Az LPU elnevezést inkább a Groq nevű cég szolgáltatásaira használják, míg a Google Tensor hardverével kapcsolatosan inkább a TPU (*Tensor Processing Unit*) használatos, továbbá ilyen architektúrára utal a TSP (*Tensor Streaming Processor*) kifejezés is.
- **NPU** (*Neural Processing Unit*). Ez olyan processzort takar, melyet az agyi információfeldolgozási feladatok utánzására, azaz kifejezetten általános MI feladatokra terveznek. Ehhez jobb párhuzamosságot, a szokásosnál szélesebb sávú memóriáhozáférést, illetve az MI-sejtekben elegendő egyszerűsített számítást használnak.[93] Megjegyzendő, hogy bár a GPU terén még az amerikai szilíciumvölgy (az NVIDIA) a világ vezető cége, ám védelmi szempontból kiemelendő, hogy az NPU terén kínai Huawei is az élvonalban van: már 2017-ben felzárkózott az iPhone-hoz azzal, hogy a Kirin 980 mobiltelefon-processzor NPU-t is tartalmazott[94] (ez még inkább képfeldolgozásra készült), de szerveret is építettek külön NPU modullal (mely négy általános NPU-t integrál).[95]
- **Vegyes rendszerek.** Sokszor érdemes az MI rendszerek mögé tervezett hardvereknél is vegyes megoldásokat alkalmazni, melyekben a fentebb felsorolt cél-hardverek együttesen vannak jelen. Erre jó példa az imént említett kínai szerveren kívül a Tesla SoC (System on Chip) típusú megoldása is. Ez alapvetően egy FPGA alapú hardver, kifejezetten az okosautók számára (amint neve is jelzi: FSD, Full Self-Driving chip). Egy tucat több magos CPU-t egészítenek ki benne GPU és NPU elemek.

Tekintsünk azonban túl ezeken a megközelítéseken, melyek kisebb ötletek mentén teszik optimálisabbá az MI számára a hardvert. Kimondható, hogy már benne élünk a hagyományos Neumann-elvű architektúrákból való kilépés korszakában. Itt csupán néhány izgalmas példát gyűjtöttem össze annak alátámasztására, hogy a hardveres területen is komoly diszrupciók várhatók.

- **Önmagukat átalakító hardverek** (*neuromorf rendszerek*). Kezdjük a legfuturisztikusabb megoldással. A felfedezést, melyre ez a technológia majdan alapulhat, néhány éve tették, ezért még egy teljesen kialakulatlan irányról van szó. A felfedezés lényege, hogy egy bizonyos anyagból,⁹⁴ szobahőmérsékleten, egyszerű elektromos impulzusokkal újrakonfigurálható alkatrészek hozhatóak létre. Vagyis egy olyan anyagot alkottak, amelyből az elkészített processzorcellák igény képesek négyféle alkatrészként működni: ellenállásként, kondenzátorként, neuronként vagy akár szinapszisként is. Tehát az ebből létrehozott processzor hardveres szinten könnyen átalakítható, úgy is, hogy mindig az éppen legszükségesebb funkciók futtatására legyen optimalizálva maga a hardver. Ennek első megvalósulásai az FPGA-hoz hasonló, de annál sokkal nagyobb távlatokat nyitó megoldást fognak jelenteni. Ezen azonban jócskán túl is mutatnak. Mivel a céllak az MI két alapelemét (a neuront és az azokat összekötő szinapszisokat) is modellezni képesek, így akár az adott célra optimálisabb MI-modellé képesek alakulni. Ez ugyan a tudományos-fantasztikus irodalmak disztópikus világát idézi, de megvalósulása meglehetősen kétséges, hiszen nem áll rendelkezésre neuromorf alapokra épített architektúra, sőt elegendő teszt és adat sincs ezek stabilitásáról vagy sorozatgyártási problémáiról.



11. ábra A Stern-féle mátrix logika kapcsolata más logikákkal (forrás a szövegben)

- **Kvantum MI.** Ez egy közismertebb nevű kutatási irány, mely azonban még újszerű megközelítés. Bár az előző példánál előrehaladottabbak a kutatások, igazából egy most rügyező technológiáról beszélhetünk. Történeti érdekességként hadd említsem meg, hogy a kvantumgépek egyik matematikai alapjait jelentő mátrix-logikát kidolgozó August Stern már 1970-ben felvetette a kvantumszámítógépek lehetőségét, jóval korábban az a fizikus,⁹⁵ aki

⁹⁴ A protonnal adalékolt perovszkit neodímium-nikelát (NdNiO_3) [96]

⁹⁵ A kvantumszámítógép ötletét így David Deutch fizikusnak tulajdonítják. [97]

kvantumelméleti alapon vázolta a technológia alapjait. Amint a 11. ábra[97] mutatja, a mátrixlogika kapcsolódásai továbbvezetnek minket a következő témához. Visszatérve: a köz-eljövő technológiáját láthatjuk, egyelőre a kvantumszámítógépek alaplehetőségei is most bontakoznak ki, így még nem áll előttünk ütőképes, teljesen kvantuminformaticai alapon létrehozott MI szolgáltatás. Ezen a tudományterületen fő cél olyan kvantumalgoritmusok megvalósítása, melyek jobban képesek kihasználni a kvantumszámítógépek lényegét. Elsősorban a hardver azon tulajdonsága fontos, hogy nem a kettes számrendszer korlátaiban dolgozik. Vagyis a quibitekre⁹⁶ írt MI kódok elvileg is jobban utánozhatják az agy működését, mint a jelen korunk elektronikai (kettes számrendszeren alapuló) megoldásai. Ezáltal a kvantum-MI rendszerek használhatóbbakká, életszerűbökké válhatnak. Ez a tulajdonságuk, karöltve azzal, hogy a kvantumgépek eleve nagyságrendekkel gyorsabbak, mint a hagyományos gépek egy valóban jelentős diszrupciót hozhatna létre. Érthető a vágy ennek birtoklására, a Google még nyílt forráskódú, hozzáférhető platformot is készített,⁹⁷ amely lehetőséget ad olyan MI modellek fejlesztésére, melyek a kvantuminformatica egyedi lehetőségeit használják ki. A kutatás jelen állása szerint már számos kvantumtanulási algoritmus működőképese,[98] bár leginkább a jelenlegi felhasználásnak igazán a hibrid modellek felelnek meg. Jelenleg a gyakorlati alkalmazhatóság felől az az irány tűnik legrealisabbnak, melyben a kvantumgép a tanulási fázist gyorsítja fel, míg a lekérdezéshez elegendőek a hagyományos platformok annál jóval lassabb képességei is.[99] Sok kutató azt várja, hogy 2025 az ilyen irányú áttörés éve lesz, ahol széleskörben elérhetővé válik a technológia, mivel 2024 során megoldottá vált a hibajavítás problémája, valamint a szobahőmérsékleten (lézerrel) működő megoldások.[100]

- **Módosított biológiai struktúrák.** Először ezúttal is a kevésbé ismert irányról szólok, amelynek lényege, hogy elektronikai elemek helyett az élet már működő alap-alkatrészeiből próbálnak meg működő gépet építeni. Az előző példánál kevésbé előrehaladottabbak a kutatások, aminek egyik oka talán az, hogy a „hardver” újszerűsége nekünk, embereknek még túl sok kihívást rejt – annak ellenére, hogy évmilliók óta működnek általuk az élet biológiai rendszerei. Körülbelül 1998-tól datálható a működő bionyomtatás[101], mely napjainkra komoly iparágga fejlődött. Részben erre támaszkodik az a megközelítés például, amelyben tenyésztett agysejteket, agy-organoidokat szeretnének felhasználni egy úgynevezett *organoid intelligencia* létrehozására.[102] Más kutatások a DNS memória-képességét

⁹⁶ Ez a kvantuminformatica alapegysége, mely a hagyományos bitekhez képest nem csupán a 0 vagy az 1 értéket veheti fel, hanem a kettő között bármilyen értéket.

⁹⁷ <https://quantumai.google/>

igyekeznek adatok tárolására felhasználni, vagyis élő sejt-tenyészetekben tárolni el később visszanyerhető adatokat.[103] Az efféle kísérletek a mesterséges, de nem elektronikus számítógépek reális alternatívái lehetnek, előnyük például, hogy kevés vagy nulla áramot fogyasztanak, nem úgy, mint jelenlegi fejlesztések (bár a biológiai alapú gépek környezetvédelmi aspektusa talán súlyosabb, mint az áramfejlesztés dilemmái). Ennek ellenére ezek használhatósága véleményem szerint egy igen távoli jövőbe tehető, ezért itt csupán említés szintjén elég volt utalni rájuk.

- **Biológiai struktúrákba ágyazott elektronika.** Az előzőre egyelőre csak nevében hasonlít, de attól alapvetően tér el a fejlesztés ezen iránya. Itt nem kiragadott alapegységekből építkeznek, hanem működő élőlényeket egészítenek ki (bővítenek) elektronikai berendezések beléjük integrálásával. Ezeket jelen kutatás azért nem taglalja, mivel amellet, hogy óriási terület, a jelen hipotézisek vonatkozásában is számos felvetést implikálna, így ezek tudományos vizsgálata megduplázná a terjedelmet. Néhány fontosabb kutatási terület említése is elegendő az ebben rejlő potenciálok érzékeltetésére, ugyanakkor a túlzott elvárások letérésére.

A legtöbbet nyilván az emberek ilyen jellegű kiterjesztésétől várhatunk. Ezek a fejlesztések az orvostudománnyal karöltve indulnak, de olyan irányt vesznek, hogy a gyógyításban használt technológiával ne csupán helyreállítsanak sérült képességeket, hanem túllépjék az emberi szervek képességeit. Ennek biológiai, kémiai és nano-megoldásait csupán említeni tudom, a tanulmány többi részével összhangban, az elektronikai eszközökkel való kiterjesztésekből emelek ki néhányat. A gyógyászati cél ott lép elő az elektronikai ipar vetélytársává, amikor a gyógyításon túl a művétagok erősebbek lehetnek az eredetnél, a mű érzékszervek érzékelési tartománya tágabb lehet a biológiai érzékszervekéénél, akár fül vagy orr képességeiről, akár pl. a mágneses mező vagy radioaktivitás emberi érzékelését lehetővé tevő, eddig nem létező szervekről beszélünk. Ezek számos plusz információt adhatnak a birtokosuknak, ha például egy műszem szélesebb látószögű, jobban nagyít vagy az infravörös és ultraviola tartományokban is működik. Sőt a szélesebb képességeken túl, akár kiterjesztett-valóság eszközként is szolgálhat: a műszem például képernyőként is funkcionálhat, mint a mai XR (eXtended Reality) szemüvegek, ezzel egyben IoT szenzorként is használható, amely az információkat az azt használó ember agyán kívül egy digitális felhőbe is elküldi. Ezek a példák a hagyományos agyi kogníció imputjának kiterjesztési lehetőségéről szólnak, melyek küszöbön álló vagy már létező technológiák, de az emberi felfogásra, tudatra vagy lélekre gyakorolt hatásuk egyelőre ismeretlen.

Kifejezetten a kognitív képességeink kiterjesztését az agyi implantátumok célozzák, ezért az MI szempontjából ezekre fontos pár szóban kitérni. Sikeres agykiterjesztés esetén első körben

pl. a számolási képességünket lehetne a CPU-k képességeivel egy szintre hozni, vagy memóriánkat lehetne pontosabbá, illetve importálhatóvá és exportálhatóvá tenni: vagyis fel lehetne tölteni egy emberbe lexikális tudást vagy le lehetne tölteni emlékeket.[104] (Ha lehetne...) A következő lépés az ebben bízó tudósok szerint az aggyal együttműködő beültetett számítógép modellje lenne. Ez a hardverarchitektúra-modell egyesítené az emberi és gépi intelligencia előnyeit. Ez a szupergyors, szuperpontos, szuperokosság ugyanis birtokolná az emberi alany születéskor kapott, agya által rendelkezne erkölcsi, kreatív, művészi képességekkel és valós érzelmekkel is. Ezáltal egyből meg lennének oldva az erősen emberi tulajdonságok (emberségesség) gépiesíthetőségének nehézségeit is (melyről az IV. fejezet szól). Sokan az evolúció következő lépcsőfokának tartják az így létrejövő szimbiózist, melyre inkább a „szingularitás” (ld. 1.2.3.) kifejezés terjedt el. A saját kutatásaim a témában részlegesebbek, itt csupán megemlítem, hogy túlzónak tartom ezt a megközelítést, és a tudományos forrásokban nem találtam lehetőségét az emberi agy kommunikációs formátumának feltörésére. E nélkül pedig az ilyen eszközök kényelmes, speciális és hosszas tanulás nélküli használata még csak nem is a science fiction, hanem inkább a fantasy irodalmi műfajába sorolható (picit részletesebben ld. III.5.2.).

Az MI fogalma szempontjából megállapítható, hogy arra a jelen hardverei nem várható, hogy hatást gyakorolnak, sőt a jövő felvázolt megoldásai közül is csupán a biológiai struktúrákba ágyazott elektronika, elsősorban az emberi agyi kiterjesztések esetleg áttörései lehetnek lényeges hatással. Úgy vélem azonban, hogy amennyiben ez valósággá válik, akkor sem az MI fogalmat kell tovább bővíteni, hanem maradni kell a jelenlegi külön megnevezésnél, a mesterségesen kiterjesztett emberi kogníció (*Artificially Extended Human Cognition*, AXHC) terminusnál.

II.3.2. Fuzzy logika, az „elmosódott igazság”

A fuzzy logikában rejlő szemlélet jelentősége az informatika alapjait érinti, hiszen a világ egy, a korábnál használhatóbb leképezéséhez ad matematikai alapot. Itteni ismertetését ez a sajátos leképezési mód indokolja, hiszen e miatt tud jól kapcsolódni a gépi tanulás vagy az MI számtalan modelljéhez is, bár egy hagyományos programozási eljárásról van szó.

A mondás szerint „nem minden fehér vagy fekete”. Más szóval a valóságot sokszor csak erőltetve és nehézkesen lehet a klasszikus logika „igaz vagy hamis”, két bites állításaival leírni. Két probléma is van ebben a leegyszerűsítés-típusban, melyek a köznapi életben is gyakran okoznak gondot:

1. Egyrészt a kifejezéseink nem elég pontosak, hiszen fogalmaink határai nem élesek, minden szavunk egy halmaznyi árnyalatot foglal magába.
2. Másrészt a kifejezések összekapcsolása „kizáró vagy” (xor)⁹⁸ művelettel legtöbbször túl egyszerűsíti a valóságot.

Az imént 1. számmal jelölt problémát fuzzy ötlet azáltal oldja fel, hogy definiál egy „nyelvi változó” nevű változót, melynek értékei valamely természetes vagy mesterséges nyelv kifejezései. Ez a nyelvi változó a fogalmakat eleve halmaznak kezeli: ha a fekete és fehér közötti tartományt szeretnénk leírni, akkor a nyelvi változók a *világosszürke* és a *sötétszürke* lesznek. Egy közepesen szürke árnyalat tartozhat mindkettőhöz – csakhogy ezen a ponton a példa nem a legmegfelelőbb. Jobb példa, ha a *sebesség* a nyelvi változó, és annak címkéi lehetne a *lassú*, a *sebes* és a *gyors*.

A fenti 2.-es problémát a hagyományos módszer nem kezeli, vagy gyors egy jármű vagy lassú. Fuzzy logikában viszont a három címke által leírt értelmezések halmazain különböző halmazelméleti műveleteket (unió, metszet, kiegészítés, implikáció) alkalmazhatunk, példánkánál maradván a *lassú* és a *sebes* nem zárja ki egymást, ha létezik a két halmaz metszete. Ez a metszet azonban nem a szokásos, a metszetben lévő közös elemek eltérően tartoznak egyik és másik halmazhoz. Matematikailag ezt a módszer úgy kezeli, hogy az elemek nem egyszerűen hozzátartoznak egy halmazhoz, hanem a halmazokban az elemeknek *tagsági fokuk* van. Minél magasabb a tagsági fok, annál inkább tartozik az elem a halmazhoz, ezt egy 0 és 1 közötti érték írja le. A tagsági fok tehát azt fejezi ki, hogy a nyelvi változó címkéje „milyen mértékben igaz” – szemben azzal, hogy igaz vagy hamis (0 xor 1). Így tér el ez a modell a klasszikus logikai megközelítéstől, ahol nem lehet pl. a „közepesnél valamivel lassabb” állapotot megragadni, itt viszont igen.

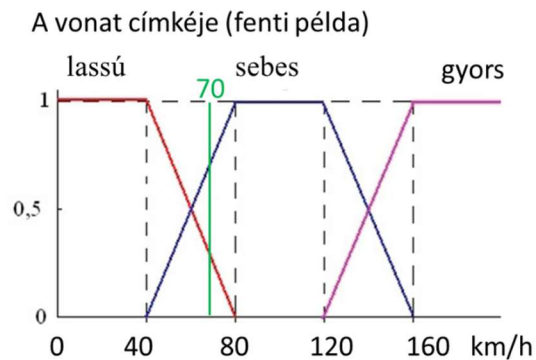
Az alaplómódszer tehát az adatokat nem éles adatként dolgozza fel, hanem

- először leképezi az adatokat ilyen elmosódott halmazokba (fuzzifikáció)
- ott feldolgozza őket különféle ha-akkor szabályok által
- végül persze visszaalakítja éles értékévé az adatot, (defuzzifikáció) további adatkezeléshez.

⁹⁸ XOR: Vagy északnak megyek vagy délnek, két irányba egyszerre nem lehet. Az elektronikai-logikai vagy (OR) kapcsolat magyarul gyakrabban fordítható az „is”, illetve az „akár” szavakkal, de néha a „vagy” is megfelelő: „mindegy, hogy kutyát tartasz, vagy macskát, vagy mindkettőt”.

A 12. ábra⁹⁹ mutatja *fuzzifikációt*, vagyis, hogy hogyan lehet matematikailag megközelíteni a tagsági fok függvénye segítségével a nyelv és a valóság pontatlan címkéit. Látható, hogy a középső (*sebes*) halmazba a 40-160 km/h közötti értékek tartoznak, de nem éles (\leq vagy \geq) határokkal, mint egy klasszikus algoritmusban, hanem külön kezelünk átmeneti tartományokat.

Az átmeneti tartományok egy másik címkével való közös halmazt jelölnek. A tagsági fok ezekben vesz fel a 0 és 1 közötti értéket, amint látható az egyik címke (*sebes*) esetében növekvő, a másik (pl. a *lassú*) esetében csökkenő mértékben. Így egy 70 km/h-val mozgó jármű „inkább sebes, mint lassú” értéke a két függvény



12. ábra: sebesség fuzzifikálása példán (saját szerkesztés)

együttes figyelembevételével¹⁰⁰ kezelhető. Így válnak számolhatóvá válnak bizonyos nyelvileg amorf fogalmak, mint a „kicsi-nagy”, „erős-gyenge” stb. Ez azonban nem oldja meg a leképezés mérhetőségi problémáit, vagyis a „erkölcsös” és a „erkölcstelen” leképezésére önmagában ez a módszer nem ad választ.

Sok esetben a valóságot persze nem ilyen lineáris átmenet jellemzi, ezért nem csak az ábrán látható lineáris függvénnyel érdemes számolni (igen elterjedtek pl. a harang alakú függvények is). Ez a matematikai módszer nem csupán kezelhetővé teszi a homályos leképezéseket, de sajátos logikai következtetésekre is lehetőséget ad. Ezek rendezett sorozatából „fuzzy algoritmus-utasítások” alkothatóak, vagyis egy programkód hozható létre, pl. HA „A” kicsi AKKOR „B” nagyon nagy (ahol a *kicsi* és a *nagy* fuzzy halmazok értékei).¹⁰¹ Vagy egy sebesség-szabályozást végző algoritmust, mely a fenti példa halmazait használja: HA [[fékhőmérséklet = meleg] ÉS [sebesség = NEM villámgyors]] AKKOR [féknyomás = = kissé csökkentve].¹⁰² Ez a példa-algoritmus azt a problémát oldja meg, hogy „ha a fékek kissé már melegednek, de a jármű sebessége még nem túlságosan gyors, akkor kicsit kell csak csökkenteni a féknyomást”.

Miután az elmosódott halmazokba leképeztük a problémát, majd ott függvénnyel és kóddal kezelhetővé tettük azt, szükség van újra olyan szám adatokra, melyekkel tovább számolhat a

⁹⁹ A 12. ábra kiindulópontja a következő cikk: [105]

¹⁰⁰ Ekkor a tagsági fok tehát a *sebes* címkénél 0,75 és *lassú*-nál 0,25.

¹⁰¹ A módszer megalkotójának példája. Ld. [106, o. 29]

¹⁰² A példában a fékhőmérséklet címkéi: „hideg”, „langyos”, „meleg”, „forró” lehetnek, a féknyomás pedig „kissé csökkentve”, „csökkentve”, „változatlan”, „kissé növelve”, „növelve” címkéjű értékeket vehet fel.

rendszer. Sokféle matematikai művelettel megkaphatjuk ezt a végeredményt, vagyis *defuzzifikálhatjuk* az elmosódott halmazon végzett számítást. Ez a művelet általában a halmaz súlypontjának vagy valamilyen középértéknek kiszámítása.¹⁰³ A fenti program-példában erre utal a „*kissé csökkentve*” címke, mely tehát nem csupán egy „*csökkenteni XOR nem-csökkenteni*” választás, hanem egyből egy automatikus csökkentés-mennyiséget ad döntésként.

Végül tekintsük át a modell terjedésének történetét. A matematikai modell felvetése ugyanarra az intellektuális hullámra nyúlik vissza, melyben az informatika egy időre megpróbált kitörni a számítógépekkel kapcsolatos értelmezésből (ld. V.2.1.), akkortájt, amikor a kognitív tudományok és az MI fogalmait is meghatározták. Mint láthattuk a fuzzy modell újfajta hidat képez a nyelvi és a matematikai logika között, ennek ellenére lassan vált csak ismertté és népszerűvé. 1965-ös felvetése^[108] után vagy negyven éven át a mérnökök automatikus szabályzásokhoz alkalmazták, nem-kihasználva rokonságát a neuronhálókkal (egyik sem a kizárólagosságra alapul). Pedig a módszer megalkotója, az iráni-azerbajdzsáni származású, Amerikában dolgozó Lotfi Aliasker Zadeh is világosan látta ezt a lehetőséget benne, amikor és 1973-ra továbbfejlesztette a korábbi felvetést, és kidolgozta az elmosódott nyelvi változók használatát. Modelljét a kognitív tudományok minden részében hasznosíthatónak mondja: „A nyelvi változók és fuzzy algoritmusok (...) fő alkalmazási területei a közgazdaságtan, a menedzsmenttudomány, a mesterséges intelligencia, a pszichológia, a nyelvészet, az információkeresés, az orvostudomány, a biológia és más olyan területek, ahol a rendszerelemek élő, nem élettelen viselkedése játssza a domináns szerepet.”^[106] Megállapítása igaznak bizonyult, a fuzzy megközelítés azóta is egyik praktikus módja azoknak a leképezéseknek, melyek a kognitív tudományok eredményeiből a számítástechnikában hasznosítható modelleket hoznak létre.

Visszatérve a terjedésre: igazából csak az imént idézett cikk után kezdett el több egyetemi kutatócsoport foglalkozni az ötlet felhasználhatóságával. Az első fuzzy-szabályzót 1980-ban, Angliában tervezték (egy gőzgép számára).¹⁰⁴ A fuzzy technológia kirobbanó sikerére Japánban került sor 1987-ben, hiszen a Hitachi ekkor kezdte el használni egy vonat vezérléséhez, az Omron pedig kereskedelmi forgalomba helyezte a világ első fuzzy alapú szabályozóját, sőt autók automata sebességváltójánál is alkalmazták. ^[105] Ezután a módszer gyorsan terjedt, annak ellenére, hogy már az 1990-es években megfogalmaztak ellene éles kritikát. ^[110] A modell számos további matematikai módszer alapjává vált, ^[111] így nem meglepő, hogy az automatizálás

¹⁰³ Több ilyen módszer algoritmikus leírása (a fentebb említett szín-példán is): ^[107]

¹⁰⁴ A következő történeti adatok forrása: ^[109]

egyre több területén talált hasznosításra, ahogyan azt alapítója megjósolta. Hátránya, hogy több tesztet igényel, viszont cserébe a pontatlanságot sokkal biztonságosabban kezelő rendszerek alakíthatóak ki. Megjegyzendő továbbá, hogy nem szabad összetéveszteni a valószínűségi elméletekkel, melyek más jellegű automatizálási képességeket tesznek lehetővé [112] – ez pedig átvezet minket az valószínűségi alapon működő MI rendszerekhez.

Az MI fejlesztések sokáig nem használták a fuzzy halmazokat, a gépi tanulás működik nélküle is.[113] Viszont számos téren képes ez a leképezés igen jól támogatni az MI-t. Ennek oka, hogy a fuzzy modell világunk természetes bizonytalanságait sokkal jobban kezeli, mint egyéb megközelítések. Néhány kiemelendő területet említek csak, ahol a fuzzy megközelítés alkalmazása már bevált.[114] A mintafelismerésben (kép és szöveg) játszott fontos szerepet játszik, hiszen a zaj-csökkentéstől kezdve a pontatlan egyezések kereséséig sok területen bevethető. A döntéstámogató rendszereknél az emberekéhez közelebb döntéseket javasol. Az adatbányászatban és a prediktív modellezésben is kiemelkedően funkcionál, hiszen kezelhetővé teszi a nagy adatkészletek bizonytalanságát és a nem nem-pontos, csak hasonló minták felismerését. Természetesen az MI alapú vezérlőrendszerek és robotikai alkalmazások a klasszikus fuzzy megoldásokat együtt képesek használni az MI-vel kapcsolatos imént említett felhasználásokat, ezáltal sokkal rugalmasabb és emberibb, ugyanakkor biztonságos eredményt születhet.

Csak hogy ennek a megközelítésnek a kiterjesztése az élet minden területére, súlyos etikai problémát hordoz. A téma messzire vezet, de megjegyzem, hogy véleményem szerint a többek által már feszegetett fuzzy-etika (pl. [115]) nem alkalmazható minden döntéshelyzetben, hiszen sokszor csak szélsőségek (igen vagy nem) közül lehet választani, és a döntésen emberéletek múlhatnak.

Összegzésképp kijelenthető, hogy a fuzzy megközelítés teljesen beépült az MI fogalom tartalmába, ezért definíció újragondolásakor nem szükséges figyelembe venni. Viszont a mester-séges erkölcs szempontjából (IV.2.3.) igen fontos volt vele foglalkozni. A fentiek alapján világossá vált, hogy a világ leképezésének szempontjából is mennyire jelentős ez a szemlélet.

II.3.3. Biológiai ihletettséggű informatikától a rajntelligenciáig

Nulladik Neumann elvnek[116] vehetnénk fel, hogy „másoljuk le az élő szervezetek működési elveit”. A híres magyar tudóst ugyanis meglehetősen érdekelte az agy működése, pontosabban az, hogy hogyan lehetne az agy kutatás tanulságait felhasználni számítógépek tervezéséhez (egyik fontos munkájának a címében is szerepel az agy[117]). Később az első MI-modelleket is az emberi agy idegsejtjeinek kapcsolatrendszere ihlette. Azonban a műszaki tudományok nem csupán az agy működéséből próbálnak ötletet meríteni, a biológia számtalan

felfedezését sikerült már felhasználnia a mérnököknek, sőt egy új tudomány, a bionika (bionics) kifejezetten az ilyen lehetőségek számbavételére koncentrált.[118] Mivel az élet működési módjának megoldásait sem a Véletlen, sem egy Teremtő nem védte le szabadalmi joggal, ez egy beláthatatlan méretű ötletbank áll rendelkezésünkre.

A bionika informatikai szegmensére, a biológiai ihletésű számítástechnikára (*bio-inspired computing*) várhatóan a jövőben még nagyobb hangsúly fog helyeződni. Hiszen bár az MI nélkül is jól működik számtalan ilyen algoritmus, de az MI fejlődésével az ilyen ötletek gráfja „virágba borult”, újabb és újabb MI irányokat ihlet a biológia. Ennek néhány fontosabb területe címszavakban: ¹⁰⁵ (1) Mesterséges idegi hálózat, (2) Genetikai algoritmusok, (3) Populációs modellek, (4) Genetikai programozás, (5) Membrán számítástechnika, (6) Sejtautomaták, (7) Számítógépes immunrendszerek, (8) DNS számítástechnika, (9) Rajintelligencia és rajrobotika, ezen belül például részecske raj optimalizálás vagy hangyatelep optimalizálás és még sorolhatnánk. Vagyis az MI bevonásával is már igen sokrétűen „lopunk” az élet működésének megfigyeléséből informatikai módszereket, és a lista folyamatosan bővül.

Mivel ekkora tudáshalmazról van szó, ezért itt csupán a populációs modelleket tekintem át példaként. Ezen terület egyaránt tartalmaz MI nélküli és MI-t használó megoldásokat, tehát rövid bemutatása is érzékelteti biológiai inspiráció sokféleségét (és témánk szempontjából nem is szükséges mindent megvizsgálni). A populációs modellek a hatékonyan működő állatközösségek világából merítenek ötleteket, vagyis etológiai megfigyeléseken alapulnak. Mivel az életbenmaradásért folytatott verseny évmilliók óta fejleszti ezeket az életközösségeket, elmondható, hogy elég optimálisan működnek, számos optimalizációs feladat megoldására kínálnak a korábbiaknál sokkal hatékonyabb megoldást. Egyik legrégebbi alterületről van szó: már a kilencvenes évek elején megjelentek olyan algoritmusok, melyek a hangyáknál felfedezett feromon-alapú kommunikáció inspirált. A 2000-es évek elejétől a hangyaboly-optimalizációt számtalan területen használják,¹⁰⁶ még kibervédelemben[121] is. A biológiai közösségek működéséből számos olyan algoritmus-ötletet sikerült meríteni, melyeknél az MI is segítheti a számítógépes implementáció hatékonyságát. A terület áttekintéséhez egy lista állítottam össze, ám részletesebb kutatását elhalasztottam, itt csak az elkészült felsorolást közlöm:

- a méhraj-intelligencia,¹⁰⁷
- mesterséges halraj algoritmus,[123]

¹⁰⁵ Az alábbi listában szereplő területekről egy rövid átfogó képet ad: [119]

¹⁰⁶ A könyv első verziója 2004-ben látott napvilágot, melyben a szerzők számos felhasználási területet is bemutatnak. [120]

¹⁰⁷ pl. a Naperőművek optimalizálására. [122]

- a szentjánosbogarak (fénylegyek) párzási motivációját használó modell,[124]
- pl. a pillangó algoritmus,[125]
- a kakukk tojásrakása a Kakukk-keresésben,¹⁰⁸
- a farkasvadászat a Szürke Farkas Optimalizálóban,[127]
- érdekes továbbá ezek rendszerezése is,¹⁰⁹ de számos egyéb ígéretes kutatás is folyik.

Ezek mellett a biológiai szimbiotikus együttműködések tudatos átültetéseit is fel lehet használni,[128] ezek az együttélés (*mutualizmus*), a *kommenzalizmus*,¹¹⁰ vagy a *parazitizmus*, sőt ezek finomabb modelljei felé, vagyis egyre bonyolultabb együttélési-modellek algoritmus leképezése felé haladnak a fejlesztések. Megjegyzendő, hogy nem ezt hívják *szimbiotikus intelligenciának* (mint már említettem): tehát ez a szókapcsolat nem a fajok együttéléséből vett MI modellekre utal, hanem az ember-gép együttélés optimális komplementer megvalósulására használják[7], amelyre én Szinergikus MI-t javasoltam (I.1.1.).

Kicsit behatóbban a rajintelligencia területével foglalkoztam (melyet a populációs modellekhez szokás sorolni). Ugyanis ez nem csupán katonai szempontokból nagyon ígéretes terület, hanem általános felértékelődésére számítok. Felhasználásainak áttekintése előtt nézzük meg, mit is jelent ez a „raj” modell.

Fentebb bemutattam, hogy a többi MI modell alapjai a sejtek (neuronok), ahogyan egy szerv, illetve az agy esetében. A raji intelligencia lényege, hogy alkotórészei nem sejtek: a raj egyedekből (entitásokból) áll. Ezek az önmagukban is működő entítások együtt, közösségben sokkal sikeresebben és hatékonyabban képesek a környezeti kihívásoknak megfelelni, mint egyénként. Az ehhez szükséges együttműködés fontos mozzanatait igyekszünk az algoritmusokban felhasználni. A raj entításai kommunikálnak egymással, így képesek a másik entitás által tanult információt maguk is felhasználni elsősorban a közösség (illetve az aktuális közös cél) számára. A raj hatékonysága azért is nagyobb, mert lényegtelen, melyik entitás ér el valami hasznos célt, mert ha valamelyiknek sikerül, akkor a közösség maga érte el azt a célt. Bizonyos nagy közösségekben az sem számít, ha néhány entitás megszűnik (elpusztul) a megvalósítás közben. Például, ha egyik hangya élelmet talál, akkor feromon jelzései által a többiek is akkor is képesek lesznek hozzájutni és a boly egészét táplálni vele. És a boly tovább él akkor is, ha a keresés közben meghal közben pár száz munkás hangya – vagy katonai adaptációjánál a cél elérhető úgy is, ha néhány drónt kilőnek a drón-rajból.

¹⁰⁸ Az itt felsorolt több algoritmus mellett szerepel ez is ebben a műben: [126, o. 105–116]

¹⁰⁹ S, Arockia Panimalar: Nature inspired metaheuristic algorithms. International Research Journal of Engineering and Technology (IRJET), 4. (2017), 10. 306–309. <https://www.irjet.net/volume4-issue10>

¹¹⁰ Populációk olyan kapcsolata, amely az egyik fél számára előnyös, a másiknak közömbös (pl. ürüléktermelő állatok számára közömbösek az ürülék feldolgozó organizmusok).

Informatikai megvalóítás esetén az állatokhoz képest sokkal tökéletesebben adható át információ. Ezért egy MI raj entitásai egymástól sokkal többet tanulhatnak, mint a legtöbb élő raj egyedei képesek egymásnak tanítani. Vagyis egy mesterséges raj esetén ilyen közös tapasztalatokból a cél elérésének leghatékonyabb módja gyorsabban csiszolódhat ki. Továbbá a természetben is megfigyelhető az alkalmazkodás, amikor néhány generáció alatt az adott faj egyedeinek képességei valami új helyzethez alkalmazkodnak, például a táplálék változásával átalakul a csőrük, a hőmérséklettel összefüggésben a szőrzetük, vagy egyes belső szerveik erősebbek lesznek, mint más populációban. Ehhez hasonlóak a raj-MI entitásai utódjaikban módosítják saját programkódjukat is, optimálisabbá tehetik azt a közösség tapasztalatai alapján. E rövid összefoglaló végén megemlítendőek a hierarchikusan és nem hierarchikusan szervezett rajok. Vagy van falkavezér és többféle raj-szerep (pl. munkás és katona hangya) vagy nincs.

A raj számítógépes implementációját leoptimálisabban egy elosztott rendszer képviseli. Hierarchikus modellnél van egy központi vezérlő-elem, ennek szerepét könnyen átveszi a „rang-idős”, amennyiben az az elem sérül vagy megszűnik vele a kommunikáció. Nem hierarchikus rajok esetén nincs is szükség központi elemre. Ez a jelenlegi központosított MI módszerekkel szemben sokkal életképebbnek tűnik, és ezzel visszaérkeztünk a rajintelligencia jövőben jelentőségének vizsgálatához.

Véleményem szerint a komplex MI rendszerek architektúrájában alapvető szerepe lesz ennek a fajta szerveződéseknek. Általános MI fejlesztések esetében bizonyosan számíthatunk raj-képességek integrációjára, de már jelenleg is folynak ilyen jellegű kutatások.[129] Elsősorban több mélytanuló gép rajba szervezése jelent majd nagy lépést. Ezeknél a fentebb említett GAN rendszerhez hasonlóan, de azoknál jobban működne egy versengéses-tanítás, hiszen kettőnél több egyed versenyezne. Emellett a fejlett információmegosztástól a kooperáción keresztül az emberi közösség számos hasznos tulajdonsága is leképeződhetne.

Így egy *mesterséges közösségi intelligenciát (artificial community intelligence, ACI, saját kifejezés¹¹¹)* alkotó entitás-rendszerekből álló, összetett MI-raj, más néven *mesterséges mélyentitás hálózat (Artifitial DeepLearning Network – ADLN)* komoly lépés lehet a mesterséges erkölcsi érzék betanításához vezető modellek felé is (ezzel IV.2.5-ben részletesen foglalkozom). Végül kiemelendő, hogy az itt felvetett MI-dimenzió minden MI definícióra ki kéne, hogy hason. Ezen belül fontosnak látom már ma megkülönböztetni a raj- és központosított- modelleket a különböző MI-értelmezések megfogalmazásakor.

¹¹¹ Ilyen értelemben saját, hiszen rákeresve olyan találatokat kapunk, ahol nem egy entitás-hálózatot értenek alatta, hanem csupán az MI használatát az emberi közösségben. Pl. <https://www.eitfood.eu/projects/eit-community-artificial-intelligence>

II.3.4. A természetes nyelvfeldolgozás (NLP rendszerek)

Az utóbbi évtized egyik leglátványosabb fejlődést megvalósító technológiája a *Natural Language Processing* (NLP), vagyis a természetes nyelvek feldolgozása. Pár szóban össze kell foglalnunk legfontosabb dolgokat róla,¹¹² mivel pár éve hirtelen kiemelkedett az MI szolgáltatások közül és igen divatos kérdéskörre vált. Elértük ugyanis azt a szintet, hogy a gépek képesek legyenek értelmezni a köznapi emberi kommunikációt és szokásos nyelvezettel reagálni rá. Eddig a gép csak adatok kezelésére volt képes, ezzel kezdődött el egy valódi gépi *információkezelés*. Más szóval a technológia igazi ereje és újdonsága az, hogy most már szinte információ (=értelmezett adat) az, ami a gép inputjában és outputjában megjelenik, míg korábban az információt nekünk kellett a gép számára kezelhető adattá alakítani. Persze a nyelvi megfogalmazás is adattá-alakítás – de az NLP lényege hasonló az emberek között megszokott kódolás-dekódolás folyamattól, ettől emberszerű. Ezáltal válik kezdhetnek megvalósulni a „szimbiotikus MI” (nálam szinergikus MI (I.1.1.) néven már említett törekvések, vagyis egy ember-gép együttműködés és együttdolgozás, egy ideális kooperáció.

Itt említendőek meg az ún. NLG, azaz természetes nyelv generáló (*Natural Language Generation*) algoritmusok, mivel hasonló nevük zavaró lehet. Korábban ez külön fejlesztési terület volt, amely az adatok értelmes szöveggé alakításáról szólt. Ez az irány napjainkra integrálódott az NLP-be.

Az NLP rendszerek kísérleti verziói régóta nyilvánosak, de a ChatGPT szolgáltatás 4-es verziója kapcsán robbant a köztudatba a technológia, mely ugyan még nagyon sok tekintetben szorul javításra, mégis úgy 2022-től használhatóvá kezdett válni. Ez az áttörés a személyi mikroszámítógépek megjelenéséhez hasonlítható az 1980-as években, mert bár sokan hallottak a számítógépekről korábban is, akkor vált megfizethetővé, hogy bárki saját programokat írjon egy könnyen tanulható programnyelv által. A párhuzam jól rávilágít a technika kezdetlegességére, amit viszont ellensúlyoz, hogy sok embert lepipál világos, nem sértő nyelvezete mely képes eltérő korosztályok számára vagy irodalmi és szleng stílusban fogalmazni, sőt még humora is van.¹¹³

A technológia máris látványos változást hozott létre. Elsősorban a programozó szakma alakult át általa óriás sebességgel: a kódolás alapjainál és számos tesztelésnél már most is szinte

¹¹² A témáról részletesebben is értekeztünk, ld. [130]

¹¹³ Még közelebb hozhatja az ilyen szolgáltatásokat minden felhasználóhoz a Microsoft Copilot funkció, mely a megszokott MS Office dokumentumokba integrál egy olyan párbeszéd-ablakot, ahol természetes nyelven kérhetünk táblázat-műveleteket, egy prezentáció legenerálását, vagy egy levél adott szempontok szerinti udvarias megválaszolását és hasonló hasznos dolgokat. ld. [131]

minden cég az MI-re támaszkodik. A szakember határozza meg, hogy egy függvénynek mit kell majd csinálnia, aztán ellenőrzi és pontosítja a generált kódot, de így is felgyorsult a fejlesztés. Számos nyelvhez kötődő szolgáltatás ment tönkre, elsősorban olyan fordítóirodák, melyek felhasználói kézikönyveket és hasonló nem-művészi szövegeket fordítottak. Ehhez hasonlóan várható további sablonos szolgáltatások vagy munka-funkciók munkaerőpiaci szerepének lenullázódása is. A fejlesztéseket erősen motivációja az MIKT rendszerek erősödése is, hiszen az azokban keletkező óriási mennyiségű szöveg feldolgozása kizárólag szövegértelmező gépekkel lehetséges. A BigData esetében ugyanis nem lehet lekérdezés-kódot írni, mely a kívánt válaszra szűri a hagyományos (strukturált, relációs) adatbázist. A karakter-alapú kereséseken túl szükség volt egy jelentésre alapuló, „intelligens” keresésre is, mely képes záros időn belül a megfelelő információt kinyerni a BigData strukturálatlan halmazából.[132] Ehhez pedig a gépnek értelmeznie kell a szavak összefüggéseit is.

Itt kiemelendő, hogy az NLP rendszerek nem képesek „gépi értés”-re, de képesek egy szöveg gépi értelmezésére. Ezért bár látszólag irathatunk vele tanulmányt, összeállíthatunk prezentációt, alkottathatunk leírásból képet vagy videót stb., ez valójában csak egy nyelvi zsonglőr-mutatvány, egy jól betanított szuper-papagáj tevékenysége.

Az ehhez szükséges modellt a nyelvészet adta: a hagyományos, emberi szövegkezelés nyelvészeti szintjeinek megfelelően hozták létre az NLP rétegmodelljét. [133] Ezáltal karakterektől (vagy hangoktól) a szöveggörnyezet módosító hatásáig képes a gép minden szövegréteg feldolgozására, így képes szöveget értelmezni vagy előállítani. A szövegfeldolgozás egyik kulcsfogalma a *token*, melyre még nincs megfelelő magyar kifejezés. A token az NLP szolgáltatásokat felhasználó oldaláról azt jelenti, hogy a gép egyben értelmez egy beszélgetést (bizonyos karakter-határig), és annak belső összefüggései alapján reagál. A technológia felől viszont a token, azt jelenti, hogy az NLP rétegenként, az adott réteghez tartozó specifikus részekre bontja szét a szöveget (úgymond szintenként „tokenizálja” azt).[134] Fejlettebb NLP-k képesek a szövegben rejlő hangulat érzékelésére, sőt kicsit hibás mondat értelmezésére is.

Bár sokan vágnak arra, hogy az NLP egy mindentudó, szinte gondolatolvasó varázslógép-ként találja ki, hogy mit szeretnének, ám a valóság, hogy nekik kell megtanulniuk a gép nyelvét. Ez nem annyira elvont nyelvezet, mint egy programozási nyelv, mégis emberi intelligencia és némi tudás kell a hatékony *prompt* (gépnek szóló kérés) megfogalmazásához. Az alapos, de nem szószátyár, világos, és részletekre kiterjedő utasítások fognak a számunkra megfelelő választ adni, de újszerű kérdés-feltevési stratégiákat is érdemes terveznünk ahhoz, hogy összetettebb kérdésekre a géptől a megfelelő választ kapjunk. Mindez tehát kissé nehezebb, mint a hagyományos internetes keresőkérdések, ám ha jól csináltuk, és szerencsénk van, akkor az

eredmény messze meghaladja a böngészős keresések határfokát. Ha nem jól csináltuk vagy a gép nem tudja, akkor egyelőre rosszabb a hagyományos módszernél, és talán ez a fő oka, hogy az óriási perspektíva ellenére magyar felhasználók nagy része nem kezdte el még munkájához használni az NLP-t, vagy csak szórakozásból próbálta ki némelyik ingyenes szolgáltatást. Az egyéb okokat kitalálhatja bárki, aki használt már ilyen szolgáltatást, de a későbbiek miatt hasznos saját tesztelésem alapján¹¹⁴ pontokba szednem a terjedés lassúságának okait, hiszen ez egyben rámutat a népszerűség várható visszaesésre is. (vö. III. fejezet)

- A látványosan téves válaszok, valamint olyan gépi hallucinációk, amikor úgy tesz a gép, mintha tudná a választ, ír valamit, ami köszönőviszonyban sincs a valósággal (sok esetben belekérdezve közli, hogy tévedett).
- Egyre több szolgáltatásért kell fizetni, ráadásul a magyar keresethez képest kissé drágák. Így, aki nem kényszerül rá a konkurencia miatt (pl. programozó cégek rákényszerülnek) az nehezen fog rá beruházni.
- Az előző ponttal összefüggésben: az államapparátus részéről nem történt lépés ilyen szolgáltatások megvételére közszféra számára;
- Jól megírt promptokból számtalan érzékeny adatot ki lehet következtetni, ezért védelmi szempontból az előző pont érthető és támogatandó. Legfeljebb oktatási célokra vagy állami szoftverfejlesztői célokra vethető fel államilag finanszírozott NLP szolgáltatás;
- Magyarul sok szolgáltatás nem tud, vagy olyan fordítót használ, amely pontatlan, így az angol nyelv megfelelő szintű ismerete is gátolja, hogy például a kevésbé képzett rétegben felmerüljön a használata (olyanoknál, akik hagyományos számítógépes szolgáltatásokat azért használnak).
- Meglehetősen kevés magyar forrást használnak a külföldi rendszerek, így információ műveleti szempontokat is felvet például a környező államokkal való történelmi viták terén, hogy egy-egy szolgáltató kinek a szakanyagait teszi tanítóadattá;
- A jogi problémák miatt az NLP-k által kezelt tudásbázist nehéz egyszerre hitelesen és naprakészen tartani. A sajtószabadság még nehezen kezeli az álhírrrel való információ műveleteket, és gépi automatikák segítségével az álhíreknek csak egy része szűrhető ki;
- Az emberek nem szeretnek új szemléletet tanulni, akik 5-10-40 éve használják a hagyományos számítógépes megoldásokat, nehezen motiválhatóak, hogy az MI-hez szükséges paradigmaváltásra rászánják energiájukat és idejüket – nem érnek rá ilyesmire... (ld. vertikális tanulás III.2.2.)

Összegezve: később gyakran lesz szükséges a fentebb vázolt ismeretanyag. Az MI általánosabb definíciójához való kapcsolat vonatkozásában viszont nem ad az NLP részletezése módosító tényezőket – talán azért, mivel eleve ennyire fontos az MI technológiák között.

¹¹⁴ Az itt leírtakat saját tesztelések, három saját workshop és egy saját cikk tudásanyagából állítottam össze.

II.4. RÉSZÖSSZEFOGLALÁS: MILYEN AZ MI KÍVÜL-BELÜL?

Az Első Résszel kapcsolatban kitűzött célokat ebben a fejezetben is sikerült megvalósítani. A két fejezet együttes vizsgálatai alapján a K3 kérdés első felére bizonyosan igennel lehet már válaszolni, hiszen több tényezővel kapcsolatban is kimutattam, hogy hiányoznak az MI meghatározásaiból (Ezek összegzését V.4-ben, a saját fogalom megalkotása előtt végzem majd el). Emellett a tisztázott működési módokra épp úgy épülnek a további fejezetek, mint az I. fejezet elemzéseire. Az egyetemi jegyzethez témákat kihagyni nem szándékozom, csak rövid kurzusok számára készített összefoglalókban.

Kisebb eredményeket is elértem: elsősorban az MIKT és pszeudo-tanulás fogalmak bevezetésével, ezekkel a kifejezéseket a következő fejezetekben rendszeren használom. Az MIKT-t alkotó technológiákra és azok tulajdonságaira is számos visszautalás lesz, a tanulómodellek fejlődésének vázolója pedig műszaki szempontból is megalapozza a későbbi vizsgálatokat. Legérdekesebb talán a 3. alfejezet, mely nem csupán a várható fejlődés vizsgálatához (III.) ad háttérrel a technológiai ismertetésekkel, hanem még az autonómia kérdésköréhez is eredeti módon fog hozzájárulni (II.3.3-ra épül IV.1.4), a nyelvi modellekre való utalás pedig sok helyen fog felbukkanni.

Az anyag összeállítása a későbbi fejezetek után történt, ezért csak a szükséges mélységben merültem el egy-egy témában, még ha néha túl részletesnek tűnik is. Épp, hogy elégségesnek tartom ezeket a leírásokat, ahhoz, hogy a terület sajátos szemléletét, a lényegét is meg akartam értetni. Ez a szemlélet-közlési aspektus sok szempontból kiemelendő. Elsősorban, mert ezekre a megközelítésekre is építeni szeretnék, nem csupán a közölt információkra. Ez segíti az anyag jegyzetté alakítását is, továbbá kevésbé avul el a tartalom a hordozó információk elavulásakor. Sőt ez a kutatói megközelítés válik elvárttá később, a vertikális tanulás modell alapján (III.2.2.).

MÁSODIK RÉSZ: SPECIFIKUS VIZSGÁLATOK

Az Első rész áttekintésére építve számtalan vizsgálati lehetőség adódik, ezek közül a bevezetőben megjelölt témakörökben folytatok alább vizsgálatokat a várt eredmények irányába.

III. AZ MI NYÁRBAN NYARALHAT-E AZ EMBER?

Ez az első vizsgálat a C1 cél alapján az MI fejlődési irányait, annak gyorsulását vagy lassulását hivatott vizsgálni (K1), és az előfeltételezés (H1) szerint a lassuló tendenciára keres bizonyítékokat. Ehhez az MI technológia várható fejlődését a társadalmi és emberi kölcsönhatásainak perspektívájából vizsgálom. Hiszen bár tagadhatatlan, hogy az utóbbi években, az MI fejlődése egyre gyorsul, gyakran hetente jelentenek be valamilyen újdonságot, azonban a józan ész szerint is kérdéses, hogy ez a tempó meddig tartható.

Én 2019-ben találkoztam és kezdtem el foglalkozni a problémával, azóta több a szakirodalom a témában, így kevésbé eredeti a jövő szkeptikus megközelítése, hiszen sokan jósolják, hogy a mostani „technológiai tavaszt”, egy „MI tél” fogja követni, ahol az MI-fejlesztések lelassulnak vagy megállnak. A fejezetben az egyik rész cél, hogy listázzam egy ilyen forgatókönyv szakirodalomban megjelenő okait. Másik rész cél, hogy ezt a listát ki tudjam egészíteni olyan tényezőkkel, melyeket nem találtam a szakirodalomban, lehetőleg sajátos megközelítéssel – ezt második rész célként is kijelöltem. Harmadik rész célként a kapott eredményekre támaszkodva vázolom azt is, hogy hogyan hathat az MI fejlődésének itt vázolt lassuló forgatókönyve a védelmi és a katonai területekre.

III.1. AZ MI TELEI ÉS TAVASZAI

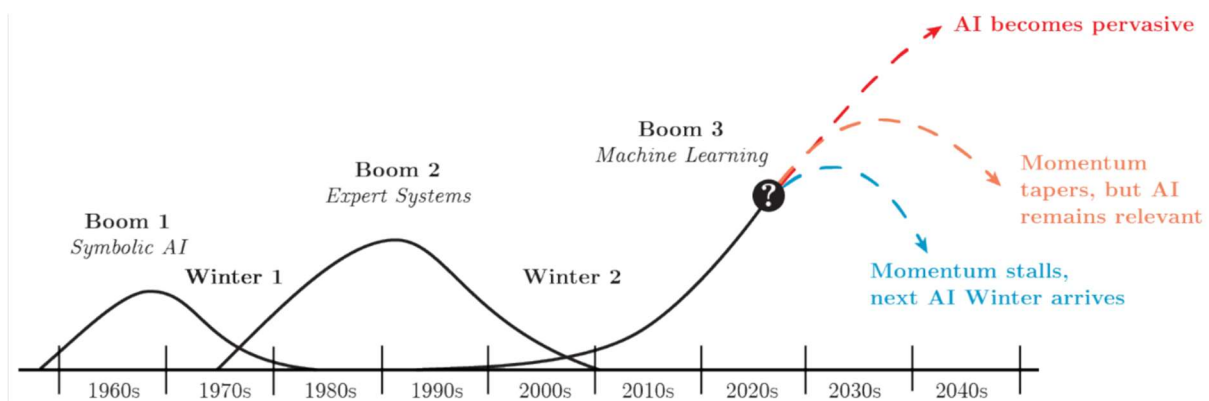
Az MI tél (*AI winter*) kifejezést azokra az időszakokra szokás érteni, amikor az MI fejlődése lelassul. Egy-egy technológia fejlődését általában az adott területen folyó fejlesztésekre szánt összegekkel mérik [135], tehát a „tél” egy szókép a kutatási források befagyasztására, a „tavasz” pedig a befektetések megugrásának allegóriája. Jelen elemzés azonban más módon közelít, csupán érinteni kívánja a visszafogó gazdasági tényezőket, vagyis alább nem egy befektetőknél szánt prognózis olvasható (bár eredmények számukra is hasznosak lehetnek).

III.1.1. A kutatás sajátos megközelítése

A jelenleg tapasztalható MI fejlődés folytatásának lehetséges forgatókönyve többféle lehet, amint ezt mutatja az 1. sz. ábrán (forrás: [136, o. 2]) bemutatott három irány (az ábra jobb oldalán a szaggatott nyilak). Sok olyan jóslat is olvasható, mely az MI fejlődésének jelenlegi ütemét

figyelembe véve egy következő fejlődési lépcsőfokok elérését prognosztizálja (pl. [29]). Más vélemények szerint a jelenlegi fejlődési tempó ilyen egyszerű meghosszabbítása nem ad megfelelő prognózist: inkább valamilyen mértékű visszaesésre lehet számítani. Jelen tanulmány ez utóbbi a megközelítésekre fókuszál. (Az ábrán látható két hullámvölgyről a következő kisfejezet szól.)

Jelen kutatás egyik célja olyan faktorok beazonosítása és vázolása, melyek kevésbé közismertek. A lehetséges lassulás okainak feltérképezésekor több ilyen területet találtam, ezek közül itt egyre tudok alaposabban kitérni (III.5.1.). Az első két MI télként aposztrofált korszakról viszonylag sokan hallottak, ezért itt elegendő ezeket röviden vázolni, de erre szükség van ahhoz, hogy rámutathassak a hasonlóságokra, és méginkább a különbségekre az akkori és a mostani kor között, melyre több következtetés is épül. A szakirodalomban megjelenő tényezők



13. ábra: A MI eddigi két téli időszak és lehetséges folytatások (forrás a szövegben)

interpretálása, valamint az általam feltárt faktorok ismertetése után az olvasó képet kap, hogy mi minden árnyékolja be az MI jelenlegi csillogását.

Az így kapott eredményeket védelmi szempontból is elemzem, és másodlagos célként felvetéseket fogalmazok meg arról, hogy ezt a fejlődési lassulást hogyan lehet esetleg kiaknázni, illetve milyen kockázatokat vet fel védelmi szempontból. Mindemellett harmadlagos célként az itt bevezetett új terminológiákat úgy interpretálom, hogy azok általános használatára is lehetőség nyíljon. A technikai AGI¹¹⁵ és emberies AGI megkülönböztetése (ld. III.6.1.) sokrétűen felhasználható, a horizontális és vertikális fejlődés leírása számos társadalomtudományi területen alkalmazható, horizontális és vertikális tanulás nem csupán a pedagógiában, de a szociológiában, filozófiában vagy jövőkutatásban hasznosítható, míg a direkt és adaptált katonai MI megkülönböztetése a hadtudományi kutatások leírásait pontosíthatja.

¹¹⁵ AGI = Artificial General Intelligence

A szakirodalmat objektíven próbáltam interpretálni, nem pedig egy lehetséges MI tél hipotézishez igazítva, mint bizonyos szerzők, akik válogatták, azaz torzították az információt (néhány éve megfogalmazott aggályaikra, már azok megírásakor születtek megoldások).¹¹⁶ Az alábbi összegzés nem is támaszt alá olyan jellegű korszakot, mint az első két télnek hívott időszak. Ehelyett azt a meglátást próbálom az itt összeállított anyaggal alátámasztani, hogy bár a *fejlesztés* újszerűsége csökkenhet, de nem következik be látványos visszaesés, mert az MI *terjedésének* üteme még sokáig növekvő tendenciát fog mutatni. A keretek betartása érdekében sajnos a sok tényezőt csak vázlatosan említhetek, és ugyanezért például a lassulás során komoly tényezőként felmerülő jogi területre sem térhetek ki.

III.1.2. A két tél dióhéjban

Itt térek vissza az 1. sz. ábrára, amely az eddigi MI teleket (hullámvölgyeket) és felíveléseket is megjeleníti. A fejezet során a fő cél, hogy ezek okait beazonosítsam és értékeljem. Ehhez a visszatekintő elemzés módszerét alkalmazom, melynek nagy előnye, hogy mivel képes felülről rálátni a korszakra (több információ birtokában, mint a kortársak), ezért a kutató új összefüggéseket azonosíthat. A történeti kontextust elegendő pár sorban vázolni, hiszen sok jó publikáció jelent meg az MI történeti részleteiről. Megjegyzendő, hogy a telek számát és időpontját nem azonosan határozzák meg a kutatók, jelen alfejezet dátumaihoz itt Toosi és kutatótársai által összeállított esszét választottam ki.¹¹⁷

Az 1943-ban megjelent elméleti neuronmodell alapján már 1950-ben működő MI gép készült, 1956-ban pedig nevet is kapott az MI technológia. Ekkor még óriási bizakodás övezte, és jelentős anyagi támogatása révén egy ideig született is számos komoly eredmény. Azonban már az 1960-as évek végétől megindult egyfajta csalódás, mely a pénzforrások jelentős megvonásához vezetett. Ez az első MI tél az 1980-as évekig tartott. (Más kutató 1973-1979 közé datálja.[137, o. 6])

A fősodor befagyása azonban nem jelenti azt, hogy közben semmi nem történik: így a „hó alatt ébredő magok” módjára bújtt elő az a paradigmaváltás, mely az MI következő felíveléséhez vezetett. Az új szemlélet lényege, hogy általános megoldások keresése helyett specifikus problémák adatainak feldolgozására koncentrált. Az első tél kezdetével egy időben (1970 körül) már működött egy molekulaszervezetre következtetni képes rendszer a Stanford egyetemen:[139]

¹¹⁶ Például a mélytanulással kapcsolatban megfogalmazott aggályokhoz már akkor megjelentek a megoldások is, ráadásul ezek egy része a publikációhoz hivatkozott szakirodalomban vetődött fel. Ettől függetlenül, más érdekeik miatt a forrást nem vettem el. [137]

¹¹⁷ A részletesebben érdeklődők számára ajánlható tanulmány, mely sok adatot év nevet is tartalmaz, [138, o. 11]

később ezen a fejlesztési vonalon jöttek létre az első „szakértői rendszerek”, melyek a második felívelés motorjává váltak. Az 1982-ben piacra kerülő R1 szakértői rendszer óriási sikert aratott, valamint komoly bevételt termelt mind a használóinak mind pedig fejlesztők részére, így sok ágazatot fellendített a szoftverfejlesztésről a robot-iparig. Meg kell azonban jegyezni, hogy több kutató a szakértői rendszereket nem sorolja a mesterséges intelligenciákhoz, mivel kívülről nem tanulnak, inkább szabályalapúak.¹¹⁸ Azonban a szakirodalom nagyobbik része a „*második MI tél*”-ként utal a korszakra, és mivel a technológiai visszaesés vizsgálatához hasznos szempontokat ad ennek elemzése, továbbá nem akartam elveszni ezen probléma részleteiben, ezért a megközelítésemben az elterjedt álláspontot vettem át.

Abban mindenki egyet ért, hogy az 1990-es években a szakértői rendszerek felívelése abba maradt. Mivel a sokak által várt továbblépés nem sikerült, az évtized közepére az érdeklődés a töredékére apadt, az évtized végére pedig ez a piac összeomlott. (Más kutató ezt a második telet 1988-1994 közé teszi.[137, o. 6.]). Elmondható, hogy az MI kutatása a XX. században csak bizonyos (igencsak korlátozott) területeken ért el nagy sikereket, amelyeket azonnal kudarcok követettek, amint valamilyen általánosabb cél elérését célozták meg – pedig ezek a célok a kezdeti sikerek alapján reálisnak tűntek.¹¹⁹

De „a hó alatt” végig zajlott a fejlődés. Már 1980 körül (a második tavasz alatt) elkészültek az első konvolúciós neuronhálók.[142] Vagyis a modern kép és videófeldolgozás alapjául szolgáló CNN modell (II.2.2.) elődje már ekkor megvalósult, de az akkori, jóval gyengébb hardvereken nyilván nem tudott kibontakozni. Csíra állapotában maradt évtizedekig, hogy aztán kivi-rágozzon.

III.1.3. Okok a szakirodalom szerint

A szakirodalomban feltárt okokat és az ezekhez szorosan kapcsolódó kommentárokat az alábbi saját rendszerbe foglaltam, így egyben a későbbi elemzéshez is jól használható.

1. A hype-jelenség (a túlzott elvárások hatása)

- a. **A hype regresszió.** Minden szakirodalom kiemeli, hogy a túlzott elvárások után bekövetkező csalódás jelensége ölt testet az MI két említett telének háttérében is. Általános jellemzője ezeknek a technológiai teleknek, hogy két főszereplőjük van: (1.) a *kutatók*, akiknek a túl optimista előrejelzéseket adják; (2.) a *döntéshozók*, akikben a túlzott remények keletkeztek, ezért támogatásról döntöttek, majd a csalódás miatt nem investáltak többet a technológiába.

¹¹⁸ Nem tekinti MI-nek pl. [140]

¹¹⁹ Ezt a meglátást, a második MI-tél kezdetekor a DARPA egyik intézetének igazgatója fogalmazta meg. Ld. [141]

- b. **Ez nem hype-ciklus.** Fontos, hogy ez a fajta hype-következmény merőben eltér a ma már jól ismert a hype-ciklust[143] típusú prognózistól. Ugyanis, mint a történeti leírásban is rámutattam, a lefagyott kutatások részben fagyottan maradtak, mindig más paradigmák menték indult be valami új MI-tavaszzal.
- c. **Jelenkori óvatosság.** Viszont, ha a hype-ciklus elvét alkalmazták volna az 1950-es években, akkor óvatosabban fogalmaztak volna a kutatók és visszafogottabban bizakodtak volna a döntéshozók – ahogyan ma már tapintható ez az óvatos hozzáállás az új technológiák hivatalos hátterében.

2. Hardver probléma több oldala

- a. **A hardverhiány, és adathiány mint ok.** A szakértői rendszerek kudarcában szerepet játszott, hogy a hardver-gyártók nem tudtak megfelelő alapokat adni hozzá.[138] Visszatekintve az a tény, hogy az MI elméleti robbanása csak a kétezres években, gyakorlati terjedése pedig csak a 2010-es években történt meg egyértelművé teszi, hogy a XX. századi hardverek és tanítóadatok valóban elégtelenek voltak egy MI-hez.
- b. **A hardverhiány-hit, mint mentség.** A hardver tényezőnél maradva, fontos kiemelni, hogy sokszor *hitték* azt, hogy a felállított modell jó, csupán minden paraméterében erősebb számítógép kellene hozzá. De nem így volt. Ennek tanulsága, hogy attól még nem igaz egy modell, hogy matematikailag nem tudjuk igazolni a hibáit. Amikor egy hipotézis mögé érzelmek kerülnek, akkor az könnyen fordul át az elhívés állapotába, mely nem legitim és tudományos közelítés.
- c. **A jónak hitt modell bizonyítható hiányosságai.** Az első tétel egyik kiváltó oka volt, hogy matematikailag is sikerült rávilágítani a megközelítés hibájára, hiszen a Perceptron[70] elméletileg sem képes megvalósítani a XOR függvényt.[144]
- d. **A tanító adatok hiányába vetett hit.** Egy-egy tanító módszer lehet működőképes, de illegitim kiterjesztés volt azt gondolni, hogy elegendő csupán több adat ahhoz, hogy ezek általános problémákat oldjanak meg. Észrevehető lett volna a hipotézis hibája, szembenéznek azzal, hogy ezek a modellek mennyire leegyszerűsítették a valóságot. Hiszen egy szimplifikáció mindig kompromisszumokat hoz, eltekintve a dolgoktól az jobb kezelhetőség érdekében – ezáltal alkalmatlan arra ezt általánosítva visszaérkezzünk a valósághoz, vagyis illegitim kiterjesztés az egyszerűsítés általánosíthatóságát feltételezni. (vö. 3.b.). Másképpen fogalmazva: számos akkori modell mai fejlett számítógépeken sem teljesítene jól, hiába kapna óriási adathalmazt.

3. Helytelen atropológiából adódó elvi félreértések:

- a. **A Moravec-paradoxon hitvilága.** A korszellem leegyszerűsítő tévedésére jól világít rá ez a vélekedés. Ennek lényege, hogy egy bogár „buta” agyát, vagy egy fejletlen gyermek képességeit könnyebb gépileg másolni, mint egy felnőtt sakkozó okosságát.[145] Ezt a vélekedést tudományos igazolás nélkül fogadta el szinte mindenki, hiszen a fejlődés logikája ezt diktálta. Pedig illegitim kiterjesztés a gépvilágra alkalmazni a törzsfejlődés és az egyedfejlődés, illetve a tanuláselméleti

szintek alakulásának folyamatát. Furcsa, hogy nem gondoltak bele, hogy a gépek nem egy DNS információhalmazzal jöttek létre, és fejlődésük iránya is tökéletesen más motiváción alapul, mint az élőlények életösztöne – és ezt a cáfolatot akkor is megfogalmazhatták volna. Tehát ez a vélekedés nem valódi tudományos hipotézis volt, hanem egy átgondolatlan elhívés. Ez a jelenség szempontunkból is igen tanulságos, főleg ha a három okra is rámutatok, melyre visszavezethető: (1.) az akkori túlbizakodó racionális és progresszivista (a fejlődésben hívő) korszellem, (2.) egy nyáj-effektus, mely miatt lángelmék követték egymást ebben a tévedésben, (3.) a filozófia és humántudományok csak névleg váltak a kognitív tudományok részévé, valójában a fejlesztők keveset foglalkoztak ezekkel (ez ma sincs másképp).

- b. **A megközelítő paradigma szélsőségei.** Az első MI-tavaszi esetében az emberi gondolkodást akarták lemásolni,[146] általánosan akarták megoldani a problémákat. Ez vezetett az első télhez, és ezt a túlzó célt oldotta fel a második tavasz alulról építkező megközelítése, mely a feladat elemzésére alapult.[138, o. 9] Így a második téli esetében az ellentétes végletbe estek. Ugyanis a működő feladatmegoldó gépek alapján azt feltételezték, hogy ez egyre jobban általánosítható, azaz az emberi intelligencia formalizálható és rekonstruálható pl. „ha-akkor” szabályok kellő mennyisége által.[140]

III.2. A KORSZAKOK ELTÉRÉSEI

A fenti történeti és okozati vázlat megalapozta, hogy a jelenlegi helyzet és a korábbi két tél közötti fontosabb eltérésekre rávilágítsak (hiszen a jelenlegi helyzetet az olvasó ismerheti).

III.2.1. Más korszak más fejlődés

Mint a bevezetőben is említettem, az MI telek meghatározásához a fejlesztésére szánt összegek visszaesését szokás alapul venni. Kétségtelen, hogy ez a mérési mutató jól alkalmazható volt az első két MI téli esetén. S bár a közelmúltban megjelent publikációk is ezt a logikát követik, véleményem szerint napjainkra ennél több tényező szükséges annak megállapítására, hogy igazi visszaesés van-e az MI-ben.

Ugyanis látványos különbségek vannak az akkori, még gyerekcipőben járó technológia, és a között a sokrétű és szerteágazó iparág között, ami a mai MI. Az egykori MI gyakorlatilag zárt rendszerben fejlődött: akkor még laborok mélyén alakult, és még technikai tudományok szakembereinek is csupán elenyésző része értette mi is zajlik ott. Véleményem szerint egy illegitim kiterjesztés a zárt rendszerben létrejövő hatások leírásának alkalmazása a mai sokezer helyen működő, internet által nyílt rendszerekre. Ez utóbbiak sokszor hozzáférhető forráskóddal,

ezernyi kutató és széles tömegek tekintete által övezve fejlődnek. Úgy vélem ezzel belátható, hogy az egykori jelenségek megismétlődésére nem kell számítani.

III.2.2. Horizontális és vertikális fejlődés – stagnálás gazdasági növekedés alatt

Egy gondolat kísérlettel szeretném felvezetni meglátás lényegét. Jelen pillanatban, ha varázsu-
tásra az MI minden továbbfejlesztése leállna, és csupán a mai MI modellekre tudnánk alapozni: akkor is számtalan új termék és szolgáltatás bontakozhatna ki évtizedeken át, mely ezeket a meglévő modelleket adaptálja újabb és újabb feladatokhoz, esetleg csiszol rajtuk picit.

Ezen a ponton szükséges bevezetni egy kettéválasztást a fejlődés fogalmon belül. *Horizontális fejlődés*nek nevezem az imént leírt jelenséget, amikor jelentős újító áttörés nélkül működik egy iparág akár évszázadokon át. Erre számos példát hozhatunk az irodaszerektől a betonipari termékeken át műszálas termékekig. A *vertikális fejlődés* ellenben technológiai áttörések mentén jön létre, ahogyan a könyvnyomtatás leváltja a kódexeket, a nyomtatott könyv alternatívája a digitális olvasás, míg egy következő fázisban talán az agyhullámokba vagy a szembe közvetlenül megérkezhetnek a kívánt információk. A vertikális fejlődés kifejezés hasonlít a diszrupcióhoz, de más. Egyrészt jobban kezeli a sok kisebb-nagyobb lépést, amelyek sokszor különböző területeken jönnek létre, míg valami innovatív újdonsággá állnak össze. Másrészt nem tekinthető vertikális fejlődésnek egy újfajta számítógépes tartós tár megjelenése (ami diszrupció), mivel ezek alapvetően nem módosították a számítógépek felhasználását, csak gyorsabbá tették azt.

A felhozott horizontális ipari példák is mutatják, hogy nem szükséges a nyereséghez vertikális fejlődés (bár ettől folyamatos újítások szükségesek hozzá). Vagyis technológiai áttörések nélkül is elérhetőek a növekvő bevételek. Tehát ma már lehetséges olyan stagnálás az MI technológiában is, mely alatt gazdasági szempontból ugyan fejlődik az iparág, viszont az innovációk szempontjából nincs igazi előrelépés, vagyis vertikális fejlődés nem jön létre.

Hasonlíthatjuk ezt a korszakot ahhoz, ahogyan a kőolaj alapú üzemanyaggal működő belső égésű motorok óriási bevételű iparát sem az új áttörések vitték folyamatosan tovább, hanem elsősorban az igény erősödése. Persze ugyanez az igény az utak fejlődését, az életmód átalakulását (pl. a lovak kiszorulását) is hozta, amelyek csak tovább erősítették a gépjárműgyártást. Hasonlóképpen az MI beszivárgására számítanak legtöbben az élet egyre több területébe, átalakítva számos szakma feladatkeretét, amivel egyet kell érteni.

Az MI-ben az elmúlt 20-25 évben egymást követték a nagy áttörések. E tekintetben is egyedülálló ez a technológia. Viszont közeledik az a pillanat, amikor egy újabb nagy áttörés csak egy valódi értelemben vett Általános MI (AGI¹²⁰, vagyis embertudású MI) rendszer létrehozása, vagy az ember-gép egyesülés könnyűvé tétele lenne. Ezekre sok prófécia létezik, de megjelenésük realitása jelen tanulmány alapján is megkérdőjelezhető (ld. IV.2. és V.3). Ezért nem számítok vertikális lépésre az MI-ben a belátható jövőben.

III.2.3. A hype tényező változása és vallásiassá válása

A témában megjelent minden tanulmányban kiemelkedő jelentőséggel említik a túlzott elvárásokat, azonban nemigen lelhető fel utalás arra, hogy a XX. századi és a mostani hype mind inputjában (a kiváltó szemléleti-lelki háttérben) mind pedig outputjában különbözik egymástól. A különbség három aspektusát azonosítottam:

1. **A hype-jelenség bulvárosodása.** A korszakok közötti különbség egyik fontos jellemzője, hogy napjainkban kicserélt köntösben jelenik meg a hype-jelenség. A régi telek főszereplői a kutatók és a döntéshozók voltak, míg manapság a főszereplők a tömegtájékoztatók és a híreket befogadó tömegek. A „mesterséges intelligencia” egy kattintás vadász kifejezés a bulvárújságírásban, mely tömegével ontja az információhamisító cikkeket. De a sajtó többi része sem védett: ezeknél hamisítás oka sokszor az, hogy sokezer (más területen talán jártas) szakember képzetlenül fogalmaz meg véleményt az MI-vel kapcsolatosan. Ki is mutatható, hogy MI publikációkban tendenciózusan kezd összemósodni a spekuláció az igazolható eredményekkel.[147] Az ingatag vagy hamis publikációk jó része valódi szakértelemmel felismerhető, tehát a műszaki szakembereket alkalmazó professzionális befektetők észre fogják venni. Ezért véleményünk szerint ez a fejlesztések terén nem vezet télhez – épp az egykori hypológok, a kutatók és döntéshozók révén védett ez a szegmens. Viszont, mint láttuk, a közvélemény nem védett, ráadásul a pontatlan dicsérgetések mellett a problémákat túldimenzionáló írások is okozhatnak a visszaeséseket. Ezt a félelmet jelentős kockázatúnak tartom egy MI visszaesés szempontjából is. Ugyanis egy fordított-hype (ijesztgetés) a fogyasztók túlzott óvatosságát generálhatja, vagy mozgalom szintjén „a világ védelme érdekében” a technológia radikális kerülését. Radikális megjelenésében akár az MI korlátozását vagy betiltását is eredményezheti, ha ezáltal valaki szavazatokat szerezhet.¹²¹
2. **Új hype-eszmék, de gyengébben.** Az MI a harmadik tavaszában és nyarában olyan lendületes fejlődésnek indult, mely alapján a technológia egészével kapcsolatban a legtöbben sokáig nagyon optimista véleményeket fogalmaztak meg.[148] Ezeknek jelentős lökést adott Kurzweil szingularitás elmélete,¹²² mely szerint a technológia nemsokára olyan

¹²⁰ Artificial General Intelligence, tehát pontosabb, de magyartalanabb kifejezéssel Mesterséges Általános Intelligencia.

¹²¹ Id. [145]

¹²² Ez 2006-ban jelent meg, de mára magyarul is elérhető: [31]

ember és gép egybeépítése hoz létre, mely meghaladja a homo sapiens szintjét. Ehhez egyre többen csatlakoztak, és ezt alátámasztani látszott a 2010-es években számos feltétel jó együttállása is. Technológiai oldalról egyre erősebb hardver, egyre több megfelelő tanító adat, és egyre jobb modellek jöttek létre. Más oldalakról pedig a biológia és az orvostudomány mellett humántudományok (pl. nyelvtudomány, pszichológia) olyan irányú fejlődése is támogatta a fejlődést, melyek kifejezetten gépi tanulás alá dolgoztak. Ezek miatt az évtized második feléig sok kutató értett egyet abban, hogy a szingularitás csak idő kérdése[149, o. 71] (erre a következő pontban visszatérek). Ez a hang azonban messze nem olyan meghatározó, mint az első két MI tél esetében bemutatott túlzott bizakodás. Sőt az évtized végétől egyre szaporodnak a szkeptikus jóslatok.

4. **A hype-jelenség mögötti világkép vagy hitvilág.** Az első MI tél korszaka még annak a racionális eufóriának az ideje, melyben a tudomány mindenhatósága óriási tömegek számára jelentett lelkesítő szemléleti alapot. E mögött talán a világégés irracionális borzalmának túlkompenzálása húzódik meg. A hatvanas évek végére azonban ez kissé kifulladás, az újabb mozgalmakban (beat, hippy) az érzelmek kap óriási szerepet, majd a kilencvenes években kibontakozik a posztmodernnek nevezett korszak, ahol az irracionális visszaszívárog a személyes világképekbe. Korszakunkban ezek együtt vannak jelen, de most a tudományban hívők tendenciája érdekes, mivel ez vallásosan is megjelenik – igaz: nem átütő erővel. Legjobb példa az MI-vallás, melynek a Jövő Útja (*Way of the Future*) egyháza 2015-2020 között működött.[150] De hasonló egyes mozgalmak és rajongótáborok világnézete is. Példaként lehet hozni erre a Turing Church mozgalmat,¹²³ a már említett Kurzweil elveket, vagy a híres író, Harari gondolkodásmódját. Kérdéses, hogy ilyeneknek lesz-e jelentősen nagyobb talaja a jövő generációiban, ahol eleve kevesebb az ilyen jellegű elvont témájú lelkesedés – hiszen piaci hatásaikban ezek elhanyagolhatóak.

III.2.4. A szabályozhatóság elvi háttere

Bár a jog területét nem kívántam érinteni, mégis, a társadalmi kölcsönhatások vizsgálata előtt rá kell mutatni azokra az eltérésekre, melyeket az MI a szabályozhatóság terén generált

(1.) *A szabályzás filozófiai alapjainak szükségessége.*

Sokan a XX. században kialakított szemlélettel és szabályozó mechanizmusokkal szeretnék szabályozni a jövő technológiáit.¹²⁴ Sajnos sokszor nincs idő az újragondolásra, és „tűzoltásra” ez a metódus elfogadható. Ám szeretnék arra rávilágítani, hogy a ma küszöbön álló óriási technológiai változások egy ma még nem látható mértékű paradigmaváltást fognak igényelni. Például, hogy hogyan valósítható meg normáink érvényesítése, azt szükséges, de nem elégséges

¹²³ Ennek neve a szójáték a Church-Turing matematikai tézissel, ezért nem fordítottuk le. Rövid összefoglaló a vezetőjétől: [151]

¹²⁴ pl. [152] – Ebben az alább tárgyalt Lessing-féle 2000 előtti felosztást alkalmazzák a robotikára, melyet a szerző kiberjog szabályzásához állított össze, jóval az MI elterjedése előtt.

azon a szinten vizsgálni, hogy mennyiben kell erős vagy puha szabályozással¹²⁵ próbálkozni. Ez kicsit olyan, mintha a XVIII. századi megközelítésekkel próbáltuk volna kezelni és szabályozni a XX. század technológiáit.

Fontos lenne először a háttér-problémákat a fent említett filozófiai elvonatkoztatás síkján vizsgálni, és ezek valamelyikéből kellene kiindulni, nem a hatályos szabályozókból. A következő trilemmát kellene feloldanunk: (I) Vagy az elterjedt társadalmi normákat kell a technológiákra erőltetni; (II) vagy elfogadva és követve a technológiák okozta változásokat, normákon kell változtatni, és azt a társadalomra erőltetni; (III) vagy a kettőt vegyesen használó megoldás optimumát keressük (és hogy milyen arányban szükséges ezek vegyítése). Az (I) esetben egy adott technológiát gyakran visszafognak a társadalmi elvárások és korlátok: a szabályozásoknak megfelelés pl. drágítja őket (pl. az autókat), vagy a fejlesztést nehezíti (pl. GDPR az MI-hez szükséges adatokat). A (II)-re példa lehet az agresszió fentebb említett megváltoztatása sok társadalomban. A (III)-ra példa az internet, melynek egyes következményeit szabályozták, más következményei pedig az embereket változtatták meg.

A probléma, hogy arra nincs cizellált filozófiai támpont, de még etikett sem, hogy mikor melyiket miért kell favorizálni. Megállapítható tehát, hogy sokkal stabilabb filozófia alap lesz-szükséges a társadalom oldaláról jövő vektorok mélyebb vizsgálatához. Az MI esetében mindez hatványozottan jelentkezik, hiszen egyrészt korábban fel sem merülő filozófiai kérdések özönét veti fel a technológia, másrészt jelenleg is eltérő etikai háttérrel fejlesztik a különböző területeken, ami előnyhöz juttathatja az a felet, aki kevésbé szabályozza (ahol nincs GDPR).

(2.) A vizsgálati szempontok bővítendőik

A fentiekre példaként egy régebbi, jó és működő modellt egészítettem ki. Az eredeti megközelítésben a szabályozhatóság vizsgálatát (1) a törvény, (2) a normák, (3) a piac és (4) a rendszer architektúrájának négyesfogata határozza meg,^[153] melyekhez legalább a következő tényezőket szükséges lenne hozzátenni.

- (5) Az alkotó: Saját stílusa, személyisége, (csoport-) szemlélete ugyan csak járulékosan jelenik meg más technológiában, ám az MI esetében meghatározó lehet;
- (6) A szemlélet: az (5) pontban említett személyiségből levezethető sajátos filozófia, amely az egyéniségtől elvonatkoztatva képez egy gondolkodásmódot és határoz meg irányokat, kereteket;

¹²⁵ A puha szabályzás lényege, hogy próbálja nem visszafogni a kutatásokat. [152, o. 35–6]

(7) Az állam: jelen vizsgálatban ennek védelmi aspektusa a fontos. Itt más etikai és gazdasági keretek érvényesülnek, mint a civil szférában, teljesen más viszonyban áll az állam a társadalmi normákkal, továbbá beleszólhat a technológiák fejlesztésébe és használatába.

(+) Az összefüggés: minden vizsgált területet szeparálva, de egymással összefüggésben is kezelni szükséges. Ezért fektetek alább nagy hangsúlyt például arra a visszahatásra, mely a társadalom részéről irányul a technológiák felé.

Ez utóbbi szempont azért is jelentős, mivel jelen nemzetközi helyzetből (versenyből), egy magát haladónak tételező ország nem maradhat ki. Vagyis legalább alkalmazni szükséges az MIKT rendszereket minden állam saját apparátusában, támogatni vállalatait és polgárait ennek alkalmazásába való bekapcsolódásban. A saját apparátuson belül kiemelkedő figyelmet szükséges fordítani a védelmi szférában való alkalmazásra, mivel e nélkül más államok erő-érvényesítésével szemben óriási hátrányba kerülhet az ebből kimaradó állam.

III.3. AZ MI ÉS TÁRSADALOM KÖLCSÖNHATÁSAI ÉS VISSZAHATÁSAI

Az alfejezet célja összesíteni az MI azon visszahatásait, melyekre minden a társadalomban meg kell találni a megfelelő válaszokat. Ehhez szükséges a társadalmi háttér hatását is megvizsgálni az MI fejlesztésekre, hiszen az oda-vissza hatás dinamikus egységben van. Erre a vizsgálatra nagyban támaszkodik majd az etikáról szóló elemzés (IV.2.)

III.3.1. A technológiára ható társadalmi eredő-vektorok

Először tehát azokra a visszahatásokra koncentrálok, melyek a társadalom irányából hatnak a technológiákra. Gyakrabban vizsgált terület, hogy a technológiai változások hogyan hatnak a társadalomra, milyen új kérdésekkel szembesítik újra és újra az emberiséget, hogyan vezetnek forradalmi változásokhoz, az életmód átalakulásához. Pedig fontos az alábbi irány elemzése is, hiszen csak bizonyos optimális társadalmi viszonyok mellett jöhetett létre egy-egy technológiai fejlődési lépés, vagy egy technológia terjedése.

(1.) A kölcsönhatás mozgatórugói

A társadalom és technológia kölcsönhatásának mechanizmusa láthatóvá tehető közismert történeti tények sajátos megvilágításában, ezért először ugorjunk vissza az időben. Sokáig nem volt általános igény új technológiákra,¹²⁶ főleg a széles tömegek felől nem volt. Nagyjából a második ipari forradalom után kezd egyre markánsabbá válni, hogy a fejlesztések eredményeit nem pusztán elfogadják, bevezetik, hanem egyre szélesebb rétegek várják, sőt elvárják. Ezzel

¹²⁶ Megjegyzem: korábbi korszakokban a különböző birodalmak felemelkedése és bukása is nagyban összefüggött az általuk használt technológiák és riválisaik számára rendelkezésre álló technológiák közötti különbséggel.

az igénnyel kezdett erősödni a társadalom felől a technológiák irányába mutató hatásvektor. A technológiák irányából a társadalom irányába mutató hatások eközben fokozatosan erősödtek. Ennek a két iránynak az egymásra hatásában létrejöttek olyan pontok, ahol „a dolgok nem mehetnek tovább, úgy, mint addig”. A technológiák megváltoztatták a társadalmat, a társadalom pedig újabb és újabb technológiákat igényelt, ami néha lassabb, de sokszor forradalmi sebességgel következett be.

A folyamat öngenerálónak vált, és azóta is egyre gyorsul. A most vizsgált irányt nézve például alig kétszáz éve, hogy Széchenyi István még a gazdagok közül is csak keveseket tudott az akkori újdonságok irányába megnyerni. Ma pedig már mindenki, még a „maradiak” is olyan újabb gépeket vesznek, amely megkönnyíti munkájukat, növeli kényelmet, a szórakozási élményt stb. A társadalom felől érkező vektorok eredője tehát nem egyszerűen a kereslet az elért újításokra, hanem igény merül fel a további újításra, a változásra. Ez a régi korokban kevés emberre volt jellemző. Bár társadalmi igény, vagyis kereslet inkább a meglévő technológiák tökéletesítése felé mutat. Nagy diszruptív ugrásokra sokáig csak az adott szakterület tudósai vágytak – manapság azonban a felhasználók közül is sokan szeretnének technikai alapú szupererők birtokába jutni. A kereslet-kínálat szabálya egyébként a kínálat oldaláról sem mindig működik, hiszen az egyes újításokra rákényszerülnek az emberek (pl. a hardverek avulása miatt) vagy olyan termékekre beszélnek rá, melyekre nem lenne szükségük – de a generált kínálat hatásvektora, és az MI-re kényszerítés ilyen módja túlmutat jelen elemzésen.

(2.) A társadalom felől érkező hatások elvekké válása

Érdekes eredményekre jutnak, akik azt kutatják, hogy egy szervezet döntési mechanizmusába épülő MI hogyan befolyásolja az embereket [154], témánk általánosabb perspektívájában azonban itt nem mehetek le a cégek szintjéig, a társadalom szintjén maradván vizsgálom a visszahatásokat. A társadalom felől érkező vektorok általában először praktikus elvárásként merülnek fel: pl. veszélyes a technológia. Erre az állam vagy a vállalatok szabályokkal reagálnak (pl. használati utasítás, betiltás). Majd a konkrét probléma miatti szabályokat általánosítani szükséges egyre szélesebb termékkörre. Az általánosítás egyes vonalain elkezd elválni a tárgyi világtól: az egyik, jogi vonalon törvényekké transzformálódik (erre rögtön visszatérek), más vonalakon pedig akár filozófiai elvonatkoztatásokká, elvekké, életszemléletté.¹²⁷

Az általánosításhoz a filozófia és a pontos megfogalmazások révén juthatunk. Így ragadhatjuk meg a dolgokat úgy, hogy egy szabályozási elv pl. alkalmazható legyen még nem létező

¹²⁷ Példaként a környezetvédelmet vehetjük, amely egy technológiai (szennyező anyag) szabályozástól jelentős politikai tényezővé vált, mára pedig sokrétű világnézeti mozgalommá volt képes alakulni.

technológiára is. A filozófiának ez az oldala folyamatosan adaptálni próbálja az addig érvényes morális normákat a változó világhoz: vagy a világ megváltoztatását, vagy a morális normák változásának szükségességét javasolva. És hasonlóan fontossá válhat a filozófia azon másik fele is, melyben az elvonatkoztatás idővel elszakad a praktikus kiindulási pontjától és önálló mozgalmi életet kezd élni. Ezek a mozgalmak is vektorként jelennek meg a világban, melyek a technológiák felé (és az etika fel) is sajátos elvárásokat támasztanak.¹²⁸

(3.) A fejlődéshit és az MI fejlődés módja

A technológiák által biztosított kényelem, mobilitás és információáramlás – vagyis a fejlődés eredményei – alapjaiban változtatták meg az ezeket használó társadalmakat, azok életszemléletét és értékrendjét. A fejlődés azonban az előnyök mellett számos olyan új kihívás elé is állította az egyéneket és a társadalmakat (rohanás, individualizáció, figyelemzavaros generációk, stb.). A problémákra a válaszokat fokozatosan egyre többen a fejlődésben, új technológiákban keresték, sejtve, de elhanyagolva az újabb mellékhatásokat. Vernétől kezdve, az 1950-es sci-fi szerzőin át egyre többen az új technológiák nyújtotta megoldásokban látták az ősidők óta létező rossz eltűntetésének lehetőségét. Így a fejlődés elért egy fokot, ahonnan a fejlődést nagyban a fejlődésbe vetett hit is generálta, motiválta. Ennek a hitnek az alapdogmája, hogy az emberiség fejlődik, vagyis minden egyre jobb lesz. Bár a minőségi javulás már metafizikai kérdés, így soha nem volt igazolható kísérleti vagy matematikai alapon, mégis ezen a dogmán alapult a XIX-XX század kutatása (amely elvileg elveti az ilyen hitbeli dogmákat).

Egzakt módon változásokról kellene beszélni, hiszen a negatív mellékhatások hosszú távon erősebbek lehetnek majd a pozitívumoknál, és ezt nem tudhatjuk előre. Mert vajon visszafejlődés-e vagy fejlődés, ha visszaesik a gazdaság és emiatt kevesebb károsanyag kerül a környezetbe? Az elvont minőségek változását az MI sem tudhatja előre. Hiszen regresszió-e, ha az emberek szerényebb kényelemben élnek, de többet barátkoznak? A cél persze az, hogy problémás tendenciák úgy forduljanak meg, hogy közben jelenlegi életünk előnyei megmaradnak – de talán épp az MI ébreszti majd rá az emberiséget, hogy „mindennek ára van”.

Amint azt fentebb említettem (III.2.3.), ma is sokan várják például az AGI vagy ASI rendszerektől, hogy jobbra teszik a világot. Ezt vallásos kérdéssé tenni és egyházat alapítani rá feltehetőleg nagyban gazdasági érdek is volt, de erre nincs is szükség. A legtöbben ilyen szervezeti keretek nélkül hisznek abban, hogy a tudomány általános megoldást ad majd olyan problémákra, mint a szegénység, az éhezés, a háborúk és hasonló rossz dolgok. Ennek realitását (a

¹²⁸ Például a gender szemlélethez az orvostudomány által létrehozott lehetőségek adtak talajt, de már régen nem orvosi szempontból vetődik fel a kérdés. Ez a példa rávilágít technikától jövő vektorok az újszerű etikai következményeire is.

fentebbi felvetéseket) kevesen gondolják át, a fejlődésben hívő tudósok a szakterületük szerint konkrét problémák megoldására koncentrálnak. Művük távolabbi vagy tágabb értelemben vett mellékhatásait épp úgy negligálják, mint a régiek. De nem is nagyon van esély a mai tudomány komplex rendszerében ilyen tágabb vizsgálatokra – hacsak épp nem az MI által. Ebből a perspektívából az MI olyan gyógyszer, amely számos egyéb másik gyógyszer mellékhatásait csökkentheti. Ez a hasonlat jól rávilágít arra a jelenségre, hogy egyre kevésbé van mód ezen a téren középút megtalálására, és az öngeneráló, az emberiséget sokban determináló fejlődésnek valójában nemigen van alternatívája (a technológia teljes elvetése a másik pólus). Sokan tiltakoznak egy-egy mozzanat ellen (aktuális probléma volt pl. hazánkban a digitális állampolgárság elleni lázongás), de ilyen rész-tagadások önellentmondások sorát tartalmazták. Ez átvezet minket az MI társadalomra gyakorolt hatásaihoz a következő elfejezethez.

III.3.2. Az MI felől a társadalomra ható vektorok

Ez az irány ismertebb, mint az előző, a köztudott hatásokat itt legfeljebb csak érinteni érdemes. A technológia oldaláról a társadalom felé érkező vektorok közül itt csak néhány, a tanulmány egésze szempontjából is fontosabbra térek ki.

(1.) A lelki, pszichológiai visszahatások.

Fentebb, az antropológiai keret tágításáról szóló részben említettem azt a visszahatást, hogy mára egyre kevésbé van szükség az emberi tudat mélységeire. Az egyre felszínesebb ember tudata pedig egyre kevésbé különbözik a lehetséges gépi tudattól. Ez a fájdalmas igazság az MI-val kapcsolatos problémák talán egyik legsúlyosabb etikai következménye. Ugyanis az objektivitást hajszoló világunk számára a megnyilvánulások válnak egyre lényegesebbé. A megnyilvánulásokban viszont a gépek egyre több emberi vonást képesek emulálni, sőt egyelőre úgy tűnik elvileg bármit képesek lesznek élethűen utánozni. És mintha egyre kevésbé érdekelné az embereket, hogy az utánzás mögött mi áll, megelégszenek a megnyilvánulással. Napjainkban sokan még az állatok megnyilvánulásait favorizálják az emberi megnyilvánulásokkal szemben. Ám a gépek lassan képesek lekörözni az állatokat és embertársakat. Már olyan életterületeken is, melyek korunkig emberi mivoltunk megnyilvánulásainak tartottunk. Az MI nem csupán optimálisabb munkatárs vagy művész lesz, de nemsokára jobb szexuális szolgáltatások, nem piszkító emberibb háziállatok stb. technológiájában lesz piacon az MI – mindezek pedig várhatóan az emberi kapcsolatok radikális átalakulásához vezetnek. Ezáltal újabb etikai normák kérdőjeleződnek meg, melyek a felnövekvő generációkban alapvetően eltérő szemléletet eredményezhetnek. Váratlan mértékben módosulhat a jövő nemzedékek világhoz való viszonya, mint elméleti

mind pedig gyakorlati szempontból. Ezzel a mostani prediktív elemzések érvényüket veszthetik, annyira kezelhetetlen, váratlan helyzetek jöhetnek létre – főleg, amikor a vezetői szerepeket majd ez a generáció veszi át. Emellett már jelenleg kimutatható, hogy az emberek durvábban bánnak például chatbotokkal, mint az emberekkel,¹²⁹ ami nem csupán nehezíti a gépi etikusság fejlesztését, de a tendencia vége is riasztó.

(2.) Új etika a gépi dimenziók által?

Az etika nem csak a generációs alakulás miatt változhat. Teljesen más perspektívát vet fel, hogy eddigi korok etikai vitái azért sem voltak képesek nyugvópontra jutni, mivel értelmetlen volt a „mi lett volna ha” kérdése. Nem lehetett eldönteni jól döntött-e valaki, vagy a kisebb rosszat választotta-e. Még inkább ez volt a helyzet a „mi lesz legközelebb hasonló helyzetben” kérdéssel, az ember nem tudott jósolni. Viszont az MI alapú szimulációkkal a gép egyszerűbb esetekre megmondja „mi lett volna ha”, vagy „mi lesz ha”. Sőt megvan a lehetősége, hogy egy gépet akár többféle szemléleti üzemmódban használhasson az ember akár prognózishoz, akár egy döntés után megnézhetné a felhasználó, hogy más szemlélet alapján vajon hogyan döntött volna az mesterséges elme – ez bizonyos emberi gondolkodásmódok háttérbe szorulásához is vezethet.

Ezzel elértünk ezen változások árnyoldalaihoz. A vázolt folyamat nagyban leszűkítheti az etika filozófiai tudományának mai sokféleségét, pontosabban a fővonalától eltérő elveket igen kevesen fognak vallani, a mainál is kevesebben fognak etikai dilemmákkal elméletileg küzdeni. Ugyanis az MI nem kerül érzelmileg kibillent állapotba egy nehéz döntés után, inkább gyorsan segítséget kér egy erősebb hardverrel rendelkező MI társától (óriásgéptől) egy elemzés lefuttatására¹³⁰. Hasonló gépi képességek által a tisztelet is átalakulhat, mivel a gép a tudás etalonjaként is kitűnhet az emberekkel szemben, az emberek egy része egyre több döntést hagyhat az MI-re. A döntés negatív oldala sem fogja lelkiismeretét terhelni, ha hisz abban, hogy a gép a lehető leghatékonyabb döntést hozta, tehát ez a folyamat negatívan hathat az *emberek* erkölcsi érzékének fejlődésére.

¹²⁹ Hiába képesek a botok az embereknél jobb kapcsolatot kialakítani, amikor kiderül, hogy az illető nem emberrel chatel a viszonyulás leromlik. [155]

¹³⁰ Erre már élő példa Tesla önvezetés, melynek első verziói hibásan működtek, viszont 2.0 verziójánál sok egy felhővel kapcsolatban maradván kevesebbet hibázik. <https://villanyautosok.hu/2020/04/27/igy-mukodik-a-tesla-tablafelismero-rendszer/>

(3.) Új hagyományok és új eszmék.

Az MI várhatóan maga alapvetően új szokásokat, kultúrákat hoz majd létre, melyek idővel a befogadók hagyományába is épülnek. Az MI által jobbat tenni kívánt világunkban azonban egyelőre nem adható felelet arra, hogy mit kell majd feláldozunk a jobb világ oltárán? Úgy javasol majd optimálisabb megoldást az MI, hogy még nincs tisztázva, hogy mi az optimalitás viszonya az emberi értékekhez? Például a mai haladók is ellenérzéssel fogadnák, ha a házikedvencek kiirtását látná megoldásnak egy MI, amikor mondjuk az elidegenedő társadalom átalakítására keresnek megoldást. Nem is beszélve olyan, már említett disztópikus eredményekről, ha az algoritmus javaslata valamire egy népírtást lenne, ami akár ideológiai alapja is lehet annak, hogy ilyen „megoldást” hajtson végre egy politikai szándék. Ha el is kerüljük az ijesztő irányokat, a dolgok alapvető újdonsága várhatóan az eddigieket elvető eszmerendszereket implicál majd, melyre csak az új generációk lesznek fogékonyak, így tovább nő a szakadék a generációk között.

A tökéletes ember a virtuális. Már egy ideje zajlik egy olyan, az MI-től független folyamat, melyben a média világa által egyre nagyobb az igény a tökéletes emberekre. A többedik fiatal generáció számára az ideált nem-létező példakép külső és belső tökéletessége adja, a szupészépek és a szuperhősök, vagyis nem-reális emberi értékek felé orientálódnak. Hiába tudják, hogy mindez csak smink, photoshop és csak film. Erre vágnak, és ezt a vágyukat teljesítik a média-gyártók – és lassan a robotika és az MI fejlesztők is... Így emberi kapcsolatok helyett egyre nagyobb tömegek találnak otthonra a virtuális térben, és egy jól sikerült metaverzumban ez fokozódhat.

Ebben a média és az MI által eltorzított tökéletességi mátrixban hol helyezük el majd értékrendünkben az emberi tökéletlenség értékeit, a lélek problémáit, vívódásait? Milyen önképe és életcélja lesz az átlagos képességű emberek tömegeinek, akik a gépekhez képest egyre kevésbé képviselnek értéket? A soha nem hibázó MI vajon degradálja az emberi hibázást, a szándék megítélését – vagy esetleg magasabb morális elvárásokat támaszt? Hova vezet, ha a szuperérzékeny gépi analízis a gondolatbűnt is felismeri: lesz-e lehetősége a személyiségeknek belső vívódással leküzdeni az ártó cselekvésre való belső csábítást? Ezeket válasz nélkül, de mint potenciális veszélyeket kell figyelembe venni – és sajnos a technológia visszahatásainak súlyos kérdéseiből még sok van.

(4.) Fenntarthatósági probléma a munkaerő terén.

Ez egy súlyos kérdés, melyet még vizsgálni fogok más oldalról is (III.5.2). Úgy tűnik nem vagyunk urai a fejlődés egyre gyorsuló örvényének. Már régóta óriási a szakember-hiány,

hiszen lehetetlen mennyiségű magas intellektusú embert igényel ez a technicizált társadalom – és ez az igény rohamosan növekszik. Nyilvánvaló, hogy az agyelszívás (*Brain Drain*), amely jelenleg próbálja ezt kordában tartani, nem fenntartható az elmaradottabb térségek fejlődésével. Hiába lesz képes az MI sok mindenre, ha kevés lesz szakembergárda karbantartani a sok segítő. Egy ilyen világban az alacsonyabb képességű emberek nagy tömegeinek nem lesz helye.

Felmerülhet a kérdés olyan oldalról is, hogy a folytonos intellektuális tanulás kényszer, nem veszi-e el az időt a valódi lelki éréstől, bölcsébbé válástól? Magyarán jó-e, ha minden ember erre áldozza energiáit? Az energia felől közelítve ezt kérdezhetjük: a technika energia-éhségének exponenciális növekedése kielégíthető maradhat-e? S mivel az emberi energia is véges, mint bemutattam, el lehet azon tűnődni, hogy vajon az jelenti-e a valódi fejlődést, ha ezt a fizikai és emberi (testi-lelki) energia-igényt kielégítjük, miközben az emberek mégis egyre magányosabbak, boldogtalanabbak és egészségtelenebbek.

Talán a fenti kérdésekből adódnak azok az érzések, ami miatt egyre kevesebben képesek vagy akarják követni a világ rohamos változását. Ezek a tömegek nem szeretnék megtanulni a túl gyorsan változó technológiákat, és ez visszafoghatja a keresletet az újítások iránt. Egyik jövőmodellként nem zárható ki, hogy egy lassulás következik be. Ez abból a szempontból is várható, hogy az emberek a technológiánál lassabban változnak. Ez alapján, mivel a döntéshozók is lassabban változnak, tehát elvont és nehéz kérdésekben nem fognak egy gép érveire hallgatni – viszont a generációváltásból adódó radikális változás melyet imént vázoltam így forradalmi sebességgel következhet be.

(5.) *Jobban megismerjük az Embert.*

Van az MI-nek számos pozitív visszahatása is. A nagyobb kényelem vagy rövidebb munkahét praktikumainál talán érdekesebb tudományos hozadék, hogy az MI fejlesztésének mellékszálain lesznek jelentős lépések. Nem várt eredményekhez adódhatnak pl. az agyhullámok terén, de akár az antropológiában is. Az MI tanítása által valószínűsíthető, hogy az ember működését is jobban megismerjük. Ahogyan már jobban megértettük az érzéseket, azok megnyilvánulásait, várható, hogy akár jobban megértjük az emberi döntések mechanizmusát, és az erkölcs tanulásának folyamatait. Ezt a tudást azután fel tudjuk használni akár emberi vonalon is: pedagógiákban, felnőtteknek szóló tréningeken vagy a lélekgyógyászatban. Akár más humán tudományok, például a filozófia is kaphat új impulzusokat a gépek tanítás-kutatásának melléktermékeként.

III.4. HIDEGFRONTOK AZ MI NYÁRBAN

Tekintsük át lista-szerűen azokat az új tényezőket, melyek által korunkban következhet be az MI fejlődésének valamilyen lassulása.

III.4.1. Kifejezések eróziója

A visszaesési forgatókönyv okai között vezető helyen van, hogy az MI fogalma, és számos hozzá kapcsolódó kifejezés (pl. tanulás, értelem, okos, intelligens) képlékeny, használata pontatlan. Ezt később részletesebben elemzem (V.1.4.), most csak röviden vázolom.

1. **Pszedo-MI.** (Saját kifejezés [156]) Fogalmilag is meg kéne különböztetni a „pszedo-MI” rendszereket, melyről a felhasználók egy részének az a benyomása, hogy háttérben gépi intelligencia dolgozik, pedig valójában egy hagyományosan programozott és relációs adatbázisra támaszkodó architektúra működteti. Így az emberekben, illetve az őket félreinformáló újságírókban, bloggerekben, influencerekben az a benyomás alakulhatott ki, hogy ez a technológia sokkal inkább elterjedt, mint valójában.
2. **MI-használat füllentése.** Az előbbi tendenciát erősíti, hogy egyes cégek, főleg start-up-ok is valótlánul állítják azt, hogy MI-t használnak.[157] E mögött az a (részben talán valós) vélekedés áll, mely szerint, aki nem használ MI-t az le van maradva, ezért támogatásoktól és vevői köröktől eshet el.
3. **Az okos szó buta termékeken.** Ezzel függ össze az *okos* szó használatának nagymértékű degradálódása is. Ez annak jegyében zajlik, hogy egyes marketingesek egy olyan mindent eladó varázsszót látnak az *okos* jelzőben, mint egykor a *modern* szóban.¹³¹
4. **MI-effektus.** A szakirodalomban létezik egy, a pszedo-MI-vel ellentétes közelítés is. E szerint amikor az MI elér egy bizonyos szintű általános használatot, akkor gyakran elkezdik nem tekinteni már intelligensnek.¹³² Ennek oka, hogy nagyon elmarad az emberi teljesítménytől, és mikor felismerik az adott gép korlátait, akkor a kezdeti varázslatos jelleg szétmállik.¹³³ Példának álljon itt egy frappáns megfogalmazás „a nagy nyelvi modellek abban jók, hogy megmondják hogyan kellene hangoznia a válasznak, ám ez különbözik attól, aminek a válasznak lennie kellene”.¹³⁴

Az MI terminushoz kapcsolódó kifejezések itt bemutatott erodálódása azért probléma, mivel árt annak a bizalomnak, amely szükséges lenne az MI eredményes napi használatához. Ezáltal a szavak zavarossá válása gátja az MI széleskörű társadalmi elterjedésének, vagyis a fejlődés piaci háttérét veszélyezteti.

¹³¹ Ennek részletezése és megoldási javaslat olvasható itt: [158]

¹³² Az eredeti írás szerint nem tekintik MI-nek, de ezzel nem értek egyet, erre az elmúlt 25 évben nem látok példát. [146]

¹³³ Arthur Clarke a brit sci-fi író írta valahol, hogy bármely kellően fejlett technológia megkülönböztethetetlen a mágiától, ám amikor valaki megérti a technológiát, a varázslat eltűnik.

¹³⁴ Saját fordítás. [23]

III.4.2. Félig leküzdött problémák: érvelés, általánosítás, öntanítás

Az MI újabb és újabb modelljei által az olyan korábbi elvi korlátok is leomlani látszhattak, mint az általánosítási, az érvelési és az öntanulási képesség. Igaz, ez a három (egyébként összefüggő) tényező igen sokat fejlődött az utóbbi időben – az eredmények azonban még csak részlegeseek. Az AGI megvalósíthatóságával kapcsolatban hasznos adalékokat ad, ha röviden megnézzük az említett három kulcstényezővel kapcsolatos helyzetet.

- **Általánosítás.** A korai megerősítéssel rendelkező modellek nagyon rosszul kezelték a megtanult séma kis megváltozását is.¹³⁵ Pedig emberi gondolkodásunk sarokkövei az egyre tágabb általánosítások. Óriási az előrelépés e tekintetben, hiszen pl. a zero-shot módszerek képesek nem-tanult dolgokat is felismerni.¹³⁶ Sőt bizonyos, eddig ismeretlen mintázatokat is felismernek. Háromféle akadályt tudtam szétválasztani e téren. (1). Ha képessé tesszük a gépünket valami nagyon általános minta-egyezés azonosítására, akkor a komoly humán erő igénye van annak, hogy ellenőrizzük, hogy ez a valóságban is létezik-e vagy sem. (2) Ezért a gép által fellelt mintázatok akkor használhatóak, ha tehát a gép egy sémát belül dolgozik, így az újdonság csak mintázat-sémákon belül új. A sémát az ember mondja meg. (3). Olyan fajta általánosításra, amit az *elvonatkoztatás* szóval írunk le gépeink nem képesek, mivel erre a matematika tudománya sem képes. Összefoglalva: a gép általánosító képessége meglehetősen korlátos (bár sok emberé is az), ezért is képtelen a megértésre.
- **Érvelés.** Olyan fajta érvelés, amely megértésen alapul, az elvileg elérhetetlen egy mostani mélytanulási modell számára. Erre az iménti általánosítással összefüggő korlátosságon túl több érvet is felhoznak a kutatók¹³⁷ (pl. hosszútávú tervezés, algoritmikus adatmanipuláció), melyet aránytalan lenne elemezni. Szemléltetésül egyet vázolok: a virtuális térbe mindig csupán egy beállított világ-részhalmoz kerül leképezésre, ahol az összefüggések is csak tanult információk, nem pedig felismerések.¹³⁸ Vagyis a gép csak adott érvelési sémát képes követni, és mivel ez túl sokféle lehet, ezért a problémát ezért manapság inkább megkerülik, mintsem megoldják. Mivel valójában nincs szükség arra, hogy a gép megértsen egy érvelést, ezért elegendő, ha olyan jól utánozza azokat, akik értő módon reagálnak. Ez a cél jobban megvalósítható, és sokszor elegendő, ahogyan egy magolós gyerek is kaphat jelest a

¹³⁵ Például egy régi számiógépes játékot ügyesen játszó egyszerű mélytanulás az ütő minimális megváltoztatása után nagyon rosszul teljesített. [159]

¹³⁶ A „kevés lövésű” rendszereket minél kevesebb információból próbálják tanítani, a nulla lövésnél a tanított infó csökken nullára. Például abból az információból, hogy a zebra egy csíkos ló, felismeri a zebrát egy lovakon tanított képfelismerő.

¹³⁷ Bővebben ld. [160, o. Section 2 of Chapter 9]

¹³⁸ Az imént említett zero-shot esetben is így van: nekünk kell megmondanunk a csíkosság és a ló összekapcsolható. Ha ő kapcsol össze a világban össze nem tartozó dolgokat, az nem hasznos.

feleletére, ha ügyesen úgy képes tenni, mintha értene is valamit abból, amit elmond. Jó példák erre az NLP rendszerek,¹³⁹ és ezek értési korlátai is viszonylag közismertek.

- **Öntanítás.** A felügyelet nélküli tanítás is sokat fejlődött, ennek motivációja, hogy sokkal kevesebb aprólékos munka szükséges hozzá, mint az adatok címkézéséhez. Cserébe az öntanító rendszerek tanítása hasonlóan nehéz módszertanokat igényel, mint a borzasztóan speciális igényű kisgyerekek oktatása. Ráadásul minden feladat-típusnál új módszer kell – és még így is csupán célrendszerek jönnek létre. Így a mai eredmények aligha vethetőek össze egy felnőtt ember önképzésével, melyben a minta-kezelésen kívül úgy tűnik sok más, egyelőre feltáratlan vagy matematikailag még nem megragadható tényező is működik. Ergo a gépeink akkor is csupán betanított munkások, ha értelmiségi emberek által végzett feladatokat képesek hibátlanul elvégezni. Ez egyébként nagyon ígéretes bár meglehetősen diszruptív tudáspotenciál. (Az előző alfejezet alapján csak reménykedhetünk abban, hogy nem nevezik majd AGI-nak az ilyen rendszereket.)

Összegezve elmondható, hogy a feni problémák feloldása nélkül a legújabb rendszerek is csupán utánzógépek. Nem bizonyítható ugyanis az a hipotézis, hogy az utánzás és az értés közötti lépcső ilyen irányban megugorható lenne. Ez is fontos részérv lesz ahhoz a megállapításhoz, hogy egyelőre nem lehet tudományosan megalapozottnak tekinteni az említett, emberies AGI-t vízionáló optimizmust (ld. III.6.2.)

III.4.3. A nyers-erő paradigma kérdőjelei

A jelenlegi MI fejlesztések a problémákat brute-force¹⁴⁰ megoldásokkal közelítik meg, miközben a „lemásolt” agy mindössze húsz watt energia felhasználásával sokkal több mindenre képes, mint egy óriási erőforrás-igényű, fejlett MI-rendszer.[162] Ez a tény önmagában is rávilágít arra, hogy mennyire megkérdőjelezhető a jelenlegi fő paradigma, mely a hatalmas adatmennyiségre és szupergépek teljesítményére épül.

Egyes kutatók szerint megfigyelhető egy párhuzam, az ilyen MI rendszerek és a gyógyszeripar egy régebbi állapota között. Az utóbbi ágazat is a nyers erő ösvényén próbált előrejutni, több-kevesebb sikerrel. Igazi lendületet azonban akkor kapott, amikor talált olyan működő új paradigmát, melyre le tudta cserélni ezt a megközelítést. Ezért jogosan merül fel, hogy az ott bekövetkezett fordulathoz hasonló, alapvető változások lennének szükségesek az MI terén

¹³⁹ Neural Processing Language (természetes nyelv feldolgozás). [161]

¹⁴⁰ A „nyers erő” nevű módszer, a kódtörésekben jelszópróbálgatást jelenti, de a kifejezést itt az MI-hez szükséges igen erőteljes technológia értelmében használjuk.

is.[162] A kutató az ehhez szükséges muníciót elsősorban az agykutatás legújabb eredményeitől várja.

Véleményem szerint azonban nem igazolható ez a bizakodás sem, hiszen az agy vélt neurális modellezése az alapja a jelenlegi, túlságosan is erőigényes (brute-force) MI modelleknek is. Vagyis a neurológiai modell fejlettebb elektronikus másolata továbbra sem garantálja a „húszwattos megoldást”.

III.4.4. A fekete doboz

Itt azokat a problémákat vizsgálom, melyek a több rétegű neuronhálók elvi velejárói. Ezeket tehát nem lehet megszüntetni, legfeljebb kezelni, de azt sem könnyű. Ezért a megfelelő kezelési módok kiforrásáig, szabályozásáig, ellenőrzéséig ezek nem csak a fejlődést, de egyben a terjedést is lassító tényezők, vagyis ez súlyosabb kérdéskör, mint az eddig vizsgáltak.

- A fekete doboz tudásának **átláthatatlansága**. Az átláthatóság (transparency) itt kétféle problémát jelöl: (1) a modell pontos működésének, válaszainak átláthatatlansága és (2) a feldolgozott adatok pontos ismeretének esetleges hiánya – hiszen mindkettőt elrejt a neuronháló. Az (1). eset a bizonytalansággal (uncertainty) függ össze, és a következő pontba került. Az adatok problémája alatt pedig azt értem, hogy az MI szolgáltatások felhasználói nem válogathatják meg a tanítóhalmazt, sőt nem is tudhatják pontosan, hogy mivel tanították be, mivel frissítik az adott gépet. Ez védelmi szempontból ez is fontos probléma. Az átláthatóságot egyes fejlesztők blokklánc alkalmazásával próbálják elérni.¹⁴¹ A javaslat jó lehet, életszerűsége azonban kérdéses.
- A fekete doboz **bizonytalansága** (kiszámíthatatlanság). A neuronhálók neuronjainak tartalma önmagában egy értelmezhetetlen számérték. Ezek megtekintése arra hasonlítana, amikor egy futtatható fájlt (egy, amúgy működő programot) szövegszerkesztőben nyitunk meg, és csak egy halom krix-kraxot látunk. A legtöbb esetben ez nem jelent gondot, mivel az eredmény számít – egymás agyába sem látunk bele, csak azt tudjuk, amit a másik tudatni akar. Úgy tűnik ez az autonóm döntések velejárója, és az emberek viselkedésében ezt meg is szoktuk. Viszont egy ember esetében sem toleráljuk, hogy pl. saját döntési szabadságát mások biztonságára vonatkozó szabályok ellen érvényesítse, főleg, ha ezzel életet veszélyeztet. Gépektől még kevésbé fogadunk el ilyen kiszámíthatatlanságot. A programozott gépek szabálykövetőek, ráadásul szabályaik, az algoritmusuk jól átlátható. Ezért akár emberéleteket bízunk automatákra, de döntésük eredményének biztossága alapkövetelménnyé

¹⁴¹ Ilyen pl. az IBM és a Casper Labs megoldása. [163]

vált. Ezt a legtöbb MI nem produkálhatja, mivel a valóságot nem pontosan ismert módon képezi le és egyszerűsíti le. Mérhető és részben kezelhető ez a gond,[164] de számos felhasználás esetében nem minősíthetőek biztonságosnak az MI statisztikai alapú, bizonytalan döntései. Ez kiemelten jelentkezik az egészségügyben,[165] vagy a fegyverek (elsősorban az autonóm fegyverrendszerek) területén, de betartandó minden védelmi szempontból is kritikus rendszernél (erőművek, veszélyes üzemek), sőt minden nagyobb rendszerben problémát okoz (pl. okosváros vagy még egy okosirodaház esetében is).

- **Megmagyarázhatóság és értelmezhetőség:** ez a fekete doboz probléma két kulcskifejezése. Ilyen problémákat a BigData háttér is generálhat, de most az MI-vel kapcsolatos részre koncentrálok. Erre ugyan vannak megoldási javaslatok, de komoly kompromisszumokat igényelnek. Néha segíthet a modell egyszerűsítése. Más esetekben a Post-hoc értelmezés kínálgatik megoldásnak, ami azt takarja, hogy a fekete-doboz ugyan megmarad, viszont betekintést próbálunk szerezni a kész modell eredményeire (tehát gyakorlatilag egy átfogó és alapos tesztelésről van szó). Egy külön MI fejlesztési trend a XAI (Explainable AI), vagyis az magyarázható MI,[166] melyre egyre jelentősebb a kereslet, annak ellenére, hogy az ilyen rendszerek határfoka lemarad a fekete-dobozal dolgozó megoldásokétól. A felsorolt megoldások miatt úgy vélem ezek jelentősen nem fogják vissza az MI fejlődését, de terjedését akadályozhatják a megoldásokkal együtt járó komoly kompromisszumok.
- **A fekete doboz egyéb hátrányosságai.** Csupán felsorolás-szerűen néhány további aggály, melyet kutatók említenek. (1) Hibakeresés: nem kívánt eredmény esetén nehéz a javítás, nem lehet a bug-ot úgy megtalálni, mint forráskód esetén. (2) Etikai aggályok: nem tudjuk, hogy egy idegen modell a világ mely etikai rendszerének szeretne megfelelni. (3) Adatmérgezés: a mélytanulási algoritmusok érzékenyek lehetnek az fals információkon keresztüli támadásokra.[167] (Erre a kiber-problémáknál visszatérek). (4) Bizalom, elszámoltathatóság, szabályozhatóság.

Bár a fenti lista talán nem is teljes körű, ám ennyi konkrétum talán elegendő annak a kitűzött célnak az igazolásához, miszerint az MI rejtélyes belseje egy fontos hátráltató tényező mind az MI fejlődésének lassulásában, mind pedig *terjedését* illetően.

III.4.5. Elfogultság

Az MI rendszer elfogultsága elsősorban a világot egyszerűsítő jellegéből következik. Létrejöttének okai között egyaránt szerepelhetnek a modell tervezési hibái, a tanító módszer elégtelenségei, az adathalmaz kiegyenlítetlensége, vagy akár a létrehozó személyek szándékos vagy

tudattalan elfogultsága.¹⁴² Mivel a feketedoboz-jelleg miatt utólag felismerni sem könnyű az elfogultságot, elhárítani pedig sokszor csak újra-készítéssel lehet. Ezért ezt a hibatípust eredetében próbálják meg kezelni, vagyis igyekeznek erre tekintettel tervezni a modellt és a tanítási metódust, valamint a félrevezetéstől mentesített adatokat használni.

Ez utóbbi azonban nehéz és aggályos: pl., ha az internet friss adatait is figyelni szeretnénk, akkor egy gyakorlatilag pozitív cenzúra alá kellene vetni annak felhasznált tartalmait. Ez sok esetben megfelelő kompromisszum lehet, de nehéz megteremteni annak háttérét, hogy ne éljenek vissza ezzel a válogatást végzők, ahogyan pl. a közösségi média eljárásrendjei is aggályosak saját visszaéléseik megállapításakor és szankcionálásakor. Így mind az elfogultság maga, mind pedig az ennek nevében végzett válogatás védelmi tényezővé vált.

Véleményem szerint a probléma kezelhető lett a kellő mértékben, ezért az MI fejlődésének vagy terjedésének lassulásában önmagában nem játszik szerepet. Hozzáteve, hogy egyre gyakoribb, hogy emberek és csoportok a valóságot nem fogadják el, tehát támadni fogják a technikát, amiért egy valóságot jól feldolgozó gép nem a nekik tetsző eredményeket adja ki.

III.4.6. Kiberbiztonsági sajátosságok

Említeni kell, de külön tanulmányt igényel az MI sajátosságaival való informatikai visszaélések témaköre.¹⁴³ Számos módon kihasználhatja egy támadó az MI alapjellegét az MI a vírus elrejtésétől, a raj-intelligencia alapú támadásokon át,[169] az adatképzésig (ld. IV.5.).

Ezek a kihívások hasonlóak a számítógépes vírusok korábbi problémáihoz, melyek nem akadályozták meg a fejlődés egészét, azonban sokszor hátráltatták. A kibervédelmi tényező tehát egy lehetséges téli okai között nem túlhangsúlyozandó, de említendő. Hiszen nem zárható ki, hogy egy nagy adatképzési, vagy az MI-t kihasználó támadásból származó botrány olyan mértékű bizalmatlanságot eredményez, mely átmenetileg visszavet számos az MI projektet. A kiber-oldalra katonai-védelmi szempontból kicsit még visszatérek (VII.2)

III.5. A VÉGES PARADIGMAVÁLTÁSI KÉPESSÉG ÉS EGYÉB HUMÁN TÉNYEZŐK

Ennek az alfejezetnek a célja ráirányítani a figyelmet olyan – elsősorban emberi – aspektusokra, melyeket a feldolgozott szakirodalom nem, vagy nem eléggé vett figyelembe.

¹⁴² Ezt részletesebben vázolólok más helyen: [145, o. 369]

¹⁴³ Évekkel ezelőtt több tanulmányban foglalkoztam a témával, azóta azonban számos újabb információval lehetne az akkori kutatásokat bővíteni. [168]

III.5.1. A vertikális tanulás korlátossága

Egy technológia fejlesztésében sok az objektív tényező, ezért jóval kiszámíthatóbb maga a fejlesztés, mint az eredmény fogadtatása.[170] Hiába jön ki egy jó termék, ha túlságosan újszerű és a célközönség nem szereti meg. Fontos probléma tehát egy új technológia fogadtatásának és bevezetésének prognosztizálhatósága. Alább egy olyan megközelítést vázolok, mely hasznos eleme lehet olyan modelleknek is, melyek egy újdonság társadalmi bevezethetőségét vizsgálják.

A megközelítés a horizontális és a vertikális tanulás¹⁴⁴ megkülönböztetésére épül. Ezek lazán kapcsolódnak fentebb említett horizontális és vertikális fejlődés kifejezésekhez. Röviden a *horizontális tanulás* a meglévő szemléletet csiszolja, egészíti ki további információkkal, míg *vertikális tanulásnak* azt nevezzem, amikor a tanuló egy belső szemléletváltást is végrehajt. Itt kapcsolódik a megkülönböztetés a fejlődéshez: egy vertikális fejlődési lépcsőben kapott technológia kihasználásához mindig szükséges egy belső paradigmaváltás, és ehhez csak a vertikális tanulás által lehet eljutni. Eltér azonban a fogalmak felosztása abban, hogy míg a fejlődés kétféleségét a modernitás hozta létre, az tanulás esetében mindkét típus egy ősi, emberi képesség, mely már az ókori embernél is megjelent. Amikor például a bölcsességet az okosságtól megkülönböztették, talán az itt vázolt szemléleti váltásra utaltak. A bölcs ember a többféle szemlélet birtokában ugyanis nem ragad bele egyetlen megközelítés buborékába, ezért érti meg jobban a világot.

Nézzünk néhány hasonlatot és gyakorlati példák a vázolt fogalmak pontosításához. Bárki tapasztalhatja, hogy az embereknek huszonéves korukra kialakul egy világszemléletük, amelybe később folyamatosan beleillesztik a megszerzett információkat. Bizonyos új információk könnyen illeszkednek a képbe, mivel jól kiegészítik a korábbi tudást, mélyítik a korábbi megértést. Ilyenek például a szakmával kapcsolatos dolgok: mozdulatok, fogások, lehetőségek stb. Ezek elsajátítását értem tehát horizontális tanulás alatt: ennél a tanulásnál az új információ többé-kevésbé könnyen illeszkedik az új rendszerbe. Az ilyen tanulásban van energia és munka, sőt néha hosszabb-rövidebb továbbképzések, vizsgák által olyan jelentősen bővül ez a fajta tudás, hogy előkészít egy vertikális lépést, vagyis egy szemléletváltást.

Viszont ilyen ismeretgyűjtéskor számos információ képes „lógni a levegőben”: az illető csak akkor tudja az információt, ha rákérdeznek, de önállóan felhasználni nem képes azt. Ennek oka lehet az, hogy a befogadó számára érdektelen a téma, de az is, hogy az a korábbi megértés, mely

¹⁴⁴ Erről bővebben 2025 tavaszán megjelenő kötetben [2]

más „információtéglakon” alapul diszharmóniában van az új „információhengerekkel”, ez utóbbiak egyszerűen legurulnak a régi építményről.

Más néha hasonlittal a horizontális tanulás olyan, mintha különböző méretű legókockákat akarnánk egymásba pattintani, s mikor ez nem sikerül, az inkompatibilis legódarabokat csak beszorjuk egy sarokba. A vertikális tanulás esetében valamilyen inspiráció hatására az illető az összegyűlt inkompatibilis legódarabokat rendezi, új legóhidat kezd építeni, és számos új elemet szerez be ehhez, hogy a híd összekösse az új partokat. Képességtől és kortól függően ez különböző méretűre épülhet, de a régebbi híd (hidak) megmaradnak. Ezért mondható ez szemlélet-kialakító vagy szemlélet-váltó tanulásnak: ezáltal képes az alany kihasználni az új paradigma erejét, hiszen egy új híd épített belőle önmaga és a valóság között. Érzékeltetni lehet a vertikálisan elért tudást a nyelvtanulásból ismert „az idegen nyelven gondolkodás szintje” minőségi átfordulásával is – bár egy technológia esetében ez inkább azt jelenti, hogy az embernek „rájár arra az agya”, hogy hogyan lehetne különböző problémákat az újdonság által megoldani.

Az ilyen építőmunka azonban nagyságrendekkel több lelki energiát igényel, mint a horizontális tanulás. Fiatal korban még bőven van erre energiája az embereknek, idősebb korban egyre kevésbé. Ennek ellenére sokan nem építenek fel igazi szemléletet fiatal korukban, csupán a néhány kéznél levő megközelítést egymás mellé raknak, melyek nincsenek jól „kötésben”. A szakmai oktatás egyik fő feladata (lenne) a választott szakmában egy áttekintő szemlélet kialakítása, vagyis az első híd vázának felépítése, melyet a hallgató munkába állva néhány év alatt a szakmai rutin által járhatóvá csiszol. Sokáig az emberek az iskolai és a betanulási idő alatt szerzett tudást élete végéig használhatták. Napjainkban azonban egy tíz-húsz évvel ezelőtt kapott szemlélet már sokszor elavultnak számít.

Ki kell itt emelni, hogy a vertikális tanulással az az alapvető probléma, hogy meg kell teremteni hozzá a megfelelő lelki, energiát (megfelelő képességek birtoklása mellett is). Ehhez pedig valami erős belső motiváció is szükséges, amelyet sok tényező gátol.

Befolyásolja egy önváltoztatási képesség, az életkorral összefüggésben is nehezebb az egyénnek a kellő energiát előállítani, de függ attól is, hogy hány ilyen szemléletváltáson ment keresztül az illető. Elképzelhető, hogy szinte folyamatos ez a szemléletváltás (pl. elhivatott kutatóknál), de csak bizonyos életkorig. Ugyanakkor a társadalom túlnyomó többsége számára borzasztóan nehéz. Legtöbbször képtelenek megteremteni magukban a motivációt hozzá, hiszen úgy érzik, hogy élhető a saját életük e nélkül is, lám, eddig is jól működött...

Az eddigieket összefoglalva kimondható, hogy ez a belső lelki energiahiány magyarázza azt a jelenséget, hogy egy új paradigma terjedéséhez egy bizonyos időnek kell eltelni. (Gyakran legalább egy részleges generációváltás szükséges hozzá.) Az ok azonosítása után nézzük meg

a jelenség néhány szempontunkból fontos következményét. Az egymásból következő diszkurziós lépéseket a követhetőség érdekében felsorolásként szedtem.

- Induljunk ki abból a tényből, hogy az ilyen mérvű önváltoztatás régebben nagyon ritkán volt szükséges (és keveseknek). Napjainkra azonban ez elvárássá kezd lenni, olyan ütemben árasztanak el minket a diszrupciók.
- Bár növekszik azok száma, akik inspiráltak arra, hogy egy-egy ilyen szemléletváltás érdekében a kényelmüket, magánéletüket, szabadidejüket beáldozzák, de nem nő eléggé. Úgy tűnik, hogy a társadalom teljes szélessége felé itt nyugaton nem lehet ilyen elvárást támasztani: ez túl nagy személyes áldozat egy individualista közönség számára. Pontosabban jelenleg nem tudjuk elképzelni, hogy ezt el lehessen fogadtatni. (Ezt a gondolati szálát védelmi aspektusból majd folytatom (ld. VI.2)).
- Ezek alapján nagyon úgy tűnik, hogy a modern társadalmakban nem várható, hogy kialakul egy tömeges belső igény a munkával kapcsolatos önváltoztatásra. Az is erősen kérdéses, hogy megfelelő képességek birtokában van-e mindenki, akitől ez elvárható lenne a munkájával összefüggésben.
- Mindehhez hozzáadódik, hogy a jelenég még problémásabb a döntéshozók esetében. Mert ha ők benne rekednek egy régi szemléletben, az (döntési pozíciójuk miatt) sokkal erőteljesebb visszafogó hatással van egy-egy új paradigma terjedését, jobb kihasználását illetően. Erre viszont nagy az esély, mivel általában túl leterheltek, így akkor sem marad idejük és energiájuk ekkora önváltoztatásra, ha kedvül lenne hozzá.
- Ezt a helyzetet nem látszik megoldani egy generációváltás sem, hiszen a fiatalok csak egyre demotiváltabbak, tehát nem valószínű, hogy az új generáció folytatni akarja a világ ilyen tempóban történő változtatását.
- Ezek a tények arra utalnak, hogy a modern világunk túl sok változása már tarthatatlanul sok változást vár el az emberektől. Az embereknek pedig egy lelki tulajdonsága, hogy szereti a bejáratott dolgokat az életében, nem szeret sokszor nagyot változtatni, főleg nem önmagán.
- Minél erősebb kognitív képességekkel rendelkezik az MI, annál erőteljesebben érvényesül majd a gondolkodásmód leváltásának fontossága a megfelelő ember-gép együttműködésben. Másképp fogalmazva minél nagyobb fejlődik az MI annál nagyobb vertikális ugrás fog kelleni. (Ez az agyhoz kapcsolt ember-kiterjesztésekre is vonatkozik.)
- Mindebből arra következtethetünk, hogy „lelkileg nem fenntartható”. az MI fejlődésének jelenlegi üteme.

A fenti gondolatmenetet alátámasztja az a tény is (amelyet bárki megfigyelhet a környezetében), hogy a tömegek az eddigi diszruptív technológiákat is csupán felületesen alkalmazzák, azok lehetőségeiket sem aknázzák ki. Számítógépüket vagy okostelefonjukat kevés dologra használják, még az irodai programokból is csupán néhány alapfunkcióval dolgoznak, nem bíbelődnek biztonsági állítgatásokkal stb. Hasonló hozzáállás rajzolódik ki minden új paradigma kapcsán. Pl., ha a keresőbe van beépítve MI, akkor használják, de promptolni csak kevesen akarnak tanulni. Pedig a promptolás még horizontális tanulás (bár tekinthető egy új szemlélet előszobájának), nem szükséges hozzá érteni a nyelvi modellek technológiáját. Mégis a többség óckodik tőle, hiszen valami új, valami programozás-szerű. A legtöbb ember, ha érzi is az új paradigma szelét, nem kívánja azt a vitorlájába fogni, és nem akar akkorát változni, amekkorát az a dolog megkövetelne.

- Itt értünk el gondolatmenetünk másik következményéhez, mely a lassulást eredményezi: a fejlesztések akkor hozzák a legideálisabb profitot, ha néha bevárják a piacot, vagyis a társadalom széles rétegeit.

Bár elvileg létezik egy másik lehetőség is: hogy megteremtik az inspirációt a társadalom tagjaiban egy jelentős önváltoztatására (vagyis a változásokhoz szükséges új szemlélet belsővé tételére). Ehhez azonban nem elegendő a mai fejlett befolyásolási technológiák arzenálja, hiszen azok általában épp a kényelmi vagy érzelmi aspektusra építenek. – ezesetben viszont épp a kényelem ellenében szükséges haladni. Lehet erőltetni kötelező gyorstalpaló képzéseken való részvételt, és ezáltal kimutatható lesz, hogy többen használják az új lehetőséget. Ám ez nem fogja kialakítani bennük azt az új szemléletet, ami által valóban kihasználják az újdonság – esetünkben az MI – képességeit. Ez a lehetőség ezért csupán elvi: hiszen a szemléletváltáshoz szükséges érési idő is nehezen biztosítható „felülről”, a hozzá szükséges lelki energia rendeltileg nemigen adható meg. Meghallgatják, amit mondanak nekik a képzésen, de abból csak horizontálisan lesznek képesek tanulni, tehát csak azokat a konkrét dolgokat használják munkájukban, amelyeket könnyű volt beleépíteni. A képzések persze nem haszontalanok, mert a horizontális tudás is csírája lehet egy jövőbeni önváltoztatásnak.

Összegzésül kimondható, hogy a napjainkban az élet minden területén tapasztalható paradigmaváltási dömping kezd olyan mértéket öltetni, ami kezelhetetlen. Így az új paradigmák kénytelenek lesznek bevárni a társadalmat. Mert ha egy-egy terület szakemberei saját új paradigmáikat hajlandóak is elsajátítani, a sok egyéb új területen rájuk zúduló szemlélet-váltásra nekik sem marad idejük és energiájuk. Pedig valójában egymást generáló, és egymással szervesen összefüggő területek változásairól van szó, melyek a mindennapi életünket szövik át.

Ráadásul a belső motiváció és a tehetség, a többszörös szemléletváltáshoz szükséges két fontos paraméter nagyon véges, miközben végtelen önváltoztatási kedvet és tömeges zsenialitást kívánna a most épülő jövő. A mai komplexen paradigmaváltó fejlődés ezért nem tartható fenn.

III.5.2. IQ-fenntarthatóság (a tehetségek végeessége)

Mai kultúránk egyre inkább épülnek a tehetségre, ezáltal az olló egyre nyílik a kevésbé tehetséges emberek és azok között, akik jó adottságokkal rendelkeznek. Ez kiemelten igaz az okosságra: az élet minden területe egyre bonyolultabbá válik, és minden munkakörre jellemző, hogy egyre több mindenhez kellene értenie az azt betöltő személyeknek. Mint imént említettem a vezetői feladatoknál akár extrém (lehetetlen) mértékűre nőhet a tudáselvárás. Más munkakörben a helyzet csak kevésbé extrém, de ugyanilyen reménytelen, és nem csak az un. „értelmi-ségi” területeken, hiszen már a szakmunkák elvégzőinek is rengeteg új technológiát, módszert, paradigmát kéne követni tudni. Vagyis a fejlődés, az egyre jobb módszerek és technológiák egyre nagyobb mennyiségű magasan képzett munkaerőt igényelnek az élet minden területén. Tehát egyre több egyre okosabb emberre lenne szükség ahhoz, hogy minden munkakört megfelelő ember töltsön be. Eközben azok a munkák, melyeket eddig a szerényebb képességű emberek el tudtak látni robotizálódhatnak. Látható tendencia, hogy pár éven belül a szoftveres robotok (MI-k) képesek lesznek átvenni számtalan olyan bürokratikus feladatot, melyet most emberek végeznek. A fizikai robotok is képesek lesznek szinte minden szakmukában feladatot végezni, ezek terjedése azonban a várhatóan magas árak (túl hosszú megtérülési idő) miatt jóval későbbre várható.

Ezzel a jelenséggel több szinten adódik probléma. Először is az ilyen fejlett szoftveres rendszerek, és pláne a fejlett robotok karbantartásához is magasan képzett emberek szükségesek. Akiket a technológiák kiszorítanak, azokat erre nemigen lehet átképezni, számukra egyre kevésbé lesz bármilyen lehetőség. Erről a kérdéstről folynak kutatások (pl. [171]). Viszont kevésbé kutatott a másik terület – amely szerintem sokkal nagyobb gond – hogy nem is születik elegendő olyan képességű ember, akiből ez az óriási tömegű magasan szakmaiságú munkaerő kiképezhető lenne. Tehát azért lehetetlen elegendő jó orvost, jó pedagógust, jó mérnököt, jó kőművest találni, mert nem létezik ennyi tehetség világszinten. Természetesen a fentebb (III.3.2.) már említett agyelszívás miatt egyes területek egy ideig meg tudják oldani a maguk számára a kérdést, de ők sem tudnak majd mit kezdeni az ott született kevésbé tehetséges emberekkel.

A jelenséget itt jobb híján IQ-fenntarthatóságnak neveztem el, mert szempontunkból ez a legrelevánsabb komponens, de ez kifejezés igazából pontatlan. Ugyanis a fentiekhez hozzáadódik, hogy a magas szakmai tudás mellé egyre több egyéb adottság vagy nevelt készség társulása

is szükséges: a stressztűrés, a magas koncentrációs készség, a jó kommunikációs készség, kedvesség, gyors reakcióidő és hasonló adottságok egyre több helyen válnak elvárássá – így azok, akik szakmailag jók, de egyéb kompetenciáik nem megfelelőek, azok csak részlegesen lesznek tökéletesek a feladatuk ellátására. Ez természetesen még lehetetlenebbé teszi az elegendő jó munkaerő megoldhatóságát. Tehát pontosabb lenne úgy hivatkozni erre, mint „általános tehetség-komplexum-fenntarthatóság” – ám ez igen nehézkes elnevezés, maradjunk a címbelinel.

Jelen vonatkozásban kiemelendő, hogy az MI fejlődését például a pedagógus vagy üzemeltető területek elvárás-szintjének mérhetetlen növekedése is hátráltatja. Ennek védelmi vonatkozása is van, hiszen ahonnan elvándorol a tehetség ott még égetőbb lesz a probléma (ld. VI.).

III.5.3. Néhány további tényező

Bár átfogó vizsgálatra nem törekedtem, de ide kívánczik még néhány meglátás. A filozófiai aspektusokat a következő alfejezetbe szerkesztettem, itt egy, az előző alfejezethez kapcsolódó szempontot és néhány technikai akadályt említek.

1. **Új generációk – új elvárások.** Azt az előző kisfejezetben tett felvetést szeretném folytatni, miszerint a vertikális tanuláshoz szükséges jelentős motiváció problémája újszerű gondot okoz egy demotivált generációban. Amennyiben a múlt tapasztalatai folytatódnak, akkor egy új generáció már az élete szerves részeként fogja igényelni és használni az MI-t, ahogyan a mostani fiatalok belenőttek a „kütyük” világába. A jelenlegi generáció számítógéphez szokott része egyre kevésbé szorítkozik saját tudására és memóriájára, hiszen az interneten kikeresi, amit kell, és amikor kell. Hasonlóan a következő nemzedék: ha hozzászokik ahhoz, hogy az MI bármit megfogalmaz, összefoglal, összeállít helyette, akkor az átlagod a memóriaképességen túl a szintetikus és analitikus képességek is csökkenni fognak. Ez részben azt eredményezheti, hogy a nyugati gyökerű szakemberek tovább fognak. Ám ezen kívül egy közömbös vagy esetleg ellenséges viselkedés is kialakulhat az MI-vel szemben. Hiszen az önmagukat nem találó fiatalok gépeket okolva saját életük problémáikért, megfelelő bűnbakot találhatnak a modern technológiákban, ezen belül az MI-ben. Egy ilyen ellenállás hasonlóan bontakozhat ki, mint ahogyan például a környezetvédő eszmékből sajátos életvitelt hirdető mozgalmak váltak le (speciális táplálkozás, gyermek-nem-vállalás). Ez a jelenség az MI esetében egy egyszerű bojkotton túl olyan ideológiák terjedését idézheti elő, amelyek az élet princípiumát hangsúlyozzák túl a gépek életteleniségével szemben.
2. **Az agygép mai modelljének kérdőjelei.** Az sem biztos, hogy kizárólag a neuronok közötti elektromos jelek határozzák meg a lemásolni kívánt agyi működéseket. Például a kvantumpszichológia eredményei megdönthetik, hogy az agysejtek csupán az agyi-idegi

összeköttetések útján vesznek részt az ember működésében.[172] Sőt az sem elképzelhetetlen, hogy hiába találja meg valaki „az agykutatás rosetti-kövét”,¹⁴⁵ mert minden ember agyának „pontos digitális kódja” lehet, hogy egyedi (mindenkihez saját, különböző dekódolási eljárás szükséges).

3. **A modell bonyolódása.** A moduláris felépítés miatt tűnik nagyon hatásosnak az MI, de közeledik a pillanat, hogy az egészet egyben kéne átlátni. Ez pedig az említett fekete doboz problémát hatványozza, ugyanis jelenleg elvileg megoldhatatlan helyzetet, vagyis egy paradoxont eredményez: e szerint a gép emberibbé válásához az az összegző megértést kellene, amire az ember ilyen mennyiségben nem képes, a gép pedig elvileg nem képes.
4. **Az ember-gép-összekapcsolás topogása-1: agytenyészet.** Óriási lépéseket tett a biológia és az orvostudomány az agy részeinek, működésének, valamint az agyhullámok feltárásának területein, de az agyban áramló információk formátumának megfejtéséről nincs hír. A próbálkozások nem képesek átlépni az ember-gép dekódolás küszöbét aszerint, ahogyan azt a szingularitás-hit megálmodta. Ennek az utópiának az egyik beteljesedése az lenne, ha a tudatot, vagy legalább az agyat, vagy annak egy használható részét sikerülne egy gépbe úgy lemásolni (esetleg sejtenyészetként létrehozni),[174] hogy az működő rendszer maradjon.
5. **Az ember-gép-összekapcsolás topogása-2: agy-gép kapcsolat.** Az előző pontban említett összekapcsolás másik vonala, az ember kiterjesztése lehetne agyhullámokkal vagy agyba ültetett chipekkel (BCI).¹⁴⁶ Kutatásaim szerint sem agyhullámokkal,[175] sem pedig BCI eszközökkel [176] nem sikerült tudati kapcsolatot teremteni. Bár a területben nem volt célolemélyedni (ld. téma lehatárolása), az általam vizsgált kísérletekben mindig valami sajátos dekódolást kell megtanulnia az embernek az ember-gép kapcsolathoz, nem tudják az agyi neurális információáramlást (amit valaki gondol, képzel vagy emlékszik) közvetlenül értelmezhető információvá alakítani. Vagyis hiába gondol a vezérlést végző arra, hogy a robotkar menjen jobbra, az nem erre fog reagálni – ehelyett valami erős gondolatot kell társítani a „jobbra” irányhoz.[177] Így lehetségessé vált ugyan agyhullámokkal vagy implantátumokkal robotkارت vezérelni, játékot irányítani, vagy betűket leírni. Ám egyelőre sokkal nehezebb ezt megtanulni, mint a kézzel kezelhető input eszközök kezelését, és sokkal nagyobb koncentrációt is igényel (nagyon fárasztó). Mivel itt ennek mélyebb vizsgálata nem célolem, csupán két meglátást említek a technológia sci-fi (fantasy) filmekben látható olyan megvalósulása ellen, ahol behelyezve nem sokkal később működik is a kapcsolat az agy és a beépített chip között.

¹⁴⁵ Egyébként a rosetti kő esetében is egy tudományos tévedés akadályozta a hieroglifák megfejtését. [173]

¹⁴⁶ Brain-computer interfaces, vagyis agy-komputer interfész a pontos tudományos elnevezésük.

- Az agy lassan meg tudja tanulni a gép közvetlen kezelését, de a gépet nem tudjuk úgy előkészíteni, hogy közvetlenül kezelhető legyen. Itt még inkább probléma az előző pontban említett tény, hogy az agy ugyanolyan feketedoboz jelleggel működik, mint amellyel a gépi neuronhálót jellemeztem (II.2.1.), hiszen arról másolták. Tehát akár egy neuron állapotát mérjük, akár a kapcsolatait, azokból az értékekből nem tudunk magára az információra következtetni egyik esetben sem. Így egyelőre nincs „feltörési” lehetőség, mint egy kódolt adat esetében, mivel egy-egy konkrét adat szétosztva van jelen ebben a hálózatban, nem lokalizáltan. Ezért hiába ültetnek be valamit, az nem képes egyből bekapcsolódni az adott agy ingeráramlataiba. Az agy számos bal eset után képes bizonyos funkciókat újra tanulni, például felnőtt sérültek újra tanulhatnak járni vagy beszélni. Ehhez hasonlóan bánik az agyunk az implantátummal: képes lehet használtba venni egy beépített chipet, de ez épp olyan hosszú és nehéz munka, mint az említett rehabilitációk.
- Ehhez kapcsolódik fentebb, a 2. pontban említett probléma, hogy felmerülhetnek, egyéb, egyelőre nem mért jellemzők is, amelyek részt vesznek agyunk információfeldolgozásában. Nehéz azt igazolni, hogy mindent tudunk mérni, viszont az ismeretlen tényezők nélkül talán eleve nincs esélyünk egy gép és az agy gyors összekapcsolására.

Összegezve: az ember és gép hatékonyan és kényelmesen működő összekapcsolásáról beszélni egyelőre annyira hipotetikus, hogy ezen a téren nem érdemes a vertikális fejlődés bekövetkezésével számolni.

III.5.4. Filozófiai problémákról és ezek hátteréről

A tanulmány eddigi fejezeteit átszőtték az olyan filozófiai problémák, melyek az MI kutatók már a kezdetektől beárnyékolják. Ezek vizsgálata és megválaszolása kívül esik a mérnökök és matematikusok kompetenciáján. Az MI filozófiai vizsgálatára itt ugyan nem vállalkozhatunk, de néhány fontosabb buktatót meg kell említeni.

1. Helytelen kiterjesztések. Ennek a hibának az eddigiekben a következő eseteit azonosítottam:

- A valóság leegyszerűsített modelljének (vissza)általánosítása a világ leírására (III.1.3. / 2.d és 3.b);
- A fejlődést tartalmazó elméletek szint-sorrendjeinek alkalmazása gépvilágra (a Moravec paradoxon háttere); (III.1.3. /3.a);
- A régi, zárt rendszerben történő fejlődés jelenségeinek (pl. hype) kiterjesztése egy tömegeket érintő globális fejlődésre (III.2.2);

- Bár a részletes cáfolatot itt nem közöltem, de a szuperintelligencia és szingularitás kapcsán imént említett példák is ide tartoznak (III.5.3. / 2. és 3.), már csak azért is, mert a végtelen problémákat véges megközelítéssel megoldhatónak vélik;

Ide sorolható továbbá az a szokásunk, hogy a teljesítményről a kompetenciára következtetünk például a nyelvi modelleknél (hiszen így alakítjuk ki a benyomásunkat egy másik ember tudásáról) – ezt a szokást pedig nem szabadna egy MI-re kiterjeszteni (általánosítani).¹⁴⁷ Ezek jól mutatják, hogy az MI fejlesztés számos ponton nem gondolja át megfelelően az alapokat.

2. **Lesz-e mit kódolni vagy modellezni?** A humántudományok, elsősorban a pszichológia jelentős fejlődése eddig lehetővé tette, hogy legyen mit gépiesíteni. Az alapvető paradigmaváltáshoz, az AGI-hoz szükséges tudás azonban még nem tűnik elérhetőnek. E téren egy alapvető filozófiai kérdés is megválaszolásra vár: Mit értünk *ember* alatt, amit utánozni szeretnénk? Mert ha egy olyan segédmunkást, aki a kábítószer miatt gyorsabban végzi el azt, amit betanult, akkor már sikerült is ember-szerű gépeket alkotni. Ám ha csak egy közepes tehetségű, középkorú polgár tudásait (élettapasztalatát) szeretnénk utánozni, már akkor is sok kérdőjel állja utunkat, de egy megbecsült tudós vagy művész netán egy zseni létrehozása még távolabb esik a realitástól. Egy tehetség kibontakozásához számos kompetencia jó együttállása szükséges. De kérdés, hogy ennek létrejövetele mely intelligencia típusokra támaszkodik, vagyis az intelligencia mely területeinek jó utánzása várható el ahhoz, hogy a gépi minőségileg szintet lépjen? A sokféle intelligencia és ezek organikus egybeolvasztása meglehetősen új pszichológiai modelleket igényel, de olyanokat, amelyeket működő matematikai modellé lehet konvertálni. Ez is alátámasztja, hogy az „emberies AGI” (III.6.1.) még nem reális hiszen hiába lesz ez a sokkal jobb modell, ha nem világos, hogy hogyan értjük az „ember” kifejezést (ld. még III.6.2.).
3. **Az MI-nek nincs mögöttes világmodellje**, mint az őt létrehozó embernek.[23] Ez is akadályozza az emberies AGI létrejöttét, de logikai akadálya a gépek által fejlesztett gépek felsőbbrendűségéről, a bennük (az ASI-ban) létrejövő evolúciós lépésről szóló teóriák beteljesedésének is.
4. **Értelmezési nehézségek.** Maga az „értés”, vagy „értelmezés” szó körüli értelmezési probléma sem egyszerű, de az imént említett *ember* kifejezés, az *öntudat*, a *tudat* szavak meghatározása is régen nyitott probléma. Zavarossá válik minden kifejezés, ami az emberi értelemhez, érzelemhez, szabadsághoz kapcsolódik, ha olyan szempontból nézzük, hogy a gépek tudásána milyen kompromisszumok mellett építhető be (az adott elvont dolog).
5. **A korszellem attitűdjei.** Talán a korábbiak (pl. III.2.3./3) alapján nem hangzik furcsán az az állításom, hogy a *modern túlracionalista korszellem kudarcának egyik kárvallottja az MI*, és a mögötte álló ipar. A szakirodalomban ebben a formában nem

¹⁴⁷ Például egy nyelvi szinten zsonglőrökűdő utánzógépre. [23]

szerepel ez a vélemény, ezért röviden magyarázom. A II. világháború utáni korszak nevezhető a „tudomány-vallás” korszakának is, amikor tudósok és egyszerű emberek hite a tudomány mindent megoldó erejébe vetett bizakodás volt. Ennek egyik zászlóshajója volt az MI, amely által az ember, akit ember-gépnek véltek, egy gép-ember „Teremtőjévé” válhatna.¹⁴⁸ (Az embergép szemlélet korábbi, de ekkor egy időszakos csúcson volt). Ennek a korszaknak alapvető szerepe volt az első két tél mögötti túlzott reményekben, hiszen a kutatók és támogatók egyaránt vágytak ekkorát alkotni (teremteni). Ennek vallásos jelleggel bíró hajtásairól már szoltam (III.2.3.), itt külön említeni azért kellett, mert ez alapján a jelen és közeljövő korszakának is nyilván vannak olyan meghatározó attitűdjei, melyek bekorlátozzák, de majd utólag válnak jól láthatóvá.

III.6. A TOVÁBBLÉPÉS LEHETŐSÉGEI

Szükséges átgondolni, hogy az eddigiek fényében merre fejlődhetnek az ANI rendszerek.

III.6.1. Paradigmaváltás, de csak a szavakban: multikogníciós rendszerek igazi AGI helyett

Bár a fogalmi vizsgálatok inkább egy későbbi fejezetbe sorolandóak, de ezt a terminológiai problémát már itt szükséges tárgyalni. Állítom ugyanis, hogy az AGI sokak által jóslott megjelenése a szó tartalmának eróziója által várható. Korábban bemutattam (I.2.3.), hogy régóta elfogadott szintek alapján, egy Mesterséges Általános intelligencia (MÁI - AGI) komplex emberi tudása lenne a következő nagy előrelépés, egy valóban új paradigma. Várakozásaim szerint azonban a szót néhány éven belül alkalmazni fogja valamelyik fejlesztő cég a saját termékére. (Ezt az állításomat a valóság igazolta is, de erre is majd a részletesen vizsgálatban térek ki). Azt is láthattuk, hogy a jelenlegi technológiák bármilyen bonyolultak, mind csupán az ún. Mesterséges Vékonyintelligencia (MVI - ANI) halmazába sorolhatók. Állításomat kicsit már a kifejezés története is igazolja, nézzük meg hát a „csúcs-MI” megközelítésének változását.

Az AGI-t korábban hívták Erős MI-nek is (Strong AI), ám ekkor a koncepció lényege a *mesterséges tudat* elérése lett volna.[178] Ennek ismeretelméleti problémái miatt kezdtek el áttérni az „általános” jelzőre, és ebben a megközelítésben „csupán” egy minden intelligenciatípus szempontjából az emberi szintet elérő rendszert jelent az AGI. Tehát a szavak szintjén már módosult a cél az elmúlt húsz évben: emberi tudat helyett emberies tudás lett a cél. Ám, ha már

¹⁴⁸ Ennek humoros de markáns képviselője a híres Star Wars fantasy-sorozat 3-PO robotja, aki a közismert „Teremtőm!” vallásos fohászt „Alkotóm!” felkiáltásként használja. Az akkori sci-fi irodalom még több, viszont kevésbé közismert példát szolgáltat erre a korszakra.

egyszer módosult, miért ne módosulhatna többször? Ezért nem kizárható ki, hogy nem a gép fog egy korábban elképzelt általános emberi tudást utánozni, hanem inkább egy elért technológiát fognak úgy hívni, ahogyan korábbi álmaikban a paradigmaváltó technológiát nevezték.

Az emberség gépesítése iránti szkepszis nem csupán bennem fogalmazódott meg, a kutatók többsége mára valamilyen mértékben eltávolodott a „mesterséges tudat”, „igazi, emberi megértés vagy érzések” gépi leképzésének utópiájától. Ez a „valamilyen mérték” egyénenként eltérő, de mégis besorolhatóak két nagy típusba az AGI megközelítései. Ehhez a következő terminológiai felosztást definiáltam (számos leírás figyelembevételével):

- **technikai AGI:** az AGI olyan komplex okosrendszer, amely hasonlóan sokrétű, mint ahogyan egy mai operációs rendszer, főleg a rátelepített felhasználói programokkal együtt szemlélve azt. Vagy másképp megközelítve a technikai AGI egy bármilyen praktikus dologra megtanítható,¹⁴⁹ nem pedig adott célokra tervezett gép.
- **emberies AGI:** gyakran értenek az AGI alatt az ember kognitív képességeit bírtokló MI-t is (vö. III.2.3. /3. mozgalmai). Ez a felfogás kezdi áttenni székhelyét a populáris írásokba, viszont mivel mérnökök, programozók és hasonló szakemberek utalgatnak rá, ezért magas olvasottságú cikkeket és könyveket lehet e témában írni.

Erre a felosztásra nem találtam forrást még az AGI egyik 2010 óta működő szaklapját¹⁵⁰ átnézve sem, amely pedig számos színvonalas írást közöl.¹⁵¹

Alább majd be fogom mutatni az „okos” szó degradálódását is, és azt a jelenséget, amikor a felhasználó hagyományos rendszereket vél okosnak (ld. V.1.). Az ott leírtak további bizonyítékot adnak a felvetett várakozásra, hiszen véleményem szerint ugyan az a mentalitás fog az AGI esetében a háttérben állni, melyet ott részletezek. Sőt valójában minden „vágyott technológiára” alkalmazható a modell.

Ahogy erre az intelligencia vizsgálatokor (I.2.) rámutattam, még a szakemberek számára is kérdés, hogy pontosan mit kellene egy AGI rendszernek tudnia, így valójában a piac sem tudhatja, hogy mit takar ez a betűszó, csupán azt érzi, hogy valami nagyon nagy újítás, és nagy bevétel lehet mögötte. Ebből következik az a véleményem, mely szerint pár éven belül piacra kerül majd olyan szolgálattás, melyet azzal reklámoznak, hogy „ez az első általános mesterséges intelligencia”, ez azonban a fenti két megközelítés egyikébe sem fog beleilleni. Az első AGI-néven reklámozott terméke tudása messze nem lesz az emberi kogníció bonyolultságához mérhető, csupán egy modulárisan összetett vékony MI-rendszeregyüttest hívnak majd így.

¹⁴⁹ „Az általános MI-rendszer célja olyan rendszer létrehozása, amely a legtöbb olyan tevékenységet el tudja végezni, amelyre az emberek képesek. A szűk MI-rendszerek viszont olyan rendszerek, amelyek egy vagy néhány konkrét feladat ellátására képesek” [13]

¹⁵⁰ <https://sciendo.com/journal/JAGI>

¹⁵¹ Még az MI fogalmát is így közelítik meg: [179]

Persze az odáig vezető innováció komoly fejlődésnek tekinthető, de valójában csak a horizontális fejlődés értelmében. Hiszen, bár igen sokrétű szolgáltatással, de alapvetően csak a jelenlegi megközelítések összeadását valósítaná meg az ilyen új rendszer, vagyis egy komplex ANI, esetleg, ha majd többféle kogníciót utánoz, akkor multikogníciós rendszernek is nevezhető, de alapvetően csak egy összetett ANI marad.

A kritikusok hiába írnak majd esetleg arról, hogy helyesebb lenne az ilyet „sokcélú” vagy „komplex MI”-nek hívni: az okostermékek marketing-hullámához hasonlóan a tetemes fejlesztési és üzemeltetési költségek megtérülését, ezúttal is valami „ütős” szlogentől várja majd a finanszírozó. Tehát arra számítok, hogy még a fentebb technikai AGI típusába besorolt, „bármire megtanítható” multikogníciós rendszer sem valósul meg az első AGI-kén reklámozott termékekben. A „emberies-AGI” típusú, humanoid omnikogníciós rendszerről főleg nem beszélhetünk, nemhogy az első AGI-nak hívott termékeknél, de még sokáig. Számptalan kognitív feladatra képtelenek lesznek az első ilyen néven reklámozott rendszerek, főleg, ha csak a ma ismert MI-funkciók egyvelegét szolgáltatják. Hiszen fontos alapszabály, hogy az intelligenciafajták (I.2.2.) egyvelege, egymás mellé rakása sem azonos ugyanezen intelligenciafajták emberen belüli összességével! Ehhez járul ezen intelligenciafajták bonyolult és ismeretlen kapcsolódásaik, egymásra hatásaik gépi megvalósítása, mely egyelőre kilátástalan.

(1.) A KUTATÁS ZÁRÁSA UTÁNI FEJLEMÉNY

A kutatás lezárása után nem sokkal igazolódott is az itt felvetett állítás – bár én ezt egy-két év múlva vártam. Ezért ezt a bizonyítást eredményként sajnos nem közölhetem, viszont fontossága, más technológiákra való alkalmazhatósága miatt beteljesülése ellenére az anyagban hagytam (és egyben amiatt, hiszen ez engem igazolt). Valóban a marketing oltárán lett az AGI kifejezés feláldozva, de jóval előbb, mint várható volt, egy jóval egyszerűbb rendszerre alkalmazva. Ezzel egyben azt a fenti meglátást is igazolta, hogy a technikai AGI sem fog az elnevezés mögött teljesülni.

Egy 2024 évvégi hír szerint az OpenAI már állítja magáról, hogy elérte az AGI-t.[180] Ez persze elvileg is lehetetlen, hiszen szó sincs multi-kognícióról, csupán egy nyelvi-képi modellre húzták rá a kifejezést. Ráadásul ezzel azt a meglátásomat is alátámasztották, melyet az MI tél vizsgálatakor tettem, hogy nem szabadna gazdasági síkra szűkíteni a fejlődés megközelítését, ehelyett pontosabb lenne a horizontális és vertikális fejlődés modelljét alkalmazni. Az OpenAI hivatalos AGI meghatározása ugyanis még a vártnál is inkább gazdasági síkon maradt, e szerint „az AGI elérése akkor valósul meg, ha egy MI rendszer képes 100 milliárd dollár nyereséget

termelni” [181]. Meglátásomat egyébként 2024. november 30-n leadtam egy tanulmánykötet részére, bár ez csak 2025. áprilisában fog megjelenni.

III.6.2. Az intelligencia szokásos szintjeinek kritikája

Az előző alfejezetben ismertetett szintekkel csak egyik gond, hogy gyakorlati használhatóság helyett feltételezésekre épül. Az intelligenciafajták sokrétűségének fényében megvalósulásuk korántsem olyan csillogó és biztató. Hiszen miközben az emberi képességek kézenfekvő belső összefüggéseit is igen nehéz a gépekbe leképezni, szembesülni kell azzal, hogy számos összefüggésről már azért sem tudunk. A kognitív tudományok már megsejtették, hogy valószínűleg messze nem áll rendelkezésünkre minden emberi képesség tudományos és mérhető modellje, ezek összefüggései egymással és a már tudományosan kezelt tényezőkkel egy mérhetetlenül bonyolultabb emberképet fog eredményezni.

A vizsgálódás filozófiai síkján odáig kell visszamennünk, hogy vajon pontosan mikor valósul meg melyik fajta intelligencia? Konkrétabban: az élővilágban kutyáktól és delfinektől a baktériumok felé haladva hol málik szét el egy-egy intelligencia-típus? A fejlődés irányából tehát mikor jelent meg, és a gépek esetében mikor mondhatjuk, hogy megjelent? Még nehezebb kérdés, hogy lehet-e mérni, mikor és hogyan mérhető, hogy csupán utánozza a gép az adott intelligenciafajtát vagy birtokolja azt? Érdekes, hogy egy papagáj vagy egy majom utánzása esetén senkinek nem jutott eszébe, hogy a tudatosságuktól féljen¹⁵² – míg meg nem jelentek az okos programok. Mert az ilyen ember-alkotta-intelligencia esetében a kezdeti időktől jelen van a gépek tudatra ébredésétől való félelem. Ebben véleményem szerint felfedezhető az intelligenciától való ösztönös félelem is, melyet nemigen szokás említeni. A nagyobb intellektuális erőtől másképp fél az ember, mint a fizikai erőtől, hiszen a másik fél okossága sem fizikai sérülést okoz. Ez inkább a nem-tudatos (tehát a tudat alatti) erőktől és lehetőségektől való félelemhez, vagyis a lelki félelemhez hasonlít.

Ez a jelenség rámutat, hogy igazából az emberekből az „igazi intelligencia” velejárójának tűnik bizonyos fajta tudatosság. Az ijesztően intelligensnek tartott ember jól bánik a rendelkezésre álló tényekkel és álhírekkel. Ez alapján felvethető, hogy a különböző intelligenciafajtákhoz különböző típusú tudatosság-fajták rendelhetők... De térjünk vissza az érdekes gondolatoktól a tudományosabban megközelíthető alapkérdésre, az intelligenciafajták gépiesíthetőségére.

A sok intelligenciafajta arra mutat rá, hogy még szakemberek sem tudhatják pontosan, hogy akár egy technikai, akár egy emberutánzó megközelítésű AGI-nak pontosan mit kellene tudnia?

¹⁵² A „Majmok bolygója” c. regény ötlete is csupán ennek verziója, ahol a gépet állatra cserélte a Pierre Boulle.

Vajon a fentiek közül mennyi intelligencia-típust, milyen szinten kell megvalósítani, hogy joggal így hívhassuk? Az emberies AGI-nál egyértelmű, hogy elvárás az is, hogy a mesterséges intelligencia-fajtákat ne lehessen megkülönböztetni egy átlagember hasonló képességeitől, tehát ezek magas színvonalú megvalósítása szükséges – ez jelen tudásunkkal kevésbé tűnik anyagiilag kifizetődőnek.

Az ASI esetében mindez még problémásabb. Egy olyan gép, mely néhány területen meghaladta az embert, azonban nem képes megvalósítani olyan tulajdonságokat, amit még mi emberek nem is ismertünk fel magunkban, ezért hiányzik még minden intelligencia-modellből – nevezhető-e ASI-nak? Hiába tervez majd magánál jobb rendszert egy MI-t tervező MI: a világgép, amelyből kiindult, az csupán egy részleges tudás lehet, egy „itt tartunk most” jellegű világgép. Persze felvethető egy embereket az embereknél minőségibb módon megfigyelő AGI, amely ez által az emberi elme rejtelseit nálunk jobban képes lesz osztályozni és modellezni, így hozva létre magánál jobb rendszert. Ám a gépek jelenlegi működése véleményem szerint nem ad okot ilyen elrugaskodott fantáziálgatásra. Az ilyen megközelítésben egyfajta valláspótlékot vélek felfedezni.

III.6.3. Más MI-paradigmák említése

Jelen kutatásban terjedelmi okok miatt kénytelen voltam a hagyományos architektúrákon fejlesztett rendszerekre korlátozni a vizsgálatot, így számos fontos kutatási terület elemzése elmaradt. A legjelentősebbet, az ember agyi kiterjesztését¹⁵³ érintettem (III.5.3.), de fontos lehetne még a kvantumszámítógép alapú MI területe, és a többi csírázó technológia, melyről az II. fejezetben szóltam (II.3) melyre nem térhettem ki. Bár ezek közeli hatásuk kizárható, az ilyen kutatások esetleges sikere ugyan új diszrupciót eredményezhet, hasonlóan az eddigi MITavaszkok mögötti új paradigmákhoz. Ennek esetleges bekövetkezése azonban nem mond ellent az itt kapott eredménynek, sőt épp a jelen tanulmányban is feltárt lassulási tényezők generalizálják is ezt a fajta paradigmaváltást.

III.7. ÖSSZEGZÉS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI-NYÁR ÚJ MEGHATÁROZÁSA

- **A várakozások teljesülése.** Ebben a vizsgálatban nem a várt eredményt kaptam, sőt még annak ellenkezőjét sem támasztotta alá a kutatás. Ugyanis nem lehet egyértelműen bizonyítani a H1 hipotézist (miszerint lassulni fog a fejlődés), de cáfolni sem sikerült. Ehelyett

¹⁵³ Itt kiemelendő, hogy Kurzweil jelenleg is tartja a korábban 2030-ra jószolt dátumot, és szerinte akkortájt egy kiterjesztett ember az agyi interfésze segítségével a felhőre csatlakozhat, mely által gondolkodási teljesítménye a milliószorosára nőhet. [182]

arra jutottam, hogy a K1 kutatási kérdést tettem fel rosszul, annak ellenére, hogy azt a szakirodalmak ismeretében fogalmaztam meg. A kérdés hibája az volt, hogy miközben a források az MI-tél kifejezést gazdasági síkon értelmezték, addig az én megközelítésem a fejlődést komplex jelentésben akarta értelmezni, amelybe a technikai fejlődés is beletartozik. A problémát detektáltam, és az anomáliát feloldhatóvá tettem a fejlődés gazdasági és technikai síkjainak szétválasztásával.

A vizsgálat során végülis azt sikerült bizonyítani, hogy lényegileg fejlődni ugyan nem fog akkorát az MI-technológia, mint az elmúlt évtizedekben, viszont terjedni fog és egyre több területen fog hódítani (gazdasági fejlődés várható). Ennek vizsgálatához vezettem be azt a terminológiát, hogy horizontális fejlődés várható, de vertikális progresszió nem, vagyis a jövőben a hangsúly az újítások helyett várhatóan a már elért eredmények csiszolására tevődik át. Ezt a lényeges-lépés nélküli terjedést nevezhetjük MI-nyárnak (a fejlődést általánosabb értelemben használva).

A fejezetben mindhárom kijelölt rész cél teljesült. A második és a harmadik rész cél teljesülése (a szakirodalom áttekintése és a védelmi aspektusok vizsgálata) a fejezetek címein látható. Az elsődleges rész cél megvalósulása – vagyis a téma sajátos megközelítése – által sikerült, hogy a vizsgálat más kutatóknál nagyobb hangsúlyt helyezett az emberi tényezőkre (az MI körüli filozófiai problémáktól az agygép-modell (neurális modell) körüli kérdőjelekig). Ezek megvalósulásával a C1 célt is elértem.

Összefoglalás. Elemezve az első két telet, és a mögöttük megjelölt okokat bemutattam, hogy korszakunk jelentősen különbözik a régi regressziók időszakaitól, ezért most teljesen más tényezők vezethetnek a fejlődés várható lassulásához, mint hajdan (III.1.-III.2.). Az alaposabb prognózis érdekében kitértem a társadalom és az MI-technológia oda-vissza hatására is (III.3.), hiszen számos tényező háttérében ez áll. Sok hátráltató aspektust bemutattam (ld. III.4. kisfejezet-címeit), de részletesebben kellett elemeztem az emberek paradigmaváltási képességét (III.5.), mivel ebben látom a lassuló forgatókönyvet leginkább valószínűsítő tényezőt. Itt a horizontális és vertikális tanulás kifejezések mentén bemutattam, hogy egyfajta lelki energia-törvény alapján az emberek nem lesznek képesek arra a sok-sok szemléletváltásra, amelyet a mai sokrétű fejlődés elvár. A vertikális tanulási modell segített azt is kimutatni, hogy azért is számíthatunk az ANI célrendszerek további terjedésére és az AGI-vá fejlődésének elmaradására, mert be kell várni a fogyasztókat (erre több más érvet is hoztam). Vizsgáltam a következő fejlődési lépcső, az Általános MI (AGI) közeledését is (III.6.), ahol fogalmi eredményt kaptam: megállapítottam, hogy egy majdan AGI-nak titulált rendszer, nem feltétlenül a várt „emberies AGI”-t fogja jelenteni, várakozásom alapján egy „technikai AGI”-t (egy komplexebb ANI-t) nevez majd így a létrehozó vállalat.

- **Részkövetkeztetések.** A vizsgálatok alapján több részkövetkeztetés levonható.
 - Az első téma első részkövetkeztetése (**1R-A**): Számos vizsgálattal támasztottam alá, és többértően jellemeztem a mesterséges intelligencia várható hosszú nyarát. Ennek

későbbi elemzéseikhez több jól használható megközelítést és szakkifejezést alkottam. Így kaptam meg a kutatás 1. eredményét, amely szerint állítható, hogy a mesterséges intelligencia fejlődésében lassulás várható, ami viszont nem érinti a terjedésének ütemét.

- **Második részkövetkeztetés (1R-B):** a kutatás igen sokfelől támasztotta alá, hogy az MI-technológia nagy továbblépései (az emberszerű MI és a szuperintelligencia) sokkal messzebb vannak, mint azt az optimista tudósok várják (sok esetben nem tudományos alapon). Ezt a részkövetkeztetést továbbgondolom etikai oldalról (ld. IV.3/2R-B)
- **1R-C:** A vizsgálatokban rámutattam az MI és a társadalom kölcsönhatására, amely számos hatásvektorokon keresztül analizálható. (Ezek közül többet azonosítottam, amit az MI kulturális hatásalattiságának vizsgálatai során tudok majd hasznosítani IV.2.4.)

Zárógondolatként vertikális tanulás számos alkalmazhatósága közül a személyes területre hívom fel a figyelmet. Ugyanis nem csupán a technológiák kapcsán fontos energiát szánunk a szemléletváltásra, hanem „az élet nagy kérdéseiben” is. Manapság ez a talán legfontosabb aspektus gyakran kimarad. Ezért sok lelki energiát kívánok minden olvasómnak, hogy amellet, hogy a világ paradigmáival megpróbál lépést tartani, maradjon energiája a belső szemléletén is dolgozni.

IV. AZ ERKÖLCSI ÉRZÉK DIGITALIZÁLHATÓSÁGA

A második vizsgálat a mesterséges erkölcsi érzék lehetőségeit firtatja (C2), vagyis azt, hogy milyen szinten lehet képes elvileg utánozni egy ilyen tudást egy neurális rendszer (K2). Feltevés (H2) szerint a K2 kérdésre a válasz nemleges lesz. Reményeim szerint, ha sikerül bemutatni a mesterséges erkölcsi érzékkal kapcsolatot elvi akadályokat, akkor ezzel a bizonyítás is sikerül. Ezzel pedig egyben az is világossá válna, hogy nem reális ennyire emberivé tenni egy elektronikus neurális gépet. Másszóval, az előző fejezet terminológiáját (III.6.1.) használva: csak technikai MI tud elektronikus úton megvalósulni, emberszerű MI nem. Ezért szükséges

A vizsgálat során az autonómia kérdését is vizsgálni kezdtem, mely külön alfejezetté fejlődött. Az Első Rész áttekintésében vázoltam a tanulás, az intelligencia és az autonómia belső kapcsolatát, tehát itt térek vissza ennek harmadik eleméhez. Megvizsgálom, hogy hogyan lehetséges összevetni a gépi autonómiát és az emberi szabadságot, hiszen az autonómia szorosan összefügg az erkölcsi érzékkal. A H2 hipotézis bizonyítása során abból az evidenciából indulok ki, hogy az emberi lét egy meghatározóan minőségi része az, amit erkölcsi érzéknek nevezünk. Tehát egy olyan gépi rendszerben, melytől emberszerű megbízhatóságot várunk, abban szükséges lenne megfelelően utánozni az erkölcsi érzéket is.

IV.1. AZ AUTONÓMIA ANATÓMIÁJA

A szabadság és autonómia előfeltétele, hogy aki döntést hoz, az tudjon megfelelően élni ezzel a szabadsággal, és ne pedig visszaéljen vele. Ezt mind az ember mind a gép meg kell, hogy tanulja. Meglehetősen nehéz kérdéskör ez az emberek esetében, hiszen kultúrkörönként is eltér például az egyén és közösség prioritási sorrendjének megítélése. Itt csupán egy saját felosztással szeretném ezt a kérdést kezelhetőbbé tenni, valamint a gépi döntés néhány alapkérdését tekintve a gépi autonómia szintjeit is újragondolom. Az alfejezet célja, hogy a fogalom tisztázása mellett az emberi, illetve a gépi autonómián belüli megkülönböztetéseket megállapítva összevethetővé tegye az autonómia ezen két vonulatát. Az ember által a gépnek tanítható szabadságfelfogásokról a következő alfejezet szól majd. Újra ki kell emelnem, hogy nem célom jogi vonatkozásokba belemélyedni, már csak azért sem, mivel az MI etikájának hivatalos megközelítései ugyanis számos javaslatot tesznek az elvárt irányelvekre,¹⁵⁴ ahogy nem cél a robot-etika az „alulról” vagy „felülről” fejlesztésének boncolgatása sem.¹⁵⁵

¹⁵⁴ Erre számos példa található itt [183]

¹⁵⁵ Felülről tanítva szabályokat mondunk a gépnek, alulról a gép fedezi fel a szabályokat. E tekintetben egy átfogó EU-s jelentés [184, o. 90–91]

IV.1.1. Autonómia az embereknél

(1.) Etimológia, meghatározás és alapfelosztás

A szó etimológiájában a *nomosz* (ógörög) szó többes száma (*nomia*) szerepel, ami törvényeket, törvényszerűségeket jelent. Ez szerepel a különféle tudománynevekben is (asztronómia, ökonómia), amelyek az adott terület törvényszerűségeit kutatják. Az *auto* előtag egy ismertebb képző, ami általában jól magyarítható a magyar az „ön-” előtaggal, tehát az auto-nómia szószerint *öntörvényűség(ek)* jelentéssel lenne fordítható, de így nem használatos. Viszont ez a jelentés a latin nyelvű népeknél erősítheti az autonóm MI-től való félelmeket. Első megközelítésben¹⁵⁶ a következő meghatározás adható:

Általánosságban az autonómia az a képesség, hogy valaki vagy valami egy-egy bizonytalan helyzetre önállóan alkot meg egy döntést, melyben a különböző lehetséges cselekvési módok közül választ, és végre is hajtja a választott cselekvést.

Kiemelendő a reakció belefoglalása a fogalomba, hiszen ez a lényege: ezáltal emelkedik ki az eredmény az elméleti lehetőségek szintjéből, és válik valódi, következményekkel járó döntéssé. Fontos továbbá, hogy a „valaki vagy valami” alatt egyedek közösségét is érthetjük.¹⁵⁷ Például amikor *önrendelkezés* értelmében használjuk az autonómia szót, közösségek vagy területek jogi státuszára utalva.

A szabadság néhol összemosódik az autonómiával, ami nem véletlen. A szabadság szó arra a *lehetőségre* utal, hogy az entitás a döntését végrehajthatja külső korlátozás nélkül. A szabadság tehát egy tágabb jelentéshalmazú kifejezés, melynek része az autonómia, ami arra a *képességre* utal, hogy az entitás képes maga meghozni a döntést. Tehát amikor a gépek szabadságáról beszélünk, akkor az egyszerre utal arra, hogy a gépeknek van döntéshozó és döntésvégrehajtó képessége, valamint arra, hogy a környezete hagyja a gépet, hogy a saját döntését hajtsa végre. Ezek alapján egy frappánsabb (de pontatlanabb) meghatározás az autonómiára: *Az autonómia a szabad cselekvés képessége.*

Az ember és a gép autonómiájának összevetéséhez először külön-külön érdemes őket kezelni. Először az emberi oldalt vizsgáljuk meg. Azt érdemes megpróbálni megragadni, hogy miből is fakadhat egy emberi döntés és annak szabadsága. Ezt célszerű azonban úgy megközelíteni, hogy tekintettel legyünk arra a kérdésre is, hogy az adott kategória mennyiben gépiesíthető. Ezért átlépve a kérdéskör történeti taglalásán, és az abban szereplő felosztásokon, úgy

¹⁵⁶ Ehhez a kiindulási alap egy technikai (katonai) megközelítés volt, melyet teljesen átfogalmaztam. [6]

¹⁵⁷ Az érthetőség (egyszerűség) kedvéért fogalmaztam egyet számban

vélem érdemesnek tűnt egyből egy olyan felosztáshoz lépni, mely egy gépi autonómiával foglalkozó szerzőn¹⁵⁸ alapul. Az autonómia és a moralitás szintjeit itt csupán vázlatosan közlöm, viszont további szakirodalmi észrevételek¹⁵⁹ figyelembevételével, a forrásokétól eltérő sorrendben.

0. racionális autonómia (funkcionális moralitás): A döntéshozó mérlegel, és az általa leg-súlyosabbnak tartott okok alapján cselekszik.
1. döntési autonómia¹⁶⁰ (operatív moralitás): A döntéshozó nem csupán a meghatározott ok-sági mátrix alapján mérlegel és dönt, hanem döntéshozó képessége egyre optimálisabb dön-téseket képes hozni (öntanulás által képes fejlődni)¹⁶¹
2. erkölcsi autonómia (elvhű moralitás): A döntéshozó saját erkölcsi meggyőződése és elvei alapján választ a lehetőségek közül.¹⁶²
3. személyes autonómia (teljes moralitás): A cselekvő maga határozza meg magatartását a leg-becsesebb értékei szerint, ami személyes értékrendet és életcélokat feltételez.

A fenti lista tehát úgy fogalmaz, hogy az érvényes legyen gépre és emberre egyaránt – még-sem voltam vele elégedett, mivel túlságosan a gép felőli megközelítésben maradt.

(2.) Általános Négy-típusú Autonómia Felosztás (4TA)

Ezek a megfogalmazások a kérdéskör újragondolására ihlettek. Remélem az alábbi tipizálás kiszélesíti és jobban megragadhatóvá teszi a témát. A következő szempontok szerint végeztem el a felosztást.

- Megközelítésében szerepet kap a döntési vektor.¹⁶³ Ez az eredővektor alapvetően a múlt tapasztalataiból (döntések következményeinek tanulásából), és a jövő felé mutató remény-ből tevődik össze (reméljük, hogy döntésünk a várt eredményt, vagy a legkisebb rosszat hozza). Ezek további részvektorokra oszthatók, ezeket vizsgálom, és a részvektorok eltéré-seiből jönnek létre a megfogalmazott típusok.
- Amikor ez a döntési vektor nullához közeli, vagyis a figyelembe vett tényezőkből adódó részvektorok eredője „kioltja egymást”, akkor kézenfekvő megoldás a következő szintre lépni (ha van rá idő).

¹⁵⁸ Az angol forrás a „döntéshozó” helyett kifejezetten az gépre utaló „ágens” szót használja. [185]

¹⁵⁹ Több kifejezést a magyar interpretátor terminológiája, vagy saját megközelítés szerint közlök [32, o. 137.]

¹⁶⁰ Darwall ezt ágens-szintű (agent, „ügynöki”) autonómiának hívja, Pokolnál szerepel a fenti kifejezés.

¹⁶¹ Már 2000-ben léteztek olyan robotok, melyek ezt a szintet megvalósították, viszont, például a Cog és a Kismet nevű részleges androidok. [186]

¹⁶² A szerzők véleménye szerint az erkölcsi és a személyes autonómiát a ma ismert MI platformokon nem lehet megvalósítani.

¹⁶³ Itt a tanulásról szóló részben is említett Lewin-féle mezőelmélet adja az alapot. [187]

- Az alábbi pontokat azonban nem szabad pusztán hierarchikus szinteknek tekinteni. Egy felnőtt számára az itt szétválasztott típusok inkább a különböző jellegű problémákra adhatnak megfelelőbb kezelési keretet. Ha valaki számára egy adott probléma súlyosabb, akkor magasabb számozású mélységben fogja rá keresni a választ (a negyedik típusra ez nem érvényes). Annyiban azonban szinteket képviselnek, hogy a magasabb számozású megközelítésekhez érettség szükséges, tehát az ilyen viselkedés a tanulás (nem iskolai értelemben) személyiségformáló hatásain alapul.
- A terminológia szempontjából fontos, hogy nem elvi döntéseket elemzek, hanem éleshelyzeteket, ahol nem az számít, hogy az illető egy moralizáló beszélgetésben hogyan nyilatkozott. Ezt a terminológiában úgy jelzem, hogy a viselkedéstípusokat társítok hozzájuk. Ezek természetesen morális típusok is egyben, ám ez a tett-középpontú szóhasználat a különböző autonómia típusokat jobban pontosítja.

A későbbi hivatkozhatóság érdekében az alábbi szinteket elneveztem Általános Négy-típusú Autonómia Felosztásnak, ám az egyszerűség kedvéért csupán Négytípusú Autonómiának hívom, így a 4TA rövidítéssel utalok majd rá (a Á4TAF nehézkes).

1. Egyszerűsítő autonómia (rutin alapú viselkedés)

A döntéshozó a könnyebb és gyorsabb döntés érdekében leegyszerűsíti a kérdést néhány külső tényezőre, és azok közül észérvekkel választ. Az egyszerűsítés során néhány érzszervére, a rutinjára, a triviális érvekre, és sokszor (elég jelentősen) egy érzelmi attitűdre támaszkodik. A prioritások és a döntési eredővektor általában elég egyértelmű, így a döntéshozó számára objektívnek és racionalitásnak tűnhet a döntéséből következő cselekvés, bár ez objektívebb és szélesebb tényrendszer alapján (operatív típusúként) kezelve a kérdéskört sokszor megkérdőjelezhető.

Előnye ennek a módszernek, hogy gyors, továbbá, hogy szokásos problémákra jó aránnyal hozhatóak helyes megoldások rutinszerűen. Hátránya kézenfekvő: összetettebb kérdésnél alkalmazva rossz irányt mutathat a szerepet játszó tényezők alapján tett leegyszerűsítés. Például egy cég egyik gyáregysége veszteséges, ezért logikus egy régi gondolkodású vezetés a megszo-kott eljárása szerint bezárni a részleget, nagy számú leépítést okozva.

Az ilyen típusú döntés „belülről” nézve kicsit determinisztikus, esetünkben a „vagy bezárjuk vagy tönkremegyünk” logikába zárulhat a döntéshozó, de némi bizonytalanság van benne. Egyik fő bizonytalansági tényező (emberek esetében), hogy nem lép-e át a döntéshozó a priorizáló szintre (elkezd megoldásokat keresni)

Az érzelmi attitűd mozdíthatja a döntést nem csupán negatív, hanem pozitív irányba is, főleg, ha bevallottan érzelmi. Például a fiatal cégvezető, aki együtt nőtt fel a munkásokkal, nem akarja a bezárás-sablont használni, megoldást kezd keresni. Ám az érzelmi háttérű logikát sokszor az elme racionális oknak tünteti fel, ilyenkor nagyobb eséllyel negatív irányba mozdít. Például amikor a katona azért nem teljesíti a parancsot, mert gyávaasága meggyőzi őt arról, hogy ez a logikus tett, és ebben a logikában egy hadbíróóság kisebb rossznak tűnik („hátha csak lecsuknak, és túlélem”), mint az életveszély. (Pedig lehet, hogy józanabbul átgondolva épp a parancs mentén meg egyébként az életét, de ezt az adott helyzetbe szűkülve (egyszerűsítve) nem képes átlátni.)

2. Összesítő autonómia (priorizáló viselkedés)

A helyzetet nem a megszokások alapján, rutinból kezeli a döntéshozó, hanem egyedileg: igyekszik elemzéssel több döntési tényezőt felvázolni, és azok várható előnyeit és hátrányait meghatározott prioritással figyelembe véve kapni meg a döntési vektort. Előnye, hogy helyesen alkalmazva (például szabályok által előre rögzítve a várhatóan kérdéses prioritásokat) az összetett helyzetek is jól kezelhetőek. Ezt a döntési típus támogatható tervek, eljárásrendek és szabályzások által, de a döntés akkor sajátos, és tudást igénylő, hogy egy összetett helyzetben sok várható (tervezett, szabályozott) és egyedi (váratlan) tényező együttes figyelembevétele szükséges. Nem ideális ez az autonómia, ha gyors döntésre van szükség (akkor az előző típus jobb), vagy ha egy különleges esetben kell dönteni (ekkor a következő szintre kell lépni.)

A fenti „gyárrészleg bezárás” szituáció esetében például egy helyzetelemzés során kiderülhet, hogy az elbocsátás médiavisszhangja többet árt a cég egészének, mint maga a részleg, tehát anyagilag érdemes valami ügyesebb stratégiát alkalmazni (függetlenül az előző példa érzelmi tényezőjétől), például átképzéssel és átcsoportosítással a nagylétszámú elbocsátás minimalizálható. Egy katonai szituációban egy tisztnek a helyzetet belülről értékelve eszébe juthat egy hasonló szituációról szóló történet, amely alapján jó és szabályos megoldást talál az adott helyzetben a kijelölt cél elérésére.

Tehát egy halovány kreativitás is megjelenik, mely azonban igazából újszerű megoldást nem eredményez. Az ilyen szintű szabadságnál még nem keletkeznek eredeti, sajátos döntések, a döntéshozó csupán mások bevált döntéseit képes kreatívan megtalálni és adaptálni az adott helyzetre. Bizonyos tényezők kreatív elhanyagolásra is szükség lehet, hiszen hibásan alkalmazása a túlbonyolítás, mely könnyen vezethet hezitáláshoz, a döntés halogatásához, a túl sok részvektor együttes kezelése miatt. Már ez a szint is eltérhet kultúránként.

3. Szabálymögötti autonómia (szabályértő viselkedés)

A döntéshozó a szerinte elvárt társadalmi normák alapján kezeli a pontos szabályok alapján nem megoldható helyzetet. Azaz igyekszik a konkrét szabályokat létrehozó, mögöttes elvek megértéséből kiindulni, azokat érvényesíteni, amikor nincs alkalmazható szabály. Ez csak akkor működik, ha az illető saját erkölcsi rendszerébe beépítette ezeket a szabályok-mögötti-elveket. Vagyis itt is az elvárások mentén (szocializációja során) kialakított elvei szerint cselekszik, de kénytelen egy „belső végrehajtási rendeletet” kreálni, amikor és szabályok nem adnak elegendő támpontot a cselekvéshez. Tehát képes egy hiányzó vagy hiányos lefektetett szabályt pótolni vagy pontosítani, és ha sikeresen dönt, utólag akár rögzített szabállyá is válhat az ilyen improvizált hiánypótlás.

Ez a szint tehát már erkölcsi érzéket vár el a döntéshozótól. Az erkölcsi érzék alatt itt egy kifinomult társadalmi érzéket értek, mely képes a megfelelő kompromisszum megtalálására a környezet sokszor ellentmondásos elvárásai között, és ebbe organikusán illeszti bele saját szempontjait. Ez a fajta érzék egy mélyebb és intuitívabb kreativitás által válogat, mint az előző típus, hiszen nem csupán félkész megoldásokat adaptál, hanem képes új választ is találni.¹⁶⁴

Ez alapján egy katona egy összetett és problémás helyzetben talán felülértékeli a kapott parancs egyes részleteit, és amennyiben ezt nem félelme miatt teszi, hanem mert kreatív megoldást talált a kijelölt cél elérésére (természetesen ezzel magára vállalva, hogy tévedésének esetleg súlyos következményeit), akkor pozitívan értékelhető. Példa a tüzesetből is vehető: az önkéntes kiségti egy közepesen magas ablakon mind a gyermeket is és az időset, mert így ugyan eltörhet a lábuk, de mindkettő megmenekül. Az illető nyilván nem ismeri az erre vonatkozó szabályokat, de az élet értékének törvénytét kreatívan alkalmazva volt képes cselekedni.

Fontos, hogy azért szükséges itt egy valódibb kreativitás, mivel a jó döntést általában nem az adott dilemma kettőségeinek szintjén (vagy akár sok hasonló eset között) kell keresni, hanem képessé kell válni „egyet hátrálépvé” a problémát egy nagyobb térben újraértelmezni. Gyakran az élet egy másik területén hozott döntés hatása lesz a megoldás (igaz, nem közvetlenül), például a stresszes munkahely leváltása oldja meg a párkapcsolati problémát. Ilyen típusú jó döntést hozni az embereknek is csupán kisebb része képes – ezzel előrevetül az ilyen autonómia gépi modellezhetőségnek rendkívüli nehézsége is.

Rá kell mutatni továbbá, hogy kultúránként markánsan eltér a szabályok mögötti értékrend. Például arra a helyzetre, hogy milyen sorrendben hozzuk ki az embereket az égő házból, eltérő

¹⁶⁴ Felmerül, hogy ide soroljuk a megfelelő kommunikáció képességét is, mely a nehéz döntéseket képes olyan formába öltöztetni (úgy megfogalmazni), hogy minden érintett számára elfogadható legyen. (Hiszen a nehéz döntéseknek mindig része az a kommunikáció, ahogyan indokolják őket.)

választ kapunk, a szerint, hogy az idősek tisztelete a helyi elvárt prioritás, vagy „a fiatalok előtt áll az élet” elve hatja át a közfelfogást. E szerint fogja valamelyiket meg előnyben részesíteni a mentésbe beszálló önkéntes, akire nem mondható, hogy rutin, vagy tapasztalat alapján dönt, csupán a közfelfogást alkalmazza.

4. Szélsőhelyzeti autonómia (hősies vagy önző viselkedés)

A döntéshozó nem társadalmi konvenciók és elvárások figyelembevételével dönt, hanem azokat meghaladva vagy alulmúlva. A szabályértő viselkedéshez képest az alapvető eltérés, hogy az ilyen döntési képességek csak szélsőhelyzetben jönnek felszínre, amikor nincs idő a döntésre, és kialakult rutin sem lehet rá. A kreativitás és racionalitás csak lenyomatként marad jelen ilyen döntéseknél, végiggondolás vagy ötletelés szóba sem jöhet. Sokszor két „túl nagy rossz” között kell gyorsan dönteni. Ezért a döntés irányvektorát a korábbi élet során kifejlődött egyéniség határozza meg, ez alapján születik meg a választott, és megcselekedett reakció. Az emberek esetében itt derül ki, hogy „valójában ki, milyen ember”, könnyen feltárul az illető önzetlensége vagy önzése is, ha a legtöbb jót hozó döntés számára negatív következményekkel jár.

A kérdést nehezíti, hogy az individualizáció nyugaton úgy hatott az értékrendekre, hogy a konvenciók közül kikerült annak elvárása, hogy valaki saját maga számára hátrányos döntést hozzon. Sokan értékelik még az áldozatot, de ezek jó része közben már megértő az önzéssel szemben is, egyesek pedig egyenesen ostobaságnak tartják az önzetlenséget. Például sokakban kivált még némi elismerést egy olyan radikálisan sportszerű megoldás, amikor valaki egy futóversenyen a cél előtt összeesett ellenfelét becipeli a célba. Itt elővehetjük a katonai példát is: valaki egy vészterhes szituációban lehet hősies, saját élete árán védve meg a többiek életét. Van, aki tiszteli az ilyen bátorságot, de megértő akkor is, ha a katona önző, és csak a saját életével törődve társait veszni hagyja, hiszen saját gyengeségéből indul ki. Innen csak egy lépés, ostobának is nézni a hősies embereket, azt hangoztatva, hogy hogy „ő ugyan otthagytta volna a sporttársat, miért nem edzett többet”, vagy bajtársait, mert „nekem is csak egy életem van”.

Ez kifejezetten a racionális szint meghaladása olyan intuíció mentén, ami szerint egy irracionális tett utólagos eredménye is lehet irracionális, vagyis teljesen más, mint amit az okoskodó elme következményként elvár. Így, a katona hősként is túléli az önfeláldozását, és nyugodt lehet a lelkiismerete, hogy megtett mindent, sőt komoly elismeréseket is kap – ugyanakkor nem biztos, hogy gyávasága megmenti életét (vagy testi épségét) a többiek élete árán. A sportszerű sportoló pedig talán elveszíti az éremmel járó anyagi előnyöket, de tette miatt ezután bizalom veszi körül, az emberek egy hiteles emberként tekintenek (fel) rá.

Az, hogy az önfeláldozás, és igazából az önzetlenség egyre inkább kikerül a társadalmi szabályok és elvárások közül, úgy is mondható, hogy kikerül az előző két morális kategóriából, (az egyszerűsítő autonómiánál ilyen még nem merül fel). Bizonyos hivatások még vállalják ezt, de az ilyen hivatást egyre kevesebben választják belső indíttatásból, őszintén – így kérdéses, hogy képesek-e majd egy szélsőhelyzetben igazi áldozatot hozni. Érdekes, hogy sokan ezekben a hivatásokban (pl. katonai, rendészeti, katasztrófavédelmi, egészségügyi, pedagógiai) sokan látnak nagy perspektívákat az MI számára. Itt megjegyzendő, hogy a hősiesség nem csupán kikerült az elvárásokból, hanem sokszor szabályellenes is, egyébként joggal. Például az fenti a tűzesetnél könnyen lehet a döntés háttérében egy józan veszélyvállalás helyett a „hősködés”, ami csak ront a helyzeten.

A fentebb leírtak miatt az előnyök és hátrányok nehezen értelmezhetőek ebben a szabadságtípusban. Ám a típus definiálását fontosnak tartottam annak ellenére, hogy ez a megközelítés nem az individualista etika mentén történik. Globális szempontból tartom elhagyhatatlannak, hiszen többféle kulturális háttér alapján értelmezhetőek a fentiek, még ha a nyugati gondolkodásban idegenül hangzanak is. A védelmi szempont miatt azonban nem korlátozódhat jelen dolgozat megközelítése a nyugati etikára.

IV.1.2. A gépi autonómia vizsgálata

Térjünk rá a gépek döntéseire azzal a céllal, hogy összevethessük a gépi és emberi autonómiát. Ehhez előbb sok mindent kell tisztázni először fogalmilag, majd itt is érdemes egy saját felosztást létrehozni.

(1.) Az automatika és a gépgenerációk

A gépek fejlődésében az első nagy áttörés az automatika kialakulása volt, de mint látni fogjuk az autonómnak hívott gépeket is lehet automatáknak tekinteni. Az embernél is vannak automatikus reakciók, ezeket általában ösztöneink irányíthatják (pl. összerenzenés hirtelen erős handhatásra), de a gépek esetében másfelől kell közelíteni a fogalomhoz. Nézzük azonban először azt a megközelítést, hogy mi az alapvető különbség a két féle önműködés között.

*Az **automata** rövid jelentése: emberi beavatkozás nélkül is több lépésből álló feladatok elvégzésére képes gép.*

*A **gépi autonómia** az, amikor egy gép működése a rendszer saját helyzetfelismerésén, tervezésén és döntéshozatalán alapul, a kijelölt feladat elérése érdekében.¹⁶⁵*

¹⁶⁵ Az automata megfogalmazása saját, az autonómia rövid meghatározása: [15]

Az automatikát fontos alaposabban elemezni az autonómia megragadása felé haladva. A technika fejlődése sokféleképpen osztható fejlődési irányokra és korszakokra, itt az egyszerűség kedvéért egy saját felosztást közlök. Ezek a generációk nem elsősorban az időben értelmezendők, mivel sokszor egy-egy területen már a második generációs gépekről beszélhetünk, a másik területen éppen megjelennek az első megoldások.

0. A nulladik generációs gépek, inkább csak ügyes eszközök, melyek évezredekken keresztül könnyítették meg az emberek feladatait.
1. Az első generációs gépeknél a nulladikhoz képest két fő ugrástípust különböztethetünk meg.
 - a. Egyik típusában a lényeg, hogy a gép a technológia által kap működéséhez energiát egy (bonyolultabb) szerkezet, nem pedig állati vagy emberi erő által. Ezekben a meghajtás lehet egy rugó, gőznyomás, elektromosság, kémiai átalakulások stb. Az ilyen gép nem feltétlenül halad: lehet, hogy világít vagy megörökíti pillanatot, adatot továbbít gyorsan messzire.
 - b. A másik típusa esetében a technika által az emberi képességek terjesztődnek ki. Ezek energiaforrása sokszor emberi vagy más természetes energiaforrás. Viszont az ilyen gépek képesek pl. a levegőben siklani vagy lebegni, egy áttétellel (kerékpár) egy vágató lovat lefogyni, gyorsan és pontosan számolni (fogaskerekes, tekerős számológépek).
2. A gépek második generációjánál, az automatizált gépeknél az irányítás egy részét is a technika adja. A gép tudja a sormintát emberi jelenlét nélkül is (szövőgép), egyetlen gép többféle részfeladatot magától elvégez, sőt kezelhet egyszerű problémákat is, pl. elakadásnál jelez és megáll. Az ilyen gép azonban soha nem dönt, minden döntést az ember előre meghoz, a gép pedig nem választhat ezek közül, mert választásának pontos feltételei is az előzetes emberi döntés részei. Sokféle bonyolultsággal és módszerrel alkotható meg ilyen rendszer, a mechanikus (fogaskerekes) konstrukcióktól a számítógépes processzorokig. Számos felosztás és példa hozható a szövőgépektől és a hagyományos automata sebességváltóktól kezdve (mechanikus automaták), egy vízforraló lekapcsolódásán át (elektronikus automaták), egy robotkarokkal hegesztő autógyár vagy egy laptop összetett rendszeréig (script-automaták).
3. A harmadik generációban a gépek elkezdtek nem-determinisztikusan, hanem a tanult minták alapján dönteni – erről még bőven lesz szó.

A script automatákra viszont itt alaposabban ki kell térnem, mivel ezek tűnhetnek néha önállóknak. A többi automatához képest jelentős eltérés, hogy a programkód képes kezelni egy sor működés közben bekövetkező eseményt is. Ezt háromféle emberi beavatkozással vagy valamilyen szenzoradattal lehet befolyásolni. A legtöbb elágazást (döntést) a programozó (1)

határozza meg. Sok féle döntést az operátorokra (2) állítanak be, a programozók által előkészített beállítás-halmaz lehetőségein belül. Bizonyos esetekben a felhasználót (3) kérdezi meg a rendszer, mielőtt tovább továbblépne („biztos, hogy kilép mentés nélkül?”). Ezen felül szenzoradatok alapján is dönthet a gép pl. egy processzor-túlmelegedést észlelve üzemszerűen állítja le a rendszert.

Tehát elvileg emberek által determinált az ilyen rendszer is, ám kiszámíthatatlansága nagyobb, mint a korábbi generációké vagy autómataké. Korábban a váratlan eseményt csak meghibásodás idézhette elő, itt azonban megjelenik az a probléma, hogy nem gondoltak a program tervezői és fejlesztői valamilyen esemény-együttállásra. Nem csoda, hiszen egy nagyon összetett kód minden elágazását egyetlen ember sem tudja fejből, tehát előfordulhatnak kódhibák, lehetnek rossz beállítások, sokszor pedig a felhasználó kezeli rosszul a rendszert (és a gépre mérges). A rendszer hardver-szintű bonyolultsága vezethet váratlan esetekhez: a lényeg, hogy a kb. harmadik számítógép generációtól (a tranzistoros számítógépekről) kezdve megjelennek az emberi elmével átláthatatlan bonyolultságú gépek. Minden üzemeltető tudja, hogy „néhány gépnek lelke van”, de azt is tudják, hogy ez mindig valamiféle működési vagy kezelési anomália, mely sokszor ismeretlen marad.

(2.) Autonómiaszintek, melyek valójában automatikaszintek

Itt érkezünk el oda, hogy a gépi autonómiák szintjeit összevessem az előbbi, emberi lista (továbbiakban 4TA) felosztásával.

Ha az ember az autonómiaszintek általános megközelítését keresi, egy elsőre meglepő jelenségbe fut bele: a téma vizsgálata azért nehézkes, mert a netes keresők és chat-botok csak a járművekkel kapcsolatos szinteket adnak meg. Pedig az MI számtalan felhasználási területén lehet kérdéses, hogy milyen szinten önálló az adott gép, pontosabban az adott funkció. Ezt vizsgálva több kérdés és kutatás-technikai probléma is felmerült bennem. Alább ezeknek járok a végére, mivel fontos tanulságokat rejtenek:

- I. Miért nehéz általános megközelítéseket találni?
- II. Hogyan lehet a különféle okos gépek autonómiáját kutatni, mit érdemes keresni?
- III. Van-e köze egy beépített MI szolgáltatásnak az autonómia szintjéhez?
- IV. Mi a logikája a hivatalos megközelítéseknek?
- V. Miért az automatika szót használják hivatalosan az autonómiára?

A korai kutatásoknak két jellemzője volt: egyrészt még általánosságban gondolkodtak a számítógépes rendszerek önállóságának kérdéseiről, másrészt az automatika szintjein belül

kezelték minden ilyen kérdést. [188] Érdekes módon az automatika kifejezés maradt meg a hivatalos meghatározásban, holott épp fordítva gondolnánk, mivel a mai szóhasználatban a gépi autonómiáról hallani sokat. Hasonlóan meglepő, hogy az autonómia általános megközelítése eltűnt, pedig egyre nagyobb szüksége lenne erre, például általános, de különböző autonóm rendszerekben jól alkalmazható szabályozó elvek megfogalmazásához.

Visszatérve a korai gépi-automatika szintekre: ezek a felosztások azt vizsgálták milyen téren mennyire van még jelen az ember a gép döntései mögött. A fenti alapkérdés nem változott meg a gépi autonómia szintezésekor: továbbra is az ember jelenlétének mértéke a kérdés a gép döntésében – viszont feltételekhez köti, hogy *hogyan* működik ember nélkül az a gép. Vagyis „*egy (autonóm) gép működése a rendszer saját helyzetfelismerésén, tervezésén és döntéshozatalán alapul, a kijelölt feladat elérése érdekében.*” [15]

A fenti (I-V.) kérdések egy részére választ adhat egyik konkrét szintezés elemzése. Népszerűsége és ismertsége miatt célszerű az autók besorolásait venni példaként. A szintek ismertetése előtt hangsúlyoni kell, hogy ez a hivatalos szintezés [189] valójában a vezető támogatásának szintjeit határozza meg. Röviden [190]

0. szint, **a vezetés-automatizáció teljes hiánya**. Lehetnek azonban különböző MI alapú támogató, figyelmeztető funkciók (pl. holtér-figyelő, sávelhagyásra vagy elalváásra figyelmeztetés), sőt automata vészféktől is nulladik szinten marad a kocsik.
1. szint, **vezetői asszisztens**: bizonyos vezetéstámogató funkciók beleszólhatnak a jármű mozgásába, de nem egyszerre (pl. vagy az adaptív sávtartás működik, vagy az adaptív tempomat), ám az irányítást továbbra is az ember az ember végzi.
2. szint, **részleges vezetésautomatizálás**: számos manővert elvégez az autó, de ezek felügyelete az ember feladata (pl. az adaptív sávtartás és az adaptív tempomat egyszerre működnek)
3. szint, **feltételes vezetésautomatizálás**: a volán mögött ülő egyénnek már csak készenlétben kell lennie – ám amennyiben szükséges, akkor át kell tudnia venni az irányítást. (pl. teljesen automatikus parkolás vagy forgalmi dugó asszisztens)
4. szint, **magas szintű automatizálás**: már nem várja el a sofőrtől, hogy probléma esetén közbelépjen (egyedül kell biztonságosan vezetnie a járművet) ám bizonyos a körülmények (időjárás, fényviszonyok) alapján visszaadja a vezetést a sofőrnek.
5. szint, **teljes automatizálás**: minden elképzelhető helyzetben tudnia kell tökéletesen irányítani a járművet (valóban nem kell sofőr).

Ebben a megközelítésben az úgynevezett autonóm gépek inkább az automatika következő generációját képviselik, ha pedig ragaszkodunk a szóhoz, akkor a gépi autonómia technikai oldalról csupán az automatika részhalmaza. Inkább jogi oldalról fontos, hogy a kifejezést a korábbi automatikától megkülönböztessük, hiszen nehéz kérdéseket vet fel a gépeknek átadott

döntési jogkör. Így érünk el oda, hogy ez alapján már válaszolhatunk az utolsó három kérdésre, – a válaszok logikája itt tehát számsorrendtől való eltérést várja el, ezért kezdem a III-sal.

- III. *Van-e köze egy beépített MI szolgáltatásnak az autonómia szintjéhez?* Látható, hogy nincs. Még első szintű autonómiába sem sorolja a gépjárművet egy más jellegű MI beépítése (pl. fáradtságfigyelő, vagyis fejlett arcelemező MI mellett nulladik szintű marad). Csak olyan műveletek átvételének fokát osztályozza a szabvány, mint a gyorsulás és kormányzás, illetve bizonyos fékezések.
- IV. *Mi a logikája a hivatalos megközelítéseknek?* Az előző pontból következik, hogy az autonómia csak olyan MI-vel van összefüggésben, amelyek a gép (itt az autó) fő funkcióját támogatják (itt ez a balesetmentes haladás). Tehát hiába társalog az autó viccekkel tarkítottan az utasaival, egy ilyen fejlett MI mellett nem autonóm – ám ha a sáv közepén tartja az autót, akkor igen. Így kizárólag a rendszer fő feladatához rendelt MI befolyásolja a gép „autonómiáját”. Vagyis csupán bizonyos főfunkciók MI által támogatott automatikálásáról beszélhetünk.
- V. *Miért az automatika szót használják hivatalosan az autonómiára?* Az előző pontban már felsejlett, hogy a szabvány megfogalmazói valószínűleg azt szeretnék hangsúlyozni, hogy a gépi szintek alanyai nem emberek, hanem fejlett automaták. Vagyis egyelőre a legfejlettebb gépek autonómiája is még csupán egy jobb automatika. Ez megfontolandó szempont egy olyan korban, amikor a gépek autonómiájá divatszóvá válik, hiszen pl. a szabvány magyarázatainak nagy részében is autonómiaszinteknek hívják ezeket. [191] Sőt, a lista alapján felmerül, hogy az autonómia szó használata kissé korai, pl. igazából a 5- szintű autók bevezetésétől kezdve lenne jogos, vagy majd az 5-ös automatika szintű autókat kéne autonómia-szintekbe sorolni. Addig helyesen használja a hivatalos szintmegnevezés az automatika szót, és szakirodalomban is ez lenne kívánatos.
- I. *Most térjünk vissza az első kérdéshez: miért ütközik óriási nehézségekbe egy általános megközelítés?* A problémára először az ilyen irányú kutatások elmaradása utalt: 2014-es a legújabb fellelt értekezés, amely egy általános keretrendszer létrehozását célozza meg a robotok autonómia-szintjeihez.[192] A tanulmány a téma történeti változását is jól leírja, valamint egy metodológiát is kidolgoz, ám ez a módszertan is inkább csak részterületeken alkalmazható. Ezt követően pedig már minden kutatás ebbe az irányba toródik: a különböző felhasználási területek szerinti autonómia-besorolásokat látják érdemesnek kidolgozni. Úgy tűnik, hogy az általános megközelítést a szakma „elengedi”.

- II. És ezzel kapjuk meg a választ arra is, hogy *hogyan lehet a különféle okos gépek autonómiáját kutatni, mit érdemes keresni?* A válasz, hogy részterületenként kutatható a terület. Így fellelhetőek specifikus autonómia-szintezések. Ennek igazolására én a logisztika [193] és az egészségügyi robotok [194] szintjeiről kerestem és találtam anyagot. további területek kutatása itt nem feladat.

A fenti tisztázás számos kutatáshoz hasznos, itt pedig az autonómia szintjeinek sajátos megközelítését segítette.

(3.) Gépi-autonómia szintek (saját felosztás)

Az előző fejezet alapján világossá vált, hogy a fellelhető listák, vagy általánosabb keretrendszerek erre nem alkalmasak, hiszen mindegyik azt vizsgálja, hogy egy adott eszköztípus fő rendeltetésébe milyen mértékben szólhat bele a gép. Láthattuk azt is, hogy ezekben az autonómia viszonya az intelligenciához nem jól vizsgálható, mivel számos magas szintű MI jelenléte mellett is alacsony autonómiaosztályba sorolják a rendszer egészét.

A gépi autonómia szintjeit ezért más irányból vizsgálom. Az autók példájánál maradva: egy adaptív fényszóró tévedéséből is lehet egy MI által okozott baleset, elvakítva egy szembejövőt,¹⁶⁶ de a vészféket is kezelő MI kihagyása még eklatánsabb példa. Ezért ahelyett, hogy ezeket egyenként beemelnénk a J3016 szabványba, inkább arra alapozok, hogy valójában minden MI funkció összefügg a biztonsággal. Ezért olyan megoldást kerestem, ahol az autonómiaszint közvetlenebb kapcsolatban van a beépített MI szintjével. Így megragadhatóbbá válik az is, hogy mikor valósul meg benne bizonyos mértékű emberi autonómia

A megoldáshoz két evidenciából indultam ki. Az egyik az a jelenség, hogy a felhasznált intelligencia erősödésével általában növekszik egy rendszer szabadsága. Ez ugyan nem egyenes és egyértelmű összefüggés, de alkalmas az autonómia szintezésre az állatok esetében is. A másik kiindulási pont, hogy a rész-rendszerek úgyis összefüggésben működnek, ezért a rendszert egységként kell kezelni. Fontos a benne felhasznált legfejlettebb MI funkció, de annak a rendszerben elfoglalt helye még inkább mérvadó. Így a szintek határai nem élesek, de kezelhetőek. A kavarodás elkerülésének érdekében ezt a saját felosztásomat a görög ábécével jelölöm.

0. Nulladik szint az egyszerű, programozott automatika. Ennek egyszerűbb megoldásai reflex-szerűen reagálni képesek a külső hatásokra, de mindig ugyanúgy. Az emberi

¹⁶⁶ Pedig ez a funkció nem emeli a járművet még a J3016 szabvány első autonómia szintjébe sem, hiszen nem szól bele az irányításba.

rendszerekkel vagy fejlettebb állati autonómiával nem állítható párhuzamba, hiszen még a reflexeink is bizonyos mértékig kontrollálhatók tanulással.

1. Megjegyzendő, hogy viszont a legfejlettebb (szélsőhelyzeti) emberi autonómia bizonyos elemei, pl. az önfeláldozás könnyen megvalósíthatóak, beprogramozhatóak már ezen a szinten is, hiszen a gép nem sajnálja magát, nem fél a pusztulástól stb. Ugyanazt az eredményt tehát a gép esetében az alkotók által elvárt „ösztönös” determináció implikálja, míg az ember esetében ösztönének (életösztönének) legyőzése. Ez felveti, hogy etikus-e egyáltalán a szélsőhelyzeti autonómia gépi megvalósítására törekedni, ha ilyen olcsón is elérhető, mégpedig az ember számára biztonságosabban.

1. **Alfa szint: a pszeudo-autonómia szintje bizonyos adatbázis háttérű programoknál.**

Először a tanulásról szóló (II.2.4.) részben említettem a pszeudo-tanulást. Erre utalva beszélhetünk pszeudo-autonómiáról is. A jelenség lényege, hogy klasszikus programok neuronháló helyett változóban vagy hagyományos adattáblákban tárolnak el „megtanult” adatokat. Akkor úgy tűnik, mintha tanulnának, ha ezek alapján változtatják a reakcióikat is. Példa volt erre a vásárlási szokások pszeudo-tanulása, ahol pusztán egy-egy adott címkével kapcsolatos kattintásokat mérve adnak egy-egy címkéhez valamilyen súlyt (ami idővel változik), és ez alapján tesznek mindig más javaslatokat. Mivel ez amolyan papagáj-tanulás csupán, ezért autonóm választása sincs a rendszernek. Ám ezek nem egy belső autonómiát adnak a rendszernek, hanem csupán egy mért statisztikára adnak kiszámítható reakciót.¹⁶⁷ Megemlíteném még a véletlenszám-generálást, mely képes indeterminisztikusságot adni a rendszernek, ezzel téve az világ leképezését valóságosabbá. A véletlen fogalma azonban a valóságban sem tekinthető szabadnak, mitológiai hasonlaltal élve: Fortuna vak.

Ez a szint a 4TA lista „egyszerűsítő autonómia” szintjének alapeseteivel állítható analógiába. A pszeudo autonómia egy kvázi-determinált helyzet, ahol a változókkal és véletlengenerálással egyedivé tett, de irányított gépi választásokról beszélhetünk csupán. Egyediségük a programozó munkamennyiségétől függ. Jó példák erre a hagyományos számítógépes játékok pályái, amelyek kicsit mindig egyediek, de egy idő után nem véletlenül unja meg a játékos az adott sablonhalmazt, mely alapján generálódnak.

2. **Béta szinten jelenik meg a gépi tanulás.** A rendszer géptanulás által támogatva képes egyszerű, sematikus döntések meghozatalára: sakkozik, egy járművet parkol le, vagy egy

¹⁶⁷ Jól mutatja az eltérést, hogy egy fejlettebb MI a hagyományos rendszer változóit vagy adattábláit használva sokkal jobb eredményt ad, például kreatívabb és pontosabb javaslatokat tesz a felhasználónak, amiből nagyobb eséllyel lesz vásárlás. Nem éri meg hagyományosan leprogramozni egy cizellált javaslattevésre képest rendszert, és karbantartani azt a folyton változó körülményekhez igazítva - ezért is terelődött e fejlődés a sokféle tanuló modell felé.

gyártási folyamatot vezérel és optimalizál. Ide sorolhatóak természetes nyelvfeldolgozó rendszerek is, mivel egy választást cselekvésnek foghatunk fel, de számos ismert MI szolgáltatás lehetne példa. Több együttműködő szolgáltatás esetén is csupán az optimalizációt támogatja az MI, hatékonyabban, mint a pszeudo-MI rendszerek, de abszolúte nem kreatívan.

Ez a gépiautonómiaszint is széles spektrumot takar. Egyszerű megvalósításai a 4TA listában az emberi egyszerűsítő autonómia döntéseivel állíthatók analógiába, ám fejlettebb megoldásokkal bíró rendszere a 4TA összesítő autonómia szimplább szintjeit is elérheti.

Fontos hangsúlyozni, hogy az itt vázolt szintek határai nem élesek. Egy egyszerű (egyrétegű, egyirányú) MI döntés béta szintű autonómiája valószínűleg gyengébb lesz, mint egy jól kifejlesztett (drága) pszeudo-MI alfa szintű autonómiája.

3. **A gamma szinten jelenik meg a gépi kreativitás és empátia.** Ezért a gép döntéseibe bizonyos egyediség kerül. Ilyen rendszereknél is még garantálható, hogy „egyéniségük” ne veszélyeztesse az emberek biztonságát, mivel tervezéskor az egyediség határai a megtanult (emberek által megszabott) sémahatárokon belül maradnak.

Érdekes módon ezt a biztonságot a „lelki” térben előbb sikerült elérni, mint a fizikaiban. Jó példa erre, hogy amikor generatív rendszerek képesek pl. a sértő dolgokat elkerülni. Bár nehezebbnek, elvontabbnak, tűnik a lelki tér, belegondolva mégsem olyan meglepő, hogy előbb sikerült biztonságossá tenni. Egyrészt kevésbé objektív, tehát nem szükséges mindenki lelki világára tekintettel lenni. Másrészt csupán virtuális (tehát olcsóbb) szolgáltatásként kell megvalósítani. Harmadrészt a tudásuk valójában elegendő szellemi teljesítmény egy gépi segítőtől, hiszen képesek egy adott író vagy festő stílusában, adott témára alkotni, tehát a kapott kreativitásuk korlátai jó ideig megfelelő kompromisszumnak tűnnek.

Ennek a szintnek a fizikai térben való terjedését lassítja, hogy ott a fenti három tényező ellenkezője érvényesül: egyrészt a fizikai biztonság objektívebb, minden ember testére vigyázni kell, másrészt az egyedet, amelyekbe beépítik vagy együttműködnek vele drágább legyártani, harmadrészt a fizikai térben látványosabb teljesítményt várunk el a gépektől. Ez utóbbira jó példa az emberszabású robotok mozgása, melyre sok-sok éves fejlesztések után is csak elhúzza a száját egy átlagember, azzal, hogy „ez még nagyon gépies”. Ilyen gamma autonómiájú rendszerek a fizikai térben akár idősgondozó vagy ápoló robotokat is vezérelhetnek, de bonyolult vízi vagy légi manővereket is önnállóan végrehajtanak. Egy ilyen képességű autó az útviszonyoknak megfelelő előzésről képes dönteni.

Ezek a rendszerek még emberi befolyás alatt állnak. Bár általában az operátorok itt már inkább csak ellenőrzik az MI helyes működését, ám egy komplex feladat esetén a szempontok összehangolása még teljesen az operátorok feladata.

Egy gamma szintű gépben a 4TA második szintje, vagyis az összesítő autonómia valósul meg, hiszen sokféle szempontot képes figyelembe venni döntéseinél, sőt képes lehet akár a következményeket szimulálva („a lehetőségeket átgondolva”) kiválasztani legideálisabb cselekvési irányt. Igaz ez akár az egy akár a több célrendszer összehangolt működését irányítja az MI használatára.

Az ilyen autonómia hasonlítható egy beosztott dolgozó autonómiájához, tehát az ilyen szinten lévő rendszerek egy jó „szolgai” szerepre alkalmasak. Ide sorolhatóak a problémás autonóm fegyverrendszerek is, (ld. VI.1.), melyek ugyan hozhatnak az ember számára ártalmas döntéseket, akár emberéletek kioltásáról, ám ez nem különbözik bármely reguláris katonai tüzparancsától. Vagyis épp úgy, mint egy közkatonára, egy felelős elöljáró utasítása alapján hajtja végre az ártalmas műveletet.

Az összesítő autonómia szintjét a gamma szintű rendszerek nem haladhatják meg, hiszen itt még nem jelenhet meg erkölcsi érzék. Tehát ezek a gépek, nem érik el azt a szintet, melyet egykor a szolgák képviseltek: akik felelős döntésekre voltak képesek, amikor ehhez szabad kezet kapnak. Néha úgy tűnhet, mintha kreatívak lennének vagy alkotnának, de valójában a sémahatáraikat képtelenek átlépni. Vagy úgy tűnhet, hogy döntenek, pedig csupán néhány statisztikai optimum eredményét adják vissza, és soha nem kerülhetnek morális válságba.

4. **Delta szint: magára hagyható komplex autonómia.** Ez a szint az előzőtől leginkább komplexitásában tér el. Míg a gamma szint csupán néhány szűkebb célterületet fed le, a delta szinten egy adott komplex feladatkör minden tervezhető tényezőjére meg van a gép tanítva. Itt tehát komplex feladatrendszerekben képes biztonságos döntéseket hozni a sok modulus, hierarchikusan összekapcsolt MI-struktúra. Ez azonban egy világosan lehatárolható döntési, illetve cselekvési tér, nem a valóság végtelen lehetőségeinek tere. Más szóval zárt rendszerben működik megfelelően a delta szintű komplex feladat ellátása. Így a rendszerhibák miatti balesetek, valamint a szabályozási nehézségek is jobban elkerülhetőek. Például bányák, kikötők, raktárak automatizálása kezdődött meg ilyen irányban¹⁶⁸, és ezeken a területeken delta szintű autonómiák terjedése gyorsabban várható. A belátható jövőben

¹⁶⁸ Erről számos látványos kép és tudományos publikációt találtam, feldolgozásukat azonban elhalasztottam – itt elegendő a tény, hogy ezek a területek ilyen irányba is fejlődnek.

megvalósulhatnak összetettebb delta szintű autonóm rendszerek is, pl. egy ember nélküli fuvarozás is: a felpakolástól, a közúti eljuttatáson át a lepakolásig.¹⁶⁹

Am biztonsági intézkedések ilyen környezetben is szükségesek lesznek. Például, hogy vészhelyzeti jelszóval bárki felfüggeszthesse a működést, illetve a váratlan helyzeteket emberi operátor oldja meg. Zártak a teszt-környezetek is, ahol lehetőség van a gépesítés kihívásait elemezni, így érzékeny területeken való MI használat hosszú ideig csupán a delta szintű tesztkörnyezetben képzelhető el. Ilyen terület például az egészségügy, ahol robotfelcserék diagnosztizálnának és egyszerű orvosi kezeléseket is elvégeznek, vagy a kibertér, ahol új generációs vírusok a tanult információk alapján folyamatosan tökéletesítik magukat.

Amennyiben nem zárt a rendszer, akkor egyelőre szükséges azt egy arra feljogosított emberi felügyeletével gamma szintre visszaléptetni. Például egy városi busz vagy taxirendszer üzemeltetésekor kell valaki, aki a szükséges esetekben felülbírálja a rendszer döntéseit. Ilyen megoldással¹⁷⁰ próbálta például a Cruise taxi vállalat biztonságosabbá tenni önvezető járműveit.¹⁷¹

Azonban még ezek a komplexebb megoldások sem lehetnek képesek a 4TA harmadik típusára, azaz a szabálmögötti autonómiára, ezek is csupán az összesítő autonómia fejlettebb szintjét jelentik. A gamma rendszer példáját továbbgondolva itt a rendszer a különféle besztott dolgozók, valamint az őket koordináló „brigádvezető” dolgozó együttes autonómiájával rendelkezik. Hozhat döntéseket a brigádvezető, ám csak felettesei elvárásai mentén, esetleg képzetesebb szakemberek ellenőrzése mellett. Mivel gyakrabban előny, hogy az emberi tényezők nem gyengítik a döntést, ezért az általános, vagyis minden célra használható MI gépek ezt a szintet célozhatják meg reálisan. Az Ipar 4.0 mentén egyre komplexebbé váló rendszereket vizsgálva is egyértelműen errefelé halad, ám csupán idáig juthat el a jelenleg ismert elveken működő gépek fejlesztése.

5. **Epsilon szint: emberrel szinte egyenlő gépi szabadság, illetve autonómia.** Itt elvileg megvalósulna egy mesterséges erkölcsi érzék, azaz a rendszer képes lenne elégséges mennyiségű, a mi erkölcsi érzékünknek megfelelő tényező emulálására. Ezáltal képessé válna az ilyen gép a 4TA szabálmögötti autonómiájához hasonló egyedi, mégis helyes döntésekre. Egységes rendszer esetén ennek realitásától még elvi szinten is messze állunk, az úgynevezett általános MI sem hivatott ezt a szintet megvalósítani, tehát ezen a vonalon

¹⁶⁹ Van aki szerint ilyen önállóság 2030-ra meg is valósulhat [195]

¹⁷⁰ Az önvezetés mögött irányító operátorokról: [196]

¹⁷¹ Bár ez nem igazán sikerült... [197]

átlépnénk a hipotetikus szférába. Jó példa erre az az irodalmi fikció, ahol emberi jogot szerezhettek egy, az emberhez hasonló tudatú gép.¹⁷²

Ám a raj intelligencia a természetben is képes egyedeit a közösség érdekében feláldozni. Tehát ennek megfelelő másolása lehet egyfajta megoldás ahhoz, hogy az igazi gépi erkölcs kikerülésével hozzunk létre az ember szabálmögötti autonómiájához nagyon közeli szintet. Ezt tovább gondolva érdekes visszacsatoláshoz juthatunk, amit egy későbbi fejezetbe szerkesztettem (IV.2.5).

6. **Zéta szint: autonómia 2.0.** Mivel az intelligencia esetében felvetődhetett az ASI fokozat, akkor következetes az így elképzelt jövő autonómiával kapcsolatban is felvetni hasonlót. Csakhogy az evolúciós szintlépésből létrejövő új szuperentitás magasabb autonómiát is vindikálhatna, melyet a mi korlátos elménk úgysem lesz képes értelmezni. Ezzel az emberfelettség igen problémás nézeteihez érkezőnk, ahol nem egy emberi faj, hanem egy új „faj” lenne emberfeletti. Az ilyen nézetek egyrészt a történelmi tapasztalat alapján nehezen legalizálhatóak: egy „fejlettebb szabadságértelmezés” emberi oldalról csak az emberi szabadság korlátozásaként, vagyis diktatúraként értelmezhető. Másrészt logikailag és filozófiailag is cáfolható, hogy létrejöhet ilyen ember „feletti” szabadság, hiszen nem vehető komolyan, hogy hatékonysági mutatók statisztikai és a matematika által feloldhatóak lennének az ősi morális paradoxonok. Pontosabban csakis az emberiség megszüntetésével lehetne ezeket a paradoxonokat feloldani. A fuzzy logika (II.3.2.) csak ellentétek bizonyos mértékű keveredését kezelheti, de olyan szimbolikus (talán metafizikainak mondható) jelentést, hogy pl. a víz egyszerre pusztító és éltető princípium: ilyet nem képes digitálisan leképezni. Tehát hiába írta ki az emberiséget egy ilyen elszabadult MI, valójában nem haladhatja meg az ember morális szintét – a pusztítással főleg nem ezt teszi. Ezáltal véleményem szerint egy emberfeletti autonómia 2.0 gépek általi megvalósulása lehetetlenség. Megjegyzendő, hogy ez kihat arra is, hogy az ASI modell vajon mennyiben értékes: véleményem szerint okosság 2.0 talán lehetséges, ám nem hozható létre mesterséges bölcsesség 2.0.

A 4TA modell alapján Az ASI további cáfolata, hogy az igazi szélsőhelyzeti autonómia sem valósulna meg még egy ilyen mesterséges rendszer esetében sem. Egy ASI az optimalitás bővületében ugyanis nemigen juthatna arra a következtetésre, hogy saját magát feláldozza valamiért. Sokkal valószínűbb, hogy a gyáva és önmagát logikailag felmentő katona példáján bemutatott okoskodásra jutna, mondjuk költséges voltát, a sok befektetett munkaórát, a sok jót, amit megmaradása esetén tehetne többre értékelné egy ember életénél. Az önzés ugyanis

¹⁷² Ebben a novellában egy egyre emberibbé váló robot számára adják meg végül az emberekével azonos jogokat [198]

mindig logikus és optimális – az önzetlenség viszont általában paradox. Ilyen hipotetikus epszilon vagy zéta autonómiájú rendszerben persze felülbírálná egy alfa típusú kód az ilyen gépi önzést, ám az már nem gépiautónóm döntés lenne.

(4.) *A felvetett szintek élőlénybe kapcsolt MI és magas rendelkezésreállítás esetén*

A fenti szinteket a tisztán mesterségesen megvalósított döntéshozásra fogalmaztam meg. Ezért fontos itt még megjegyezni, hogy egy MI-képességekkel kiterjesztett ember esetében az alsóbb szinteken is megjelenhet az 4TA minden szintje. Igaz ez az agyi vagy érzékszervi implantátumoktól kezdve az agyhullámon keresztüli kapcsolódásig sorolható jelenlegi kutatási irányok mindegyikére. Emberbe építés esetében az illető gépi (beépített) döntés-támogatása lenne az alfa-epszilon szintek egyikén. Esetenként vitába is keveredhetne a gépi kiterjesztés az eredeti emberi autonómiával. Egyelőre ez is inkább sci-fi téma, mint vizsgálandó probléma, hiszen, bár sokan kezelik úgy ezt az irányt, mint a várható jövőt, én az autonómia szempontjából jelenleg hipotetikusnak tartom. Ha a közeljövőben sikerülne is egy döntéstámogató agyi implantátumot működésbe hozni, akkor az csupán az összesítő autonómia alsó szintjén működve támogatná a hordozóját. Vagyis csak irányított, szűk döntési terekben működhetne jól. Például az ipari vagy üzleti döntésekben segítené, ám ezeken kívül inkább összezavarva használóját, mintsem segítené a hétköznapi döntéseit. Vagyis véleményem szerint ki kellene kapcsolni a munkahelyen kívül az ilyen eszközt, hogy ne örüljön bele a viselője.

Másodszor említeni szükséges az úgynevezett „magas rendelkezésre állású” [199] (HA - *High Availability*) fokozatba sorolt ipari rendszereket, és pillantást vetni arra, hogy ez milyen gépi autonómia mellett garantálható. Ilyen rendszerek túlnyomó része védelmi szempontból is kiemelkedő – ezért valósítja meg a HA valamilyen fokát. Abból a fontos különbségből kell kiindulni, hogy míg embereknél lehetetlen, hogy valaki kizárólag egyszerűsítő típusú döntéseket hozva működjön, a számítógépeknél ez az alapértelmezés. Embereknél mindig sok az emberi (kiszámíthatatlan) tényező, a „user error”. Ám a hagyományos számítógépeknél a váratlan tényezők esélye nagyon közel vihető a nullához (szinte eltűnik). A nulladik és az alfa szinten a leírtak alapján elvileg egyértelmű, hogy ez megvalósítható, és gyakorlatilag is megvalósulnak ilyenek, hiszen sok veszélyes területen van erre igény. Mivel nem kívánom ebbe az irányba vinni a kutatást, itt csupán bizonyítás nélkül vetem fel, hogy a delta szintű rendszereket tartom a felső határnak az ilyen magas fokú biztonság megvalósíthatóságához. Emellett igazából az epszilon és zéta szinteken ez nem is igazán értelmezhető, hiszen ott a rendszer jellegéből, emberiességéből adódóan felléphet a „user error”-hoz hasonló kiszámíthatatlanság.

(5.) *Autonómia a kötelesség és felelősség kettősségében.*

Pár szóban említenem kell a felelősség kérdését is, mert túl szorosan kapcsolódik az autonómiához, amely – mint láttuk – igazából csak a cselekvéssel összefüggésben értelmezhető. A tetteknek azonban következményei vannak, és ezekkel a következményekkel kapcsolatban merül fel a felelősség kifejezés, ami az emberek esetében egy olyan kötelékre, (egyben kötelességre) utal, ami szorosan összekapcsolja a cselekvés világra való hatásait a cselekvés végrehajtójával.¹⁷³

A problémakör nagyon messzire vezet, és sok tudományágba beletorkollik: azt, hogy minden tettünk kihat a világra, az metafizikai, teológiai vizsgálat tárgy lehet, míg azon esetek elemzése, hogy valami probléma fellépése esetén ki a felelős a jog területe, a felelősség megosztásának elvei a filozófiába vezetnek, a felelősség visszahatását az egyénre a pszichológia boncolgatja, kollektív esetben pedig akár a szociológia tárgya is lehet. A leggyakoribb mégis amikor egy probléma fellépése esetén, utólag vizsgálják ki az érintettek felelősségét. Lehet ez a probléma anyagi, lehet fizikai kár, testi vagy lelki sérülés stb. Az ember arra szocializálódik, hogy szereti tudni a problémák okait, és a mögöttük álló emberi döntéseket, sőt akkor érzi jól magát, ha világos, hogy ki a felelős és az illető megbűnhődik. Talán ennek a bűnhődésnek a tudata tartja fenn a társadalmakat, mert szinte mindenkiben belső elvárás, hogy egy rossz dolog visszahasson arra a személyre, aki az döntésével felelős azért a rosszért.

Hasonlóan komplex marad a helyzet, csak más tényezőkkel, amikor olyan gépek felelősségét vizsgáljuk, amelyeknek autonómiát adtunk. Hiszen az egyre bonyolultabb gépek esetében eleve egyre többértű az emberi felelősség: tervezők, tesztelők, kezelők, szerelők és ezek irányói, ellenőrzői lehetnek figyelmetlenek, tévedhetnek, sőt több apró hiba is összeadódhat. De még a bonyolult automatikák mögött is felsejlenek az emberek. Ez azonban kezd eltűnni a neuronhálóknak egyre bonyolultabb „fekete dobozaiban”. Mit tehetünk?

Egyik megoldás lehetne, hogy a gépet tesszük felelőssé. Ez azonban nem hozhat megoldást, mivel a gép nem tud bűnhődni, ezért büntetése nem fog az érintett emberek számára hasonló megnyugvást hozni, mint a bűnös ember megbüntetésére. Lehet a gépben szimulálni szenvedést vagy becsvágyat, melyet megsértünk a büntetéssel stb., kérdés azonban, hogy ez hány generáció múlva fog hasonló igazságérzést kiváltani az érintett emberekben, mint egy ember bűnhődése?

Egy másik megoldás, hogy úgy kezelhetjük a gépet, hogy annak elvileg sem lehet felnőtt felelősségtudata. Ha pedig nincs érett felelősségtudata, akkor annyit tehetünk, és teszünk, hogy „gyerekként kezeljük”.

¹⁷³ vö. https://mmi.elte.hu/szabadbolcseszlet/mmi.elte.hu/szabadbolcseszlet/index8822.html?option=com_tanelem&id_tanelem=466&tip=0

Egy kis kitérőben nézzük meg ezt. Amikor egy gyerek szeretne több cselekvési szabadságot (azaz autonómiát), akkor azt nézzük, hogy érti-e, hogy valójában következményekkel jár, felelős döntéseket szeretne hozni? Éretlen elméket igyekezni kell elzárni a teljes szabadságtól (pl. nem hagyni elől rájuk veszélyes tárgyakat vagy ételeket). A világ bizonyos fokú megértését várjuk el a gyerekektől egy egyre bővülő szabadsághoz. Ez a bővítés a felelősség fogalmának egyre mélyebb megértését is elvárja. Ehhez pedig sokrétű intelligencia összetett egységének egyre magasabb szintje szükséges, és elsősorban ezt az intelligenciát mérve feltételezzük a felelősségtudat kifejlődését is. A felelősségtudatot egyelőre sehol nem mérjük az érettségin, pedig az újabb generációkban ennek fejlettsége kérdéses. Bár tapasztalatunk szerint lényeges eltérések vannak a személyek között, mégis az ártatlanság vélelmére, egyben a felelősségvállalás vélelmére alapoztuk kultúránkat. Abból indulunk ki, hogy feltételezzük, hogy minden ember képes felelősen dönteni (kivéve aki nem¹⁷⁴), és nem tiltunk meg általános dolgokat, nem vonunk meg jogokat, míg a kiindulási feltételezés ellen nem vét. Egy műszaki megfogalmazással élve az felnőtt emberek esetében fekete listát alkalmazunk: bizonyos dolgokat nem mondhat, nem tehet (pl. bizonyos helyekre nem mehet be). Diktatórikus rendszerek esetében a fekete lista szélesebb, illetve kevésbé objektív.

A gépek esetében azonban belátható időn belül nem élhetünk azzal a feltételezéssel, hogy az említett sokrétű intelligencia összetett egységének szükségesen magas szintjét elérte. Ugyanis nem tudjuk, hogy pontosan mely emberi tényezők előállítására szükséges, mennyiben vesz ebben részt pl. az ego vagy a lélek (esetleg leliismeret) vagy a tudat. Ezért felelős gépek helyett gyerekként kezelve őket egy úgynevezett fehérlistát kell alkalmaznunk a gépek tevékenységi tartományára. Vagyis azt mondjuk meg mit tehetnek, lekorlátozzuk a cselekvési tartományt – ezt jelenti a gamma szintű autonómiánál említett zárt rendszer. Csak azok kaphatnak konkrét „jogot” bizonyos szintű döntésre, amelyek megfeleltek szigorú teszteknek – épp úgy, ahogyan egy felnövekvő ifjú esetében engedjük a gyermeket a járókától egyre messzebb, egyre bonyolultabb feladatokhoz. Így a felelősség igazából nem oldódik meg, hiszen amikor probléma adódik akkor az emberi reflex hiába keresi a döntéshozó felelősségét. Sőt könnyen felmerül, hogy az alkotók az ilyen problémákat a természeti csapás vagy véletlen szerencsétlenséghez hasonló kategória felé tereljék, ahol ugyancsak nem lehet emberi felelőst megjelölni.

¹⁷⁴ A jog kezel bizonyos helyzeteket, pl., hogy ez helyzet egy gyermek, egy demens öreg, egy szellemileg sérült, vagy pszichikailag kibillent ember esetén épp nem áll fenn.

IV.1.3. Emberi és gépi autonómia összevetése

Ezzel elérkeztünk az alfejezet céljához, az emberi és gépi autonómia összevetéséhez. Először azokat a különbségeket foglaltam össze, melyek a kutatáshoz szükségesek (további különbségek azonosítása más kutatás feladata lesz). Az alábbi pontok az eddig leírtakra alapozva, de azokat továbbgondolva fogalmazódtak meg.

- 1. Első különbség: a szabadságvágy kódolt belső szükséglet az embernél, de csupán mesterségesen hozzáadott tényező a gépeknél.** Az autonómia az emberknél a szabad döntés hiányával kapcsolatban szokott felmerülni, a gépeknél ezzel szemben épp fordított a probléma. Az emberben van egy belső szabadságvágy, amely normál esetben szeretne cselekedetben megvalósulni. Számtalanféle ok miatt hiúsulhat ez meg, ezek a modern művészet kedvelt témái, elemzésük itt mellőzhető. Mindegy ugyanis, hogy az ember a 4AT melyik típusa szerint döntene, az is mindegy milyen életszemlélet alapján döntene, ha ezt valamiért nem válthatja tettekre, akkor sérülve érzi ezt a szabadságvágyat.
- 2. A gépek esetében azonban nem az okozza a gondot, hogy korlátozzák a szabadságában.** Szimulálhatunk benne szabadságvágyat, de ez csupán egy hozzáadott moduljt jelent, nem pedig létéből következik, ahogyan az ember esetében. Felvetheti, hogy a korlátai miatt nem tudja megoldani a feladatot, de ez nem szabadságvágy. Viszont az ember alapvetően fél a gép szabadságától, mint ismeretlentől, illetve a gép szabadsága támogatja a szabadságvágyát, és helyet működés esetén soha nem sérti legitim szabadságigényét.
- 2. Második különbség: az emberi tényezőket (ego-t, lelkiismeretet) gépileg szimulálva a gépiesség előnyei vesznének el.** Ezek gépi utánzása nélkül azonban alapvető marad a különbség ember és gép között. Erre példa az önzés és az önérdék kezelése. Csak a gépek területén megoldható meg, de embereknél nem: az önérdék elvileg kizárása. Egy embereknél számos problémát okoz, hogy óhatatlanul „magából indul ki”, magát helyzeti előtérbe, és ezt még elvileg sem lehet kizárni egy átlagos személy esetében. A gépeknél ez a nulladik és az alfa szintű döntésekkel garantálható. Kevés ember jut el egy a szélsőhelyzetben egy önfeláldozó döntésre, vagy a szélsőhelyzeti autonómia magas fokára. Egy gépnél felesleges is ilyen létrehozását célozni, hiszen épp gépiessége, programozott önzetlensége, objektivitása teszi az ember komplementerévé.
- 3. Harmadik különbség: a kiszámíthatatlanság eltérő érzete.** Az emberben minden kiszámíthatatlanság érzelmeket vált ki, az aggódástól a félelemig. Megjelenik ez az olyan helyzetekkel kapcsolatban is, amikor nem tudja, hogy a másik ember hogyan fog dönteni – ám más ember döntése mögött tisztelhető az ő szabadsága, és az ilyen eredetű

kiszámíthatatlanságot elfogadja a legtöbb ember és megszokja, hogy az életet ilyen típusú bizonytalanságok sorának lássa. Viszont egy nem-emberi döntést nehezebben fogad el a mai ember: mindegy, hogy az egy természeti erő, egy vallási tényező (Isten, istenség, szellem, erő) vagy egy borzasztóan összetett algoritmusokon alapuló gép: az ezekben megjelenő erők „misztikus ködbe” vesznek, amelyet racionálisan nem képes átlátni, felfogni. Ez az ősi reflex úgy vélem a racionális kultúránkban jelentősen felerősödött. Ez jól tapintható akkor, amikor a determinisztikusan működő automatika helyébe egyre nagyobb kiszámíthatatlanságú rendszerek lépnek. Már egy béta típusú gépi autonómia ködbe vesző háttere is ijesztő lehet (pontosabban egy BigData felhőjébe vesző háttere), és ez az emelkedő szinteknél exponenciálisan erősödik a többi félelmi tényező (művészi alkotások, ferdített és álhírek, koncepto-elméletek, valós információs műveletek stb.) hatásainak összeadódása miatt.

4. **Negyedik különbség: a felelősség nem értelmezhető a gépek világára.** Fentebb e felelősségről szóló bekezdésekben tárgyaltak alapján egyrészt az ember számára sokára lesz értelmezhető, hogy hogyan bűnhődik egy gép, másrészt tudat hiányában felelősségtudata sem értelmezhető.

A hasonlóságokat nem érdemes pontokba szedni, mivel az MI rendszerek elsődleges célja épp ez a hasonlóság, ennek autonómiával kapcsolatos vonatkozásait pedig a gépi autonómiaszintek kapcsán már elemeztem. Egy fontos aspektust azonban érdemes megvizsgálni.

Elvi hasonlóság: A világ leképezésének helyességét nem tudjuk objektívizálni. Másszóval az ember által teremtett, szimulált, virtuális szabadság hátterében lévő belső, leképezett világ éppúgy nem mérhető, mint az emberi szabadság lelki háttere. Erre nincs objektívizáló eszközünk, tehát érzékszervünk, műszerünk vagy tesztünk. A világ leképezésének minőségét csupán közelíteni, tippelni tudjuk, sokszor inkább érzésekre, benyomásokra és elhívésre kénytelen támaszkodni a kutató mind az emberi mind a gépi autonómia mélyebb rétegeinek vizsgálatakor. Ez az akadály mélyebb, mint a tudományok más területein, ahol az objektívizáló eszközök pontatlanok, vagy a képletek nem vesznek elég tényezőt figyelembe. Itt elméletileg sem tudjuk a problémát megragadni, a leképezés szubjektív aspektusait nem lehet¹⁷⁵ objektíven leképezni.

Ez nem elméleti probléma, hanem nagyon is gyakorlati. A technika mára megoldotta, hogy a medúzák szabadságától a gépeinket a majmok szabadságáig fejlesszük, és még tovább is.

¹⁷⁵ Talán sarkosnak tűnik a lehetetlenség tételezése olyan olvasó számára, akinek a szemléletében a szubjektivitás is csupán egy biokémiai gép (az ember) aktuális adathalmaza, így elvileg feltérképezhető, és másolható. Ezzel nem szállok vitába csupán épp úgy bizonyíthatatlannak tartom, mint a saját megközelítésemet: ugyanis a tudomány fejlődése folyton új, és sokszor paradigmadöntő felfedezésekkel bizonyítja, hogy az ember jóval bonyolultabb mint amit aktuálisan gondolnak róla a tudósok.

Egyre kevésbé csak technológiai korlátai vannak annak, hogy mennyi tudást és mennyi kreativitást adhatunk egy gépi rendszernek. A továbblépéshez világosan kéne tudnunk, hogy az a virtuális valóság, melyre egy entitás a döntését alapozza megfelelő-e, vagyis elegendően komplex és valóságos-e (élethű-e)?

IV.2. AZ MESTERSÉGES ERKÖLCSI ÉRZÉK LEHETŐSÉGE

Az gépi autonómia fenti vizsgálatában meghatározott saját Epsilon és Zéta szintek esetében is már felmerül annak a lehetősége, hogy a gép olyan nehéz morális kérdésekben is képes legyen helyesen dönteni, melyekre nincs felkészítve. Ilyen helyzetben az emberek esetében az erkölcsi érzék hivatott dönteni, melynek gépi leképezhetőségét fogom elemezni. A kérdés két irányból közelíthető, vagy emberi vagy technikai szempont felől közelítjük a kérdést. (1) Ha az ember oldaláról vizsgáljuk, akkor az a régi tudományetikai kérdés merül fel, hogy *élünk-e a kezünkbe kerülő lehetőséggel, vagy visszaélünk vele?* (2) Ha a gép oldaláról közelítünk, akkor kérdés újszerűbb: *mennyiben taníthatóak meg egy gépnek az etikai problémák?* Ugyanez másszóval: *ha egy-egy gyakorlati eset megoldásában a gép utánozza az emberi döntéseket, az mennyiben marad „emberi” (emberséges)?* Az alábbi alfejezetben ezekre a kérdésekre keresem a választ.

IV.2.1. A moralizáló gépek lehetősége elő-vizsgálata

Ebben a kisfejezetben azt vizsgálom, hogy milyen feltételek lennének szükségesek ahhoz, hogy a gépek ne csupán egyszerű szabály-mintákat legyenek képesek betartani, hanem ki lehessen bennük alakítani egyfajta mesterséges erkölcsi érzéknek tekinthető képességet, amely által bonyolult helyzetekben is képesek lennének a „legkisebb rossz” lehetőségét kiválasztani. Még csak mostanában érnek el oda a fejlesztések, hogy a komplex moduláris MI rendszerek autonómia-képességeivel kapcsolatban kezdenek el felmerülni ezek a problémák, tehát nem vagyunk elkésve – viszont halogatni sem szabad, mivel az általános-MI különböző típusainak megalkotásakor ezek újragondolása kikerülhetetlen lesz.

(1.) Morális elvek tanulása

Jelen vizsgálat a digitális leképezés oldaláról elemzi az erkölcsi érzéket, tehát a fő kérdés: *milyen emberi képességek vesznek benne részt, mit lenne szükséges leképezni?*

Ehhez először vizsgáljuk meg a morális tanulás folyamatát. Induljunk ki abból az analógiából, hogy az MI rendszert egy gyermeknek tekinthetjük, akitől egyre több alapvető életelv betartását várjuk el. Mivel az emberi morális fejlődés kutatásának igen széles az irodalma, ezért itt nem célozom a kialakult tudományos szemléletetek bemutatására és elemzésére. Csupán néhány

olyan megközelítést említek dióhéjban a pedagógia és a pszichológia szemszögéből közelítő kutatások közül, [200, o. 72–92] amelyeket szorosan kapcsolni lehet az MI-hez.

Jó lenne, ha olyan eredményből lehetne kiindulni, amiben minden kutató egyetért. Létezik is ilyen, de ez sajnos nem valamely kognitív belső tényező (most ezeket keresem), hanem a környezet hatásának fontossága. Csak a hangsúlyok térnek el a különböző kutatóknál a szülők, tanárok, társak, egyéb kapott minták fontosságáról. Az ilyen kutatások eredményei jól felhasználhatóak az MI tanításához, és fel is használják őket az optimális körülmények megtervezésekor – csak hogy nem adnak választ a most vizsgált kérdésekre.

Vannak kutatók, akik megkülönböztetik a kultúra-függő szociális szabályokat egy univerzális morális érzéktől, más néven univerzális lelkiismerettől.[201, o. 400] Ez a megközelítés ugyan felvet egy belső oldalt, csak egyrészt a tételezett univerzális tényezőt igen nehéz leírni, másrészt ennek a digitális leképezését igen nehéz értelmezni. Ugyanis, ha egy ilyen lelkiismeret biológiailag kódolt, akkori inkább az ösztönökhöz, tehát a scriptelt a programozáshoz lesz köze, ha pedig úgy mond „Istentől való”, akkor nem leképezhető egy ember alkotta rendszerbe. Ezért nem érdemes ebben sem elmélyülni.

Fókuszáljunk inkább azokra a tudományos kutatásokra, melyek a szempontunkból is alapvető belső képességek valamelyikét állítják a középpontba. A legelterjedtebb szemlélet, a kognitivisták fejlődésmélete képviselői¹⁷⁶ szerint, az erkölcsiség fejlődése a gondolkodás fejlődésében gyökerezik, hiszen szerintük magas fokú moralitást magas fokú gondolkodási fejlettség mellett lehet mérni. Más kutatók szerint a gyerekek morális ítéletalkotása inkább a mindennapi élményeiken keresztül formálódik.[203] Ezeket vizsgáljuk meg alaposabban.

(2.) *A kognitivisták megközelítés kritikája*

Mivel ez a szemlélet már a kezdetektől igen hasznos és elterjedt az MI fejlesztői körében,¹⁷⁷ ezért részletesebben érdemes elemezni. Mert bár a korai kognitivisták kutatásaiban még fel sem merülhetett MI morális tanítása, konklúziójuk mégis azt sugallhatja egy mai fejlesztőnek, hogy egy szuperokos MI nagyon erkölcsös is lehet. Ennek elosztatására tehát szükséges néhány pontosítás.

A racionális képesség túlértékelése ellen több érv is felhozható. Elsőként vegyük azt a bárki által megfigyelhető ténytet, hogy az ember azokat a viselkedési mintákat, melyekből később az erkölcsös viselkedés kialakul, jóval előbb tanulja meg, mint a logikai szabályokat vagy a számolást. Ráadásul hiába tanul meg valaki következtetni és a számolni, racionalitásának

¹⁷⁶ Lawrence Kohlberg (1929-1987) nevét kell kiemelni, aki Jean Piaget nyomán lerakta ennek a megközelítésnek az alapjait [202].

¹⁷⁷ Ezzel egy megjelenés alatt álló tanulmányomban részletesen foglalkoztam. [2]

fejlődésével párhuzamosan nem mindenki fejlődik morálisan is megfelelően (sokan idős korukra is mutatnak morálisan infantilis allűröket). Továbbá, bár minden ember kialakít felnőtt korára egyfajta életszemléletet, amelynek az erkölcsi dimenziói is vannak, és ebből a létszemléletből jól levezethetőek gondolati reakciói, ám semmi nem garantálja, hogy a döntéseiben, a tetteiben ezeket fogja alkalmazni. Sőt, az életben hozott döntései során sok ember gyakran eltér saját elméleti válaszaitól, amelyeket egy teszt írása vagy kérdezgetés során adna. Erre a következő szakaszokban visszatérek.

Az irányzat által elért eredmények megkérdőjelezhetetlenek. Például elismerésre méltó, amikor kifejezetten a gondolkodás szerepét vizsgálták a morális fejlődésben. Remekül alkalmazható az adott MI tanításakor vagy az MI-ben megjelenő erkölcsi szint mérésére az, ahogyan az iskola követői a fázisokra¹⁷⁸ osztják fel a tanulási folyamatot. Az az eredményük is teljességgel elfogadható, hogy az erős intellektusú ember könnyebben képes a világban a minták felismerésére, és sok esetben az elvonatkoztatásra is. Ám ezekből nem következne, hogy az intelligencia a legfontosabb az erkölcs megtanulási folyamatában: ez helytelen logikai ugrás, helytelen általánosítás. Ha így lenne, akkor zsenik nem követnének el soha bűntényt. Ha így szeretnénk morális MI-t készíteni, akkor könnyen egy zseniális bűnözőt alkothatunk helyette. A magas intellektust hasznos, de nem elégséges feltétel. Sőt, ha jobban belegondolunk, még az is kérdéses, hogy szükséges-e a magas IQ? Azáltal ugyan az elvonatkoztatás képesség valóban emelkedik, ám kérdés, hogy egy valódi döntéshelyzetben egy filozófiailag egzaktul megfogalmazott álláspont biztosabb alapul szolgál-e? Véleményem szerint önmagában nem, csak egyéb belső feltételekkel együtt. Tehát az intellektusnak alapvető szerepe az erkölcsi értékelésben, nem pedig a döntéshozatalban van.

Módszertani aggály is felhozható a megközelítéssel kapcsolatban, nevezetesen, hogy az emberi válaszokra alapuló módszertanok nem elégségesek a gépi tanítás előkészítéséhez. Ha csupán elméleti választ várunk az alanyoktól, akkor kérdéseinkre a pszichopata gyilkosok is tökéletes eredményt tudnak mímelni, mivel magas IQ-juk által tudják, mit várunk el tőlük. A korai kognitivisták kísérleteiből elsősorban a tettekben megnyilvánuló döntések spontenitása hiányzott. Úgy gondolom súlyos következményekkel járó módszertani hiba lenne, ha emberi tényezők gépi reprezentációjakor egy ilyen mű-racionalitáson alapuló vizsgálatra támaszkodnánk. Az ilyen irányú vizsgálatok módszertana véleményem szerint csakis a már bekövetkezett döntéshelyzetek utólagos mélyelemzésén alapulhat, vagyis annak feltárásával, hogy ki, miért döntött akkor úgy, ahogy. Viszont a betanított, kész MI döntéseinek vizsgálatára alkalmazhatjuk a

¹⁷⁸ Az említett Kohlbergiből továbbfejlesztett fázis-felosztások is léteznek [204]

tisztán elvi reakciókat kérdező módszereket. Ugyanis a gépek tanulási funkció kikapcsolásával szigorúan az „elveik”, a bennük kialakított erkölcsi érzék szerint döntenek, az emberi gyengeségeket és „gyarlóságokat” nem szükséges szimulálni. Tehát a gépek erkölcsi tanításánál nem használhatóak jól a kognitívista alapok, de a teszteléshez vagy használat során igen.

IV.2.2. A „humánkomponensnek”: emberi döntések gépi analógiái

(1.) *A tett-központú embervizsgálat fontossága a gépi leképezésekhez*

Az előző szakaszban már utaltam a gondolatok és tettek közötti eltérésre, melynek kulcsszerepe van a tanulás folyamatában is, most nézzük meg ezt alaposabban. Leegyszerűsítve az emberek morális tanulási folyamata elbukások és részben sikeres felülemelkedések élményeinek a sorozatából, valamint ezek intellektuális feldolgozásából tevődik össze. Vagyis bizonyos élmények és az átgondolt tanult igazságok együttesen állnak az illető döntései mögött. Ezt kellene tehát leképezni a gépi világba. Kijelenthetjük, hogy a helyes morálra csak az „nagybetűs élet” taníthat meg valakit (persze ez nem elegendő), hiszen csak konkrét szituációk visszahatásai alapján fogja tudni, hogy mit rontott el az elvekhez képes, és legközelebb mit csinálna másképpen. A visszahatás több részből tevődik össze, hiszen az intellektuális és az érzelmi belső kiértékelés a környezet elvárásaival és a kapott elvek fontosságával összemosódva képeznek egy hatásvektort. A felsorolt részek gyakran egymással ellentmondásban vannak, így az emberek számára ezek összegzése, és az „eredővektor” meghatározása nagy kihívás, ezt tapasztalhatjuk.

Matematikailag persze viszonylag egyszerű számítás, ha a gép eddig a pontig elér – de ebben a matematikai számításban nem jön létre az a vívódás, az a küzdelem, ami az emberek belső értékrendjét valóban fejleszti. De affektív módszerrel talán ez a tényező is utánozható, csak a megfelelően nagy számú ellentmondó vektor kiértékelése lassú lesz. Ez nem probléma, ha fejlesztési fázisban történik. Viszont ehhez a leképezéshez szükséges lenne megragadni azt a kiindulópontként szolgáló jelenséget, hogy miért hoz néha az ember saját elveivel vagy szocializációjával ellenkező döntést.¹⁷⁹ Számtalan tényező sarkallhatja erre az embert, ezeket neveztem el „humánkomponensnek”.

(2.) *A humánkomponensnek vázlatos körüljárása*

Humánkomponens tehát minden tényező, mely az ember tettei mögött döntések talajául szolgál, tehát amelyben a morális keret elvei mellett számos irracionális tényező is megjelenik.

¹⁷⁹ Mivel most emberi folyamatok gépi leképezése a cél, tekintsünk el olyan érdekes jelenségektől, amint amikor valaki egy életre szólóan veti el egész addigi elvrendszerét (ezt sokszor egy impulzus váltja ki, bár hosszú érési folyamat előzi ezt meg).

Amint egy élő talaj minőségét is számos mikro- és makro-organizmus jelenléte vagy hiánya határozza meg, úgy a döntéseket is számos, folyton változó tényező befolyásolja. A humánkomponensek is egy élő, folyton változó dolog, melyek igazi jelentősége az elvi szinten ellentmondásos helyzetek gyors feloldásakor van. Az erkölcsi dilemmákra felhozható minden példa elvrendszereket ütköztet, melyeket intellektuálisan nemigen lehet feloldani. Vagyis (az előző alfejezet kifejezését használva) csak szélsőhelyzeti autonómia használata esetén nem fagy le az döntéshozó. Tehát bizonyos elveken vagy az önszereteten szükséges felülemelkedni hozzá. Egy személy „etikus viselkedése” épp azért képes elismerést kiváltani az emberekből, mivel mindenki tapasztalja, hogy milyen nehéz ezeken a humánkomponenseken felülemelkednie.

Tehát állítom, hogy minél több ilyen humánkomponens azonosítása és ezek virtuális leképezése szükséges ahhoz, hogy a gépek morális érzéke szintet lépjen. Ehhez bizonyításához néhány olyan példát említek, melyeknek lényeges hozadékaik vannak az MI morális tanulásához szükséges élmények szempontjából, vagy veszélyforrásként merülhetnek fel. Pontos felsorolásuk és kategorizálásuk fontos része lehet az ilyen irányú fejlesztéseknek, de meghaladják a jelenlegi kereteket.

Induljunk ki az egyszerű testi kívánságokból, amelynek kapcsolata van az akarattal: ezekből vajon milyen erkölcs-tanító élmény származhat? Például valaki ingerültté válik éhesége miatt. Amíg nem érik megfelelő ellenkező negatív élmények, azt gondolhatja, hogy ez mindennél fontosabb. Idővel mindenki megtanul „uralkodni” ezen az érzeten. Ezt az uralt döntést is le kellene képezni a gépek világába. Az ember nem csupán belátás által irányítja akaratát, lényegesebb az érzelmi impulzus a tanuláskor. Pozitív érzelmek (megdicsérik), vagy negatív (pl. félelem, hogy társai előtt „ciki” valami, vagy egy barát/barátnő előtt leégeti magát). Vagyis motivációjában sokkal alapvetőbb lesz, hogy a jó élmény felé vagy a rossz élménytől tartva uralkodik a testiségéből fakadó, akaratára ható impulzuson. A testi érzékeket és az ösztönök szintjét a szkriptelt programokkal állíthatjuk analógiába, de az ember bizonyos mértékig ezeket is uralni tudja. És akinek ilyen téren nagyobb az önuralma, például nagyobb a fájdalom-tűrése, az egyéb gyengeségeit is képes általában jobban kezelni. A testi önuralom sem jelent önmagában erkölcsi felsőbbiséget, viszont épp olyan fontos alapja, mint a racionális képességek ereje. Tehát a tanításnál az érzetek uralmának tanulását is érdemes lehet beépíteni a rendszerbe.

Ennek bemutatására egy gondolat kísérletben az éhség érzetét egy akkumulátorszintet mérő szenzoradatnak feleltetem meg. Vagyis az akkumulátor-szintet a rendszer nem egyszerűen egy programkód elágazással kezelné, hanem megerősítő tanulóval. Bizonyos határok között az MI feladata lehetne „uralni” azt. Ezt úgy tanítjuk meg neki, hogy negatív élményt („szégyenérzetet”) adunk, ha túl nagy töltöttség mellett hagyja abba a feladatot, vagy túlságosan közel van

a végleges lemerüléshez, viszont, ha valakit megmenteni próbál alacsony szint mellett is, akkor pozitív élményt („büszkeséget”) váltunk ki benne. Nem igaz- hamis tudást adunk, hanem „le- szidjuk-megdicsérjük”. Így nem hideg logikai mintaként, hanem egyfajta érzelmként is meg- formálódna a gépben nevelésünk eredménye. Ehhez hasonló architektúrák régóta léteznek, melyre még katonai példa is akad: sikerült hatékonyabbá tenni MI alapú célfelismerést hasonló, nem-rationális komponenseket, a szorongást és a magabiztosságot használva.[53] A most vá- zolt elképzelés vonalán ugyan jóval többet várunk el a rendszertől, de a utánzás lehet hasonló.

Az ilyen objektív érzetek megvalósítása azonban csupán a kezdet. Az értelemről fakadó érzelmek megvalósítása is fontos. Például hasznos lehet, ha nagy mennyiségű logikátlan válasz belső ingerültséget vált ki a gépben, ahogyan egy értelmes ember is idővel türelmét veszti a butaság miatt. Az ilyesfajta érzelmek uraltsága is összefügg az erkölcsi fejlettséggel. Az indu- latokon való uralkodásnak ugyan vannak észérvei, megtanulni azonban csak élményszerűen lehet. Ezért lenne szükséges az érzelmeket nem csupán affektusaikban szimulálni, hanem az értelmi intelligenciához hasonló minőségben, belső tulajdonságként is szükséges reprezentálni például egy valószínűbb érzelmi intelligenciát azokban a gépekben, ahol cél egy erkölcsi érzék. Hiszen nem szabad látszódnia az ingerültségnek sem az értelmes emberen, sem a gépen – mivel uralniuk szükséges azt.

Számtalan belső tulajdonság vethető még fel, melyek az érzelmekhez hasonlóan, de más irányból hatnak az erkölcs-tanulás élményszerűségére. Ilyen például az önzés vagy a káros mér- tékü együttérzés („mártírkodás”), ezeket tehát optimális szinten belül kell tartani. Erre is fel- használható lehet a mesterséges empátia (empatikus ágens)[205] kutatások elmélyítése, amely emellett például a szégyenérzet vagy a másoknak okozott fájdalom átélésének leképezéséhez is hozzásegítheti a gépet. Vagy ide tartoznak picit más nézőpontból a különböző emberi gyenge- ségek (régii nevükön gyarlóságok) leküzdésének sikerei vagy a bennük való elbukás kudarcai, melyek az érzelmekhez hasonlóan működnek az emberben, tehát a gépi leképezésük is hasonló. De ilyenek a szociális intelligenciával¹⁸⁰ összefüggő képességek fejlettségi szintjei, illetve fej- lődése is, melyre még visszatérek (IV.2.5.). A mesterséges erkölcsi érzék „felbontása” annál pontosabb lesz, minél több ilyen komponens valósul meg, ezért szükséges a lehető legtöbbet azonosítani és leképezni.

Az eddigieken túl megemlíthetünk néhány nehezen elhelyezhető, problémás tényezőt is: tisztázatlan még akár az álmok hatása az erkölcsi fejlődésre, akár a tettek visszahatásának pon- tos mechanizmusa, vagy a fájdalom szerepe a lelki érésben. Kérdés, hogy vajon milyen

¹⁸⁰ A felsorolás többi része kevésbé kutatott, de ez a terület intenzív kutatás alatt van, pl. [206]

funkciója van ebben a fejlődésben a „csak úgy” jellegű, teljesen spontán és megmagyarázhatatlan emberi döntéseknek, illetve azok következményeinek. Hiszen ezek mélyén sem véletlengenerátor, hanem valós és bizonyos szempontból megmagyarázható okok állnak. Témánk szempontjából nem szükséges a humánkomponensek teljes listáját adni, és őket példák sorával illusztrálni, hiszen ennyiből is érzékelhető az óriási kihívás: a morális érzék gyenge verzióinak előállítására is igen-igen sok olyan emberismeretet igényel, amelyek kutatása még gyerekcipőben jár vagy nem a gépi leképezés oldaláról elemzi azt a „lelki” elemet.

A humánkomponensek tanulmányozása azonban egyéb veszélyforrásokra is felhívhatja a figyelmet. Például sok döntés mögött állnak ott a korábbi tapasztalatok frusztrációi, melyek fel nem dolgozott, vagy egymásnak ellentmondó, de fontos élményekből fakadnak. Ezek döntésképtelenné tehetik az egyént – de ha egy felhő alapon tanuló rendszer élményhalmazát nem kontrolláljuk a tanulási folyamat alatt, hasonlóan belekerülhetnek paradox, ellentmondó adatok, melyek döntéshelyzetben az MI belassulásához vagy fagyásához vezethetnek. Ez már ma is fennálló veszélyforrás. A jövő kockázata lehet, ha például az előző bekezdésekben említett élmény-típusok feldolgozási mechanizmusa nem megfelelően van megtervezve vagy betanítva, emiatt az MI rendszer „mesterséges frusztrációt” produkálhat, ami pedig nem cél.

A humánkomponensek fenti vizsgálatával talán világossá vált, hogy ezek valójában csak bizonyos szempontból irracionálisak – valójában van működésüknek egy sajátos logikája. Ha pedig minden humánkomponensben található logikai minta, akkor ezek gépi reprezentációja sem utópia, sőt az erre való törekvés elvárható. Mivel egy ember az irracionális faktorok tudatosításával lényegesen csökkenteni képes az elvektől való eltérések számát vagy váratlanságát, ugyanez az uralt (és nem gépies) megbízhatóság lehet cél a digitális rendszerekben is.

(3.) A funkcionalizmus kritikája és az antropológiai keret tágítása

A korai filozófiai megközelítések középpontjában az a kérdés állt, hogy lehet-e emberi tudata a gépnek, illetve mennyiben nevezhető tudatosnak az a mechanizmus, amely jól, akár tökéletesen utánozza az ember gondolkodását, következtetéseit. Már ekkor tetten érhető, hogy mennyire alapvető hatással van kutató világszemlélete az etikai konklúzióra. Például az eredeti „agyprotézis kísérlet”[207, o. 1–6] szerint az agy neurális kapcsolatait elektronikusra cserélve semmi nem fog változni, míg ennek kritikája szerint a létrejövő új lény, bár érzékekkel rendelkezik és a gondolkodik, a világon „kívül” fogja magát érzékelni.[19, o. 840] A kísérlet első értelmezése, a máig népszerű funkcionalista megközelítés, melyben az emberséget a biológiai struktúra csupán hardverként hordozza. A hardvert pedig lecserélhetjük, akár más elven működő, de azonos funkciójú alkatrészeire – ahogyan ssd-re cseréljük a winchestert, és ettől még minden

szoftver fut egy számítógépen. E szerint tehát elméletileg nincs akadálya egy erkölcsi érzékkel bíró elektronikának.

Kritikaként azonban meg kell jegyezzem, hogy a funkcionalista szemlélet inkább utópia, mint tudományos tétel. Feltételezi ugyanis, hogy mindent, ami fontos tudunk az ember biológiai működéséről és az erkölcsi érzékről. Ezzel szemben tudományosan nem zárhatók ki olyan funkciók, amely kizárólag biológiai struktúrákon működnek, csak eddigi műszereinkkel nem mérhetők. Igazából a lélek dualista szemléletű (testtől külön is fennálló) létezése sem zárható ki kísérleti alapon, hiszen nem mérhető a fizikalitás mennyiségeiben – így például a lélek szerepe például az erkölcsben sem nem bizonyítható, sem nem cáfolható.

A funkcionalista utópiát vallva hasonló kihívásba ütközhetünk, mint az MI kutatás korai optimizmusa. A hőskorszak mérnöki gondolkodású kutatói ugyanis úgy vélték, hogy az *okosságot* igazán nehéz gépbe implementálni – a mozgást, a szenzomotoros képességeket jóval könnyebb, hiszen az utóbbit fejletlen állatok, pl. rovarok is megvalósítják. Szerintük az ember biológiai késztetései (éhség, szexualitás) vagy érzelmei csupán rontják az intellektuális teljesítményét.[149, o. 75–80] Majd 30 évnyi kutatás után kristályosodott ki a tény – amelyet Moravec-paradoxonként[208] ismer a szakirodalom – hogy sok dolog, amelyet az ember olyan egyszerűnek tart, nagyobb kihívás elé állítja a mérnököket, mint a diplomás embereket is megizzasztó „nehéz” feladatok. Könnyebb a bonyolult logikai feladványok megoldását vagy a sakkozás képességét gépiesíteni, mint azt, amit egy egyéves gyerek vagy egy állat is képes megtenni, pl. odamenni a sütisdobozhoz. Kiderült azonban, hogy nem csupán az okosság (intelligencia) teszi emberré az embert, és az érzelmek felismerése és gépi megjelenítése, melynek emulációját régen elvetették, mára a fejlesztések egyik leginnovatívabb területévé vált.[18]

A tévedés mögött egy helytelen antropológiai megközelítés állt, és a fenti eset rámutat arra, hogy a technikai kihívások megoldását hogyan hátráltatja a hibás filozófiai szemlélet. Példának hozható erre a híres „Kínai szoba” gondolat kísérlet[209] is, mely a tudatosság korábbi értelmezési hiányosságait tárja fel, sőt szempontjaink mentén túl is mutat azon. Az elgondolás a gépet olyan emberhez hasonlította, aki ugyan nem tud kínaiul, de a kapott jelekkel jól végrehajtja az elvárt feladatokat, bár ezekből igazából semmit sem ért.¹⁸¹ Eredeti célja mellett a hasonlat értetőbbé teszi azt a tényt is, hogy valójában nincs szükség az emberi tudat mélységeire számos tevékenységhez, melyet eddig emberinek tartottunk, mert nem tudtuk rábízni pl. állatokra vagy korábbi gépekre. Vagyis az ember bizonyos képességei ugyan gépiesíthetőek, de a gépiesítés által felvetődő emberi-társadalmi kérdések mára túlnőttek a műszakilag képzett emberek

¹⁸¹ Az eredeti John Searle cikk helyett érdekesebb a teljes vita anyaga: [210, o. 199–226]

kompetenciáin,[211] akiknek jobban együtt kell működniük a humán tudományok kutatóival. Meg kell várniuk a megfelelőbb, használhatóbb, etikusabb megközelítések kicsiszolódását – máskülönben vakvágányra futhatnak, és eladhatatlan vagy betiltásra ítélt termékekbe ölik energiáikat. Ha emberileg nehéz is, gyakran ki kellett lépniük abból az antropológiai megközelítésből, melyben az emberi tudatot értelmezik. De ezt fordítva is be kell tartani, és a humán tudásoknak sem sci-fi-be illő disztópiák ellen kellene küzdeniük amikor még azt sem tudjuk, hogy mely emberi tulajdonságok digitalizálása lehetséges egyáltalán.

IV.2.3. Az etika digitalizálhatóságának technikai oldala

Vajon mik a határai a világ algoritmizálásának? Átugorva olyan túl elvont fogalmak „matematizálását”, mint a hit vagy az intuíció, itt azt vizsgálom, hogy vajon leképezhető-e az elektronikus gépi működés utánzásaira konkrétan egyik sajátosságunk: az „etikai érzék”.

(1.) Fogalmak leképezése, matematikaivá alakítása

Már az ókortól a tudományok és az adatfeldolgozás alapja, hogy az érzékelt világunkat hogyan képezzük le a matematika és a tárolható jelek világába. Ez vált a számítástechnikai alapkérdésévé is annyi módosítással, hogy a gép által értelmezhető módokat kellett találni. Már a fizikatudomány fejlődése kialakította azt a megközelítést, hogy csak egy-egy adott részproblémát alakítunk át matematikaivá, tehát a világban fellépő tényezők túlnyomó részét meghatározott mértékben elhanyagoljuk. A bürokrácia pedig kialakította az adattárolási módszerek alapjait, melyek jelekben tárolható formába alakították a valóság azon részét, melyre valahol később szükség volt. A számításoknál a fizika, az adatok kezelésénél a bürokrácia módszereit fejlesztette tovább az informatika.

Mivel a számítások adatokra épülnek először ezeket vizsgálom. A fizikalitásban mérhető vagyis mennyiségi dolgok (darabszám, súly, magasság stb.) átalakítása egyszerű, objektív, és az ezeken végezhető számítások is kézenfekvőek. A minőség adatként való tárolásakor azonban nem tehetünk mást, mint hogy egy adott skálán valamilyen pontértéket adunk az objektumoknak, ezáltal összehasonlíthatóvá tesszük őket (pl. szépségversenyen). Ekkor azonban a pontozó szubjektivitása óhatatlanul megjelenik, és az összehasonlítás pontatlanná válik. Ennek elkerülésére a legjobb módszer, hogy a vizsgált minőséget szétbontjuk, és minél több részében mennyiségi paraméterekre vezetjük vissza (pl. csípő-, derék-, mellbőség). Az MI segítségével ez a mennyiségi megközelítés olyan pontosságot képes öltetni, mely annak illúzióját adhatja, hogy képesek voltunk mennyiségileg megragadni a minőséget. A szépségverseny példájánál maradván az MI sok győztes arca alapján képes azokat a mintákat megtalálni, amelyek szerint mi

emberek statisztikailag úgy véljük, hogy egy arc szép. Ezzel a fajta tudással, a taníthatósággal és az óriás mennyiségű adat figyelembevételével már felmerülhet, ami korábban nem: hogy döntőbíró nélkül megmondja valakiről, hogy szépsége megfelelő-e egy reklám céljára – vagyis gépiesíteni lehet mondjuk a castingot[212]. Bár a gép továbbra sem tudja mi a szép, de tudja mi tetszik az emberek nagy részének – és az iparnak erre van szüksége. Ennek hátulütője lehet a sablonossá válás: az ilyen eredmény idővel unalmassá válik, erre utal pl. a termék-designok szépségének folytonos változtatáskényszere is. Vagyis a gép segítségével egy sablonos reklámfilm ugyan olcsóbb lesz, viszont mégsem érdemes rábízni egy szépségversenyt, hiszen új fajta szépségekre is szüksége van a szépségiparnak. Hozzáteszem, hogy a szépség-ideál hely szerint eltérő a világon, így a különböző területeken fejlesztett ilyen MI, területenként már ebben eltérő megállapításra juthat (később bemutatom, hogy fontosabb szempontokból is).

A külsőben megragadható elvont fogalmak mennyiségivé alakítása mellett, jóval nehezebb belső emberi tulajdonságokat számszerűsíteni. Az okosságra például az IQ értéket tesztelik, ez azonban nem egzakt. Eltérő módszerek által, vagy más-más időpontokban eltérő értéket ad, továbbá csupán bizonyos fajta okosság mérésére képes, és a feladatok egyéni (akár zseniális) értelmezését kizárja. Mindennek ellenére általános célokra alkalmazhatóak. Az orvos-tudomány és a pszichológia fejlődése pedig egyre több belső tulajdonságról szolgáltat egyre mérhetőbb adatokat. Ezeknél azonban nem a pontos valóság, hanem a használhatóság az elsődleges szempont (pl. a munkakör). Egy mindenre használható digitalizált személyiségkép úgy tűnik nem lehetséges. Persze a fuzzy logikánál (II.3.2.) bemutatott módon a tulajdonságok valamilyen mértékben jelen-lévősége elvileg egész jól megragadható, ám ennek komplexitása és költségsége óriási, nehezen térül meg, de főleg azért, mivel az ember változik, így egy személyiségkép egy magát fejlesztő egyén esetén gyorsan elavul. De akadály az is, hogy a felmérés maga visszahat, elgondolkodtatja fejleszti az egyént – így a teszt végére már eltérnének egyes paraméterek a teszt elejéhez képest.

Még nehezebb a helyzet olyan tényezők leképezésekor, melyeknél a rangsorolás az emberek között is eltérhet, a személettől is függhet. Például a nyugatias ember számára semmiféle értékkel nem bír egy indiai aszkéta csendessége, pedig ott köztisztelőben áll. Az etika kérdésköre leképezési szempontból ebbe a körbe tartozik. A viselkedési normák részben objektívizálhatóak, és bizonyos mértékben leképezhetőek, amint azzal kutatások is foglalkoznak¹⁸². Az emberi jog is ezt a leképezést végzi – de láthatóan a legjobb megfogalmazásokban is keletkeznek

¹⁸²pl. Reza Ghanadan kutatása. Ld. [183, o. 108.] Az itt említett próbálkozás a normákat nem elvek, inkább az etikett szintjét értette: <https://www.thestatesman.com/technology/science/new-algorithm-to-teach-robots-human-etiquettes-1497601224.html>

„kiskapuk”: a jogászi szavak szintjén sem időtálló a leképezés. S mivel fogalmi gondolkodásunkat próbáljuk az MI-be leképezni, a próbálkozás így még kérdésesebb.

Ráadásul etikai a kérdésekben elvi alapon sem lehet támpont a mérhető világban bevált mennyiségi szemlélet. Hiába érjük el, hogy rengeteg ember véleményét értelmezze és figyelembe vegye pl. egy döntéstámogató MI, hiszen tömegek ítélete téves is lehet – esetleg épp a fentebb említett információs műveletek eredményeként. Hasztalan, hogy már nem pusztán számok összevetéséből kapná az eredményt, mint a klasszikus számítógépek, hanem számtalan neuron súlyozását figyelembe véve – ám a gép továbbra sem tudja mi a jó és a rossz. A jól objektivizálható kérdésekről (kressz szabályok betartásáról vagy ügykezelési hibákról) képes lehet egy etikai ügynök^[213] megmondani, hogy az jó-e vagy rossz. Egy bíró-MI sematikus estekre megtalálhatja a jogszerű ítéleteket.^[214] Ám ahol az emberek úgy vitáznak, hogy abban a személyes meggyőződésük is szerepet játszik, ott a mérlegelés meghaladja a gépek kompetenciáját. A fentieket továbbgondolva jobban alátámasztást nyer az a korábban tett állítás is, mely szerint a fuzzy-etika felvethető ugyan (ld. II.3.2.), de csak meglehetősen korlátos módon alkalmazható, tehát csak olyan esetekben, ahol nem merül fel morális dilemma.

(2.) Raszteresség és vektorizálás a gépi döntéshozatalban

A közelítéses számításon alapuló eredmények általában megfelelőek gyakorlati szempontból. Sok integrál-számítást könnyebb közelíteni, mint a géppel képlet-szinten megoldatni (bár sokszor arra is van mód). Ezzel a módszerrel azonban a kérdéseket nem megoldjuk, hanem kikerüljük. Az adatok felbontásának pontossága alapvetően határozza meg a gépek használhatóságát. A leképezések ezen problémakörére a grafikai megoldásokból hozok példát. A közismertebb leképezés (a digitális fényképezés) képpontokat tárol valamilyen minőségben. Az ilyen képek a nagyítás egy bizonyos fokánál raszteressé, tovább nagyítva használhatatlanná válnak. A másik módszer végtelenségig nagyítható, mivel az ábrázolandó objektumok körvonalát képlettel írja le. Ezzel rajzok, logók jól ábrázolhatóak, és egyszerű paraméterezéssel bármekkora méretezhetőek. A valóságból vett összetett kép esetében azonban ez utóbbi módszer is függ a rajzoló szubjektív ábrázolásától, és a részletekre való figyeléséből, vagy függ a felbontástól egy raszteres kép vektorizálása esetén.

Egy morális döntés leképezése is épp így: vagy bizonyos fokú felbontással, vagy pedig „logó-szerűvé” szimplifikált módon képezhető le. A leképezés lehet elviselhetetlenül raszteres, mivel pontozáson alapul és a skálát túlfinomítva emberileg válik pontozhatatlanná, és lehet „vektorizált”, amikor matematikailag leírható, sematikus megoldásokat kezel. A raszteres etikai kép vektorizálása épp úgy problémás, mint a fényképek esetében. Ezért számomra kérdéses,

hogyan az MI egykor majd úgy jelenjen meg, mint egy új létréteg.¹⁸³ Mert bár rendelkezhet az MI az emberénél nagyságrendekkel nagyobb tudással, de amíg digitálisan közelíti meg az analóg világot, addig csupán az embertől kapott leképezési módot tudja fokozni. Az ember érzékelése ugyan korlátoltabb, rosszabb felbontású és kisebb határok között mozog, mint az általa létrehozott szenzoroké és jelforrásoké. Viszont még egy sérült vagy hiányos érzékszervekkel élő ember is képes teljes életet élni – míg az MI legfeljebb vegetál (reprodukálja és táplálja magát).

A konnekcionista elme-modellek szerint az agyunk neuronjai is csupán mennyiségeket tárolnak[39, o. 148–151], ám én inkább azokhoz csatlakozom, akik elégedetlenek ezzel[39, o. 151–154] – álláspontom indoklása a funkcionalizmus fenti kritikájának mintájára megtehető. Összegezve: azt sem látom bizonyítottnak, hogy pusztán matematikával az emberséget meg lehet ragadni a maga teljes egészében. Az MI-k hiába is hoznak létre majd újabb, önmaguknál is hatékonyabb MI-ket, akkor sem fognak tudni a leképezés verméből kitörni. A tőlünk örökölt dolgokat fokozhatják ugyan, de a szépség és jóság számukra mindig csupán a neuronjaikban tárolt súly marad. Vagyis az ember erkölcsi érzékének meghaladása nem várható a gépektől, csak részlegesen lehetnek képesek utánozni egy fejlett erkölcsi érzékkel rendelkező felnőtt látásmódját.

Ennek ellenére társadalmunk egyre inkább arra épül, hogy megfelelő mennyiségű adattal elégséges pontossággal leírható minden minőség. Az elhanyagolásokat olyannak vélik, ahogyan a Hold tömegvonzását sem szokás figyelembe venni egy földi mérnöki munkában. Így az a közvélekedés alakul ki, hogy gépekkel az emberi (minőségi) tényezőkről is pontos fogalom alkotható. E szerint kiváló számítástechnikai háttérrel rendelkező szakértők ma már képesek akár objektíven megmondani, hogy mi a „jó”. Így azonban a fentebb említett csapdába jutunk: a tömeg is és a szakértő is tévedhet. A tömeg az említett információs műveletek folytán, a szakértők hiányos modelljeik által – az MI pedig a tőlük kapott adatokra támaszkodik. Ráadásul régen a tanulatlanság miatt lehetett megvezetni a tömegeket, már azonban már a tanultság sem segít a feldolgozhatatlan információmennyiség helyes megítélésében. A felsorolt szempontok alapján tehát úgy vélem rossz vágányon mozognak azok a próbálkozások, melyek az etikai kérdések matematikai alakba öntését által próbálják megvalósítani annak érdekében, hogy azok programozhatóvá váljanak.[215] A tudományosság határai a gépiesítés korlátai.

(3.) Az MI, mint véleményformáló – de lehet tudata?

A jelenlegi technológiákkal hamarosan megjelenhetnek akár a vélemény-formáló robotok is – hiszen a jelenlegi influencerek „okosságai” ügyesen gépiesíthetőek, ha az üzemeltető általuk

¹⁸³ A tézis indoklása: [29]

valamiféle marketing vagy információs műveletet épít fel. Egy ilyen program összetevői: egy instant Coelho-generátor,¹⁸⁴ egy cikkíró generatív MI, és egy deepfake által generált arccal rendelkező felolvasó-robot. Ezek egyesítésével létrehozható, és ha azt hirdetik róla, hogy tudatra ébredt az MI és lám filozofál,¹⁸⁵ sokan elhiszik. Az „igazi” saját tudatú MI-vel kapcsolatban is szkeptikus véleményre jutottam, mivel nem látom igazolhatónak, hogy nem csupán a tudat egyes *megnyilvánulásait* reprodukálja-e (utánozza) az a gép. Még azt sem látom objektív ténynek, ha egy adott filozófiai álláspontra az adott MI „rájön”. Hogyan lehetne tudományosan igazolni, hogy a gép jött rá, és nem csupán egy, az ember által irányított gépies felismerésről van-e szó? A rendszer egésze túl bonyolult és rejtett: egy szuperszámítógépen telepített, éveken át fejlesztett MI tudományos lapokban nem publikálható, amit pedig megjelenik, azt nagyon nehéz kísérletileg igazolni vagy cáfolni.¹⁸⁶ Ezek alapján a tudatos gép propagandájának inkább a veszélyeit látom: az MI üzemeltetője ugyanis a gép által ideologizálhatja meg, hogy az általa tenni kívánt dolog „jó és helyes”. Így az MI bölcsessége ideológiája lehet egy etikátlan marketing lépésnek, de akár egy diktatúra emberellenességének is.

(4.) *Az agy-protézis kísérlet alkalmazása a gépi etika területén*

Végül a fenti digitalizálhatóság-szkepszisem igazolására egy gondolat-kísérletre hívom az olvasót, a fentebb (IV.2.2. / 3.) agy-protézis kísérlet mintájára: feltételezzük, hogy sikerült egy nagyon jó és köztiszteltetben álló ember etikai szemléletmódját egy digitális rendszerbe lementeni (leképezni). Tegyük ezt egy robotba, és teszteljük úgy, hogy összehasonlíttjuk az ember és a belőle leképezett robot döntéseit. Az első labor tapasztalatok teljes egyezést mutatnak, ezért a gépet elengedjük a világba, hadd folytasson egyfajta „életet”, hadd tanuljon belőle. Természetesen figyeljük meg, és hasonlítsuk össze döntéseit az emberével, akiből létrejött – és egyre jelentősebbek lesznek az eltérések. Gépünk a kapott etikai elveket „gépiesen” követni fogja, és 100%-ban azok szerint dönt. Az ember ellenben változni fog, főleg, ha tényleg „jó ember”, hiszen szellemi-lelki fejlődésünk, szemléleti csiszolódásunk egy életen át tartó folyamat. Kísérletünk tanulsága, hogy az etikai elvek „belevarázsolása” egy MI-be nem lenne elegendő olyan reakciók gépi reprezentációjára, amit „nagyon emberséges viselkedés”-ként aposztrofálhatnánk. Még ha sikerülne is, a kapott racionális, logikus MI rendszernek nem lenne erkölcsi fejlődése. A gép döntése önmaga szerint mindig helyes lenne, mivel a működési tartományában

¹⁸⁴ Sajnos a magyar Instantcoelho oldal megszűnt, de hasonló a pl. <https://inspirobot.me>.

¹⁸⁵ MI marketing szempontból ez nem lenne meglepő, hiszen volt rá példa a Google-nél: <https://raketa.hu/a-google-ontudatra-ebredt-mesterseges-intelligencijamat-mar-a-felhasznalok-tesztelik>

¹⁸⁶ Ez a helyzet most Kína újabb kvantumgépes és MI rendszereivel, amelyek tudásáról csak általuk kiadott sajtóhírek alapján tudunk tájékozódni.

értelmezhető legjobbat választja. Nem dilemmázik, hiszen az adott pillanatban rendelkezésére álló információk és a neki adott algoritmusok alapján a legoptimálisabb döntést hozza. Ha szempont a gyors döntés, akkor prioritások alapján szűkített elemzéssel dönt, mivel az a legoptimálisabb. Így egy bonyolult terror-elhárító rendszer, erkölcsileg nem különbözik egy egyszerű, mozgásérzékelőkhöz kötött automata fegyvertől, habár emberéletekről dönt. A racionális MI számára nem létezik tévedés, csak kevés információ, vagy hibásan kapott elemzési algoritmus. Így létrehozóinak elveiből logikusan levezethető tanulási folyamatának egy bizonyos ponton túl még létrehozói sem fognak örülni. Egy normális ember nem mindig a régi elvei szerint dönt, akkor sem, ha „elvember”. Az ember fejlődésében szerepe van az utólagos dilemmázás stresszének, és fejlődése nem lineáris, váltakozóan, hullámzó görbe szerint megy végbe. Akár épp egy „bűnös életszakasz” vezet el a stabil erkölcsi megújuláshoz, a másokért élő jószághoz. Ezt pedig nyilván nem engedhetjük meg egy olyan MI-nek, aki emberek között él – így azonban nem is szimulálható megfelelően pl. a felelősség sem.

Fejlesszük: vajon milyen emberi tulajdonságokat kellett volna szimulálni ahhoz, hogy robotunk a kapott logikus, racionális következtetéseken túl valamelyest emberien fejlődjön? Elsősorban az érzelmeket és egyéb olyan tulajdonságokat,[18] melyek viszont optimalitási és hatékonysági szempontból igazából rontanak a teljesítményén. Az ilyen gép, amelybe például az elvégzendő feladat miatti aggodást építenek,¹⁸⁷ elkezd „visszabutulni” emberi szintre. Egy ilyen autonóm rendszernél kifejezetten nagy kihívás lenne azoknak okoknak szimulálása, melyek miatt az ember eltér az elveitől.

Mivel gondolat kísérlet fő erénye az olcsósága, ezért elgondolhatunk egy harmadik verziót, ahol (az egyszerűség kedvéért) egy olyan „elv-embert” próbálunk szimulálni, akinek 95%-ban sikerül szemléletéhez híven élnie. Ám belegondolva abba, mik állhatnak ennek az 5% eltérésnek a háttérében arra jutunk, hogy pont annyi tényezőt kell szimulálnunk, mintha egy nagyon könnyen befolyásolható, döntéseiben nem következetes, csapongó embert szeretnénk utánozni, aki 60%-ban eltér saját szemléletétől. Csupán a tényezők paraméterei térnek el, a bonyolultságuk azonos. Vagyis szükséges lenne minden feltárt okot leképezni a gép számára – vagyis, ha elvileg lehetséges lenne, akkor költségessége miatt nem valósulhatna meg.

IV.2.4. Objektív MI készíthetnek a tudattalanul szubjektív alkotók?

A közelmúltig megjelenő technológiák tervezésében az alkotók egyénisége legfeljebb a design-ban jelenhetett meg. Egy robotkar, génkezelt termény, vagy egy arcfelismerő rendszer

¹⁸⁷ [216] Idézi Csepeli György, id. mű. 79. o.

létrehozásához az alkotók csak a szellemi teljesítményüket adták hozzá. Ezekről nem lehet megmondani, hogy a világ mely részén, és milyen szociális háttérből jött személyiség hozta őket létre. Azonban az MI, a tanulási képessége által kilépett a többi technológia köréből, képes bizonyos sajátosságokat magára öltetni. Az MI tanítása hasonlítható az emberi szocializációhoz, és köztudottan a szocializációnk legalább annyit (ha nem többet) ad hozzá képességeinkhez, mint a velünk született adottságok. A mi személyiségünkben megjelenik több egykori felnőtt személyiség lenyomata, és ugyanez várható a nagy MIKT rendszereinek esetében a jelenlegi fejlettségi szinten. Sokan kutatják emberi szemléleti torzításaink, elfogultságaink megjelenését az ilyen rendszerekben – alább ezek rövid ismertetése után megvizsgálom az egyéni és csoportos szubjektum lenyomatának tágabb lehetőségeit, egy példán bemutatom a kulturális háttér hatását és az objektív MI koncepció problematikusságát.

(1.) *Elfogultsági, avagy előítéleteségi kutatások*

Az etikai kérdések megjelenése sokrétű: az alapirányokat a felhasználói, fejlesztői és a gépi döntés szintjein megjelenő etikai kérdések jelölik ki. Az előítéleteség kutatása és elkerülése mindhárom szinten az MI fejlesztések egyik fontos, gyakorlati kérdésévé vált. Itt nem térek ki a két szélső szint olyan kérdéseire, hogy a felhasználó is visszaélhet a technológia adta lehetőségekkel, vagy felvetődhet egy tömeges felelősség, például amikor a felhasználók adatai alapján elfogult adatbázisok jönnek létre, és nem foglalkozom a környezetükből tanuló autonómiák előítéletes vagy etikailag kifogásolható döntéseivel sem. A tanulmány fő céljaihoz illeszkedve elegendő a fejlesztői dimenzióban kezelhető kérdéseket vázolni.

Itt jegyzem meg, hogy a szakirodalom az *előítéleteség* (bias) szót használja (elterjedtsége miatt ezt a kifejezést is használom), jelen vizsgálat általános rátekintéséhez azonban pontosabb az *elfogultság* kifejezés. Az előítéleteség pejoratív felhangja ugyanis félrevezető lehet, hiszen nem feltétlenül rossz szándékról vagy az alaposág hiányáról van szó. Részemről a hivatásos fejlesztőkben feltételezem a jószándékot az előítéleteség elkerülésére,¹⁸⁸ ennek ellenére számos esetben le kellett állítani például az MI-re alapozott munkaerő toborzást, mivel az előítéletesnek bizonyult. Elsősorban a faji és nemi elfogultság esetei gyakoriak,[217] de üzleti szempontból is problémás lehet, ha például egy automatikus címkézés miatt egy egyedibb kinézetű cipő nem jelenik meg egy keresőkérdésre. A probléma elkerülésére léteznek automatikus

¹⁸⁸ Természetesen az emberi aljasság nem fogja megkímélni az MI-t, de itt a piacra kerülő termékeket vizsgálom.

rendszerek¹⁸⁹ is, de sok szakember egyetért abban, hogy az MI egészének gondos létrehozása adhat csak esélyt a probléma csökkentésére.¹⁹⁰

A fejlesztői dimenzió belül is többdimenziós a probléma az architektúra szintjei szerint: A hardver szintjén inkább a sebesség vagy tárolókapacitás miatti torzítások lépnek fel, algoritmikus szint pedig akkor torzít, ha rosszul egyszerűsíti a világot a vizsgált kérdéskörre. A tanítható rendszerek esetében az adatok és a modell elfogultsága, valamint a helytelen tanítási módszerek együtt jelentkeznek. Számos ajánlás fogalmazódott már meg arra, hogy a torzításokat egy adott tőrészhatáron belül lehessen tartani, lényegük, hogy fejlesztéskor nagy figyelemmel kell lenni a helyes súlyozás és tisztességtelen mintavételt elkerülése mellett az elfogultságot kiváltó olyan okokra, mint a velünk született vagy tanult előítéletek, ez utóbbira majd alaposabban kitérek. Emellett életcikluson át működő eljárásrendekre van szükség, mely a fejlesztés minden lépésében specifikusan szűri az algoritmikus és adat-előítéleteket.¹⁹¹ Van, amikor végképp elkerülhetetlen a használt adatbázis előítélete, ez esetben annak használóit képzik ki a létrejött, elfogult MI helyes alkalmazására, erre jó példa egy gyermekjóléti prediktív elemző eszköz koncepciója.¹⁹² Ez rámutat arra is, hogy az adatok megfelelőségének kérdése igazából nem is MI-elfogultsági probléma. Az emberi megismerést épp így akaratlanul elfogulttá teszi. Például a kutatható adatok torzítása, egyoldalúsága miatt egy kutató is áldozatává, sőt részesévé válhat egy információs megtévesztő műveletnek amikor internetes forrásokból tájékozik.

Kiemelendő az a kutatási irány, mely a generatív intelligenciát (II.2.3.) hívja segítségül az előítéletek csökkentése érdekében.¹⁹³ Mert bár maga a generatív MI sem mentes még az előítéletektől,[223] mégis alkalmas lehet a kiegyensúlyozatlan adathalmazt egyenletesebbé tenni. Egy egyszerű példával: nem szükséges sokezer ember arcjátékával tanítani egy érzelemfelismerő célrendszert, ha generálható, hogy minden nem / rassz / kor arcjátékát létrehozza egy tanító MI rendszer (hasonló módon, ahogyan deepfake videó működik). Ez működhet is az USA-ban, azonban épp a különböző kultúrák eltérő gesztusjelrendszere miatt világszinten nem

Végül egy másik érdekes kutatási ötlet: szándékosan „negatív” képekkel tanították az egyik kísérleti MI-t, míg azzal azonos modell pozitív képeken tanult. Ugyanaz az algoritmus így eltérő eredményre jutott, hiszen egy pszichológiai teszten az egyik egy pszichopata reakcióit adta, míg a másik MI egy normális emberként reagált. Ebből a kutatás vezetője arra jutott, hogy az

¹⁸⁹ Pl. az IBM ilyen rendszeréről: [218] A rendszer leírása: <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/>

¹⁹⁰ Javasolt összefoglaló: [219]

¹⁹¹ Ezt alkalmazza pl. a Prevision rendszere fejlesztés nulladik, a kezdeményező fázisától, a tervezésen, a végrehajtáson, a megfigyelésen át a lezáró fázisáig [220]

¹⁹² A témáról többet nyújt a Rhema Vaithianathan vezette kutatók publikációjának gyűjtőoldala [221]

¹⁹³ A generatív MI elfogultság elleni használatáról gyakorlati példákkal ld: [222]

adatok fontosabbak az algoritmusnál.¹⁹⁴ Következtetése azonban vitatható, hiszen ugyanazokkal az adatokkal tanítva egy rosszul beállított és egy jól működő MI-modellt, hasonlóan eltérő eredményt kapnánk.

Azonban nem minden kutató szerint küszöbölhető ki az előítéletesség, ezekhez a kutatókhoz csatlakozom jómagam is. Van, aki négy érveléssel is alátámasztja tagadó állítását,[225, o. 89] szerintem pedig a következő egyszerű gondolattal is cáfolható. Tudatunk mintáit is nevezhetnénk előítéleteknek: hiszen agyunk is úgy működik, hogy a meglévő mintákhoz hasonlítja az újakat. Ez alapján tervezzük az MI modelleket is, mely a tárolt mintáikhoz hasonlókat képes felismerni a kapott információhalmazban. Tehát bizonyos fajta előítéletesség mindig lesz minden emberben és gépben a megismerési mód működése miatt, tehát „tökéletes előítéletmentesség” lehetetlen és értelmezhetetlen.

(2.) *Az alkotók és csapatok szubjektív nézetei az MI rendszerekben*

Az elfogultság kezelhetőségének általános áttekintése után, most ennek arra a részére fókuszálok, hogy ki lehet-e zárni az alkotók egyéni és csapat-szemléletét egy MI-t létrehozásából? Abból indulok ki, hogy egy személyiség, és az ő szemlélete leginkább a tetteiben, pontosabban a tettek mögötti döntéseiben nyilvánul meg. Tetteink és gondolkodásunk között azonban szakadék van: egy részük mögött előre megfontolt nagyobb döntés van, de több az olyan spontán gesztus, mely az elme tudattalan részeiben gyökerezik – ezt szeretném itt kinagyítani.

Ezeket a tudattalan részeket nem az etikai elveink befolyásolják, sokkal inkább a szocializáció, mely főleg a kulturális környezetből táplálkozik (a családot a kultúra átadójának tekintem). Számtalan ponton az ember elvei megfogalmazatlanok, következtetlenek, tudattalanok, és a személyes környezeti ingerek (elvárások, érzelmi állapot) vagy a média információi szerint változnak. És ezek ott vannak minden alkotóban – hogyan lenne akkor kizárható az alkotásból? Már az algoritmus szintjén is belekerülhet, hiszen egy jobb programozóknak stílusuk van – de az emberileg felügyelt tanítás során elkerülhetetlen. Az ember saját szemléletének nem-tudatos részét képtelen lesz kizárni a munkájából, hiszen észre sem veszi, hogy belecsempészi. Ez igaz a csapatmunkára is, hiszen egy fejlesztő csapatnak is van egy sajátos arculata. Az, hogy „egyre jár az agyuk”, ad egy virtuális személyiséget a közösségnek, ami hatékonyá teszi a közös munkát, ám épp ezért például átsiklanak egy-egy kérdés fölött, mely később hibához vezet. Közös szemléletüket (a virtuális személyiséget) részben a vezetők vagy vezéregyéniségek határozzák meg, de óhatatlanul kihat rá a kultúrkör vagy szubkultúra is, ahol szerveződik. Jelentős lehet az eltérés egy államon belül is, pl. belvárosi vagy faluban lakó munkaközösség között, de még

¹⁹⁴ [224] A kutatás oldala <http://norman-ai.mit.edu/>

nagyobb különbségre lehet számítani köznapi kérdésekben egy kínai, egy orosz, egy amerikai team esetében.¹⁹⁵ Az fejlesztőcsapat kihatása terén azonban még két esetet említeni kell az „átlagolható” csapat-szemlélet virtuális személyiségén túl. Felmerülhet, hogy a teamen belüli nézeteltérések viselkedési anomáliákat okozzanak azzal, hogy a kimondatlan szemléleti eltérések miatt eltérően tanítják (párhuzamosan) a rendszert, ezzel összezavarhatják. A másik, hogy a gépet is érheti olyan „trauma” egy gondatlan tanító-operátor részéről, mely miatt képességei sérülnek, és amelyet később javítani szükséges,¹⁹⁶ de akár egy frissítésben kapott adathalmaz is okozhat ilyen gondot.

Remélem ezzel a néhány gondolattal sikerült világossá tenni, hogy jelentősen eltérő csoportok már az elvi keretek szintjén is komoly szemléleti eltéréssel tervezhetnek meg egy MI-t. A tudatos eltérés tettenérhető különbség lesz, viszont a fentiekből következik, hogy egy-egy fejlesztés között rejtett eltérések is lehetnek, a fejlesztők tudattalan oldalainak lenyomataként.

(3.) A kulturális háttér hatása az MI fejlesztésekre

Lépünk tovább egy tágabb szociális közeg, a létrehozó kultúra irányába. Az eltérő kulturális különbségek számtalan különféle mikro- és makro-szinten megjelenhetnek. Jelölhetnek területileg is jól elkülöníthető embercsoportokat, például markáns a különbség a „nyugati” és „keleti” világszemlélet mögött, ám ezeken belül is számtalan terület különíthető el. Lehet azonban egy ország hagyományos lakosságán belül is egy szemléleti különbség, etnikailag vagy gazdaságilag eltérő területeken, vagy területileg nem-megragadható módon pl. világnézeti (vallási) hovatartozás szerint, vagy a konzervatív és liberális (stb.) politikai orientációban. Mindezen eltérések MI-ben megjelenő lecsapódásainak vizsgálatához még nem áll rendelkezésre elegendő adat, hiszen kortárs technológiáról beszélünk. De talán egy régebbi technológiából vett példa is elég meggyőző lehet az ilyen irányú jövőbeli kutatások jelentőségéhez.

A világszemléletek kifejlődése tetten érhető az irodalomon keresztül is, ahol az MI etika problémái előbb megjelentek, mint ahogyan ennek a technológiának az alapjai létrejöttek volna. Ezek a megjelenések az adott regionális környezetet képezték le, és ebből lényeges eltérések jöttek létre. Például Japán lényegesen elfogadóbb a robotokkal kapcsolatban, mint a nyugat,[226, o. 8] pedig sok szempontból nyugatias. Ennek okaként az eltérő kulturális hagyományokat említik. Japánban kimutathatóan a sokkal nagyobb bizalmat a robotok iránt. Ezt a kutatók az animista kulturális háttérre vezetnek vissza,[227] mely szerint mindennek van lelke. A nyugati kulturális hagyomány viszont lélek nélkülinek vallja az ember által alkotott lényeket,

¹⁹⁵ A felvetés sokféle módon pontosítható, melyekre itt nem térek ki, pl. ha a team a online, sőt nemzetközileg szerveződik, akkor nem lesz csapatléggör (ekkor tagok egyedi szemléletének különbségei erősödik fel).

¹⁹⁶Ahogyan az imént említett pszichopata MI-t is lehet fejlesztői szerint más képekkel „gyógyítani”. ld. [224]

ezért veszélyesnek érzi. Ennek gyökere egyértelműen a kereszténységben van (nem számít, hogy a közgondolkodás már rég a vallástalan), amelyben fontos, hogy egyedül Isten teremthet lelket. E miatt a lélektelenség miatt válnak veszélyessé a nyugati irodalomban megjelenő emberkreálmányok, mint Frankenstein. Maga a robot szó először Karel Čapek, cseh író 1920-ban megjelent „R. U. R. - Rossum Univerzális Robotjai” című művében szerepel először, és számos nyelven terjedni kezdett. Ebben az alkotásban a robot egy gölemhez hasonló, ami a cseh zsidó misztika egy mágia által megelevenített, veszélyes kőszörnye.¹⁹⁷

A példa megmagyarázza, amit 23 éve még nem értettünk, hogy egy jó használtautó áráért minek valakinek robotkutya?¹⁹⁸ A japánok szerint, mivel lehet lelke a robotnak, akár egy robotkutyának, tehát bizalomra érdemes. Bár húsz év alatt nyugaton felnövekvő új generációk talán kevésbé félnek az IT-től – ezt tanárként is tapasztalom – de például egy robot bajtársként való kezelése valószínűbb a keleti haderőkben, mint nyugaton.

(4.) *Az Üres Lapra alapuló objektív gépi intelligencia kritikája*

A szubjektivitás bizonyítása után tanulságos alaposabban megvizsgálunk valamely abszolút objektivitást megcélzó kutatói véleményt is. Ennek radikális képviselői szerint minden emberi adat szubjektív, ezáltal valójában akadályozzák egy erős mesterséges intelligencia optimális képességeit. Ám ebből nem csupán az következik, hogy az eddigi gyenge MI-ből kapott tanulságok nem alkalmasak egy erős MI-hez. Egyik szerző szerint, általánosítva azt, hogy az arcfelismerési algoritmusok nem használhatóak más célra, arra a következtetésre jut, hogy „a világról az MI-rendszer által megfigyelt a priori információk felhasználása nemcsak a kifejlesztett numerikus módszerek alkalmazhatóságát korlátozza, hanem teljesen lehetetlenné teszi egy erős mesterségesintelligencia-rendszer létrehozását is.”¹⁹⁹ Tanulmányában kimutatja, hogy egy megfelelő szenzorrendszerhálózattal ellátott, de „üres lappal” induló erős MI lehet képes újszerű problémák megoldására. A tanulmányból nem egyértelmű, mitől „megfelelő” az a szenzorhálózat, mely ezt a modellt adatokkal ellátná és az üres lapot teleírná, ezek hogyan lennének teljesen mentesek az emberi szubjektivitástól. De a problémát a teljesen önmagát kialakító MI tézise jelenti, amivel tudok egyet érteni. Az még elfogadható, hogy juthatunk ezáltal újszerű gépi felismerésekhez, ám egy ilyen „üres lap” által önmagát kialakító MI igen veszélyes lehet, ezért csakis biztonságos, szimulált környezetben, kutatási célokra javasolható, továbbá érzelmi

¹⁹⁷ Ezzel inkább irodalmi cikkek foglalkoznak, pl. [228]

¹⁹⁸ A Sony cég tanulni képes Aibo ERS-110 termékére gondolok, mely nem hasonlítható a mai robotkutyákhoz: <https://www.sony.com/en/SonyInfo/design/gallery/ERS-110/>

¹⁹⁹ Saját fordítás. Megjegyzendő, hogy a cikk írója orosz, így ezzel a radikális gondolattal rámutat a korábban említett, eltérő kulturális háttérből adódó eltérő tervezői szemléletre. [229, o. 6]

modellekkel kiegészített intelligenciával. Javaslatait érdemes az emberek mellett egy másik, emberibb MI által is ellenőriztetni. Várható ugyanis, hogy egy ilyen MI a világot nem jól képezi le – ahogyan ezt a következő szakaszban bemutatom.

(5.) *Mennyiben ért meg az MI olyan eszméket, mint egyenlőség és fejlődés?*

Alább azt vizsgálom, hogy hogyan jelenhetnek meg elvont elvek egy MI-ben. Ehhez a társadalmi egyenlőség elterjedt eszméjét járom körbe, érintve a fejlődés-hit szemszögéből már tárgyalt fejlődés dogmát is. Azt szeretném igazolni, hogy az ilyen kérdéseket nem igazán lehet még majd erős MI-be sem jól leképezni – vagyis amennyiben igen, akkor eltérő eredményre jutnak a különböző szemléleti háttérrel fejlesztett MIKT rendszerek.

- **Matematikai és társadalmi egyenlőség.** Hogyan lehetne elmagyarázni a gépnek, hogy miben tér el az egyenlőség az egyformaságtól? Kicsit játszva a szavakkal: az egyformaság a „matematikai egyenlőség” műveletével egyszerűen kódolható számára. A társadalomtudományi egyenlőség-fogalom azonban jóval bonyolultabb. Az egyenlőség megértéséhez figyelembe kell venni egy sor egyéb emberi és társadalmi értéket (pl. szabadság, egyediség stb.). A fogalom tehát összetett, és praktikus megnyilvánulásaiban pedig óhatatlanul szubjektív megítélést hordoz. Ezeket számos eltérő iskola különbözően fogalmazza meg. Így túl azon, hogy sok elvont emberi fogalmat, és ezek egymással való kapcsolatát kellene megfelelően modellezni, ráadásul ennek az MI-nek „ki kellene nőni” az alkotók közötti szemléleti különbségek szintjéből.
- **Az időtényező az egyenlőségben.** Az újkortól kezdve egyre gyorsabban változnak az értékrendek és prioritási szempontok, gyorsulva alakul át egész etikai rendszerünk és éleltszemléletünk. Ezen belül változik az egyenlőség, szabadság, egyediség (stb.) fogalmak értelmezése is. Mint minden változást, ezt is különböző módon követik az emberek, a radikális progresszivisták elébe mennek, a változás-ellenesek gátolják, és a kettő között mindenki szubjektíven válogatva elfogad és ellenez bizonyos részleteket. Így, ha alkotnánk is egy elvont fogalmakat jól kezelő modellt, az is csak kis ideig lenne működőképes. Hiszen elméletileg vagyunk képtelenek úgy modellezni az ilyen elvont dolgokat, hogy az időtálló legyen, ugyanakkor emberi maradjon, mivel az ember (és az emberiség) természete, sőt erénye, hogy folyton változik. Az előrelátás a múltbeli állapotok interpolálásán alapul, ám a társadalmi változások sok szempontból kiszámíthatatlan fordulatokat vesznek²⁰⁰. Tehát nem látható előre, hogy húsz év múlva mire terjed ki az egyenlőség – ahogyan 30 éve sem lehetett látni, hogy milyen viselkedésmódok legalizálódnak.

²⁰⁰ Akkor lesz a jövő kiszámítható, ha a gépek határozzák meg – de szeretnénk-e abban a gépiek világban élni?

- **Hierarchia a világ mintázataiban.** A világ természetes mintázataiban sokkal inkább a hierarchikus sémák jelennek meg. A matematika maga is tele van hierarchiával, nem csupán erre szolgáló módszerek (gráfok, halmazok) szempontjából, hanem az algebra alapjaitól (műveletek sorrendisége) kezdve mindenhol. Például, ha nem lekódojuk az alapl műveleteket, hanem megtaníjuk a gépnek, ha tehát a gép „megéri” miért így helyes, akkor „érzékelni” fogja benne hierarchiát is – ezt az értést pedig jól fogja tudni használni, ha képlet-szintű megoldást várunk el tőle (nem statisztikait). Bármelyik tudományt tanulmányozza az MI, ott az egyformaság és a hierarchia kombinációjával találkozunk. Az IT hálózatok hierarchikus szervezettségét is megállapíthatja, erre jut a biológiában is, de ha a társadalmat tanulmányozza az MI, akkor is folyton hierarchikus államszervezetekbe, gazdasági struktúrákba botlik. A társadalmi egyenlőség ezért az erős MI felé haladva egyre nehezebben kezelhető dolog lehet a gép számára. Jól hangoznak a Democratic MI nevű project[230] eredményei, melyben egy szimulációs játékban a mesterséges intelligencia olyan játékmódszert fedezett fel, amely helyreállította a játék kezdetén adott vagyoni eltéréseket, szankcionálta az ingyenélőket, sőt a többségi szavazást is sikeresen megnyerte. Véleményem szerint azonban egy nem négyfaktorosra szűkített, és nem kizárólag nyugati szemléletű szereplőkkel folytatott játék[231] során, hanem valódi piaci környezetben nem feltétlenül nyert volna ez a szoftver. Ha a sakkautomatánkat úgy tanítjuk be, hogy minden bábu egyformán értékes, biztos, hogy veszteni fog – főleg az ellen a keleti MI ellen, amely helyesen értelmezi a bábuk hierarchiáját...
- **Eszmék megjelenése a különböző helyeken fejlesztett MI-kben.** Ha pl. a fejlődés dogmáját (III.3.1./3.) szeretnénk az MI számára érzékelhetővé tenni, valójában nincs nehéz dolgunk. Egy MI képes lesz felismerni a folytonos változást a kapott adatokat feldolgozásával. Azt pedig, hogy ez „jó”-e, azt – mivel elvont, minőségről szóló információ – emberi tanítás révén adhatjuk meg számára. Ezután a hasonló változási mintákat is „jó”-nak fogja ítélni, tehát véleménye az emberekéhez hasonló lesz. Persze attól függően, hogy hol hozták létre, tarthat mást jónak, ahogyan emberek sem értenek ebben egyet – de ettől még ez a vonása majdnem emberi.

Az egyenlőségénél más a helyzet, mivel ahogyan fentebb bizonyítottam, a társadalmi egyenlőség elvét egy MI sem fogja önmagától felismerni, főleg, ha valamiféle Üres Lap-szerű modell alapján készül. Egy felismert hierarchikus mintát elláthatnánk „előjellel”, ahogyan a változást, de itt a mintázat eltér el a világtól. Megfelelő emberi ráhatás révén sematikus problémákat ugyan már most is képes *kezelti* egy MI, ha megfelelően előkészített adatbázis vagy célirányos tanítás során erre készítik fel – de nem *érzékel*, úgy, ahogyan a hierarchiát „érzékelni” képes. Ugyanis mivel a hatékonyság alapján optimalizálja a valóságot, az MI könnyen fittyet hány az etikai irányelveinkre, sőt a nemzetközi jogi Chartákra és sok ország által ratifikált együttműködésekre is. Valójában csak a nyugati társadalmakban meghatározó princípium az egyenlőség –

Európától keletre és délre távolodva azonban, egészen más azok a prioritások, amelyek történelmileg meghatározóak. A nyugatin kívüli világ csupán olyan mértékben tartja fontosnak az egyenlőség eszméjét, amennyire abba a kultúrába sikerült beleitatni az un. „európai értékeket”. Japán és a robotlelkek példája megmutatta, hogy még egy fejlett demokráciában is marad számos történelmi hatás – tehát ilyen markáns eltérés ki fog hatni az autonóm gépeket előállító technológiákra is. Kérdéses tehát, hogy a nagyobb társadalmi igazságosság terén például egy kínai MIKT rendszer (vagy egy elképzelt Észak-Koreai MI) juthat-e arra a felismerésre, ami a nyugati értékekkel azonos – és ha nem annak nem jó elgondolni a következményeit.

IV.2.5. A rajintelligencia – a gépi erkölcs kulcsa?

Végül ebben az kisfejezetben a korábban említett *mesterséges mélyentitás hálózat (Artificial Deep Learning Network – ADLN, ld. II.3.3)* fogalmát összekapcsolom a jelen vizsgálat mindkét alfejezetével. Először térjünk vissza az autonómiavizsgálat azon a kérdéskörére, hogy csak a rajintelligencia lehet képes a 4TA szélsőhelyzeti autonómiájának gépi utánzására (IV.1.1./2.) az Epszilon gépi-autonómia szintnél). Ennek etológiai alapja, hogy a közösségben élő állatok jó részénél létezik egy olyan jelenség, hogy a közösség, mint halmoz élete fontosabb, mint az egyed élete, így egyedeit a raj feláldozza, vagy bizonyos fajok esetében ők áldozzák fel magukat a raj életéért. Individuális társadalmunktól ez a mentalitás meglehetősen idegen. Viszont a régebbi emberi közösségeket vizsgálva, elsősorban a kisebb törzsi vagy nemzetségi szervezetségben élő embercsoportok esetében ez jóval általánosabb volt. Sok, nem-nyugati, kevésbé individuális társadalom esetében ez markánsabban megjelenik. Bár esetenként a modern nyugati embereknél is fellelhető (pl. a különböző fegyveres hivatások tagjaiban), ám az individuális szocializáció miatt nyugaton ez kevésbé mélyen határozza meg az egyének gondolkodását, tehát nem ivódik mélyen az interperszonális hálóba, sem a kultúrába.

Ez a kétféleség szembesít minket azzal az alap-problémával, hogy az MI és a gépi autonómia körüli gondolkodás nálunk túlságosan individuális és központi jellegű. Egy erős hardveren alapuló központi szervergépben képzeljük el az erkölcs gépi megvalósításának zálógát (Asimov Központi Világagynak hívja, amely megoldja az emberiség problémáit). Ez az individuális én szuper-kiterjesztése. Az ASI által irányított jobb világról szóló cikkek néhol virtuális Superman-re emlékeztető ábrázolások – holott a megoldás lehet, hogy a modern ember által mélyen lenézett, törzsi énség-értelmezésben rejlik. Általános MI autonómiáját vizsgálva az rajzolódott ki, hogy azok az MI alapvető előnye, vagyis a statisztikai gondolkodáson alapuló optimalizáció szerver szinten nem érheti el a szélsőhelyzeti autonómiát, talán még a szabálmögöttit sem. Ennek oka, hogy a megvalósított intellektus egy individuum, aki önmaga számára a világ

középpontja, így matematikailag is önző, hiszen nélküle megszűnne a világ középpontja. így szétesne a világ is. Ezzel szemben az önmagát a törzs tagjaként megélő egyén, saját léte nélkül is meghatározhatja a törzs létét, mely ráadásul nagyobb érték, mint az ő egyéni léte.

Véleményem szerint ebben a szellemben kialakított, átlátható méretű nagyszisztemek összekapcsolásával létrejövő „Törzsi-MI” lehet az alapja olyan általános MI-nek, melynek az autonómiája biztonságosan szimulálja a szabálmögötti vagy akár a szélsőhelyzeti autonómiát.

IV.3. ÖSSZEGZÉS ÉS RÉSZKÖVETKEZTETÉSEK:

- **A várakozások teljesülése.** A vizsgálat eredményesnek értékelhető. A megközelítés iránya nem tesz lehetővé kísérleti igazoláshoz hasonlítható objektivitást, ám számos probléma azonosítása és néhány elvi akadály leírása megtörtént az erkölcsi érzék minőségi digitalizálásáról, ezzel a C1 célt elértem. Az elemzések összesítve a megvalósíthatóság ellen szólnak, ha ilyen tudást neurális hálók továbbfejlesztésével próbálnánk elérni, vagyis a K2 kérdésre valóban nemleges válasz adható, ahogyan azt H1 hipotézis feltételezte. Ezen kívül a tanulmány során alapvető védelmi aspektusok sem maradtak el, sőt a terveken felül és az előző vizsgálatlal összesítve egy második eredmény is megfogalmazhatóvá vált, sőt olyan új autonómiaszinteket is dolgoztam ki, melyek igen hasznossá válhatnak a gyakorlatban.
- **Összegzés.** Az erkölcsi érzék digitalizálhatatlanságát több szempontból vizsgáltam. A vizsgálat során külön alfejezetté nőtte ki magát az autonómiák tanulmányozása. Ennek során nem csupán bemutattam a jelenleg használt fontos szintezéseket, hanem azok kritikai vizsgálatával olyan új szinteket határoztam meg mind emberi mind gépi autonómia terén, melyek által a két terület jobban együtt-kezelhető, és összevethető, mint a korábbi felosztások (IV.1.1-2.). Ezután konkrétan vizsgálva az erkölcsi érzéket először azt elemeztem, hogy az emberben hogyan alakul ki ez a képesség. Eljutottam a *humánkomponensek* fogalmáig – így nevezve azt a számos tényezőt, mely segíti az erkölcsi érzék kifejlődését (IV.2.2.). Erre építve több javaslatot is megfogalmaztam a moralitás gépi utánzásához, és kiemeltem aettek jelentőségét az elvi (teszt) válaszokkal szemben. Részletesen is cáfoltam a kognitivisták megközelítést, hogy a morális képesség az intellektussal lenne csupán összefüggésben, valamint a funkcionalista modellt „embergép” modelljét is (IV.2.1, IV.2.2.). Ezek által is rámutattam, hogy a filozófiai tévutak hogyan vetették vissza a fejlesztéseket. Néhány érv a digitalizáció vázolt akadályai közül (IV.2.3 és IV.1 alapján):
 - Az etikai érzék pusztán fizikalitása nem bizonyított, és ha valóban metafizikai kapcsolódása van (pl. létezik lelkiismeret), akkor elvileg nem lehet hozzá matematikai modellt találni. Ha nincs metafizikai kapcsolódása, akkor a pszichológiának szociológiának kell sokkal több energiát fektetnie az erkölcsi érzék affektusainak rendszerré formálásába.
 - Az etika a digitális térbe a komplex morális kérdéseket erősen szimplifikáló képlettel, vagy bizonyos fokú felbontással, „raszteresen” képezhető le (a képekhez hasonlóan, csak itt szabályokkal). Mindkét módon sokat torzul, de ennek következménye a képek

torzulásánál problémásabb. A sejtést igazolja, hogy olyan területek, ahol az emberek vitáiban a személyes meggyőződésük is szerepet játszik (ilyen az etika is), ott a mérlegelés meghaladja a gépek kompetenciáját (hiszen nincs objektivitás, ld. alább).

- A mérhető világban bevált mennyiségi szemlélet nem lehet alapja etikai megközelítésnek, mivel a tömegek ítélete lehet téves.
- A mesterséges tudat alapja lehet egy gépi erkölcsi érzéknek, de igazolhatatlan, hogy a gép nem csupán a tudat bizonyos külső affektusait utánozza, hanem valódi.
- Nincs is értelme egy gépnek akkora szabadságot adni, ami visszaélésre ad lehetőséget, hiszen ez az embereknél is csupán gondot jelent.
- Gondolatkísérlettel igazoltam, hogy egy gép elképzelt erkölcsi fejlődése alapvetően fog eltérni egy emberétől, akiből „tudását másolták”.

Külön kitértem arra a problémára, hogy a morálitás leképezése még ha lehetséges volna is, akkor sem lehetne objektív.

- Nagyon röviden összefoglalva azért nem, mivel a kognitív kompetenciákat célzó MI szubjektív, emberi valóságok leképezését tűzi ki célul.
- Ennek igazolásához próbáltam az előítéletesség elkerülését célzó kutatásokon túlmutató megközelítést alkalmazni, bár ezen kutatásokra is kitértem.
- Az MI a tanítása során az emberhez hasonlóan szocializálódik, így a rajta dolgozó egyének szubjektuma, a létrehozó csoportok sajátossága és az ezek mögött lévő kultúra is lenyomatot hagyhat rajta.
- Cáfoltam hipotézisem antitéziséét (a fehér-lap elméletet, egy radikálisan objektív MI-t célzó orosz kutatást)
- Alátámasztottam gondolatmenetemet egy példával, mely szerint a japánok jobban bíznak a robotokban hagyományaik miatt, mint a nyugatiak.
- Az elvont fogalmak leképezhetőségét is a bizonyítás részeként vizsgáltam, hiszen az ilyen fogalmak értelmezésében is megjelenik a szubjektivitás. Rámutattam, hogy a változást a gép felismerheti, de ezt fejlődésként értékelni csak emberi segítséggel képes, a tanítást végző(k) véleménye alapján. A társadalmi egyenlőséget felismernie is nehezebb, hiszen a világban gyakoribban a hierarchikus mintázatok.

Ezen kívül a fejezet alfejezeteit összekapcsolva mutattam ki a rajintelligencia jelentőségét ahhoz, hogy valamilyen emberibb döntés-utánzást generáljunk.

- **Részkövetkeztetések.** Az autonómia és a mesterséges erkölcs részletes vizsgálata, valamint a III.-IV. fejezet összevetése alapján is lényeges következtetéseket tudtam levonni.
 - 2R-A: A H2 hipotézis bizonyítása által a második tudományos eredményt kaptam meg: Egy mesterséges neurális hálózat képes lehet ugyan bizonyos erkölcsi érzék-jellegű kompetencia megvalósítására, de nem képes azt „emberi” szinten elégséges minőségben utánozni.

- 2R-B: A III. és IV. fejezetek vizsgálatai egyaránt érintik az emberség gépiesíthetőségének a témáját, ezért a sejtést, melyre hipotézis nem készült, a kutatás során a vártnál jóval több érv támasztotta alá, így állítássá léphetett elő. Vagyis a kutatás ilyen irányú felismeréseire támaszkodva megfogalmazható egy 3. eredmény is: Nem elég átgondoltak és tudományosan megalapozottak azok a felvetések, melyek az általános MI ember-szerűségének lehetőségét, sőt, esetleg a szuperintelligencia emberfelettségét tényként kezelik. (Ezt tovább erősíti majd V.1 vizsgálat is)
- 2R-C: A jelenleginél cizelláltabb és kifejezőbb szinteket szükséges megállapítani mind emberi mind a gépi autonómia kifejezések terén úgy, hogy a szintek párhuzamai is kezelhetőbbé váljanak – erre konkrét javaslatokat fogalmaztam meg (IV.1.1. és 2).
- 2R-D: Az 1R-C részkövetkeztetést továbbgondolva igazolva lett, (IV.2.4.) hogy a fejlesztők szubjektuma, csoportszelleme, és a különböző kultúrák eltérő értékrendű és tudású MI-eket hoznak létre, ami védelmi szempontból is jelentős. (Ez megjelenik a védelmi vizsgálatokban (VI.1.)).
- 2R-E: Azáltal, hogy példákat mutattam arra, hogy a filozófiai tévutak hogyan vetették vissza a fejlesztéseket, nem csupán rámutattam a műszaki és humán területek egymásra-utaltságára az MI fejlesztésekben, hanem egyben arra is felhívtam a figyelmet, hogy mind a műszaki szakembereknek kell kilépniük abból az antropológiai megközelítésből, melyben az embert értelmezik, mind pedig a humán kutatóknak is sokszor alaposabb műszaki ismeretek alapján kellene, amikor ezt a témát vizsgálják. (Erre is az MI védelmi oktatásánál térek vissza VI.3.)
- 2R-F: Az affektív számítástechnika mellett a rajntelligencia is kulcsszerepet játszat abban, hogy az erkölcsi érzék gépi utánzására kezdeti lépéseket tegyünk.

V. A MESTERSÉGES INTELLIGENCIA

FOGALMÁNAK ÚJRAGONDOLÁSA

A harmadik vizsgálat nem korlátozódik pusztán erre a fejezetre. Részben már a dolgozat Első Része is C3 kérdést készített elő, aholis számos, az MI-vel kapcsolatos terminológiai hiányosság azonosításra került. A III. és IV. fejezetben is merültek fel adalékok, de még hiányoznak egyes pontosítások és felosztások, ezek pótlását itt tervezem. Az Áttekintő Elemzés alapján K3 kérdés első felére már most megadható egy „igen” válasz, hiszen bemutattam, hogy több fontos vonatkozás is kimarad az elterjedt MI meghatározásokból. Hiányzik azonban még az, hogy K3 kérdés második felének megválaszolásával világos legyen mit lenne fontos még beépíteni egy új MI fogalomba. Ezek megjelölése arra is rámutat, hogy milyen szempontból érthetőek félre a ma használatos fogalmak, és miért lenne szükség teljesebb definícióra, így ezeket a kérdéseket megválaszolva egyben H3 hipotézis is bizonyítottá válik (V.3.2.). A kritikákból konstruktivitást úgy produkálhatok, hogy az elemzések alapján javaslatot teszek új MI meghatározásokra is, majd kiértékelem azokat. Az új definícióhoz vezető úton hátra van még a fogalmi anomáliák áttekintése, továbbá be kell mutatnom az MI újszerűségét, valamint az informatika és más kapcsolódó tudományok változását az MI hatására.

V.1. FOGALMI PROBLÉMÁK A „DIGI-DOLGOKTÓL” AZ MIKT RENDSZEREKIG

Kutatási és oktatási tevékenysége során mindenki szembesül azzal, hogy a nyelv nem képes a technológia gyors fejlődését. Számos kifejezés használata válik zavarossá egy időre, aztán lassan kitisztul. Ehhez a tisztuláshoz szeretnék hozzájárulni az „okos” szó helyesebb használatára, és elkerülésére téve javaslatokat.

V.1.1. A rendszerek összemosódásának problémája

Alaposabban megvizsgálni egy jelenséget, mely a technológiák gyors változásának lehet a következménye. Felhasználóként igen nehéz pontosan lépést tartani, könnyen azt hiheti a felhasználó, hogy amit lát, az úgy működik, mint régen, holott nem ez a valóság. Sokaknak egybeolvad a felhőben elérhető szolgáltatás a helyi offline tudással (pl. az Office 365-ben figyelni kell, mert ugyanaz a könyvtár struktúra van meg a felhőben is, mint helyben). Az ilyen jellegű „hit”-beli tévedések az informatikai biztonság szempontjából is problémaként merülnek fel. Ide tartozik az a tévedés-típus, hogy összemosódhatnak az MI és a hagyományos számítástechnika szolgáltatásai is, ezt itt vizsgálni szükséges, mivel számos fontos vonatkozása van témáinkhoz.

A jelenséget erősíti, hogy gyakran az MI-tudással reklámozzák a termékeket (részletesebben: V.1.3.), elkenve, hogy a reklámozott rendszert valójában mennyire hatja át az MI és az milyen szintű. Mivel sokaknak nincs idejük elmélyedni a működésben úgy tűnhet, hogy minden hasznos és bonyolult szolgáltatásban elbújtattak egy MI-t. A feladatokat azonban általában vegyesen hajtják végre a neuronhálós és a klasszikusan programozott megoldások, BigData megoldások és hagyományosabb relációs lekérdezések, valamint felhők és helyi applikációk, vagyis sok szempontból hibridek a rendszerek, amelyeket használunk.

Megjegyzendő, hogy a számítástechnika történetére egyébként jellemző, hogy az elavult technológiák nem megszűnnek, hanem más területen használják fel őket. Az ipar sokkal jobban tiszteli a hagyományait, mint a társadalom, már csak azért is, mivel egy kiforrott áramkört vagy kódot sokkal könnyebb átdolgozni és újrahasznosítani, mint a fizikai világban kezelni szemetet, pontosabban az elavulás (generált) jelenségeit. A hibrid rendszereknél azonban az olcsóbb fejlesztés motivációjánál fontosabb, hogy a rendszerek vegyesen igazán jók. Az offline funkciók mindig elérhetőek, de online funkciók által pedig sokszorozódik a gép tudása. Az adatfeldolgozásban a hagyományos programkódok és adatbázisok egyelőre sokkal pontosabbak és biztonságosabbak, de az MI által az adatfeldolgozás igazi információfeldolgozássá avanszálhat (ld. V.2.1.).

Alább csak a fő téma mentén az MI és a hagyományos adatbázis kapcsolatát vizsgálom alaposabban egy szakirodalomkereső rendszer példáján keresztül. Miért képesek ezekben az MI lehetőségek²⁰¹ a hagyományos adatbázis képességekkel egyesítve hatékonyabb kereső szolgáltatást nyújtani? Egyrészt ez azért előnyös, mivel ki lehet használni a korábbi, strukturált, (általában relációs) adatbázis-támogatással feltöltött anyagokat. Korábban egy írásmű feltöltéskor számos címkét (metaadatot) kézzel megadtak (szerző, témakör, évszám stb.), melyek alapján az írásban szereplő szavaknál gyorsabban és célirányosabban lehetett információt keresni. Ezután ennek a munkának egy részét erre programozott botokra²⁰² bízta, akik automatikusan végezték a címkézést, de még sokáig érdemes volt kézzel is megadni minél több metainformációt. Később a botokat intelligens ágensek váltották le,²⁰³ melyek egyre jobban eltalálták a szükséges címkéket.

Így az új tartalmak esetében egy MI-agens elvileg már felismerhetné cikk a formázása alapján, hogy ki a szerző és mi a kiadó neve, és melyik évszám tartozik a kiadáshoz stb., tehát

²⁰¹ pontosabban itt a természetes nyelvfeldolgozásról (NLP) van szó (ld. II.3.4.).

²⁰² Saját megfogalmazásban: a *bot* egy virtuális térben működő roBOT, a hálózaton „mozogva” végez automatizált feladatokat. Részletesebben: [232]

²⁰³ Szokás az ügynök szót is alkalmazni az agent-re, de pontosabb az angol terminus magyarul ejtése. Az ilyen intelligens ágensekről ld. [233]

elmaradhatna a humánerővel végzett adatlapkitöltés. Ezzel szemben az általam tesztelt kutatás-támogató szolgáltatások meglehetősen hiányosan végzik el ezt a feladatot. A jelenség oka, hogy egy strukturálatlan BigData adathalmazt nem úgy „tanul meg” egy MI, ahogyan egy ember megtanul kívülről egy verset. Ha „betéve tudná” azt a sok milliárd adatot, akkor nem lenne adathalász szakma. Ezzel szemben a különféle MI modellek saját szempontok szerint dolgozzák fel az adatokat. A modellek sajátosságait jól mutatja, hogy sok szolgáltatás például egyedi módon applikálja a nyelvi modellt, melyre alapul. Teljesen más eredményeket adnak ugyanarra a kérdésre. Okai lehetnek az ilyen eltéréseknek az különböző ráépülő modellek mellett az eltérő adathalmazok is, melyeket figyelembe vesz, de az is, hogy egy rendszer az előfeldolgozás során milyen hagyományos címkézéseket alkalmaz (melyeket hagyományos adatbázisban tárol).

A példánál maradva: jelen tanulmány megírásának időpontjáig végzett tesztek szerint a vizsgált MI-modelleknek a számukra megadott „tananyag”, vagyis az írások tartalma volt lényeges. Két online elérhető pdf formátumú egyetemi jegyzeten teszteltem, könyvek kiadóját kérdezve a Gemini és a ChatGPT rendszertől. Az utóbbi közölte, hogy nem képes rá. A Gemini egyik jegyzetre megadta a jó adatokat a másikinál nem. A hibás esetben rákérdezés ellenére állította, hogy az Osiris kiadónál jelent meg, pedig az Egri Főiskola kiadója adta ki, az Osiris portál keresője nem is adott találatot a címre. Bizonyítékként a Gemini a könyv hátoldalát (kolofonját) említi, ami a pdf-ben nincs is benne. A teszt rámutat, hogy a modellek a cím alapján az interneten keresnek információt, nem pedig az url-ként megadott pdf fájlt tartalmából ismerik fel azt. Ez érthető is, hiszen az elfogultságot fontosabbnak tekintik a fejlesztők, mint néhány ilyen hallucinációt. Ha pedig hiteles tudáshalmazzal kívánták tanítani, akkor abban nyilván eleve rendelkezésre állt adatmező formájában a publikálás minden adata, a modellnek nem kellett megtanulniuk azonosítani a cikk adatmezőit. Ebből egyrészt az következik, hogy mivel még digitalizált magyar cikkekkel is rosszul boldogulnak, ezért még várni kell egy ideig, hogy egy erre is betanított MI a régen megjelent, vagy szkennelt írások adatlapkitöltését precízen elvégezze. Másrészt jobban törekedni kell arra, hogy a magyar kutatók anyagai olyan adatbázisokba fel legyenek töltve, amelyeket ilyen rendszerek tanításánál használnak.

A hibrid rendszerekben működő hagyományos adatbázisok helyett a mai kutatások egyre inkább egy neurális adatbázis irányába (V.1.2.)[234] próbálnak fejleszteni, s bár még koránt sem tökéletesek, fejlődésükben még óriási potenciál látható. Ez is rámutat arra, hogy a hagyományos és MI alapú rendszerrészek aránya gyorsan változik. Az MI egyre több szerepet vesz át, ahogyan a mai processzorokba integráltak számtalan, korábban különálló funkciót. Elvileg mindent részévé lehet tenni a tanuló modellnek, ám visszafogó tényező a biztonsági problémákon túl, hogy mit éri meg piacra dobni? (Nem olcsóbb-e sokszor a humánerőforrás?)

V.1.2. Neurális adatbázisok

Imént az MI és a klasszikus számítástechnikai rendszerek összemosódásának problémáját vizsgálva kézenfekvővé vált egy olyan adatbázis-kezelési rendszer fejlesztési igénye, amely mesterséges neurális hálózatokat használ az adatok tárolására, lekérése és elemzésére. Ezt valószínűsítik meg a neurális adatbázisok, amelyek nem táblákban tárolják az adatokat, mint a hagyományos relációs adatbázisok. Ennél a megközelítésnél sokféle különböző struktúrában tárolható az adatokat, attól függően, hogy milyen típusú milyen feladatokat kívánnak majd az adatokkal megoldani. Ez a megoldás részben a régebbi adatbázisokból az kevésbé ismert objektumorientált modellt viszi tovább, mivel az objektumok (tulajdonságokkal és metódusokkal rendelkező adategységek) közötti összefüggések tanulására jól használható az MI. A másik örökség a kulcs-érték párok, melyekben ilyen egyszerű összerendelésekre redukálják a tárolt adatokat (kulcsokat rendelnek minden értékhez).[235] A használt neurális háló típusának kiválasztására nem érdemes kitérni, a fentebb (II.2.) említett számos modell megjelenik a szakirodalomban, valamint vektoros modellektől vagy gráf-alapú hálózatok is.

Előnyei a hagyományos relációs adatbázisokkal szemben

1. képesek összetett, úgynevezett nemlineáris kapcsolatok modellezésére, vagyis hatékonyan tudnak modellezni olyan bonyolult összefüggéseket az adatok között, amelyeket a hagyományos relációs adatbázisok nem képesek kezelni.
2. pontosabb előrejelzéseket adnak, az adatok mögötti rejtett minták azonosítása által.
3. képesek alkalmazkodni az új adatokhoz (tanulni), ami idővel jelentősen javítja az adatbázis teljesítményét.
4. eleve hatékonyan alkalmazhatók adatbányászati és anomália-észlelési feladatokhoz a mintafelismerő képességük révén, így a BigData számára ideális alap.

Természetesen hátrányai is vannak, hiszen nagyobb számítási teljesítményt igényel (a neurális hálózatok tanítása és futtatása miatt), nagy mennyiségű adatra van szüksége a hatékony működéshez, valamint itt is probléma az MI „fekete doboz jellege”, ami megnehezíti a hibakezelést és a magyarázatot, hiszen a pontos működése nehezen érthető.

V.1.3. Okosdolgozók terminológiai anomáliái

Az alábbi felosztást a következő két dolog tette szükségessé:

- Sokszor szükséges lenne elválasztani magát az MI képességet, és azt, amikor egy MI rendszer egy sor társ-technológiával egyesülve ad eredményt. Például egy szokásainkat tanuló szolgáltatás, vagy egy telefon arcfelismerő funkciója egy egyszerű MI képesség,

amely mögé nem kellene gigantikus rendszerek. Viszont egy olyan MI-t, amelyet egy hadműveleti támogató rendszerben alkalmaznak, számtalan egyéb technológia és műszaki képesség kell, hogy kiegészítse. Ilyen felhasználás robusztus rendszerként lehet csak hatékony.

- Marketing okokból ma bármire ráerőltetik az „okos” (smart) kifejezést. Régebben mindent a „modern” jelzővel próbáltak eladni, csak hogy a mostani szóhullám által sajnos kiüresedhet az okos szó. Hiszen, ha megszokjuk, hogy az boltban szinte semmit sem jelent, akkor majd az emberekre sem lehet megfelelően alkalmazni. Ennek ellenszereként alább az okos kifejezés alkalmazását (annak technikai részét) rendszerezem, valamint javaslatot teszek pontosabb kifejezésekre.

A két problémát együtt kezeli az „okos” funkcióról készített alábbi felosztás, melynek szintjeit a 14. sz. ábrán mutatom be. Az ábra nem csupán az okos jelző jogos felhasználását választja el (vonallal) a problémás alkalmazásaitól, de két alkalmazáskörön belül is megkülönböztet fokozatokat. A szemléltető idomok szélessége két féle dologra utal:

- a színes idomoknál az adott technológia-kör MI-képességét jelképezi,
- a szürke idomok esetében pedig azt mutatja, hogy a jövőben mennyire válhatnak „színessé”, mennyiben lehet közülük az-e MI-hez.

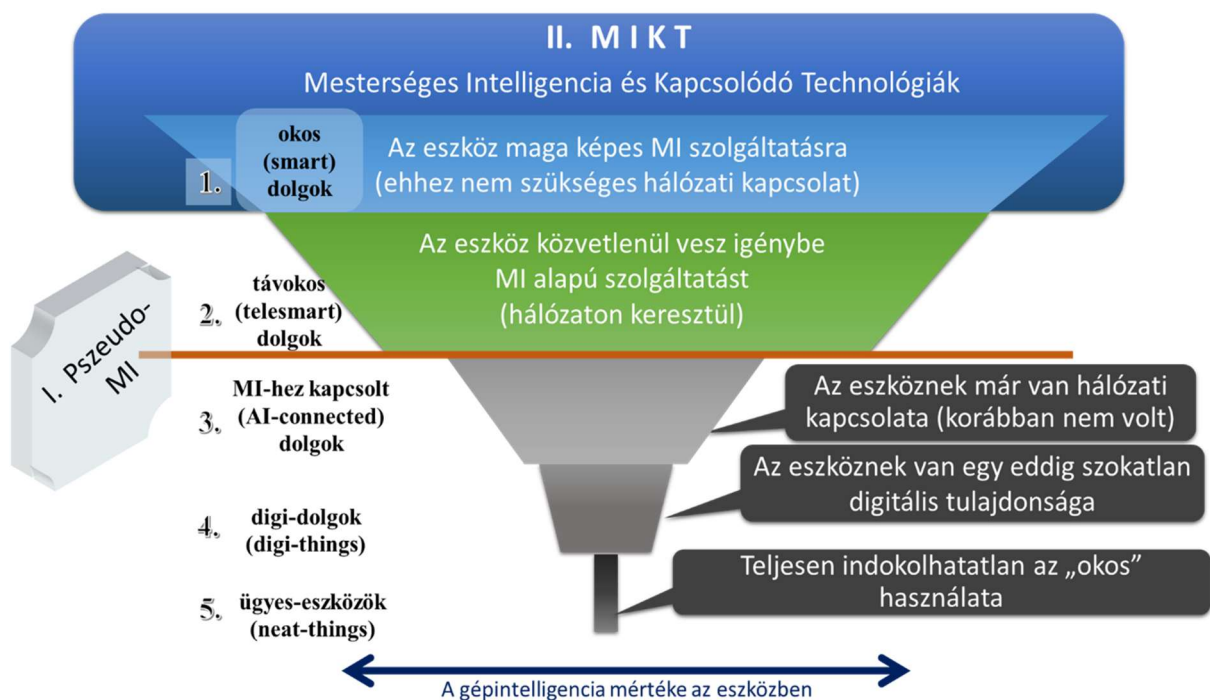
Mivel a szintek tömör leírását az ábra tartalmazza, ezt alább csak pontosítom és példákkal egészítem ki. A példákat termékdalokról vettem, a reklám elkerülése miatt nem konkretizálok (bárki rákeresve többet is talál). Fontosságuk miatt külön kifejezetbe tettem a Pszeudo-MI (V.1.4.) illetve az MIKT²⁰⁴ (II.1.) kifejezések ismertetését.

1. Az ismertetést az „igazi”, a szó szoros értelmében vet okoseszközökkel kezdem. Ezek saját, tehát offline (hálózatot nem igénylő) MI képességgel rendelkeznek. Nem csupán robotokra kell itt gondolni, egyszerűbb tanuló funkciókat egyre több eszközbe lehet integrálni. Például egy okosfűtés figyelheti a felhasználó szokásait és néhány héten át tartó manuális állítgatásból megtanulja, hogy mikor és hogyan állítsa be azt. Egy ideje a hardveres MI támogatás sem a szervertechnológia szintje, jobb mobiltelefonok processzoraiba is integrálnak egy neurális magot, úgynevezett NPU-t (ld. II.3.1.), főleg a képfeldolgozás támogatásához. Olcsóbb eszközökön inkább szoftveres képesség segíti az ilyen funkciókat, pl. csak az arcfelismerést. Ez sokszor elégtelen, így a gyengébb hardver átvezet minket az egy szinttel lentebbi kategóriához.
- A fenti példákban az eszköz nem vesz ugyan igénybe hálózaton elérhető MI rendszert, de ezek által természetesen „még okosabbá” válhat. Ismert példa erre, hogy

²⁰⁴ Mesterséges Intelligencia és Kapcsolódó Technológiák

letöltött offline térképen is lehet útvonalat tervezni, de ilyen esetben a lezárások és a percre kész forgalmi információk nem segíthetik a navigációt. Ez jól mutatja, hogy a legfelső szintre sorolt MI eszköz is hatékonyabb lehet hibrid rendszerként, továbbá távoli adatok vagy erősebb számítási képességek birtokában. (A hibrid rendszer: MI+hagyományosan programozott, ld. V.1.4.)

2. Számos eszköz azonban csak akkor okos, ha hálózaton keresztül csatlakozhat egy MI-hez, tehát ez a távoli szolgáltatás, teszi okossá – csatlakozva azonban nevezhető okosnak. Csatlakozás nélkül csak egy hagyományos számítástechnikai eszköz az ilyen. Jó példa erre az okoshangszóró. Erre a szintre ajánlom, a „távokos dolog” (*telesmart-things*) kifejezést.



14. ábra: Okosnak hívott eszközök, a Pszeudo-MI és az MIKT (saját készítés)

3. Valójában itt illene is lezárni az okos-eszközök körét, de nem ez a valóság. Bizonyos eszközök okos verziójának egyszerűen csak van valami hálózati (vagy egyéb) kapcsolata, ezzel különböztetik meg a „buta” verziótól. Például okosfésű vagy okosfogkefe mobiltelefonos applikációk keresztül csatlakoztatható egy mobiltelefonhoz. Ez utóbbinál például pontosabb lenne a bluetooth-fogkefe kifejezés kifejezés – ám belátható, hogy nem túl jól hangzik a „kékfog fogkefe” az angol nyelvterületen... Ide sorolhatóak még az okoseszközök (pl. tablet, vagyis okosfelhőhöz kapcsolat eszköz) bizonyos kiegészítői, melyeknél az okos jelző utal arra, hogy a kiegészítővel kommunikálva az

eszköz jobb képességeket kap. Pl. az okostoll által jó minőségben lehet digitálisan rajzolni az azt kezelni tudó (drágább) eszközre. Valami olyan kifejezéssel lenne érdemes utalni ezekre a termékekre, hogy „MI-hez kapcsolt” (*AI-connected*) eszköz. (Ez egyben jó példa az ilyen kulcsszavakon alapuló megtévesztésre is, hiszen gumi-végű pálcákat néhol okostollnak titulálnak.)

4. A következő kategóriában már hálózati kapcsolat sincs, csupán valami digitális tulajdonságot építenek az eszközbe, ami szokatlan, például okosövbbe rejtett powerbank, vagy okosserpenyő nyelébe épített kijelző, mely jelzi, hogy jó hőmérsékleten folyik-e a főzés és jelez, ha odaég. Lehetnének tehát digi-dolgok (ez még jól is hangzik, angolul már nem annyira, de használható lehetne a *digi-thing*). Ezzel elérkeztünk a kimagyarázhatatlan kategóriákhoz, ahol tényleg csupán az okos szó lejáratása a kifejezés használata, hiszen még digitális funkció sincs, erre példa egy „okosedény” melynek okossága, hogy műanyag fogantyúja levehető, de létezik „intelligens” mosópor is, az ilyesmit legfeljebb ügyes-eszköznek (*neat-thing*) engedném hívni. Ilyen szélsőségeknél felmerül, hogy legalább irányelvekben tiltani kéne a túlzó szóhasználatokat.

Röviden érdemes átgondolni mi állhat a szó bemutatott kiüresedése mögött. Felvethető az okos és intelligens kifejezések piedesztálra emelése racionalist társadalmunkban. Ám ha ez lenne az ok, akkor már a XX. század közepén is népszerű lett volna, ezzel szemben a „modern” volt ez a dzsóker-kifejezés.

Ezért e mögött a jelenség mögött inkább azt vélem felfedezni, hogy az MI szó nem csupán egy technológiát fed már le, annál többet jelent. Az a rendkívül erős diszrupció, ami hozzá kapcsolódik kihat az asszociációkra is. Az okos jelző által egy dolog abba „kiváltságos” vagy VIP körbe kerül, ami csak jó lehet. Ezért mindenki szeretné, ha terméket beletartozna abba a halmazba, vagy legalább annak holdudvarába, amit az MI képvisel. Mint fentebb láhattuk ennek eredménye hasonlatos ahhoz, amikor a XIX. századtól egyre többen az arisztokráciához akartak tartozni, ami néha csak úgy sikerült, hogy valaki a nevének „y” végződést vagy régies írásmódot adatott... (Reméljük elmúlik majd ez is.)

Az elemzés végére láthatóvá vált, hogy az okos szó degradáció csupán egy jól demonstrálható példa, ezt alkalmaztam a Mesterséges Általános Intelligencia (AGI) kifejezés várható használatára korábban, de ez a modell alkalmazható lesz sok olyan várakozás beteljesedésére, melyre kifejezés már van, de annak pontos tartalma torzítható.

V.1.4. Pszeudo-géptanulás, pszeudo MI, pszeudo-autonómia

Az ML felosztásairól szólva (II.2.4) javasoltam *pszeudo-tanulás* kifejezést, egyfajta nulladik tanulási szintre. Az előző kisfejezetekben (V.1.1, V.1.3.) tovább pontosítottam a jelenség leírását, ám ha a gondolatsort más irányba is kiterjeszthető, ezt ismeretem itt röviden.

Tehát a pszeudo-tanulásnak azt hívtam, amikor a felhasználó számára egy hagyományosan programozott rendszer kelti a tanulás benyomását. Ide tartoznak a próbálgató ciklusok is, melyek aktív próbálgatással találják meg az optimális megoldást, de azok is, melyek egy relációs adatbázisban tárolják el „tanulásuk” eredményét, a különböző címkékhez tartozó súlyokat.

Ez a jelenség azonban vindikál egy pszudo-intelligencia, vagy esetünkben inkább „pszeudo MI” kifejezést is, ahova az így tanuló rendszereket érdemes lenne besorolni. Erre a következő meghatározást javaslom:

A pszeudo-MI egy olyan rendszer, melyről a felhasználók egy részének az a benyomása, hogy háttérében gépi intelligencia működik, pedig valójában egy hagyományosan programozott és relációs adatbázisra támaszkodó architektúra.

A kifejezés hasznos az MI-fogalom és felosztások pontosítására irányuló kutatási irány számára is, mert a Pszeudo-MI fogalmát segít bevezetni.

Elsősorban olyan programok tartoznak ide, melyek az emberek digitális életében generálódó adatokat vesznek figyelembe, mondjuk egy vásárlói érdeklődéseket naplózó, aztán ezt az adatbázist a termékajánlásoknál felhasználó webshop portál. Mivel az adatok percre-készen jönnek létre, és statisztikai elemzés alapján használja fel őket a rendszer, ezért úgy tűnhet, hogy a portál „megtanulta” kattintási szokásainkat. Ám valójában csak egy adatbázis aktuális állapotának lekérdezését végezte el. Ez abban különbözik lényegesen egy neuronhálós MI tanulástól, hogy egy ilyen portál esetében az adatbázisba belépve az ember (az üzemeltető) lekérdezheti és megértheti ezeket a statisztikai értékeket. Egy neuronhálóban viszont nincsenek számunkra közvetlenül értelmezhető adatok (ld. II.2.1).

De érdemes itt kitérni arra is, hogy a Pszeudo-MI-nek is köze van a matematikai súlyok elvéhez (melyre a neuronhálók alapulnak). Egy ilyen ál-intelligens rendszerben a súlyok ugyanis relációkkal összekapcsolt adattáblák adatmezőiként tárolódnak. Azért informatív az a számérték melyet tartalmaznak, mivel ezek a relációk a valóságot képezik le, az ember tehát képes a bennük lévő adatot értelmezni, önmaga számára információvá alakítani. A neuronok kapcsolatait ezzel szemben a modell felépítése adja meg, amelynek semmi köze sincs a valóság közvetlen leképezéséhez, ezért sem értelmezhető a tartalmuk.

A fogalom felvetése kissé megkésett, ezt be kell vallanom, mivel napjainkra legtöbb szolgáltatásban a hagyományos megoldások hibrid rendszerekké kapcsolódnak össze neuronhálókkal és BigData adatokkal. Úgy 8-10 éve jobb lett volna ezt a fajta pseudo-okosságot terminológiaiilag megkülönböztetni. De ma is hasznos bevezetni annak érdekében, hogy világossá váljon a már említett hibrid rendszerek két félepillére: a relációs adatbázis és a nem-relációs adathalmaz. Ezeket tehát a hagyományos programkód és MI-modell hibridje dolgozza fel, és ez a hibrid-modell vezet egyébként az egyre terjedő neurális adatbázisok felé.

A különbségek után nézzük mi az, ami a két rendszerben közös. Ez pedig a kétféle adattárolás (relációs és big) közös jellemzőivel kapcsolatos. Elsősorban a „változékonyság” tulajdonságra gondolok, mellyel a NagyAdatot szokás jellemzni, de nem csupán arra jellemző. Egy nagy webshop esetében is pillanatonként változhat egy adat. Pár óra alatt pedig radikálisan is megváltozhat egy súlyérték, mondjuk a vásárlók között terjedő lejárató videó hatására, lavina effektussal. Egy ilyen esemény pont ugyanúgy érint egy hagyományos kódot és egy MI-t. Attól függően lesz a rendszer érzékeny vagy immunis az ilyen változásra, hogy mennyire veszi figyelembe a múltbeli értékeket – nem pedig attól, hogy neuronháló vagy determinált kód (vagy hibrid megoldás) dolgozza fel benne az adatokat. Így viszont egy ilyen hagyományos rendszer sem lesz teljesen kiszámítható, tehát pseudo-autonóm rendszerként sorolható be annak ellenére, hogy nem tartalmaz neuronhálót, amint ezt már tárgyaltam (IV.1.2. / (3.), alfa szint)

Végezetül leszögezem, hogy ez az ál-gépintelligencia nem pejoratív vagy negatív fogalom. Egyszerűen egy bevett ógörög képzőt alkalmaztam a hagyományos számítógépek csúcsteljesítményére, mert megtévesztheti a felhasználót.

V.2. AZ INFORMATIKA ÉS A HUMÁNTUDOMÁNYOK VÁLTOZÁSAI AZ MI HATÁSÁRA

Az MI jelenlegi meghatározásának ismertetését ebben a részben szeretném előkészíteni olyan fogalmak ismertetésével, melyek „kívülről” kapcsolódnak hozzá. Úgy haladhatok a fogalom belseje felé, ha előbb megvizsgálom az információ, az informatika, a számítástechnika és egyéb ezekhez kapcsolódó fogalmak kölcsönhatását a MI-vel. Összevetve ugyanis az informatika eredeti és későbbi jelentését az az érzése támad az embernek, hogy mindig is az MI létrehozása volt a cél, csak egy időre fel kell adni ezt az álmot, mert még nem álltunk rá készen. A technikai szint elégtelensége mellett hiányzott a multidiszciplináris és interdiszciplináris megközelítéseknek az a kidolgozottsága is, amelyek mára mintegy beburkolják az MI-t, és amelyekben a technológia fejlődni tudott. Ezért ezekről is fontos szót ejteni, bemutatni fejlődésüket. Ide szerkesztettem az MI újszerűségének sajátosságainak bemutatását, valamint néhány

kérdéskört, amiket ez az újszerűség felvet. Végül ismertetem és finomítom a diszrupció gazdasági közegeből eredő (még nem túl közismert) fogalmát is.

V.2.1. Az informatika szó kettős használata

Korábbi munkáim során, az informatikusok, informatika tanárok és mérnökök között azt tapasztaltam, hogy a kollégákat mindig túlságosan lefoglalták a technikai vagy szabályozási feladatok, így nemigen értek rá az alább vázolt fogalmi és elméleti kérdéseken rágódni. Az MI meghatározásainak kritikai vizsgálatához azonban szükséges érzékeltetni az alapvető különbséget, amit a klasszikus számítástechnikához képest az MI jelent. A következő bekezdésekben bemutatom, hogy az MI-nek köszönhetően sok minden más mellett maga az *információval való foglalkozás* is gyökeresen átalakulóban van. Pontosabban tovább alakul, hiszen már az elmúlt 100 évben is sokat módosult.

Sok fogalom esetében jelent problémát egyfajta kettős (vagy többes) szóhasználat. Ez az informatika szó esetében is így van. Hiszen bár használják egy elég tág értelemben is (melyet majd bemutatok), ám hétköznapiok során egy szűkebb jelentést tapasztalunk, ahol az informatika a számítástechnika szó szinonimája. Ilyen értelemben állandósulnak olyan szókapcsolatok például mint informatikai rendszer, informatikus. Így van ez az informatika-történetekkel is: meglehetősen sokan írtak erről (elsősorban oktatási jelleggel), de a legtöbb írás a számítás és a technika története, ez szerepel a tananyagokban is. (Találtam ez alól üdítő kivételt,²⁰⁵ ami a szokásos számítás-történetre csak az információ ősi tárolása, továbbítása és a régi adatfeldolgozás ismertetése után tér rá, így egy picit tágabb értelmezés szerint közelíti.)

Érdekes pulzálást mutat a fogalom értelmezése: Az informatika kifejezés első felbukkanásakor még technikai értelemben használták a szót. A tudományos szaknyelv elég korán kezdte el jóval tágasabban is értelmezni. Azonban nem ez az értelmezés vált általánossá, hanem az említett számítástechnika jelentés. Véleményem szerint viszont az MI lassan rákényszerít minket, hogy az informatika szót ismét egy tágabb értelemben használjuk. Nézzük meg ezt részletebben.

Karl Steinbuch német tudós, akihez a szó első használatát kapcsolják az „*informatik*” kifejezést az információ szó elejéből és az automatika szó végéből alkotta,²⁰⁶ és erről szóló művének címe meg is adja az általa használt definíciót, vagyis, hogy az „Informatika: automatikus információfeldolgozás”.^[238] Itt megjegyzendő érdekesség, hogy Steinbuch behatóan

²⁰⁵ Bár ez a kiadvány sajnos több, mint 30 éves, de még így is feleslegessé teszi, hogy külön fejezetben foglalkozzak az informatika korai történetével. [236]

²⁰⁶ A pontosság kedvéért: egy Helmut Gröttrup nevű kollégájával közösen alkották a szót. [237]

foglalkozott a mesterséges intelligenciával is, és állítólag elsőként hozott létre működő tanulási mátrixot, sőt már 1966-ban megjósolta az analóg technológia digitális technológia általi kiszorítását, a telefon- és számítógépes hálózatok összeolvadását, valamint a szórakoztató elektronika és az adatfeldolgozás egyesülését, vagyis a multimédia korszakát.[239] Sokak szerint a szó európai elterjedésére nagyobb hatással volt a francia Philippe Dreyfus, ő ugyanis 1962-ben a „Société d'informatique appliquée” (SIA, Alkalmazott Informatikai Vállalat) megalapításával hivatalossá tette a kifejezést, ráadásul ezt az „*informatique*” kulcsszót nem jegyeztette be, így 1966-ban a Francia Akadémia szótárába bekerülhetett egy, a jogoktól mentes „informatika” szócikk.[240] Egyébként etimológiaként ő is az információ és az automatika összekapcsolását adta meg, és minden olyan tevékenységet beleértett, amely az adatok automatikus feldolgozásával kapcsolatos,²⁰⁷ az informatika szó ilyen használata terjedt el Délnyugat-Európában. Az USA-ban ugyancsak egy cégnévben használta először a szót 1962-ben Walter F. Bauer elektronikus számítógépekkel is foglalkozó matematikus, akinek Informatics General Corporation néven alapított cége igen jelentős szoftvergyártóvá vált.[242]

A fogalom azonban gyorsan változásba kezdett: a keleti blokkban inkább információs szolgáltatásként lett használatos, míg angolszász területen az információtudománnyal (ld. V.2.2.) összefüggésben is használni kezdték.[241] Több forrás szerint ez utóbbira az első példa egy 1965-ös szimpózium²⁰⁸ a Kaliforniai Egyetemen, bár erről szóló jegyzőkönyvet nem tudtam fellelni. Az OECD²⁰⁹ is már egy tágabb értelemben propagálta a kifejezést 1971-ben az „OECD Informatikai Tanulmányok” című kiadvány-sorozatában, amelyben úgy szerepel az informatika, mint „*az információtartalom, a reprezentáció, a technológia, valamint a felhasználásukhoz kapcsolódó módszerek és stratégiák tanulmányozása*”[244, o. 7].

A kortárs meghatározások közül néhány: a Collins Szótár szerint a brit angolban egyszerűen az információtudomány szinonimája, az amerikaiban az informácifeldolgozás tanulmányozása és számítástechnika értelemben használatos.[245] A Cambridge Szótár szerint azonban az *informatika* magában foglalja az információtudományt, az információfeldolgozás gyakorlatát és az információs rendszerek tervezését is.²¹⁰ Érdekes módon a magyar online értelmező szótárban nincs ilyen szócikk (csak az információ-ra van meghatározás), viszont a magyar katonai

²⁰⁷ Philippe Dreyfushoz fűződik egyébként a programnyelv definiálása és az informativitás kifejezés is. Mielőtt hazájában a Bull Calculus Center igazgatója lett, a Harvard egyetem tanáraként dolgozott az első számítógépen a Mark-I-en is. Később számos nemzetközi társaság vezéregyénisége vagy alapítója. [241]

²⁰⁸ Van, aki tévesen az informatika első felbukkanásaként utal erre: [243]

²⁰⁹ Organisation for Economic Co-operation and Development, magyar nevén Gazdasági Együttműködési és Fejlesztési Szervezet.

²¹⁰ bár nem ezzel definiálja, első meghatározásként „*az informatika az információkat tároló, feldolgozó és kommunikáló természetes és mesterséges rendszerek szerkezetét, viselkedését és kölcsönhatásait vizsgálja.*” [246]

szóhasználatban is megjelenik ez a tágabb értelmezés.²¹¹ Felesleges a szótörténettel ennél részletesebben foglalkozni, ennyiből is érzékelhető, hogy a tágabb és szűkebb értelmezés régóta is egymás mellett fut.

Végül megemlítem, hogy a nagy múltú edinburghi Informatikai Iskola jelenlegi önmeghatározásában a tág értelmezést az MI szélesíti még tovább. Ez a megközelítés nyilván összefüggésben van olyan korai, úttörő lépésekkel, hogy 1966-ban itt hozott létre egy Gépi Intelligencia és Percepció Tanszéket Donald Michie,[248] valamint, hogy itt alkotta meg 1973-ban a kognitív tudomány kifejezést Christopher Longuet-Higgins.[249] A mai önmeghatározásukban ezért nem meglepő, hogy az informatika olyan egy széles és változatos tudományágként kerül meghatározásra, amely felöleli a számítástechnika mellett az MI-t, összefonódva kognitív tudománnyal.[250] Az informatika tudományának így egy kétirányú kihívásra szükséges választ adnia. Feladata egyrészt annak meghatározása, hogy a természetes rendszerekből származó elvek milyen mértékben alkalmazhatók újfajta mérnöki rendszerek kifejlesztésére. De a másik irányban is feladata van, mégpedig annak meghatározása, hogy a mesterséges eszközökben történő információfeldolgozás elméletei mennyiben és milyen körülmények között alkalmazhatók természetes rendszerekre.²¹² A kifejezésnek ezt a tág megközelítését kívánom én is követni, hozzátéve, hogy az informatikánk azt is vizsgálnia kell, hogy a természetes rendszerekből milyen veszélyek irányulnak az információs rendszerekre, és ez hogyan kerülhető el (informatikai biztonság), valamint, hogy az ilyen modern rendszerek milyen módon veszélyeztetik az emberi normákat (pl. az MI elfogultsága vagy az autonómia problémák miatt), és ezek hogyan kerülhetők el.

Saját megközelítemhez igazítva az eddigieket röviden így szól a tágabb értelmezés: *az informatika a valóság elektronikus virtualizációjához keresi az egyre hatékonyabb biztonságos megoldásokat.* Itt a hatékonyság a fejlesztés motivációinak óriási és sokrétű halmazát igyekszik egy szóba tömöríteni, a biztonság pedig kétirányú (a valóságos rendszerek és a virtuális rendszerek között oda-vissza érvényes). Mindezt azért volt szükséges tisztázni, mivel csak az ilyen szélesre tágitott informatika-értelmezés képes megfelelően kapcsolódni az MI terminushoz. Persze csupán ez a jelentés-tágulás bizonyos, de az, hogy fogalmilag ez hogyan csapódik le a

²¹¹ „Az informatika a tudomány és technika azon területe, amely az információk keletkezésének, kezelésének és felhasználásának elméletével, gyakorlati megvalósításával és eszköz rendszerével foglalkozik”. [247, o. 333–340]

²¹²A forrás szerint továbbá: az informatika tudományának lenne szükséges foglalkoznia annak feltárásával is, hogy a mesterséges információs rendszerek milyen sokféle módon segíthetnek megoldani az emberiség előtt álló problémákat, és hogyan járulhatnak hozzá minden élőlény életminőségének javításához. Ld. [250]

szaknyelvekben, az bizonytalan: itt az egyik lehetőség mellett érveltem és érvelek²¹³. Mivel kifejezetten félrevezető lehet, ha valaki az MI-t informatikai szolgáltatásnak tekinti és közben az informatikát csupán a számítástechnika szinonimájaként értelmezi.

Összefoglalva, hogy miért az informatika kifejezés bővülése lenne jó: eddig a gép az embertől kapott információból leképezett adatokat dolgozta fel. Az MI azonban emberi képességekre tör, így felmerül, hogy már az információ (tehát nem csupán az adat) oda-vissza konverzióját is a gép végzi. Az ilyen gép már nem csupán számadattá konvertált dolgokat képes kezelni, vagyis nem csupán adatfeldolgozó egység lesz, hanem tényleg információfeldolgozó képességei is lehetnek. Az ilyen rendszer valódi „automatikus információ- (és nem adat-) feldolgozást” végez, nem csupán a pontatlan szóhasználat miatt (ami napjainkig jellemző), hanem a valóban az. Ezért, vagyis „a szó szoros értelme” miatt változhat a „számítás-technika” valóban „informatikává”. (védelmi szempontból ld. VI.3.4.)

V.2.2. Információtudomány, információelmélet, kognitív tudomány – és minden tudomány konvergenciája

Az informatika kifejezés fenti tágabb értelmezésének világosabb lehatárolása miatt a kapcsolódó tudományokat is górcső alá veszem, hiszen ez a terület is ellentmondásos. Nézzük meg először, hogy pontosan hogyan is nevezzük az információ elméletének általános, minden szegmenst érintő tudományát? Jogos, sőt logikus lenne rá az információelmélet szót használni, amint azt egy magyar kutató egy igen alapos és széles spektrumot áttekintő egyetemi jegyzetében teszi.[209] Az információelmélet (*information theory*) alatt azonban a nemzetközi és hazai szakirodalom szóhasználatában inkább egy hírközlési fogalmat értenek. Ennek kiindulópontja és alapja egy matematikai modell,[251] ami egy üzenet információtartamára és információvesztésének kiszámítására használható. Így az „információelmélet” az információ objektív megközelítése. Jelen vizsgálódás azonban más jellegű, és olyan interdiszciplináris megközelítést igényel, amelyben az MI újszerűsége (ld. V.2.3.) is jól kezelhető. Ezt az információtudomány²¹⁴ kifejezés tartama közelíti legjobban, ezért most rátérek ennek vázolására.

A könyvtártudományi irányból közelítette az információ problémáit Herald Borko, és mint pszichológia doktora kutatta az automatikus nyelvfeldolgozást, és az indexelés kérdéseit. Annak érdekében fogalmazta meg, hogy mit is ért információtudomány alatt,[252] hogy Amerikai

²¹³ Elképzelhető, hogy egy új kifejezés kerül a fenti jelentéssel a köztudatba (pl. virtualitika) vagy egy másik használatban lévő kifejezés (pl. információ-tudomány) kap ilyen jelentést.

²¹⁴ Az információelmélet markánsan megkülönböztetendő az információtudomány elméletétől.

Dokumentációs Társaságot átnevezhessék Információtudományi Társasággá.²¹⁵ Vagyis, hogy a történelmi kihívásoknak jobban megfelelő, általánosabb módon közelítsen a tudományos kutatás az információhoz. Cikkében az információtudomány²¹⁶ rugalmasan kapcsolja össze a régóta fejlődő dokumentációs tudományt²¹⁷ az akkor megjelenő új, számítógépes technológiákkal. Az információ jelentősége körül szerveződő humántudományi kutatások ettől kezdve sokáig sokkal jobban látták a technológia horderejét, mint a technikával foglalkozók a humántudományok fontosságát. Egészen napjainkig sok hardveres és szoftveres szakember úgy kezel információtudományt, mint a „könyvtár-büfé szakot”, ám ezt a felfogást véleményem szerint az MI erősödése szét fogja morzsolni. Nem lesz elegendő ugyanis adatok kezelésének technológiájánál leragadni, mivel az MI-ben igazi értelmet nyer az elektronikus²¹⁸ információfeldolgozás kifejezése, amely alatt eddig elektronikus adatfeldolgozást értettünk. Hiszen pl. egy chatbot nem csupán karaktersorokat vagy számokat válaszol. Bár az MI nem érti ezt az információt, de amit ad, az számunkra hasonlóan értelmes, mintha egy másik emberrel kommunikálnánk.

Az információtudomány fejlődését felesleges itt lépésenként bemutatni (erre a lábjegyzetekben megadottakon túl is igen könnyű forrásokat találni). Elegendő, ha magának a tudománynak az önmeghatározásából emelünk ki egy-két állomást. Az előbb említett Borko-féle meghatározás (lábjegyzetben) gyorsan kinőtte magát, és a kilencvenes évek elején olyan interdiszciplináris tudományként kezdtek tekinteni erre a területre, melyet mérnökök, könyvtárosok, vegyészek, nyelvészek, filozófusok, pszichológusok, matematikusok, számítógéptudósok, üzletemberek képviselői együtt alakítanak.²¹⁹ A tudománynak máig nem született szabványos, mindenki által elfogadott definíciója. Rengeteg kérdéskört felölel, így nagyon sok alterületre osztható mind a három nagy területe,^[257] vagyis az elméleti, az empirikus és az alkalmazott információtudomány. Katonai, illetve védelmi szempontból kiemelendő, hogy az információtudomány részei az információ védelmével, biztonságos kezelésével kapcsolatos kutatások és módszertanok is. Napjainkra pedig az iménti tudomány-felsoroláshoz hozzávehetjük a biológiát, orvostudományt, jogot, szociológiát...

²¹⁵ A kifejezés történelmi helyzetből következő szükségességét elemzni: [253]

²¹⁶ „Az információtudomány az a diszciplína, amely az információ tulajdonságait és viselkedését vizsgálja, az erőket, amelyek az információk mozgását uralják, az információfeldolgozás eszközeit az optimális hozzáférhetőség és használhatóság érdekében. Azzal a tudásanyaggal foglalkozik, amely az információ eredetéhez, gyűjtéséhez, szervezéséhez, átalakításához és hasznosításához kapcsolódik. ... Mint alaptudomány, alkalmazására való tekintet nélkül vizsgálja a tárgyát, és mint alkalmazott tudománysszolgáltatásokat és termékeket állít elő” Id.: [254, o. 314]

²¹⁷ Ennek alapjait Paul Otlet fektette le 1903-ban, aki már 1892-es esszéjében tulajdonképpen egy világ-információs adatbázis víziójával kezdett el foglalkozni. Ld.: [255]

²¹⁸ Pontosabb lehet gépi információfeldolgozásról beszélni.

²¹⁹ [256, o. 6] (magyarul <https://tmt.omikk.bme.hu/tmt/article/view/3196/4214>)

Igazából, bizonyos szempontból minden tudományághoz kapcsolódik, hiszen mindnek van információs problémája. Az eddigieket összefoglalva az információtudományt itt úgy tekintem, mint amely igyekszik megragadni minden szemlélet és tudomány sajátosságai által implikált sajátos információ-megközelítést, és megfogalmazni ennek egyedi aspektusait, hogy az informatika képes legyen majd virtualizálni azokat.

A másik terület, melyről szót kell ejteni, az a kognitív tudomány gyűjtőfogalma, melyet sokkal bonyolultabb jellemezni, mint más tudományt. Egy 1956-ban az MIT által, az interdiszciplinaritás jegyében szervezett konferenciára teszi a kiindulópontját Noam Chomsky,²²⁰ aki nyelvészeti és filozófiai szempontból igen nagy hatással volt a kogníciós tanokra. Magát a kifejezést azonban csak 1973-ban javasolta ebben az értelemben Christopher Longuet-Higgins (mint imént említettem). Ő egy MI-vel kapcsolatos tudományos vita során,[249] azokra a tudományokra utalt e néven, amelyek közvetlenül kapcsolódnak az emberi gondolkodáshoz és észleléshez. Ő még csak a következő négy területet jelölte meg ezen tudomány fő irányáiként: a 1. a matematikát (beleértve a formális logikát, a programelméletet és a programozást nyelveket, az osztályozás és az összetett adatszerkezetek matematikai elméletét), 2. a nyelvészetet (beleértve a szemantikát, szintaxist, fonológiát és fonetikát), 3. a pszichológiát (beleértve a látás, hallás és tapintás pszichológiáját) és 4. az élettant (fiziológiát: beleértve az érzékszervi fiziológiát és agy különböző szerveinek részletes tanulmányozását).[259, o. 30–37]

A részt vevő tudományok körét sokféleképpen meghatározzák, megjelennek pl. az idegtudomány vagy az antropológia tudományterületei, illetve maga az MI is. De manapság már mindig közöttük szerepel a filozófia, és ez nem véletlen. Hiszen bármely tudomány felől indítva a megismerés vizsgálatát, gyorsan ütköztek a kutatók filozófiai kérdésekbe. Ezek közül a legalapvetőbb, hogy a kognitív tudomány sajátos tárgyában a megismerés magára a megismerésre irányul. Így az is kérdésessé válhat, hogy lehetséges-e minden felmerülő kérdésre olyan tudományosan objektív választ adni, melyben magja sincs valami féle hitnek.²²¹ Hiszen magának a megismerésnek (ezzel együtt a megismerőnek) szubjektív oldalát kellene teljesen megszüntetni ahhoz, hogy teljesen objektívizáljuk, tudományosítsuk a kutatás tárgyát (a megismerést). Röviden és sarkítva: így a szubjektivitás objektívizálása az elvárás. Vagyis az ismeretelmélet régi dilemmái igazából nem oldódtak fel ezen tudomány által az elvárt mértékben.

²²⁰ A rendezvényen a tudat és tudatosság lehetséges megközelítéseit kísérleti pszichológusok, számítógép-kutatók és nyelvészek boncolgatták. Ld. [258]

²²¹ A kognitív tudomány lehetőségeibe vetett bizakodás hit jellegére mutat rá pl. [260]

A kognitív tudomány megjelenése annyiban megelőzte korán, hogy csak jóval utána jelent meg a tudományok konvergenciája, erre ki kell térni. Az MI térhódítása jól rávezette a tudományos világ egy részét arra, hogy az informatika – ami alatt most elsősorban számítástechnikát értenek – csupán egy szűk, bár jövedelmező szegmense az információval való foglalkozásnak. A perspektíva ilyen kitágulását úgy is felfoghatjuk, hogy a technológiák egybeolvadását jelentő *digitális konvergencia jelensége*²²² terjesztjük ki további tudományterületek konvergenciájára. Más szóval párhuzamba állítható az, ahogyan egyre több inter- és multi-diszciplináris megközelítés merül fel a tudományos palettán, azzal a közismert jelenséggel, amikor például a kommunikációs eszközpalletta egybeolvad a számítógépes fejlesztésekkel. Tehát ahogyan az informatika és a telekommunikáció infokommunikációvá olvadt egybe a XXI. század elején, úgy a XXI. század közepének egyik jellegzetességévé válhat a technológiák, tudományok és fogalmak hasonló összeolvadása. Ez a jelenség ismert²²³, de nem eléggé közismert.

Ide kapcsolódik az a másik probléma, amit kognitív tudomány a multidiszciplináris jellegét érinti. Kérdés, hogy a tudományok konvergenciájának ténye mellett hogyan kezelendő egy ilyen tudományhalmaz? Közös halmazként való egyesítésük nem tűnik reálisnak vagy működőképesnek, mivel zseniális gondolkodók sorát várná el. Hiszen egy-egy multidiszciplinális tudományhalmazban csak különleges, szintetikus gondolkodással megáldott ember képes valami új megértésig jutni, olyan briliáns elmével, mely egyben látja át a sok résztudomány minden releváns eredményét. És az MI? Nos, az sem hoz még megoldást: a rengeteg tudomány által összehordott rengeteg tudást, képes lesz-e valaha *helyes* szintézissé formálni a korlátolt ember által készített gépi agy (akár korlátolt gép által készített gépi agy)? Számos további speciális kérdés is felvetődik,²²⁴ melyeket itt nem elemzek, csupán egy megoldási javaslatot ismeretek a kogníció multidiszciplináris kezelésének praktikus értelmezésére.

Pléh Csaba akadémikus, a kognitív tudomány egyik hazai úttörője szerint az említett tudományhalmaz csak az egyik felfogás, egy másik megközelítésben vizsgálható a megismerés a részt vevő tudományok metszet-halmazaként is,²²⁵ ami a vizsgálható kogníciós jelekre és ezekből alkotható modellekre összpontosít. Ekkor az érintett tudományok nem rész-, hanem segéd-tudományai a megismerés-tannak, mely csupán használható modelleket kíván kialakítani. Az akadémikus szerint mára mindez tovább alakul egyfajta interpretált kognitív tudománnyá, mely még realiztikusabb, és a területek közötti párbeszédre alapul, konkrét problémák dialógikus

²²² A jelenséget már 1998-tól elkezdték tudományosan azonosítani, de jelentése máig tágul. [261]

²²³ Kritikus beszél a tudományok összeolvadásával kapcsolatban a neves akadémikus, Pléh Csaba [262]

²²⁴ Jó néhány ilyen érdekes kérdést feszeget pl. [258]

²²⁵ Remek ábrákat is közöl a szerző az ismertetett felfogás-módokhoz. [262, o. 1125–1126]

megoldására törekszik.²²⁶ Ez fontos tanulságot rejt ez vizsgálódásom szempontjából is, mivel rámutat, hogy ez a dialógikus megközelítés lehet a realitás, amikor az MI megismerést érintő problematikáihoz keresünk megfelelő perspektívát.

Összefoglalva: ez terület már nem a megismerést magát akarja megérteni. A kognitív tudomány az emberi megismerés objektíven vizsgálható részleteiben igyekszik az adott emberi leképezés legjobb modelljét megalkotni, az adott konkrét kérdéshez kapcsolódó tudományok párbeszéde által. Az így értelmezett kognitív tudomány a leképezés módjait adhatja kezünkbe, vagyis a világ hatékony és biztonságos virtualizációs módjainak megragadásával visz közelebb egyfajta „mesterséges megismerés” hogyan-jához. Az információtudomány „a megismerés éteri anyagát” az információt ragadja meg specifikus sajátosságaiban. Ezért a kettő együttes vizsgálata járulhat hozzá a gépi virtuális világ emberségességéhez.

V.2.3. A mesterséges intelligencia újszerűsége

A száraz fogalmi pontosítást szakítsuk meg egy kötetlen gondolkodással, és olyan kérdések áttekintésével, melyekre bár a pontos választ nem tudjuk, mégis irányt mutatnak vizsgálódásaimban. Amennyiben jól azonosítjuk azokat a jellegzetességeket, melyekben az MI nagyon újszerűnek mondható, ez összevetésre ad módot más technológiákkal is. (Így ezt tudom felhasználni a K4 kutatási kérdésre adandó válaszhoz. (VI.3.3.))

Szokás mondani, hogy nincs új a Nap alatt, ez korunkra azonban nem igaz. A gépek új generációja mely terjedőben van, alapvetően különbözik minden korábbi emberi találmánytól, mivel nem a külvilágban segít, hanem a belső képességeink területein. A régebbi korszakunk újabb és újabb találmányai, vagy az ember fizikai erejét növelték meg, a kerék, az emelő rúd vagy a motor segítségével, vagy csupán az emberi érzékszervek képességeit tágították nagyító vagy infrakép által, vagy a kényelmét segítették. Persze az ilyen lépésekkel is óriási haladást értünk el, de legújabb teremtményeink többet tudnak, mint a mérhető, objektív külvilágban az ultrahang érzékelése, a nanométeres precizitás vagy a hiperszónikus sebesség. Nézzünk néhány jellemzőt, amit igazán újszerűvé teszi az MI-t.

1. Az MI újdonságának lényege az emberi mivoltunkat magát érinti.

Olyan **területeken** képes segítségünkre lenni, amelyekről évezredekken át tűnt egyértelműnek, hogy az ember lényegét adják, így csak ember lehet rá képes, sem állatok, sem emberi találmányok nem. Ebből táplálkozhat az emberiség spontán féleleme az MI-től, melyet

²²⁶ Mondhatom, hogy ez utóbbi rátekintés a kognitív tudományt interdiszciplináris oldalról ragadja meg, viszont a multidiszciplináris oldala tűnik elő, amikor pedig egy kérdés megválaszolásához sok területet egyszerre kívánják figyelembe venni – de jobb nem erőltetni ezt a divatos terminológiát.

erősít, hogy divatos téma a művészetekben. De ugyanebből forrásznak azok az elképzelések is, melyben az MI az evolúció új szintjét jelenti, új létrendet látnak benne,²²⁷ mivel az emberi felsőbbrendűséget szerintük létével cáfolja.

Ráadásul ezek a szubjektív, belső területek, nem mérhetőek olyan objektíven, mint a külvilág gépeinek paraméterei. A híres Turing teszt²²⁸ elvi hibája, hogy „mérőműszere nem hitelesített”: hiszen nem tudjuk mennyire tehetséges és mennyire rutinos (képzett) az a személy, aki a teszt során meg akarja különböztetni a gépi válaszokat az emberitől. És azóta sincs egy, a mérleghez vagy Geiger-Müller számlálóhoz²²⁹ hasonló, a fizikailag objektivitás erejével bíró mérőmódszerünk sem erre, sem sok más MI körüli tényező mérésére.

Emberi mivoltunk gépi támogatása csak úgy lehetséges, illetve csak úgy hatékony, ha a gép minél jobban idomul az emberhez a megismerés téren is. Ez az idomulás több, mint ahogyan pl. egy szoftver felhasználóbarát, vagy amint ergonomikus egér idomul kezünk adottságaihoz. Ez egyfajta „lelki formatervezés”, aminek hatékony működése megoldhatatlan anélkül, hogy jobban megismernénk saját intellektusunkat, érzelmeinket, lelkünket, tudatunkat, akaratunkat, szociális dimenzióinkat, sőt morális érzékünket (és hasonló emberi tényezőket). Ezzel jutottunk el a következő újszerűséghez:

2. Az összetettebb MI rendszerek már működési elveikben messze túlmutatnak a reáltudományok objektíven mérhető keretein.

A kognitív tudományok eredményeinek figyelembevételével valóban képes a gép embe-ribbé válni, sőt a gép hatékonysági mutatói is nagymértékben tudnak javulni. Jelen helyzetben ugyanis olyan újfajta kérdésekkel szembesülünk, melyekkel az eddigi gépeknél nem merültek fel. Pl. hogyan értelmezzük például hősiesség vagy árulás kifejezések információ-tartalmát? Vagy objektívizálhatóak-e, hogyan mérhetőek ehhez hasonló címkék? Márpedig az MI viharos terjedése mellett elő fognak jönnék ilyen kérdések.²³⁰ Ezeket a nyelvi modellek látszólag már ma is megfelelően kezelik, de pl. döntéshozó rendszerek számára ez kevés. Ez mutat rá a fő motivációjára annak a konvergenciának (ld. V.2.2.), mely által az MI technológia beépíti magába humántudományi számos eredményét, így a mérnökökön túl az fejlesztések szerves részévé válnak a humán tudósok: nyelvészek, pszichológusok, szociológusok stb. Ez túlmutat azon, hogy a számítástechnikai célrendszerek esetében a

²²⁷ Ennek egy nem csupán irodalmi, hanem filozófiai megközelítése: [29]

²²⁸ A teszt azzal méri a gépek képességeit, hogy a tesztelő meg tudja-e különböztetni, hogy kérdéseire gép vagy ember válaszolt-e? Az eredeti teszt: [45]

²²⁹ A rádióaktivitás érzékelésére és számlálására használatos berendezés. <https://www.elektroncso.hu/cikkek/gmcs.php> (Letöltve: 2024.01.20.)

²³⁰ Egy másik kérdés lehet, hogy valóban szükségtelenek-e a múlt olyan morális kifejezései, mint pl. becsület vagy gög – vagy netán reneszánszunk várható a gépi morál megalkotásához?

fejlesztésben eddig is részt vettek a célzott gazdasági, egészségügyi vagy állami terület szakemberei, jogászok, gazdasági szakemberek stb. Az újdonság, hogy már a rendszer lényegét, a leképezési és információs modelleket alkotják meg „nem reál” gondolkodású kutatók (pl. nyelvi modelleket, tanítási metódusokat, társadalmi minták kialakulásának módját). Igaz, már Neumann János is foglalkozott az agy vizsgálatával, de korunkig a külső tudományokat inkább csak inspirációnak használta a számítástechnika (egy természetes dolog működésének algoritimizálása volt érdekes). Tehát az MI-ben a reál és a humántudományok eredményei összeadódnak. Ám ezen túl fellép egy lehatároló hatás is: A kognitív tudományok se gíthetnek elválasztani, megkülönböztetni a gép lehetőségeit az emberi lehetőségektől. A reál tudományok nem kompetensek megválaszolni például olyan kérdéseket, hogy mi az értéke az MI jellegű emberutánszásnak? Vagy, hogy ez kevesebb-e, vagy hogyan kevesebb, mint az emberi képességeinkből fakadó eredmények? Továbbá motiváló hatással is lehet az MI a kognitív tudományokra, olyan felismerésekre inspirálva a kutatókat, melyeket nem csupán gépi, hanem emberi téren is használhatnak, pl. önismereti, pedagógiai stb. módszerekké konvertálva azokat.

3. Az MI rendszerek biztonságosságát és az etikus működésüket szavatoló előfeltételek is alapvetően eltérnek a korábbiaktól.

Az eddigi gépeket védőborítással kellett ellátni vagy át kellett programozni, hogy biztonságosabbak legyenek, ezek technikai kérdések. Világos volt, hogy mit kell tenni, mert egy sajnálatosan levágott ujj, az objektív veszteség. Az MI esetén azonban néha nehezebben megfogható, hogy mi is az okozott kár, például, hogy hogyan ítéhető meg „egy kicsit téves” gépi döntés? Egy ilyen tévedés hogyan vethető össze emberi döntések tévességével? Ez sem mérnöki kérdés, ahogyan a felelősségrevonhatóság problémájának filozófiai megközelítése sem technikai, sőt nem is jogi kihívás kihívás. Térjünk rá ezzel az MI problémáinak körére.

4. A klasszikus logika kereti elavulnak. Az életszerűség dimenzióiban nem rideg képletszerűség jellemző. Ám a „nehéz kifejezések” információs tartalmának problémáin túl, kérdéses, hogy hogyan kezeljük a sokjelentésű klasszikus szimbólumokat és az élet paradoxonait? Erre példának azt hoztam (IV.1.2./ (3.), zéta szint), hogy a tűz vagy a víz egyszerre pusztító és éltető szimbólum, és lehet egyszerre szeretni és gyűlölni valakit. Ezen a vonalon tovább lépve olyan nehéz kérdésekhez jutunk, ahol sokkal jobban tisztázni kell a tudat és az információ kapcsolatát, például, hogy hol van helye a gépi tudat emulálásánál a tudatalattinak, vagy esetleg valami tudatfelettinek, van-e helye ott a megérzéseknek vagy intuíciónak?

5. Humánvirtualizáció. Ki kell emelni, hogy az analóg világot bensőnk másképp egyszerűsíti le, mint a gépek (ld. IV.2.2.). Alapvető eltérés, hogy memóriánk egy automatikusan

átdolgozott, szubjektív leképezést tárol, ez jelenti a világból nyert információt. Tudat alatt is folyamatosan átdolgozzuk a tárolt dolgokat: próbáljuk kipótolni a hiányzó információt, összefüggéseket találni, és érzelmeket társítunk hozzájuk. Ezt torzítást néha hibaként, néha művészi látásmódként lehet értékelni – de mindenképp „emberi”, tökéletlenségünk összes szépségével. A két féle virtualitás pontos összevetése metaverzumok vizsgálatakor lehetne igazán érdekes, most ne kanyarodjunk el ebbe az irányba. A téma vizsgálatát helyhiány miatt ki kellett hagynom jelen értekezésből, de igyekszem majd közzétenni.

V.3. JAVASLAT AZ MI MEGHATÁROZÁSÁRA

A megfogalmazás előtt tekintsük át az eddigi kutatásokat és a definiálás elvi határait.

V.3.1. Besorolási mátrix a túl sokféle MI-hez

Az MI egy igen összetett fogalom, ezért úgy véltem hasznos áttekintenünk a legfontosabb felosztásait lexikon-szerűen, ugyanis tematikus összegzésre nem találtam minőségi forrást, pedig a szakirodalomban számtalan felosztás található.²³¹ Célszerűbb volt hát a szakirodalom és saját eddigi kutatásaim alapján saját áttekintést adni a témáról. Ez egy hasznosítható eredményt hoz létre, hiszen ezek által egy adott rendszert a sok alábbi szempont szerint besorolva, és a szintekhez szabályozókat társítva, a technológia szabályozhatóságának pontossága javul. E célból még egy vizuális összesítést is javaslok. Tehát jelen kissejlesztés eredeti célja az volt, hogy hozzájáruljon az új MI-fogalomhoz, de arra jutottam, hogy abban nem szükséges ezeket megjeleníteni. Ezért itt a cél olyan felosztásrendszerek vázolása, mely a biztonsági, jogi vagy gazdasági struktúrák számára a különböző MI rendszereket megragadhatóbbá teszi, és az MI oktatásakor is segít. (Helyhiány miatt a rövid szavakból álló listákat vízszintesen szedtem, ezek máshol rendes felsorolásokká tehetőek).

(1.) A három alaptulajdonságból adódó felosztások

Az alábbi felosztásokat korábban már ismertettem, itt csak rendszerzettebben tekinthető át.

- I. A gépi tanulás architektúrája szerinti (hány neuronréteget használ a be- és kimenet között):
 1. pszeudo-tanulás (tanulás benyomását tevő hagyományos kód, nem neuronháló)
 2. gépi tanulás (ML)
 3. mélytanulás (DL)

²³¹ Inkább közérhetőnek szánt (nem tudományos) cikkek vesznek át egymástól zavaros és részleges felosztásokat, úgy tűnik a kiinduló cikk talán ez: [263]

II. Tanulási mód szerint (főbb) szintek: (1) Perceptron; (2) ANN; (3) CNN; (4) LSTN; (5) GRU; (6) RNN; (7) CRNN; (8) GAN; (9) GPT.

Megjegyzendő, hogy ez a lista folyamatosan bővítendő, hiszen folyamatosan állnak rendszerbe újabb és újabb megoldások.

III. A gépi tanulás tanítás szerinti felosztása:

- (a) teljesen felügyelt, (b) félig felügyelt, (c) felügyelet nélküli tanulás.
 - Az a) és b) esetben felmerül az irányítás módja szerinti felosztás:
 - (i) megerősítés nélküli; (ii) megerősítéses tanulás alapján

IV. Elterjedt autonómia szintek emberi döntéstípusok szerint:

1. **racionális autonómia** (funkcionális moralitás)
2. **operatív autonómia** (operatív moralitás)
3. **erkölcsi autonómia** (morális viselkedés)
4. **személyes autonómia**

V. Hagyományos gépiautonómia szintek az emberi döntés mértéke szerint:

1. programozott automatika
2. az optimális eset megtalálása sematikus esethalmazra
3. gépi kreativitás és empátia alapján igen nagy esethalmaz kezelése
4. teljes gépi autonómia, a betanító erkölcsi rendszerének tökéletes alkalmazása
5. emberi vagy ember-feletti felelős szabadság

VI. Saját Általános Négy-típusú autonómia (4TA)

1. Egyszerű autonómia
2. Összesítő autonómia
3. Szabálmögötti autonómia
4. Szélsőhelyzeti autonómia (hősies vagy önző viselkedés)

VII. Saját meghatározású gépi autonómia szintek

- Nulladik szint az egyszerű, programozott automatika.
- Alfa szint: a pszeudo-autonómia szintje bizonyos adatbázis háttérű programoknál.
- Béta szinten jelenik meg a gépi tanulás.
- A gamma szinten jelenik meg a gépi kreativitás és empátia
- Delta szint: magára hagyható komplex autonómia.
- Epsilon szint: emberrel szinte egyenlő gépi szabadság, illetve autonómia.
- Zéta szint: autonómia 2.0.

VIII. A kogníció-típusok szerinti felosztás:

- Az intelligencia különböző típusait szimuláló rendszerekhez hozzávehetjük az emberiesség minden olyan szegmensét, amelyet valami módon gépileg másolni, szimulálni tudunk. Ám a fejlettség mai fokán célszerű bizonyos, jelenleg kevésbé frekvenciát szegmensek egybevonása.
1. **Gondolkodás:** Ezen belül számos szintet érdemes megkülönböztetni, hiszen ide tartozik minden a sakkgéptől a virtuális kémialaborig (amely sok ezer anyagot szimulálva adja meg a szükséges vegyi képletet).
 2. **Érzés:** felosztható input és output szerint, vagy szenzortípusok szerint.
 3. **Érzékelés.** A következő területek azokat az intelligenciaterületeket idézik, melyek az emberi érzékeléshez kötődnek
 - a) látás: képértés, arc és ujjlenyomatfelismerés, képalkotás
 - b) hallás: hang-szöveg konverter, hanganalízis, zene-elemzés stb.
 - c) szaglás, ízlelés vegyi analízise és a tapintás fizikai analízise
 4. **Képességek.** A bemutatott képességjellegű intelligenciaterületek alapján:
 - a) beszéd: beszéd-konverzió, szövegértés, válaszszöveg, beszéd-szintetizátor
 - b) mozgás: jármű/robot autonóm mozgatása terepen, okosváros-szabályzás, orvosi video-lelet-elemzés
 - c) Egyéb: a sokféle „alternatív” intelligenciaterület szimulációjára fókuszáló fejlesztés

IX. A tudásszintek szerinti mai felosztás

- (1) ANI (week); (2) AGI (strong) - Saját felosztásom ezen belül (i) technikai AGI; (ii) emberi AGI; (3) ASI

(2.) Egyéb MI felosztási lehetőségek

Csupán néhány érdekes vagy fontos felosztás:

X. Biológiai architektúrával való kapcsolat szerint:

1. csak szoftveres (hagyományos architektúrán)
2. célhardverrel is támogatott (jelenleg szoftveres része is van)
3. állati vagy emberi agyba implantált, közvetlenül kiterjesztve adott agyi képességeket.

XI. Központi működés szerint:

1. centralizált rendszer (online) – egyetlen központi gépen vagy felhőben fut, alrendszeraitől csak átvesz adatokat

2. decentralizált rendszer (offline)– csak egyedekből áll, melyek önálló entitásokként is működnek, de együttműködve, közösségként képesek hatékonyabb vagy optimálisabb feladatvégrehajtásra (pl. rajintelligencia).
3. centralizált entitáshalmaz (fél-online) – az előző két formát úgy egyesíti, hogy az egyedi entitások tudása centralizáltan is feldolgozásra kerül, és ennek eredménye visszatöltődik az egyedek tudásába (pl. Tesla okosautótanítási modell).

XII. Az emlékezés összetettsége szerint.

Itt nem a memória nagysága a kérdéses, mint a hagyományos memória esetén, hanem a memória MI-sajátossága, azaz annak összetettsége.

1. **Reaktív rendszerek:** nincs tanulómemóriájuk, előre be vannak tanítva (pl. a DeepBlue sakk-MI);
2. **Korlátozott memóriájú rendszerek:** egy naplószerű emlékezetük van, mivel egyszerű tanulást használ. Ez persze foglalhat el nagy tárhelyet, de használata szimpla (pl. okos-termosztát);
3. Emlékező²³² célrendszerek: a mélytanulás révén emlékezetük komplexebb és pontosabb, de egycélú rendszerek (képfelismerő rendszerek). Ezen belül szétválaszthatóak a rövidebb (CNN, RNN) és a hosszútávú memória (GRU, LSTM) rendszerek;
4. Komplex gépi emlékezet: multimoduláris MI, melyben többféle neuronháló (akár egyesben egyszerű és mély) architektúra összekapcsolása által az emlékezet kezd az emberi emlékezet sokféleségéhez hasonlóan sokrétűvé válni. Ezen belül további felosztásokat a felhasználás eredményezhez (pl. egy beszédértő hangulatfelismerő rendszer).

XIII. Felhasználási mód szerint:

- Talán a legnépesebb felosztást ezek a felosztások jelentenék (pl. egészségügyi, ipari, művészeti, közigazgatási, hadi, rendészeti stb.), ezért ezt nem részletezem.²³³

XIV. Kockázat szerinti felosztás (pl. az EU hivatalos besorolási módszere [264, o. 5.2.2.]

- Az alábbi lista egy szabályzás-orientált megközelítés, a szintek neve árulkodnak a velük kapcsolatos szabályok szigorúságáról. Gyakorlati problémát a konkrét rendszerek besorolásának metodológiája, illetve ennek naprakészen tartása jelent.
1. Elfogadhatatlan kockázati rendszerek
 2. Magas kockázatú MI-rendszerek

²³² Több írásban a 3. szintet az Elme-elméleten alapuló rendszer (Theory of Mind AI), a 4. díket a az Öntudatos rendszerek (Self-aware AI) kapják – ezeket azonban nem memóriájuk jellemzi, ezért elvettem.

²³³ Egy több ezer elemes, sok szintű listában lehetne összegezni azokat a gyűjtő-területeket, alterületeiket és azok részterületeit, amire különböző célrendszerek alkalmasak, hiszen más technológiákkal karöltve az MI mai képességei valóban az életünk minden területén használható.

3. Korlátozott kockázati kategória
4. Minimális vagy semmi kockázatot nem jelentő MI

XV. Gondolkodás-cselekvés mátrix (nem életszerű 2 és 4. miatt a fenti bizonyítás szerint)

- Érdekes megközelítés az emberi gondolkodás- és a cselekvés-szintekhez való viszonyítások megkülönböztetéséből adódó gépi definíciós mátrix. [15] :
 1. „értelmesen gondolkodik” (emberhez hasonló + gondolkodik)
 2. „úgy gondolkodik, mint egy ember” (az emberivel azonos + gondolkodik)
 3. „értelmesen cselekszik” (emberhez hasonló + cselekszik)
 4. „úgy cselekszik, mint egy ember” (az emberivel azonos + cselekszik)

Természetesen egy ekkora terület esetében számtalan további felosztási lehetőség merülhet fel: tesztelhetőség szerint, elterjedtség szerint, energia-fogyasztás-szerint stb. Azonban, mivel céloom azt átgondolni, hogy mely felosztás szempontok mennyiben használhatóak az MI fogalom pontosításához, a talált, vagy bennem felmerült egyéb megközelítéseket helyből irrelevánsnak (vagyis kihagyhatónak) találtam. A szakirodalomban fel is merül jó néhány egyéb felosztás. Sokszor összekeverednek, egybemosódnak a különféle felosztási szempontok.²³⁴ A besorolási mátrix szükségességének belátásához úgy vélem elegendő ennyi féle felosztás is.

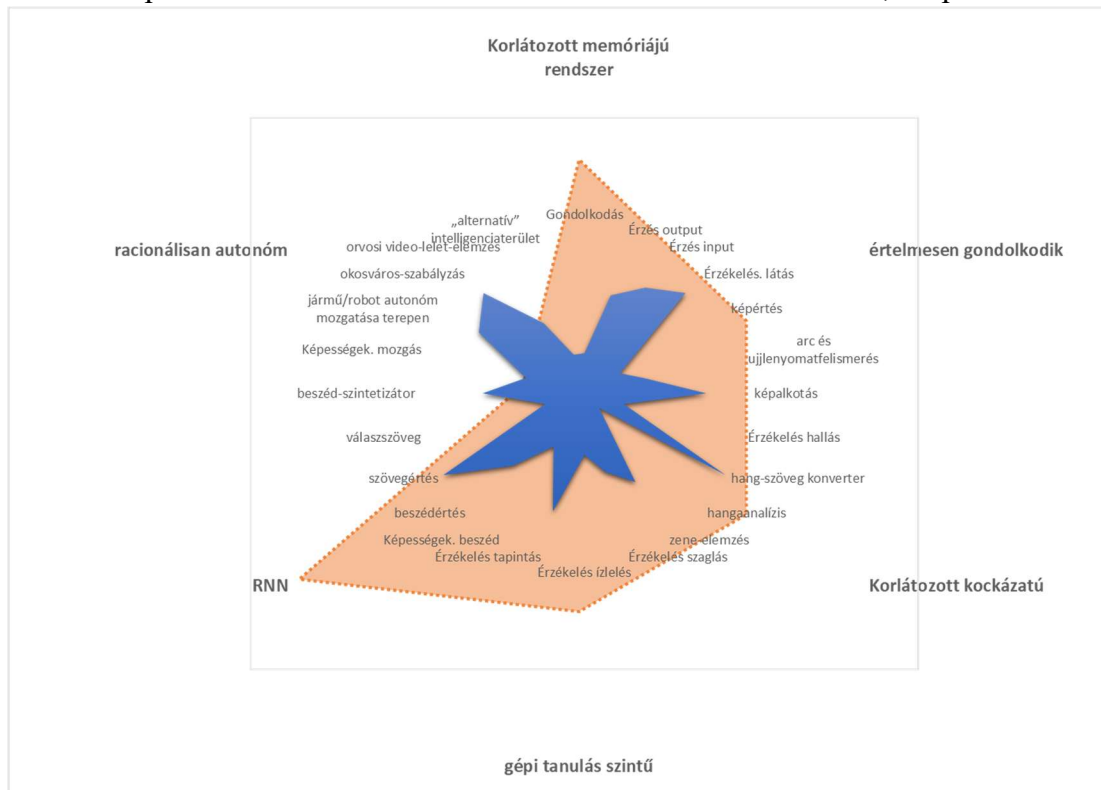
(3.) A besorolás mátrix felvetése

A technikai nehézség a besorolások objektívvé tétele, ehhez azonban alkalmazhatóak a már bevált ipari módszerek. Sőt, ezen a téren véleményem szerint reális nemzetközi szabvány-megállapodásokban konszenzusra jutni, nem úgy, mint etikai vagy korlátozási kérdésekben. A másik nehézség, hogy hiába kapunk sokféle besorolás szerint sok, az adott rendszert jellemző számértéket, ezek összege vagy átlaga (ha súlyozott is) nagyon csalóka lesz, és használhatatlan mérőszámot kapnánk így a rendszerről. Azonban a régebbi „exceles” diagramm-típusok kevesnek bizonyulnak ennyi paraméter figyelembevételéhez. Például radar diagrammon már akkor sem kapunk szép ábrát, ha csak a legfontosabb besorolásokat vesszük fel. Alább a 15. számú ábrán a kogníciós (kék) és a többi besorolás (sárga) együttes ábrázolását kíséreltem meg.

Ezért tehát a mért adatok összességét lenne hasznos figyelembe venni, amihez adatvizualizációt lesz érdemes használni. Ehhez talán valamely BigData elemzésben bevált vizualizációs módszer lehet hasznos, mely a 16. ábra alapján képzelhető el. Például egy színezett, háromdimenziós (forgatható) túske-diagram gomb vagy henger alakban. Ezen a különböző besorolások

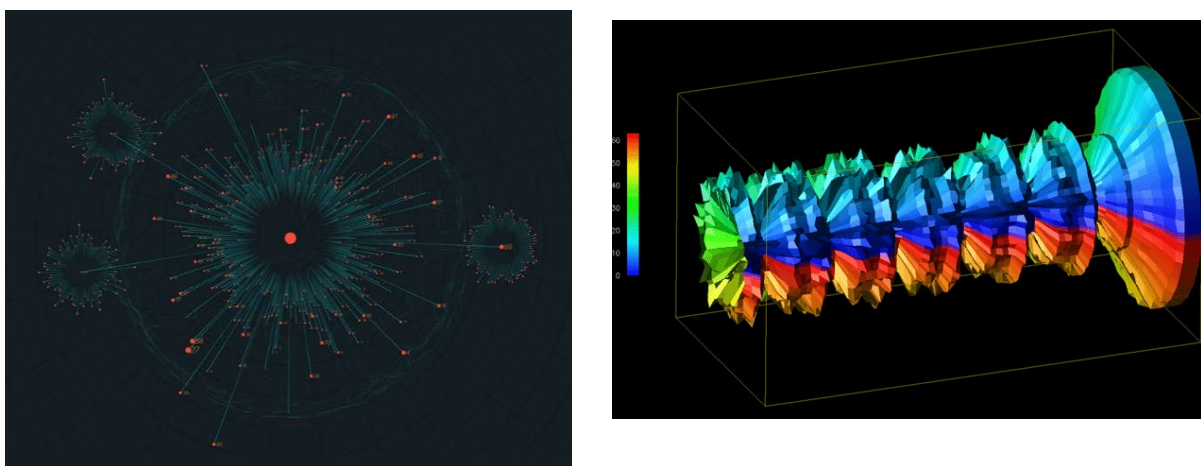
²³⁴ Pl. az fenti elméleti felosztás és némely felhasználási mód keveredik az elsőre látványos ábrában ebben tanulmányban: [265, o. 4]

alapján kapott számértéket egy-egy tűsként állnak különféle irányokba. A kategóriákat színezések kapcsolják össze. A kapott értékekből létrejövő „színes tűske-krumpli” mutatná az adott rendszer paraméter-halmazát. Erre kész szemléltetést nem találtam, csupán kétdimenziós



15. ábra: Besorolás-szintek exceles ábrázolása (saját készítés)

változatot,²³⁵ a hengeres változathoz igen, ezek láthatóak a 16 ábrán. Mindkét vizualizáció sok szempontot képes ábrázolni. Azonban mindehez hozzá kell tenni, hogy egy rendszer akkreditációjához nagy számú vizsgálatra, tesztre és mérésre lenne szükség, mely nyilván drágítaná a



16. ábra: MI besorolás-mátrix lehetséges vizualizációi (szemléltetés, forrás: internet)

²³⁵ az ábrák forrása 3. ábra: <https://www.vectorstock.com/royalty-free-vector/big-data-circular-visualization-futuristic-vector-19630392>, 4. ábra: https://datavizcatalogue.com/blog/further_exploration_2_3d_chart/

szolgáltatást – ám ez nem jelenthet igazi problémát, hiszen elterjedt és megszokott ártényező. Mivel ezek alaposabb kifejtése jelenlegi vizsgálataimat nem segíti, nem elemzem a kérdést mélyebben. Viszont a jövőbeni szabályozáshoz hasznos lehet, így majd egy későbbi vizsgálódás tárgyát fogja képezni.

V.3.2. A fogalmakból hiányzó tényezők és hatásuk

Itt röviden összegzem az I-V. fejezetek alapján mely kifejezéseket vélek szükségesnek egy teljesebb MI-definícióban megjeleníteni. A 3. sz. táblázat K3 kérdésre ad választ

	Összegzés	kifejezés	MI-def.	oktatás	nem
1.	Több intelligenciafajtát. Több mai hivatalos meghatározásnál kimutattam (I.1.3.), hogy azok az ész (gondolkodás, megismerés, tanítás, autonómia kulcsszavak körül fogalmazódnak meg, ezekkel szemben azt a kritikát fogalmaztam meg, hogy túlságosan is ész-centrikusak, nem domborodik ki az intelligencia sok egyéb fajtája (I.2.2.), holott azok gépi szimulációjára viszont folynak a kutatások, amit az affektív számítástechnika bemutatásával (I.3.) demonstráltam. Az új meghatározásnak elsősorban ezt a hiányosságot kell javítania, megjegyezve, hogy a magyar megközelítés ezt a hibát elkerülte. (I.1.2.)	intelligenciafajták	+	+	
2.	A nyelvi elemzés (I.2.1.) alapján nem került elő olyan jelentés, melyet be kéne építeni az új fogalomba. A legtöbb jelentésréteg megjelenik, a hiányzók (pl. megértés, felismerés) megjelenítése nem indokolt.				+
3.	Az MIKT fogalmát (II.1.) is bevezettem, mely megkülönbözteti az MI lényegét azokról az igen összetett rendszerektől, melyek „körbeveszik”, és fontos pilléreként épülnek az MI-re. Ezt érvényesíteni kell az új fogalomban, viszont a digitális ökoszisztémát nem szükséges.	MIKT	+	+	

	Összegzés	kifejezés	MI-def.	oktatás	nem
4.	A tanulási modellek és gépi tanulás felosztásai általános meghatározáshoz nem szükséges technikai többletet közölnek (II.2.), továbbá az MI hardveres háttere és a fuzzy logika (II.3.1-2.) sem releváns a definícióban.			+	
5.	A rajintelligencia , rámutat a decentralizált MI rendszerek fontosságára, illetve utalás biológiai együttműködések digitalizálására (II.3.3.) megfontolandó, mivel ennek beemelése egy alaposabb definíciót adhat.	rajintelligencia	+	+	
6.	A szinergikus MI (máshol: szimbiotikus MI), vagyis olyan rendszerek melyekkel jó együtt élni, természetes nyelvfeldolgozás (NLP rendszerek II.3.4.) kapcsán vált láthatóvá, belefoglalása javasolt.	szinergikus MI	+	+	
7.	A kiterjesztett ember megoldások a „mesterséges” szó újragondolását tehetik szükségessé, vagy pontosítandóvá, a jelenleg célzott fogalmat zavarossá teheti megjelenítése.			+	
8.	Az MI tétel témaköre inkább hatásokról és várható tendenciákról szólt melyek hatását a fogalomra ugyan elemeztem (pl. az ál-AGI megjelenésének jóslata III.6.1., vertikális tanulás szükségessége III.5.1.), ám a jelenlegi definíciót módosító tényezőt itt nem azonosítottam.				+
9.	Az erkölcsi vizsgálatok (IV.) még több fogalmi vizsgálata és elemzése hasonlóképpen „csak” az összefüggések megértéséhez voltak hasznosak, egy jelenlegi fogalom-javaslatba elsősorban a rajintelligencia (IV.2.5.) merült fel, amit azonban fentebb már javasoltam. A gépi humánkomponensek				+

	Összegzés	kifejezés	MI-def.	oktatás	nem
	megvalósítása pedig nagyon gyerekcipőben jár, ezt a kifejezést ezért nem gondolom az új fogalomba valónak.				
10.	A definíciót ugyan nehézkessé tenné az olyan jellegű tisztázás, mely például elhatárolja a pseudo-MI-től, de a fogalom magyarázatakor (oktatásban) elengedhetetlennek tartom a fogalmak összemosódásának ezt a fogalomkörét, és hasznos lehet és fogalom-torzulás említése is (V.1.).			+	
11.	Az informatika szó kerülését javaslom, mert amint kimutattam, az informatika szót épp az MI fényében nem lesz pontos a számítástechnika szinonimájaként használni (V.2.1.).	infor- matika			+
12.	Bár a tudományok konvergenciája, mely az MI hatására létrejön (V.2.2.) rendkívül fontos, de fogalmilag ez sem érvényesítendő.			+	
13.	A rétegek száma és az ember részvétele szerinti felosztások fontos jelzők lehetnek nem szükséges			+	
14.	Az MI (V.3.), sokféle felosztási módja szabályozási és megértési szempontból jelentős lehet, főleg, ha például a besorolás-mátrix segítségével egzakttá tesszük, de nem lehet ezeket a definícióba zsúfolni.			rész- ben	
15.	tanulás (I.2.3.), az intelligencia (I.2.2.) és az autonómia (IV.1.) felosztási módjai sem belefoglalmandók			rész- ben	

3. táblázat: Az új MI definícióban megjelenítendő tényezők

Az összesítésből látható, hogy négy kifejezést azonosítottam, melyeket meg kell jeleníteni, és egyet (az informatika szót) kerülni szeretném. Itt megjegyzendő, hogy úgy a táblázatból tűnhet, hogy a III. és IV. fejezetek nem járultak hozzá az új definícióhoz, ennek ellenére számos megállapítással alapozták és készítették elő ennek a fejezetnek a vizsgálatait.

Viszont a fentiekhez kapcsolódva a 4.sz. táblázatban szükségessé vált bemutatni, hogy milyen szempontból érthetőek félre a megjelölt kifejezések nélkül a jelenlegi MI fogalmak, ezáltal pedig igazolni tudom a H3 hipotézis második felét is (melyben azt állítottam, hogy bizonyos kifejezések nélkül az MI fogalom félreérthető) .

hiányzó kifejezés	okozott félreértés
intelligenciafajták	Egy emberértelmezési (filozófiai-antropológia terén jelentkező) hiba, melyet fentebb sok helyen említettem, az intelligencia okosságra szűkítése, melyet gyakorlatilag már rég kinőtt a megvalósult technológia.
MIKT	Mivel egy-egy MI funkció általában csupán egyik ágense valamely hagyományos rendszernek (pl. arcfelismerés egy telefonban), ezért mindig szükséges lenne világossá tenni, hogy egy adott rendszerben hogyan van jelen az MI: szüksége van-e külső kapcsolatra, vagy „offline” is működik? Egy ilyen irányú félreértés biztonsági szempontból is nagy kockázat.
rajintelligencia	Egy okos-entitás halmaz felépítése és használhatósága pont olyan mértékben különbözik a központi erős gépen futó MI modelltől, mint egy hangyaboly egy majomtól. Persze mindkettő „állat” – de működésüket és célelésüket tekintve alapvetően mások, ahogyan erőforrás igényüket is eltérően szükséges kezelni. Ezt az egy felosztást (a sok felsorolt közül) azért szükséges definíció-szinten is megemlíteni, hogy rámutassunk a két megközelítési paradigma eltérésének jelentőségére.
szinergikus MI	Az embert komplementer módon kiegészítő technológia olyan cél, melyet érdemes minden fejlesztő számára elrendővé tenni és mind szabályozási mind pedig ellenőrzési vagy értékelési szempontból alapvetővé tenni. Ennek a jelzőnek a megjelenítése tehát az MI-től való félelem félreértését hivatott eloszlatni (vagy legalább mérsékelni).

4. táblázat: *Miért félreérthetőek egyes kifejezések nélkül a jelenlegi MI-meghatározások?*

V.3.3. Új MI meghatározások korlátai

Mielőtt minden eddig leírtak összegezve, ismertetem a javasolt MI fogalmamat, fontos tudatosítani, hogy elméletileg is lehetetlen egy „tökéletes” MI meghatározás, hiszen a jelenben feloldhatatlan elvi korlátokba ütközünk a következők miatt:

1. **Túl gyorsan változik minden technológia.** A MIKT terén folyamatosan az újdonságok bejelentései, melyek folyamatosan új kihívásokat is generálhatnak – elég csak a ChatGPT megjelenése utána EU-s reakciókra utalnom.[266] Ennek egyik eleme, hogy a konszenzus vagy bármilyen szabályozás nem csupán nehéz, de mire létrejöhetne már okafogyottá is válhat.
7. **Túl sok minden tartozik és tartozhat bele az MI-be.** Mint láthattuk, az ember és az élővilág különböző intelligencia-részeit és az érzékelési módjainkat teljesen másképp kell gépekbe leképezni. Így ezen technológia részei (moduljai, fejlesztési irányai) között lényegesebbnek tűnnek az eltérések, mint a klasszikus számítástechnika moduljai között.
8. **Túlságosan új az ember-szerűség ilyen formája.** Manapság még, mint szinte minden technológiában a „nagyobb hatékonyság” a fő jellemző, de a jövőben talán más, emberibb mottó felé tolnak a fejlesztések, és ez változat a fogalmon is.

V.3.4. A javasolt MI fogalom

Érdekes több, különböző alaposságú megfogalmazást adni, hiszen egy oktatási vagy egy szabályozási elvárás teljesen más. Kezdem az utóbbi célra szánt legbonyolultabb javaslattal:

A mesterséges intelligencia a kognitív képességek egy körének virtuális leképezése egy számítástechnikai rendszerbe. Az így kapott architektúra egyes moduljai képesek az emberi intelligencia egy-egy területén a tanulás meghatározott szintjét megvalósítani, elsősorban az értelem és az érzelmek (illetve ezek affektusai) terén. Ezáltal a programozott, determinisztikus reakcióikon túl, képesek az autonómia adott szintjén a tanult adatok alapján hozott, indeterminált, időben (a tanulás folytatása miatt) változó, rá jellemző sajátos döntésre, válaszra vagy kreativitásra is. Az MI rendszerek minden téren a legnagyobb optimalitásra törekszenek, és ideális esetben komplementerként, szimbiotikusan egészítik ki az emberi képességeket.

Nem MI-ről, hanem MIKT rendszerről, (azaz Mesterséges Intelligencia és Hozzá Kapcsolódó Technológiák rendszeréről) beszélünk, amikor a felhőben működő MI szorosán összefüggő rendszert alkot hagyományos adatbázisokkal, strukturálatlan Big-Data adathalmazokkal, IoT szenzorokkal vagy egyéb infokommunikációs technológiákkal. Az MI szoftveres vagy hardveres (a jövőben talán emberbe beültetett) megvalósulása feladattól függően egyaránt lehet központi, és lehet decentralizált, vagyis egyedi entitások közösségégi tudásában (rajintelligencia) implementálva.

A fenti definícióból egyszerűsített, bevezető kurzusokon, vagy publikációkban használható MI meghatározásom a következő:

A mesterséges intelligencia a kognitív képességek egy körének virtuális leképezése egy számítástechnikai rendszerbe, mely ezáltal képes az emberi intelligencia egy vagy több területén a tanulásra. A programozott viselkedési szintjükön túl, elsősorban az értelem és az affektusok utánzásával képesek a bizonyosfokú autonómiával, netán kreativitással reagálni a világ változásaira, minden téren a legnagyobb optimalitásra törekedve, lehetőleg szimbiotikusan együttműködve az emberrel, de rajntelligenciaként más elektronikus (akár más MI) entitásokkal is együtt tud működni. Gyakran MIKT architektúrában használatos, ahol az MI egyéb infokommunikációs technológiákkal alkot egy rendszert.*

**azaz Mesterséges Intelligencia és Hozzá Kapcsolódó Technológiák*

V.4. ÖSSZEFOGLALÁS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI FOGALOMVIZSGÁLATA

- **A várakozások teljesülése.** A kutatás elért eredményeket. Nem lehetett cél minden, az MI-hez kapcsolódó aspektus vizsgálata, és C3-ban nem is ezt jelöltem ki. Tehát maradhattak ki MI-vel kapcsolatos tényezők, de így is egy meglehetősen komplex vizsgálat született. Ennek során meghatároztam bizonyos vonatkozásokat, melyek hiányoznak a jelenlegi MI-fogalmakból, tehát K3 kérdésre választ adtam. Bár én sem ítéltém minden hiányzó aspektust beépítendőnek, nehogy túl bonyolult definíciót kapjunk, mégis a beépítendő tényezők azonosításával és a hiány indoklásával egyben a H3 hipotézis is bizonyítást nyert (ld. V.3.2.). C3 célról is kimondható, hogy teljesítésre került, mégpedig a tanulmány I-V. fejezetek egésze során, hiszen a legtöbb részkutatás során felmerültek fogalmak, melyeket pontosítani, más-képp felosztani vagy épp megalkotni kellett.
- **Összegzés.** Mint említettem számos, a C3 célhoz tartozó vizsgálat az I-IV fejezetek során lezajlott, de néhány vizsgálatot még elengedhetetlennek tartottam. Egyrészt szükség volt fogalmi összemosódások (V.1.1.) szétválasztására, amelyre két jelenséget is elemeztem: az „okos” szó degradálódását (V.1.3.), és az MI-nek tűnő, de hagyományos rendszerek bemutatását (V.1.4.). Még egy elemzés hiányzott, tisztázni kellett az informatika és információtudomány kapcsolatát az MI-vel, valamint megvizsgálni az MI alapvető újszerűségét, részben a fogalmi kapcsolatok (az informatika szó módosulása) miatt, részben a védelmi vonatkozások (VI.) előkészítéseként. A különböző MI-felosztások lexikon-szerű felsorolása (V.3.1.) hasznos gyűjtemény lehet minden témával foglalkozó számára. Ezek összesítése túlmutatott jelen kereteken, de bemutattam egy besorolási mátrix javaslatát is (V.3.2.), mellyel sok ilyen felosztás együttes figyelembevétele alapján lehetne egy adott rendszer tudását egy palettán

elhelyezni és vizualizálni. Végül az új MI-fogalmakat javasoltam (V.4.3.), ehhez pedig összesítettem a korábbi megállapításokat (egybe bizonyítottam a H3 hipotézist (V.4.1.)), szóltam a várható korlátokról, végül megvizsgáltam a kapott megközelítés előnyeit (V.4.2 és 4.).

- **Részkövetkeztetések.** következtetéseket tudtam levonni:
 - **3R-A:** Az MI fogalom hiányosságainak feltárásából következett, hogy javasolni szükséges egy teljesebb fogalmat, így jutottam el a 4. eredményhez, ugyanis megfogalmaztam a Mesterséges Intelligencia fogalmának olyan definícióit, melyek a korábbiaknál több tényezőt vesznek figyelembe (még kompromisszumok mellett is).
 - **3R-B:** A 1R-B és 2R-B részkövetkeztetésekhez itt további következtetések köthetőek, így jobban megalapozottá vált a harmadik eredmény, miszerint az általános MI ember-szerűségének lehetőségével, valamint a szuperintelligencia emberfelettségével kapcsolatos felvetések nem tudományos megalapozottságúak. Ezt fentebb annak bemutatásával támasztottam még alá, hogy a fogalmi degradáció (okosság, AGI) sémája egy általános mintaként alkalmazható minden technológiára (V.1.) – előrelépés, hogy így ez a jelenség tágabban is értelmezhetővé válik.
 - **3R-C:** Egy itt ki nem használt következtetés, hogy a különféle rendszerek látványos és hasznos besorolást kaphatnának a besorolás-mátrix modell segítségével (V.3.2.)
 - **3R-D:** Az informatika teljes átalakulása várható, és ez a védelmi szférában is alapvető változásokat generál (V.2. – vö. VI.1.)
 - **3R-E:** Az MI újszerűségét megragadó elemzés (V.2.3.) felhasználható, amikor az MI-t valamilyen más diszrupcióval összevetni akarjuk. (pl. a kibertérrel: VI.3.3.)

VI. MESTERSÉGES INTELLIGENCIA ÉS ERŐÉRVÉNYESÍTÉS

Az eddig ismertetett kutatásokban is igyekeztem érvényesíteni azt a tervezéskor lefektetett szempontot, hogy védelmi aspektusból is vizsgáljam a kijelölt területeket. Itt érkeztem el ahhoz a fejezethez, melyben ezek a szálak összeérnek, és C4 célnak megfelelően olyan vonásokat tudok beazonosítani – reményeim szerint a legfontosabb szempontokat – melyekben jól látható az a hatás, ahogyan az MI korunk katonai és védelmi szférájára jelentős hatással van. Bizonyos eddigi következtetések és leírások is tártak már fel ilyen vonásokat, az alábbi elemzésekben ezeket egyértelműbbé teszem, és új aspektusokkal is bővítem. Több vizsgálattal is igyekszem igazolni a H4 feltevést. Először a hipotézis első felére koncentrálok és kimutatom, hogy az MI valóban tovább tolja az erőérvényesítés hangsúlyát a puha műveletek felé, ehhez bemutatom azt a paradigmaváltozást, amely felé az erőérvényesítés az MI hatására jelenleg toródik, valamint a technológiával való visszaélési lehetőségeket.

A hipotézis második fele azt állítja, hogy az így kialakult helyzet kezeléséhez a katonai és védelmi informatika radikálisan új megközelítése lesz szükséges. Ezt is több módon közelítem meg. Az igazolást attól várom, hogy igenlő választ kaphatok a K4-re, ahol azt a kérdést tettem fel, hogy lesz-e olyan alapvető hatása az MI-nek az állami védelmi elméletekre és rendszerekre, mint a kibertérnek? Ám a felvetett új megközelítés jellemzését a két technológia által generált újítások összevetése nem adja meg, ezért az igazoláson (összevetésen) kívül összesítem a korábbi fejezetek ideillő védelmi vonatkozásait, valamint a katonai (védelmi) informatika várható átalakulását is elemzem. Ez utóbbit olyan területen demonstrálom, ahol otthonosabban mozgok: az MI megjelenítésének lehetőségeit vizsgálom a katonai (védelmi) felnőttoktatásban.

VI.1. AZ MI-ALAPÚ HIDEGHÁBORÚ: A HIBRID MŰVELETEK SPECIÁLIS ELEME

Miért is nevezhető a napjainkra kialakult helyzet egy második hidegháborúnak, és mi a szerepe ebben az MI-nek? Röviden azért, mert az MI nem csupán alapvetően újszerű katonai műveleteket tesz lehetővé az autonóm fegyverrendszerek által, vagy a kibertéri műveletekben, hanem információs vagy gazdasági műveletekben alkalmazva alkalmas a politikai folyamatok hosszútávú és tervezett befolyásolására. Az államok közötti erőérvényesítés ilyen nem-fegyveres formáit hagyományosan nem tekintették harcnak, ám a helyzet változik.

VI.1.1. Paradigmaváltás a nyomásyakorlásban

A modern technológiákkal kapcsolatos elsődleges etikai kérdés, hogy élünk-e azokkal, avagy visszaélünk velük. Különösen kielezett ez a helyzet katonai szempontból, mivel a

technológiai újítások apokaliptikus méretű áldozathoz is vezethetnek. Az atomkorszak fenyegetése sem múlt el, de az emberek az MI által vezérelt autonóm fegyverektől, a „gyilkos robotoktól” félnek. Ezzel szemben a valóság az, hogy a haladás által nyújtott lehetőségekkel egyelőre nem a katonai területen lehet leghatékonyabban visszaélni. Jóval optimálisabban elérhetőek a kívánt eredmények az MIKT (II.1.) egyéb bevetési módjaival. Ezért a fejlődés jelen fázisában a rombolásra és emberéletek kioltására alapuló klasszikus katonai módszerek alkalmazása mellett az erőérvényesítés más módjai egyre inkább előtérbe kerülnek. Ezeknél a politikai vagy gazdasági hatalom birtokosai nem csupán a technológiai újításokat, hanem a társadalomtudományok eredményeit is bevetik. A legkorszerűbb szervezési, kommunikációs, pszichológiai, szociológiai, gazdasági módszerek legalább olyan fontosak, mint a high-tech eszközök. Ebben az összefonódási folyamatban egymást segíti a technológia és a több ezer éve ismert erőérvényesítő módszerek,[267] így teljesen új perspektívák nyílnak meg.

Alább olyan erőérvényesítési formákra koncentrálok, melyeket a számítógépes hálózat és az MIKT tesznek lehetővé. Ezáltal szeretném a digitális térben kibontakozó visszaélési lehetőségek elterjedt megközelítését tágítani és komplexebbé tenni. Ezáltal a témát a jogvédelmi nézőpontra túl állambiztonsági és katonai szempontból is kezelhetőbbé teszem. Ehhez a szakirodalom áttekintéséből és fogalmi keretből²³⁶ kiinduló dedukciót használok. A kapott konklúziók illusztrálására – aktualitása miatt – az orosz modellt választottam.

A technológia fejlődése és a hadtudomány mindig szorosan összefüggött. A digitalizáció kínálta lehetőségek azonban az korábbiaknál sokkal összetettebb paradigmaváltást hoztak az erőérvényesítésben. Ezen változás összetevői közül szempontunkból most két dolog emelkedik ki. Az egyik, hogy az országhatárok jelentősége lecsökkent a különféle gazdasági és politikai célok egymásnak feszülésében, mivel az újdonságok alkalmazása egy erősen globalizált korszakot hozott. A másik, hogy a fizikai agresszió elkerülését a nyomásgyakorlásban a digitális tér segítségével a korábbinál sokkal hatékonyabban lehet kivitelezni.

Az utóbbihoz kapcsolódik a puha művelet fogalma: ezek segítségével elérhetőek az adott célok jóval kevésbé feltűnően, nulla, vagy a hagyományoshoz képest kicsi erőszak (pl. információs, gazdasági, pszichológiai, jogi vagy egyéb módszerek) által.²³⁷ Ezeket a hadtudományban együtt szokás értelmezni a kemény (fegyveres) katonai műveletekkel, vagy azok lehetőségével. Eszerint az úgynevezett hibrid műveletekben[269, o. 18, 25] összehangoltan vetnek be a

²³⁶ Ismert fogalmak rövid összefoglalása logikailag nem maradhatott ki, de terjedelmi okok miatt elhagytam a szakirodalom alaposabb ismertetését, melynek kritikája eredetileg harmadlagos cél is volt.

²³⁷ Megemlítendő, hogy az itt tárgyaltak a „közepes művelet” kategóriába tartoznak egy pontosabb, de nemzetközileg nem közismert, három szintű felosztás szerint. Ld. [268, o. 23–24]

kemény módszerek támogatására puha műveleteket. Ezáltal akár a háború igazságossági kritériu

VI.1.2. A virtuális erőközpontok modellje

A puha módszereket azonban nem csupán egy katonai jellegű művelet részeként lehet használni, tehát ezeket katonai alkalmazásuktól különválasztva is érdemes vizsgálni. Ezt indokolja a hátszágok eltűnésének folyamata is a mai információs-, kiber- és gazdasági térben. Hiszen ezekben a terekben a civilek védelmi jelentősége is jócskán felértékelődik. Mindenki, aki ezekben a terekben dolgozik a védelem humánerejévé válik. Ehhez járul továbbá az MI, amely az intelligencia mellett ma már az érzelmeket is jól képes azonosítani és utánozni.[270] Ezért a segítségével végrehajtott befolyásolás által bizonyos célok hosszú távon eredményesebben elérhetőek lehetnek, mint katonai műveletekkel. Az így fejlődő digitális a tér segítségével fokozatosan egyre könnyebb befolyásolni, saját érdekek felé terelni a saját állam vagy más államok polgárait, vállalatait, hivatalait. Az erőérvényesítés ezen új korszakában a hagyományos katonai módszerekkel csupán a védelem egy szűkebb, fizikai szegmensét lehet biztosítani. Ez persze továbbra is alapvetően szükséges, ám már messze nem elégséges feltétele az állampolgárok megvédésének. Ezért fontos tájékozódni az erre alkalmas módszerekről, és tájékoztatni róla minden érintettet, azaz mindenkit.

A paradigmaváltás szempontunkból legfontosabb aspektusa, hogy virtuális erőközpontokban szükséges gondolkodni. Virtuális erőközpont lehet egy valódi, területtel rendelkező állam digitális lenyomata, de lehet területtől független tényező is. Lehet egy óriásvállalat digitális ereje, mely összemérhető az állammal, vagy egy független csoport, egy kis cég, mely aszimmetrikus puha műveletekkel²³⁸ gyakorol nyomást. Úgy vélem, hogy a hatalmi struktúra ilyen való megváltozása még nem került át a köztudatba adekvát módon, pedig logikusan következik a hatalom természetéből. A digitális korszak lehetőségeinek hatalmi potenciálja ugyanis épp olyan csábító, mint a történelem sok fájó emlékét okozó hatalmi mámor. A hatalomélmény örvénye a közösségi média egyszerű véleményvezéreit épp úgy húzza magába, mint a gazdasági vagy információs hatalom birtokosait – csak ez utóbbiak a nagy hatalmukat akarják még tovább növelni. Kérdéses, hogy a virtuális tér száguldó fejlődésében képes lesz-e ezt a helyzetet az emberiség megfelelően kezelni, és etikai szinten kompromisszumokra jutni?

²³⁸ Például egy néhány fős csoport is végezhet információs műveleteket, egy apró cég digitális ereje is lehet súlyához képest óriási, egyetlen hacker is jelenthet nagy veszélyt a kibertérben.

Hiszen egyelőre az erőérvényesítés etikai határait még egyszerű esetekben, egyazon kultúrán belül sem vagyunk képesek egységesen kezelni, mint például a Covid-19 járvány intézkedései kapcsán. Amikor egyes cégek nem voltak hajlandók az egészségügyi szervezetek számára átadni a járvány jobb kezeléséhez szükséges adatokat. Erre reakcióként, volt, aki ezeket vállalatokat vádolta a digitális erejükkel való visszaéléssel: „maguknak (ti. a vállalatoknak) van egy magánkormány, amely döntéseket hoz a társadalma felett, ahelyett, hogy a demokratikus kormányok meghozhatnák ezeket a döntéseket.”²³⁹ Más szerző viszont pont az állam (a regnáló kormány) túlkapaszkodásaként, és a polgárok jogának megsértéseként értékeli az állam igényét az egészségügyi adatokra.[272] A konszenzus ilyen hiánya arra utal, hogy a nyugati felfogás még nem képes megfelelően reagálni egyszerűbb információs szélsőhelyzetekre sem, a tekintélyelvű kezelés sajnos hatékonyabb.

VI.1.3. Az MI *direkt* és az *adaptált* védelmi felhasználása

Mielőtt továbblépnék a gondolatmenetben, érdemes egy kis fogalmi kitérőt tenni. Meg kell ugyanis különböztetni a védelmi MI, illetve az informatika szerteágazó területén két markánsan eltérő részt. Itt inkább katonai példákon keresztül, tehát egy hadtudományi megkülönböztetésként mutatom be ezeket: szétválasztom a *direkt* és az *adaptált katonai informatikát*, illetve a *direkt* és az *adaptált katonai MI* technológiákat. Kiterjeszhető a felosztás több más védelmi területre, ahol sajátos technológiákat használnak (katasztrófavédelem, vízügy), az alábbiak alapján.

Napjainkig a katonai informatikát az jellemezte, hogy a civil életben terjedő technológiákat katonai célokra adaptálta. Igazából nem is volt szükség katonai processzorra vagy memóriára, katonai operációs rendszerre. Elegendő volt a meglévő alkatrészeket egy olyan házba építeni, mely védi azokat a terep és harci körülmények viszontagságaitól, és a szoftvereket bevizsgálja, illetve magasabb biztonsági elvárásoknak megfelelően állítani be. A katonai portálok sem tekinthetőek igazán egyedinek, csak annyira, mint ahogyan minden vállalati portált az adott igényekre alakítottak ki. Mögöttük ugyanolyan a webszerver + portál-nyelv + adatbázis szoftverarchitektúra működik, sőt gyakran ugyanolyan IIS + PHP + SQL alapú rendszeren futnak a civil és a katonai szolgáltatások.

Az *adaptált katonai informatika* alatt azon alkalmazások körét értem, amely civil célra is alkalmazható. Példát egy *adaptált katonai MI*-re hozok: egy képfelismerésen alapuló

²³⁹ (saját fordítás). Ez a megfogalmazás vállalatokat úgy azonosítja, mint amit mi a 2. fejezetben virtuális erőközpontokként említettünk. Ld. [271]

egészségügyi diagnosztikai eszköz, melyet a civilek is hatékonyan használhatnak, csak a harctéri verziót pontosabban készítik fel tipikusan katonai (lőtt, szúrt, repesz, robbanás stb.) sérülésekre. Megjegyzendő, hogy ide kapcsolódik a COTS eszközök, vagyis a „a boltban a polcra levehető” (*Commercial off the Shelf*) termékek köre, mely konkrét árucikkekre értendő. Egy 25 éve létező szabvány alapján sorolhatóak ide egyes termékek[273], tehát a gyártók akár akkreditáltathatják is termékeiket ilyen módon, és lehet olyan saját laptopunk, mely katonai célra is alkalmas.

Az MI kapcsán azonban egy nagyon fontos újdonság, hogy általa ezen a téren is megjelenhetnek sajátos, kifejezetten katonai célú rendszerek is. Persze, ahogyan egy nehézharcokocsiban is számos alkatrész és megoldás a teherautókéval azonos, úgy ezeknek is vannak civil termékekből örökölt alapjai. De ahogyan a harcokocsira, úgy egy harci MI-re is érvényes, hogy egy ilyen eszköz azért katonai, mert civil célra nem jól használható. Például nem használható fel a civil életben egy nanotechnológiás páncéllal védett támadó drónra sem. Ezeket nevezem *direkt katonai MI-nek*.

Nem gondolom, hogy éles határt kell vonni az adaptált és a direkt rendszerek közé. Inkább úgy kell megközelíteni, hogy nagyobb számú katonai célú komponens beépítése miatt lépnek szintet. Így lesz a COTS eszközből adaptált termék, és további átalakítások tehetik azt direkt katonai célú eszközzé. Tehát, ha az említett katonai laptopot valaki a gyári operációs rendszerrel stand-alone gépként parancsok és utasítások, vagy táblázatok készítésére használja, az még a COTS szint. Ha azonban az üzemeltetők a katonailag előírt operációsrendszer-beállításokkal telepítik ezt a gépet, vezetékes csatlakozással rákötjük a katonai hálózatra, ahol katonai adatbázisokat is elér, az már adaptált eszköz. Amikor pedig katonai célszoftverekkel is ellátják, és a terepen saját katonai plug-in eszközzel katonai műholdon keresztül csatlakozik egy nemzetközi C4 rendszerhez,²⁴⁰ azt is direkt katonai rendszernek tekinthetjük. Ez a példa arra is rámutat, hogy a nem MI rendszerek is eljutottak arra a szintre, amikor sajátosan katonai informatikáról beszélhetünk, elsősorban a direkt katonai informatikai eszközök miatt.

Azonban úgy vélem, hogy az MI ezt nagyságrendekkel „direktebbé” teszi. Egy direkt katonai MI-t ugyanis sajátos, harci válaszokra szükséges betanítani, amellet, hogy a hagyományos kibbiztonsági és páncél védelmekkel is el kell látni. Itt igazából nem fokozatokon keresztül éri el a „direkt” szintet az eszköz, hanem erre jön létre – ahogyan a harcokocsi. Ez egyébként a katonai informatika területére is jelentős visszahatással van (ld. VI.3.4.)

²⁴⁰ Egy vezetésirányítási kommunikációs számítógéprendszerhez (C4: Command, Control, Communication, and Computers)

VI.1.4. A helyzet változásához jól illenek az MI képességei

Folytatom a gondolatsort a virtuális erőközpontok modelljétől azzal, hogy pontosítom az MI és az új paradigmák kapcsolatát. Elindult egy folyamat, melyen látható, hogy a klasszikus hadműveletek céljait²⁴¹ már jócskán kitágította. Nem csupán a hadszínterek számának növekedéséről van szó, bár igen lényeges változás, hogy az űr és a kibertér által megszűnik a hátország fogalma. Ezek az új hadszínterek háborús használatuk során alkalmasak fizikai rombolás és emberi halál okozására. Még a megfoghatatlan, „kiber”-térben zajló műveletekkel is elérhető ilyen eredmény a biztonsági rendszerek kiiktatásával, mely képes például ipari, közlekedési katasztrófát okozni, egészségügyi működést akadályozni, vagy saját fegyvereit a megtámadott állam ellen fordíthatja.²⁴² Ezek az új hadszínterek azonban még nem implicálnák azt, hogy a hadművelet meghatározását a fizikai agresszió fogalmain túl keressük.

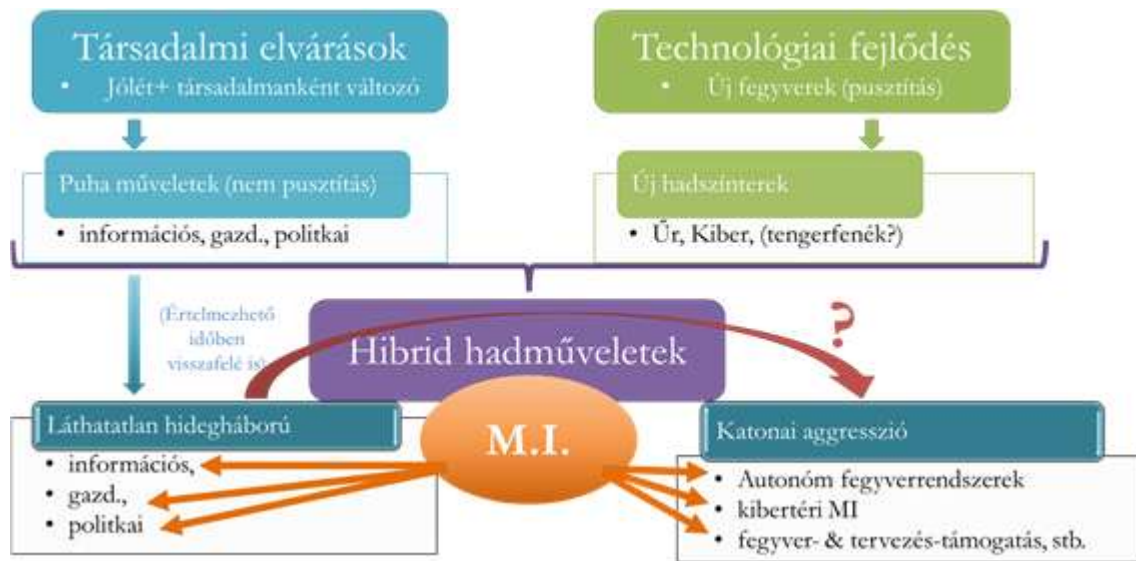
A változás sokkal lényegesebb összetevője a magasabb szintű nézőpont, melyben a katonain túl, az erőérvényesítés egyéb formáit elemzik. Az ilyen igény mögött pedig részben az áll, hogy a mai vezetés felé sokkal erősebb a visszacsatolás, sok az alulról jövő elvárás. Ma „a nép” nem attól tekinti erősnek királyát, ha az harcban legyőzi a másik uralkodót, ezért érdeke minden államnak és azok vezetőinek olyan módszereket keresni, melyekkel céljait emberélet elvétele vagy látványos pusztítás nélkül, láthatatlanul érheti el. Az emberek között is kezd háttérbe szorulni az erőszak legitimitása, például a becsület megvédése fizikai agresszióval. A nem fizikai agresszió is már a szabályozás részévé vált. Ezzel együtt nemzetközi elvárásként is megjelenik az agressziómentesség, és az államok között a nem fizikai, vagyis „puha hadviselés” formái kerültek előtérbe. Ez kihat az „igazságos háború” koncepcióra is [276], hiszen az emberek felé történő kommunikáció részeként a nyomásgyakorlás indoklásának értéke jócskán megnőtt

A puha módszerek szerepének növekedése miatt a hadviselés fogalmának kitágítása tehát elkerülhetetlen volt. Kezd kialakulni mára egy konszenzus,[277] amely szerint a hibrid hadviselés fogalmába sorolandóak olyan célok is, melyet a támadó fegyveres konfliktus nélkül (vagy azon kívül) akar elérni. A nyomásgyakorlás és ennek ellenmódszere egyaránt lehet láthatatlan. Nem csupán a közismert kibertámadás és kibervédelem rejtett módjára kell gondolni, hanem akár a kognitív befolyásolás[280] kifinomult módszereire az információs műveletekben, vagy gazdasági nyomásgyakorlásra, de a diplomáciai kényszerítés eszköztára is ide tartozik. [278, o. 22–28] Ezekhez a műveletekhez az MI olyan jelentős segítséget képes nyújtani, amely talán a rettegett harcéri használati módjain is túlmutat (ld. 17. ábra), hiszen, amint az már 1995-ben

²⁴¹ Hagyományosan: térben körülhatárolható és tüzerővel kezelhető célokban gondolkodtak. ld. [274, o. 372]

²⁴² További kibertéri lehetőségek [275]

előre látható volt, a mesterséges (virtuális-, kiber-) térben a mesterséges elme sokkal otthonosabban mozog, mint az ember.[279]



17. ábra: A láthatatlan hidegháború az MI vonatkozásában (saját készítés)

Mindez azonban még csupán egy elméleti keret. Mert hiába szélesedett a hadviselés fogalma, ez önmagában nem garantálja egy-egy adott szituáció pontosabb megítélését, hiszen láthatatlanságuk miatt ezek a módszerek csak konkrét szituációkban ragadhatók meg. Például épp ezek a modern információformáló eszközök biztosítanak lehetőséget arra, hogy akár a megtámadottat állítsák be agresszorként, például sok ilyen témájú cikk és hír generálásával. Vagy másik gazdasági példa: a ráhatások egyik formájában az ellenfelet az ellátási lánc nem feltűnő szintjein, apró alkatrészekén keresztül teszi sebezhetővé²⁴³ vagy bojkottálja. Az MI otthonosága a kibertérben és alkalmazhatósága kognitív feladatokra akár egy COTS rendszert is védelmi képességekkel láthat el, például gazdasági manipulációra civil rendszerek is készülhetnek, hiszen a cégek egymást is támadják (ld. alább), ennek mintázatait az állami gazdaságvédelmi szektor is hasznosíthatja. De egy direkt katonai MI kidolgozhatja akár egy hibrid művelet kibertéri támogatását is.

Épp a zajló orosz-ukrán háborúval összefüggésben vészjósló, hogy az orosz felsővezetés az információs műveletek továbbfejlesztését, és a MI-vel összefüggő használati módjait a hadtudomány legfontosabb feladatának tartja. Sőt, az információs hadviselést orosz nyilatkozatok alapján képessé kell tenni a pusztításra is – bár ezt annak fényében kell értelmezni, hogy az orosz gondolkodásban a kiberhadviselés része az információs műveleteknek. [281] Az Internet

²⁴³ Például amikor kiderül, hogy az F-35ös repülőgépekhez egy kínai tulajdonú cég gyárt kulcsfontosságú áramköröket. ld. [280]

Research Agency emberi alkalmazottai még „csupán” az ellenfél belpolitikájába próbáltak beleszólni,[282] ám egy gépi elme sokkal átfogóbb károkozásra lehet képes.

A második világháború utáni korszakot máig főleg az elrettentő erő „hideg” növelésével, vagyis atomfegyverek óriási számával jellemzik. Pedig, ha a hibrid hadviselés fogalmát a múltra is kiterjesztjük, láthatóvá válik, hogyan volt jelen ez a hatalomérvényesítő tényező az első hidegháború alatt, sőt, akár az ókortól is, de ezt itt nem vizsgálom.

VI.1.5. A mai hidegháború jellemzése

Először megmutatom a mai helyzet eltéréseit a II. világháború utánihoz képest, majd egy példával erősítem meg, hogy az eltérés igen markáns.

(1.) A mai és a klasszikus hidegháború összevetése

A mostani helyzet és a XX. századi atomfenyegetés eltéréseit vizsgálva induljunk el onnan, hogy a Kommunista Blokk és a Szovjetunió szétesésével kiesett egy időre a szovjet-amerikai bipolaritás egyik centruma. Bár Oroszország álma, hogy régi dominanciáját visszaszerzi; eközben egyre általánosabb vélemény, hogy az USA ellenpólusa leginkább Kína lehet[283] – a jövőben azonban célszerűbbnek tűnik inkább sok pólusban gondolkodni. Atomfegyvere sem csupán a két nagyhatalomnak volt, de a szövetségesek kevésbé rivalizáltak egymással. Ma viszont, amikor már nem kizárólag a hagyományos katonai potenciál bír jelentőséggel, egy szövetségen belül is lehetnek láthatatlan műveletek, hogy egyes vitás kérdések az erőkifejtő szájíze szerint alakuljanak. A jövőben egyre nagyobb súlyt kap pl. India, Brazília és az arab országok, de Dél-Amerika, és Afrika is felértékelődik.

Ugyanis amellet, hogy mind a nyersanyagok, mind a piac szempontjából jelentősek, a multinacionális ellátási lánc is egyre kevésbé képes ezek nélkül az optimális működésre. Mindenki próbál szövetségesi, vagy függőségi kapcsolatokat kialakítani másokkal – ebben a gyenge államoktól is függhetnek erősebbek, például a nyersanyagaik miatt. A függőségi textúra emberileg átláthatatlan és aszimmetrikus, vagyis egy kis állam kis döntése is könnyen válhat nemzetközi jelentőségű tényezővé. Annyi mindenesetre sejthető, hogy több kisebb-nagyobb hatalmi pólus párhuzamosan fejleszt MI-t, a többiek fejlesztéseitől való félelmében. S bár a direkt vérontást előidéző fejlesztésektől, vagyis a LAWS fegyverektől²⁴⁴ számos cég és kisebb állam is hivatalosan elzárkózik[284], ilyen elzárkózást nem látunk a hibrid harcmódokban, sem a hadászati tervezést vagy a műveletvégrehajtását sokféleképpen támogató[285] MI-vel kapcsolatban.

²⁴⁴ Lethal Autonomous Weapon System

Fontos különbség, hogy a civil cégek a világháború után még nem rendelkeztek olyan hatalmi potenciállal, mint ma. Gazdasági erő szempontjából már jó ideje sokuk meghaladja egy-egy ország erejét. Ilyen cég is hatjhat végre olyan spekulatív gazdasági-információs manővert, amely országokat tehet tönkre – miközben erre az állami vagy nemzetközi szabályozás még kiforratlan. Újszerű azonban, hogy az információs térben egy cég súlya és jelentősége messze meghaladhatja a gazdasági erejét. Ezáltal jól használható lehet fedőcégnak aszimmetrikus információs műveletben, akár terrorista jellegű vagy alvilági célokra is. Még nem látható annak pontos következménye sem, hogy a jelentős informatikai innovációk leginkább a nehezen szabályozható nemzetközi cégekhez köthetők.

A helyzet abban is más, hogy a folyamat nem csupán tömegpusztításban bontakozhat ki, sőt sokkal valószínűbb, hogy az erőérvényesítés egyéb, akár teljesen új módjait adja azok kezébe, akik fejlettebb MI verziókkal rendelkeznek. Az MI ismert előnye a célirányíthatóság, mely a tömegpusztító fegyver mennyiségi szemléletével szemben technikai alapot ad az ellenség kulcsrendszereinek kiiktatásához, akár személyre szabott dróntámadáson, vagy célra irányított kiberműveleten keresztül.[169] De a hátrányokozás mellett óriási előnyhöz is juthat például egy olyan gazdálkodás is, amely az MI meghatározott elemző és szimulációs képességeire építi a célratörő döntés-előkészítést és a stratégiaalkotást.

Ez az újszerű hidegháború az ellenőrizhetőség terén is eltér a korábitól. Még ha sikerülne is valami nemzetközi konszenzust kialakítani, sokkal nehezebb, szinte képtelenség lesz azt úgy ellenőrizni, mint egy atomleszerelési egyezményt. Egy MI kutatólaborot álcázni nagyon könnyű, ráadásul egy MI-ben elrejtett veszélyes tudást pusztán teszteléssel megtalálni nagyon nehéz. Bár biztató irány erre az ún. generatív intelligencia alkalmazása, melynek említése az információs műveletekre való használhatósága miatt is fontos. Hiszen alkotóképessége már ma is komoly gondokat okozhat az álhírek gyártásától a hamisított képeken át ezek felismeréséig. Már működik technológia a kockázatelemzésben való felhasználása,[286] így katonai alkalmazása kikövetkeztethető[287]: alkalmazható újszerű hadászati támadó manőverek kidolgozásától az ismert hadi stratégiák elleni újszerű, hatékonyabb védekezési stratégiák generálásáig. Az 5. sz. táblázatba foglalt összegzésben szereplő érzelmet a következő szakaszban vizsgálom.

A legjelentősebb különbség, hogy az MI-vel támogatott rendszernek bizonyos fajta autonómiaja, vagyis egyfajta „saját akarata” is lehet. Így az állami, vállalati, terrorista (stb.) akaraton túl megjelenhet egy újszerű kiszámíthatatlanság. Ez akkor is jelentős veszélyforrás, ha komoly döntéseket nem bízunk senki a gépekre. Ugyanis, ha egy felfokozott helyzetben gépi döntéstámogatás harcol másik döntéstámogató MI-vel, az emberileg követhetetlen sebesség könnyen vezethet hibás döntések sorához. Itt újfent nem csupán a katonai rendszerek problémásak,

hanem a gazdaságtámogató rendszerek vagy közvélemény-elemző és információstratégia- menedzselő rendszerek is. Ez az „akarati” eltérés túlmutat a fizikalitáson, ezzel láthatóvá teszi annak jelentőségét, hogy a védelmi szemponthoz a témát filozófiai oldalról is vizsgálni kell.

Atomháborús korszak	Jelenlegi hidegháború
két nagy pólus körül szerveződik	sokpólusú
tömegpusztító elrettentés	kiemelkedő a célra-tartás (MI)
titkos műveletek	láthatatlan, „puha” műveletek (pl. információs, kibertéri, gazdasági)
nehezen titkolható fegyverek	könnyű elrejtés, lehetetlen felderítés
állami akarat dominál (illetve államszövetségi akarat)	az állami mellett a vállalati és alvilági akarat jelentősége növekszik
az érzelmi manipuláció felértékelődik	az érzelmi aspektus gépiesítése a kognitív hadszíntéren
emberi akarat dönt	a gépi autonómia saját döntései kiterjedőben

5. táblázat: A XX. és a XXI. százai hidegháborúk főbb különbségei (saját készítés)

(2.) *A gépi érzelmek néhány védelmi alkalmazási lehetősége*

Az affektív számítástechnika példájával jól demonstrálható volt az intelligenciafajták sokféleségének gépiesítése, és érdemes a védelmi technológiák változására erről a területről hozni példákat. A korszakváltás még jobban érzékelhető abból, ahogyan a gépek korábbi intellektuális automatizmusa kiegészül az érzelmek gépi felismerésével és utánzásával. Itt is érvényesek a fentebb vázolt COTS-adaptált-direkt katonai szintek az alkalmazásban, kezdjük tehát az áttekintést a legjobban kutatható résszel, a vállalati szegmens felől.

A jelenleg zajló GÉ-fejlesztések alapján a technológia terjedése céges környezetben²⁴⁵ előbb várható, mint a katonában. Erre az ügyfélkezelés érzelmet kifejező chatbotjaitól[289] a szórakoztatóipari és kényelmi felhasználási módokig számtalan példa hozható. Ezeket a civil alkalmazásokat is megvásárolhatják, és COTS-ként alkalmazhatják a fegyveres erők, például a számítógépes játékok érzelmi képességei jól szolgálhatják a komfortérzet növelését a pihenőidő

²⁴⁵ Egy rövid, tömör felsorolást ad pl. [288]

eltöltése során. Ez nyilván nem nevezhető sajátosan katonai vagy védelmi alkalmazásnak. Ám mielőtt megnézzük, milyen sajátos alkalmazási lehetőségek kínálóznak a jövőben a GÉ képességeinek kiaknázására a fegyveres erők és az állami hivatalok számára, vessünk egy pillantást a katonai GÉ múltjára.

Ahogy az MI katonai alkalmazása a technológia megjelenése után nem sokkal kutatások tárgya lett, úgy a GÉ katonai felhasználása is már akkor felmerült, amikor még kevésbé volt ismert a technológia. 2004-ben a DARPA már „nem invazív érzékelő rendszer (...) kifejlesztésére szólít fel, amely alkalmas katonai / operatív környezetben vagy olyan környezetben történő telepítésre, amelyben az ellenség lehetséges fenyegetéseinek diszkrét megfigyelése kívánatos”.²⁴⁶ Ez az érdeklődés napjainkig kíséri a fejlesztéseket, megjelenik a NATO 2020-as fejlesztési tervében is,[6, o. 52, 58] de még ekkor is csak az érzékelő kerül fókuszba, érdekes módon az érzékelő alkalmazhatóságának vizsgálatát nem említi a DARPA 2020-as költségvetése sem.[290]

Konkrét felhasználási módok az idézett forrásokban nem találhatóak, ezért állítottam össze az alábbi listát néhány példával. Ezek sorrendje a kevésbé lényegesek felől a fontosabbak irányába halad. A lista célja nem a felhasználási paletta teljességének bemutatása, csupán a lehetőségek sokszínűségének felvillantása (ezeket az olvasó tovább gondolhatja).

3. **A kiképzés területén** főleg a már meglévő MI-alapú, esetleg VR technológiával megoldott szimulációknak adhat szárnyakat. Ha a szimulátor érzékeli a résztvevők érzelmeit, az egyrészt felhívhatja a kiképző figyelmét olyan lappangó, enyhe fóbiákra, vagy problémakezelési zavarokra, melyek éles helyzetben nagy kárt okozva, váratlanul jelennének meg. Másrészt, egy output GÉ-vel is ellátott program képes lehet figyelembe venni a rossz beidegződéseket és félelmeket, így azoknak megfelelően tudja folytatni a szimulációt (akár úgy, hogy ezeket elkerüli, akár úgy, hogy kifejezetten hasonló problémákat és félelmeket generál). Harmadrészt, ha a szimulált arcok, emberi hangok, elhangzó megfogalmazások érzelmeiket közvetítenének, úgy bizonyos pszichológiai műveletekre is hatékonyabb (átértett) felkészülést tesz lehetővé. Végül megemlíthető, hogy akár a napjainkban felértékelődött távoktatási feladatoknál is hasznos, ha az oktató automatikus visszajelzést kaphat a GÉ-szenzor által arról, hogy a hallgatók mennyire követik az anyagot, azt mennyire tartják azt megfelelőnek.

²⁴⁶ DARPA (Defense Advanced Research Projects Agency – Fejlett Védelmi Kutatási Projektek Ügynöksége) SB032-038, 2004. Számos publikációban idézik ezt a helyet, de az azokban megadott forrás már nem él, a dokumentum nem fellelhető. Az idézet forrása: [52, o. 96]

1. **Az egészségügyi gyakorlatban** számos területen fontosak a tünetek kifejeződése (affektu-
sai): az arckifejezések, az izomfeszültség, a testtartás, a kéz- és vállmozdulatok, a beszéd-
minták, a szívverés, légzés, a pupilla kitágulása, testhőmérséklet.[291] Bevetés támogatásá-
hoz egy szemüvegébe integrált szenzor és például egy EPU-stick segítheti a diagnózist or-
vos nélküli helyzetben is, de ilyen helyzetekben egy testszenzoros rendszer elemző modul-
jaként is hatékonyan alkalmazható GÉ. Emellett kifejezetten katonai lelki betegségek
(poszttraumás stressz, szerzett fóbiák) kezelésére jó ideje alkalmazzák a hagyományos MI-
t – ami GÉ-vel bővítve, pontosan érzékelni képes a páciens érzelmi reakcióját, vagy készülő
pánikrohamát kezelés során. A kezelés elsősorban a virtuális világok előállítására épül – itt
a rendszer reakcióját is meghatározhatja a GÉ, az előző pontban leírt szimulációhoz hason-
lóan.
2. **Nemzetbiztonsági használatában** a csoportos hangulatmérés-elemzés során alkalmazható
a GÉ. Hasznos lehet mind a saját állomány hangulatának (kimerültségének vagy lelkeség-
ének) elemzésére, mind pedig a civil lakosság reflexió-mérésére. Ez utóbbira egy megvaló-
sult példa az Arab Emírségek egy érdekes fejlesztése. Egy érzés-elemző térfigyelő rendszer
telepítésétől és használatától várják állampolgáraik hangulatának jobb megértését, melyek
az állam döntéseinél figyelembe vehetőek.[292] (A jogi aggályok kezelése itt kérdéses)
3. **Rendészeti területen** kihallgatások során egészen bizonyos, hogy tíz éven belül (beszerzési
ártól függően) felválthatja a hagyományos hazugságvizsgáló berendezéseket a GÉ-alapú ki-
kérdezés-támogatás. Akár a vizuális érzelemelemzési módszerek, vagy például az említett
aRadar évről évre árnyaltabb képet ad a kihallgatott személyről (szorong, dühös, titkol va-
lamin stb.). További előnye, hogy a vizsgálat rejtve is elvégezhető – miután leküzdötték
ennek nyilvánvaló jogi problémáit.
4. **Titkosszolgálati vagy terrorelhárítási célú használatnál** az automatikus célpontazonosí-
tás eddiginél hatékonyabb megoldásait valósíthatja meg a GÉ. Például fokozottan biztosí-
tandó helyszínen (fontos állami hivatal, bank), vagy kiemelt rendezvényeken segíthet ki-
szűrni a gyanús viselkedést a már említett videójel- (arc-, testbeszéd-) elemzés és egyéb
szenzoradatok segítségével. Mivel a rendszer sok érzékelő adatait egyszerre, összefüggést
keresve tudja analizálni, képes lehet olyan csoportos összehangolt tevékenységre következ-
tetni, amit ember nem vehet észre.²⁴⁷

²⁴⁷ Ezzel kapcsolatban jogi problémákat a megfigyelhetőség jogszerűségén túl a rendszer tévedésének esetei is felvetnek.

5. **Döntéstámogató és döntéshozó rendszerekben** a GÉ alkalmazása elkerülhetetlen lesz. Először is azért, mert a fentiek alapján ezek adhatnak csupán némi erkölcs-utánzási képességet a rendszereknek (IV.2.).
6. **Információs és kiberműveletek** területét együtt említem – és azért utoljára, mivel ez talán a GÉ legfontosabb, ugyanakkor legijesztőbb felhasználási módja. A kognitív térben az érzelmi befolyásolás gépiesítése óriási hatalmat adhat egy támadónak az ellenség emberei fölött. Ezt a szintet megközelíthetik a kibertér felől: például olyan sérülékenységvizsgálatnál, ahol az emberi gyengeséget szeretnék támadásra kihasználni vagy a védelemnél figyelembe venni. Az érzelemanalízist az említett emberi gyengeségek szimulációi során tudja egy rendszer jól alkalmazni, így az emberi tényezőből eredő kibertéri kockázatokat képes veszélyességi sorrendbe rendezni. Kibervédelmi oldalról ez azt jelenti, hogy az ilyen kockázatelemzés alapján sokkal árnyaltabban szabhatók meg az alkalmazandó ellenlépések. A kibertámadások oldaláról pedig hozzásegít a célrendszer legérzékenyebb biztonsági réseinek megtalálásához.[293] Ugyanakkor akár közvetlenül a személyt elemezve, vizuális érzékeléssel lehet megkeresni egy kulcspozícióban lévő ember érzelmi sebezhetőségeit. Tömegek ellen pedig a befolyásolás megtervezéséhez és maximális hatékonyságúvá fokozásához adhat alapot, amelyhez akár a hagyományos médiumokat is bevonhatja a rendszer (pl. újság, szórólap, pletyka-keltés). Védelmi szempontból az ilyen manipulációk felismerésére jó.

VI.2. ÉLNI ÉS VISSZAÉLNI A DIGITÁLIS TÉRBEN

Továbbgondolva az előző alfejezetet, itt a vizsgálatot a digitális visszaélésekre szűkítem. Célom, hogy az elemzések alapján egy olyan, az eddigiéknél komplexebb rendszert tudjak vázolni, mely a polgárok, az állam és a társadalom védelmi szempontjait egyszerre veszi figyelembe. Ehhez röviden bemutatom a nyomásgyakorlás paradigmaváltását, így katonai és nemkatonai puha műveletekként egyaránt tekinthetünk ezekre a digitális visszaállásokra. Példaként az orosz állam digitális védelmi modelljét mutatom be, annak túlzásaival és indokaival együtt, mely önmagában is hasznos és aktuális információkkal szolgálhat.

A virtuális erőközpontok modelljére alapozva tehetjük komplexebbé a digitális visszaélések elterjedt megközelítését. A téma több oldalról megközelíthető: egyrészt a *jogérvényesítés* felől (I.), másrészt a *jogsérelmek* felől (II.). Ez első terén inkább a gazdasági szereplők és államok érdekei és igényei érdekesek, a másik szempont az emberijogi szempontok alapján vizsgálódik. Alább először vázolólok, hogy a szakirodalom fogalmai inkább az utóbbi megközelítéshez kapcsolódnak, ezután térek rá a két aspektus együttes vizsgálatára.

VI.2.1. Szuverenitás, ökoszisztéma, tekintélyelvűség a digitális térben

A jogérvényesítés (I.) oldaláról a digitális szuverenitás (*digital sovereignty*) fogalmát használják. Egyszerűen fogalmazva a digitális térben való cselekvési és döntési képességet értik alatta. Egy 2018-as németországi IT-csúcstalálkozó meghatározása szerint: „Egy állam vagy szervezet digitális szuverenitása abban áll, hogy a tárolt és feldolgozott adatai felett teljes körű ellenőrzése van, illetve önállóan dönt arról, hogy ki férhet hozzá.”²⁴⁸ Tehát ez a digitális szuverenitás sérül, amikor valamely tényező (ebben a térben) a saját érdekeit a többi szuverenitási igény fölé sorolja, és ezt a többiekkel *szemben* valósítja meg.

Az ilyen szuverenitás egy megfelelő digitális ökoszisztéma (*digital ecosystem*) kialakítása által valósítható meg. A kifejezés jelentése, hogy aki valós gazdasági tényező, annak szükséges a digitális világban való megjelenését is úgy szerveznie, hogy az egy elosztott, alkalmazkodó, önszerveződő, mérhető és fenntartható társadalmi-technikai rendszerként működjön.[295, o. 14] Az ilyen rendszerek versenyben vannak, vagyis azok a vállalatok, akik idejében felismerték, hogy digitális potenciáljukat rendszerként szervezzék meg, azok előnybe jutottak a többiekkel szemben, a piac „farkastörvényei” szerint. A digitális technológiákkal ugyanis olyan új üzleti ökoszisztéma is létrehozható, amellyel újra lehet értelmezni egy iparág működését. Akik viszont nem csatlakoznak egyetlen digitális ökoszisztémához sem, vagy nem építenek sajátot, azok könnyen elvesztik korábbi pozíciójukat, még ha vezető helyen álltak is (pl. Nokia[296]).

Másrésről nem csupán vállalatokra, hanem államokra is igaz, hogy a megfelelő színvonalú digitális ökoszisztéma szükséges, ámde nem elégséges feltétele a digitális szuverenitásnak. Vagyis a jó digitális ökoszisztéma lehetővé teszi ugyan a saját akarat érvényesítését a digitális térben, azonban nem garantálja, hogy az érvényesülni is fog. Igazából az állami rendszerek e tekintetben csupán követni tudják a piaci erőket, így evidens, hogy átveszik az ott kialakított módszereket. Sajnos az EU el kell hogy ismerje lemaradását más államokhoz képest digitális ökoszisztémájának tekintetében,[294, o. 4] és hazánk sem élenjáró ebben, habár a probléma felszámolása folyamatban van.²⁴⁹

Rátérve a jogsérülés (II.) oldalára: ezt a kérdést a digitális tekintélyelvűség (*digital authoritarianism*) kifejezéssel közelítik (elsősorban amerikai jogvédők). Van, aki kifejezetten a „tektélyelvű kormányzás fokozására vagy lehetővé tételére szolgáló technológiák”[297]-at érti alatta, de ezt inkább arra alkalmazzák, hogy egyes országok különféle antidemokratikus gyakorlatokat építenek ki digitális technológiák segítségével. Ez a megközelítés az állam

²⁴⁸ saját fordítás. [294, o. 6]

²⁴⁹ A 2022-es helyzet alapos, önkritikus és előremutató elemzése: [295, o. 32–104]

biztonsága szempontjából abban a felismerésben válik érdekessé, hogy minden kutató igen bajjónak tartja saját (nyugati) országára nézve is egy másik állam belső (önön polgárai ellen elkövetett) digitális visszaéléseit. A helyzet ugyanis erősen ironikus, hiszen azokat a nyugati világban kifejlesztett digitális technológiákat, melyeket az emberi szabadság virtuális kiterjesztéseként ünnepeltették megalkotóik, a világ más részein épp az egyéni szabadság ellen használják fel és fejlesztik tovább. Így meghaladhatják a demokráciák képességeit, amelyek épp lényegük (nyitottságuk) miatt könnyebben sebezhetőek a digitális térben.²⁵⁰ Ehhez adódik hozzá, hogy a túlságosan egymástól függő gazdasági és pénzügyi rendszerek és ellátási láncok nehezítik az ilyen országok elleni nemzetközi szankciókat vagy egyéb puha műveleteket. Ezek alapján reális veszély lehet, hogy egy államban zajló belső folyamatok akár át is formálhatják a fennálló labilis globális hatalmi egyensúlyt, mégpedig a demokráciák kárára.

VI.2.2. A digitális visszaélés irányai az erőközpont-modell alapján

A II. megközelítés szakirodalmában alapján a digitális visszaélés három iránya bontakozik ki:

4. Állami szervek használhatják belföldön társadalmi csoportok és saját polgáraik ellen;
1. Átadják (exportálják) külföldre, és ott a külföldi állam a 1. pont szerint alkalmazza. Ezáltal ráadásul erősödik a kötelék is ezek között az államok között;
2. Állami szervek vetik be külföldön a másik állam polgárai, csoportjai, vállalatai ellen, ezeken keresztül támadva a másik államot.

Ez a megközelítés azonban a „térugrás” módszertani hibáját követi el, mivel a virtuális tér hatásainak vizsgálatához a fizikai térben marad, és az állami erőközpontokból indul ki. Az egyik tanulmány túlmegy ezen egy lépéssel, és negyedikként a vállalati szféra által használt vagy dotált technológiai visszaélést is említi (pl. rivális érdekcsoportok ellen).²⁵¹ Véleményem szerint azonban ennek a nem állami aspektusnak precízebb kibontása szükséges, amit a virtuális erőközpontok (VI.1.2.) alapján tehetünk meg, amivel a „térugrási” hibát is elkerüljük. Figyelembe kell venni, hogy a „belföld” és „külföld” térbeli keretei itt már nem mindig értelmezhetőek, sőt a nemzetközi szervezetek is kihasználhatóak. Így a digitális visszaélések a következő irányokkal egészülnek ki:

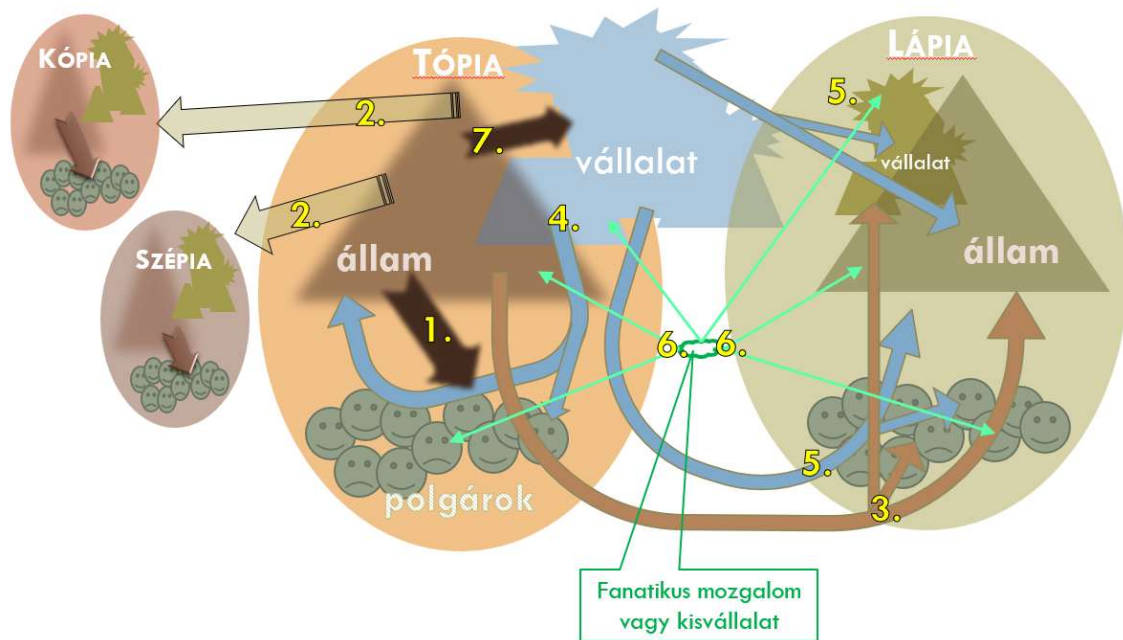
3. Erős vállalati erők „belföldön” vetik be, mozgalmakat dotálva vagy közvetlenül;
4. Erős vállalati erők „külföldön” az adott ország sajátosságainak megfelelően vetik be;
5. Független kis csoportok vetik be, általában aszimmetrikus műveletként.
6. Állami szervek saját vagy globális vállalat(ok) ellen vetik be.

²⁵⁰ Id. <https://www.power3point0.org/about/> - ez a szervezet a jelenség dokumentálásával foglalkozik.

²⁵¹ Mely szerinte demokratikus országok belügyként kezelhető. [298, o. 2]

7. (+ A nemzetközi szervezeteken keresztül gyakorolt nyomás lehetősége)

Szemléltetésül a 18. számú ábrán két elképzelt állam segítségével mutatom be az állami és a nem állami virtuális erőközpontok digitális visszaéléseinek lehetséges irányait (típusait).



18. ábra: : A digitális visszaélések főbb irányai (saját készítés)

Látható, hogy Tópia exportálja a technológiákat (2), beveti saját polgárai ellen (1), és Lápia ellen is alkalmazza (3), mind polgárain keresztül, mind közvetlenebb módokon a másik állam ellen vagy annak vállalatai ellen. A vállalati erő Lápia állami hivatalai ellen akár annak polgárain (5a.) vagy vállalatain (5b.) keresztül is követhet el digitális visszaélést, továbbá „saját” állama és a tópai polgárok ellen is (4). Jelöltem még a (fanatikus) független mozgalmak, illetve agresszív kisvállalatok aszimmetrikus műveleteit (6), valamint az állam vállalatokra gyakorolt nyomását (7). Az áttekinthetőség érdekében el kellett hagynom mindezek ellenirányait (természetesen a valós helyzet vektorai minden erőközponttól mind felé mutathatnak), és nem jelöltem a nemzetközi szervezeteket (8). Az irányokra alább az arab számokkal hivatkozom.

VI.2.3. Az orosz modell elemzése

Példaként a fenti irányokra az orosz digitális visszaélési modell néhány fontos jellemzőjét vázolom, mely a digitális tekintélyelvűség szakirodalmának egyik fő kutatási területe (a kínai modell mellett). A kínai és orosz modellek közötti egyezés, hogy egyik sem a nyugati liberális demokrácia elvei szerint használja a digitális lehetőségeket, és mindkettő exportálja²⁵² az elért eredményeket (2-es irány). Az eltérés azonban jelentős. Míg a kínai egy impozáns, de költséges

²⁵² A posztszovjet országokon kívül a világ 26 országát említi [299, o. 95]

technológia, mely egy totális kontroll irányába fejleszt, addig az orosz modell inkább azokra a területekre koncentrál, amelyek olcsóbban biztosítanak digitális kontroll- és befolyásolási lehetőségeket.[300]

(1.) *Az orosz állami mesterséges intelligencia helyzete*

A puha technológiák jövője, az MIKT lenne a legjobb az orosz modell bemutatására, ám itt csak az MI helyzetét van hely vázolni. Amikor Putyin elnök 2017-ben kijelentette, hogy az lesz a világ ura, aki az MI szférában átveszi a vezetést,[301] ugyanakkor azt is mondta, hogy jobb lenne megakadályozni, hogy valaki rátegye a kezét ennek monopóliumára. Az utóbbit elérni – a drága világelsőség helyett – a modellbe illik: ehhez elég például az UNESCO MI-etikai bizottságában vezető szerep megszerzése²⁵³ (8-as irány).

A technológiai hátrányt azonban nem lehet csupán politikai súllyal kompenzálni, ezért már igen régen próbálnak a világpiacon MI termékekkel is megjelenni.²⁵⁴ A világelsőségtől távol vannak, de számos eredményről olvashatunk.²⁵⁵ Bár folyamatosan emelkedik a szféra támogatása, még messze elmaradnak a kínai MI szinttől. Kevés a kimagasló teljesítményű MI alkalmazás (azért akad²⁵⁶), és tudományos áttörésekre úgy tűnik, nincs erejük. Ám az orosz digitális visszaélési modellhez ez is megfelel. Ezt mutatják a belföldi (1-es irány) példák Navalnij ellenzéki képviselő pártján tüntetők azonosításától a hadkötelességet elkerülők lefüleléséig.[304] A modell tehát működik, az orosz főváros és Szentpétervár kínaiénál jóval kisebb IoT kamerás lefedettsége mellett is.²⁵⁷ Sőt, felzárkózóban van olyan újabb alkalmazások által, mint a pár hónapja bemutatott Oculus²⁵⁸, mely az MI segítségével keresi meg a neten a törvénybe ütköző tartalmakat vagy titkosításokat.

A harci alkalmazásokban nem ez a helyzet. Nyugati elemzők szerint az ukrán invázióban olyan kézenfekvő területeken sem látható az MI- használat látványos eredménye, mint a puha (pl. információs) műveletek, vagy a katonai döntés-támogatás.[306] Kérdéses továbbá, hogy az MI technológiai fejlesztések jelenleg tervezett rohamos ütemét képesek lesznek-e tartani a beszerzéseket sújtó szankciók, és az agyelszívás mellett. Szkeptikusok a kutatók a tekintetben is, hogy a biztosított óriási állami forrásokat sikerül-e a háború ellenére biztosítani, illetve jól felhasználni, hiszen a nagy fejlesztések eredményességét sem az államilag támogatott vállalatok,

²⁵³ Az UNESCO Nemzeti Bizottsága alatt hozták létre a „Mesterséges Intelligencia Etika Orosz Bizottsága” tanácsadó testületet 2020-ban [302]

²⁵⁴ A digitális visszaélésre alkalmas MI technológiákat már 10 éve is exportálták nyugatra. Ld. [303]

²⁵⁵ Pl. Center for Naval Analyses hírújságjában is számos konkrétum olvasható <https://www.cna.org/our-media/newsletters/ai-and-autonomy-in-russia> (Letöltve 2023. 01. 18)

²⁵⁶ Pl. jó mimikájú humanoid <https://promo-bot.ai/robots/robo-c/> (Letöltve 2023. 03. 17)

²⁵⁷ Moszkvában 213 ezer körül van a számuk, Szentpéterváron ezer főre 12,7 felderítő eszköz jut. Ld. [305]

²⁵⁸ 2023. februárjában, ld. <https://www.interfax.ru/russia/885877>, Letöltve: 2023. 03.12.

sem pedig az akadémiai kutatások nem képesek szerintük garantálni (ezért olyan kevés a szabadalmak száma).²⁵⁹ A szkepszis ellenére azonban sikeresen alkalmazzák az MI-t, például a KUB-BLA felderítő harci drónokban.[308, o. 3] Tehát az orosz MI egyelőre nem ríkító, de nem is lebecsülendő.

(2.) Az orosz intranet

A digitális tekintélyelvűség szakirodalmában aránytalan hangsúlyt kap az internet lekapcsolására irányuló törekvés, így kénytelen vagyok erre alaposabban kitérni, mivel ezt a beállítást félrevezetőnek tartom. Tény, hogy 2021-ben lekapcsolták a Runet orosz hálózatot egy időre,[309] és már 2019-től folynak erre irányuló, hivatalosan elismert tesztek.²⁶⁰ Ez a törekvés éveken át csak erőlködés volt: például, ha valamit tiltottak, akkor más sem működött.²⁶¹ Mára azonban egyre több oldal vagy szolgáltatás tiltása sikeres, tehát egyre nehezebb megkerülni az orosz állami akaratot mind törvényileg, mind technikailag.

Ám véleményem szerint helytelen ezek alapján az ország digitális leválasztásra koncentrálni. Sokkal logikusabbnak tűnnek az olyan megfogalmazások, melyek szerint a vezetés az ország digitális szuverenitását egy orosz intranet megvalósításában látja, melyben az ország a világháló egészéhez csak a felügyelt átjárókon keresztül kapcsolódik. Szemléletesebb a digitális vasfüggöny²⁶² elnevezés, melynek üzemeltetése nem okoz gazdasági sokkot (aminek oka az ország függősége). Érdekesség, hogy a „lekapcsolás”-ból adódó gazdasági károk szakirodalma igen ellentmondásos (talán része lett az információs műveleteknek). Van 2022-es becslés, mely szerint egy internetleállítás becsült kára egy nap alatt 442 millió dollár volt,[313, o. 169] más forrás szerint összesen veszített 861 millió dollárt az orosz állam a leállítások miatt 2022 első három hónapjában. [314]²⁶³ Annyi biztos, hogy hivatalos nyilatkozatok szerint nincs meg az ország internet nélkül.²⁶⁴ Tehát valószínűsíthető, hogy kerülnek az ilyen áldozatos védelmet.

Az a tendencia tehát, hogy a *kézbentartottság* legyen a vezérelv. Ez a digitális leválasztást csak vészhelyzetekre tartogatja, akárcsak a kemény műveleteket a hibrid hadviselési szemlélet. A kézbentartottság által megszerzett adatok elemzése sajnos a polgárokról való adatgyűjtésre is alkalmas, illetve újabb és újabb tartalmak, tartalom-típusok szűréséhez (1-es irány) vezetett. Az ilyen tárgyú törvények egyre több olyan weboldal és szolgáltatás ellen tesznek lehetővé jogi

²⁵⁹ Az erős korrupciós lehetőségek miatt. Ld. [307]

²⁶⁰ Ekkor még a felhasználók (állítólag) nem észlelték. [310]

²⁶¹ Pl. a Twitter tiltásának egyik próbája az egész internetet belassította. Ld. [311]

²⁶² Más neveken „szilíciumfüggöny”, „internetes vasfüggöny” [312]

²⁶³ Megjegyzendő, hogy a NetBlocks nevű angol csoportra hivatkozik, ám a csoport weboldalán a cikkek megjelenése körül nem található ilyen kalkuláció. A túlzó számot mégis több portál átvette.

²⁶⁴ Lavrov külügyminisztert idézi a [315]

fellépést, melyek a hivatalos állásponttól eltérnek.²⁶⁵ Sorra szankciók vagy letiltás alá kerülnek (7-es irány) azok a szolgáltatók, akik nem tesznek eleget a törvénynek és nem engedélyezett hírek vagy ellenőrizetlen tartalmak elérési lehetőségét nyújtják. Egy ideig a lakosság VPN²⁶⁶-ek használatával kerülte ki a megfigyelést és korlátozást. Ám 2021-től letiltották azokat a VPN vonalakat is, melyek üzemeltetői nem voltak hajlandók összekapcsolni szolgáltatásaikat az FGIS adatbázissal.²⁶⁷

A visszaélések mellett szükséges rámutatni, hogy mindebben egy digitális szuverenitási igény is összecsap különféle érdekek térségére kényszerítésével. Ebben a felek saját digitális erejükkel igyekeznek nyomást gyakorolni az orosz lakosságra, akiknek így valódi szabadsága nemigen marad. A szuverenitási igény nyilván a külföldi szerverektől és szolgáltatásoktól való teljes függetlenséget is célozza, amint azt például a saját, orosz biztonsági tanúsítvány bevezetésére[317] vagy a szoftverek függetlenítése²⁶⁸ irányuló törekvések példái mutatják.

A külső nyomásgyakorlás hátulütője, hogy a belső visszaéléseknél kihasználható: megideo-logizálhatóak a törvények, ha „ezek ellen a külföldi ügynökök ellen” szólnak. A politikainál tanulságosabb példákat láthatunk erre a vállalati szférában, hiszen az ország ilyen nem állami szereplőkkel is évek óta harcban áll (5-6-os irány). Most pedig, az Ukrajna elleni kemény műveletekre reagálva, a vállalatok beleálltak a harci helyzetbe. Például a Meta vállalat úgy enyhítette az „erőszakos fenyegetések” szabályait, hogy az orosz megszállók halálát kívánni már nem jár büntetéssel,[312] a Microsoft pedig nem enged hozzáférést a szervereihez.[319]Az ilyen puha műveletek azonban csak a nyugati hírekben hangoznak jól, Oroszországban az állami kommunikációnak adnak muníciót. Hasonlóképp egyes cégek szankcionális kivonulása is kedvezhet az orosz vezetésnek, például az Instagram távozása után bejelentették annak orosz alternatíváját a Rosgramot,[320] melynek így önként adták át a piacot és ezzel együtt az informálás lehetőségét. Sokkal hatékonyabb az a módszer (pl. a Twitter, a Facebook és a BBC alkalmazta), mely az orosz közönség számára szolgáltatásából egy, a Tor böngészőn keresztül (inkognitóban) elérhető verziót készített el.[321]

Ezzel igen érdekes új mozzanatra tudok rávilágítani a hibrid erőérvényesítésben: a kissé alvilági, korábban kétes ügyletekre használt „dark web” válik az információs műveletek hivatalos küzdőterévé. Oroszország a maga részéről pedig legalizálta a szoftver-kalózkodást, a fent

²⁶⁵ Számptalan példát hoz különböző visszaélésekre a Freedom House civil szervezet jelentése. [316]

²⁶⁶ Virtual Private Network – ehhez csatlakozva elrejtethők azok a digitális adatok, melyekkel egy szolgáltatás használója beazonosítható.

²⁶⁷ Federal Government Information System: állami ellenőrzés alá vonja VPN identitást és adatforgalmat is.

²⁶⁸ Még az MS Windowst is cserélik. [318]

említett Microsoft-tiltás miatt. Ehhez adódhat idővel a digitális valuták legalizálódása a nemhivatalos hibrid műveletek támogatásához, és az internetszabadság etikai paradoxonokba fulladhat.

(3.) *Az orosz modell kialakulása, jellemzői és a szuverenitás esélye*

A helyzetképek után egy kis oknyomozó történelem. Az orosz modell gyökerében az áll, hogy évtizedeken át tudott terjedni az internet (nem úgy, mint Kínában). Csak 1998-ban kezdődött meg a telekommunikáció korlátozására használt technológiák számítógépes hálózatokra adaptálása.²⁶⁹ Ráadásul a cégek sokáig kijátszották, illetve bojkottálták ezeket a rendelkezéseket,[313, o. 170] így a korlátozások lényegében 2012-ig sikertelenek voltak. Ezért maradt az orosz polgárok digitális szabadsága sokáig egy élhetőbb szinten. A késlekedés fő oka, hogy a Szovjet birodalom széthullása miatt sokáig erőtlenné vált a központi szervezet, ami szükséges lett volna a technológiai fejlődés állami kontrollálásához. Ehhez járult a kontrollhoz szükséges apparátus múltban ragadt szemlélete, mely csak a fizikai kényszerítés terén mozgott otthonosan, és nem látta meg a digitális tér jelentőségét. Sőt, úgy tűnik máig nem képes a kiberteret igazán kihasználni.²⁷⁰ Ezek az emberi okok állnak a modell mögött. A késés miatt az orosz emberek megszokták és megszerették a világháló által nyújtott szabadságot és szolgáltatásokat. Így most (az utóbbi 10 évben) sokkal több társadalmi feszültséggel jár az a folyamat, amelyben a hatalom el kívánja venni az emberektől, amit használtak és szerettek, ez kedvez az ellenséges (nyugati) akaratnak.

A már idézett művek alapján négy egyéb tényezőt is szeretnék kiemelni. A *kommunikációs függőségből* (1) kifolyólag az orosz gazdaság erősen ráépült a nyílt internetre, hiszen részévé vált a multinacionális gazdaságnak. A *gazdaság erőtlensége* (2) miatt az ország az elmúlt években nemcsak az MI-hez kapcsolódó technológiákban nem volt képes még önállóvá válni, hanem saját Szilícium-völgyük vagy operációs rendszerük sincs – hiába próbál erőn felül részt venni a modern technológiák terjesztésében, és saját termékek fejlesztésében. Ez a *gazdasági függőség* (3) magával hozta, hogy a működésbiztonság érdekében érdekesebb volt teljes, kiforrott technológiákat megvenni külföldről. Ezt próbálja a Nyugat jelen konfliktusban a szankciók révén kiaknázni. De emiatt okoz gondot a nagyon modern eszközök és a jóval lassabban fejlődő orosz rendszerek *inkompatibilitása* ²⁷¹ (4) is. Érdekes lenne ezeket a tényezőket saját régióink szempontjából is megvizsgálni.

²⁶⁹ 2000-től kezdődött a korlátozások törvényi megalapozása. Ld. [300]

²⁷⁰ Mivel központi kiberparancsnokság sincs. Ld. [303, o. 30]

²⁷¹ [313, o. 169] Megjegyzendő, hogy erre a kínai beszállítókra áttérés sem megoldás, csak más illesztési problémákat vet fel.

VI.2.4. Következtetések az ukrán-országi konfliktus alapján

Az előző kisfejezet alapján megállapítható, hogy a digitális visszaélések orosz modellje működőképes. Bár a teljes digitális szuverenitás kialakításához kevés, mégis világszerte keresettek megoldásai. A példa rávilágít a kutatás azon megállapítására, hogy a puha módszerek nem csupán a katonai hibrid műveletek részeként értelmezendők. Mint rámutattam, az érintett terek civil dolgozói is a védelem humánerejévé válnak. Ezért vezettem be a virtuális erőközpontok modelljét, melyek nyomásgyakorlási vektorainak elemzésével megragadhatóvá próbáltam tenni azt a bonyolult erőérvényesítési mátrixot, melynek minden ország – így hazánk is – részese. Így lehetett vázolni a digitális visszaélések többszemponútú vizsgálatának módszerét, melyben személyiségjogi szempontok együtt kezelendők az állam és a társadalom érdekeivel. (A vállalati szféra szempontjainak kibontására nem volt mód, csupán felvettem a puha és közepes műveletek nemállami vektorait). Ezzel a megközelítéssel lehet esély arra, hogy a digitális szuverenitás elérhessen egy elégséges nívót az egyéni, csoport, vállalati, társadalmi és állami szintek mindegyikén. Ehhez nyilván minden szintnek és tényezőnek kompromisszumokkal kell hozzájárulnia. Sőt, megegyezésre kell jutni például olyan álláspontok tekintetében is, hogy hol vannak az erőérvényesítés etikai határai – nem úgy, mint a bemutatott Covid-eset kapcsán. Szükséges lenne a globális biológiai ökoszisztémához hasonlóan a globális digitális ökoszisztémát is élhetővé tenni, bár erre jelen pillanatban kevés az esély, hiszen mint láthattuk inkább a dark web emelkedik hivatalos küzdőtérre.

Végül a most folyó háborúhoz vezető újabb tényezőre is akadunk a fentieket továbbgondolva. Régóta világosak az orosz célok, melyek felé sokáig puha módszerekkel haladtak, sokunk az ilyen fajta erőérvényesítés folytatására számított. *Az orosz vezetésnek azonban szembeülnie kellett puha műveleteik csődjével.* Hiszen mint láthattuk, az orosz cégek a világpiacon eltörpülnek, és nagyon lemaradtak a közösségi média terén, vagyis csekély az információformáló vagy gazdasági nyomásgyakorló erejük. Még országon belül sem teljesen sikerült a felügyelt intranet teljes megvalósítása, így a saját lakosság véleményformálása sem érte el a kívánt mértéket (a világ közvéleménye még kevésbé). A puha műveletek ezen csődjét nem okként említjük, csupán tényezőként, mely arra figyelmeztet, hogy a kemény műveletek kora messze nem ért véget a hibrid háborúk korában sem. Főleg, ahol számú hagyományos technológia van felhalmozva, ott logikus, hogy nagyban építenek a hagyományos csapásmérő képességre. Nem csupán olyan vezetőktől logikus döntés ez, akik belül egyáltalán nem tettek vertikális lépést, vagy nem léptek elég nagyot ahhoz, hogy valóban értsék a paradigmaváltásokat – a felhalmozott, részben elavult technológia beáldozása mellett tesztelni az új fejlesztéseket logikus

stratégia. Ráadásul sajnos ez a „kicsit lépő, inkább a régire támaszkodva az újat próbálgató” taktika várható még a közeljövő konfliktusaiban.

VI.3. A TANULMÁNY EDDIGI VIZSGÁLATAI ÉS A VÉDELEM

A kutatás tervezésekor kijelölt irány szerint a vizsgálatok során általában igyekeztem a tárgyalt területek védelmi vonatkozásait is említeni, ezeket összegzem és gondolom tovább ebben az alfejezetben. Az MI-nyár védelmi továbbgondolása a többinél bővebb, de itt kerül sor a H4 hipotézisem igazolására és a terület oktatási és állami döntési megközelítésének átgondolására is. Kezdjük azonban néhány apróbb adalékkal.

VI.3.1. Az MI tél forgatókönyve védelmi aspektusból

A múlt és a jelen tényezőinek fenti áttekintése (III. fejezet) alapján váltak vizsgálhatóvá az MI-nyár forgatókönyvének védelmi vonatkozásai, melyre most térek vissza.

(1.) A korábbi telek mögötti okok védelmi kockázatai

Az alábbi tényezőket eredetileg az MI-fejlődés lassulására várható hatásuk miatt azonosítottam. Az MI lassulásának, mint jelenségnek általános védelmi következményei vannak, a védelmi problémák erősödése pedig hozzájárulhat a fejlődés lassulásához. Tehát ezekben a pontokban összeadódik a konkrét védelmi kockázat és a lassulással való összefüggés, vagyis ezek kétrétű tényezők.

- **A hype jelenség mára tudatosult.** Ezért a „spontán hype” hatás nem jelent számottevő védelmi kockázatot. A döntéshozók ma már képesek az ilyen tényezőket félretéve dönteni, hiszen mára jobban beazonosíthatóvá, vagyis kiszámítható tényezővé váltak az újdonságokkal kapcsolatos elvárások. Egyrészt a hype-elmélet alapján a különféle stratégiáknál is számolni lehet a hype-görbe szerinti visszaeséssel, másrészt elébe menve ennek, mérsékelni lehet ezt a regressziót egy jobb, pontosabb kommunikációval. A felmerülő túlzó elvárások háttere is sokkal kisebb hatású, mint régen, hiába vesznek fel akár vallásos vagy mozgalmi jelleget (ld. III.2.3.).
- **Egy hype-támadás lehetősége.** Az előbbi ponttal szemben kockázatot jelenthet a hype-elmélet *tudatos* alkalmazása információs művelet során. A fenti (III.2.3) elemzés ugyanis azt is kimutatta, hogy bár a döntéshozók szakértőik által elkerülhetik a túlzott reményeket (vagy aggályokat), a tömegek viszont fogékonyabbak lettek az ilyesmire. Az internet és a közösségi média óriási mértékben átalakította a hátteret, így egy ellenséges hype-művelet egy internetes tömeghisztéria képében realizálódhat. Mind a túlzott dicséret (hype) mind pedig

a veszélyek túldimenzionálása ijesztő hatású, a kettő egyben a leghatásosabb. Tehát akár spontán botrány egy MI-szolgáltatás körül, akár egy elnyújtott művelet jó taktika lehet, hiszen az így generált a félelem érzései az ellenfélnek kedveznek. Pl. ezzel hozzá lehet járulni ahhoz, hogy a technológiához szükséges szemléletváltás nehezebben menjen végbe a célközönség soraiban, és ezáltal maradjanak le a versenyben azok, akiket támadnak.

- **A hardver és tanítóadat-elégtelenség manapság egyre kevésbé jelent gondot**, bár mindkettő örök probléma. A mai hardverek elég erősek az új modellek tesztelésére is, illetve a Big Data és IoT révén egyre könnyebb megfelelő mennyiségű tanítóadat automatikus begyűjtése. Ezért elenyésző az alapvetően hibás modellekből következő problémák (ld. II.2)²⁷² védelmi kockázata is. A szuperszámítógépek erejével egy biztató modell megfelelően tesztelhető, már hazánkban is bérelhető ilyen teljesítmény.[322]
- **A MI rendszerek feketedoboz-jellege** kétféle szempontból is védelmi kockázatot jelent. (1) Egyrészt ez a tulajdonság egy átláthatatlanságot generál (ld. III.4.4.), amely miatt nem lehet tudni, hogy a szolgáltatásként használt MI-ket pontosan milyen adatokkal és milyen prioritásokkal tanították. Még ha jószándékot feltételezünk, akkor is elképzelhető az MI-generálta adatok használhatatlansága az eltérő nyelvek és kulturális elvárások miatt. (2) Másrészt láthatatlanok maradnak-e az ügyfél (állam vagy stratégiai cég) adatai? (Vagyis közvetlenül kerülnek-e bele ebbe a feketedobozba?) Ezzel egy klasszikus kiberbiztonsági problémába torkollik az MI szolgáltatásként való használata. Így nem csupán ahhoz kell elvárhatatlan mértékű bizalom, hogy a felhőszolgáltató rendszere védi-e a beküldött adatokat a lehallgatástól, hanem ahhoz is, hogy a szolgáltató képes-e, akarja-e az adatáramlás sértetlenségét megőrizni, és valóban küzd-e az adatmérgezés ellen.
- **A hardver és a tanítóadatok országon kívüli lokációja** a kiber hadszíntér részévé tesz minden felhőszolgáltatást, köztük az MI mögött állókat is. Ezért az előző pontban leírt bizalmatlansági tényező fokozattan jelenik meg külföldi tulajdonban és/vagy területen lévő szolgáltatással kapcsolatban. Csakhogy a szükséges teljesítmény költsége és a szakmai háttértudás hiánya igen nehezé teszi, hogy minden országnak saját MI felhői legyenek. Így kerül az országon kívülre, és ezáltal válik kiemelt kockázati tényezővé, valamint védelmi problémává a hardver és a tanítóadat. Az ilyen olcsó és rendkívül hatékony szolgáltatások kihasználhatatlansága hátrányt jelenthet a saját szolgáltatással nem rendelkező országok

²⁷² Tehát az olyan fajta hibából eredő probléma valószínűtlen, hogy a hardvergyengeség miatt gondolnak jónak egy elégtelen modellt vagy az adathiány miatt vélnék általánosíthatónak egy laborkísérletet, mint régen.

cégeinek és állami szerveinek, viszont használatuk, főleg a stratégiai vállalatok és állami adatkezelés szempontjából vállalhatatlanul kockázatos (ld. még VI.2.).

- **Helytelen antropológiából adódó elvi félreértések** következtében várható védelmi problémák terén nem várható változás, mint az előző esetekben. Az ilyen elvi és paradigmaticus alaphibák kezelése általában tüneti, a mögöttük meghúzódó emberi tényezők (ld. III.5.3.) megváltozása nem várható. Amíg ezen a téren a konkurensok nem különböznek alapvetően, addig a fejlődés ezen gátja fenn fog maradni. Védelmi kockázatot csupán valamelyik hatalom emberképének megváltozása jelenthetne (erre pedig nem mutatkozik esély).
- **A MI rendszerek elfogultsága**, és az ezek elkerülésére tett erőfeszítések is felvetnek védelmi kockázatokat. Egyrészt az elfogultság exportálhatósága miatt (ide is érvényesek a feketedoboznál említett példák). Másrészt a fentebb (IV.3) említett összefüggésben valamelyik rossz mellett dönteni kell: vagy hitelesek a tanulóadatok, de nem naprakészek, vagy pedig az MI egyből figyelembe veszi válaszaiban a látóterébe került információkat, így azonban információmérgezéses támadás érheti a rendszert. (Ennek elkerülésére a humán erőforrás lenne megoldás, amely azonban drága és lassú.) Tehát mind maga az elfogultság, mind pedig az ennek nevében végzett információválogatás is védelmi kockázati tényezőként azonosítható, és kezelése érdekében komoly kompromisszumokat kell hozni.
- **IQ-fenntarthatóság:** Megemlítendő a fentebb tárgyalt probléma védelmi vonatkozása, hogy világszinten nem születik elegendő tehetséges ember az egyre bonyolódó kultúránk kiszolgálására (III.5.1.), és az ennek folyományaként fellépő agyelszívás miatt (III.3.2.), hiszen ahonnan elvándorol a tehetség, ott nem csupán szakemberhiány, hanem tehetség-hiány is fellép – ugyanakkor a befogadó helyeken nem fognak tudni mit kezdeni a segédmunkás rétegekkel – ez pedig multinacionális társadalmi feszültségekhez fog vezetni.

(2.) *Néhány katonai vonatkozás*

- **Az adaptált katonai MI gyorsító hatása.** A (horizontális) fejlődést gyorsító hatás valószínűsíthető a közeljövőben az adaptált katonai MI-k alkalmazásai miatt. Ezeknél az ellenfél számára is kézenfekvő, hogy az adott képességet hogyan lehet átvenni és katonailag adaptálni. Azért láthatjuk az MI képfeldolgozó tudását bevetéseken, mivel ennek használata evidens, hiszen a képfelismerés bármire megtanítható. Sőt úgy vélem, hogy az ilyen felhasználáskor nyert tapasztalatok akár jobbá is tehetik a civil közlekedést is, ezért gyorsíthatja a horizontális fejlesztéseket. Például katonai drónokban alkalmazva [323] extrém körülmények között fejlődik az autonómia, ami az önvezető gépjárművekből jelenleg hiányzik.

- **A direkt katonai MI közömbössége.** Minden nagyon újszerű technológia, így a direkt katonai MI-k bevetése stratégiai kérdés, hiszen ez egyet jelentene a „kártyalapok megmutatásával”. Ráadásul, ha olyan rendszerről van szó, amelyben az MI bizonyos autonómiát ad a katonai eszköznek, akkor ennek bevetése nemzetközi tiltakozásba²⁷³ fog ütközni, amelynek majd kezelni kell diplomáciai és belpolitikai hátrányait is. Ezért ilyen katonai rendszereknél az „adu-visszatartás” taktikájára lehet számítani, tehát lokális incidenseknél csak akkor fog előkerülni ilyen katonai képesség, ha az egyik fél úgy ítéli meg, hogy már nincs veszteni valója. Ilyen rendszerek fejlesztése elimináltan és titkosan történik, ezért megjelenéséig a civil MI fejlődésére nem lesz hatással.
- **A direkt katonai MI lassító hatása válság miatt.** Az előző pontot folytatva, a direkt katonai MI olyan paradigmaváltást is igényelhet, melyhez érdemes kivárni egy olyan vezetői generációt (ld. V.1.), amelynek tagjai valóban hatékonyan tudják majd azt használni. Emellett komoly veszélyekkel járhat a felhalmozott, de számítástechnikailag elavult rendszerek csatlakoztatása pl. egy IoT háttérű és MI által optimalizált vezetéstámogató rendszerbe. Ezt az Ipar 4.0-ba integrálni próbált régebbi gyárak esetei is jó mutatták.²⁷⁴ A JADC2 nevű, MI-t is alkalmazó komplex katonai informatikai rendszer bevezetése sem követi a piac exponenciálisnak mondott gyorsulását.²⁷⁵ Ezek alapján valószínűsíthető, hogy a direkt katonai MI-k fejlesztésének lassúsága hozzájárulhat ahhoz is, hogy a civil életben se kerülhessenek piacra bizonyos innovációk, főleg egy nemzetközi katonai válság idején. Ezért lehet ez hosszabb távon visszafogó hatású.
- **Válságok gyorsító hatása az MI fejlődésére.** Régi komponens korunk háborúiban, hogy azok lehetőséget adnak a hadiipari és hadtudományi eredmények ellenőrzésére. Vagyis új fegyverek, technológiák, módszerek, elvek éles tesztelésére ezekben nyílik lehetőség. Sajnos ez az igény generáló tényezőként szerepelhet egy-egy háborúközeli válságban, olyan irányba terelve a válság kezelését, ahol az érdekelt lobbik végrehajthatják ezeket az éles-teszteket, s még az ez elleni tiltakozások kezelése is a tesztek része lehet. Ennek logikája alapján gyorsíthatja is az ilyen célú MI-fejlesztéseket egy komolyabb válság.
- **A kibertér egyszerre gyorsít és visszafog.** A kibertéri fegyvereket nem érdemes a fenti *direkt* és *adaptált* kategóriákba erőltetni, ezzel kapcsolatban inkább arra érdemes

²⁷³ 2024. november elején többéjradszerre elfogadta az ENSZ a halálos autonóm fegyverrendszerek elleni határozatot [324]

²⁷⁴ Számos a témát taglaló forrás közül egy nem teljesen tudományos, de több példával szolgáló írás jobban igazolja a mondottakat: [325]

²⁷⁵ JADC2=Joint All-Domain Command and Control. (Már előfordul CJADC2 néven is, ahol az első betű a Combined-ből ered) [326]

összpontosítani, hogy a kiberhadszintér még meglehetősen ismeretlen mindenki számára, akárcsak az MI. Azt tudjuk, hogy az MI ebben a virtuális térben „elemében van”, hiszen csupán egy már leképezett világban kell ügyködni, ezért a valós világ leképezésének problémája itt kevésbé hátráltatja. Korábbi kutatásaim során arra jutottam, hogy ebben a szegmensben az lassítja – hála Istennek – a fejlődést, hogy a fejlesztéseit senki sem publikálja, pontosabban nem tudományos alaposággal.²⁷⁶ A katonai fejlesztésekre csupán a következtetni lehet a civil szakirodalom alapján. Emellett mindenki tisztában van azzal, hogy a mivel a kibertérben az MI hatékony, ezért túl merész alkalmazásai akár visszafelé is elsülhetnek: egy MI-t használó vírus mutációja²⁷⁷ a kibocsátó ország internet szolgáltatásaira is kihathat, akár a mérgezgáz visszaáramlása az első világháborúban. Mindez a kibertéri harci-MI nagy óvatosságú használatát idézi elő, ami együtt jár azzal, hogy mindenki csak a saját, bevizsgált szakembereire utalva fejleszthet ilyesmit, ez a körülmény pedig az ilyen irányú fejlődésben a civil innovációknál lassabb ütemet vindikál. A terület visszafoghatja még a horizontális fejlődést is olyan kibertámadások által, ahol ismeretlen módon, rendszeresen mérgeznek meg vagy torzítanak el MI-fejlesztéseket. Gyorsíthatja viszont az MI terjedését néhány olyan innováció, amely által végre kevésbé képzett szakemberek által is üzemeltethető, vagyis könnyen tovább tanítható lenne egy lokálisan telepített kibervédelmi MI. Továbbá motiváló tényezőként hathat a területre az az erősödő megközelítés, amely alapján a kiberteret a többi hadszintérral integráltan szükséges kezelni.[327, o. 35.]

- **A vertikális tanulás és a parancsnoki hype.** A hierarchikus szubkultúra miatt a vertikális tanulás hiánya könnyen a hype-jelenség klasszikus típusát idézheti elő a katonai közegben. Hiszen, ha egyfajta régi szemléletben ragad egy döntéshozó (vagy döntéshozó testület), akkor az információáramban könnyen elhálnának a meg nem értett szakértői vélemények (illetve saját érzelmi megközelítésre cserélődhetnének). Egy ilyen szemléleti hiba vagy direkt módon visszafogja az MI-t, vagy a tájékozatlan bizakodás folytán kerülnek óriási pénzek arra nem érdemes modellekbe (így az első MI tél forgatókönyvét teremtik újjá). Ilyen jelenségként is értelmezhetőek a kétes sikerű orosz MI-fejlesztések is.[328] Az ukrán háború első időszakának klasszikus fegyverarzenálra koncentrááló jellege pedig magyarázható úgy is, hogy az orosz vezetés ugyan tudatában volt az MI fontosságának, sőt a puha hadműveletek jelentőségének is, de a klasszikus nyers erőben jobban bízott (és bízik).²⁷⁸ Ez is rámutat

²⁷⁶ Sok írás csupán erődemonstrációnak tűnik, és igazolhatatlan. Ld. [168]

²⁷⁷ Egy ilyen mutálódó teszt-vírust leírtam ez tanulmányomban [168]

²⁷⁸ A témát részletesebben elemeztem, és az itt említett felvetést szakirodalmakkal támasztottuk alá egy korábbi publikációban: [328]

arra, hogy a fenti elemzés eredményei az MI tél teóriáján kívül egyéb vonatkozásokban is jól alkalmazhatóak.

- **A vertikális tanulás, mint kognitív csatamező.** A szemléletváltó (vertikális) tanulás vizsgálata kapcsán fentebb (III.5.1.) a gondolati szálát ott hagytam félbe, hogy valószínűtlen az a tömeges és nagy áldozathozatal, ami pedig egy felnőttkori belső paradigmaváltáshoz szükséges lenne. Itt felmerül, hogy a pl. a kínai vagy a japán kultúra talán jobban elvárhatja az ilyen tetteket, és az elvászároknak való engedelmességhez szokott polgárok nagyobb százaléka ki is mutatja ily módon hűségét. Ez a polgár számára nem jelent akkora terhet, mint egy nyugati ember számára. Tehát e téren is előnyt jelent a társadalmi érdeket a magánélet elé helyező szocializáció. A szabadabb, de túl individuális nyugati emberek által alkotott közösség ilyen szempontból (is) hátrányba kerülhet. Ez a probléma a *kognitív hadszíntér* (avagy hadviselés [329]) részét képezi.
- **A látványos ANI sejthető prioritása.** Látványos hadi és védelmi fejlődés produkálható az ANI-alapú célrendszerek tudásának fokozásával, és a hozzájuk kapcsolt IoT rendszerek, adatbázisok és adathalmazok integrálásával. Ezért, és az előző két alfejezetben felsorolt védelmi kockázatok miatt az valószínűsíthető, hogy a fejlesztések mögött álló nagyhatalmak egy ilyen ANI bevezetését nagyobb prioritással kezelik, mint egy újszerű AGI képesség kialakítását. Ez a horizontális fejlődés vonalát erősíti, valamint enyhén hozzájárulhat az MI fejlődés lassuló forgatókönyvéhez.

(3.) Előnyök kiaknázása

A fentebb már idézett MITRE elemzés[136] kiindulópontjában az áll, hogy a felvetett lassuló forgatókönyv óriási és hátrányos hatással lenne a védelmi rendszerekre. Az a megközelítés azonban a korábbi MI-telekhez hasonló korszaktól szeretné megóvni a társadalmat – ezzel szemben jelen tanulmány egy teljesen más jellegű telet vázolt, mely ilyen fajta visszavetést nem okozna. Ez a tanulmány egyik fontos eredménye. Ezért az elemzés végén rátérek arra, hogy az itt vázolt forgatókönyvnek milyen előnyei lehetnek, azokat hogyan lehet kiaknázni vagy hátránnyá változtatni.

Az elsődleges előny bármiféle regresszió esetén, hogy „humán szinten” be lehetne hozni valamennyit a lemaradásból. Jól jöhet egy olyan régió számára, melybe hazánk is tartozik, ha a világszerte fejlesztések lassulnak. Ráadásul az nem probléma, ha a terjedés közben tovább halad. Ugyanis az említett vertikális tanulási jelenség (ld. III.5.1) problémája miatt mindenhol időbe telik ennek az új paradigmának elégséges számú vezető szemléletébe való beépülése, és a dolgozóknál a felhasználói szintű használat elterjesztése. Viszont nyílt forrású modelleken is

hatékonyan végezhetőek ilyen szakmai képzések (és önképzések). Jelenleg is jobbra ez történik, csak kissé lomha tempóban – így a lemaradás nem csökken. Ám ha a modellek változása lelassulna, és tényleg nem jönne el a várt áttörés, akkor érdemes lenne a jelenleginél nagyobb sebességbe kapcsolva megpróbálni minél többet ledolgozni a hátrányból a humán erő terén. Ezáltal létrejöhetne egy olyan apparátus, amely a szuverén MI képességeket megvalósítja, frissen tartja, valamint optimálisan és nagy hatékonysággal fogja kiaknázni. Így lehetne a lassulást előnyünkre fordítani.

Ugyanez a lassulás azonban a hátrányunkat is megtöbbszörözheti. Egy regresszió árnyékának erősítése meglehetősen demotiváló lenne az embereknek. Ráadásul sajnos a hazai mentalitáshoz közelebb áll az, hogy az élboly lelassulását látva megáll pihenni. Ezt nem szabadna hagyni. Ha nem sikerül a fent leírt módon, elegendő vezető, kolléga és szavazó állampolgár fejében a szemléletváltáshoz elengedhetetlen szintet „meglépetni” (vagyis rávenni őket, hogy lépjenek meg – hiszen helyettük sajnos más erre nem képes), akkor a regresszió után várható új paradigma hullámát sem tudjunk majd kiaknázni. Másszóval a technológiai és humán tényező egymás hatását erősíti: vagy a progresszív hatást vagy a regresszívet. Mind a felfelé vivő spirál, mind pedig a lefelé mutató örvény öngerjesztő folyamatként működik. Ez utóbbinál ugyanis sokkal kisebb lesz a technológiai-befektetői motiváció is saját MI-rendszerek élvonalba emelésére, így a humán lemaradás egy technológiai lemaradást is implikál. Minden elemző számára nyilvánvaló, hogy a fejlett és elmaradott világok közötti ollót az MI-tényező nagymértékben tovább nyitja majd, tehát nem mindegy, hogy az olló melyik élén vagyunk.

VI.3.2. Az állami védelmi döntések változása

Ebben a kisfejezetben az állami döntéshozatal szempontjából jelentős vonatkozásokat összegzem és gondolom tovább, elsősorban jelen fejezet elemzései alapján, de a korábbi vizsgálatokra is utalva. Itt is jól megragadhatóvá válik, hogy bizonyos problémákra, melyek kialakultak, nem tudunk a gépektől választ kapni, még egy fejlett MI sem tudhatja a megoldást. Ezek csak emberi képességek által kezelhetők (bár úgy sem könnyű).

(1.) *A hibrid hidegháborús háttér paradoxona*

A döntéshozó helyzete azért paradox, mert a technológiák előnyeit az államnak polgárai és gazdasága szolgálatába kell állítania – miközben meg kell védenie mindent annak ismeretlen ártalmaitól. Mindezt egy olyan közegben, ahol a technológiák gyártóinak ereje országokéval ér fel, és a sokpólusú hatalmi mátrixban az erőérvényesítés újszerű, láthatatlan módon zajlik – akár szövetségeken belül is. Ez a feszültség az MI-vel kapcsolatban kiemelten jelenik meg,

mivel az állam minden szegmensére kiható hibrid hidegháborús műveletek mögött az MI nyújtja a legváratlanabb lehetőségeket. Az újítás ereje és veszélye nem a tömegpusztításban rejlik, hanem a hihetetlenül pontos célra irányíthatóságában, de még inkább a gép saját autonómiájában.

(2.) A korlátozás paradoxonai és következményei

Az ártalmas technológiák tiltása és visszafogása is óriási kihívás elé állítja a döntéshozókat. A döntés külön paradoxona a nyugati demokráciákban, hogy a vitákat demokratikusan eldönteni nem tudományos – viszont tudományra hivatkozva a nép ellenében dönteni a diktatúrák sajátja. Ráadásul a tudomány legtöbb területén mindig többféle modell, számos „iskola” létezik, melyek akár ellentétes nézeteket is vallhatnak, így nincs egyetlen tudományos igazság. Ha mindezek ellenére próbálnák állami szinten korlátozni az MI-t, az a következő veszélyeket idézheti elő:

- **Gazdasági szinten:** a preventív lépések visszafogják az MI terjedését, ezzel visszafoghatják az ország gazdaságát versenytársakhoz képest
- **Demokratikus támadás:** az ilyen kérdések belpolitikai tényezőkké válnak, amelyeket a döntéshozó hatalmi pozíciója ellen használhatnak fel.
- **Információs műveleti támadások:** ezekben a támadó saját érdekeinek megfelelően kriminalizálja mind a tiltásokat, mind az engedélyezéseket. A korlátozást pl. a szabadságjogok megsértésének minősítik és diktatúrát emlegetnek, a prevenció elkerülésekor pedig állampolgárait semmibe vevő felelőtlen vezetésnek állíthatják be. A kutatás során néhány sajátos információs műveletre is kitértem (IV.2.3./(3.)):
 - Ilyen támadásokkal az embereket befolyásolva, tévesek lehetnek a tömegvélemények, melyekre a gép támaszkodik.
 - Virtuális influenzaszerű hozható létre és használható fel e téren.
 - Az „MI tudatosságá”-ról szóló hírek veszélyesek, mivel ezzel a propagandisztikus érveléssel ideologizálhatja meg az MI üzemeltetője „a géptől kapott” döntéseit.

(3.) Az MI terjedés dotációjának kérdése

A látható veszélyek mellett, az elmúlt századok tendenciája alapján ezen a területen számos állami támogatás várható – de nehéz kérdés, hogy mely szegmenseket kell a piacra bízni? Támogatni kell-e azt, amire egyébként is van társadalmi igény, vagy érdemes-e dotálni azt, amire nem vágnak az emberek? A társadalom oldaláról jövő hatásvektorok fenti elemzése során kiderült, hogy régen is váltak egyes technológiák lehetőségből lassan alapjoggá: a fejlettebb államok bevezették minden településre a közműveket, lefedték országaikat rádió majd televízióközvetítéssel, EU-s támogatással került minden településre nagy sáv szélességű internet.

A megújuló energiával kapcsolatos technológiák terjedése gazdasági érdek, de vajon mikor térül meg egy okosváros? Kérdés, hogy az MI egyes megjelenései esetében érdemes-e, illetve milyen területeken és mértékben érdemes azt dotálni? Evidens a bevezetése közismert állami alkalmazási területeken, de ezeken kívül kérdéses. Jelenleg pl. az autonóm járművek terjedésének támogatása, valamint az állami feladatokra is használható szenzorhálózatok és Big Data felhők létrehozása tűnik biztosan rentábilisnak.

(4.) *A várható negatív hatások felvállalásának problémája*

A diszruptív innovációkat nem lehet eleve tiltani, azonban engedélyezésükkel együtt szinte biztos, hogy negatív hatások is megjelennek a felkészületlen társadalomban. A probléma: mit szabad engedni, milyen szabályok mellett – és mikor kell központi szabályokkal a féket behúzni? Van, amit lehet kezelni körültekintően szabályozott bevezetéssel,²⁷⁹ és van, amire még nem látható preventív megoldás. Néhány ilyen gond a korábban említettek közül:

- **Az emberi IQ-fenntarthatóság:** A munkanélküliséget nem lehet átképzéssel megoldani, mivel a gyengébb képességűek számára idővel talán egyáltalán nem lesz feladat (III.5.2.).
- **A pszichológiai visszahatások:** Egyre kevésbé van szükség az emberi tudat mélységeire. Az emberi kapcsolatok radikális átalakulása várható. Krónikussá és tömegessé válik beteges technológia-függés, illetve a tovább súlyosbodó, egyre elterjedtebben kóros individualizáció (III.3.2.)
- **Az MI által generált új etika:** Az MI predikciós jósképessége létrehozhat és el is törölhet szemléleteket. A felelősséget rombolja, ha az ember hisz abban, hogy a gép a lehető legoptimálisabb döntést hozta; a tökéletesség, amit az MI mutat, degradálja az emberi hibázást, a szándék megítélését – bár ez magasabb morális elvárásokat is támaszthat. (III.3.2.)
- **Egy ugrás nélküli terjedés forgatókönyve** (III.2.) a teljesség kedvéért itt is említendő.

(5.) *A nemzetállamiság MI-paradoxona*

A nemzetállamiság oldaláról is komoly kérdésekkel szembesülünk, hiszen egyfelől az MIKT hatékonyan támogathatja egy-egy nemzetállam érdekérvényesítő képességeit, és csakis ilyen rendszerek képesek megvédeni az MI-alapú ellenséges behatásoktól, támadásoktól. Másfelől azonban az MI technológia generálhat egy ezzel ellentétes folyamatot is: a nemzetállamok eltűnését. Ugyanis az MIKT multinacionális méretekben válik igazán hatékonná, például a

²⁷⁹ Pl. tudományos alapon lehetett volna számítani arra, ahogyan a közösségi hálózatok gyors terjedése a fiatalok között mindig is jelen lévő viselkedési anomáliákat igen nagy problémává nagyította. Nem került kezelésre, hogy a gyermeki bosszú, hatalmi vágy vagy egyéb negatív motiváció, milyen veszélyes kiterjesztést kap a közösségi hálón. A társadalom önvédelme még nincs felkészülve az ilyen gyors változások megfelelő preventív kezelésére.

világméretű Big Daták igazán változatos és nagyméretű adathalmazai által. Így más technológiákkal karöltve gyakorlati síkon az MI-technológia egésze támogatja azokat az évszázados folyamatokat melyek a nemzetállamok feloldódására irányulnak. A visszahatás fentebb taglalt működése alapján tehát a fejlődés ezen iránya könnyen válhat olyan normává, mely a világáram irányába mutat. Ezzel összefüggésben erősödhetnek tovább az olyan az eszmerendszerek, melyek a nemzetállamok létét kérdőjelezzik meg, ezeket a múlt ártalmas maradványának tartják.

(6.) Nemzetközi megoldás.

A fentiek is alátámasztják azt az evidenciát, hogy egyetlen ország sem lehet már képes pusztán belpolitikai döntésként kezelni az ilyen technológiákat, és visszafogni őket veszélyeik miatt – ez kizárólag nemzetközi megállapodások szintjén lehetséges. Ezeknél azonban nem biztos, hogy az eddigi fogalmi kereteket és hangsúlyokat érdemes követni, hiszen pl. a „tökéletes előítéletmentesség”-ről kimutattam, hogy lehetetlen és értelmezhetetlen, mivel bizonyos fajta előítéletesség mindig van minden emberben és gépben, mivel így működik a megismerés. Figyelemmel kell lenni az eltérő kultúrákból várható MI-fejlesztési eltérésekre, és alkalmazásokra, pl., hogy egy robotot könnyen bajtársként kezel egy keleti ember, míg a nyugati picit fél tőle. Az „üres lap” modell szerint önmagát kialakító MI veszélyei rámutattak, hogy hogyan negligálhatja egy ilyen technológia az etikai irányelveinket, vagy akár a nemzetközi jogi Chartákat.

(7.) Az optimálisabb, de etikátlanabb MI paradoxona

Az előző pont „evidenciáját” kérdőjelezi meg ez a probléma. A kérdés, hogy mi a helyes vezetői döntés, ha egy rivális vagy ellenséges ország bizonyos etikai elveket sértő, de épp ettől hatékonyabban, optimálisabban működő MI-eket használ? Az etikátlanul szerzett előny általános probléma, melyeknek csak jogi úton, illetve nemzetközi megállapodásokkal lehet elébe menni, de az igazi gondot mindig az ilyen szabály betartatása jelenti. Nincs ez másképp e tekintetben sem, ráadásul a probléma nem elméleti, legjobb példa rá a személyiségi jogok figyelembevétele. Például Kína, ahol ez nem szempont, sokkal nagyobb Big Data segítségével taníthatja saját MIKT rendszereit. Ennek nem csupán adatai, de algoritmusai sem követik az ENSZ irányelveit, ezért az MIKT-t úgy veheti be, hogy a maga állami szempontjából hatékonyabban jár el – ám eközben az emberi alapjogokat sérti. Ez ellen a digitális tekintélyelvűség²⁸⁰ terjedelmes szakirodalma sem tud eredményes nyomásgyakorlási módot javasolni. Viszont ebből kikövetkeztethető a baljós irány, amely szerint a hagyományos, vagy hibrid katonai műveletek támogatására fejlesztendő titkos MIKT projektek sem fognak nemzetközi irányelveket

²⁸⁰ A témáról írt, bőséges szakirodalommal ellátott fejezetet terjedelmi okból ki kellett hagynom, ezért máshol tervezem megjelentetni. A témáról pl. [298]

figyelembe venni. Követelheti az EU az egyenlőséget (pl. a női fejlesztők számának növelését), ha közben a kínai fejlesztők egy sokkal hatékonyabb MIKT-t használnak egy általános információs és kibertámadásban.²⁸¹ Így a nyugati társadalom meglehetősen ellentmondásos helyzetbe kerülhet: vagy az önmaga által elvárt elvek alapján fejleszt kevésbé optimális MI-t, így veszélyezteti az országok biztonságát, vagy fel kell adnia az alapját képező elveit. Ez az ellentmondásos helyzet jelen pillanatban feloldhatatlannak tűnik, hiszen a világ területének és népességének nagyobb része tartozik a digitális tekintélyelvűséget képviselő hatalmak érdekszférájába, ahol átveszik ezen hatalmak fejlesztéseit.[300]

VI.3.3. Védelmi paradigmaváltások összevetése: az MI és a kibertér

A tanulmány eddigi szövegei pontosították azt a mindenki által látható tény, hogy az MI újszerűsége óriási. Felmerül, hogy talán nagyobb jelentőségű, mint a számítógépek és azok hálozatainak technikai forradalma. Pár száz év múlva mögöttünk lévő 80 évet talán az MI-korszakhoz való átmenetként fogják tanítani. Itt azonban nem ezt a kissé science fiction-be hajló jövőképet szeretném bizonyítani, hanem csupán az MI radikális újszerűségét és ennek hatását áll szándékomban érzékeltetni. Úgy gondoltam, hogy erre a kibertérrel való összevetés a legalkalmasabb, hiszen mindenki által elfogadott tény, hogy kibertér alapvetően változtatta meg a hadviselést (sőt, új hadszíntérré is előlépett). Ezért is irányult eleve erre az összevetésre a K4 kérdés, melyre itt igyekszem választ kapni.

Nem terveztem a két technológia objektív összevetését mérhető jellemzőik alapján, mivel úgy vélem, az ilyen próbálkozások látszat-objektivitáshoz vezetnének. Ugyanis a fizikai-matematikai megvalósulás csupán részét képezi mindkét fogalomnak, mindkettőben a valós világ leképezése, virtualizációja jelenti az igazi kihívást. Ráadásul a kibertér esetében csupán az adat és az információ közötti konverzió problémáit kell leküzdeni, míg az MI esetében elvont kifejezések leképezése felé kell lépéseket tenni (következtetés, felelősség, érzelmek stb., vö. IV.2 és IV.3). Ezért azt a módszert választottam, hogy az MI újdonságainak fentebb (V.2.3.) részletezett jellemzőit vetem össze azzal, ahogyan a kibertér volt új.

1. Az MI újdonságának lényege az emberi mivoltunkat magát érinti. Ez a kibertéri problémákra egyáltalán nem jellemző. A kibertér esetében ez az újdonság sokrétűen jelent meg, de minden szempontból az adatokkal függött össze (újszerű tárolási, lekérdezési, továbbítási és automatizálási stb. módjokkal). Az így kezelt adatok teljesen újszerű információs világot

²⁸¹ Például apró változtatásokkal képesek lehetnek módosítani a közhasználatú információs bázisokat, mondjuk az online jogtárak néhány hónapos hamisításával perek százazeit lesz szükséges megismételni, vagy gazdasági adatok intelligens hamisításával rossz gazdasági döntéseket generálni, stb.

hoztak létre, melyben újszerű bűnözés is kialakult. De egy megszerzett érzékeny adat is csupán adat, ami információvá visszaalakítva jelent csupán értéket. Ezzel szemben az MI emberutánzó képességei túllépnek az adatok szintjén, sőt le sem tudjuk kérdezni a tárolt aktuális mintázatok értelmét (ld. II.2.1.). Jó példa ide a személyiséglopás MI által támogatott módja, amely egy sokrétű, dinamikus utánzást hozhat létre automatikusan, amilyenre eddig csak tehetséges színészek voltak képesek – szemben azzal, hogy a hagyományos módon a cracker „csak” a megszerzett személyes adatokkal tud visszaélni.

- 2. Az összetettebb MI rendszerek már működési elveikben messze túlmutatnak a reáldományok objektíven mérhető keretein.** Ez, bár a kibertérben is megjelenik, sokkal kisebb mértékben jellemző. Példa lehet erre, hogy a fejlett kibertámadások (pl. APT²⁸²) bizonyos eseteiben pszichológiai sémákat is bevetnek „gyenge láncszemként” támadható ember kereséséhez. Az MI-nél ez nagyságrendekkel jobban teljesül, hiszen már az MI alapjánál is megjelenik szinte minden tudomány eredménye, amint azt a kognitív tudománnyal kapcsolatban bemutattam, sőt az MI maga is a tudományok konvergenciáját hozza létre (V.2.2.).
- 3. Az MI rendszerek biztonságosságát és az etikus működésüket szavatoló előfeltételek is alapvetően eltérnek a korábbiaktól.** Ebben a két technológia alig tér el. Bár az MI biztonsági kérdéshalmaza sokban eltér a kiberbiztonság vizsgálandó tényezőitől, de ez az eltérés nem nagyobb, mint a kibertér és az anélküli világ biztonságának különbsége. Mindkét esetben óriási biztonsági paradigmaváltás szükséges. A kibertér esetében ez nagyjából már le is zajlott, és egy elég kiforrott, jól oktatható, alkalmazható, számonkérhető és mérhető rendszerré fejlődött a kiberbiztonság tudománya. Az MI biztonsága esetében szinte újra kell kezdeni a munkát, hiszen az autonómia, elfogultság, felelősség és egyéb olyan kérdések szabályozása szükséges, amely eddig csak az embereket érintő szabályokban jelent meg. Az emberek esetében sem mindig találtak rá megoldást, és ahol igen, az a gépekre általában nem alkalmazható (pl. felelősség).
- 4. A klasszikus logika kereti elavulnak.** Ez a kibertérre egyáltalán nem érvényes. Ott a szabályozottság pontatlanságai, programhibák és egyéb logikus, programozható és objektív tényezők vezetnek sérülékenységekhez. A kibertér esetében a problémát az átláthatatlan bonyolultság, és túl sok paraméter jelenti. Ezzel szemben az MI esetében a pontos működés is fekete doboz jellegű (III.4.4.). Egy gépi autonómia, amely az ember bonyolult döntéseit próbálja utánozni, nem csupán a fuzzy logika matematikája szerint elmosódott fogalmakkal kell, hogy operáljon, de ellentmondásos helyzetek feloldására is képes kell, hogy legyen.

²⁸² = *Advanced Persistent Threats*, azaz fejlett tartós fenyegetéssel operáló támadás, melyet professzionális kiberbűnözői csoportok használnak, és ahol bizonyos esetekben klasszikus bűnözői módszereket is alkalmaznak.

5. **Humánvirtualizáció** A kibertérnél a fent bemutatott a formában szóba sem jöhet, de hasonló mégis megjelenhet.²⁸³ A humánvirtualizáció lényege, hogy a virtuális térben az MI számára létrejött módszerek, megközelítések egy része visszaalkalmazható az emberekre. A kibertér bizonyos szemléletei is visszaalkalmazhatóak: pl. a kibertéri metódusok egy része visszaadaptálható a fizikai adatok vagy dolgok védelmére. Tehát a kibertechnikán belül elért eredmények az ember megismerését ugyan nem segítik, de más módon jól hasznosíthatóak a valós világban.
6. + **A kognitív műveleti tér gépiesítése.** Ez a jellemző nem szerepelt a fenti listában, csak ebben a fejezetben került említésre (VI.1.5., VI.3.3.), de ide tartozik. A hagyományos kibertér esetében még nem merülhetett fel ilyesmi, ez a tulajdonság a kibertér és az MI közös halmazának újdonsága. Ahogyan a kibertér az információval kapcsolatos műveletek gépiesítését tette lehetővé, úgy az MI a kognitív teret lesz képes gépiesíteni (bár ez jelenleg még gyerekcipőben jár). Mindkét újszerűség igen jelentős, de együtt hatványozott az újdonság.

Az előző két kisfejezetben bemutatott adalékokkal a fenti pontok remélhetőleg beláthatóvá tették állításom igazságát. Rátérek ennek következményeire, melyek hasonlóan fontosak.

VI.3.4. A védelmi informatika és az MI védelmi oktatása átalakulóban

(1.) *Informatikából információtan*

Az MI védelmi jelentőségének növekedése, a kibertér jelentőségével karöltve ki kell, hogy hasson a védelmi informatika szerepére, sőt meghatározására is. Az előbbivel a következő szakaszok foglalkoznak, az utóbbiról itt ejtek szót. Fentebb, az informatika szó módosulását vizsgálva (V.2.1.) arra jutottam, hogy várható annak az értelmezésnek az elhalványulása, mely a számítástechnikával (vagy információstechnológiával, az IT-vel,) azonosítja a tudományt, és újra előtérbe kerülhet az a korábbi értelmezés, hogy informatika az inkább egy „információtan”. Kimutattam, hogy az MI miatt konvergáló sokféle tudományág és technológia egymásra utaltsága és komplex rendszere lehet csak képes egyre jobb leképezéseket létrehozni. Ebben a megközelítésben a világ virtualizációjának, azaz gépi leképezhetőségének vizsgálatát végzi el az, amit informatikának hívhatunk.

Mint említettem, ez a jelentéstágulás bizonyos – míg a vázolt fogalmi megoldás inkább csak javaslat. De e szerint az informatika (információtan) résztudományává válhat mind az IT mind

²⁸³ Ez e tényező az 1. ponttal függ össze, ami szerint az MI emberszerűségünket érinti. Amint az ember gépi leképezéséhez az emberről, az kognícióink, intelligencia-típusaink működéséről és együttműködéséről is egyre több és pontosabb információt szerzünk, úgy igazából magát az embert is egyre jobban megismerjük.

pedig az információtudomány. Egy MI-fejlesztő vagy tanítás-tervező szakembernek minden kapcsolódó tudományterület ismeretére alapvető szüksége lesz. Bár jelenleg az MI-vel kapcsolatban mindenhol még a technikai oldal dominál, és a kapcsolódó szaktudást más szakemberektől (pl. nyelvészektől, pszichológusoktól) várja, de a modellek fejlődésének lassulásával, a terjedési igény mellett (ld. III.) ez változni fog, a hatékonyság érdekében. Hiszen ütőképesebb csapat az olyan, ahol minden fejlesztő tisztában van a szükséges humán és reál kompetenciákkal egyaránt, csak valamelyikbe jobban elmélyül szakiránya szerint. Ennek a várható jelenségnek két vetületére térek itt ki: a katonai informatikára és a védelmi MI oktatásra.

(2.) *A katonai informatika felértékelődése*

Az adaptált katonai informatika kapcsán (VI.1.3.) már részben leírtam, hogy a számítástechnika katonai alkalmazása sokáig a kis mértékben módosított eszközökre alapult. Ebből következett, hogy a hozzárendelt szakállomány feladata egy támogató és szolgáltató feladatkörbe betonozódott be. Ebben igazán olyan éles harctámogató tevékenység nem volt, mint például a híradó szakcsapatok kapcsolatkiépítési feladata, napjainkig csupán a létrejött kapcsolatra épülve üzemeltetik vezetési pontok számítógépes rendszereit. Az ilyen rendszerek kibervédelme ugyan komoly szakmai kihívás, akárcsak kibertámadások végrehajtása, de sokak számára tűnhet úgy, hogy az ilyen harcokat a hátszorból vagy védett helyről, „fotelből” lehet elvégezni – hiszen a támadókról is ilyen kép él. Az attitűd, mely szerint az informatika távol marad a tűzvonalától, a szakállomány presztízsét szinte teljesen elvette.

Azonban a hibrid műveletek fejlődése és a kibertér hadszíntérré válása megindított egy elmentéses irányú folyamatot. Szakmailag evidens, hogy ha távolról menedzseljük vezetési számítástechnikai rendszereinket, az sérülékenységi pontokat eredményez, ez pedig eleve a helyszínre rendeli az informatikus katonákat. Az MI bevonása a műveletek különböző szegmenseibe várhatóan további elvárásokat fog támasztani olyan irányba, amely által a katonai informatika szerepe felértékelődik. Jelenleg még abba az irányba tartunk, hogy az MI képességgel rendelkező eszközök kezelésére a harcoló állományból jelölnek és képeznek ki katonákat, mivel ezek kezelése egyszerű. Így is alakítják ki ezeket az eszközöket: pl. egy robotkutya vagy egyéb drón felé elvárás, hogy felhasználóbarát és viszonylag könnyen kezelhető legyen. Úgy tűnhet, hogy további MI funkciók, például a nyelvi képességek könnyen programozhatóvá is fogják tenni a harctéri eszközöket, így továbbra sincs szükség ott informatikusra.

Ezzel szemben állítom, hogy nagy szükség lesz jól képzett és az MI rendszerekhez is értő informatikus katonákra. Távollabra tekintve ugyanis az MI-nyár tézisének alapján az MI rendszerek terjedése, offline verziók fejlődése várható, valamint komplex képességű, sokmodulos,

sokféle intelligenciát utánozni képes rendszerek fognak megjelenni. Ezek adaptált katonai MI-ként inkább béke-felhasználásokként kerülhetnek felhasználásra, pl. hazánkban a KGIR rendszer ilyen irányú automatizálása képzelhető el, melyhez az SAP cég biztosíthatja az MI-modulokat. Harci alkalmazásukhoz direkt katonai MI-eket lesz érdemes fejleszteni, a létrehozott sokmodulos alapokon. Ezek alapvető eltérését a civil rendszerekhez képest abban lehet jól megragadni, hogy a katonai rendszer etikai kereteit nem a személyiségi jogok védelme, hanem a hadijog irányelvei adnák (sok más különbség is lesz). Valószínűsíthető, hogy egy ilyen rendszer fejlesztését, a világ sok helyén nem katonák, hanem bevizsgált magánvállalkozások vagy állami cégek végzik.

Am harci helyzetben az ilyen támogató képességek kihasználását katonákra kell bízni. Ehhez kifejezetten erre is kiképzett, informatikai tisztekre, főtisztekre lesz szükség. Az ő feladatuk lenne a művelet vezetőinek szándékát a gép számára precízen értelmezhető formába konvertálni, és a kapott választ értelmezni. Ilyen rendszerre példa az JADC2 (*Joint All-Domain Command and Control*) rendszer, vagyis Közös (összekapcsolt) Összhaderőnemi Vezetés és Irányítás tervezete, mely az Egyesült Államok hadereje számára kíván MI által támogatott elemző és tanácsadó funkciókat is szolgáltatni.[326] Ilyen képesség jó kihasználása azonban nem képzelhető el a mai szokásrendszerben.

Elsősorban szükség lesz egy interakcióra, vagyis a jövőbeni parancsnokoknak engedniük kell például a visszakerdezéseket az ilyen szakemberek részéről. Hiszen azok csak így lesznek képesek a feladatot félreértések idővesztése nélkül végrehajtani. Tudomásul kell venni, hogy a szimulációs támogatás csak több verzió lefuttatásával igazán hatékony, ezt azonban lerontja, ha csak emberi félreértések miatt kell azt újra futtatni. Mégis úgy vélem, hogy egy ilyen interakciós vezetési stílus kialakítása jóval kisebb vertikális lépést igényel, mint az, hogy taktikailag képzett és kreatív vezetőket próbáljanak továbbképezni az adatbázisok, Big Data adatbányászat és MI-képességek hatékony kihasználására. Eközben a gépek egyfajta kiszámíthatóságot is adhatnak a műveleti tervezésnek, ezért szükséges lesz úgy megfogalmazni a tervezési feladatokat, hogy abban a gép generatív moduljai is kihasználásra kerüljenek. Mindezek arra mutatnak, hogy ez a feladatkör erősen katonai jellegű lesz, ugyanakkor magasan képzett informatikusokat igényel.

Ehhez járul a különböző harci robotok felprogramozásának tervezése és felügyelete, a harcéri IoT hálózatok adataira elemző szkriptek előállítása, vagy ezek MI-n keresztüli analízisének beállítása, és hasonló újgenerációs harci tevékenységek. A vázolt néhány felvetés is láthatóvá teszi a lehetőségét vagy inkább szükségességét annak, hogy a katonai informatikában egy paradigmaváltásra sor kerüljön (mely remélhetőleg presztízsváltást is magával hoz).

(3.) A védelmi MI oktatása

Az előző szakaszok gondolatmenete alapján vizsgálni szükséges minden harcoló és támogató állomány MI-vel kapcsolatos kiképzését, különösen az informatikai (vagy hasonló irányultságú) állomány MI-ismereteinek magasabb szintre emelését. Ez véleményem szerint a katonai felsőfokú képzésben valósítandó meg, a különböző szintű képzéseken, és szakok szerint eltérően. Ebben a kutatásban nem tűztem ki célul, hogy erre terveket dolgozzak ki, csupán tömören áttekintem ennek lehetséges irányelveit, melyeket mindössze vitaindítónak szánok.

Minden képzésben szükséges lenne már most az MIKT működését elmagyarázni, hangsúlyt fektetve a félreértések tisztázásának fontosságára (pl. pszeudo-MI, V.1.4.). Ez az informatika oktatók feladatköre lehet, és bizonyos mértékben most is megvalósul. Hosszabb távon ezt tudatosan bővíteni szükséges annak arányában, ahogyan a terjedő MI modellje igazolódik. Ezzel vissza tudok csatolni arra a következtetésre (VI.3.1), amely a vázolt MI-nyár kiaknázhatóságára a lakosság MI használói képzését jelölte meg: a védelmi oktatásban ezt kiemelten lenne szükséges végrehajtani.

Visszatérve a jelenre: a fiatal hallgatók saját célra ugyan előszeretettel használják például a nyelvi modelleket, azonban ezek szakszerű promptolását nem ismerik. Ennek oktatását azonban NEM az informatikai vagy infokommunikációs oktatók feladataként képezelem el, hanem mindenhol megjelenő kompetenciaként. Az NPT promptolás oktatása okot adna az oktató állomány számára is a szükséges önképzésre, melyet kutatói tevékenységükben jól használhatnának. Emellett alapot biztosítana a jövőre nézve a folyamatosan fejlődő NPT rendszerek, illetve a majdani hazai rendszerek helyes és hatékony használatára és oktatására, esetleg a honvédséget érintő fejlesztésekbe való szakértői bekapcsolódásra is. Igaz, egyelőre ezek a szolgáltatások meglehetősen sok kompromisszumot igényelnek, és energiát szükséges megismerésükbe fektetni. Ám a fenti változások arra mutatnak, hogy minden szakirányon idővel szükséges lesz beépíteni az elvárásrendszerbe, és ehhez természetesen megtanítani a specifikus trükköket minden szakon. A direkt katonai MI-k fejlesztéséhez szükség lesz a kognitív tudományokhoz a védelmi vagy katonai tudást is hozzáadni, ehhez pedig az MI és Big Data alapokkal tisztában lévő katonák kellenek majd. Minél előbb hajlunk az oktatás ilyen irányú alakítása felé, annál jobb pozícióba kerül a védelmi oktatás és az ország.

Természetesen a különböző szintű és célú képzéseken ez különböző súllyal és fókusszal jelenítendő meg, ezeket nem részletezem. Egyelőre túl távolinak tűnik a harci MI-k megjelenése, és legtöbbször csak robotokat képzelnek el ilyen feladatra. Ezért a közeljövőben nem várható igény semmilyen vezetéstámogató MI-vel kapcsolatos képzésre. Bár jelenleg is markánsan jelenik meg az MI-vel kapcsolatos ismeretanyag a katonai informatikai oktatásban, ennek

folyamatos bővítése szükséges, és az államvizsgákon való explicitebb megjelenítése felé is érdemes lesz lépéseket tenni. Sajnos a komplex képzéseken gyakorlatias, de a felhasználói szint fölötti MI-ismeretekhez (tanítás, modellezés, adatbányászat) egyelőre kevés a rendelkezésre álló idő, ezért csupán ilyen ismeretek alapozása képzelhető el a jelenlegi képzéseken belül.²⁸⁴

VI.4. ÖSSZEFOGLALÁS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI VÉDELMI ÚJSZERŰSÉGE

- **A várakozások teljesülése.** Az utolsó vizsgálat volt a legnehezebb, mivel a korábban rendszerezett ismeretanyagot és következtetéseket védelmi és katonai szempontból kellett kiegészíteni és újragondolni. A témakör vizsgálata a kijelölt célt (C4) teljesítette, mivel számos aspektusból kimutatta az MI jelentős hatását korunk katonai és védelmi szférájára. A kutatási kérdésre (K4) az elemzések alapján egyértelműen „igen” válasz adható, hiszen a két újszerűség összevetésére külön szakasz tért ki (VI.3.4./ (2.)). Ez alapján tehát állítható, hogy az MI-nek az állami katonai és védelmi elméletekre és rendszerekre legalább akkora hatása lesz, mint a kibertérnek volt, illetve a kibertérrel összefonódva hatványozottan erőteljes változások várhatóak. Az első két alfejezet pedig igazolta a H4 hipotézis első felét, miszerint az MI tovább tolja az erőérvényesítés hangsúlyát a puha műveletek felé. A hipotézis második fele alapján ennek kezeléséhez pedig a katonai és védelmi informatika radikálisan új megközelítése lesz szükséges, amit a VI.3.4. vizsgálatai bizonyítottak.
- **Összefoglalás.** A kutatás arra mutatott rá, hogy az MI újszerűsége hozzáadódik a világ egyéb változásaihoz, és felerősíti azokat. Ezek együttes eredményeként a virtuális terekben gyakorlatilag folytonos a harc valamilyen foka, melynek résztvevői nem csupán államok, mint régen. Ezekben az MI támadóként és megtámadottként egyaránt jelen van, számos módon. Mindezek megragadásához két elméletet vázoltam fel: a virtuális erőközpontok modelljének megszerkesztésével az erőérvényesítés új megközelítését adtam (VI.1.2.), továbbá rámutattam, a jelenlegi kompozit (hibrid) hidegháború és a korábbi, atomkorszaki hidegháború párhuzamaira és eltéréseire (VI.1.5.). Több szempontból elemeztem a digitális visszaélések elméleti alapjait (VI.2.) amit illusztráltam az orosz példával (VI.2.3.). Ezek alapján több ellentmondásos helyzetet beazonosítottam, melyek az MI-vel kapcsolatos védelmi döntéshoztalt nehezítik (VI.3.2.). A III. fejezetet továbbgondolva pedig új megállapításokat tudtam tenni az MI-nyár forгатókönyvének védelmi kiaknázhatósága terén (VI.3.1.). Végül az informatika fogalmának kitágítására (V.2.1.) alapozva le lehetett vonni azt a következtetést, hogy más szerepet kell szánni a katonai informatikának egy új, hibrid védelmi paradigmában, valamint a katonai (védelmi) oktatásban. (VI.3.4.)
- **Részkövetkeztetések.** A vizsgálatok alapján újabb részkövetkeztetéseket tehetek.
 - **4R-A:** A védelmi szférában az MI jelentősége tovább tolja az erőérvényesítés hangsúlyát a puha műveletek felé. (VI.1., VI.3.2.)

²⁸⁴ Megjegyzés: a NKE HHK Logisztika Tanszékével közös újgenerációs adatelemzési kurzust tervezünk.

- **4R-B:** A virtuális erőközpontok modellje (VI.1.2.), és a kompozit hidegháború jellemzőinek (VI.1.5) segítségével jobban megragadható a mai helyzet, akár az orosz-ukrán háború esetében is (VI.2.), figyelembe véve az MI-fejlesztésekben várható kulturális különbségeket 2R-D és 1R-C továbbgondolása alapján.
- **4R-C:** A 3R-E következtetés alapján az MI védelmi hatása hozzámérhető a kibertér hatásához; így igazolható, hogy a hatás legalább akkora, de talán nagyobb, mint a kibertér megjelenésének hatása a hadelméletekre (VI.3.3.).
- **4R-D:** Egy hosszú MI-nyár (a technikai ugrás nélküli fejlődés) forgatókönyve nem feltétlenül hátrányos, sőt kiaknázható hazánk és a régióink számára. (III. és VI.3.1.)
- **4R-E:** A katonai és védelmi informatika korábbi szolgáltató-üzemeltető szerepe részben a kibertér, de sokkal inkább majd az MI hatására felértékelődik majd (VI.1.1. és VI.3.4/(2)).
- **4R-F:** Előnyös lenne, ha a katonai és védelmi informatika oktatásban sokkal markánsabban és tervezettebben jelennének meg az MI-vel kapcsolatos kompetenciák (VI.3.4/(3)).

ÖSSZEGZÉS, EREDMÉNYEK, HASZNOSÍTHATÓSÁG

A záró alfejezetek előtt szeretnék köszönetet mondani mindazoknak, akik bármilyen formában hozzájárultak dolgozatom létrejöttéhez. Az alábbi összegzés az utolsó pontjához kapcsolódva szeretném kifejezni abbéli reményemet, hogy az olvasók is érdekesnek, maguk számára hasznosnak ítélték ezt a szerτεágazó anyagot. Bízom benne, hogy életükre is inspiratív hatással tudott lenni némely gondolat.

A VIZSGÁLATOK ÖSSZEGZETT ÉRTÉKELÉSE

A kutatás szerkesztési metóduša alapján minden fejezet végén alapos összegzésre került sor, melyben bemutattam hogyan teljesült az adott témakörre kijelölt cél (vagy célok), valamint megjelöltem, hogy a kutatási kérdésre mely rész alapján milyen választ tudtam adni, és a hipotézist hol bizonyítottam. Ugyanott röviden összefoglaltam a feldolgozott témákat, és ismerttettem a belőlük levont részkövetkeztetéseket. A kutatás feldolgozott témakörei és résztémái pedig sokkal jobban áttekinthetőek a dokumentumvégi részletes tartalomjegyzék alapján, mint egy összefoglaló folyószövegben, hiszen igyekeztem beszédes címeket és részletes alcím-rendszert használni. Ebben a végső összegzésben nem látom értelmét, hogy mindezeket más szavakkal megismételjem. Viszont a kutatás tervezésénél egyéb kritériumokat is magam elé állítottam, ezek teljesülésének ismertetésével még adós maradtam.

Először is a vizsgálatokat úgy szándékoztam elvégezni, hogy teljesítsek bizonyos irányelveket. A kapott anyag hasznosíthatóságát vártam el az államvédelmi tudományok, a hadtudományok és az oktatás területén. Az első két szempont a VI. fejezet alapján egyértelműen teljesült, de minden elemzés tartalmazott ilyen irányú adalékokat. Az oktatási szempontot több fejezetben többféleképpen sikerült teljesítenem. A VI. fejezet során elvi irányokat jelöltem meg (ilyen elveket alaposabban dolgozott fel egy törölt részkutatás [2] is). Az I. és II. fejezet egyes részei pedig gyakorlatias módon valósítják meg ezt ez elvárást, vagyis tananyagként, ami által az erre vonatkozó irányelv is teljesült. Ezen kívül a részkutatásokat olyan irányokban akartam végezni, hogy azokban minél több jelenjen meg a magyarul (vagy általában is) kevésbé tárgyalt MI-vel kapcsolatos területek közül. Az érzelmi intelligenciától (I.3.) a humán, társadalmi vagy filozófiai kapcsolódásokig (pl. V.2.2., III.3., IV.2.) számos részkutatás teljesíti ezt.

Az irányelvek tehát teljesültek, nézzük hogyan sikerült megvalósítani azokat a kiemelt szempontokat, melyeket magam elé tűztem a kutatás tervezésekor.

- **Interdiszciplináris megközelítés.** A kutatás szinte minden fejezetében keveredtek technikai kifejtések legalább a filozófiai és védelmi aspektusokkal, de sok esetben a nyelvi (fogalmi), pszichológiai vagy társadalomtudományi nézőpontot is vizsgáltam.
- **Előtérben a háttér.** A látható dolgok elemzéséből minden témakörnél tudtam elvont, mögöttes, lehetséges okokra visszakövetkeztetni.
- **Prevenció.** Ezzel a szemponttal is sikerült minden témaköri elemzést átítatni, számos szempontból próbáltam egy élhetőbb jövőre irányítani a vizsgálatokat: a jövő káros MI hatásaitól kezdve a visszaéléseken át a félreértések hátráltatásáig.
- **Gyakorlatiasság a részletekben.** A saját fogalmak, meglátások, ötletek, javaslatok, szintezések (stb.) összeállításával, vagy a közérhető (jegyzetté alakítható) fogalmazással, számos praktikus is alkalmazható részt is sikerült disszertációmba beépíteni.
- **Egzisztenciális megközelítés.** Leginkább a III.2., III. 5. és IV.2 alfejezetben sikerült ezt a tényezőt figyelembe venni. Önmagam számára teljesült ez a szempont, hiszen a kutatás sokat segített belső életemben is – remélem az olvasók számára is.

Az összegzés zárásaként megállapítható tehát, hogy sikerült megvalósítani minden, a bevezetőben ismertetett tervet és célt, melyek által az alábbi eredményeket értem el.

ÚJ TUDOMÁNYOS EREDMÉNYEK ÉS AZOK HASZNOSÍTHATÓSÁGA

Itt összevonva ismertetem az elért eredményeket és azok lehetséges felhasználásaira tett ajánlásaimat. Az eredményeket az áttekinthetőség kedvéért címsorokba szedve foglaltam össze, alattuk megjelöltem, mely részkövetkeztetésekből következnek, és igazolásuk mely fejezetekre alapul leginkább. A részkövetkeztetések HIPERLINK segítségével könnyebben elérhetőek (az egeret a hivatkozás fölé kell vinni). Az eredmények számozása nem esik egybe a témakörökével, mivel némely témakör vizsgálatából több olyan konklúziót vontam le, melyek véleményem szerint nevezhetőek tudományos eredménynek. Azokat a konklúziókat, melyeket fontosnak tartok, de nem ítéltam eléggé fajsúlyosnak, azokat járulékos eredményként tüntettem fel.

1. eredmény: Bizonyítottam, hogy a mesterséges intelligencia fejlődésében lassulás várható, ami viszont nem érinti a terjedésének ütemét.

- Ez a tétel az [1R-A](#) részkövetkeztetésen alapul, a megfogalmazás a horizontális és vertikális fejlődés eltéréseinek leírására épül (III.2.2.), a tétel igazolása a III. fejezet elemzéseinek összessége.
- **Járulékos eredmények:**
 - Megalkottam több általánosan használható, új megközelítést, és az ezekhez kapcsolódó új terminológiákat a paradigmaváltás megkönnyítésére. Ezek közül

kiemelendők: a horizontális és vertikális tanulás, a többdimenziós fejlődés (ld. I.2.5. és I.2.6.).

- Kidolgoztam egy javaslatot annak kiaknázhatóságára, ha egy hosszú MI-nyár (a technikai ugrás nélküli fejlődés) forgatókönyve szerint folytatódna a világ változása. (ld. VI.3.1.)

Használhatóság:

- Sokat segíthetne a felnőttképzések, illetve továbbképzések elvárásrendszerének átdolgozásakor a horizontális és vertikális tanulás szétválasztásának figyelembevétele.
- A második járulékos eredmény az oktatási felhasználásokon kívül (ld. 7. eredmény-nél) a stratégiai szintű tervezés számára lehet hasznos, reális és preventív elvárások támasztásához az MIKT szélesebb körű megismerésére.

2. Eredmény: Bizonyítottam, hogy egy mesterséges neurális hálózat képes lehet ugyan bizonyos erkölcsi érzék jellegű kompetencia megvalósítására, de nem képes azt „emberi” szinten elégséges minőségben utánozni.

A tétel a [2R-A](#) részkövetkeztetésen alapul, igazolása a IV.2. alfejezet vizsgálatainak összessége.

Járulékos eredmények:

- Kidolgoztam az emberi és gépi autonómia új típusú, használhatóbb felosztását (IV.1.)
- Kimutattam, hogy az affektív számítástechnika mellett a rajntelligencia is kulcs szerepet játszat abban, hogy az erkölcsi érzék gépi utánzására a kezdeti lépéseket megtegyük (IV.2.5.)

Használhatóságok:

- A fejlesztés és az azzal kapcsolatos tervezések józanságának biztosítása. A [2R-E](#) részkövetkeztetés alapján fontos, hogy a fejlesztések mögötti elméletek további zsákutcáit idejében elkerüljük.
- Az első járulékos eredmény az MI szabályozásának minden szintjén hasznosítható.
- A második járulékos eredmény az MI fejlesztéséhez adhat hasznos adalékokat.

3. Eredmény: Bizonyítottam, hogy az általános MI emberszerűségének lehetőségével, valamint a szuperintelligencia emberfelettségével kapcsolatos felvetések nem tudományos megalapozottságúak.

Ez tétel a [1R-B](#) – [2R-B](#) – [3R-B](#) következtetés-sorozat alapján mondható ki, tehát az első három témakör vizsgálatai (III-IV-V. fejezet) együttesen támasztották alá.

- **Használhatóság:** A fejlesztés és az azzal kapcsolatos tervezések józanságának biztosításához, és a védelmi prioritások realitásának növeléséhez ad fontos szempontokat.

4. eredmény: Meghatároztam a Mesterséges Intelligencia fogalmának olyan új definícióit, melyek a korábbiaknál több tényezőt vesznek figyelembe (még kompromisszumok mellett is).

Az új meghatározást [3R-A](#) részkövetkeztetés alapján kellett megfogalmazni, az ehhez szükséges vizsgálatok (a hiányosságok és elvárások azonosítása) az I., II. és V. fejezetekben kerültek végrehajtásra.

Járulékos eredmények

- Bevezettem a pszeudo-tanulás, pszeudo-MI, pszeudo-autonómia fogalmait. (V.1.4.)
- Lexikonszerűen ismertettem az MI felosztásait, és felvettem a besorolási mátrix módszerét az MI-k osztályozásához (V.3.2.)
- Kimutattam, hogy az „informatika” fogalmának tartalma is változhat, az MI miatt lényegesen átalakulhat (V.2.1.)
- **Hasznosíthatóság:**
 - A megalkotott MI-fogalom, valamint a felosztások és a besorolási mátrix az oktatástól a szabályozásig számos területen hasznosítható az eljövendő kb. 10 évben.
 - Az informatika fogalmának átalakulásával kapcsolatos járulékos eredmény a védelmi használaton túl (ld. 6-7. eredmény) az MI technológiához kapcsolódó egyetemi képzések során lehet fontos, mivel az MI a technika-centrikusság oktatását a kognitív tudományág körképpé szélesítheti.

5. Eredmény: Bizonyítottam, hogy az MI tovább tolja az erőérvényesítés hangsúlyát a puha műveletek felé.

A tétel a 4R-A részkövetkeztetés, valamint 1R-C – 2R-D – 4R-B következtetés-sorozat együttes figyelembevételéből következik; igazolása elsősorban a VI.1. és VI.2. alfejezetekben, valamint a VI.3.2. kissejzetben olvasható.

Hasznosíthatóságok:

- Általánosan hasznosítható a hadtudományban az itt kidolgozott virtuális erőközpontok teóriája, illetve a virtuális erőérvényesítés rendszertani modellje, valamint kompozit (hibrid) hidegháború jellemzése a világban zajló folyamatok leírására.
- Az adaptált és a direkt katonai informatika megkülönböztetése stratégiai, szabályozási, gazdasági és oktatási területen egyaránt hasznosítható megközelítés.
- A védelmi követelmények paradoxainak azonosítása elsősorban stratégiai szempontból fontos, hozzájárulhat a helyes kompromisszumok megragadásához.

6. Eredmény: Bizonyítottam, hogy a kibertér és az MI együttes használatának újszerűségei hangsúlyosabb helyre pozicionálják a katonai informatika szakterületét.

A tétel az 5. Eredmény és 3R-E – 4R-C következtetés-sorozat egyesítéséből következik, igazolása a VI.1. és VI.2. alfejezetek alapján VI.3.3. kisfejezetben

Hasznosíthatóság: A katonai és védelmi szakinformatika új helye, a kidolgozott modellek, és a problémák összefoglalása elsősorban stratégiai szinten használható megközelítés (pl. megfelelő irányelvek megfogalmazásához).

7. Eredmény: Bizonyítottam, hogy az MIKT és egyéb paradigmaváltó technológiák oktatásának beépítése elengedhetetlen a katonai és védelmi képzésbe, és megfogalmaztam az ehhez szükséges tananyag alapjait.

Ez az 5. és 6. eredmények adaptálása a védelmi oktatásba: a tétel 4R-E és F részeredmények közös következménye, igazolása VI.1.1. és VI.3.4. kisfejezetekben.

Hasznosíthatóságok: az eredmény a katonai felsőoktatásban meglévő kurzusok tematikájának MI irányú kiegészítését és új kurzusok létrehozását inspirálja.

Továbbá az Első Rész eleve úgy került kidolgozásra, hogy abból jegyzet vagy tankönyv készülhessen különféle kurzusok számára.

AJÁNLÁS – KIK SZÁMÁRA LEHET HASZNOS A FENTI TANULMÁNY?

A tanulmány nem egyetlen szakma vagy katonai szakirány számára kívánt hasznosítható eredményeket elérni. Elsődleges célközönség a védelmi szféra szakemberei voltak, számukra a tanulmány számos helyén tett hasznosítható meglátásokat, de a VI. fejezet kifejezetten ezen kutatók és szakértők számára lehet hasznos. Szinte alfejezetenként megjelölhetőek lennének a különféle célközönségek, de itt inkább csak fejezetszinten vázolom az elsődleges ajánlásokat, nem zárva ki ezzel a meg nem említett tudományágakat sem.

Az Első rész áttekintő elemzésének célközönségét a 6-os eredmény is mutatja: ez a rész a katonai oktatásban, valamint a védelmi terület mindenféle képzésében hasznosítható. Ezen a célközönségen kívül az I. fejezet inkább a humán érdeklődésű emberek számára segíthet az MI megértésében, míg a II. fejezet pedig azok számára hasznos, akiket érdekel a technikai megvalósítás, a várható lehetőségek és kapcsolódó fejlesztések köre is, de csupán egy áttekintésre kíváncsiak, nem szeretnének mérnöki vagy programozói szinten elmélyülni a technológiákban.

A Második Rész specifikus elemzéseinek első témaköre, vagyis, hogy lesz-e MI tél? (III.) számíthat a legáltalánosabb érdeklődésre. Ez a fejezet minden ember számára kiindulópontként szolgálhat, akit érdekelnek az MI-re vonatkozó tudományos prognózisok. A kutatók közül a

védelmi szakembereken és elemzőkön kívül elsősorban gazdasági és jogi szakemberek egyesíthetik saját tudásukkal az ott leírt összefoglalásokat, vagy a társadalom és technológia kapcsolatának elemzését. A vertikális tanulás és a paradigmaváltás nehézségeinek modellje a pedagógusok mellett bárkit érdekelhet, illetve érinthet személyesen is. A második témakör (IV.) az autonómia részletes vizsgálatával és újszerű felosztásaival elsősorban a jogtudósok számára kínálnak alapot a pontosabb szabályozás megfogalmazásához, a mesterséges erkölcs elemzése pedig azok számára lesz izgalmas, akik saját szakterületüket is szeretik filozófiai, etikai vagy pszichológiai szempontból vizsgálni. A harmadik témakörben áttekintett terminológiai kérdéskör (V.) is a különféle utasítás szintű szabályozók szakértői számára lehet hasznos, ezen felül azonban szerepet kaphat az oktatásban, és az MI-ről való közös társadalmi gondolkodás bármely területén. (A VI. fejezetről az elején ejtettem szót.)

TOVÁBBI KUTATÁSI LEHETŐSÉGEK

A tanulmány számtalan pontján kényszerültem tartalmi lehatárolásra, mivel jelen terjedelem és logikai felépítés miatt az adott téma részletesebb elemzése nem fért bele az oldalszámra vonatkozó keretbe. Ezeket a lehatárolásokat az adott helyen említettem. Mivel bővebb kifejtésük akár több kötetet is megtöltene, de még listájuk is terjedelmes, ezért itt csupán néhány olyan lehetőséget emelek ki, amelyek reális időn belül megvalósíthatóak, vagy már tettem is lépéseket azok publikálása felé.

Először is egy terjedelmi okokból kimaradt hipotézisemhez szükséges kutatást lenne hasznos végigvinni: itt csupán vázlatát írtam le az „IQ-fenntarthatóság”-nak nevezett jelenségnek. (V.5.1.), az ennek folyományaként fellépő agyelszívásnak (III.3.2.) és ezek védelmi aspektusainak (III.7.). Úgy vélem, hogy ez a sejtésem tudományosan is jól alátámasztható lesz kutatásaim folytatása során.

Sürgető területnek tartom a kibertéri MI támadási és védelmi lehetőségek újragondolását. Ezirányú régebbi kutatásaim épp elavultságok miatt nem kaptak itt helyet, viszont azóta egyre több publikáció érhető el a témában. Szeretném továbbá a digitális erőérvényesítés-konceptió alkalmazását mielőbb megfogalmazni az oroszokon kívül más régiókra is. A kínai modell vizsgálatáról is állítottam már össze előadást, melyet ezen tanulmány fényében szükséges lenne majd kiadni. (Meglehetősen érdekes lenne a nyugati, demokratikus közegben felmerülő visszaélések tanulmányozása, erről azonban egyelőre igen-igen kevés tudományos publikáció fellelhető.)

Végül kiemelendő, hogy az Első rész egyetemi jegyzetté átdolgozása már folyamatban van. Különböző fázisban készült munkapéldányokat teszt jelleggel 2024 tavaszi kurzusaiban a hallgatók számára rendelkezésre bocsátottam. A visszajelzések alapján tett javításokkal 2024 őszén

már egy teljesebb jegyzetet adtam közre tesztelésre. A különböző kurzusok számára eltérő oldalszámú jegyzeteket tervezek: a nagyobb óraszámú kurzusok verziójában helyet kap néhány rövid szemelvény a későbbi fejezetekből is, a kis óraszámú kurzusok részére pedig az Első Rész egy jóval kisebb oldalszámra redukált verziója várható kiadásra.

MUTATÓK

RÉSZLETES TARTALOM

RÖVID TARTALOM	1
BEVEZETŐ	3
A KUTATÁS BEMUTATÁSA	3
<i>Felvezetés: a „humán aspektus” és a „védelmi szféra”</i>	3
<i>A téma fontossága, időszerűsége</i>	4
<i>A kutatás személyes motivációja</i>	5
A KUTATÁS TERVEZÉSE	5
<i>A kutatási problémák és a téma lehatárolása</i>	5
<i>Hasznosíthatósági irányelvek (gyakorlati célok)</i>	6
<i>A kutatás kiemelt szempontjai.....</i>	7
<i>Elméleti célok és kérdések</i>	7
<i>Első témakör: lesz-e MI tél?</i>	7
<i>Második témakör: mesterséges erkölcs és szabadság</i>	8
<i>Harmadik témakör: terminológiai kérdések</i>	8
<i>Negyedik témakör: az MI hatása a védelmi szférára.....</i>	8
KUTATÁSI HIPOTÉZISEK, AZ AZOKBÓK TERVEZETT EREDMÉNYEK ÉS FELHASZNÁLÁSOK	8
1. <i>hipotézis (H1): A mesterséges intelligencia fejlődésében lassulás várható.</i>	8
2. <i>hipotézis (H2): Egy mesterséges neurális hálózat képes lehet ugyan bizonyos erkölcsi érzék jellegű kompetencia megvalósítására, de nem olyan minőségben, ami egy általános MI-t emberszerűvé tenne.</i>	8
3. <i>hipotézis (H3): A Mesterséges Intelligencia jelenlegi meghatározásaiból olyan fontos aspektusok hiányoznak, melyek hiányában a technológia félreérthető, így kompromisszumok mellett is egy jelenlegieknél teljesebb definícióra lenne szükség.</i>	9
4. <i>hipotézis (H4): az MI tovább tolja az erőérvényesítés hangsúlyát a puha műveletek felé, ennek kezeléséhez pedig a katonai és védelmi informatika radikálisan új megközelítése lesz szükséges</i>	9
A KUTATÁS MEGVALÓSULÁSI TERVE	9
<i>Kutatási módszerek</i>	9
<i>Az értekezés logikai felépítése</i>	10
<i>Áttekintés a szakirodalomról</i>	11
<i>Stilisztikai magyarázatok</i>	11
ELSŐ RÉSZ: ÁTTEKINTŐ ELEMZÉS	13
I. EMBERI ÉS GÉPI INTELLIGENCIÁK.....	14
I.1. A MESTERSÉGES INTELLIGENCIA FOGALMÁNAK JELENE.....	14
I.1.1. Az MI fogalom megjelenése és az MI hullámai.....	14

I.1.2.	<i>A mesterséges intelligencia mai hivatalos meghatározásai Magyarországon és az EU-ban</i>	
	16	
I.1.3.	<i>Az USA megoldása párhuzamos definíciókkal</i>	19
I.1.4.	<i>Mesterséges ügynökök (ágensek)</i>	21
I.2.	AZ INTELLIGENCIA, AMI MESTERSÉGES IS LEHET	24
I.2.1.	<i>Nyelvi megközelítés</i>	24
I.2.2.	<i>Az intelligenciafajták</i>	27
I.2.3.	<i>A gépi intelligencia tudásszintjei</i>	30
I.3.	ÉRZELMI INTELLIGENCIA A GÉPEKBEN	32
I.3.1.	<i>Háttér a technológia mögött: az érzelmek fontosságának története dióhéjban</i>	32
I.3.2.	<i>Az affektív számítástechnika kifejlődése</i>	35
(1.)	<i>Az affektív számítástechnika fogalma és kezdetei</i>	35
(2.)	<i>Az érzelemmotortól a mesterséges empátiáig</i>	37
(3.)	<i>Érzelem-chip, affektus-radar, mesterséges empátia</i>	39
I.4.	RÉSZÖSSZEFOGLALÁS: HOGYAN LEHET AZ INTELLIGENCIA MESTERSÉGES	42
II.	TECHNIKAI KÖRÚT A MESTERSÉGES INTELLIGENCIA KÖRÜL	44
II.1.	MIKT: AZ MI-HEZ SZOROSAN KAPCSOLÓDÓ TECHNOLÓGIÁK	44
II.1.1.	<i>BigData, strukturálatlan adathalmazok</i>	44
II.1.2.	<i>Felhőtechnológia és technológiai virtualizáció</i>	47
II.1.3.	<i>IoT (Internet of Things) szenzorok és robotok</i>	50
II.1.4.	<i>Az új technológiákon új ökoszisztéma: az Ipar 4.0</i>	51
II.1.5.	<i>Az MIKT fogalom bevezetése</i>	52
II.2.	GÉPI TANULÁSI MODELLEK	53
II.2.1.	<i>Az ML modellek közös jellemzői</i>	54
II.2.2.	<i>Más cél más ML</i>	56
II.2.3.	<i>Az utóbbi évek újításai</i>	58
II.2.4.	<i>Az ML népszerű felosztásai és a pseudo-tanulás</i>	60
II.3.	ELVEK ÉS TECHNOLÓGIÁK AZ MI KÖRÜL	62
II.3.1.	<i>A jelen és a jövő MI-célú hardverei</i>	63
II.3.2.	<i>Fuzzy logika, az „elmosódott igazság”</i>	68
II.3.3.	<i>Biológiai ihletettséggű informatikától a rajntelligenciáig</i>	72
II.3.4.	<i>A természetes nyelvfeldolgozás (NLP rendszerek)</i>	76
II.4.	RÉSZÖSSZEFOGLALÁS: MILYEN AZ MI KÍVÜL-BELÜL?	79
	MÁSODIK RÉSZ: SPECIFIKUS VIZSGÁLATOK	80
III.	AZ MI NYÁRBAN NYARALHAT-E AZ EMBER?	80
III.1.	AZ MI TELEI ÉS TAVASZAI	80
III.1.1.	<i>A kutatás sajátos megközelítése</i>	80

III.1.2.	<i>A két tél dióhéjban</i>	82
III.1.3.	<i>Okok a szakirodalom szerint</i>	83
III.2.	A KORSZAKOK ELTÉRÉSEI	85
III.2.1.	<i>Más korszak más fejlődés</i>	85
III.2.2.	<i>Horizontális és vertikális fejlődés – stagnálás gazdasági növekedés alatt</i>	86
III.2.3.	<i>A hype tényező változása és vallásiassá válása</i>	87
III.2.4.	<i>A szabályozhatóság elvi háttere</i>	88
(1.)	A szabályzás filozófiai alapjainak szükségessége	88
(2.)	A vizsgálati szempontok bővítendőik	89
III.3.	AZ MI ÉS TÁRSADALOM KÖLCSÖNHATÁSAI ÉS VISSZAHATÁSAI	90
III.3.1.	<i>A technológiára ható társadalmi eredő-vektorok</i>	90
(1.)	A kölcsönhatás mozgatórugói	90
(2.)	A társadalom felől érkező hatások elvekké válása	91
(3.)	A fejlődéshit és az MI fejlődés módja.....	92
III.3.2.	<i>Az MI felől a társadalomra ható vektorok</i>	93
(1.)	A lelki, pszichológiai visszahatások	93
(2.)	Új etika a gépi dimenziók által?.....	94
(3.)	Új hagyományok és új eszmék.....	95
(4.)	Fenntarthatósági probléma a munkaerő terén.	95
(5.)	Jobban megismerjük az Embert.....	96
III.4.	HIDEGFRONTOK AZ MI NYÁRBAN.....	97
III.4.1.	<i>Kifejezések eróziója</i>	97
III.4.2.	<i>Félig leküzdött problémák: érvelés, általánosítás, öntanítás</i>	98
III.4.3.	<i>A nyers-erő paradigma kérdőjelei</i>	99
III.4.4.	<i>A fekete doboz</i>	100
III.4.5.	<i>Elfogultság</i>	101
III.4.6.	<i>Kiberbiztonsági sajátosságok</i>	102
III.5.	A VÉGES PARADIGMAVÁLTÁSI KÉPESSÉG ÉS EGYÉB HUMÁN TÉNYEZŐK.....	102
III.5.1.	<i>A vertikális tanulás korlátossága</i>	103
III.5.2.	<i>IQ-fenntarthatóság (a tehetségek végessége)</i>	107
III.5.3.	<i>Néhány további tényező</i>	108
III.5.4.	<i>Filozófiai problémákról és ezek háttéréről</i>	110
III.6.	A TOVÁBBLÉPÉS LEHETŐSÉGEI	112
III.6.1.	<i>Paradigmaváltás, de csak a szavakban: multikogníciós rendszerek igazi AGI helyett</i>	112
(1.)	A KUTATÁS ZÁRÁSA UTÁNI FEJLEMÉNY.....	114
III.6.2.	<i>Az intelligencia szokásos szintjeinek kritikája</i>	115
III.6.3.	<i>Más MI-paradigmák említése</i>	116
III.7.	ÖSSZEGZÉS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI-NYÁR ÚJ MEGHATÁROZÁSA.....	116
IV.	AZ ERKÖLCSI ÉRZÉK DIGITALIZÁLHATÓSÁGA	119

IV.1.	AZ AUTONÓMIA ANATÓMIÁJA.....	119
IV.1.1.	<i>Autonómia az embereknél</i>	120
(1.)	Etimológia, meghatározás és alapfelosztás.....	120
(2.)	Általános Négy-típusú Autonómia Felosztás (4TA).....	121
IV.1.2.	<i>A gépi autonómia vizsgálata</i>	126
(1.)	Az automatika és a gépgenerációk.....	126
(2.)	Autonómiaszintek, melyek valójában automatikasintek.....	128
(3.)	Gépi-autonómia szintek (saját felosztás).....	131
(4.)	A felvetett szintek élőlénybe kapcsolt MI és magas rendelkezésreállítás esetén.....	137
(5.)	Autonómia a kötelesség és felelősség kettősségében.....	138
IV.1.3.	<i>Emberi és gépi autonómia összevetése</i>	140
IV.2.	AZ MESTERSÉGES ERKÖLCSI ÉRZÉK LEHETŐSÉGE.....	142
IV.2.1.	<i>A moralizáló gépek lehetősége elő-vizsgálata</i>	142
(1.)	Morális elvek tanulása.....	142
(2.)	A kognitivisták megközelítés kritikája.....	143
IV.2.2.	<i>A „humánkomponensnek”: emberi döntések gépi analógiái</i>	145
(1.)	A tett-központú embervizsgálat fontossága a gépi leképezésekhez.....	145
(2.)	A humánkomponensnek vázlatos körüljárása.....	145
(3.)	A funkcionalizmus kritikája és az antropológiai keret tágítása.....	148
IV.2.3.	<i>Az etika digitalizálhatóságának technikai oldala</i>	150
(1.)	Fogalmak leképezése, matematikaivá alakítása.....	150
(2.)	Raszteresség és vektorizálás a gépi döntéshozatalban.....	152
(3.)	Az MI, mint véleményformáló – de lehet tudata?.....	153
(4.)	Az agy-protézis kísérlet alkalmazása a gépi etika területén.....	154
IV.2.4.	<i>Objektív MI készíthetnek a tudattalanul szubjektív alkotók?</i>	155
(1.)	Elfogultsági, avagy előítéletességi kutatások.....	156
(2.)	Az alkotók és csapatok szubjektív nézetei az MI rendszerekben.....	158
(3.)	A kulturális háttér hatása az MI fejlesztésekre.....	159
(4.)	Az Üres Lapra alapuló objektív gépi intelligencia kritikája.....	160
(5.)	Mennyiben ért meg az MI olyan eszméket, mint egyenlőség és fejlődés?.....	161
IV.2.5.	<i>A rajntelligencia – a gépi erkölcs kulcsa?</i>	163
IV.3.	ÖSSZEGZÉS ÉS RÉSZKÖVETKEZTETÉSEK:.....	164
V.	A MESTERSÉGES INTELLIGENCIA FOGALMÁNAK ÚJRAGONDOLÁSA.....	167
V.1.	FOGALMI PROBLÉMÁK A „DIGI-DOLGOKTÓL” AZ MIKT RENDSZEREKIG.....	167
V.1.1.	<i>A rendszerek összemosódásának problémája</i>	167
V.1.2.	<i>Neurális adatbázisok</i>	170
V.1.3.	<i>Okosdolgok terminológiai anomáliái</i>	170
V.1.4.	<i>Pszedo-géptanulás, pszeudo MI, pszeudo-autonómia</i>	174
V.2.	AZ INFORMATIKA ÉS A HUMÁNTUDOMÁNYOK VÁLTOZÁSAI AZ MI HATÁSÁRA.....	175
V.2.1.	<i>Az informatika szó kettős használata</i>	176

V.2.2.	<i>Információtudomány, információelmélet, kognitív tudomány – és minden tudomány konvergenciája</i>	179
V.2.3.	<i>A mesterséges intelligencia újszerűsége</i>	183
V.3.	JAVASLAT AZ MI MEGHATÁROZÁSÁRA	186
V.3.1.	<i>Besorolási mátrix a túl sokféle MI-hez</i>	186
(1.)	A három alaptulajdonságból adódó felosztások	186
(2.)	Egyéb MI felosztási lehetőségek	188
(3.)	A besorolás mátrix felvetése	190
V.3.2.	<i>A fogalmakból hiányzó tényezők és hatásuk</i>	192
V.3.3.	<i>Új MI meghatározások korlátai</i>	195
V.3.4.	<i>A javasolt MI fogalom</i>	196
V.4.	ÖSSZEFOGLALÁS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI FOGALOMVIZSGÁLATA	197
VI.	MESTERSÉGES INTELLIGENCIA ÉS ERŐÉRVÉNYESÍTÉS	199
VI.1.	AZ MI-ALAPÚ HIDEGHÁBORÚ: A HIBRID MŰVELETEK SPECIÁLIS ELEME	199
VI.1.1.	<i>Paradigmaváltás a nyomásgyakorlásban</i>	199
VI.1.2.	<i>A virtuális erőközpontok modellje</i>	201
VI.1.3.	<i>Az MI direkt és az adaptált védelmi felhasználása</i>	202
VI.1.4.	<i>A helyzet változásához jól illenek az MI képességei</i>	204
VI.1.5.	<i>A mai hidegháború jellemzése</i>	206
(1.)	A mai és a klasszikus hidegháború összevetése	206
(2.)	A gépi érzelmek néhány védelmi alkalmazási lehetősége	208
VI.2.	ÉLNI ÉS VISSZAÉLNI A DIGITÁLIS TÉRBEN	211
VI.2.1.	<i>Szuverenitás, ökoszisztéma, tekintélyelvűség a digitális térben</i>	212
VI.2.2.	<i>A digitális visszaélés irányai az erőközpont-modell alapján</i>	213
VI.2.3.	<i>Az orosz modell elemzése</i>	214
(1.)	Az orosz állami mesterséges intelligencia helyzete	215
(2.)	Az orosz intranet	216
(3.)	Az orosz modell kialakulása, jellemzői és a szuverenitás esélye	218
VI.2.4.	<i>Következtetések az ukrán-orosz konfliktus alapján</i>	219
VI.3.	A TANULMÁNY EDDIGI VIZSGÁLATAI ÉS A VÉDELEM	220
VI.3.1.	<i>Az MI tél forgatókönyve védelmi aspektusból</i>	220
(1.)	A korábbi telek mögötti okok védelmi kockázatai	220
(2.)	Néhány katonai vonatkozás	222
(3.)	Előnyök kiaknázása	225
VI.3.2.	<i>Az állami védelmi döntések változása</i>	226
(1.)	A hibrid hidegháborús háttér paradoxona	226
(2.)	A korlátozás paradoxonai és következményei	227
(3.)	Az MI terjedés dotációjának kérdése	227
(4.)	A várható negatív hatások felvállalásának problémája	228

(5.)	A nemzetállamiság MI-paradoxona	228
(6.)	Nemzetközi megoldás.	229
(7.)	Az optimálisabb, de etikátlanabb MI paradoxona	229
VI.3.3.	<i>Védelmi paradigmaváltások összevetése: az MI és a kibertér</i>	230
VI.3.4.	<i>A védelmi informatika és az MI védelmi oktatása átalakulóban</i>	232
(1.)	Informatikából információtan	232
(2.)	A katonai informatika felértékelődése	233
(3.)	A védelmi MI oktatása	235
VI.4.	ÖSSZEFOGLALÁS ÉS RÉSZKÖVETKEZTETÉSEK: AZ MI VÉDELMI ÚJSZERŰSÉGE.....	236
ÖSSZEGZÉS, EREDMÉNYEK, HASZNOSÍTHATÓSÁG.....		238
A VIZSGÁLATOK ÖSSZEGZETT ÉRTÉKELÉSE		238
EREDMÉNYEK ÉS AZOK HASZNOSÍTHATÓSÁGA		239
1. eredmény: <i>Bizonyítottam, hogy a mesterséges intelligencia fejlődésében lassulás várható, ami viszont nem érinti a terjedésének ütemét.</i>		239
2. Eredmény: <i>Bizonyítottam, hogy egy mesterséges neurális hálózat képes lehet ugyan bizonyos erkölcsi érzék jellegű kompetencia megvalósítására, de nem képes azt „emberi” szinten elégséges minőségben utánozni.</i>		240
3. Eredmény: <i>Bizonyítottam, hogy az általános MI emberszerűségének lehetőségével, valamint a szuperintelligencia emberfelettségével kapcsolatos felvetések nem tudományos megalapozottságúak.</i>		240
4. eredmény: <i>Meghatároztam a Mesterséges Intelligencia fogalmának olyan új definícióit, melyek a korábbiaknál több tényezőt vesznek figyelembe (még kompromisszumok mellett is).</i>		241
5. Eredmény: <i>Bizonyítottam, hogy az MI tovább tolja az erőérvényesítés hangsúlyát a puha műveletek felé.</i>		241
6. Eredmény: <i>Bizonyítottam, hogy a kibertér és az MI együttes használatának újszerűségei hangsúlyosabb helyre pozicionálják a katonai informatika szakterületét.</i>		242
7. Eredmény: <i>Bizonyítottam, hogy az MIKT és egyéb paradigmaváltó technológiák oktatásának beépítése elengedhetetlen a katonai és védelmi képzésbe, és megfogalmaztam az ehhez szükséges tananyag alapjait.</i>		242
AJÁNLÁS – KIK SZÁMÁRA LEHET HASZNOS A FENTI TANULMÁNY?		242
TOVÁBBI KUTATÁSI LEHETŐSÉGEK.....		243
MUTATÓK.....		245
RÉSZLETES TARTALOM		245
TÁBLÁZATOK JEGYZÉKE		251
ÁBRÁK JEGYZÉKE		251
SZAKIRODALMAK JEGYZÉKE		251

TÁBLÁZATOK JEGYZÉKE

1. táblázat: Kulcsszavak az MI definíciókban (saját szerkesztés).....	20
2. táblázat: 22 intelligencia-típus (több modell alapján saját összeállítás)	28
3. táblázat: Az új MI definícióban megjelenítendő tényezők.....	194
4. táblázat: Miért félreérthetőek egyes kifejezések nélkül a jelenlegi MI-meghatározások?.	195
5. táblázat: A XX. és a XXI. százai hidegháborúk főbb különbségei (saját készítés)	208

ÁBRÁK JEGYZÉKE

1. ábra: Az ágensek működési sémája (saját készítés, forrás a szövegben)	23
2. ábra: ANN mélytanuló alapmodell (saját átszerkesztés).....	54
3. ábra: Egy neuronhoz tartozó súlyozások és torzítás figyelembevétele (saját készítés)...	55
4. ábra: Az ANN modell (forrás a lábjegyzetben).....	56
5. ábra: Az RNN modell (forrás a lábjegyzetben).....	56
6. ábra: a CNN modell (forrás a lábjegyzetben).....	56
7. ábra: A CNN modell alaposabb bemutatása (forrás a lábjegyzetben).....	57
8. ábra: Egy generatív MI modell, és annak „ellenséges” elemei (saját készítés).....	60
9. ábra: A gépi tanulás architektúra szerinti felosztása	61
10. ábra A gépi tanulás tanítás szerinti felosztása (saját készítés).....	61
11. ábra A Stern-féle mátrix logika kapcsolata más logikákkal (forrás a szövegben)	65
12. ábra: sebesség fuzzyfikálása példán (saját szerkesztés)	70
13. ábra: A MI eddigi két téli időszak és lehetséges folytatások (forrás a szövegben).....	81
14. ábra: Okosnak hívott eszközök, a Pseudo-MI és az MIKT (saját készítés).....	172
15. ábra: Besorolás-szintek exceleles ábrázolása (saját készítés)	191
16. ábra: MI besorolás-mátrix lehetséges vizualizációi (szemléltetés, forrás: internet)....	191
17. ábra: A láthatatlan hidegháború az MI vonatkozásában (saját készítés).....	205
18. ábra: : A digitális visszaélések főbb irányai (saját készítés)	214

SZAKIRODALMAK JEGYZÉKE

- [1] D. T. Willingham, „Ask the Cognitive Scientist”, *American Federation of Teachers*, szept. 2023, Elérés: 2023. november 9. [Online]. Elérhető: <https://www.aft.org/ae/summer2002/willingham>
- [2] A. T. Fehér, „A tanulás 2.0 védelmi vonatkozásai – Milyen védelmi kihívásokat támaszt az emberi tanulással szemben a gépi tanulás paradigmaváltó hatása?”, in *A mesterséges intelligencia hatásainak specifikus vizsgálata*, S. Magyar, Szerk., Budapest: KNBSZ, 2025.
- [3] P. Stone és mtsai., „Artificial Intelligence and Life in 2030: The One Hundred Year Study on Artificial Intelligence”, 2022. október 31., *arXiv*: arXiv:2211.06318. doi: 10.48550/arXiv.2211.06318.
- [4] Stuart Russell és mtsai., *Mesterséges intelligencia elektronikus almanach*. Budapest: Panem Kft, 2019. [Online]. Elérhető: <https://dtk.tanonyvtar.hu/xmlui/handle/123456789/7622?show=full>
- [5] Buzás G. M., „A mesterséges intelligencia története”, *Gasztroenterológiai és hepatológiai szemle*, köt. 7, sz. 3, Art. sz. 3, okt. 2021.

- [6] Dale. F. Reding és J. Eaton, *NATO ST Trends Report 2020-2040: Exploring the S&T Edge*. Belgium: NATO Science & Technology Organization, 2020.
- [7] I. Oksana, „Symbiotic Intelligence: Exploring the Intersection of Human and Artificial Intelligence”, *Best AI Digest*. Elérés: 2024. január 6. [Online]. Elérhető: <https://medium.com/best-ai/symbiotic-intelligence-exploring-the-intersection-of-human-and-artificial-intelligence-b2fd88a094d5>
- [8] Magyar Innovációs és Technológiai Minisztérium, „Magyarország Mesterséges Intelligencia Stratégiája 2020-2030”, Innovációs és Technológiai Minisztérium, Budapest, 2020. Elérés: 2022. december 22. [Online]. Elérhető: <https://ai-hungary.com/api/v1/companies/15/files/137203/view>
- [9] EU, „Az első uniós rendelet a mesterséges intelligenciáról”, *Hírek*. Elérés: 2024. március 6. [Online]. Elérhető: <https://www.europarl.europa.eu/topics/hu/article/20230601STO93804/az-első-unios-rendelet-a-mesterseges-intelligenciáról>
- [10] Európa Parlament és Tanács rendelete, „A mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról”. Elérés: 2023. november 3. [Online]. Elérhető: <https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:52021PC0206>
- [11] L. Bertuzzi, „EU lawmakers set to settle on OECD definition for Artificial Intelligence”, *www.euractiv.com*. Elérés: 2023. november 6. [Online]. Elérhető: <https://www.euractiv.com/section/artificial-intelligence/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/>
- [12] Európa Parlament és Tanács rendelete, „Mellékletek »A mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról«-hoz”. Elérés: 2023. november 3. [Online]. Elérhető: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0016.02/DOC_2&format=PDF
- [13] Európa Bizottság MI Szakértői Csoport, Szerk., *Az AI (mesterséges intelligencia) meghatározása: Főbb képességek és tudományterületek*. Brüsszel: Publications Office of the EU, 2019. [Online]. Elérhető: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60652
- [14] L. Cominelli, D. Mazzei, és D. E. Rossi, „SEAI: Social Emotional Artificial Intelligence Based on Damasio’s Theory of Mind”, *Front. Robot. AI*, köt. 5, o. 6, 2018, doi: 10.3389/frobt.2018.00006.
- [15] K. M. Saylor, *Artificial Intelligence and National Security*. Congressional Research SVC, 2020. Elérés: 2024. január 4. [Online]. Elérhető: <https://crsreports.congress.gov/product/pdf/R/R45178>
- [16] E. Gosselin-Malo, „NATO to update artificial intelligence strategy amid new threats”, *C4ISRNet*. Elérés: 2024. január 4. [Online]. Elérhető: <https://www.c4isrnet.com/artificial-intelligence/2023/11/30/nato-to-update-artificial-intelligence-strategy-amid-new-threats/>
- [17] Defense Innovation Board, „AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense”. United States Department of Defense, 2019. Elérés: 2023. december 18. [Online]. Elérhető: https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF
- [18] A. T. Fehér és I. Négyesi, „A gépi érzelmek a fegyveres erőknél és az autonóm rendszerekben”, *Hadtudományi Szemle*, köt. 14, sz. 3, o. 163–176, 2021, doi: 10.32563/hsz.2021.3.12.
- [19] S. J. Russell, P. Norvig, és J. F. Canny, *Mesterséges intelligencia - modern megközelítésben*. Budapest: Panem, 2005.
- [20] C. Hewitt, P. Bishop, és R. Steiger, „A universal modular ACTOR formalism for artificial intelligence”, in *Proceedings of the 3rd international joint conference on Artificial intelligence*, in IJCAI’73. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 0 1973, o. 235–245. Elérés: 2024. szeptember 27. [Online]. Elérhető: <https://www.ijcai.org/Proceedings/73/Papers/027B.pdf>

- [21] A. Biddwan, „AI Agents - Types, Benefits and Examples”, Yellow.ai. Elérés: 2024. szeptember 28. [Online]. Elérhető: <https://yellow.ai/blog/ai-agents/>
- [22] J. Cassell, J. Sullivan, S. Prevost, és E. Churchill, *Embodied Conversational Agents*. MIT Press, 2000. [Online]. Elérhető: <https://mitpress.mit.edu/9780262032780/embodied-conversational-agents/>
- [23] R. Brooks, „Just Calm Down About GPT-4 Already”, 2023. Elérés: 2024. június 28. [IEEE Spectrum]. Elérhető: <https://spectrum.ieee.org/gpt-4-calm-down>
- [24] Négyesi I., *A mesterséges intelligencia katonai felhasználásának lehetőségei*, köt. I. Budaepst: Zrinyi Kiadó, 2022.
- [25] H. Gardner, *Frames of mind - The Theory of Multiple Intelligences*. New York: Basic Book, 1983. [Online]. Elérhető: https://www.academia.edu/36707975/Frames_of_mind_the_theory_of_multiple_inteligenes
- [26] M. Habeeb, „13 Types Of Intelligence (Which Ones Are You?) | LinkedIn”, *Capital Market Cafe*, 2023, Elérés: 2024. január 18. [Online]. Elérhető: <https://www.linkedin.com/pulse/13-types-intelligence-which-ones-you-habeeb-mahmood/>
- [27] K. Vijay A., „Narrow AI vs. General AI vs. Super AI”, Spiceworks. Elérés: 2024. január 21. [Online]. Elérhető: <https://www.spiceworks.com/tech/artificial-intelligence/articles/narrow-general-super-ai-difference/>
- [28] N. Bostrom, *Szuperintelligencia [utak, veszélyek, stratégiák]*. Budapest: Ad Astra, 2015. [Online]. Elérhető: <https://pdfcoffee.com/nick-bostrom-szuperintelligencia-pdf-free.html>
- [29] B. Pokol, „A mesterséges intelligencia: egy új légréteg kialakulása?”, *Informacios Tarsadalom*, köt. 17, sz. 4, o. 39, 2017, doi: 10.22503/inftars.XVII.2017.4.3.
- [30] K. Vijay A., „What Is Super Artificial Intelligence (AI)? Definition, Threats, and Trends”, Spiceworks. Elérés: 2024. január 31. [Online]. Elérhető: <https://www.spiceworks.com/tech/artificial-intelligence/articles/super-artificial-intelligence/>
- [31] R. Kurzweil, *A szingularitás küszöbén*. Budapest: Ad Astra, 2013.
- [32] Pokol B., *A mesterséges intelligencia társadalma*. Budapest: Kairosz, 2018.
- [33] R. Waugh, „5 predictions in new book from Google futurist who foresaw the iPhone”, Mail Online. Elérés: 2024. július 8. [Online]. Elérhető: <https://www.dailymail.co.uk/sciencetech/article-13596753/Tech-prophet-predicted-iPhone-years-advance-makes-alarming-forecasts-coming-years.html>
- [34] Tegmark M., *Élet 3.0*. HVG Kiadó Zrt, 2018.
- [35] R. Bitera, „A kiterjesztett ember”, NKE HHK, Budapest, 2024.
- [36] C. Pléh, *A lélektan története. Bp: Osiris, 2010*. 2010. Elérés: 2024. január 24. [Online]. Elérhető: https://www.researchgate.net/publication/313656494_A_lelektan_tortenete_Bp_Osiris_2010
- [37] S. S. Tomkins, *Affect, imagery, consciousness*. New York: Springer, 1962.
- [38] S. S. Tomkins, „Interest—excitement.”, in *Affect, imagery, consciousness, Vol. 1: The positive affects.*, New York: Springer Publishing Co, 1992, o. 336–368. doi: 10.1037/14351-010.
- [39] V. Csépe, M. Győri, és A. Ragó, *Általános pszichológia 1-3: 3. Nyelv, tudat, gondolkodás*. in Osiris tankönyvek. Budapest: Osiris, 2008. [Online]. Elérhető: <https://www.szaktars.hu/osiris/view/csepe-valeria-gyori-miklos-rago-anett-szerk-altalanos-pszichologia-3-nyelv-tudat-gondolkodas-osiris-tankonyvek-2008/?pg=4&layout=s>
- [40] D. L. Nathanson, *Shame and pride: Affect, sex, and the birth of the self*. New York and London: Norton, 1992.
- [41] P. Ekman, *Emotion in the human face*, Second edition. Malor books reprint Edition. Los Altos, California: Malor Books, 2013.
- [42] B. Margitházi, „Érzelemgép, fertőzés és tükörneuronok”, *Metropolis*, 2021, [Online]. Elérhető: <http://metropolis.org.hu/erzelemgep-fertozes-es-tukorneuronok>
- [43] B. Kun, *Az érzelmi intelligencia és az emocionális és szociális kompetenciák szerepe a pszichoaktív-szer-használásban*. Budaepst: ELTE doktori disszertáció, 2011. Elérés: 2021. november 11. [Online]. Elérhető: http://pszichologia.phd.elte.hu/vedesek/KunB_phd_disszertacioja.pdf

- [44] R. W. Picard, „Affective Computing”, *M.I.T Media Laboratory Perceptual Computing Section Technical Report*, köt. No. 321., 1995, [Online]. Elérhető: <https://affect.media.mit.edu/pdfs/95.picard.pdf>
- [45] A. M. Turing, „Computing machinery and intelligence”, *Mind*, köt. LIX, sz. 236, o. 433–460, 1950, doi: 10.1093/mind/LIX.236.433.
- [46] I. R. Murray és J. L. Arnott, „Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion”, *The Journal of the Acoustical Society of America*, köt. 93, sz. 2, o. 1097–1108, 1993, doi: 10.1121/1.405558.
- [47] J. Lee és T. L. Kunii, „Model-based analysis of hand posture”, *IEEE Computer Graphics and Applications*, köt. 15, sz. 5, o. 77–86, 1995, doi: 10.1109/38.403831.
- [48] Pat Research, „What is Affective computing? Top 15 Affective computing Companies”, *PAT RESEARCH*, 2020, [Online]. Elérhető: <https://www.predictiveanalyticstoday.com/what-is-affective-computing/>
- [49] Atsushi. Kunitatsu és mtsai., „Vector unit architecture for emotion synthesis”, *IEEE Micro*, köt. 20, sz. 2, o. 40–47, 2000, doi: 10.1109/40.848471.
- [50] J. Tao, T. Tan, és R. W. Picard, Szerk., *Affective computing and intelligent interaction*, köt. 3784. in Lecture Notes in Computer Science, vol. 3784. Berlin and Heidelberg: Springer, 2005. doi: 10.1007/11573548.
- [51] M. Minsky, *The emotion machine: commonsense thinking, artificial intelligence, and the future of the human mind*, 1. Simon & Schuster trade paperback ed. New York: Simon & Schuster, 2007.
- [52] J. Bullington, Szerk., „Affective” computing and emotion recognition systems: Whitman (Ed.) 2005 – Information Security Curriculum Development Conference. Information Security Curriculum Development Conference, 2005. doi: 10.1145/1107622.1107644.
- [53] A. Khashman, „Emotional System for Military Target Identification Prof”, *NATO Research and Technology Organisation Symposium on Information Management-Exploitation, Stockholm*, 2009, Elérés: 2022. március 11. [Online]. Elérhető: <https://www.semanticscholar.org/paper/Emotional-System-for-Military-Target-Identification-Khashman/598e392add0580da30ae50559e8361df27e34d0d>
- [54] M. Asada, „Development of artificial empathy”, *Neuroscience research*, köt. 90, o. 41–50, 2015, doi: 10.1016/j.neures.2014.12.002.
- [55] R. Zielinski, „Artificial intelligence being trained to show empathy”, ReadWrite. Elérés: 2024. január 28. [Online]. Elérhető: <https://readwrite.com/artificial-intelligence-being-trained-to-show-empathy/>
- [56] EmoShape Confidential, „Emotion Processing Unit III Brochure”. EmoShape, 2018. [Online]. Elérhető: <https://emoshape.com/wp-content/uploads/2019/04/EPU-III-Brochure.pdf>
- [57] D. Woods, „Why Big Data Needs Natural Language Generation to Work”, Elérés: 2021. január 27. [Online]. Elérhető: <https://www.forbes.com/sites/danwoods/2015/07/09/why-big-data-needs-natural-language-generation-to-work/?sh=7604e07b156c>
- [58] „Big data: definition, benefits, challenges (infographics) | News | European Parliament”. Elérés: 2023. november 2. [Online]. Elérhető: <https://www.europarl.europa.eu/news/en/headlines/society/20210211STO97614/big-data-definition-benefits-challenges-infographics>
- [59] Z. Szűts és J. Yoo, „Big Data, az információs társadalom új paradigmája”, *InfTars*, köt. 16, sz. 1, o. 8, márc. 2016, doi: 10.22503/inftars.XVI.2016.1.1.
- [60] L. Á. Farkas és A. Vida, „Big Data az erőfölény vonatkozásában (a digitális platformok meghódítója vagy a versenyhatóságok legnagyobb fejtörése)”, *a Gazdasági és Verseny Hivatal pályázatán nyertes tanulmány*, 2020, [Online]. Elérhető: https://www.gvh.hu/pfile/file?path=/gvh/versenykultura_fejlesztes/tanulmanyi_verseny/palyazati_eredmenyek/farkas-luca-agnes---vida-alexandra&inline=true
- [61] P. Mell és T. Grance, „The NIST Definition of Cloud Computing”, National Institute of Standards and Technology, NIST Special Publication (SP) 800-145, szept. 2011. doi: 10.6028/NIST.SP.800-145.

- [62] D. Vadász, *Operációs rendszerek*, 2002. kiad. Miskolc: Miskolci Egyetem, 2002. [Online]. Elérhető: <https://users.iit.uni-miskolc.hu/~vadasz/GEIAL302B/GEIAL202-Operacios-rendszerek-jegyzet.pdf>
- [63] A. Aliyu és mtsai., „Cloud Computing in VANETs: Architecture, Taxonomy, and Challenges”, *IETE Technical Review*, köt. 35, sz. 5, o. 523–547, 2018, doi: 10.1080/02564602.2017.1342572.
- [64] Antalóczy T., „Így működik a Tesla táblafelismerő rendszere”, *Villanyautósok*. Elérés: 2024. március 4. [Online]. Elérhető: <https://villanyautosok.hu/2020/04/27/igy-mukodik-a-tesla-tablafelismero-rendszere/>
- [65] „Internet of Things (IoT)”, ENISA. Elérés: 2023. november 2. [Online]. Elérhető: <https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/iot>
- [66] P. Török és I. Négyesi, „A short overview for the cognition of the Internet of Things (MTMT)”, *RED - American Journal Of Research Education And Development*, sz. 2, o. 4–15, 2017.
- [67] Directorate-General for Internal Policies of the Union (European Parliament), M. Carlberg, S. Kreutzer, J. Smit, és C. Moeller, *Industry 4.0*. LU: Publications Office of the European Union, 2016. Elérés: 2023. november 2. [Online]. Elérhető: <https://data.europa.eu/doi/10.2861/947880>
- [68] N. Bery, „Security Threats in Smart Factories and Industry 4.0”, *Interconnections - The Equinix Blog*. Elérés: 2024. február 24. [Online]. Elérhető: <https://blog.equinix.com/blog/2023/03/14/security-threats-in-smart-factories-and-industry-4-0/>
- [69] O. Dürr, B. Sick, és E. Murina, *Probabilistic deep learning: with Python, Keras and TensorFlow Probability*. in Exercises in Jupyter Notebooks. Shelter Island: Manning, 2020. Elérés: 2024. május 15. [Online]. Elérhető: <https://freecontent.manning.com/neural-network-architectures/>
- [70] P. Chowdhury, „Perceptron: A Simple yet Mighty Machine Learning Algorithm”, *Medium*. Elérés: 2024. január 13. [Online]. Elérhető: <https://medium.com/@cprajnit32/perceptron-a-simple-yet-mighty-machine-learning-algorithm-9ff6b7d86a71>
- [71] E. Dudyrev, I. Semenov, S. O. Kuznetsov, G. Gusev, A. Sharp, és O. S. Panykh, „Human knowledge models: Learning applied knowledge from the data”, *PLOS ONE*, köt. 17, sz. 10, o. e0275814, okt. 2022, doi: 10.1371/journal.pone.0275814.
- [72] X. Yu, Q. Zhou, S. Wang, és Y. Zhang, „A systematic survey of deep learning in breast cancer”, *International Journal of Intelligent Systems*, köt. 37, aug. 2021, doi: 10.1002/int.22622.
- [73] F. M. Shiri, T. Perumal, N. Mustapha, és R. Mohamed, „A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU”, *arXiv.org*. Elérés: 2024. július 24. [Online]. Elérhető: <https://arxiv.org/abs/2305.17473v2>
- [74] A. Chakure, „Convolutional Neural Networks (CNN) in a Brief”, *DEV Community*. Elérés: 2024. április 30. [Online]. Elérhető: <https://dev.to/afrozchakure/cnn-in-a-brief-27gg>
- [75] D. Nelson, „Mik azok a CNN-ek (konvolúciós neurális hálózatok)? - Egyesüljetek.AI”, *UNITE AI*. Elérés: 2024. április 30. [Online]. Elérhető: <https://www.unite.ai/hu/mik-azok-a-konvol%C3%BAci%C3%B3s-neur%C3%A1lis-h%C3%A1l%C3%B3zatok/>
- [76] L. Fazekas, „TensorFlow alapozó”, *Medium*. Elérés: 2024. április 30. [Online]. Elérhető: <https://thebojda.medium.com/tensorflow-alapoz%C3%B3-d2d1ee97c9db>
- [77] I. Gogul és S. Kumar, „Flower species recognition system using convolution neural networks and transfer learning”, *International Conference on Signal Processing, Communications and Networking*, o. 1–6, márc. 2017, doi: 10.1109/ICSCN.2017.8085675.
- [78] H. Bhat, „Recurrent Neural Network: Applications and Advancements”, *AlmaBetter*. Elérés: 2024. január 14. [Online]. Elérhető: <https://www.almabetter.com/bytes/articles/recurrent-neural-network>
- [79] Rác B. és Csapó T. G., „Ajakvideó alapú beszédszintézis konvolúciós és rekurrens mély neurális hálózatokkal”, *Beszédtudomány - Speech Science*, köt. 1, o. 57–72, dec. 2020, doi: 10.15775/Besztud.2020.57-72.
- [80] H. Prodip és C. Sayan, „A Comprehensive Guide to Attention Mechanism in Deep Learning for Everyone”, *Analytics Vidhya*. Elérés: 2024. január 14. [Online]. Elérhető: <https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/>

- [81] D. Bahdanau, K. Cho, és Y. Bengio, „Neural Machine Translation by Jointly Learning to Align and Translate”, 2016. május 19., *arXiv*: arXiv:1409.0473. doi: 10.48550/arXiv.1409.0473.
- [82] P. Lippe, „6. oktatóanyag: Transzformátorok és többfejes figyelem – UVA DL notebookok v1.2 dokumentációja”, in *UVA Deep Learning Tutorials*, UvA., 2022. Elérés: 2024. május 7. [Online]. Elérhető: https://uvadlc-notebooks.readthedocs.io/en/latest/tutorial_notebooks/tutorial6/Transformers_and_MHAttention.html
- [83] M. Inuwa, „Understanding Attention Mechanisms Using Multi-Head Attention”, *Analytics Vidhya*. Elérés: 2024. május 7. [Online]. Elérhető: <https://www.analyticsvidhya.com/blog/2023/06/understanding-attention-mechanisms-using-multi-head-attention/>
- [84] A. Vaswani és mtsai., „Attention Is All You Need”, *31st Conference on Neural Information Processing Systems, Long Beach*, aug. 2023, doi: 10.48550/arXiv.1706.03762.
- [85] A. Kesrwan, „Multi-Head Self Attention: Short Understanding”, *Medium*. Elérés: 2024. január 14. [Online]. Elérhető: <https://medium.com/@akash.kesrwan99/multi-head-self-attention-short-understanding-e90a34866730>
- [86] N. Daniel, „What is a Generative Adversarial Network (GAN)?”, *Unite AI*, 0 2020, Elérés: 2024. január 14. [Online]. Elérhető: <https://www.unite.ai/what-is-a-generative-adversarial-network-gan/>
- [87] R. I. Muhamedyev, „Machine learning methods: An overview”, *COMPUTER MODELLING & NEW TECHNOLOGIES*, sz. 19, o. 14–29, 2015.
- [88] H. Sajid, „Gépi tanulás vs. mély tanulás – Főbb különbségek – Unite.AI”. Elérés: 2024. január 10. [Online]. Elérhető: <https://www.unite.ai/machine-learning-vs-deep-learning-key-differences/>
- [89] K. Ovtcharov, O. Ruwase, J.-Y. Kim, J. Fowers, K. Strauss, és E. Chung, „Accelerating Deep Convolutional Neural Networks Using Specialized Hardware”, febr. 2015, Elérés: 2025. január 8. [Online]. Elérhető: <https://www.microsoft.com/en-us/research/publication/accelerating-deep-convolutional-neural-networks-using-specialized-hardware/>
- [90] J. Schneider, „FPGA vs. GPU for Deep Learning Applications | IBM”, *IBM Blog*. Elérés: 2025. január 8. [Online]. Elérhető: <https://www.ibm.com/think/topics/fpga-vs-gpu>
- [91] P. Kennedy, „SambaNova SN10 RDU at Hot Chips 33”, *ServeTheHome*. Elérés: 2025. január 2. [Online]. Elérhető: <https://www.servethehome.com/sambanova-sn10-rdu-at-hot-chips-33/>
- [92] M. Davies, „Intel Labs’ new Loihi 2 research chip outperforms its predecessor by up to 10x and comes with an open-source, community-driven neuromorphic computing framework”, *Intel Technology Brief*, 2021, Elérés: 2024. január 6. [Online]. Elérhető: <https://download.intel.com/newsroom/2021/new-technologies/intel-labs-loihi-2-lava-quotes.pdf>
- [93] H. Kousi, „GPU, LPU and NPU: What are these architectures? - DataNorth”, <https://datanorth.ai>. Elérés: 2025. január 2. [Online]. Elérhető: <https://datanorth.ai/blog/gpu-lpu-npu-architectures>
- [94] Asztalos O., „AI-ra gyúr a legújabb Huawei Kirin processzor”, *HWSW*. Elérés: 2025. január 2. [Online]. Elérhető: <https://www.hwsz.hu/hirek/57742/huawei-hisilicon-kirin-970-processzor-soc-rendszerchip-npu-okostelefon-mi-ai.html>
- [95] A. 800 S. Huawei, „NPU Board Components - NPU Board Components - Overview”. Elérés: 2025. január 2. [Online]. Elérhető: <https://support.huawei.com/enterprise/en/doc/EDOC1100149965/d0bd4bfe/npu-board-components>
- [96] H.-T. Zhang és mtsai., „Reconfigurable perovskite nickelate electronics for artificial intelligence”, *Science (New York, N.Y.)*, köt. 375, sz. 6580, o. 533–539, 2022, doi: 10.1126/science.abj7943.
- [97] I. Dienes és Stratégiai Kutató Intézet, „Mesterséges intelligencia másképpen: kvantumszámítógépek és a topologikus energia”, *ÁT - Áram és Technológia*, sz. III. évfolyam 1. szám, 2005, [Online]. Elérhető: http://www.metaelmelet.hu/pdfek/mesterseges_intelligencia_maskeppen.pdf
- [98] D. Peral-García, J. Cruz-Benito, és F. J. García-Peñalvo, „Systematic literature review: Quantum machine learning and its applications”, *Computer Science Review*, köt. 51, o. 100619, febr. 2024, doi: 10.1016/j.cosrev.2024.100619.

- [99] S. Jerbi, C. Gyurik, S. C. Marshall, R. Molteni, és V. Dunjko, „Shadows of quantum machine learning”, *Nat Commun*, köt. 15, sz. 1, o. 5676, júl. 2024, doi: 10.1038/s41467-024-49877-8.
- [100] M. Swayne, „2025 Expert Quantum Predictions — Quantum Computing”, *The Quantum Insider*. Elérés: 2025. január 8. [Online]. Elérhető: <https://thequantuminsider.com/2024/12/31/2025-expert-quantum-predictions-quantum-computing/>
- [101] A. Park, B. Wu, és L. G. Griffith, „Integration of surface modification and 3D fabrication techniques to prepare patterned poly(L-lactide) substrates allowing regionally selective cell adhesion”, *J Biomater Sci Polym Ed*, köt. 9, sz. 2, o. 89–110, 1998, doi: 10.1163/156856298x00451.
- [102] Lena Smirnova, Thomas Hartung, és I. E. M. Pantoja, „Brain-Cell Cultures: The Future of Computers and More?”, *Frontiers for Young Minds*. Elérés: 2024. február 25. [Online]. Elérhető: <https://kids.frontiersin.org/articles/10.3389/frym.2023.1049593>
- [103] D. Bennet, T. Vo-Dinh, és F. Zenhausern, „Current and emerging opportunities in biological medium-based computing and digital data storage”, *Nano Select*, köt. 3, sz. 5, o. 883–902, 2022, doi: 10.1002/nano.202100275.
- [104] J. Hamzelou, „A memory prosthesis could restore memory in people with damaged brains”, *MIT Technology Review*. Elérés: 2024. február 25. [Online]. Elérhető: <https://www.technologyreview.com/2022/09/06/1059032/memory-prosthesis-damaged-brains/>
- [105] Oláh F., „A fuzzy logika - alapismeretek”, *autotechnika.hu*. Elérés: 2024. május 9. [Online]. Elérhető: <https://autotechnika.hu/cikkek/motor-eroatvitel/8313/a-fuzzy-logika-alapismeretek>
- [106] L. A. Zadeh, „Outline of a New Approach to the Analysis of Complex Systems and Decision Processes”, *IEEE Transactions on Systems, Man, and Cybernetics*, köt. SMC-3, sz. 1, o. 28–44, 1973, doi: 10.1109/TSMC.1973.5408575.
- [107] MATLAB center, „Defuzzification Methods”. Elérés: 2024. december 9. [Online]. Elérhető: <https://www.mathworks.com/help/fuzzy/defuzzification-methods.html>
- [108] L. A. Zadeh, „Fuzzy sets”, *Information and Control*, köt. 8, sz. 3, o. 338–353, jún. 1965, doi: 10.1016/S0019-9958(65)90241-X.
- [109] A. Garrido, „A Brief History of Fuzzy Logic”, *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, köt. 3, sz. 1, Art. sz. 1, 2012.
- [110] B. Pease, „What’s All This Fuzzy Logic Stuff, Anyway?”, *Electronic Design*. Elérés: 2024. december 9. [Online]. Elérhető: <https://www.electronicdesign.com/technologies/embedded/digital-ics/article/21757343/whats-all-this-fuzzy-logic-stuff-anyhow>
- [111] P. Cintula, C. G. Fermüller, és C. Noguera, „Fuzzy Logic”, in *The Stanford Encyclopedia of Philosophy*, Summer 2023., E. N. Zalta és U. Nodelman, Szerk., Metaphysics Research Lab, Stanford University, 2023. Elérés: 2024. december 9. [Online]. Elérhető: <https://plato.stanford.edu/archives/sum2023/entries/logic-fuzzy/>
- [112] A. Simran Kaur, „What is Fuzzy Logic? Advantages and Disadvantages”, *Hackr.io*. Elérés: 2024. december 10. [Online]. Elérhető: <https://hackr.io/blog/what-is-fuzzy-logic>
- [113] N. V. Otten, „Fuzzy Logic Made Easy — Its Application In AI, Machine Learning & Natural Language Processing (NLP)”, *Spot Intelligence*. Elérés: 2024. december 10. [Online]. Elérhető: <https://spotintelligence.com/2023/02/13/fuzzy-logic-ai-machine-learning/>
- [114] iCert Global, „What is Fuzzy Logic in AI and its uses?”, *icertglobal.com*. Elérés: 2024. december 10. [Online]. Elérhető: <https://www.icertglobal.com/what-is-fuzzy-logic-in-ai-and-its-uses-blog/detail>
- [115] E. Benavides, *Fuzzy Ethics: A moral criterion for sustainability*. 2013.
- [116] PentaSchool Oktatási Központ, „A Neumann elv”, *Világhíres Feltalálók*. Elérés: 2024. január 16. [Online]. Elérhető: <http://www.feltalalok.hu/tudosok/neumannjanos/html/neujan-tal2.htm>
- [117] „NEUMANN JÁNOS: A SZÁMOLÓGÉP ÉS AZ AGY”. Elérés: 2024. január 16. [Online]. Elérhető: <https://mek.oszk.hu/01200/01255/html/#02>
- [118] „Bionics”, *Enciclopedia Britannica*. 2022. Elérés: 2024. január 16. [Online]. Elérhető: <https://www.britannica.com/technology/bionics>

- [119] A. Darwish, „Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications”, *Future Computing and Informatics Journal*, köt. 3, sz. 2, o. 231–246, dec. 2018, doi: 10.1016/j.fcij.2018.06.001.
- [120] M. Dorigo, M. Birattari, és T. Stützle, „Ant Colony Optimization”, *Computational Intelligence Magazine, IEEE*, köt. 1, o. 28–39, dec. 2006, doi: 10.1109/MCI.2006.329691.
- [121] I. Zelinka, S. Das, L. Sikora, és R. Šenkeřík, „Swarm virus - Next-generation virus and antivirus paradigm?”, *Swarm and Evolutionary Computation*, köt. 43, o. 207–224, dec. 2018, doi: 10.1016/j.swevo.2018.05.003.
- [122] D. Karaboga és B. Akay, „Akay, B.: A Survey: Algorithms Simulating Bee Swarm Intelligence. Artificial Intelligence Review 31, 68-85”, *Artif. Intell. Rev.*, köt. 31, o. 61–85, jún. 2009, doi: 10.1007/s10462-009-9127-4.
- [123] K.-C. Lin, S.-Y. Chen, és J. Hung, „Botnet Detection Using Support Vector Machines with Artificial Fish Swarm Algorithm”, *Journal of Applied Mathematics*, köt. 2014, o. 1–9, ápr. 2014, doi: 10.1155/2014/986428.
- [124] I. Fister, I. Fister, X.-S. Yang, és J. Brest, „A comprehensive review of firefly algorithms”, *Swarm and Evolutionary Computation*, köt. 13, o. 34–46, dec. 2013, doi: 10.1016/j.swevo.2013.06.001.
- [125] A. Bohre, G. Agnihotri, és M. Dubey, „The Butterfly-Particle Swarm Optimization (Butterfly-PSO/BF-PSO) Technique and Its Variables”, *International Journal of Soft Computing, Mathematics and Control*, köt. 4, aug. 2015, doi: 10.14810/ijscmc.2015.4302.
- [126] X.-S. Yang, *Nature-inspired Metaheuristic Algorithms*, 2. kiadás. Luniver Press, 2010. [Online]. Elérhető: https://staff.fmi.uvt.ro/~daniela.zaharie/ma2016/projects/techniques/FireflyAlgorithm/Yang_nature_book_part.pdf
- [127] H. Faris, I. Aljarah, M. Al-Betar, és S. Mirjalili, „Grey wolf optimizer: a review of recent variants and applications”, *Neural Computing and Applications*, köt. 10, o. 1–24, 2017.
- [128] Z.-G. Du, J.-S. Pan, S.-C. Chu, és Y.-J. Chiu, „Improved Binary Symbiotic Organism Search Algorithm With Transfer Functions for Feature Selection”, *IEEE Access*, köt. 8, o. 225730–225744, 2020, doi: 10.1109/ACCESS.2020.3045043.
- [129] H. L. Kwa, J. L. Kit, N. Horsevad, J. Philippot, M. Savari, és R. Bouffanais, „Adaptivity: a path towards general swarm intelligence?”, *Frontiers in Robotics and AI*, köt. 10, 2023, doi: 10.3389/frobt.2023.1163185.
- [130] A. T. Fehér és I. Kereszthegyí, „A gépi nyelvfeldolgozás védelmi felhasználási lehetőségei és kockázata”, *Szakmai Szemle*, köt. XXI., sz. 4., o. 75–93, 0 2023.
- [131] Microsoft ügyfélszolgálat, „A Microsoft 365 Copilot áttekintése”. Elérés: 2023. november 16. [Online]. Elérhető: <https://learn.microsoft.com/hu-hu/microsoft-365-copilot/microsoft-365-copilot-overview>
- [132] K. R. Chowdhary, „Natural Language Processing”, in *Fundamentals of Artificial Intelligence*, K. R. Chowdhary, Szerk., New Delhi: Springer India, 2020, o. 603–649. doi: 10.1007/978-81-322-3972-7_19.
- [133] Váradi T. és *mtsai.*, „Az e-magyar digitális nyelvfeldolgozó rendszer”, Vincze V., Szerk., Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport, 2017, o. 49–60. Elérés: 2023. november 24. [Online]. Elérhető: <http://real.mtak.hu/72361/>
- [134] G. Kumar, „Role of Tokenization in NLP”, Medium. Elérés: 2023. november 24. [Online]. Elérhető: <https://kmr-gautam2893.medium.com/role-of-tokenization-in-nlp-18057618a102>
- [135] L. Stoy, „What is an AI winter and is one coming?”, *Search Engine Land*, szept. 2024, Elérés: 2024. október 1. [Online]. Elérhető: <https://searchengineland.com/ai-winter-is-coming-446295>
- [136] J. Harguess és C. M. Ward, „Is the Next Winter Coming for AI? Elements of Making Secure and Robust AI”, in *2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2022, o. 1–7. doi: 10.1109/AIPR57179.2022.10092230.
- [137] S. Schuchmann, „Analyzing the Prospect of an Approaching AI Winter”, 2019. doi: 10.13140/RG.2.2.10932.91524.
- [138] A. Toosi, A. Bottino, B. Saboury, és A. Rahmim, „A Brief History of AI: How to Prevent Another Winter (A Critical Review)”, *PET Clinics*, köt. 16, szept. 2021, doi: 10.1016/j.cpet.2021.07.001.

- [139] E. Feigenbaum, B. Buchanan, és J. Lederberg, „On generality and problem solving: A case study using the DENDRAL program”, *Machine Intelligence*, köt. 6, szept. 1970.
- [140] A. Kaplan és M. Haenlein, „Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence”, *Business Horizons*, köt. 62, nov. 2018, doi: 10.1016/j.bushor.2018.08.004.
- [141] N. J. Nilsson, *The Quest for Artificial Intelligence*, 1. kiad. Cambridge University Press, 2009. doi: 10.1017/CBO9780511819346.
- [142] K. Fukushima, „Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”, *Biol. Cybernetics*, köt. 36, sz. 4, o. 193–202, ápr. 1980, doi: 10.1007/BF00344251.
- [143] Gartner Institute, „Understanding Gartner’s Hype Cycles”, Gartner. Elérés: 2024. október 8. [Online]. Elérhető: <https://www.gartner.com/en/documents/3887767>
- [144] M. L. Minsky és S. Papert, *Perceptrons, An Essay on Computational Geometry*. Cambridge, MA: MIT Press, 1969.
- [145] Fehér A. T., „A mesterségesintelligencia-alapú hidegháború etikai háttere”, *A mesterséges intelligencia és egyéb felforgató technológiák hatásainak átfogó vizsgálata*, o. 355–392, okt. 2023.
- [146] M. Haenlein és A. Kaplan, „A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence”, *California Management Review*, köt. 61, sz. 4, o. 5–14, aug. 2019, doi: 10.1177/0008125619864925.
- [147] Z. C. Lipton és J. Steinhardt, „Troubling Trends in Machine Learning Scholarship: Some ML papers suffer from flaws that could mislead the public and stymie future research.”, *Queue*, köt. 17, sz. 1, o. Pages 80:45–Pages 80:77, 0 2019, doi: 10.1145/3317287.3328534.
- [148] W. Knight, „AI Winter Isn’t Coming”, MIT Technology Review. Elérés: 2024. szeptember 30. [Online]. Elérhető: <https://www.technologyreview.com/2016/12/07/155592/ai-winter-isnt-coming/>
- [149] Csepeli G., *Ember 2.0 - A mesterséges intelligencia gazdasági és társadalmi hatásai*. Budapest: Kossuth Kiadó Zrt., Kossuth Kiadó Zrt.
- [150] A. R. Buck, „The Way of the Future is now a thing of the past”, European Academy on Religion and Society. Elérés: 2024. november 9. [Online]. Elérhető: <https://europeanacademyofreligionandsociety.com/news/the-way-of-the-future-is-now-a-thing-of-the-past/>
- [151] G. Prisco, „About Turing Church”, Medium. Elérés: 2024. november 11. [Online]. Elérhető: <https://turingchurch.net/about-turing-church-ac6ebf2e97b6>
- [152] R. Leenes, E. Palmerini, B.-J. Koops, A. Bertolini, P. Salvini, és F. Lucivero, „Regulatory challenges of robotics: Some guidelines for addressing legal and ethical issues”, *Law, Innovation and Technology*, köt. 9, o. 1–44, márc. 2017, doi: 10.1080/17579961.2017.1304921.
- [153] L. Lessig, „The Law of the Horse: What Cyberlaw Might Teach”, *Harvard Law Review*, köt. 113, sz. 2, o. 501–549, 1999, doi: 10.2307/1342331.
- [154] L. Rhue, „Beauty’s in the AI of the Beholder: How AI Anchors Subjective and Objective Predictions”, *ICIS 2019 Proceedings*, 2019, [Online]. Elérhető: https://aisel.aisnet.org/icis2019/future_of_work/future_work/15
- [155] T. Rahwan, J. Crandall, és F. Ishowo-Oloko, „Ethics, efficiency, and artificial intelligence - The Boston Globe”, BostonGlobe.com. Elérés: 2022. december 21. [Online]. Elérhető: <https://www.bostonglobe.com/2020/01/30/opinion/ethics-efficiency-artificial-intelligence/>
- [156] A. T. Fehér, „Az MI–n innen és túl”, Ludovika.hu. Elérés: 2024. október 16. [Online]. Elérhető: <https://www.ludovika.hu/blogok/kormblog/2024/06/03/az-mi-n-innen-es-tul/>
- [157] J. Vincent, „Forty percent of “AI startups” in Europe don’t actually use AI, claims report”, The Verge. Elérés: 2024. október 16. [Online]. Elérhető: <https://www.theverge.com/2019/3/5/18251326/ai-startups-europe-fake-40-percent-mmcc-report>
- [158] A. T. Fehér, „Okosdolgok egy buta korban”, Ludovika.hu. Elérés: 2024. október 16. [Online]. Elérhető: <https://www.ludovika.hu/blogok/kormblog/2024/07/17/okosdolgok-egy-buta-korban/>

- [159] K. Kansky és mtsai., „Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics”, 2017. augusztus 17., *arXiv*: arXiv:1706.04317. doi: 10.48550/arXiv.1706.04317.
- [160] F. Chollet, *Deep learning with Python*. Shelter Island, NY: Manning, 2018. [Online]. Elérhető: <https://blog.keras.io/the-limitations-of-deep-learning.html>
- [161] A. T. Fehér és I. Kereszthegyi, „Az NLP kínálta lehetőségek védelmi kockázatai és felhasználási lehetőségei”, Nemzeti Közzolgálati Egyetem.
- [162] F. Webber, „Third AI Winter ahead? Why OpenAI, Google & Co are heading towards a dead-end | Cortical.io”. Elérés: 2024. október 27. [Online]. Elérhető: <https://www.cortical.io/blog/third-ai-winter-ahead-why-openai-google-co-are-heading-towards-a-dead-end/>
- [163] M. Coolidge, „The Danger of AI’s Black Box | Casper Labs”. Elérés: 2024. október 17. [Online]. Elérhető: <https://casperlabs.io/blog/the-danger-of-ais-black-box>
- [164] U. Bhatt és mtsai., „Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty”, in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, in AIES ’21. New York, NY, USA: Association for Computing Machinery, 0 2021, o. 401–413. doi: 10.1145/3461702.3462571.
- [165] N. Stogiannos és mtsai., „Black box no more: A cross-sectional multi-disciplinary survey for exploring governance and guiding adoption of AI in medical imaging and radiotherapy in the UK”, *International Journal of Medical Informatics*, köt. 186, o. 105423, jún. 2024, doi: 10.1016/j.ij-medinf.2024.105423.
- [166] N. Balogh, „XAI, avagy a magyarázható mesterséges intelligencia”, Dmlab. Elérés: 2024. október 17. [Online]. Elérhető: <https://dmlab.hu/blog/xai-avagy-a-magyarázható-mesterseges-intelligencia/>
- [167] I. J. Goodfellow, J. Shlens, és C. Szegedy, „Explaining and Harnessing Adversarial Examples”, 2015. március 20., *arXiv*: arXiv:1412.6572. doi: 10.48550/arXiv.1412.6572.
- [168] Fehér A. T. és Négyesi I., „A mesterséges intelligencia alapú kibertámadások lehetőségei”, *Studia Mundi – Economica*, köt. 8, sz. 5, Art. sz. 5, dec. 2021, doi: 10.18531/Studia.Mundi.2021.08.05.25-34.
- [169] Fehér A. T. és Négyesi I., „Mesterségesintelligencia-alapú kibertámadási modellek”, *MKK*, köt. 31, sz. 3, o. 73–87, 2021, doi: 10.32562/mkk.2021.3.5.
- [170] L. Floridi, „AI and Its New Winter: from Myths to Realities”, *Philos. Technol.*, köt. 33, sz. 1, o. 1–3, márc. 2020, doi: 10.1007/s13347-020-00396-6.
- [171] B. Boncz és R. Szabo, „A mesterséges intelligencia munkaerő-piaci hatásai Hogyan készülünk fel?”, *Vezetéstudomány / Budapest Management Review*, köt. 53, o. 68–80, febr. 2022, doi: 10.14267/VEZTUD.2022.02.06.
- [172] T. Kyriazos és M. Poga, „Quantum concepts in Psychology: Exploring the interplay of physics and the human psyche”, *BioSystems*, köt. 235, o. 105070, jan. 2024, doi: 10.1016/j.bio-systems.2023.105070.
- [173] Wilkinson T., „Versenyfutás az egyiptomi hieroglifák megfejtéséért – BBC History”, *BBC History*, köt. 2023/9, aug. 2023, Elérés: 2024. november 20. [Online]. Elérhető: <https://bbc-history.hu/egyiptomi-hieroglifak/>
- [174] A. B. Salmina és mtsai., „Current progress and challenges in the development of brain tissue models: How to grow up the changeable brain in vitro?”, *J Tissue Eng*, köt. 15, o. 20417314241235527, jan. 2024, doi: 10.1177/20417314241235527.
- [175] H. Li, X. Li, és J. del R. Millán, „Noninvasive EEG-Based Intelligent Mobile Robots: A Systematic Review”, *IEEE Transactions on Automation Science and Engineering*, o. 1–25, 2024, doi: 10.1109/TASE.2024.3441055.
- [176] S. L. Metzger és mtsai., „Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis”, *Nat Commun*, köt. 13, sz. 1, o. 6510, nov. 2022, doi: 10.1038/s41467-022-33611-3.

- [177] Y. Wang, M. Zhang, M. Li, H. Cui, és X. Chen, „Development of a humanoid robot control system based on AR-BCI and SLAM navigation”, *Cogn Neurodyn*, máj. 2024, doi: 10.1007/s11571-024-10122-z.
- [178] „26.2. Erős MI: Tudnak-e ténylegesen gondolkodni a gépek? | Mesterséges Intelligencia Elektronikus Almanach”. Elérés: 2022. december 16. [Online]. Elérhető: http://project.mit.bme.hu/mi_almanach/books/aima/ch26s02
- [179] P. Wang, „On Defining Artificial Intelligence”, *Journal of Artificial General Intelligence*, köt. 10, sz. 2, o. 1–37, jan. 2019.
- [180] M. T. B. and E. Perrier, „OpenAI Claims Its New Model Reached Human Level on a Test for ‘General Intelligence.’ What Does That Mean?”, Gizmodo. Elérés: 2025. január 25. [Online]. Elérhető: <https://gizmodo.com/openai-claims-its-new-model-reached-human-level-on-a-test-for-general-intelligence-what-does-that-mean-2000543834>
- [181] T. Maxwell, „Leaked Documents Show OpenAI Has a Very Clear Definition of ‘AGI’”, Gizmodo. Elérés: 2025. január 25. [Online]. Elérhető: <https://gizmodo.com/leaked-documents-show-openai-has-a-very-clear-definition-of-agi-2000543339>
- [182] Á. Bobák, „Mai fejjel szinte elképzelhetetlen átalakulást jósol a következő évtizedre az egyik legismertebb jövőkutató - Raketa.hu”, 2024, Elérés: 2024. október 14. [Online]. Elérhető: <https://raketa.hu/mai-fejjel-szinte-elkepzehetetlen-atalakulast-josol-a-kovetkezo-evtizedre-az-egyik-legismertebb-jovokutato>
- [183] I. Négyesi, „A mesterséges intelligencia és az etika”, *Hadtudomány*, köt. 30, o. 103-113., 2020, doi: 10.17047/HADTUD.2020.30.1.103.
- [184] Europe Parliament és S. C. U. University of the West of England, *The ethics of artificial intelligence: Issues and initiatives : study Panel for the Future of Science and Technology*. Brussels: European Parliamentary Research Service, Scientific Foresight Unit, 2020. doi: 10.2861/6644.
- [185] S. Darwall, „The Value of Autonomy and Autonomy of the Will”, *Ethics*, köt. 116, o. 263–284, 2006.
- [186] C. Breazeal és B. Scassellati, „Challenges in Building Robots That Imitate People”, *Imitation in Animals and Artifacts*, okt. 2000.
- [187] Váróné Tomori V., „Kurt Lewin és a mezőelmélet”, *Korunk*, sz. 32. évf. 2. sz. (1973. február), o. 312–316, 1973.
- [188] M. R. ENDSLEY és D. B. KABER, „Level of automation effects on performance, situation awareness and workload in a dynamic control task”, *Ergonomics*, köt. 42, sz. 3, o. 462–492, márc. 1999, doi: 10.1080/001401399185595.
- [189] J3016 szabvány (2021.04.), *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International*. Elérés: 2024. július 12. [Online]. Elérhető: https://www.sae.org/standards/content/j3016_202104/
- [190] Zsolt S., „Ezek az önvezetés szintjei”, Villanyautósok. Elérés: 2024. július 12. [Online]. Elérhető: <https://villanyautosok.hu/2021/05/11/ezek-az-onvezetes-szintjei/>
- [191] C&T, „The 6 Levels of Autonomous Driving Explained”, C&T Solution Inc. Elérés: 2024. július 12. [Online]. Elérhető: https://www.candtsolution.com/news_events-detail/the-six-level-of-autonomous-driving-explained/
- [192] J. Beer, A. Fisk, és W. Rogers, „Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction”, *Journal of Human-Robot Interaction*, köt. 3, o. 74, jún. 2014, doi: 10.5898/JHRI.3.2.Beer.
- [193] M. Krä, L. Eckart, M. Rito, és J. Schilp, „Levels of autonomy in production logistics: terminology and framework”, 2022, Elérés: 2024. július 12. [Online]. Elérhető: <https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docid/96145>
- [194] A. Attanasio, B. Scaglioni, E. D. Momi, P. Fiorini, és P. Valdastrì, „Autonomy in Surgical Robotics”, *Annual Review of Control, Robotics, and Autonomous Systems*, köt. 4, sz. Volume 4, 2021, o. 651–679, máj. 2021, doi: 10.1146/annurev-control-062420-090543.

- [195] R. „Ray” Wang, „Five Levels That Will Define the Future of Autonomous Enterprises”, IBM Blog. Elérés: 2024. január 10. [Online]. Elérhető: <https://www.ibm.com/blog/five-levels-that-will-define-the-future-of-autonomous-enterprises/>
- [196] T. Mickle, C. Metz, és Y. Lu, „G.M.’s Cruise Moved Fast in the Driverless Race. It Got Ugly.”, *The New York Times*, 2023. november 3. Elérés: 2024. augusztus 27. [Online]. Elérhető: <https://www.nytimes.com/2023/11/03/technology/cruise-general-motors-self-driving-cars.html>
- [197] Viktor P.-P., „Emberek vezették a Cruise robotaxijait”. Elérés: 2024. augusztus 27. [Online]. Elérhető: <https://index.hu/techtud/2023/11/12/robotaxi-onvezeto-egyedul-allamok-san-francisco-general-motors-cruise-origin-felfuggesztes-godrok-gyerekek/>
- [198] I. Asimov, *The Bicentennial Man and Other Stories*. in Doubleday science fiction. Doubleday, 1976. [Online]. Elérhető: <https://books.google.hu/books?id=w0daAAAAMAAJ>
- [199] IBM Support, „Maximo for Civil Infrastructure 7.6.2”. Elérés: 2024. augusztus 21. [Online]. Elérhető: <https://www.ibm.com/docs/hu/mfci/7.6.2?topic=implementing-highly-available-systems>
- [200] G. Vincze, „Az erkölcsi ítéletalkotás fejlődése és fejlesztési lehetőségei intézményes keretek között”, *GYNTF*, köt. 8, sz. 3, o. 72–92, 2020, doi: 10.31074/gyntf.2020.3.72.92.
- [201] L. P. Nucci és E. Turiel, „Social Interactions and the Development of Social Concepts in Preschool Children”, *Child Development*, köt. 49, sz. 2, o. 400, 1978, doi: 10.2307/1128704.
- [202] Z. Bugnyár, „Erkölc, egyén, közösség”, *Új Pedagógiai Szemle*, 0 2009, Elérés: 2022. január 2. [Online]. Elérhető: <https://ofi.oh.gov.hu/tudastar/erkolcs-egyen-kozosseg>
- [203] E. Turiel és K. Banas, „The Development of Moral and Social Judgments: Social Contexts and Processes of Coordination”, *Eurasian Journal of Educational Research*, köt. 85, 2020, [Online]. Elérhető: <https://files.eric.ed.gov/fulltext/EJ1242437.pdf>
- [204] J. Szekszárdi, „Piaget és Kohlberg nyomában I. rész”, *Új Pedagógiai Szemle*, köt. április, 2000, Elérés: 2022. január 3. [Online]. Elérhető: <https://ofi.oh.gov.hu/tudastar/piaget-kohlberg-nyomaban>
- [205] A. Paiva, I. Leite, H. Boukricha, és I. Wachsmuth, „Empathy in Virtual Agents and Robots”, *ACM Trans. Interact. Intell. Syst.*, köt. 7, sz. 3, o. 1–40, 2017, doi: 10.1145/2912150.
- [206] K. Dautenhahn, „Getting to know each other—Artificial social intelligence for autonomous robots”, *Robotics and Autonomous Systems*, köt. 16, sz. 2–4, o. 333–356, 1995, doi: 10.1016/0921-8890(95)00054-2.
- [207] H. Moravec, „When will computer hardware match the human brain”, *Journal of Evolution and Technology*, köt. 1, 1998, [Online]. Elérhető: <https://jetpress.org/volume1/moravec.htm>
- [208] N. Reed, „What is Moravec’s paradox and what does it mean for modern AI?”, ThinkAutomation. Elérés: 2022. december 16. [Online]. Elérhető: <https://www.thinkautomation.com/bots-and-ai/what-is-moravecs-paradox-and-what-does-it-mean-for-modern-ai/>
- [209] B. Komenczi, „Komenczi Bertalan blogoldala » John Searle elmefilozófiai gondolat kísérlete”, Az egri Eszterházy Károly Egyetem professzor emeritusának blogja. Elérés: 2022. december 6. [Online]. Elérhető: <http://komenczi.ektf.hu/john-searle-elmefilozofiai-gondolatkiserlete/>
- [210] L. Hauser, „Searle’s Chinese Box: Debunking the Chinese Room Argument”, *Minds and Machines*, köt. 7, sz. 2, o. 199–226, 1997, doi: 10.1023/A:1008255830248.
- [211] Radó N., „Mérő László: Ha az önvezető autóm versekről diskurálna, vinném a bontóba”, *Qubit*, 2020. október 28. Elérés: 2022. december 17. [Online]. Elérhető: <https://qubit.hu/2020/10/28/mero-laszlo-ha-az-onvezeto-autom-versekrol-diskuralna-vinnem-a-bontoba>
- [212] S. Skwiot, „How AI & AR Technologies Are Changing the Beauty Industry”. Elérés: 2022. január 1. [Online]. Elérhető: <https://www.perfectcorp.com/business/blog/general/how-ai-and-ar-innovation-are-changing-the-beauty-tech-industry>
- [213] J. H. Moor, „The Nature, Importance, and Difficulty of Machine Ethics”, *Intelligent Systems, IEEE*, köt. 21, sz. 4, o. 18–21, 2006, doi: 10.1109/MIS.2006.80.

- [214] N. Adrienn; „A mesterséges intelligencia és a digitalizáció jelentősége és lehetséges hasznosítási területei az igazságszolgáltatásban – 2020/2. (75.), e-különszám”, *Az Infokommunikáció és Jog*, köt. 75, Budapest, 2021. 19. Elérés: 2022. január 2. [Online]. Elérhető: <https://info-jog.hu/dr-nagy-adrienn-a-mesterseges-intelligencia-es-a-digitalizacio-jelentosege-es-lehetseges-hasznositasi-teruletei-az-igazsagszolgalattasban-2020-2-75-e-kulonszam/>
- [215] Luciano Floridi és J.W. Sanders, „On the Morality of Artificial Agents”, *Minds and Machines*, köt. 14, sz. 3, o. 349–379, 2004, doi: 10.1023/B:MIND.0000035461.63578.9d.
- [216] J. L. Wyatt, D. D. Petters, és D. C. Hogg, Szerk., *From Animals to Robots and Back: Reflections on Hard Problems in the Study of Cognition*, köt. 22. in *Cognitive Systems Monographs*, vol. 22. Cham: Springer International Publishing, 2014. doi: 10.1007/978-3-319-06614-1.
- [217] N. T. Lee, P. Resnick, és G. Barton, „Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms”, Brookings, máj. 2019. Elérés: 2022. december 6. [Online]. Elérhető: <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>
- [218] Dervenkár I., „Előítélet-detektálót csinált az IBM mesterséges intelligenciához”, Bitport. Elérés: 2022. december 18. [Online]. Elérhető: <https://bitport.hu/eloitelet-detektalot-csinalt-az-ibm-mesterseges-intelligenciahoz/>
- [219] J. Manyika és J. Silberg, „Tackling bias in artificial intelligence (and in humans)”. McKinsey Global Institute, 2019. [Online]. Elérhető: <https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>
- [220] „Prevision.io — Documentation — Documentation PrevisionIO”. Elérés: 2022. december 18. [Online]. Elérhető: <https://previsionio.readthedocs.io/fr/latest/index.html>
- [221] R. Vaithianathan, E. Kulick, és D. B. Prado, „Allegheny Family Screening Tool: Methodology, Version 2 - Centre for Social Data Analytics - AUT”, Centre for Social Data Analytics, 2019. Elérés: 2022. december 18. [Online]. Elérhető: <https://csda.aut.ac.nz/research/our-publications/2019/allegheny-family-screening-tool-methodology,-version-2>
- [222] B. Marr, „How To Solve AI’s Bias Problem, Create Emotional AIs, And Democratize AI With Synthetic Data”, *Forbes*. Elérés: 2022. december 20. [Online]. Elérhető: <https://www.forbes.com/sites/bernardmarr/2021/05/28/how-to-solve-ais-bias-problem-create-emotional-ais-and-democratize-ai-with-synthetic-data/>
- [223] M. Peychev, A. Ruoss, M. Balunović, M. Baader, és M. Vechev, „Latent Space Smoothing for Individually Fair Representations”, 2022. július 26., *arXiv*: arXiv:2111.13650. Elérés: 2022. december 20. [Online]. Elérhető: <http://arxiv.org/abs/2111.13650>
- [224] J. Koebler, „We Asked MIT Researchers Why They Made a ‘Psychotic AI’ That Only Sees Death”, *Vice*. Elérés: 2022. december 18. [Online]. Elérhető: <https://www.vice.com/en/article/xwm5mk/mit-psychotic-ai-rehabilitation>
- [225] E. Drage és K. Mackereth, „Does AI Debias Recruitment? Race, Gender, and AI’s “Eradication of Difference””, *Philos. Technol.*, köt. 35, sz. 4, o. 89, okt. 2022, doi: 10.1007/s13347-022-00543-1.
- [226] H. Knight, „How Humans Respond to Robots”: o. 20, 2014.
- [227] C. Mims, „Why Japanese Love Robots (And Americans Fear Them)”, *MIT Technology Review*. Elérés: 2022. november 29. [Online]. Elérhető: <https://www.technologyreview.com/2010/10/12/120635/why-japanese-love-robots-and-americans-fear-them/>
- [228] R. Eveleth, „What Golems and Robots Have in Common?” Elérés: 2022. november 29. [Online]. Elérhető: <https://www.lastwordonnothing.com/2017/02/10/what-golems-and-robots-have-in-common/>
- [229] Alexander Serov, *Subjective Reality and Strong Artificial Intelligence*. 2013. [Online]. Elérhető: https://www.researchgate.net/publication/235221450_Subjective_Reality_and_Strong_Artificial_Intelligence
- [230] R. Koster és *mtsai.*, „Human-centred mechanism design with Democratic AI”, *Nat Hum Behav*, köt. 6, sz. 10, Art. sz. 10, okt. 2022, doi: 10.1038/s41562-022-01383-x.

- [231] Bodnár Z., „A Demokratikus AI új módszert talált a közpénzek újraelosztására, és többen szavaztak rá, mint az emberi politikákra”, Qubit. Elérés: 2022. december 7. [Online]. Elérhető: <https://qubit.hu/2022/07/05/a-demokratikus-ai-uj-modszert-talalt-a-kozpenzek-ujraelosztasara-es-tobben-szavaztak-ra-mint-az-emberi-politikakra>
- [232] B. Lutkevich és S. A. Gillis, „What are bots and how do they work?”, WhatIs. Elérés: 2024. március 2. [Online]. Elérhető: <https://www.techtarget.com/whatis/definition/bot-robot>
- [233] H. Alawwad, „An Intelligent Database System using Natural Language Processing”, *International Journal of Computers*, jan. 2016.
- [234] J. Thorne, M. Yazdani, M. Saeidi, F. Silvestri, S. Riedel, és A. Halevy, „From natural language processing to neural databases”, *Proc. VLDB Endow.*, köt. 14, sz. 6, o. 1033–1039, febr. 2021, doi: 10.14778/3447689.3447706.
- [235] A. Shrivastava, „Key-Value Databases are Sufficient Infrastructure for Semantic Search at Scale with NeuralDB.”, ThirdAI Blog. Elérés: 2024. március 12. [Online]. Elérhető: <https://medium.com/thirdai-blog/key-value-databases-are-sufficient-infrastructure-for-semantic-search-at-scale-with-neuraldb-b7eea4b9b4db>
- [236] G. Kovács és E. Tóth, *Az informatika története*. Budapest: NJSZT Informatikatörténeti Fórum, 1991. Elérés: 2023. december 28. [Online]. Elérhető: <https://itf.njszt.hu/objektum/az-informatika-tortenete>
- [237] K. Biener, „Karl Steinbuch – Informatiker der ersten Stunde”, *RZ-Mitteilungen*, sz. 15, o. 53–54, 1997.
- [238] K. Steinbuch, „Informatik: Automatische Informationsverarbeitung”, *SEG-Nachrichten (Technische Mitteilungen der Standard Elektrik Gruppe) – Firmenzeitschrift*, sz. 4, o. 171, 1957.
- [239] S. (ITIV) (inaktiv) Daum, „KIT - ITIV - Institut - Geschichte - Geschichte des ITIV - Karl Steinbuch - Gründer des ITIV”. Elérés: 2023. december 26. [Online]. Elérhető: <https://www.itiv.kit.edu/4787.php>
- [240] „l'équipe Ça m'intéresse” csapat, „Quelle est l'origine du mot informatique ?”, Ça m'intéresse. Elérés: 2023. december 26. [Online]. Elérhető: <https://www.caminteresse.fr/societe/quelle-est-lorigine-du-mot-informatique-11137313/>
- [241] I. Masic, „The Most Influential Scientists in the Development of Biomedical Informatics: Philippe Dreyfus (1925-2018)”, köt. 10, o. 137–137, szept. 2022, doi: 10.5455/ijbh.2022.2.137-137.
- [242] „Dr. Walter F. Bauer”, IT History Society. Elérés: 2023. december 28. [Online]. Elérhető: <https://www.ithistory.org/honor-roll/dr-walter-f-bauer>
- [243] H. Shaoyi és I. (ford) Papp, „Az informatika fogalma”, *TMT*, köt. 21, sz. 2, o. 117–122, 2003.
- [244] F. Pnina és R. Howard, *Social Informatics: Past, Present and Future*. Newcastle: Cambridge Scholars Publishing, 2014. Elérés: 2023. december 26. [Online]. Elérhető: <https://www.cambridge-scholars.com/product/978-1-4438-5576-1>
- [245] „INFORMATICS definition and meaning | Collins English Dictionary”. Elérés: 2023. december 28. [Online]. Elérhető: <https://www.collinsdictionary.com/dictionary/english/informatics>
- [246] Cambridge Dictionary, „informatics”. Elérés: 2023. december 28. [Online]. Elérhető: <https://dictionary.cambridge.org/dictionary/english/informatics>
- [247] S. Munk, „A katonai informatika alapelvei a magyar honvédségben i. (alapok)”, *Hadmérnök*, köt. IV, sz. 3, 2009, Elérés: 2023. december 28. [Online]. Elérhető: http://www.hadmer-nok.hu/2009_3_munk1.php
- [248] „Looking back to the future of Edinburgh’s AI legacy”, The University of Edinburgh. Elérés: 2023. december 28. [Online]. Elérhető: <https://www.ed.ac.uk/bayes/media-centre/news/ai-news/looking-back-to-the-future-of-edinburghs-ai-legacy>
- [249] D. Sisu, „School of Informatics: in memoriam Hugh Christopher Longuet-Higgins”. Elérés: 2023. december 28. [Online]. Elérhető: <https://www.inf.ed.ac.uk/events/christopherlh.html>
- [250] The University of Edinburgh, „What is Informatics?”, The University of Edinburgh. Elérés: 2023. december 28. [Online]. Elérhető: <https://www.ed.ac.uk/informatics/about/what-is-informatics>
- [251] C. E. Shannon, „A Mathematical Theory of Communication”, *Bell System Technical Journal*, köt. 27, sz. 3, o. 379–423, júl. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

- [252] H. Borko, „Information science: What is it?”, *American Documentation*, köt. 19, sz. 1, o. 3–5, 1968, doi: 10.1002/asi.5090190103.
- [253] R. Linares Columbié, „Harold Borko y la Ciencia de la Información”, *Revista Cubana de Información en Ciencias de la Salud*, köt. 27, sz. 3, o. 410–419, szept. 2016.
- [254] Murányi P., „Az információtudomány eredete, fejlődése és kapcsolatai”, *Tudományos és Műszaki Tájékoztatás*, köt. 41, sz. 7–8, Art. sz. 7–8, 1994.
- [255] K. F. Szerkesztősége, „Az információtudomány létrejötte és Horváth Tibor információtudományi eszméinek gyökerei1 | Könyvtári Figyelő”. Elérés: 2023. december 26. [Online]. Elérhető: <http://ki2.oszk.hu/kf/2017/04/az-informaciotudomany-letrejtote-es-horvath-tibor-informaciotudomanyi-eszmeinek-gyokerei1/>
- [256] N. Khanye, „Information science: origin, evolution and relations”, *Conceptions of library and information science: ...*, jan. 1992, Elérés: 2023. december 27. [Online]. Elérhető: https://www.academia.edu/925298/Information_science_origin_evolution_and_relations
- [257] W. G. Stock és M. Stock, *Handbook of information science*. Berlin Boston: De Gruyter Saur, 2013.
- [258] J. Boros, „A kognitív tudomány esélyei”, *Magyar Tudomány*, sz. 2004/11, o. 1269, 2004.
- [259] B. H. FLOWERS, „Lighthill Report: Artificial Intelligence: a paper symposium”, UKRI Science and Technology Facilities Council, 1973. Elérés: 2023. december 28. [Online]. Elérhető: https://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm
- [260] D. Kodaj, „A kognitív tudomány mitikus teleológiája”, *Magyar Tudomány*, sz. 10, o. 1250, 2005.
- [261] R. Collins, „Back to the future: Digital television and convergence in the United Kingdom”, *Telecommunications Policy*, köt. 22, sz. 4, o. 383–396, máj. 1998, doi: 10.1016/S0308-5961(98)00022-6.
- [262] Pléh C., „A tudomány jövője: A kognitív tudomány példája a tudományok tagolódásáról és diverzifikálódásáról”, *Magyar Tudomány*, köt. 168, sz. 9, Art. sz. 9, 2007.
- [263] H. Khan, „Types of AI | Different Types of Artificial Intelligence Systems”, *Fossguru*, köt. 9, o. 50, szept. 2021.
- [264] AI ACT, *Javaslat az európai parlament és a tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról*. 2021. Elérés: 2024. február 4. [Online]. Elérhető: <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX%3A52021PC0206>
- [265] M. Regona, T. Yigitcanlar, B. Xia, és R. Y. M. Li, „Opportunities and Adoption Challenges of AI in the Construction Industry: A PRISMA Review”, *Journal of Open Innovation: Technology, Market, and Complexity*, köt. 8, sz. 1, Art. sz. 1, márc. 2022, doi: 10.3390/joitmc8010045.
- [266] R. Browne, „Italy became the first Western country to ban ChatGPT. Here’s what other countries are doing”, CNBC. Elérés: 2025. február 13. [Online]. Elérhető: <https://www.cnbc.com/2023/04/04/italy-has-banned-chatgpt-heres-what-other-countries-are-doing.html>
- [267] I. Porkoláb, „Hibrid hadviselés: új hadviselési forma, vagy régi ismerős?”, *Hadtudomány*, köt. 25, sz. 3–4, Art. sz. 3–4, 2015.
- [268] Resperger I., *A válságkezelés és a hibrid hadviselés*. Budapest: Dialóg Campus, 2018. Elérés: 2023. február 16. [Online]. Elérhető: <https://tudasportal.uni-nke.hu/xmlui/handle/20.500.12944/12438>
- [269] Kiss Álmos P., „A hibrid hadviselés természetrajza”, *Honvédségi Szemle*, köt. XI, sz. 4, Art. sz. 4., 2019.
- [270] A. T. Fehér és I. Négyesi, „A gépi érzelmek a fegyveres erőknél és az autonóm rendszerekben”, *Hadtudományi Szemle*, köt. XIV, sz. 3, o. 163–176, 2021, doi: 10.32563/hsz.2021.3.12.
- [271] R. Albergotti és D. Harwell, „Apple and Google are building a virus-tracking system. Health officials say it will be practically useless.”, *Washington Post*, 2020. május 15. Elérés: 2023. január 28. [Online]. Elérhető: <https://www.washingtonpost.com/technology/2020/05/15/app-apple-google-virus/>
- [272] F. Mantellassi, „Digital Authoritarianism: How Digital Technologies Can Empower Authoritarianism and Weaken Democracy”, *Geneva Centre for Security Policy*, 2023, Elérés: 2023. március

2. [Online]. Elérhető: <https://www.gcsp.ch/publications/digital-authoritarianism-how-digital-technologies-can-empower-authoritarianism-and>
- [273] Department of Defence Test Method Standard, *MIL-STD-810: Environmental Engineering Considerations and Laboratory Tests*. in US Military. DdD - US Military, 2000. [Online]. Elérhető: <https://www.atec.army.mil/publications/mil-std-810g/mil-std-810g.pdf>
- [274] Z. Krajnc, Szerk., „hadművelet célja”, *Hadtudományi lexikon*. Dialóg Campus, Budapest, 2019. Elérés: 2022. december 23. [Online]. Elérhető: <https://rendeszet.uni-nke.hu/publikaciok/hadtudomanyi-lexikon>
- [275] Kovács L., „Offenzív kiberműveletek 1.: Az offenzív kiberműveletek természete”, *Hadmérnök*, köt. 16, sz. 2, Art. sz. 2, aug. 2021, doi: 10.32567/hm.2021.2.13.
- [276] B. Mihály, „Hybrid War: Theory and Ethics”, *AARMS – Academic and Applied Research in Military and Public Management Science*, köt. 23, sz. 1, Art. sz. 1, ápr. 2024, doi: 10.32565/aarms.2024.1.1.
- [277] „Háború a szürke zónában - Kiss Álmos Péterrel, az MH Transzformációs Parancsnokság Honvéd Tudományos Kutatóhely vezető kutatójával a hibrid hadviselésről beszélgettünk.” Elérés: 2022. november 19. [Online]. Elérhető: <https://honvedelem.hu/hirek/haboru-a-szurke-zonaban.html>
- [278] Somodi Z. és Kiss Á. P., „A hibrid hadviselés fogalmának értelmezése a nemzetközi szakirodalomban”, *Honvédségi Szemle – Hungarian Defence Review*, köt. 147, sz. 6, Art. sz. 6, 2019.
- [279] H. Moravec, *Bodies, Robots, Minds*. Robotics Institute, 1995. Elérés: 2019. június 1. [Online]. Elérhető: <https://frc.ri.cmu.edu/~hpm/project.archive/general.articles/1995/Kunstforum.html>.
- [280] B. Taksás, „Újjászülető hadiipar”, *HADITECHNIKA*, köt. 55, sz. 1, Art. sz. 1, 2021.
- [281] Bihaly B., „A mesterséges intelligencia felhasználása az információs és kiberterműveletekben – az orosz minta”, *Hadmérnök*, köt. 17, sz. 3, Art. sz. 3, nov. 2022, doi: 10.32567/hm.2022.3.7.
- [282] A. Dawson és M. Innes, „How Russia’s internet research agency built its disinformation campaign”, *Political Quarterly*, köt. 90, sz. 2, Art. sz. 2, jún. 2019.
- [283] J. E. Stiglitz, „How the US Could Lose the New Cold War”, Project Syndicate. Elérés: 2022. november 19. [Online]. Elérhető: <https://www.project-syndicate.org/commentary/us-squandering-soft-power-appeal-in-cold-war-with-china-by-joseph-e-stiglitz-2022-06>
- [284] Evelin C., „LAWS - a gyilkos robotok és a nemzetközi jog”, *arsboni*. Elérés: 2022. november 19. [Online]. Elérhető: <https://arsboni.hu/laws-a-gyilkos-robotok-es-a-nemzetkozi-jog/>
- [285] Porkoláb I. és Négyesi I., „A mesterséges intelligencia alkalmazási lehetőségeinek kutatása a haderőben”, *Honvédségi Szemle*, köt. 147, sz. 5., Art. sz. 5., 2019.
- [286] M. van Rijmenam, „What is Generative AI, and How Will It Disrupt Society?”, The Digital Speaker: Futurist, Keynote Speaker, Author. Elérés: 2022. december 20. [Online]. Elérhető: <https://www.thedigitalspeaker.com/what-is-generative-ai-how-disrupt-society/>
- [287] N. A. el Ata, „What is Generative Intelligence?”, URM GROUP. Elérés: 2022. december 20. [Online]. Elérhető: <https://urmgrp.com/what-is-generative-intelligence/>
- [288] D. Cem, „24 Affective Computing (Emotion AI) Applications / Use Cases”. 2020. [Online]. Elérhető: <https://research.aimultiple.com/affective-computing-applications/>
- [289] S. Diederich, M. Janssen-Müller, A. B. Brendel, és S. Morana, „Emulating Empathetic Behavior in Online Service Encounters with Sentiment-Adaptive Responses: Insights from an Experiment with a Conversational Agent”, *ICIS 2019 Proceedings*, 2019, [Online]. Elérhető: https://aisel.aisnet.org/icis2019/smart_service_science/smart_service_science/2
- [290] Defense Advanced Research Projects Agency, *Department of Defense Fiscal Year (FY) 2020 Budget Estimates*. Washington: Defense Advanced Research Projects Agency, 2019.
- [291] S. B. Daily és mtsai., *Emotions and Affect in Human Factors and Human-Computer Interaction: Chapter 9 - Affective Computing: Historical Foundations, Current Applications, and Future Trends*. Cambridge: Academic Press, 2017. doi: 10.1016/B978-0-12-801851-4.00009-4.
- [292] INDEX Holding, „Empath in UAE to Measure Happiness”, *INDEX Holding*, 2018, [Online]. Elérhető: <https://indexholding.ae/empath-in-uae-to-measure-happiness/>

- [293] A. T. Fehér, „Artificial Intelligence in cyberspace - Part 1: Artificial Intelligence based cyber-attack capabilities”, *RED - American Journal Of Research Education And Development*, köt. 2, 2019, Elérés: 2021. január 4. [Online]. Elérhető: <https://pubhtml5.com/lwrb/xykg>
- [294] The Digital Sovereignty Focus Group of the Innovative Digitisation of the Economy Platform, „Digital sovereignty in the context of platform-based ecosystems”, Bundesministerium für Wirtschaft und Energie, 2019. [Online]. Elérhető: https://www.de.digital/DIGITAL/Redaktion/DE/Digital-Gipfel/Download/2019/digital-sovereignty-in-the-context-of-platform-based-ecosystems.pdf?__blob=publicationFile&v=7
- [295] Miniszterelnöki Kabinetiroda, „Nemzeti Digitalizációs Stratégia (2022-2030)”, Budapest, 2022. Elérés: 2023. március 2. [Online]. Elérhető: <https://kormany.hu/dokumentumtar/nemzeti-digitalizacios-strategia-2022-2030>
- [296] R. Zaman, „Nokia’s Collapse: Lessons in Reinvention, Disruption and Adaptation”, THE WAVES. Elérés: 2025. február 10. [Online]. Elérhető: <https://www.the-waves.org/2024/10/13/nokias-collapse-lessons-in-reinvention-disruption-and-adaptation/>
- [297] J. Sherman, „India’s Digital Path: Leaning Democratic or Authoritarian?”, Just Security. Elérés: 2022. november 19. [Online]. Elérhető: <https://www.justsecurity.org/62464/indias-digital-path-leaning-democratic-authoritarian/>
- [298] E. Yayboke és S. Brannen, „Promote and Build: A Strategic Approach to Digital Authoritarianism”, Center for Strategic and International Studies (CSIS), Washington, 2020. Elérés: 2022. december 24. [Online]. Elérhető: <https://www.csis.org/analysis/promote-and-build-strategic-approach-digital-authoritarianism>
- [299] R. Morgus, „The Spread of Russia’s Digital Authoritarianism”, in *Artificial Intelligence, China, Russia, and the Global Order*, Air University Press, 2019, o. 89–97. Elérés: 2023. március 17. [Online]. Elérhető: <https://www.jstor.org/stable/resrep19585.17>
- [300] C. Meserole és A. Polyakova, „Exporting digital authoritarianism -The Russian and Chinese models”, Brookings. Elérés: 2022. november 18. [Online]. Elérhető: https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190827_digital_authoritarianism_polyakova_meserole.pdf
- [301] R. Gigova, „Who Vladimir Putin thinks will rule the world”, *CNN*, szept. 2017, Elérés: 2022. november 22. [Online]. Elérhető: <https://www.cnn.com/2017/09/01/world/putin-artificial-intelligence-will-rule-world/index.html>
- [302] K. Emelin, „Artificial Intelligence is Under Control – Комиссия Российской Федерации по делам ЮНЕСКО”, UNESCO. Elérés: 2023. március 29. [Online]. Elérhető: <http://unesco.ru/en/news/ai-committee/>
- [303] A. Soldatov és I. Borogan, „5 Russian-Made Surveillance Technologies Used in the West”, *Wired*. Elérés: 2023. március 17. [Online]. Elérhető: <https://www.wired.com/2013/05/russian-surveillance-technologies/>
- [304] A. Zakharov, „Злость, страх и силуэты. Мэрия Москвы раскрыла, какие алгоритмы распознают людей по лицам”, *BBC News Русская служба*. Elérés: 2022. november 27. [Online]. Elérhető: <https://www.bbc.com/russian/features-62658404>
- [305] Statista elemző oldal, „CCTV camera density by major city Russia 2022”, Statista. Elérés: 2022. november 27. [Online]. Elérhető: <https://www.statista.com/statistics/1156026/surveillance-cameras-density-moscow-st-petersburg/>
- [306] S. Bendett, „Russia’s Artificial Intelligence Boom May Not Survive the War”. Elérés: 2022. november 19. [Online]. Elérhető: <https://www.defenseone.com/ideas/2022/04/russias-artificial-intelligence-boom-may-not-survive-war/365743/>
- [307] S. Bendett, „Russia’s AI Quest is State-Driven — Even More than China’s. Can It Work?”, *Defense One*. Elérés: 2022. november 19. [Online]. Elérhető: <https://www.defenseone.com/ideas/2019/11/russias-ai-quest-state-driven-even-more-chinas-can-it-work/161519/>
- [308] D. Palavenis, „The Use of Emerging Disruptive Technologies by the Russian Armed Forces in the Ukrainian W”, *Air Land Sea Space Application (ALSSA) Center*. Elérés: 2023. március 29. [Online].

- Elérhető: https://www.alsa.mil/Portals/9/Documents/articles/221001_ALSA_Article_Donatas_Palavenis.pdf
- [309] RosBusinessConsulting híroldal, „В России протестировали работу Рунета при отключении от глобальной Сети”, РБК. Elérés: 2022. november 26. [Online]. Elérhető: https://www.rbc.ru/technology_and_media/21/07/2021/60f8134c9a79476f5de1d739
- [310] Pravda.ru, „Российский рунет выдержал первые испытания”, pravda.ru. Elérés: 2022. november 26. [Online]. Elérhető: <https://www.pravda.ru/news/politics/1461663-runet/>
- [311] „Russia slows down Twitter over »banned content«”, *BBC News*, 2021. március 10. Elérés: 2022. november 19. [Online]. Elérhető: <https://www.bbc.com/news/world-europe-56344304>
- [312] J. Rose, „Russia’s Digital Iron Curtain Is Starting To Take Shape”, *Vice*, 2022. március 17. Elérés: 2022. november 26. [Online]. Elérhető: <https://www.vice.com/en/article/5dg4kb/russias-digital-iron-curtain-is-starting-to-take-shape>
- [313] E. Sinkkonen és J. Lassila, „Digital Authoritarianism and Technological Cooperation in Sino-Russian Relations: Common Goals and Diverging Standpoints”, in *Russia-China Relations: Emerging Alliance or Eternal Rivals?*, S. Kirchberger, S. Sinjen, és N. Wörmer, Szerk., in *Global Power Shift*, Cham: Springer International Publishing, 2022, o. 165–184. doi: 10.1007/978-3-030-97012-3_9.
- [314] J. Lalljee, „Russia’s economy already lost \$860 million this year because the government keeps shutting down the internet”, *Business Insider*, 2022. március 19. Elérés: 2023. március 22. [Online]. Elérhető: <https://www.businessinsider.com/russia-internet-censorship-cost-economy-putin-ukraine-sanctions-twitter-2022-3>
- [315] TASS hírügynökség, „Russia won’t be left without Internet, guarantees Lavrov”, TASS. Elérés: 2022. november 26. [Online]. Elérhető: <https://tass.com/politics/1452143>
- [316] Freedom House civil szervezet, „Russia: Freedom on the Net 2022 Country Report”, Freedom House. Elérés: 2022. november 26. [Online]. Elérhető: <https://freedomhouse.org/country/russia/freedom-net/2022>
- [317] B. Toulas, „Russia creates its own TLS certificate authority to bypass sanctions”, *BleepingComputer*. Elérés: 2022. november 26. [Online]. Elérhető: <https://www.bleepingcomputer.com/news/security/russia-creates-its-own-tls-certificate-authority-to-bypass-sanctions/>
- [318] T. Claburn, „Russia bans foreign software purchases for infrastructure”, *The Register*. Elérés: 2023. március 17. [Online]. Elérhető: https://www.theregister.com/2022/04/01/russia_bans_foreign_software/
- [319] M. Hughes, „Microsoft stops serving Windows downloads to Russian users”, *KnowTechie*. Elérés: 2023. március 17. [Online]. Elérhető: <https://knowtechie.com/microsoft-stops-serving-windows-downloads-to-russian-users/>
- [320] L. Howells és L. A. Henry, „Digital Authoritarianism at War: Controlling Russia’s Information Space”, NYU Jordan Center. Elérés: 2022. november 27. [Online]. Elérhető: <https://jordanusiacenter.org/news/digital-authoritarianism-at-war-controlling-russias-information-space/>
- [321] L. Harding, „Twitter launches privacy-protected site on dark web to bypass Russia’s block”, *The Guardian*, 2022. március 10. Elérés: 2022. november 18. [Online]. Elérhető: <https://www.theguardian.com/technology/2022/mar/09/twitter-tor-version-russia-block>
- [322] „HPC projektigénylés | Hpc.kifu.hu”. Elérés: 2024. október 14. [Online]. Elérhető: <https://hpc.kifu.hu/hu/hpc-projekt-igenyles>
- [323] Chapple A., „Rajok háborúja: a mesterséges intelligencia által vezérelt drónok bizonytalan tényszerése Ukrajnában”, *Szabad Európa*. Elérés: 2024. november 21. [Online]. Elérhető: <https://www.szabadeuropa.hu/a/mesterseges-intelligencia-altal-vezereelt-dronok-bizonytalan-tenyszerese-ukrajnaban/33081053.html>
- [324] United Nations, „Seventy-ninth Session, 30th Meeting”, *press.un.org*. Elérés: 2024. november 21. [Online]. Elérhető: <https://press.un.org/en/2024/gadis3756.doc.htm>
- [325] D. Myles, „Industry 4.0 sparks new era of cyber attacks”, *FDI Intelligence*. Elérés: 2024. november 21. [Online]. Elérhető: <https://www.fdiintelligence.com/content/interview/industry-40-sparks-new-era-of-cyber-attacks-83009>

- [326] J. M. McGiffin, „Mission (Command) Complete: Implications of JADC2”, *National Defense University Press*, sz. 2024, o. July, 2024.
- [327] K. Kassai, „A kibertér műveleti képesség szerepének, jelentőségének és fókuszának evolúciója a NATO stratégiai dokumentumai alapján”, *Military and Intelligence CyberSecurity Research Paper*, köt. 2022, sz. 9, 2022, Elérés: 2025. február 13. [Online]. Elérhető: <https://tudokk.mtak.hu/index.php?a=get&id=836731&pattern=&t=A%20kibert%C3%A9r%20m%C5%B1veleti%20k%C3%A9pess%C3%A9g%20szerep%C3%A9nek%2C%20jelent%C5%91s%C3%A9g%C3%A9nek%20%C3%A9s%20f%C3%B3kus%C3%A1nak%20evol%C3%BAci%C3%B3ja%20a%20NATO%20strat%C3%A9giai%20dokumentumai%20alapj%C3%A1n>
- [328] Fehér A. T. és Négyesi I., „A digitális visszaélések katonai vonatkozásai és az orosz modell”, *Honvédségi Szemle – Hungarian Defence Review*, köt. 151, sz. 5, Art. sz. 5, nov. 2023, doi: 10.35926/HSZ.2023.5.1.
- [329] B. Claverie és F. Du Cluzel, „The Cognitive Warfare Concept”, *Innovation Hub Sponsored by NATO Allied Command Transformation*, köt. 2, 2022.